

REDUÇÃO DE RETICULADOS CONCEITUAIS

SÉRGIO MARIANO DIAS

REDUÇÃO DE RETICULADOS CONCEITUAIS

Tese apresentada ao Programa de Pós-Graduação em Ciência da Computação do Instituto de Ciências Exatas da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Doutor em Ciência da Computação.

ORIENTADOR: NEWTON JOSÉ VIEIRA

Belo Horizonte

Maio de 2016

© 2016, Sérgio Mariano Dias.
Todos os direitos reservados.

Ficha catalográfica elaborada pela Biblioteca do ICEx - UFMG

Dias, Sérgio Mariano

D541r Redução de Reticulados Conceituais / Sérgio Mariano
Dias. — Belo Horizonte, 2016
xxiii, 151 f. : il. ; 29cm

Tese (doutorado) — Universidade Federal de Minas
Gerais — Departamento de Ciência da Computação.

Orientador: Newton José Vieira

1. Computação — Teses. 2. Análise formal de conceitos
3. Teoria dos reticulados.. I. Orientador. II. Título.

CDU 519.6*54 (043)



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

FOLHA DE APROVAÇÃO

Redução de reticulados conceituais

SÉRGIO MARIANO DIAS

Tese defendida e aprovada pela banca examinadora constituída pelos Senhores:

Newton José Vieira

PROF. NEWTON JOSÉ VIEIRA - Orientador
Departamento de Ciência da Computação - UFMG

Henrique Pacca Loureiro Luna

PROF. HENRIQUE PACCA LOUREIRO LUNA
Instituto de Computação - UFAL

P & P

PROF. LOIC PASCAL GILLES CERF
Departamento Ciência da Computação - UFMG

Luís Enrique Zárate

PROF. LUÍS ENRIQUE ZÁRATE
Departamento de Ciência da Computação - PUC-MG

Wagner Meira Júnior

PROF. WAGNER MEIRA JÚNIOR
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 16 de maio de 2016.

Aos meus pais Dionisio e Delfna.

Agradecimentos

Primeiramente gostaria de agradecer ao professor Newton José Vieira pela orientação. Seu apoio foi fundamental para minha formação.

Agradeço ao Serviço Federal de Processamento de Dados (SERPRO) pela liberação formal de vinte horas semanais para realização do doutorado. Em especial aos colegas de trabalho José Maurilio Gonçalves Dias e Carlos Alberto Moreira Guedes. Ambos foram responsáveis por aprovar e apoiar minha participação no programa de pós-graduação do SERPRO.

Agradeço também ao professor Luiz Enrique Zárate pela oportunidade de coorientar os alunos da graduação em ciência da computação e do mestrado em informática da Pontifícia Universidade Católica de Minas Gerais em trabalhos relacionados à análise formal de conceitos.

Agradeço ainda aos colegas do SERPRO e do doutorado pelo apoio, em especial: Gustavo da Gama Torre, Isabel Barbosa, Marcos Carrijo e Marcelo Pita.

“Machines take me by surprise with great frequency.”
(Alan Turing)

Resumo

Um formalismo importante para representação, extração e análise de conhecimento é proporcionado pela chamada análise formal de conceitos (AFC), uma abordagem baseada na matematização da noção de conceito, bem como na organização dos conceitos em uma hierarquia conceitual.

Na AFC, a complexidade do reticulado conceitual, em função do número de conceitos formais e/ou cardinalidade da relação de cobertura, vem a se mostrar um dos problemas mais significativos. O fato de que todos os relacionamentos entre os conceitos abstraídos do contexto formal estejam presentes no reticulado é apropriado em termos de completude, mas resulta, em geral, num grande volume de relacionamentos, sobrecarregando em demasia o reticulado. Na verdade, a AFC induz a uma complexidade combinatória potencialmente alta, e as estruturas obtidas, até mesmo de um pequeno conjunto de dados podem tornar-se proibitivamente grandes. Em particular, os aspectos essenciais, aqueles que efetivamente são procurados, podem estar imersos em um emaranhado de detalhes menos importantes.

Nesta tese, quarenta técnicas de redução de reticulados conceituais, selecionadas dentre as mais representativas, são analisadas e divididas em três classes. A análise é feita com base em sete dimensões, cada uma delas consistindo de um conjunto de características. Efetuando-se a análise através da própria AFC, considerações são realizadas acerca de complexidade computacional, viabilidade e qualidade do reticulado conceitual resultante.

A análise das principais técnicas mostrou uma lacuna em técnicas com a capacidade de abstrair e generalizar o conhecimento expresso por um contexto formal. Dada essa constatação, propõe-se aqui uma técnica de redução com tais capacidades, e que não exige, ao contrário de outras técnicas, a computação dos conceitos formais do reti-

culado conceitual original. A técnica, que tem uma complexidade computacional satisfatória, substitui grupos de objetos similares por objetos representativos, sendo a similaridade medida com base nas relevâncias dos atributos.

O estudo das técnicas existentes mostrou também a ausência de uma metodologia para análise das mesmas que fosse independente de suas características e do domínio de aplicação que se tivesse em mente. Propõe-se aqui uma metodologia de análise independente de aplicação. Ela é baseada no uso de implicações como expressão do conhecimento retratado por um contexto formal. A metodologia permite a identificação do conhecimento que é preservado, eliminado, inserido e/ou transformado por uma técnica de redução. Quatro índices complementares são indicados para integrar a metodologia.

Palavras-chave: Análise Formal de Conceitos, Reticulados Conceituais, Redução, Implicações.

Abstract

An important formalism for knowledge representation, extraction and analysis is provided by the so-called formal concept analysis (FCA), an approach based on the mathematization of the notion of concept and on the organization of concepts in a conceptual hierarchy.

In FCA, the complexity of the concept lattice, as a function of the number of formal concepts and/or cardinality of the cover relation, is one of the most significant problems. The fact that all relationships between the concepts extracted from a formal context are present in the concept lattice is appropriate in terms of completeness, but generally results in a large number of relationships, thus overloading too much the lattice. In fact, FCA induces a potentially high combinatorial complexity, and the structures obtained, even from a small dataset, may become prohibitively large. In particular, key aspects, those that are indeed important, may be immersed in a maze of less relevant details.

In this thesis, forty reduction techniques of concept lattices, selected from the most representative, are analyzed and divided into three classes. The analysis is based on seven dimensions, each consisting of a set of characteristics. Performing an analysis through FCA itself, considerations are made about computational complexity, feasibility and quality of the resulting concept lattice.

The analysis of the main techniques showed a gap in techniques with the ability to abstract and generalize the knowledge expressed by a formal context. Given this finding, it is proposed here a reduction technique with such capabilities, and which does not require, unlike other techniques, the computation of the formal concepts of the original concept lattice. The technique, which has a satisfactory computational complexity, replaces groups of similar objects by representative objects, the similarity being measured

on the basis of the relevance of attributes.

The study of existing techniques also showed the absence of a methodology for analysis which was independent of its characteristics and intended application domain. It is proposed here an application independent methodology of analysis. It is based on the use of implications as an alternative expression of the knowledge portrayed by a formal context. The methodology allows the identification of which knowledge is preserved, deleted, inserted and/or transformed by a reduction technique. Four complementary indexes are indicated to integrate the methodology.

Keywords: Formal Concept Analysis, Concept Lattices, Reduction, Implications.

Lista de Figuras

2.1	Diagrama de linhas do conjunto potência de {a,b,c}.	15
2.2	Reticulado conceitual originado do contexto formal apresentado na Tabela 2.1.	24
3.1	Reticulado conceitual do contexto formal da Tabela 3.1.	45
3.2	Reticulado conceitual para cada classe.	46
4.1	Reticulados original e reduzido obtido pela técnica JBOS.	58
4.2	Exemplo de laminador utilizado no processo siderúrgico de laminação a frio. Fonte: [178]	60
4.3	Mapeando a relação entre carga (P) e espessura de saída (ho) através de implicações próprias.	67
4.4	Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias.	68
4.5	Reduções do número de conceitos formais em função dos parâmetros ε e α para a técnica JBOS utilizando o contexto formal #1.	70
4.6	Mapeando a relação entre carga (P) e espessura de saída (ho) através de implicações próprias obtidas após redução pela técnica JBOS.	72
4.7	Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica JBOS.	73
4.8	Reduções do número de conceitos formais em função de uma estabilidade mínima para o contexto #1.	73
4.9	Mapeando a relação entre carga (P) e espessura de saída (ho) através de implicações próprias obtidas após redução pela técnica de estabilidade.	75

4.10	Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica de estabilidade. . . .	76
4.11	Reduções do número de conceitos formais em função do suporte mínimo para a técnica baseada em conceitos formais frequentes para o contexto formal #1.	76
4.12	Mapeando a relação entre carga (P) e espessura de saída (ho) através de implicações próprias obtidas após redução pela técnica de conceitos formais frequentes.	78
4.13	Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica de conceitos formais frequentes.	79
4.14	Reduções do número de conceitos formais em função da relevância mínima para a técnica de redução baseada em geradores mínimos utilizando o contexto formal #1.	80
4.15	Mapeando a relação entre carga (P) e espessura de saída (ho) através de implicações próprias obtidas após redução pela técnica de seleção baseada em geradores mínimos.	81
4.16	Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica de seleção baseada em geradores mínimos.	82
5.1	Estrutura de reduções através de implicações.	92
5.2	Modelos básicos de redução.	94
5.3	Conceitos c_1 e c_2 , tais que $c_1 \prec c_2$	104
5.4	Reticulados conceituais utilizados para análise das técnicas de redução. . .	106
5.5	Reticulado conceitual animais reduzido pela técnica JBOS.	112

Lista de Tabelas

2.1	Exemplo de contexto formal.	19
2.2	Contexto formal obtido a partir de contexto formal multivalorado.	21
3.1	Contexto formal para as principais técnicas de redução.	44
3.2	Implicações com condições suficientes e concisas para cada classe.	50
4.1	Contexto formal reduzido obtido pela técnica JBOS.	58
4.2	Valores máximos e mínimos para cada parâmetro do processo de laminação.	61
4.3	Discretização aplicada no parâmetro carga.	63
4.4	Resumo dos resultados relativos às implicações próprias dos contextos de validação originais.	65
4.5	Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica JBOS.	71
4.6	Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica baseada no índice de estabilidade.	74
4.7	Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica de conceitos formais frequentes	77
4.8	Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica de geradores mínimos.	80
5.1	Características das três bases de dados.	107
5.2	Implicações próprias para o reticulado animais.	107
5.3	Conjunto de implicações próprias para o reticulado tuberculose.	108
5.4	Conjunto de implicações próprias para o reticulado laminação.	108

5.5	Transformação do conhecimento utilizando seleção do conjunto minimal de atributos para o reticulado animais.	110
5.6	Transformação do conhecimento utilizando seleção do conjunto minimal de atributos para os reticulados tuberculose e laminação.	111
5.7	Transformação do conhecimento utilizando a técnica JBOS para o reticulado animais.	113
5.8	Transformação do conhecimento utilizando a técnica JBOS para os reticulados tuberculose e laminação.	115
5.9	Transformação do conhecimento utilizando a seleção de conceitos formais baseada no índice de estabilidade para o reticulado animais.	116
5.10	Transformação do conhecimento utilizando a técnica de seleção de conceitos baseada no índice de estabilidade para os reticulados tuberculose e laminação.	118
5.11	Conhecimento prévio aplicado ao contexto formal do reticulado animais. . .	119
5.12	Transformação do conhecimento utilizando seleção de conceitos formais baseada em restrições determinadas a partir dos dados.	120
5.13	Transformação do conhecimento utilizando restrições determinadas a partir dos dados para os reticulados tuberculose e laminação.	121

Sumário

Agradecimentos	ix
Resumo	xiii
Abstract	xv
Lista de Figuras	xvii
Lista de Tabelas	xix
1 Introdução	3
1.1 Motivação	5
1.2 Objetivos	6
1.3 Metodologia	7
1.4 Contribuições	8
1.5 Organização da Tese	10
2 Análise Formal de Conceitos: Conceitos Fundamentais	13
2.1 Ordens Parciais	13
2.1.1 Conjuntos Ordenados	13
2.1.2 Diagrama de Linhas	14
2.2 Sistemas de Fecho	16
2.3 Análise Formal de Conceitos	18
2.3.1 Contextos Formais	18
2.3.1.1 Contextos Formais Multivalorados	20
2.3.2 Conceitos Formais	21

2.3.3	Reticulados Conceituais	22
2.3.3.1	Complexidade do Reticulado Conceitual	23
2.3.4	Regras	25
2.4	Conclusões	27
3	Redução de Reticulados Conceituais: Características e Classificação	29
3.1	Classificação	30
3.1.1	Eliminação de Informação Redundante	31
3.1.2	Simplificação	34
3.1.3	Seleção	37
3.2	Dimensões para Análise das Técnicas de Redução	42
3.2.1	Analisando as Técnicas de Redução através da AFC	44
3.3	Trabalhos Relacionados	50
3.4	Conclusões	51
4	Redução Baseada na Similaridade entre Objetos	53
4.1	Alguns pressupostos	54
4.2	Junção Baseada na Similaridade de Objetos - JBOS	55
4.3	Estudo de Caso	59
4.3.1	Processo Siderúrgico de Laminação a Frio	59
4.3.2	Dados para Análise do Processo de Laminação	60
4.3.3	Representando o Processo de Laminação através da AFC	62
4.3.3.1	Extraindo Conhecimento através da AFC	63
4.3.4	Aplicando Técnicas de Redução	67
4.3.4.1	Agrupamento Baseado na Técnica - JBOS	68
4.3.4.2	Estabilidade	71
4.3.4.3	Conceitos Formais Frequentes	75
4.3.4.4	Geradores Mínimos	78
4.4	Conclusões	83
5	Análise da Qualidade de Reduções	85
5.1	Qualidade das Reduções	85
5.2	Análise de Reduções através de Implicações	88
5.2.1	Características de cada Classe de Redução	94

5.3	Qualidade de um Conjunto de Implicações	96
5.3.1	Conteúdo de Informação	97
5.3.2	Fidelidade e Representatividade	100
5.4	Perda Descritiva	101
5.5	Extração de Implicações de um Conjunto de Conceitos Formais	102
5.6	Estudo de Caso	105
5.6.1	Seleção do Conjunto Minimal de Atributos	109
5.6.2	Agrupamento pela Técnica JBOS	111
5.6.3	Seleção de Conceitos Baseada no Índice de Estabilidade	114
5.6.4	Restrições Determinadas a partir dos Dados	117
5.7	Conclusões	121
6	Conclusões	123
6.1	Contribuições	123
6.2	Publicações e Prêmios	127
6.3	Recomendações para Trabalhos Futuros	127
	Anexo A Demonstrações das Proposições	129
	Referências Bibliográficas	133

Lista de Símbolos

- M - conjunto de atributos
- G - conjunto de objetos
- I - conjunto de incidências ou relação de incidências
- $\mathcal{B}(G, M, I)$ ou L - reticulado conceitual
- \mathcal{I} - conjunto de implicações
- Pr - conhecimento preservado
- El - conhecimento eliminado
- In - conhecimento novo
- Ge - conhecimento generalizado
- Sp - conhecimento especializado
- \mathcal{I}_p - implicações preservadas
- \mathcal{I}_n - implicações novas
- \mathcal{I}_e - implicações eliminadas
- \mathcal{I}^s - implicações especializadas em \mathcal{I}
- \mathcal{I}^g - implicações generalizadas em \mathcal{I}
- \mathcal{I}_r^g - implicações generalizadas em \mathcal{I}_r

- \mathcal{I}_r^s - implicações especializadas em \mathcal{I}_r
- F - fidelidade
- R - representatividade
- PD - perda descritiva
- P - carga ou pressão exercida pelos cilindros
- h_i - espessura de entrada
- h_o - espessura de saída
- μ - coeficiente de atrito
- tb - tensão à ré
- tf - tensão à frente
- \bar{y} - tensão de escoamento
- E - módulo de Young da tira
- R_c - raio do cilindro

Capítulo 1

Introdução

A capacidade de gerar e armazenar dados, observada nos últimos anos, seria inimaginável há algum tempo atrás. Por exemplo, dados provenientes de interações sociais, governo eletrônico, atividades empresariais nos mais diversos setores, conhecimento científico em biologia, química, etc., estão sendo gerados continuamente, e em enorme quantidade, todos os dias. Existe um volume tão grande de dados que é cada vez mais difícil localizar e extrair o conhecimento do qual se necessita. Além disso, esses dados, muitas vezes, demandam transformações antes da extração, propriamente dita, de conhecimento. Por outro lado, os processos de representação, extração e análise de conhecimento demandam um formalismo que proporcione, além da expressividade necessária, um posterior processamento computacional com desempenho satisfatório.

Um formalismo importante para representação, extração e análise de conhecimento, introduzido em 1982 por Wille [168], é aquele propiciado pela análise formal de conceitos (AFC)¹, que é uma abordagem baseada na matematização da noção de conceito e na organização dos conceitos em uma hierarquia conceitual na forma de um reticulado [68].

Os dados iniciais para a AFC se provêm sob a forma de uma relação binária, que envolve um conjunto de objetos e um conjunto de atributos a que se denomina *contexto formal*. O grande potencial da AFC é propiciado pela organização do conhecimento presente num contexto formal em uma hierarquia, denominada *reticulado conceitual*,

¹Do inglês, *formal concept analysis*. Será utilizado no texto análise formal de conceitos, análise formal ou simplesmente AFC.

constituída de um conjunto parcialmente ordenado de *conceitos formais*. Um corpo de conhecimento possível de se obter a partir do contexto formal ou do reticulado conceitual é o conjunto de *implicações* entre os atributos que caracterizam o contexto formal. As principais aplicações da AFC fazem uso, de forma não exclusiva, do contexto formal, de conceitos formais, reticulado conceitual e implicações.

Desde sua concepção, a análise formal de conceitos vem sendo aplicada em diferentes áreas, como recuperação de informação, representação e descoberta de conhecimento, ontologia, mineração de dados, inteligência artificial, etc. Por exemplo, em recuperação de informação, a AFC foi inicialmente empregada para refinamento de consultas, visualização e navegação [30, 72, 90]. Recentemente, observam-se interessantes trabalhos em criação de ranques de páginas web [52], web semântica e ontologia [58–60]. O uso da análise formal em recuperação de informação foi se alterando ao longo do tempo, observando-se desde trabalhos para visualização de documentos até aplicações relativas à web semântica. Um resumo das principais aplicações da AFC em recuperação de informação é apresentado em Codocedo & Napoli [35].

No contexto de representação e descoberta de conhecimento, Wille [169] identifica 10 funções de processamento de conhecimento (explorar, pesquisar, reconhecer, identificar, analisar, investigar, decidir, melhorar, reestruturar e memorizar) e investiga como estas são suportados pela AFC. Wille [170] descreve como a análise formal suporta a descoberta de conhecimento em bases de dados. Ganter [63] e Stumme [147] demonstram a extração de conhecimento através da exploração de atributos. Ainda relacionados à representação e descoberta de conhecimento, observam-se diversos trabalhos de raciocínio baseado em casos [48, 118, 124, 150], os quais empregam a análise formal de conceitos como arcabouço principal.

Na área de mineração de dados, observam-se trabalhos em classificação [14, 50, 133, 140, 154, 171], regras de associação [122, 158, 176, 177], árvores de decisão [11], mineração de texto [111], etc. Existe uma grande interdisciplinaridade entre as áreas. Em geral, mineração de dados faz uso do arcabouço matemático da AFC para conceber novas soluções.

Em inteligência artificial e lógica, também são observados diversos trabalhos [31, 131, 142, 154, 157, 182]. Em especial, no contexto de redes neurais artificiais, encontram-se trabalhos que procuram representar o conhecimento adquirido pela rede mediante regras extraídas a partir de reticulados conceituais [46, 181, 182].

Stumme [148], já em 2002, apontava que nos anos anteriores a AFC vinha migrando cada vez mais de pesquisas teóricas, apresentadas em congressos matemáticos, para aplicações práticas, expostas principalmente em congressos da área de computação. De lá para cá essa tendência prosseguiu com bastante intensidade. Poelmans et al. [128, 129] e Kuznetsov & Poelmans [105] apresentam resumos das principais aplicações da AFC nos últimos anos.

1.1 Motivação

Na análise formal de conceitos, a complexidade do reticulado conceitual em função do número de conceitos formais e/ou da cardinalidade da relação de cobertura mostra-se um dos problemas mais importantes [91]. O fato de que todos os relacionamentos existentes entre os conceitos abstraídos do contexto formal estejam presentes no reticulado é apropriado em termos de completude, mas, em geral, resulta num grande volume de relacionamentos, sobrecarregando demais o reticulado. Na verdade, a AFC induz uma complexidade combinatória potencialmente alta e as estruturas obtidas até mesmo de um pequeno conjunto de dados podem se tornar proibitivamente grandes [89]. Apesar de o pior caso ser raramente encontrado em casos práticos [73], o custo computacional ainda pode ser proibitivo para muitas aplicações. Além disso, o número resultante de conceitos formais e a hierarquia conceitual obtida podem tornar difícil a análise subjacente [137]. Em particular, os aspectos essenciais, aqueles que efetivamente são procurados, podem estar imersos em um emaranhado de detalhes irrelevantes. Ademais, o reticulado pode conter conceitos muito semelhantes entre si devido a pequenas diferenças irrelevantes para a aplicação em mente [138].

De fato, o problema de se obter um reticulado conceitual de complexidade e tamanho apropriados, e que exponha os aspectos relevantes, constitui um dos problemas mais relevantes da análise formal de conceitos [13, 18, 19, 102, 116, 137, 149, 185]. Note-se, ainda, que o problema da geração de um expressivo número de padrões consiste num problema mais geral, enfrentado em muitos métodos de análise de dados que procuram certos tipos de padrões nos mesmos [13]. Felizmente, na vida real uma gama de dados desnecessários pode vir a ser detectado como tal, sendo possível, e vital, descartá-los [126].

Para fazer frente aos problemas assinalados pelo excesso de informação pre-

sente nos reticulados conceituais, trabalhos com diferentes características são observados. Determinadas técnicas procuram eliminar informações consideradas redundantes. Elas, em geral, têm como objetivo encontrar o número mínimo de objetos ou atributos capaz de manter a estrutura do reticulado conceitual inalterada [108, 116, 126, 161, 163, 164, 166, 185]. Outras técnicas procuram obter uma abstração do reticulado conceitual, em outras palavras, um resumo em alto nível que exponha apenas os aspectos considerados essenciais do mesmo [19, 34, 36, 40, 62, 98, 130, 144]. Por último, algumas outras técnicas atuam selecionando conceitos formais, objetos ou atributos, com base em algum critério de relevância [4, 70, 99, 136, 137, 149].

Esta tese aborda os problemas decorrentes do excesso de informação presente nos reticulados conceituais, através da redução destes.

1.2 Objetivos

O tema central desta tese consiste na redução de reticulados conceituais de forma a preservar e/ou expor as informações mais relevantes em determinado contexto ou aplicação. A investigação aqui relatada teve como objetivos principais dentro desse tema:

1. Analisar e classificar as principais técnicas para redução de reticulados conceituais a partir de características bem definidas abstraídas das técnicas existentes.
2. Propor uma técnica para redução de reticulados conceituais capaz de abstrair o excesso de conhecimento presente nos reticulados, que não necessite de uma pré-computação dos conceitos formais, que possua um desempenho computacional satisfatório, e que leve, necessariamente, a um reticulado conceitual reduzido.
3. Propor uma metodologia independente de aplicação para análise de reticulados conceituais reduzidos, a qual permita identificar claramente os conhecimentos preservado, eliminado, inserido ou transformado pela técnica de redução, bem como índices para mensurar os conhecimentos original e reduzido.

Técnicas com diferentes características para redução de reticulados conceituais são observadas na literatura; entretanto, até o presente momento elas ainda não foram analisadas, comparadas nem classificadas, de forma satisfatória. Assim, o primeiro

objetivo é apresentar um estudo, análise, comparação e classificação das principais técnicas de redução.

A análise das principais técnicas para redução de reticulados conceituais demonstrou que existe uma concentração na concepção de técnicas para seleção de conceitos formais. Uma desvantagem dessas técnicas reside na incapacidade de preservar as propriedades do reticulado conceitual; i.e., o subconjunto de conceitos formais selecionados pode não formar um reticulado conceitual [138]; nesse caso, pode ser necessário um pós-processamento para formar um reticulado conceitual [84, 184]. Além disso, determinadas técnicas necessitam do reticulado conceitual completo para seleção dos conceitos formais considerados relevantes; característica que, em muitos casos, impede o uso da técnica. Frequentemente, não é possível gerar todos os conceitos formais, seja pela dimensão tempo, espaço ou ambos. Observadas essas lacunas, o segundo objetivo é propor uma técnica para redução de reticulados conceituais capaz de gerar um reticulado conceitual reduzido (sem a necessidade de um pós-processamento), que abstraia o excesso de conhecimento presente no reticulado original, possua desempenho computacional satisfatório, e, ainda, não faça uso dos conceitos formais (reticulado conceitual original).

Determinadas técnicas para redução de reticulados conceituais não vêm acompanhadas de índices que demonstrem a qualidade do reticulado resultante, ficando a qualidade dos resultados obtidos a ser interpretada de forma um tanto subjetiva. Por outro lado, algumas técnicas apresentam índices específicos (indicados apenas para a técnica em análise), esses índices sendo incompletos por não identificarem todas as características do reticulado reduzido, mesmo quando levada em conta apenas a técnica em análise. Desta forma, o terceiro objetivo é propor uma metodologia que permita, independentemente da aplicação, analisar comparativamente reticulados conceituais identificando em que grau o conhecimento é preservado, eliminado, inserido e/ou transformado pela técnica de redução, além de propor índices complementares para mensurar os conhecimentos original e reduzido.

1.3 Metodologia

Os objetivos definidos foram alcançados através da seguinte metodologia:

1. Um total de quarenta técnicas, selecionadas entre as principais da literatura, foram analisadas e classificadas com base em sete dimensões: ponto de partida, uso de conhecimento prévio, alterações no contexto formal, reticulado conceitual resultante, índices de qualidade, aspectos algorítmicos, e a classe a que a técnica pertence. Um contexto formal foi construído resumindo os resultados e uma análise foi conduzida utilizando a própria análise formal de conceitos.
2. Foi desenvolvida uma técnica de redução de reticulados conceituais denominada JBOS (*junction based on objects similarity*), que se trata de uma evolução do trabalho de mestrado apresentado por Dias [39]. A técnica permite gerar um reticulado conceitual reduzido, sem a necessidade de um pós-processamento. Ela provê um meio de abstrair o excesso de conhecimento presente no reticulado original por meio de generalizações, possui desempenho computacional satisfatório e não faz uso dos conceitos formais do reticulado conceitual original. A técnica foi classificada juntamente com as demais existentes na literatura. A técnica JBOS e três outras técnicas clássicas foram utilizadas em uma aplicação real.
3. Uma metodologia independente de aplicação foi proposta. Essa metodologia é fundamentada na análise dos conjuntos de implicações originados dos contextos formais (reticulados conceituais) original e reduzido. A visão alternativa do conhecimento expresso através de implicações é utilizada para evidenciar os conhecimentos preservado, eliminado, inserido e/ou transformado pela técnica de redução. Quatro índices foram indicados para mensurar os conhecimentos original e reduzido. A metodologia e os índices foram utilizados para analisar quatro técnicas de redução aplicadas a três contextos formais.

1.4 Contribuições

A complexidade inerente à AFC, a qual justifica o uso de técnicas de redução de reticulados conceituais, foi identificada através dos algoritmos clássicos da área, os resultados encontrando-se publicados em Dias & Vieira [41]. As contribuições dessa tese estão resumidas a seguir.

1. Análise e classificação de técnicas para redução de reticulados conceituais.

As principais técnicas existentes para redução de reticulados conceituais são divididas em três classes. A análise das técnicas é elaborada empregando sete dimensões, consistindo, cada dimensão, num conjunto de características. Analisando-se as principais técnicas para redução, através da análise formal de conceitos, considerações são realizadas sobre complexidade computacional, viabilidade e qualidade do reticulado conceitual resultante do processo de redução. Apesar da grande relevância, até onde se tem conhecimento não são encontrados trabalhos satisfatórios dessa natureza na literatura da AFC. Essa contribuição é relevante por prover um conjunto de características e uma classificação a partir dos quais se poderá situar qualquer técnica nova ou já existente dentro do quadro geral dos métodos de redução. Os resultados relativos à análise e classificação de reticulados conceituais foram publicados em Dias & Vieira [44].

2. Proposta de uma técnica para redução de reticulados conceituais.

A técnica para redução de reticulados conceituais, JBOS, inicialmente discutida na dissertação de Dias [39] é avaliada e comparada com o estado da arte. A técnica JBOS substitui grupos de objetos similares por objetos representativos. A similaridade é mensurada baseando-se, não apenas na presença e/ou ausência dos atributos conforme feito em muitos trabalhos, mas na relevância dos atributos. Embora existam outros métodos que também levam em conta a importância dos atributos, a técnica JBOS traz uma característica que pode torná-la mais adequada para muitas aplicações: o agrupamento de objetos semelhantes com base em uma medida de relevância diretamente apresentada pelo usuário. Uma segunda característica da técnica JBOS, detectada quando da análise de qualidade do reticulado obtido (vide a terceira contribuição, a seguir), é o fato de ela simplificar o reticulado conceitual generalizando o conhecimento. Por fim, a técnica apresenta um desempenho computacional satisfatório, uma vez que não necessita dos conceitos formais originais para realizar a redução. A técnica JBOS foi publicada em Dias & Vieira [40] e algumas aplicações da técnica foram apresentadas em Dias et al. [46, 47] e Dias & Vieira [42].

3. Metodologia para análise de reticulados conceituais reduzidos.

Propõe-se uma metodologia de análise independente da aplicação, que é baseada na expres-

são do conhecimento retratado pelo reticulado conceitual por meio de implicações próprias. Com ela identifica-se os conhecimentos preservado, eliminado, inserido e/ou transformado pela técnica. Proposições que fundamentam o uso do modelo são demonstradas. Para os casos em que haja acesso a um conjunto limitado de conceitos formais, apresenta-se um algoritmo, o qual, através de geradores mínimos, extrai um conjunto de implicações relativos aos conceitos. Indica-se, para mensurar o conhecimento existente nos contextos (ou reticulados) original e reduzido, quatro índices complementares: conteúdo de informação, fidelidade, representatividade e perda descritiva. A metodologia permite o entendimento do processo de redução levado a efeito por uma técnica e proporciona uma análise comparativa com outras técnicas. O índice de fidelidade foi apresentado em Dias & Vieira [40], e utilizado também em Ch. et al. [33]. O índice de representatividade foi publicado em Dias & Vieira [44]. Um artigo apresentando a metodologia proposta para análise de reticulados reduzidos encontra-se em processo de revisão [43].

1.5 Organização da Tese

O restante do texto está organizado da seguinte forma:

- Capítulo 2: *Análise Formal de Conceitos: Conceitos Fundamentais* - apresenta os fundamentos da AFC suficientes para fixar a terminologia a ser utilizada e fundamentar os principais conceitos, definições e resultados.
- Capítulo 3: *Redução de Reticulados Conceituais: Características e Classificação* – divide as técnicas existentes para redução de reticulados conceituais em três classes. Caracteriza as técnicas de redução por meio de em conjunto de dimensões proposto e aplica a própria AFC para analisar e comparar as diferentes técnicas.
- Capítulo 4: *Redução Baseada na Similaridade entre Objetos* – apresenta a técnica para redução de reticulados conceituais JBOS. Exibe um exemplo que mostra a preservação de um comportamento esperado de um processo após aplicação da técnica.

- Capítulo 5: *Análise da Qualidade de Reduções* – propõe uma metodologia, baseada na visão alternativa do conhecimento expresso através de implicações, para análise da qualidade de reduções. A técnica JBOS e outras três técnicas relevantes da literatura são analisadas utilizando-se a metodologia proposta.
- Capítulo 6: *Conclusões* – sumariza as conclusões e contribuições da tese e discute eventuais trabalhos futuros.
- Anexo A: *Demonstrações das Proposições* – apresenta demonstrações para as proposições que fundamentam a metodologia de análise desenvolvida no Capítulo 5.

Capítulo 2

Análise Formal de Conceitos: Conceitos Fundamentais

Este capítulo apresenta os principais conceitos da análise formal de conceitos (AFC). Uma vez que a AFC é fundamentada na teoria dos reticulados completos [37], os conceitos de ordens e reticulados serão primeiramente apresentados. Em seguida, a definição de sistemas de fecho será feita e os principais conceitos da AFC discutidos.

A notação e a terminologia são baseadas no excelente livro de Ganter & Wille [68]. O leitor interessado no aprofundamento dos conceitos apresentados, visto que aqui será feita apenas uma revisão suficiente para fundamentar a tese, poderá consultar os livros de Axler & Ribet [6], Davey & Priestley [37] e Ganter & Wille [68], além das referências indicadas ao longo do texto.

2.1 Ordens Parciais

A seguir, serão apresentadas as noções básicas de conjuntos ordenados e de diagramas de linhas, que são essenciais para o entendimento da AFC.

2.1.1 Conjuntos Ordenados

Primeiro, segundo, terceiro, ..., décimo. Maior, menor, igual. A noção de ordem surge, naturalmente, quando se deseja comparar objetos de um conjunto. Em termos

matemáticos, uma ordem é um tipo de relação binária. Segue a definição.

Definição 1. *Seja P um conjunto. Uma ordem (ou ordem parcial) em P é uma relação binária \leq em P tal que, para todo $x, y, z \in P$:*

- $x \leq x$;
- $x \leq y$ e $y \leq x$ implica $x = y$; e
- $x \leq y$ e $y \leq z$ implica $x \leq z$. □

As condições da Definição 1 afirmam, respectivamente, que uma relação de ordem parcial é *reflexiva*, *anti-simétrica* e *transitiva* [6, 37, 76]. Um conjunto P com uma ordem \leq , (P, \leq) , é chamado conjunto ordenado ou conjunto parcialmente ordenado¹. Se a relação \leq satisfizer uma quarta propriedade chamada *comparabilidade* (para todo $x, y \in P$, $x \leq y$ ou $y \leq x$), então ela é uma *ordem total*.

Um exemplo de *ordem total* é a relação “menor ou igual” nos números naturais. E um exemplo de *ordem parcial* (não necessariamente total) é a relação subconjunto em uma família de conjuntos. Seja X um conjunto e $\wp(X)$ o conjunto potência de X , definido por $\wp(X) = \{Y | Y \subseteq X\}$. A relação subconjunto em $\wp(X)$ tal que, para $A, B \in \wp(X)$, $A \leq B$ se e somente se $A \subseteq B$, é uma de ordem parcial. Esse tipo de ordem é utilizado na análise de conceitos formais, como será visto posteriormente.

Dada uma ordem parcial (P, \leq) , será usada a notação $x < y$ para significar que $x \leq y$ e $x \neq y$; e será usada a notação $y \geq x$ para significar o mesmo que $x \leq y$.

2.1.2 Diagrama de Linhas

Um conjunto ordenado finito P pode ser representado por meio de um desenho que descreve como os elementos de P estão relacionados. Tal desenho, denominado diagrama de linhas, é criado com base na *relação de cobertura* (P, \prec) subjacente à ordem (P, \leq) [37].

Definição 2. *Seja uma ordem parcial (P, \leq) e $x, y \in P$. Então x é coberto por y , $x \prec y$, se e somente se $x < y$ e $x \leq z < y$ implica $z = x$. □*

¹Do inglês *ordered set* ou *partially ordered set*.

Em outras palavras, $x \prec y$ se não existe elemento z entre x e y . Nesse caso, x é chamado vizinho inferior de y e y é vizinho superior de x . Veja que (P, \leq) é o fecho reflexivo e transitivo de (P, \prec) .

A relação de cobertura dos elementos de um conjunto ordenado pode ser representada por um grafo em que cada vértice é um elemento do conjunto e as arestas são os relacionamentos: há uma aresta de x para y se e somente se $x \prec y$. Com isto, $x \leq y$ se e somente se existe um caminho em tal grafo de x para y . Se, ao desenhar o grafo, cada aresta correspondente a $x \prec y$ for desenhada traçando-se uma linha unindo x e y com x abaixo de y , então se tem o que se denomina um *diagrama de linhas*². Com isto, $x \leq y$ se e somente se existe uma linha subindo de x para y (eventualmente passando por vértices intermediários).

A Figura 2.1 apresenta o diagrama de linhas do conjunto potência do conjunto $\{a, b, c\}$ ordenado pela relação *subconjunto*; observe-se que $A \prec B$ se e somente se $B = A \cup \{e\}$ para $e \in \{a, b, c\} \setminus A$. Veja-se que $A \subseteq B$ se e somente se existe uma linha subindo de A para B .

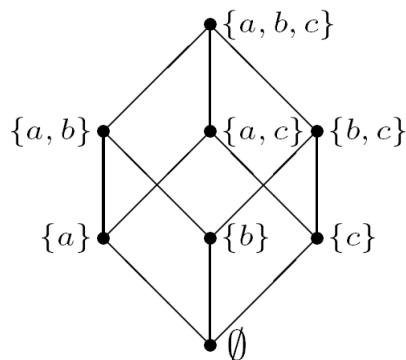


Figura 2.1. Diagrama de linhas do conjunto potência de $\{a, b, c\}$.

Associado a um conjunto ordenado, tem-se os conceitos de limite inferior e superior. Primeiramente, note-se que nem todo conjunto (P, \leq) contém um menor elemento, ou seja, um elemento x tal que $x \leq y$ para todo $y \in P$. De forma dual, ele pode não ter o maior elemento, um elemento x tal que $x \geq y$ para todo $y \in P$.

²Na literatura também é utilizado o termo *diagrama de Hasse*.

Definição 3. Seja (P, \leq) um conjunto ordenado e $S \subseteq P$. Um elemento $x \in P$ será um limite superior de S se $s \leq x$ para todo $s \in S$. De forma dual, um elemento $x \in P$ será um limite inferior de S se $x \leq s$ para todo $s \in S$. \square

O conjunto dos limites superiores de S é $S^u = \{x \in P \mid \forall s \in S: s \leq x\}$. O dual, isto é, o conjunto dos limites inferiores de S é S^l : $S^l = \{x \in P \mid \forall s \in S: x \leq s\}$. Se S^u possui um menor elemento, então esse elemento é chamado *o menor limite superior* de S . Dualmente, se S^l possui um maior elemento, então esse elemento é *o maior limite inferior* de S . O menor limite superior e o maior limite inferior também são chamados de *supremo* e *ínfimo*, respectivamente. As notações $\bigvee S$, para o supremo, e $\bigwedge S$, para o ínfimo, também são utilizadas. E no caso em que $S = \{x, y\}$, usa-se também $x \vee y$ e $x \wedge y$.

A partir dos conceitos de supremo e ínfimo, define-se o que é um reticulado completo:

Definição 4. Seja (P, \leq) um conjunto ordenado não vazio.

- Se $x \vee y$ e $x \wedge y$ existem para quaisquer $x, y \in P$, então P é um reticulado.
- Se $\bigvee S$ e $\bigwedge S$ existem para todo $S \subseteq P$, então P é dito ser um reticulado completo. \square

Note-se que o conjunto ordenado $(\wp(\{a, b, c\}), \subseteq)$, cujo diagrama de linhas está mostrado na Figura 2.1, é um reticulado completo em que as operações para determinação de ínfimo e supremo são a união e a interseção. De forma semelhante, a AFC utiliza um reticulado completo, construído sobre conjuntos fechados sob interseção, para representar a noção abstrata de conceitos e seus inter-relacionamentos.

2.2 Sistemas de Fecho

Um segundo grupo de definições importantes é o de *sistemas de fecho*, *operadores de fecho* e *conexões de Galois*. Esses conceitos são relacionados com os de operadores de derivação e de reticulado conceitual, conceitos estes fundamentais na AFC.

Sistemas de fecho³ são famílias de subconjuntos de um dado conjunto com algumas propriedades. Segue a definição formal [66].

³Em inglês, *closure systems*.

Definição 5. Um sistema de fecho do conjunto X é um subconjunto $C \subseteq \wp(X)$ que satisfaz:

- $X \in C$; e
- Se $D \subseteq C$, então $\bigcap D \in C$. □

O conjunto C é chamado sistema de fecho já que ele é fechado sob interseção, ou seja, a interseção de dois conjuntos provenientes de C sempre resulta em um conjunto de C .

Um sistema de fecho ordenado (C, \subseteq) é um reticulado completo em que o supremo e o ínfimo, para todo $D \subseteq C$, são obtidos da seguinte forma:

- $\bigvee D = \bigcap \{Y \in C \mid \bigcup D \subseteq Y\}$; e
- $\bigwedge D = \bigcap D$.

Associado ao conceito de sistemas de fecho, existe o conceito de operador de fecho.

Definição 6. Seja H um conjunto. Uma função $\varphi: \wp(H) \rightarrow \wp(H)$ é operador de fecho sobre H se para todo $A, B \subseteq H$:

- $A \subseteq \varphi(A)$;
- se $A \subseteq B$, então $\varphi(A) \subseteq \varphi(B)$; e
- $\varphi(\varphi(A)) = \varphi(A)$. □

As três propriedades na Definição 6 são chamadas de *monotonicidade*, *extensibilidade* e *idempotência*, respectivamente.

A partir de um operador de fecho, pode-se obter um reticulado. Seja φ um operador de fecho sobre um conjunto X . O conjunto $L = \{C \subseteq X \mid C = \varphi(C)\}$ ordenado pela inclusão de conjuntos é um reticulado completo em que o ínfimo e o supremo são obtidos por $\bigwedge Y = \bigcap Y$ e $\bigvee Y = \varphi(\bigcup Y)$, respectivamente [37, 67].

Um último conceito importante é o de conexão de Galois. Segue sua definição⁴.

⁴A seguir, dadas duas funções f e g , fgx abrevia $f(g(x))$.

Definição 7. Sejam (P, \leq) e (Q, \leq) dois conjuntos ordenados, $p, p_1, p_2 \in P$, $q, q_1, q_2 \in Q$ e φ, ψ dois operadores de fecho, em que $\varphi: P \rightarrow Q$ e $\psi: Q \rightarrow P$. Então, o sistema criado entre P e Q será uma conexão de Galois se:

- $p_1 \leq p_2$ implica $\varphi p_1 \geq \varphi p_2$;
- $q_1 \leq q_2$ implica $\psi q_1 \geq \psi q_2$; e
- $p \leq \psi \varphi p$ e $q \leq \varphi \psi q$. □

Para uma conexão de Galois entre P e Q , como definido acima, o mapeamento $\psi \varphi$ é um operador de fecho sobre P e o mapeamento $\varphi \psi$ é um operador de fecho sobre Q [29].

Na próxima seção, os conceitos formais da AFC são construídos através dos denominados *operadores de derivação*, que formam uma conexão de Galois.

2.3 Análise Formal de Conceitos

A análise formal de conceitos é baseada em três conceitos fundamentais: *contextos formais*, *conceitos formais* e *reticulados conceituais*. Além destas entidades, são importantes também as *regras* que podem ser extraídas a partir de contextos formais ou reticulados conceituais. Nas próximas subseções apresenta-se esses conceitos.

2.3.1 Contextos Formais

Um contexto formal consiste de dois conjuntos e uma relação binária entre eles. Mais precisamente, um contexto formal é uma tripla (G, M, I) ⁵ em que $I \subseteq G \times M$, sendo os elementos do conjunto G chamados de *objetos*, os elementos do conjunto M chamados de *atributos* e I chamada de *relação de incidência*. Para se dizer que $(g, m) \in I$ será usada, às vezes, a notação gIm .

A Tabela 2.1 ilustra um contexto formal. Cada linha da tabela representa um objeto, cada coluna um atributo (característica) e cada marcação (\times) especifica que o objeto da linha respectiva tem o atributo da coluna em que se encontra. Um contexto

⁵As letras G e M , tradicionalmente usadas em AFC, têm origem no alemão *Gegenstände* e *Merkmale* respectivamente, que significam objetos e características.

formal representa o universo de estudo. Em outras palavras, ele é a base de dados a partir do qual se pretende extrair informação e gerar conhecimento.

Tabela 2.1. Exemplo de contexto formal.

Obj/Att	a	b	c	d	e	f
1	x	x				
2	x					
3	x	x	x	x		x
4		x		x	x	x
5	x					

Dado um conjunto $A \subseteq G$, de objetos de um contexto formal (G, M, I) , pode-se pedir quais os atributos de M são comuns a todos os objetos de A . Similarmente, pode-se pedir, para um conjunto $B \subseteq M$, quais são os objetos que possuem os atributos de B . Essas questões são respondidas pelos *operadores de derivação*, assim definidos:

- $A' = \{m \in M \mid gIm \text{ para todo } g \in A\}$; e
- $B' = \{g \in G \mid gIm \text{ para todo } m \in B\}$.

Um caso especial dos operadores de derivação ocorre quando os conjuntos de objetos e atributos são vazios. Nesse caso, tem-se: se $A = \emptyset$ então $A' = M$; e se $B = \emptyset$ então $B' = G$.

Exemplo 1. Para o contexto formal apresentado na Tabela 2.1 tem-se, por exemplo, $\{a\}' = \{1, 2, 3, 5\}$, $\{a, b\}' = \{1, 3\}$, $\{1\}' = \{a, b\}$ e $\{1, 2\}' = \{a\}$. \square

O operador de derivação e suas composições têm importantes propriedades, enunciadas a seguir [66]:

Propriedades 1. Sejam $A, A_1, A_2 \subseteq G$. Os operadores de derivação satisfazem as seguintes propriedades:

- $A \subseteq A''$;
- $A_1 \subseteq A_2$ implica $A_2' \subseteq A_1'$; e

- $A' = A'''$.

De forma análoga, para $B, B_1, B_2 \subseteq M$ tem-se:

- $B \subseteq B''$;
- $B_1 \subseteq B_2$ implica $B'_2 \subseteq B'_1$; e
- $B' = B'''$. □

Essas propriedades são utilizadas por alguns algoritmos para gerar os conceitos formais, construir a relação de cobertura e extrair implicações entre atributos [39]. Por exemplo, vejam-se os algoritmos analisados em Dias & Vieira [41].

2.3.1.1 Contextos Formais Multivalorados

Um atributo que pode assumir vários valores é denominado *multivalorado*. A definição de contexto formal apresentada não prevê esse tipo de atributo. Contextos formais em que o conjunto de atributos é composto por atributos multivalorados são chamados *contextos formais multivalorados*. Tais contextos formais são quádruplas (G, M, V, I) em que G é um conjunto de objetos, M é um conjunto de atributos multivalorados, V é o conjunto de valores possíveis para os atributos e I é uma relação ternária entre G , M e V ($I \subseteq G \times M \times V$) [68].

Contextos formais multivalorados podem ser transformados em contextos formais simples [68]. Uma maneira de fazer isso é criar um novo atributo para cada valor possível de um atributo multivalorado, como mostrado na Tabela 2.2. Nessa tabela, é apresentado um contexto formal em que T_i^j representa o par (atributo i , valor j) e w_i é o número de valores possíveis para o atributo i . Ela é tal que $(g, T_i^j) \in I$ se, e somente se, para o objeto g o atributo i tem o valor j .

É importante ressaltar ainda que um atributo multivalorado pode assumir valores em um intervalo contínuo (por exemplo, um intervalo de números reais) ou então, de um conjunto muito grande (por exemplo, o conjunto de todos os números naturais). Nesse caso, cada j em T_i^j deve representar um intervalo de valores, e não um valor apenas, de tal forma que $\{T_i^1, \dots, T_i^{w_i}\}$ represente uma partição do conjunto de valores possíveis, para o atributo i .

Tabela 2.2. Contexto formal obtido a partir de contexto formal multivalorado.

Objetos	Atrib n° 1		...	Atrib n° n			
	T_1^1	...		T_1^{w1}	...	T_n^n	...
1	×			...			×
2			×	...	×		
...				...			

2.3.2 Conceitos Formais

Os *conceitos formais* obteníveis a partir de um contexto formal (G, M, I) são pares ordenados (A, B) , em que $A \subseteq G$ e $B \subseteq M$, sendo que cada objeto em A possui todos os atributos em B e cada atributo em B é atributo de todos os objetos em A . Em outras palavras, (A, B) é um conceito formal se e somente se $A' = B$ e $B' = A$. Os conjuntos A e B são denominados *extensão* e *intenção* do conceito, respectivamente. Para se referenciar o conjunto de todos os conceitos formais existentes em um contexto formal (G, M, I) , será usada a notação $\mathcal{B}(G, M, I)$ ⁶.

Exemplo 2. Um exemplo de conceito formal extraído do contexto formal apresentado na Tabela 2.1 é o par $(\{1, 3\}, \{a, b\})$, visto que $\{1, 3\}' = \{a, b\}$ e $\{a, b\}' = \{1, 3\}$. A extensão do conceito formal é $\{1, 3\}$ e sua intenção é $\{a, b\}$. \square

Conceitos formais podem ser representados de forma compacta através dos chamados *geradores mínimos* [68].

Definição 8. Seja (A, B) um conceito formal obtido do contexto (G, M, I) . Um *gerador* de (A, B) é um subconjunto $D \subseteq M$, tal que $D' = A$. O conjunto de todos os geradores de (A, B) é $ger(A, B) = \{D \subseteq M \mid D' = A\}$. O conjunto dos *geradores mínimos* de (A, B) é $mger(A, B) = \{D \in ger(A, B) \mid \nexists m \in D : (D \setminus \{m\})' = A\}$. \square

Geradores mínimos, em particular, são importantes por permitirem a obtenção de um conceito formal a partir de um conjunto de atributos de cardinalidade mínima. Eles têm ganho grande destaque recentemente na AFC. TekayaK et al. [153] e Nehmé et al. [119] discutem algoritmos para obtenção de geradores. Já Hermann & Sertkaya

⁶A letra \mathcal{B} possui origem no alemão *Begriff*, que pode ser entendido como conceito, noção ou percepção.

[77] apresentam uma proposta para extração de geradores mais representativos, e Dong et al. [51] discutem aspectos computacionais relacionados. Sua importância se deve ainda ao fato de favorecer o princípio do *minimum description length* (MDL), segundo o qual a melhor hipótese para um conjunto de dados é aquela capaz de prover a melhor compressão para os dados [75].

2.3.3 Reticulados Conceituais

Dado um contexto formal (G, M, I) , o conjunto $\mathcal{B}(G, M, I)$ é ordenado de forma tal que $(A_1, B_1) \leq (A_2, B_2)$ se e somente se $A_1 \subseteq A_2$ ($B_2 \subseteq B_1$). O conjunto $\mathcal{B}(G, M, I)$ assim ordenado é chamado *reticulado conceitual*. Como os conceitos formais são obtidos dos operadores de derivação, que formam uma conexão de Galois, os reticulados conceituais também são referenciados na literatura como reticulados de Galois.

Exemplo 3. Do contexto formal da Tabela 2.1 obtém-se os conceitos $(\{1, 3\}, \{a, b\})$ e $(\{1, 2, 3, 5\}, \{a\})$, dentre outros. Como $\{1, 3\} \subseteq \{1, 2, 3, 5\}$, segue-se que no reticulado conceitual $(\{1, 3\}, \{a, b\}) \leq (\{1, 2, 3, 5\}, \{a\})$ e, portanto, no diagrama de linhas deverá haver um caminho do vértice correspondente a $(\{1, 3\}, \{a, b\})$ àquele correspondente a $(\{1, 2, 3, 5\}, \{a\})$, com o primeiro desenhado abaixo do segundo. \square

Em um reticulado conceitual, o ínfimo e o supremo podem ser obtidos pelo seguinte teorema, conhecido na literatura como teorema fundamental da análise formal de conceitos⁷:

Teorema 1. Seja um contexto formal (G, M, I) . Então o reticulado conceitual $\mathcal{B}(G, M, I)$ é um reticulado completo no qual, para todo conjunto de conceitos $C \subseteq \mathcal{B}(G, M, I)$, o ínfimo e o supremo são dados por:

- $\bigwedge C = (\bigcap X, (\bigcup Y)''),$ e
- $\bigvee C = ((\bigcup X)'', \bigcap Y),$

em que $X = \{A \mid (A, B) \in C\}$ e $Y = \{B \mid (A, B) \in C\}$. \square

⁷A prova do teorema pode ser obtida em Davey & Priestley [37] e Ganter & Wille [67].

Ao representar uma hierarquia conceitual através de um diagrama de linhas, convencionou-se etiquetar os objetos e atributos apenas uma vez, obtendo-se o que se denomina um *diagrama de linhas reduzido*. Desta forma, a análise visual do diagrama torna-se mais fácil. A seguir, os passos para a representação de uma hierarquia conceitual por meio de um diagrama de linhas reduzido:

1. para cada conceito formal, crie um vértice e posicione-o mais alto que seus sub-conceitos;
2. trace uma linha de cada vértice para seus vizinhos inferiores;
3. para cada atributo m , coloque seu nome no conceito formal $(\{m\}', \{m\}'')$; e
4. para cada objeto g , coloque seu nome no conceito formal $(\{g\}'', \{g\}')$.

Nos passos 3 e 4, são usados os chamados *conceito atributo* $\mu m = (\{m\}', \{m\}'')$ e *conceito objeto* $\gamma g = (\{g\}'', \{g\}')$, respectivamente. Veja-se que os nomes de atributos e objetos têm ocorrência única no diagrama, o que facilita a análise visual do mesmo. O diagrama de linhas do reticulado conceitual relativo ao contexto formal da Tabela 2.1 está exibido na Figura 2.2⁸. Nele, obtém-se o conjunto de todos os atributos de um objeto, coletando-se todos os atributos acessíveis a partir do nó rotulado com este, navegando-se de baixo para cima. Já o conjunto de todos os objetos que contenham certo atributo, obtém-se coletando-se todos os objetos acessíveis a partir do nó rotulado com o atributo, navegando-se de cima para baixo. Note-se, em particular, que a análise do diagrama de linhas pode revelar, mais claramente, informações que não sejam tão visíveis na tabela original.

2.3.3.1 Complexidade do Reticulado Conceitual

A complexidade de um reticulado conceitual pode ser expressa como uma função do número de conceitos formais. Entretanto, determinar o número de conceitos formais de um contexto formal (G, M, I) é um problema #P-completo, como mostrado por Kuznetsov [100].

O número de conceitos formais passíveis de serem obtidos a partir de um contexto formal é de ordem exponencial no pior caso. Para um contexto formal (G, M, I) ,

⁸Os diagramas de linha desse trabalho foram elaborados utilizando o software Conexp [174].

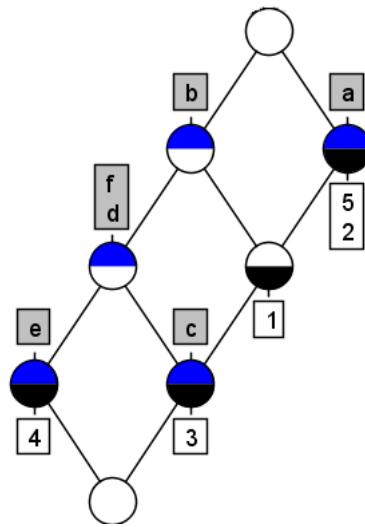


Figura 2.2. Reticulado conceitual originado do contexto formal apresentado na Tabela 2.1.

o número máximo de conceitos formais é $2^{\min(|G|, |M|)}$ [55]. Para contextos formais que apresentem um número de conceitos formais abaixo desse limite superior, diversos modelos matemáticos são observados na literatura. Esses modelos procuram contornar o fato do problema da determinação do número de conceitos formais ser #P-completo fornecendo um valor aproximado do número de conceitos formais.

Para um contexto formal multivalorado (G, M, V, I) , em que V é o conjunto de valores possíveis para os atributos, o limite superior do número de conceitos formais é $(|V| + 1)^{|M|}$. Outra possibilidade muito interessante, independentemente do tipo de contexto formal, é expressar esse limite em função do número de incidências $|I|$. Neste caso, Carpineto & Romano [29], citando Schutt [141], apresentam um limite superior de $\frac{3}{2} \times 2^{\sqrt{|I|+1}} - 1$. Kuznetsov [100] afirma que esta é uma das melhores aproximações existentes. Recentemente, em Albano [1], os limites superiores encontrados na literatura são discutidos, e um novo modelo é proposto.

Observe-se que a escolha de qual modelo utilizar torna-se uma questão difícil. Entretanto, algumas características do contexto formal e da aplicação podem auxiliar na escolha. Se o contexto formal possui uma quantidade maior de objetos, como frequentemente acontece, ou a aplicação (algoritmo) mostra-se mais sensível a esse parâmetro, um modelo que considere a relevância do número de objetos $|G|$ é mais indicado. Por

outro lado, se para a aplicação o número de atributos é mais relevante, deve-se escolher um modelo que utilize esse parâmetro.

Note-se que, apesar de existir um limite superior exponencial para alguns domínios específicos, esse comportamento é raramente encontrado na prática. Godin et al. [73] apresenta um modelo que descreve o comportamento do número de conceitos formais para uma aplicação de recuperação de informação. No domínio de recuperação de informação, se porventura um atributo representa um termo (palavra) e os objetos representam documentos, não é relevante para o modelo um termo comum a todos eles.

2.3.4 Regras

Dado um contexto formal ou um reticulado conceitual, deles podem ser extraídas regras exatas ou aproximadas (regras que possuem valores estatísticos, por exemplo, suporte e confiança) que expressem de forma alternativa o conhecimento subjacente. As regras exatas podem ser classificadas em regras de implicação e de dependências funcionais, enquanto as regras aproximadas dividem-se em regras de classificação e regras de associação [156].

É particularmente importante no trabalho desenvolvido nesta tese o uso de regras de implicação, a partir de agora chamadas de implicações apenas. Apesar da existência de uma forte relação entre implicações e dependências funcionais, estas últimas são mais exploradas no contexto de banco de dados, onde se lida com atributos multivalorados [156]. As implicações podem ser obtidas, tanto de contextos formais tradicionais, quanto a partir dos reticulados conceituais respectivos. Elas constituem uma forma alternativa de se expressar o que é expresso mediante contextos formais e reticulados conceituais, no sentido de que um reticulado isomorfo ao original pode ser obtido do conjunto de implicações. Segue a definição de implicação [71]:

Definição 9. *Seja um contexto formal cujo conjunto de atributos é M . Uma implicação é uma expressão $P \rightarrow Q$, em que $P, Q \subseteq M$. □*

Uma implicação $P \rightarrow Q$, extraída de um contexto formal, ou reticulado conceitual respectivo, deve ser tal que $P' \subseteq Q'$. Em outras palavras: todo objeto que possui os atributos de P possui os atributos de Q .

Exemplo 4. A implicação $\{f\} \rightarrow \{b, d\}$ segue do reticulado da Figura 2.2. Ela diz que todo objeto que tem o atributo f tem, também, os atributos b e d . \square

Define-se, a seguir, o que significa uma implicação *seguir de* um conjunto de implicações. Se X é um conjunto de atributos, então se diz que X *respeita* uma implicação $P \rightarrow Q$ se e somente se $P \not\subseteq X$ ou $Q \subseteq X$; e X *respeita* um conjunto de implicações \mathcal{I} se e somente se X respeita todas as implicações em \mathcal{I} . Com isto, define-se que $P \rightarrow Q$ *segue de* um conjunto de implicações \mathcal{I} se e somente se para todo conjunto de atributos X , se X respeita \mathcal{I} , então respeita também $P \rightarrow Q$.

Terão importância especial as implicações de um contexto formal. Uma implicação $P \rightarrow Q$ *se verifica* em um conjunto $\{X_1, \dots, X_n\} \subseteq \mathcal{P}(M)$ se e somente se cada X_i respeita $P \rightarrow Q$; e $P \rightarrow Q$ *é uma implicação do contexto* (G, M, I) se e somente se ela se verifica no conjunto de intencões dos objetos do contexto⁹.

Agora, definem-se os conceitos de completude e redundância de um conjunto de implicações de um contexto formal. Um conjunto de implicações \mathcal{I} de um contexto formal (G, M, I) *é completo* para (G, M, I) se e somente se qualquer implicação de (G, M, I) segue de \mathcal{I} . E \mathcal{I} *é redundante* se e somente se existe pelo menos uma implicação $i \in \mathcal{I}$, tal que i segue de $\mathcal{I} \setminus \{i\}$.

Um conjunto de implicações *é fechado* se e somente se toda implicação que segue dele já pertence a ele. Seja \mathcal{I} o conjunto fechado das implicações que seguem de um contexto formal. Tal conjunto é normalmente imenso e, altamente redundante. Em particular, todas as implicações que sigam dos axiomas de Armstrong [5] estarão presentes:

- $P \rightarrow P \in \mathcal{I}$ para todo $P \subseteq M$;
- se $P \rightarrow Q \in \mathcal{I}$, então $P \cup R \rightarrow Q \in \mathcal{I}$ para todo $P, Q, R \subseteq M$; e
- se $P \rightarrow Q \in \mathcal{I}$ e $Q \cup R \rightarrow S$, então $P \cup R \rightarrow S \in \mathcal{I}$ para todo $P, Q, R, S \subseteq M$.

Tais axiomas são denominados *reflexividade*, *aumento* e *pseudotransitividade*. Estarão presentes ainda as implicações que seguem de:

- se $P \rightarrow Q \in \mathcal{I}$ e $P \rightarrow R \in \mathcal{I}$ então $P \rightarrow Q \cup R$ para todo $P, Q, R \subseteq M$ (*união*); e
- se $P \rightarrow Q \cup R \in \mathcal{I}$ então $P \rightarrow Q$ para todo $P, Q, R \subseteq M$ (*decomposição*).

⁹Conjunto de intencões dos objetos: $\{\{g\}' \mid g \in G\}$.

O cálculo e uso direto de um conjunto fechado completo para um contexto formal é inviável na maioria das aplicações.

Em geral, pode haver vários conjuntos completos não redundantes para um contexto formal, inclusive de várias cardinalidades. Um conjunto completo não redundante com um número mínimo de implicações é conhecido na literatura como *Duquenne-Guigues-Base* ou *Steam Base* [68]. Esse conjunto não é calculável em tempo polinomial no tamanho da entrada [49, 103] e o problema de determinar sua cardinalidade é #P-completo [104]. Existem também várias propostas de extração de conjuntos de implicações completos com pouca redundância, adequados para uso em certas circunstâncias [27, 28, 112, 135, 152]. Em certos casos, procura-se transformar um conjunto completo em um mínimo não redundante para o mesmo contexto formal [21, 114, 160]. É conhecido que redundância em um conjunto de implicações pode ser removida em tempo linear [114]; contudo, o conjunto não redundante obtido é dependente da ordem de processamento das implicações.

2.4 Conclusões

Este capítulo apresentou os principais conceitos da AFC utilizando a notação e a terminologia expostas no livro de Ganter & Wille [68]. Os conceitos apresentados são suficientes para fundamentar o conteúdo dos próximos capítulos. Em especial, implicações obtidas do contexto formal (reticulado conceitual) vão assumir notável importância para análise e caracterização do conhecimento expresso, mediante contextos formais e reticulados conceituais reduzidos.

Capítulo 3

Redução de Reticulados Conceituais: Características e Classificação

A análise formal de conceitos (AFC) é considerada um importante formalismo para a representação, extração e análise de conhecimento com aplicações em diferentes áreas. Um problema identificado em diversas aplicações é o custo computacional, devido ao grande número de conceitos formais gerados. Mesmo quando esse número não se mostra de grande vulto, os aspectos essenciais, aqueles efetivamente procurados, podem estar imersos em um emaranhado de detalhes irrelevantes. De fato, o problema de obtenção de um reticulado conceitual de complexidade e tamanho apropriados é um dos problemas mais significantes da AFC. Na literatura, diferentes técnicas visando controlar a complexidade e o tamanho de um reticulado conceitual foram descritas, todavia, até o presente momento elas não foram devidamente analisadas, comparadas e ou classificadas. Assim, este capítulo apresenta um levantamento das principais técnicas existentes e propõe uma análise e classificação das mesmas. A análise das técnicas é elaborada empregando sete dimensões, cada uma delas constituindo um conjunto de características. Analisando-se as principais técnicas para redução, por meio da análise formal de conceitos, considerações são realizadas sobre complexidade computacional, viabilidade e qualidade do reticulado conceitual resultante do processo de redução.

Frente ao problema de lidar com a complexidade do reticulado conceitual com

respeito ao número de conceitos formais e ao emaranhado de inter-relacionamentos, emergem técnicas que buscam um equilíbrio entre a qualidade da informação representada, o custo computacional para gerar e manter o reticulado, e a facilidade de análise do mesmo. Essas técnicas são definidas, aqui, como técnicas para *redução de reticulados conceituais*.

Definição 10. *Uma técnica para redução de reticulados conceituais é aquela que tem como objetivo reduzir a complexidade de um reticulado conceitual, tanto em termos de magnitude quanto de inter-relacionamentos, procurando manter as informações relevantes.* □

A complexidade do reticulado conceitual é apontada na literatura como um problema chave para a análise formal, enquanto técnicas com diferentes características são descritas [13, 16, 69, 102, 116, 137, 144, 149, 184, 185]. Por exemplo, Pei & Mi [126] citam que na vida real existem muitos dados desnecessários, e que a redução torna-se vital para desconsiderá-los. Por outro lado, Liu & Mi [113] ressaltam que um problema chave na descoberta de conhecimento consiste na redução dos atributos. Ademais, conforme já dito, além de poder ser computacionalmente caro calcular todos os conceitos formais, a quantidade de inter-relacionamentos destes pode vir a inviabilizar uma análise em tempo razoável [137].

3.1 Classificação

Este trabalho propõe a divisão das técnicas para redução de reticulados conceituais em três grupos: *eliminação de informação redundante*, *simplificação* e *seleção*.

As técnicas de *eliminação de informação redundante* procuram obter um reticulado conceitual isomorfo ao original. Em geral, elas têm como objetivo encontrar um contexto formal com um número mínimo de objetos ou atributos, mantendo inalterada a estrutura do reticulado conceitual [107–110, 113, 116, 125, 126, 134, 161–164, 166, 167, 185]. Por outro lado, técnicas de *simplificação* procuram obter uma abstração do reticulado conceitual, ou seja, um resumo em alto nível que exponha apenas o essencial [10, 15, 34, 40, 42, 62, 69, 96, 97, 144, 145]. Por último, as técnicas de *seleção* atuam selecionando conceitos formais, objetos ou atributos com base em um critério de relevância [4, 13, 16–19, 25, 70, 74, 99, 101, 127, 136–138, 146, 149, 155, 184].

Este trabalho considera apenas técnicas que pressupõem o contexto formal construído; em outras palavras, não são levadas em conta técnicas que se utilizam dos dados originais que, porventura, sejam usados para obtenção do contexto formal. Trabalhos nessa linha poderiam, eventualmente, pertencer a uma quarta classe; exemplos, são aqueles trabalhos descritos por Ganter & Kuznetsov [64], Kaytoue et al. [86] e Brito & Polaillon [26]. Além disso, apenas trabalhos da AFC clássica foram considerados; extensões não foram tratadas. Por exemplo, técnicas que trabalham com contextos formais ou reticulados conceituais fuzzy não foram exaustivamente analisadas. Técnicas dessa natureza são definidas, aqui, como *extensões da AFC* e são caracterizadas na dimensão *ponto de partida*.

Nas próximas seções, cada classe será definida e caracterizada e as principais técnicas de cada classe serão discutidas. O contexto formal apresentado na Tabela 2.1 e o reticulado conceitual correspondente (Figura 2.2) serão utilizados para ilustrar as principais técnicas.

3.1.1 Eliminação de Informação Redundante

Inicialmente, o conceito de informação redundante é definido.

Definição 11. *Seja $\mathcal{B}(G, M, I)$ um reticulado conceitual. Um objeto $g \in G$, atributo $m \in M$ ou incidência $i \in I$ são considerados **informação redundante** se sua remoção ou transformação resultar em um reticulado isomorfo a $\mathcal{B}(G, M, I)$.* \square

De acordo com essa definição, se após a eliminação ou alteração de qualquer objeto $g \in G$, atributo $m \in M$ ou incidência $i \in I$, o reticulado conceitual resultante for isomorfo ao original, considera-se tais elementos redundantes. Entretanto, mesmo redundantes, eles podem, eventualmente, ser considerados relevantes em algumas aplicações. Por exemplo, objetos que possuem os mesmos atributos que outro são redundantes; mesmo assim, eles podem ser usados, por exemplo, para comporem o número de objetos da extensão de um conceito, número este importante em algumas aplicações. Isto terá sentido quando objetos com os mesmos atributos puderem, ainda assim, ser considerados distintos.

A substituição de objetos com exatamente os mesmos atributos por um único objeto e a de atributos que ocorram exatamente nos mesmos objetos por um único atributo

é uma primeira simplificação possível. Mais formalmente, se $g, h \in G$ e $g' = h'$ então g e h podem ser representados por um único objeto. De forma dual, se $m, n \in M$ e $m' = n'$, então m e n podem ser representados por um único atributo. Um contexto formal com tais características é chamado de *contexto formal clarificado* [68].

Exemplo 5. No contexto formal da Tabela 2.1 tem-se $2' = 5'$ e $d' = f'$. Com isso, na representação do reticulado conceitual correspondente àquele contexto, mediante o diagrama de linhas reduzido mostrado na Figura 2.2, os referidos objetos, 2 e 5, rotulam os mesmos nós do diagrama, assim como os atributos d e f . Como resultado, eles aparecem sempre juntos em extensões e intensões, sendo, portanto indistinguíveis do ponto de vista do reticulado conceitual em si. \square

Outro tipo de redução que também não altera a estrutura do reticulado é a eliminação de atributos representados como a combinação de um conjunto de outros atributos. Supondo que $m \in M$, $X \subseteq M$ e $m \notin X$, se $m' = X'$ então o conceito atributo μm é o ínfimo dos conceitos atributo μx , tais que $x \in X$. Com isto, se o atributo m for eliminado do contexto formal, o reticulado obtido do contexto formal resultante é isomorfo ao original. Tal atributo m é denominado *atributo redutível*. De maneira análoga, têm-se os *objetos redutíveis*, aqueles $g \in G$ tais que o conceito γg é o supremo dos conceitos objeto γy tais que $y \in Y$, sendo $Y \subseteq G$ e $g \notin Y$. A eliminação de tais objetos também resulta em um contexto formal a partir do qual se obtém um reticulado isomorfo ao original [68].

Exemplo 6. No contexto formal apresentado na Tabela 2.1, tem-se que $c' = \{a, d\}'$ e, portanto, μc é o ínfimo de μa e μd , como se pode constatar na Figura 2.1. Logo, se o atributo c for eliminado, o reticulado obtido do contexto formal em que é eliminada a coluna referente a c é isomorfo ao original; no diagrama de linhas reduzido desaparece o rótulo c , simplesmente. \square

Após a remoção de todos os objetos e atributos redutíveis de um contexto formal clarificado, resulta um contexto formal denominado *contexto formal padrão*. O reticulado conceitual relativo a um contexto padrão é isomorfo ao reticulado relativo ao contexto formal original [68].

Uma série de técnicas foram desenvolvidas objetivando reduzir contextos formais, de modo a preservar as estruturas dos reticulados conceituais correspondentes [107,

108, 113, 116, 125, 126, 143, 161, 163–167, 185], algumas se originando diretamente das considerações descritas anteriormente. Em geral, essas técnicas atuam através da redução dos atributos, e são chamadas técnicas de *redução de atributos*.

Definição 12. *A redução de atributos em um reticulado conceitual $\mathcal{B}(G, M, I)$ é a busca de um conjunto $X \subseteq M$, com a menor cardinalidade possível, tal que:*

$$\mathcal{B}(G, X, I') \text{ é isomorfo a } \mathcal{B}(G, M, I)$$

sendo $I' = I \cap (G \times X)$. Tal X é denominado **conjunto minimal** de atributos. \square

Embora exista a definição análoga de *redução de objetos*, na prática as técnicas normalmente lidam com a redução de atributos. Computacionalmente, uma aplicação direta da Definição 12 tem um custo elevado, uma vez que há um número exponencial de subconjuntos de atributos. Para determinar um conjunto minimal de atributos visando a redução referida na Definição 12, Zhang et al. [185] utilizam o conceito de matriz de discriminação.

Definição 13. *Seja (G, M, I) e $(A, B), (C, D) \in \mathcal{B}(G, M, I)$. A discriminação entre os conceitos (A, B) e (C, D) é dada pela diferença simétrica entre suas intenções B e D , ou seja, $Dis_{(A,B),(C,D)} = (B \cup D) \setminus (B \cap D)$. \square*

Dada uma matriz de discriminação Dis (a matriz simétrica indexada por conceitos) resultante de um contexto formal (G, M, I) , um conjunto minimal de atributos $X \subseteq M$ é um dos de menor cardinalidade tal que $X \cap Z \neq \emptyset$ para todo $Z \in Dis$ [185].

Em Qi [134] são apresentados critérios que permitem reduzir o número de cálculos de discriminação entre conceitos, de forma a preservar a possibilidade de obtenção de um conjunto minimal.

Wang et al. [162] apresentam uma técnica heurística para determinar um conjunto minimal de atributos. Já Wang & Zhang [164] estendem o trabalho de Zhang et al. [185] para conjuntos aproximados. A junção de objetos ou de atributos é baseada na estrutura do reticulado conceitual. Partindo desse ponto de vista, outras técnicas para redução de reticulados foram propostas [113, 163, 164].

Explorando conceitos do reticulado, Liu & Mi [113] são capazes de determinar famílias de atributos comuns aos objetos. A combinação de alguns elementos dessas

famílias pode produzir um reticulado conceitual reduzido. Um algoritmo para aplicação da técnica é apresentado. Os resultados de Liu & Mi [113] são idênticos aos apresentados por Zhang et al. [185]. A principal diferença para os trabalhos de Wei et al. [167], Zhang et al. [185] e Wang & Zhang [166] é a não utilização de uma matriz de discriminação.

Técnicas de redução que mantêm a estrutura do reticulado isomorfa ao reticulado original também são observadas em extensões da AFC. Por exemplo, em Wang & Zhang [161] a redução fundamentada em matriz de discriminação proposta por Zhang et al. [185] é empregada em contextos formais de decisão.¹ Em Pei & Mi [126] uma redução baseada em álgebra também é proposta para contextos formais de decisão. Por outro lado, Pei et al. [125] utilizam uma técnica similar de redução para contextos formais de decisão *fuzzy*. Ainda utilizando contextos formais de decisão, Li et al. [107] aplicam o algoritmo heurístico proposto em Wang et al. [162].

Alguns problemas específicos também envolvem redução de reticulados conceituais. Por exemplo, no contexto de tomada de decisões baseada em implicações, Li et al. [108] tratam o problema de redução sob a perspectiva de extração de implicações. Já Li et al. [110] empregam técnicas para redução em contextos formais com dados ausentes.

3.1.2 Simplificação

Todos os relacionamentos entre conceitos estão presentes no reticulado conceitual, incluindo aqueles entre conceitos muito “similares” [69]. Por exemplo, aquele entre dois conceitos formais distintos apenas com respeito a um único atributo. Nesse caso, considerando-se a necessidade de expor apenas o essencial e que o atributo que distingue os referidos conceitos não seja relevante (de acordo com um critério de relevância), pode-se simplificar ou abstrair essa diferença. Técnicas que seguem essa linha são denominadas aqui técnicas de *simplificação*.

Definição 14. *Uma técnica de **simplificação** é aquela que, a partir do contexto formal ou reticulado conceitual, abstrai diferenças não essenciais (de acordo com algum critério) entre conceitos, objetos ou atributos.* □

¹Um contexto formal de decisão é uma extensão de um contexto formal no qual o conjunto de atributos é dividido em atributos de decisão e condição [108].

Técnicas de simplificação baseadas em métodos de agrupamento são muito naturais [79]. Agrupamento pode ser aplicado no conjunto de objetos, atributos ou conceitos formais. Em Cheung & Vogel [34] o agrupamento de objetos se realiza no intuito de reduzir a dimensionalidade de reticulados conceituais no contexto de recuperação de informação. O trabalho usa decomposição em valores singulares (SVD)² para construir classes de equivalência de objetos. Nesse novo espaço é possível, por exemplo, empregar um método linear para indução de classes de equivalência. Cheung & Vogel [34] aplicam a relação de equivalência, utilizando matrizes reduzidas pela técnica SVD, assim definida: dois objetos g_1 e g_2 (representando documentos) são considerados equivalentes se e somente se o cosseno do ângulo entre os dois for superior a certo limiar preestabelecido quando se considera um espaço vetorial [34]. Eles definem um novo objeto h formado pela junção de todos os objetos de uma classe de equivalência $[g]$, tendo como atributos a união dos atributos dos elementos de $[g]$; i.e., se I é relação de incidência do contexto formal original, $h' = \{m \mid kIm \text{ para algum } k \in [g]\}$.

Exemplo 7. Seja o contexto formal apresentado na Tabela 2.1 e suponha-se que as classes de equivalência sejam todas unitárias, exceto a classe constituída dos objetos 1 e 3. Nesse caso, um novo objeto g substituiria 1 e 3 contendo os atributos de ambos: $g' = 1' \cup 3' = \{a, b, c, d, f\}$, enquanto que os objetos restantes permaneceriam inalterados. \square

A união dos atributos dos objetos pertencentes a uma classe de equivalência não é adequada para todo tipo de aplicação. Por exemplo, em um contexto formal que relacione animais e características, a união do conjunto de características de um animal com o de outro pode resultar em uma aberração, um “animal inexistente”.

Em paralelo ao trabalho de Cheung & Vogel [34], e na mesma linha de redução baseada em SVD, encontram-se os trabalhos de Gajdos et al. [62] e Snásel et al. [144]. Recentemente, Codocedo et al. [36] empregam a técnica SVD para descrever o conhecimento em um domínio de arquitetura de software.

Técnicas para redução de contextos formais baseadas em métodos algébricos normalmente apresentam um custo computacional elevado [97]. Como alternativa, Kumar & Srinivas [97] realizam o agrupamento de objetos através do algoritmo *fuzzy k-means*,

²Decomposição em valores singulares (do inglês *singular value decomposition*) é uma das várias técnicas de decomposição de matrizes originada na álgebra linear numérica capaz de projetar uma matriz de alta dimensionalidade em um espaço com poucas dimensões [151].

uma extensão do algoritmo *k-means*, em que cada registro pode pertencer a mais de um grupo com certa probabilidade. Kumar [96] emprega a mesma técnica para mineração de regras de associação de reticulados conceituais, em que aplicações na área médica são conduzidas. Cheung & Vogel [34] e Kumar & Srinivas [97] apontam que o reticulado reduzido pela técnica proposta seria homomorfo ao original.³ Recentemente, entretanto, Krupka [94] mostra que isso não é verdade; i.e., o reticulado resultante pode não ser homomorfo ao original.

Conhecimento prévio a respeito do domínio do problema pode ser incorporado na técnica para guiar o processo de simplificação [13, 16–19, 34, 40, 42, 184]. No contexto de técnicas relacionadas com agrupamento e incorporando conhecimento adicional ao do contexto formal, tem-se a técnica JBOS (*junction based on objects similarity*), a ser apresentada no Capítulo 4, que substitui grupos de objetos similares por objetos que os representem, com base em uma avaliação qualitativa de seus atributos. A cada atributo $m \in M$, é associado um peso w^m tal que $0 \leq w^m \leq 1$, que pode ser imaginado como representando a *relevância* do atributo, desde 0 (relevância nenhuma) a 1 (relevância máxima). A similaridade entre dois objetos g e h é dada pela soma ponderada dos pesos dos atributos em que ambos concordem (ambos possuam ou ambos não possuam). Aplicando-se um limiar de similaridade escolhido pelo usuário, são formados grupos de objetos similares com um certo número máximo de objetos, também escolhido pelo usuário. Os atributos de cada grupo H de objetos similares são os do conjunto $\bigcap \{g' \mid g \in H\}$, sendo o operador de derivação aplicado no contexto formal original.

A técnica JBOS difere de outras que utilizam conhecimento prévio, como as apresentadas em Cheung & Vogel [34], Codocedo et al. [36], Kumar & Srinivas [97], Snásel et al. [145] e Kumar [96], pelo fato de preservar apenas os atributos da interseção dos objetos similares. Essa característica faz com que haja, efetivamente, uma simplificação de objetos, não a criação de objetos fora do domínio do contexto formal. JBOS é apresentada em detalhes no Capítulo 4.

O reticulado conceitual pode ser visto como um grafo em que cada conceito formal corresponde a um vértice [69, 83]. Gély [69] relaciona decomposição modular usada em teoria dos grafos com a AFC. A ideia geral é que um conjunto de vértices de um grafo que formam um módulo pode ser substituído por um único vértice. Dado

³O mapeamento $f : L \rightarrow L_r$ de dois reticulados L e L_r é chamado homomorfo se ele preserva o supremo e o ínfimo.

um grafo (V, E) , um subconjunto X de V é um módulo se cada vértice em $V \setminus X$ é adjacente a todos ou a nenhum vértice de X . Gély [69] apresenta um algoritmo para agrupar módulos (conjuntos de conceitos formais) em vértices (conceitos formais).

Uma forma natural de controlar a complexidade do reticulado conceitual é através do controle dos detalhes representados no contexto formal. Com esse propósito, Belohlavek & Sklenar [15] utilizam a seleção de níveis de granularidade dos atributos do contexto formal. Para cada atributo, diminui-se a representatividade compondo atributos multivalorados em um atributo único ou se amplia o nível de detalhes desmembrando um atributo em novos atributos. Recentemente, Belohlavek et al. [10] discutem aspectos computacionais da proposta.

Dentre as técnicas de simplificação que lidam com extensões da AFC, destacam-se as de Krupka [92, 93] e de Belohlavek et al. [12], que empregam contextos formais *fuzzy*.

3.1.3 Seleção

Muitos conceitos formais, particularmente em grandes reticulados conceituais, podem ser considerados irrelevantes em algumas aplicações específicas. A “relevância” de um conceito formal pode ter relação com a cardinalidade de sua intenção ou extensão [25], com o relacionamento entre alguns atributos [16], etc. Técnicas que selecionem conceitos, ou mesmo objetos ou atributos, mediante algum critério de relevância, serão chamadas técnicas de *seleção*.

Definição 15. *Uma técnica de seleção é aquela que, a partir do contexto formal ou reticulado conceitual, seleciona um subconjunto de conceitos formais, objetos ou atributos que satisfaçam a um conjunto de restrições.* □

Em muitos cenários existe um conhecimento adicional acerca dos conjuntos de objetos e de atributos. Algumas técnicas de seleção empregam esse conhecimento para guiar o processo de redução. Algumas delas empregam a atribuição de pesos [13, 184], hierarquias [17], lógica [16, 19], etc., para dar sustentação a seus métodos de seleção.

Belohlavek et al. [16] e Belohlavek & Vychodil [19] propõem utilizar o conhecimento do usuário para criar restrições que expressem certas dependências entre os atributos do contexto formal. Durante a geração dos conceitos formais e construção do

reticulado conceitual, apenas conceitos formais que satisfaçam as restrições são levados em conta. De forma similar, Belohlavek et al. [17] exploram a incorporação de conhecimento previamente existente, de modo a reduzir o número de conceitos formais. Esse conhecimento é representado por intermédio de uma hierarquia entre os atributos previamente definida pelo usuário. Novamente, apenas os conceitos formais que atendam às restrições são selecionados.

A técnica proposta por Belohlavek & Macko [13] atribui um peso a cada atributo do contexto formal no intuito de explicitar sua relevância e, então, selecionar apenas conceitos formais considerados relevantes, i.e., conceitos formais com intenções compostas de atributos relevantes. A atribuição de pesos se assemelha à proposta apresentada por Dias & Vieira [40] (embora esta última seja uma técnica de simplificação), procurando capturar a relevância dos conceitos a partir do conhecimento existente para os atributos. A fim de facilitar o processo de aplicação dos pesos, Belohlavek & Macko [13] atribuem a mesma relevância a todo atributo derivado de um atributo multivalorado. A importância de um conceito formal é mensurada utilizando-se uma função de agregação (soma, média ou máximo). Por exemplo, na função média tem-se a soma dos pesos dos atributos de sua intenção dividido pela cardinalidade da intenção (similar ao trabalho de Zhang et al. [184]). Apenas conceitos com valor superior ao definido pelo usuário são selecionados.

Belohlavek & Macko [13] também sugerem que geradores ou geradores mínimos dos conceitos formais possam ser utilizados no lugar da intenção para determinar a relevância dos conceitos. No caso, a relevância de um conceito é tomada como a maior dentre as de seus geradores ou geradores mínimos. O maior problema da técnica reside na necessidade de computar o conceito formal e então verificar se sua relevância é superior ao limiar estabelecido, uma vez que a geração de todos os conceitos formais nem sempre é viável. Em particular, para a análise baseada em geradores e geradores mínimos faz-se mister enumerar os geradores e identificar aqueles que atendam ao limiar utilizado.

Exemplo 8. Seja a técnica de Belohlavek & Macko [13] baseada em geradores mínimos. Suponha-se que os atributos mostrados na Tabela 2.1 tenham os pesos $a=1$, $b=1$, $c=0$, $d=0.5$, $e=0$ e $f=0.5$. Suponha-se, ainda, que a função utilizada na seleção seja a média e que o limiar de corte seja 0.75. Nesse caso, são selecionados os conceitos

$(\{1,2,3,5\}, \{a\})$, $(\{1,3,4\}, \{b\})$, $(\{1,3\}, \{a,b\})$ e $(\{3\}, \{a,b,c,d,f\})$. Os geradores mínimos dos três primeiros conceitos são $\{a\}$, $\{b\}$ e $\{a,b\}$, respectivamente. Logo, a relevância de cada conceito é igual a 1. O conceito $(\{3\}, \{a,b,c,d,f\})$ possui três geradores mínimos $\{a,d\}$, $\{a,f\}$ e $\{c\}$ com relevâncias 0.75, 0.75 e 0, respectivamente. Logo, a relevância do conceito é a maior das três: 0.75. \square

Recentemente, Zhang et al. [184] também empregaram a atribuição de pesos a atributos como meio de associar relevância a conceitos formais. A relevância dos conceitos formais é mensurada como por Belohlavek & Macko [13]; entretanto, os autores completam o reticulado conceitual através da criação de conceitos formais virtuais.

Em uma comparação com técnicas que trabalham com pesos para extração de conjunto de itens fechados [159, 175], que é um sinônimo para a intenção de um conceito formal, a técnica proposta por Zhang et al. [184] e Belohlavek & Macko [13] não provêm uma propriedade anti-monotônica para o peso dos conceitos formais. Essa propriedade é importante por permitir a seleção de conceitos infrequentes, mas relevantes. Por outro lado, cria-se a necessidade de percorrer todo o conjunto de conceitos formais para determinação daqueles considerados relevantes.

Muitas técnicas para redução de reticulados conceituais utilizam conceitos de mineração de dados. Um dos principais trabalhos nesse sentido é aquele apresentado por Pasquier et al. [123], que relaciona itens frequentes e conceitos formais. Para relacionar itens frequentes e conceitos formais, inicialmente definem-se os conceitos de *suporte* e de *conjuntos frequentes*. O suporte de um conjunto $S \subseteq M$, $sup(S, G)$, é o número de objetos em G que contêm todos os atributos em S . O conjunto $S \subseteq M$ é dito frequente se e somente se $sup(S, G) \geq minSup$, em que $minSup$ é um suporte mínimo fornecido como parâmetro.

Definição 16. *Dado um conceito formal (G, M, I) e um suporte mínimo, $minSup$, um conceito formal (X, Y) é dito frequente se e somente se $sup(Y, G) \geq minSup$. \square*

A relação entre conceitos formais frequentes e itens frequentes faz-se possível devido ao suporte de um item Y ser igual ao suporte de seu fecho Y'' ([66] cap. 4, Lema 15), garantindo assim, que todos os itens frequentes sejam determinados, unicamente, por conceitos formais também frequentes.

Reticulados conceituais construídos sobre itens frequentes são chamados *reticulados de conceitos frequentes*⁴. Note-se que, nesse caso, o reticulado obtido é parcial; apenas os conceitos formais frequentes são representados.

Stumme et al. [149] mostram a utilização do algoritmo *Titanic* para geração de reticulados conceituais frequentes, bem como o potencial desses reticulados em diversas aplicações, tais como análise de grandes bases de dados, extração de implicações, extração de regras de associação, visualização de implicações, dentre outras.

Relacionado a reticulados conceituais frequentes, há os *reticulados conceituais alfa* [127]. Suponha-se que, no contexto formal da Tabela 2.1 sejam acrescentados novos atributos que designem a que “classe” pertence um objeto, e que cada objeto pertença a uma e apenas uma das classes. A classe a que pertence um objeto, nesse caso, é o atributo principal do mesmo. Por exemplo, em um contexto formal descrevendo a taxonomia dos animais, pode-se ter como classes, mamífero, inseto e ave. Pode-se aplicar um suporte mínimo por classe, criando-se, assim, restrições locais [155]. O reticulado conceitual construído utilizando-se restrições por classe é chamado reticulado conceitual alfa [127]. Recentemente, Soldano et al. [146] discutem a construção de reticulados alfa.

Algumas técnicas fazem suas seleções baseadas na frequência de algumas características específicas dos conceitos formais. Por exemplo, um conceito (A, B) pode ser selecionado se ele satisfaz restrições como $|A| > \alpha$, $|B| > \alpha$, $(|A| \times |B|) > \alpha$, etc., em que α é um parâmetro fornecido pelo usuário. Boulicaut & Besson [25] discutem restrições dessa natureza sob a perspectiva de mineração de dados. Recentemente, Belohlavek & Vychodil [20] consolidam um conjunto de restrições que podem vir a serem aplicadas diretamente no operador de derivação, e avaliam a utilização dessas restrições no reticulado conceitual e na extração de implicações.

A qualidade do contexto formal consiste num outro aspecto considerado, quando se reduz o número de conceitos formais. Diversos trabalhos apontam a dificuldade de se construir reticulados conceituais consistentes quando os dados originais apresentam algum tipo de ruído [89, 102, 138]. Muitas técnicas de seleção levam em consideração esse aspecto [74, 89, 101, 102, 136].

Uma técnica comumente utilizada é a seleção baseada no índice de estabilidade [99, 101]. A ideia é criar um índice para os conceitos que indique o quanto a intenção

⁴Do inglês *iceberg concept lattices*

do conceito depende do conjunto de objetos. Utilizando-se esse índice, pode-se assumir um limiar a partir do qual todos os conceitos formais com valor inferior são removidos. Contudo, o cálculo desse índice é #P-completo [101], devido à necessidade de gerar todos os subconjuntos de cada conceito formal. Para contornar esse problema, Kuznetsov et al. [102] apresentam duas heurísticas. Recentemente, Krupka [94] propõe um cálculo aproximado do índice de estabilidade, utilizando cadeias de Markov.

Klimushkin et al. [89] identificam dois tipos de erros que podem ocorrer em um contexto formal: alteração de uma incidência (tipo I) e inserção de objetos ou atributos (tipo II). Tipicamente, esses erros ocorrem durante a modelagem do contexto formal. Nesse cenário, contextos formais com informações incorretas, Klimushkin et al. [89] discutem três técnicas para seleção de conceitos formais: uma baseada em estabilidade, outra com base em uma análise probabilística, e uma terceira baseada na noção de separação. A seleção que se baseia tão somente em estabilidade pode não selecionar conceitos de baixa estabilidade que sejam relevantes. A baseada em probabilidade cria um índice probabilístico que torna possível a seleção desses conceitos. O índice de separação é a razão entre a área abrangida por um conceito ($|A|x|B|$) e a área total coberta por seus objetos e atributos. Klimushkin et al. [89] destacam o fato de o índice de estabilidade ser muito adequado para ruídos do tipo II. Já para ruídos do tipo I, a estabilidade nem sempre apresenta bons resultados. A separação deve ser utilizada como um índice auxiliar. Por último, a probabilidade, assim como a separação, não deve ser empregada isoladamente.

Assim como Klimushkin et al. [89], Riadh et al. [136] expõem a necessidade de se selecionar conceitos relevantes provenientes de dados com ruídos. Com esse propósito, os autores discutem dois índices: o *relacionamento* e a *proximidade*. O relacionamento de um objeto $g \in G$ indica a proporção de objetos em G que possuem propriedades em comum com g . Já a proximidade, indica a proporção de objetos comuns entre g e os objetos relacionados.

Algumas técnicas norteiam a seleção de conceitos formais com base na noção de distância ou similaridade. Rice & Siff [137] propõem uma medida de similaridade entre extensões de conceitos formais, como segue. Dados dois conceitos formais (A_1, B_1) e (A_2, B_2) , a similaridade entre A_1 e A_2 , e, conseqüentemente, entre os respectivos conceitos, é dada por $sim(A_1, A_2) = 1 - \frac{|A'_1 \cap A'_2|}{|A'_1 \cup A'_2|}$. A medida de similaridade é usada para selecionar um subconjunto de conceitos formais próximos, que pode ser substancial-

mente menor que o conjunto original de conceitos formais. Nesse contexto, medidas de similaridade com diferentes características podem ser empregadas, como, por exemplo, aquelas descritas por Alqadah & Bhatnagar [3], Formica [60], Muangprathub et al. [118] e Yan et al. [172].

Dentre as técnicas de seleção, existem aquelas que necessitam de acesso prévio a todos os conceitos formais que compõem o reticulado, e aquelas que não necessitam. As primeiras geram os conceitos formais ou o reticulado conceitual e posteriormente selecionam os conceitos formais considerados relevantes (mediante o uso de um limiar definido pelo usuário) [7, 13, 101, 127, 184]. As demais, em geral, aplicam restrições durante a construção do reticulado conceitual, e geram apenas os conceitos formais relevantes para a tarefa em mente [18–20, 25, 74, 137, 146, 149].

3.2 Dimensões para Análise das Técnicas de Redução

Identificamos sete dimensões para a análise das técnicas de redução, cada uma consistindo de um conjunto de características, como segue:

1. **Ponto de partida:** a) contexto formal, b) conceitos formais, c) reticulado conceitual e d) extensão da AFC.
2. **Utiliza conhecimento adicional:** a) sim e b) não.
3. **Alterações no contexto formal:** a) altera o conjunto de objetos, b) altera o conjunto de atributos e c) altera o conjunto de incidências.
4. **Reticulado conceitual resultante:** a) isomorfo ao reticulado original, b) subconjunto do reticulado original e c) não isomorfo e não subconjunto do reticulado original.
5. **Índices de qualidade:** a) perda de fidelidade, b) perda de representatividade e c) perda descritiva.
6. **Aspecto algorítmico:** a) exato, b) heurístico e c) alta complexidade.
7. **Classe:** a) eliminação de informação redundante, b) simplificação e c) seleção.

A primeira dimensão diz qual é o ponto de partida da técnica de redução: o contexto formal [10, 15, 16, 144, 145], o conjunto de conceitos formais [99, 185], o reticulado conceitual [69, 127, 149] ou alguma extensão da AFC como, por exemplo, contexto formal *fuzzy* [125], contexto formal de decisão [126, 161], etc.

A segunda dimensão informa se a técnica utiliza informação adicional (conhecimento prévio)⁵ ao contexto formal para auxiliar no processo de redução [40] ou não [69, 113, 144].

Durante o processo de redução determinadas técnicas alteram o conjunto de objetos [36, 40, 62, 98, 144], o conjunto de atributos [164, 185] e/ou o conjunto de incidências [13, 62, 184]. Essas possibilidades são identificadas na terceira dimensão.

Enquanto algumas técnicas geram um reticulado que é isomorfo ao original [107, 109, 116], outras não; elas simplesmente selecionam conceitos formais específicos considerados relevantes por algum critério preestabelecido [99, 136, 137, 149, 184] ou modificam o contexto formal e obtêm um reticulado que não é isomorfo ao original nem um subconjunto de conceitos formais [40, 62]. Essas três possibilidades caracterizam a quarta dimensão.

Na quinta dimensão, medidas de qualidade do reticulado reduzido são levadas em consideração. As técnicas podem induzir perdas de fidelidade [16, 17, 19, 40, 144], perdas de representatividade [40, 98, 144] e/ou perdas descritivas [16, 17, 19, 40, 144, 184]. Tais índices são definidos no Capítulo 5.

A sexta dimensão indica se a técnica utiliza um algoritmo exato [16, 137] ou um algoritmo baseado em alguma heurística [7, 107]; e se o algoritmo possui alta complexidade [99, 185] ou não [40, 96, 149]. Aqui, por exemplo, técnicas que requerem a geração de todos os conceitos formais antes da redução são consideradas técnicas cujo algoritmo tem alta complexidade.

Finalmente, na última dimensão, a classe a que pertence a técnica de redução é identificada. Existem três classes, conforme visto na Seção 3.1: eliminação de informação redundante [107, 116, 126, 143, 161, 163, 185], simplificação [19, 34, 36, 40, 62, 98, 130, 144] e seleção [4, 70, 99, 136, 137, 149].

⁵Do inglês, *background knowledge*.

3.2.1 Analisando as Técnicas de Redução através da AFC

A Tabela 3.1 sumariza as principais técnicas para redução de reticulados conceituais mediante um contexto formal em que os objetos representam as técnicas e os atributos representam as características das técnicas em cada uma das sete dimensões descritas na seção anterior.

Tabela 3.1. Contexto formal para as principais técnicas de redução.

	Características																Classes					
	1a	1b	1c	1d	2a	2b	3a	3b	3c	4a	4b	4c	5a	5b	5c	6a	6b	6c	c1	c2	c3	
Zhang et al. (2005)		x				x		x	x	x				x	x	x		x		x		
Wang & Ma (2006)		x				x		x	x	x				x	x	x		x		x		
Wang & Zhang (2008b)		x				x		x	x	x				x	x	x		x		x		
Wei et al. (2008)				x		x		x	x	x				x	x	x		x		x		
Liu & Mi (2008)		x				x		x	x	x				x	x	x		x		x		
Wang & Zhang (2008a)				x		x		x	x	x				x	x	x		x		x		
Qi (2009)		x				x		x	x	x				x	x	x		x		x		
Wang & Zhang (2010)		x				x		x	x	x				x	x	x		x		x		
Wang et al. (2010)		x				x		x	x	x				x	x		x		x		x	
Pei & Mi (2011)				x		x		x	x	x				x	x	x		x		x		
Pei et al. (2011)				x		x		x	x	x				x	x		x		x		x	
Li et al. (2011b)				x		x		x	x	x				x	x	x		x		x		
Li et al. (2011a)				x		x		x	x	x				x	x		x		x		x	
Medina (2012)		x				x		x	x	x				x	x	x		x		x		
Li et al. (2012)				x		x		x	x	x				x	x		x		x		x	
Li et al. (2013)				x		x		x	x	x				x	x	x		x		x		
Gajdos et al. (2004)	x					x		x				x	x	x	x	x						x
King (2004)	x				x		x		x			x	x	x	x	x						x
Belohlavek & Sklenar (2005)	x				x			x	x				x	x	x	x						x
Snásel et al. (2007)	x					x		x				x	x	x	x	x						x
Kumar & Srinivas (2010a)	x					x		x	x			x	x	x	x		x					x
Gél (2011)		x	x			x		x					x	x	x	x						x
Dias & Vieira (2010)	x				x		x						x	x	x	x		x				x
Kumar (2012)	x					x		x	x			x	x	x	x		x					x
Kuznetsov (2007)		x	x			x	x	x	x		x			x	x	x	x		x			x
Babin & Kuznetsov (2012)		x	x			x	x	x	x		x			x	x	x		x				x
Arévalo et al. (2007)	x					x	x	x	x		x			x	x	x	x					x
Rice & Siff (2001)	x					x	x	x	x		x			x	x	x	x					x
Stumme et al. (2002)				x		x	x	x	x		x			x	x	x	x					x
Pernelle et al. (2002)				x		x	x	x	x		x			x	x	x	x					x
Belohlavek et al. (2004a)	x				x		x	x	x		x			x	x	x	x					x
Belohlavek et al. (2004b)	x					x		x	x	x				x	x	x	x					x
Ventos & Soldano (2005)			x			x	x	x	x		x			x	x	x	x					x
Belohlavek & Vychodil (2006)	x					x	x	x	x		x			x	x	x	x					x
Boulicaut & Besson (2008)	x					x	x	x	x		x			x	x	x	x					x
Belohlavek & Vychodil (2009)	x				x		x	x	x		x			x	x	x	x					x
Riadh et al. (2009)	x					x	x		x			x		x	x	x	x					x
Soldano et al. (2010)			x			x	x	x	x		x			x	x	x	x					x
Belohlavek & Macko (2011)	x	x				x		x	x	x				x	x	x		x		x		x
Zhang et al. (2012)		x	x			x		x	x	x				x	x	x		x				x

Ponto de partida: 1a contexto formal, 1b conceitos formais, 1c reticulado conceitual e 1d extensão da AFC.
 Utiliza conhecimento adicional: 2a sim e 2b não.
 Alterações no contexto formal: 3a altera o conjunto de objetos, 3b altera o conjunto de atributos 3c altera o conjunto de incidências.
 Reticulado conceitual resultante: 4a isomorfo ao reticulado original, 4b subconjunto do reticulado original e 4c não isomorfo e não subconjunto do reticulado original.
 Índices de qualidade: 5a perda de fidelidade, 5b perda de representatividade, e 5c perda descritiva.
 Aspecto algorítmico: 6a exato, 6b heurístico e 6c alta complexidade.
 Classe: c1 eliminação de informação redundante, c2 simplificação e c3 seleção.

A Figura 3.1 mostra o diagrama de linhas do reticulado conceitual originado do

contexto formal da Tabela 3.1, enquanto as Figuras 3.2 (a), 3.2 (b) e 3.2 (c) apresentam os diagramas de linhas dos reticulados conceituais correspondentes a cada classe de técnicas: eliminação de informação redundante, simplificação e seleção.

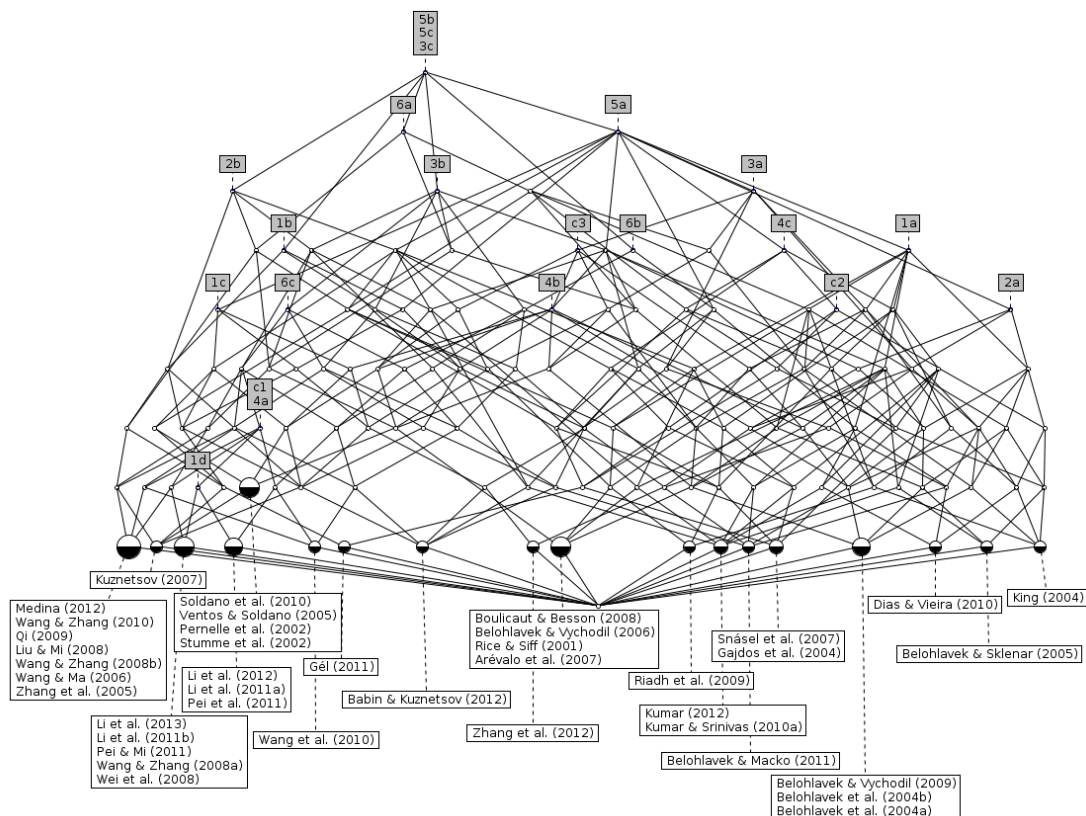
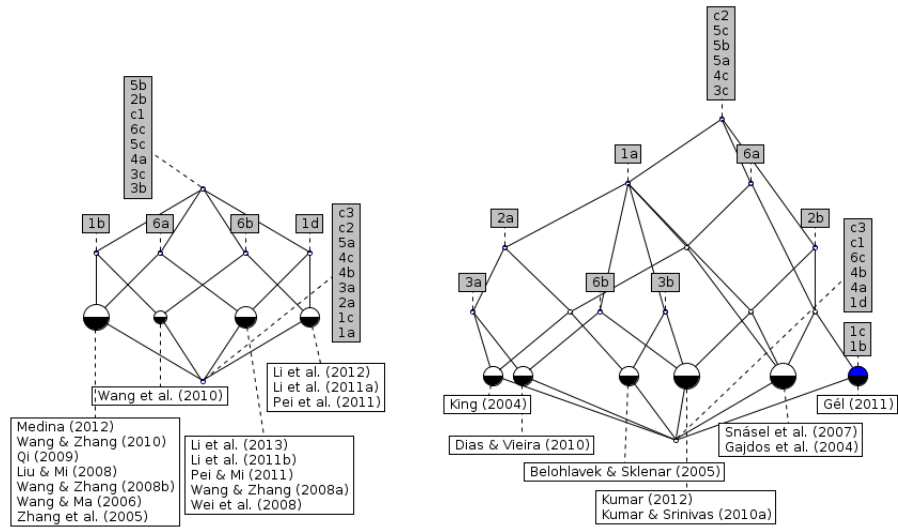


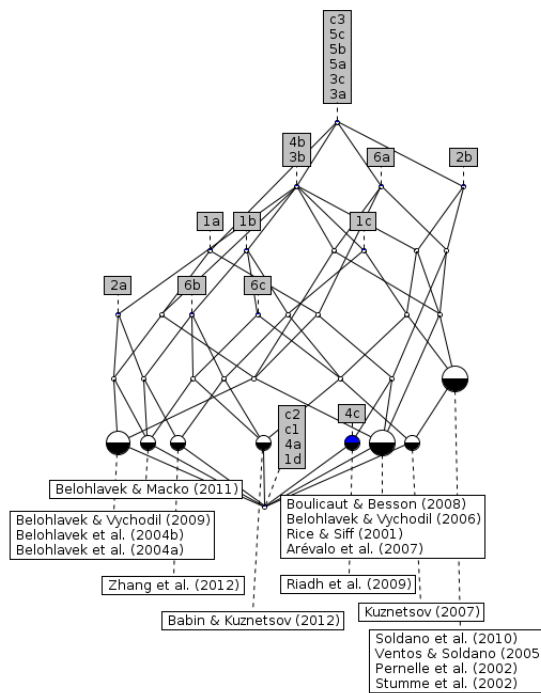
Figura 3.1. Reticulado conceitual do contexto formal da Tabela 3.1.

Inicialmente, note-se, na Figura 3.1, que todas as técnicas de redução analisadas alteram o conjunto de incidências (atributo 3c) e causam alguma perda descritiva (5c). Já nas técnicas de eliminação de informação redundante, a eliminação de atributos ou objetos obviamente altera o conjunto de incidências. De forma similar, técnicas da classe de simplificação reduzem a relação de incidência; por exemplo, através do agrupamento de objetos “similares”. Técnicas da classe de seleção também alteram o conjunto de incidências quando selecionam conceitos formais, objetos ou atributos, uma vez que alguns desses elementos são desconsiderados. Toda técnica de redução



(a) *Eliminação de informação redundante.*

(b) *Simplificação.*



(c) *Seleção.*

Figura 3.2. Reticulado conceitual para cada classe.

apresenta algum grau de perda descritiva, mesmo técnicas da classe de eliminação de informação redundante, pois estas podem cortar atributos.

A Figura 3.2 (a) mostra o reticulado conceitual restrito à classe de eliminação de informação redundante. Alguns grupos (de objetos e atributos) são visualmente perceptíveis na figura. Eles mostram, claramente, que as 16 técnicas dessa classe ostentam as seguintes características:

dimensão 1: o ponto de partida não é o contexto formal ou o reticulado conceitual; um grupo de 8 técnicas tem o conjunto de conceitos como seu ponto de partida, e as outras 8 técnicas partem de uma extensão da AFC;

dimensão 2: nenhuma técnica utiliza conhecimento prévio;

dimensão 3: todas as técnicas alteram o conjunto de atributos e a relação de incidência, mas nenhuma altera o conjunto de objetos;

dimensão 4: o reticulado resultante é isomorfo ao reticulado original;

dimensão 5: nenhuma técnica tem perda de fidelidade, embora todas apresentem perdas de representatividade e descritivas; e

dimensão 6: todas as técnicas apresentam alta complexidade; 12 delas possuem algoritmos exatos e outras 4 possuem algoritmos heurísticos.

Portanto, sendo semelhantes em outras características, eis o que distingue as técnicas desta classe:

- metade das técnicas têm como ponto de partida os conceitos formais e a outra metade se inicia de alguma extensão da AFC; e
- doze técnicas possuem algoritmos exatos e outros quatro possuem métodos heurísticos.

A Figura 3.2 (b) mostra o reticulado conceitual restrito a técnicas da classe de simplificação. São 8 técnicas, com as seguintes características:

dimensão 1: todas as técnicas têm como ponto de partida o contexto formal, exceto uma técnica cujo ponto de partida é o conjunto de conceitos formais e o reticulado conceitual;

dimensão 2: 3 técnicas utilizam conhecimento adicional ao contexto formal e 5 não utilizam;

dimensão 3: todas as técnicas alteram o conjunto de incidências; 2 delas alteram o conjunto de objetos e 3 o conjunto de atributos;

dimensão 4: nenhuma técnica possui o reticulado conceitual resultante isomorfo ou subconjunto do reticulado conceitual original;

dimensão 5: todas as técnicas apresentam perdas de fidelidade, representatividade e descritiva; e

dimensão 6: 3 técnicas possuem métodos heurísticos e todas as outras 5 possuem algoritmos exatos.

Em resumo, a maioria das técnicas de simplificação utiliza o contexto formal como ponto de partida, produzem reticulados conceituais não isomorfos nem subconjuntos do reticulado conceitual original, e possuem algoritmos exatos. Algumas técnicas utilizam conhecimento adicional ao contexto formal para o processo de simplificação e outras não.

Finalmente, a Figura 3.2 (c) apresenta o reticulado conceitual restrito a 16 técnicas da classe de seleção. A análise do reticulado mostra as seguintes características:

dimensão 1: 9 técnicas utilizam o contexto formal como ponto de partida, 7 técnicas utilizam o reticulado conceitual e quatro, o conjunto de conceitos formais; nenhuma técnica utiliza extensão da AFC;

dimensão 2: apenas 5 técnicas utilizam conhecimento adicional ao do contexto formal;

dimensão 3: todas as técnicas alteram o conjunto de objetos, atributos e incidências, exceto uma delas que não altera o conjunto de atributos;

dimensão 4: nenhum reticulado conceitual resultante é isomorfo ao reticulado original, mas todos são subconjuntos do mesmo;

dimensão 5: todas as técnicas apresentam perdas de fidelidade, representatividade e descritiva; e

dimensão 6: 13 técnicas possuem algoritmos exatos, 3 possuem métodos heurísticos e 3 apresentam alta complexidade.

Consequentemente, a maioria das técnicas se inicia do contexto formal, não utiliza conhecimento adicional e utiliza algoritmos exatos.

Seguem algumas observações finais acerca das três classes de técnicas. Muitas técnicas de eliminação de informação redundante utilizam o conceito de matriz de discriminação entre conceitos formais, porém tal característica aumenta proibitivamente a complexidade das técnicas. Para contornar o problema, elas utilizam extensões da AFC associadas a métodos heurísticos. As técnicas de eliminação de informação redundante, não reduzindo o tamanho do reticulado conceitual, são consideradas de pouca praticidade. De fato, nenhuma aplicação prática de utilização dessas técnicas foi observada na literatura. Com relação às técnicas de simplificação, até onde se tem conhecimento apenas o trabalho de Gély [69] emprega técnicas de agrupamento de conceitos formais utilizando a estrutura do reticulado conceitual, enquanto as demais utilizam o contexto formal. Já com relação às técnicas de seleção, a de Zhang et al. [184] seleciona conceitos cuja relevância é medida a partir da atribuição de pesos aos atributos, assim como muitas outras técnicas; entretanto, conceitos formais virtuais são criados para manter a completude do reticulado conceitual. Também cabe ressaltar que o que distingue o trabalho sobre o índice de estabilidade, originalmente proposto por Kuznetsov [99] e também explorado em trabalhos subsequentes [101, 138], do trabalho mais recente de Babin & Kuznetsov [7], é que o último emprega métodos heurísticos. A complexidade do cálculo do índice de estabilidade é alto devido à necessidade de se calcular os subconjuntos da extensão ou intenção de cada conceito formal. Babin & Kuznetsov [7] propõem uma aproximação utilizando cadeias de Markov.

Com vistas a finalizar a análise a partir do contexto formal da Tabela 3.1, determinou-se todas as implicações mais concisas do contexto formal que indicassem condições suficientes (em termos dos atributos das técnicas em análise) para que cada técnica pertencesse a uma das três classes de técnicas de redução. Mais precisamente, sendo \mathcal{I} o conjunto das implicações do contexto formal da referida tabela, para cada $i \in \{1, 2, 3\}$ todas as implicações $P \rightarrow ci \in \mathcal{I}$ tais que $ci \notin P$ e não existe $Z \subset P$ tal que $Z \rightarrow ci \in \mathcal{I}$ são apresentadas na Tabela 3.2.

O primeiro conjunto diz que, para as técnicas em análise, se o reticulado concei-

classe $c1$	classe $c2$	classe $c3$
$4a \rightarrow c1$	$1a, 6b, 2b \rightarrow c2$	$3a, 2b \rightarrow c3$
$1d \rightarrow c1$	$3b, 4c \rightarrow c2$	$3a, 3b \rightarrow c3$
	$4c, 6b \rightarrow c2$	$4b \rightarrow c3$
	$2a, 4c \rightarrow c2$	
	$1b, 4c \rightarrow c2$	
	$1c, 4c \rightarrow c2$	

Tabela 3.2. Implicações com condições suficientes e concisas para cada classe.

tual obtido é isomorfo ao original ou se é utilizada alguma extensão da AFC, então a técnica é da classe de eliminação de informação redundante. A primeira implicação para $c2$ mostra que se o ponto de partida é o contexto formal e conhecimento adicional não é utilizado e um método heurístico é utilizado, então a técnica é de simplificação. As outras implicações para $c2$ dizem que se o reticulado resultante não é isomorfo ao original e nem seu subconjunto e (1) o conjunto de atributos é alterado ou (2) é utilizada heurística ou (3) conhecimento adicional é utilizado ou (4) o ponto de partida é o conjunto de conceitos formais ou (5) o ponto de partida é o reticulado conceitual, então a técnica é da classe de simplificação. As últimas três implicações dizem que uma das seguintes condições é suficiente para implicar que uma técnica de seleção é utilizada: (1) o conjunto de objetos é alterado e conhecimento adicional não é utilizado ou (2) os conjuntos de objetos e de atributos são alterados ou (3) o reticulado resultante é subconjunto do reticulado conceitual original.

A análise foi feita para um conjunto específico de técnicas, aquelas que se considerou mais representativas até o momento em que foi feita a análise. Evidentemente, o acréscimo de novas técnicas pode alterar um pouco os resultados apresentados.

3.3 Trabalhos Relacionados

Priss & Old [132] identificam quatro classes de técnicas para redução de reticulados conceituais, denominadas de *redução visual*, *dimensionamento*, *poda* e *decomposição*. Os trabalhos considerados são aqueles restritos à geração de “boas” visualizações do reticulado conceitual no contexto de ontologias. As técnicas de redução são lá chamadas

de técnicas para *capina de dados*. As classes de *redução visual* e de *dimensionamento* são similares ao que se denomina aqui classe de eliminação de informação redundante, enquanto que as classes de *poda* e de *decomposição* são similares às aqui denominadas de classes de simplificação e de seleção. Note-se que, ao contrário da classificação realizada nesta tese, a deles é restrita a um tipo de aplicação.

Kuznetsov & Poelmans [105], Poelmans et al. [129] e Poelmans et al. [128] exibem resumos atuais das principais aplicações da AFC nos últimos anos; alguns trabalhos são relacionados a redução. O primeiro capítulo de Cerf [32] apresenta o estado da arte de restrições utilizadas em conjuntos de itens fechados, que são equivalentes a conceitos formais.

3.4 Conclusões

Neste capítulo foi feita uma classificação das técnicas de redução de reticulados conceituais, na qual foram identificadas três classes: eliminação de informação redundante, simplificação e seleção. Técnicas de eliminação de informação redundante produzem reticulados conceituais isomorfos ao original. Já as técnicas da classe de simplificação produzem abstrações do reticulado conceitual, ou seja, simplificações que objetivem preservar apenas aspectos considerados relevantes. Finalmente, técnicas de seleção atuam selecionando conceitos formais, objetos ou atributos com base em algum critério.

A exigência do isomorfismo limita o nível de redução possível de ser obtido pelas técnicas da classe de eliminação de informação redundante, o que inviabiliza o uso de tais técnicas em muitas aplicações.

Técnicas da classe de simplificação transformam o espaço dos conceitos formais em um espaço menor e, por conseguinte, possuem um potencial de aplicação maior em situações que exijam reduções mais drásticas. Obviamente, que essa característica pode levar a uma degradação da qualidade do reticulado conceitual reduzido com relação à do original. Por outro lado, como as técnicas são usualmente aplicadas diretamente no contexto formal, elas têm uma complexidade que as tornam viáveis para manipulação de conhecimento representável por meio de contextos formais realmente grandes, que levem a reticulados bastante complexos.

Técnicas da classe de seleção trabalham podando o espaço dos conceitos formais, utilizando algum critério de relevância; por exemplo, aplicando alguma função objetivo

que elimina caminhos no reticulado conceitual que sejam considerados irrelevantes. Um representante típico dessa classe é técnica de conceitos formais frequentes proposta por Stumme et al. [149]. A desvantagem é que, eventualmente, conceitos formais importantes possam ser perdidos. Algumas técnicas enumeram todos os conceitos formais e, em seguida, aplicam algum critério para escolha dos conceitos formais considerados relevantes [13, 101, 184]. Como todo o espaço de conceitos formais é explorado, uma técnica com base nessa abordagem pode apresentar um alto custo computacional. Além disso, o subconjunto de conceitos formais tido como relevantes, pode não compor um reticulado conceitual.

Um total de 40 técnicas, selecionadas dentre as principais da literatura, foram analisadas e classificadas com base em sete dimensões: ponto de partida; uso de conhecimento prévio; alterações no contexto formal; reticulado conceitual resultante, índices de qualidade; aspectos algorítmicos; e, classe a que a técnica pertence. Um contexto formal foi construído sumarizando os resultados relativos às técnicas selecionadas, e uma análise foi conduzida utilizando a própria AFC. Em particular, determinou-se todas as implicações mais concisas do contexto formal que indicassem condições suficientes para que cada técnica pertencesse a uma das três classes de técnicas de redução.

É óbvio que as maiores reduções, aquelas que podem reduzir substancialmente o reticulado conceitual, são as obtidas pelas técnicas de simplificação e seleção. Entretanto, as técnicas de simplificação podem ser consideradas um tanto “perigosas”, pois podem modificar substancialmente o conjunto de conceitos formais. É necessário assegurar que o processo que leva a tais modificações preserve as partes essenciais do reticulado conceitual. Mesmo que as técnicas de seleção sejam muito interessantes, visto que reduzem o espaço conceitual, o caminhar nesse espaço deve ser feito de forma que os conceitos verdadeiramente relevantes sejam os atingidos.

Capítulo 4

Redução Baseada na Similaridade entre Objetos

A análise das principais técnicas para redução de reticulados conceituais, realizada no Capítulo 3 e por Dias & Vieira [44], mostra a existência de uma lacuna em técnicas de redução que, efetivamente, simplifiquem o reticulado conceitual com base em conhecimento acerca da aplicação específica, e que apresentem um desempenho computacional satisfatório. A análise indica uma concentração na concepção de técnicas para seleção de conceitos formais. Nestas, o subconjunto de conceitos formais selecionados pode não formar um reticulado conceitual [138]. Em alguns casos, um reticulado completo é formado por meio de um pós-processamento [84, 184]. Por outro lado, determinadas técnicas necessitam do acesso a todo o reticulado conceitual original para seleção dos conceitos formais relevantes [13, 101, 184]. No entanto, nem sempre é possível gerar, previamente, todos os conceitos formais, seja por problema de tempo, de espaço ou ambos.

Este capítulo apresenta uma técnica de redução que gera um reticulado conceitual simplificado, sem a necessidade de um pós-processamento, que não exige a geração prévia dos conceitos formais (ou reticulado conceitual) originais, e que possui desempenho computacional satisfatório. Ao basear a simplificação no agrupamento de objetos similares, a técnica proporciona o descarte de conhecimento não relevante para a aplicação específica. A análise das implicações pré- e pós-redução, usando a metodologia a ser apresentada no próximo capítulo, aponta que as simplificações proporcionam generali-

zações. A técnica configura um desdobramento de resultado apresentado por Dias [39], cujas características, virtudes e limitações são apresentadas em Dias & Vieira [40, 42] e Dias et al. [46] e sumarizadas a seguir.

4.1 Alguns pressupostos

No desenvolvimento da técnica assumem-se alguns requisitos comuns a todas as técnicas existentes, e mais alguns que, conjuntamente, buscam cobrir lacunas observadas no estado da arte. Os primeiros são:

- evitar criar novos objetos, atributos ou incidências;
- preservar a hierarquia conceitual ao máximo; e
- manter o conhecimento representado no reticulado conceitual consistente com o original.

Em suma, todos os métodos de redução buscam cortar elementos, todavia procurando não introduzir inconsistências com relação ao contexto formal original. A técnica aqui proposta atende, adicionalmente, ao conjunto dos seguintes requisitos:

1. evitar acesso ao reticulado conceitual completo;
2. obter um reticulado conceitual simplificado;
3. cortar características não essenciais do reticulado original, chegando a um reticulado que exponha apenas o essencial;
4. permitir o uso de conhecimento prévio, relativo ao domínio do problema, para discernir o que é essencial do que não é; e
5. apresentar uma complexidade computacional razoável.

O primeiro requisito é um dos pontos importantes com relação a atender ao requisito 5, o qual exige que a técnica seja computacionalmente viável, visto que o tamanho do reticulado original é o que justifica a aplicação de uma técnica de redução em muitos casos. De modo a se obter um reticulado conceitual (e não um conjunto desestruturado

de dados) viável de ser posteriormente explorado, qualquer que seja a aplicação, tem-se o requisito 2. Os requisitos 3 e 4 afirmam a necessidade de simplificar o contexto formal, com base em conhecimento do que é relevante na aplicação que se tenha em mente.

4.2 Junção Baseada na Similaridade de Objetos - JBOS

A técnica de *junção baseada na similaridade de objetos* (JBOS)¹ é aplicada a partir do contexto formal, atendendo, portanto, ao requisito 1. O contexto formal resultante é obtido pela junção de objetos “similares”, i.e., pela substituição de cada grupo de objetos similares, por um único representante. Com isto, decorre uma simplificação do contexto formal e, conseqüentemente, do reticulado conceitual correspondente, o que atende ao requisito 2. Naturalmente, o aspecto chave aqui é a noção de similaridade. Para sua definição, um peso é criado para cada atributo, que represente a relevância do mesmo. Como a estrutura do reticulado conceitual é derivada do arranjo dos atributos e objetos, é esperado que a junção de objetos similares preserve, em certo grau, uma porção do reticulado original, enquanto proporcione a preservação do conhecimento mais relevante. Ambos os aspectos dependem dos pesos atribuídos aos atributos. Tal processo de aplicar pesos nos atributos é um meio de se inserir conhecimento prévio, conforme requerido pelo requisito 4, de forma a preservar o que se considere mais relevante, como enunciado pelo requisito 3. JBOS difere de outras técnicas que fazem uso de conhecimento prévio [13, 18, 19, 184] com relação às seguintes virtudes: realiza a redução a partir do contexto formal e obtém, necessariamente, um reticulado conceitual simplificado.

No restante desta seção, seja (G, M, I) o contexto formal a partir do qual um contexto formal reduzido deve ser produzido.

A cada atributo $m \in M$ é associado um peso w^m tal que $0 \leq w^m \leq 1$. Tal peso, definido pelo usuário, deve representar a *relevância* do atributo, variando de 0 (relevância nenhuma) a 1 (relevância absoluta). A similaridade entre dois objetos $g, h \in G$ também se expressa mediante um número real entre 0 (totalmente distintos) e 1 (totalmente similares), assim definido:

¹Do inglês: *Junction Based on Objects Similarity*.

$$\text{sim}(g, h) = \frac{\sum_{m \in M} \mu(g, h, m)}{\sum_{m \in M} w^m}; \mu(g, h, m) = \begin{cases} w^m, & \text{se } (g, m) \in I \leftrightarrow (h, m) \in I \\ 0, & \text{caso contrário.} \end{cases} \quad (4.1)$$

Em outras palavras, a similaridade entre dois objetos g e h ($g, h \in G$) é determinada pela soma ponderada dos pesos dos atributos em que ambos concordem (ambos possuam ou ambos não possuam).

O Algoritmo 1 usa uma matriz de similaridade, como definida por 4.1, para fazer agrupamentos de objetos similares, de tal forma que dois objetos fiquem no mesmo grupo somente se tiverem um certo “grau de similaridade”. Tal grau é expresso pelo usuário mediante um índice de similaridade ε . Ademais, um índice α , também definido pelo usuário, impõe um limite superior para o número de objetos de cada grupo. Dois objetos $g, h \in G$ são considerados similares pelo algoritmo, e, portanto, passíveis de permanecer em um mesmo grupo, se e somente se $\text{sim}(i, j) \geq \varepsilon$. O número α serve apenas como mais uma opção de corte para o usuário. A chamada de função *EscolheObjetoEm*(G), na linha 3 do algoritmo, seleciona qualquer um dos objetos do conjunto de objetos G (sendo, portanto, um ponto de não determinismo). Já a chamada *ObjetoMaiorSimilaridade*(sim, g, G), na linha 7, retorna um objeto de G dentre os mais similares ao objeto g ; se houver mais de um, escolhe qualquer um (outro ponto de não determinismo). O conjunto de agrupamentos de objetos similares é denotado por γ no algoritmo. Note-se que, mesmo a *similaridade* (dois objetos g e h são *similares* um ao outro se e somente se $\text{sim}(g, h) \geq \varepsilon$) sendo uma relação de tolerância², não necessariamente transitiva, o algoritmo calcula um conjunto “maximal” H , tal que g seja similar a h para todo $g, h \in H$. Portanto, o algoritmo produz, a partir de uma relação de tolerância, uma relação de equivalência em que as classes de equivalência são os agrupamentos que retorna.

Seja (G, M, I) o contexto formal original, γ o conjunto obtido pelo Algoritmo 1 a partir de G , sim (este, definido a partir de pesos dos atributos em M), ε e α . O contexto formal reduzido (G_r, M_r, I_r) obtido pela técnica JBOS é aquele em que:

- $G_r = \gamma$ (cada grupo é considerado um objeto);

²Uma relação de tolerância é uma relação binária reflexiva e simétrica [37].

Algoritmo 1 Determinação de grupos de objetos similares

Entrada: conjunto de objetos G , matriz de similaridade sim , índice de similaridade ε e índice de corte α .

Saída: Conjunto de grupos de objetos similares γ .

```

1:  $\gamma = \emptyset$ 
2: enquanto  $G \neq \emptyset$  faça
3:    $g = \text{EscolheObjetoEm}(G)$ 
4:    $G = G \setminus \{g\}$ 
5:    $H = \{g\}$ 
6:   enquanto  $G \neq \emptyset$  e  $|H| < \alpha$  e R faça
7:      $h = \text{ObjetoMaiorSimilaridade}(sim, g, G)$ 
8:     se  $sim(h, f) \geq \varepsilon$  para todo  $f \in H$  então
9:        $G = G \setminus \{h\}$ 
10:       $H = H \cup \{h\}$ 
11:    senão
12:      pare R
13:    fim se
14:  fim enquanto
15:   $\gamma = \gamma \cup \{H\}$ 
16: fim enquanto
17: retorne  $\gamma$ 

```

- $M_r = \bigcup \{ \bigcap \{g' | g \in H\} | H \in \gamma \}$ (cada objeto do contexto reduzido $H \in \gamma$ terá os atributos de M que são comuns a todos os objetos de H ; em outras palavras, os atributos de H serão $\bigcap \{g' | g \in H\}$, em que o operador de derivação obedece a I); e
- $I_r = \{ (H, m) \in \gamma \times M_r | m \in \bigcap \{g' | g \in H\} \}$.

Exemplo 9. Seja o contexto formal apresentado na Tabela 2.1, $\varepsilon = 0.65$, $\alpha = 3$ e o seguinte conjunto de pesos para os atributos: $w^a = 1$, $w^b = 0$, $w^c = 1$, $w^d = 0.3$, $w^e = 0.2$ e $w^f = 0.5$. Observe-se que, independentemente da importância traduzida pelos pesos dos atributos, a técnica agrupa os objetos 2 e 5, pois eles possuem exatamente os mesmos atributos ($2' = 5'$). Como o atributo b é o único que distingue o objeto 1 dos objetos 2 e 5 e ele tem peso 0, ele estará no mesmo grupo que os dois. Fazendo-se os cálculos, vê-se que $H = \{1, 2, 5\}$ é um dos grupos obtidos, com $H' = \{a\}$.

O conjunto completo de agrupamentos obtido é $\gamma = \{\{1, 2, 5\}, \{3\}, \{4\}\}$, redundando no contexto formal reduzido da Tabela 4.1. A Figura 4.1 mostra o reticulado

original (à esquerda) juntamente com o reduzido (à direita). Note-se como diminuiriam os números de objetos, de incidências e de conceitos, tornando o contexto formal e o reticulado correspondente mais simples. O atributo b , que distinguia o objeto 1 dos objetos 2 e 5, é desprezado (sua relevância, segundo o usuário, é zero para a aplicação que ele tem em mente). No reticulado reduzido, isto redundou na substituição dos conceitos $(\{1, 3\}, \{a, b\})$ e $(\{1, 2, 3, 5\}, \{a\})$ do reticulado original pelo conceito $(\{[1, 2, 5]\}, \{a\})$. Aqui, $[1, 2, 5]$ é um objeto que substitui os três objetos 1, 2 e 5. \square

Tabela 4.1. Contexto formal reduzido obtido pela técnica JBOS.

Obj/Att	a	b	c	d	e	f
[1, 2, 5]	x					
3	x	x	x	x		x
4		x		x	x	x

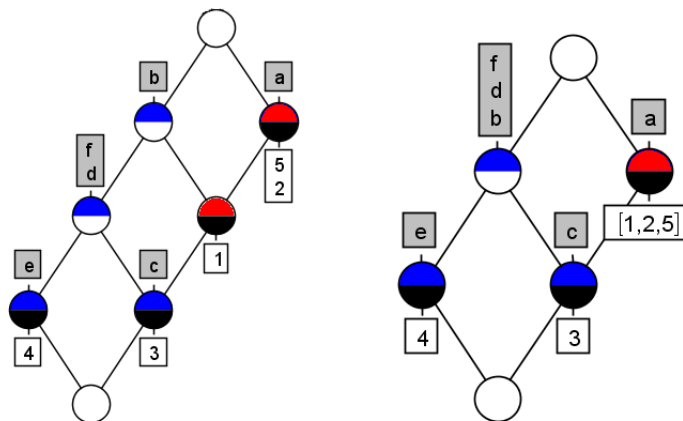


Figura 4.1. Reticulados original e reduzido obtido pela técnica JBOS.

Determinar a matriz de similaridade requerida pelo Algoritmo 1 tem uma complexidade $O(|G|^2|M|)$. Como na linha 7 do algoritmo cada objeto g é comparado com os demais, e no teste da linha 8 um objeto é comparado com menos de α outros, segue-se que a complexidade do algoritmo é $O(|G|^3)$ no pior caso, se dada a matriz de similaridade.

4.3 Estudo de Caso

A seguir, realiza-se um estudo de caso em que se mostra a aplicação da técnica JBOS, assim como outras três técnicas de redução muito citadas na literatura, em uma aplicação real. A aplicação consta da extração de regras que modelem alguns aspectos do processo siderúrgico de laminação a frio. O objetivo aqui é mostrar que as regras (no caso, implicações) refletem o comportamento do processo, mesmo após uma redução.

Além da técnica JBOS, são estudadas as viabilidades de se usar as técnicas de seleção baseadas no índice de estabilidade, em conceitos formais frequentes e em geradores mínimos. As três técnicas foram selecionadas por apresentarem bons resultados em aplicações relacionadas à representação de conhecimento [13, 61, 81, 138, 149]. Das quatro técnicas, apenas JBOS e seleção de conceitos frequentes não pressupõem a construção do conjunto completo de conceitos. As vantagens e desvantagens de cada técnica de redução são ressaltadas, em especial as da técnica JBOS.

4.3.1 Processo Siderúrgico de Laminação a Frio

O processo siderúrgico de laminação a frio tem como objetivo a produção de tiras de aço, cuja qualidade é medida principalmente pela sua espessura final. Para o processo de produção existem dois equipamentos, a desenroladeira, que fornece a matéria-prima de uma bobina ao laminador, e a enroladeira, que retira a tira de aço, colocando-a novamente em forma de bobina. A partir de um conjunto de cilindros, localizados acima e abaixo da tira de aço, aplica-se uma pressão sobre a tira, provocando sua deformação. Cada cilindro encontra-se preso em uma estrutura de aço chamada carcaça; o conjunto cilindro e carcaça forma uma caldeira, conforme ilustrado na Figura 4.2. Uma descrição detalhada sobre o processo de laminação a frio pode ser encontrada em Zárate [178].

Existem muitos fatores que influenciam no processo de laminação, entender do comportamento das variáveis envolvidas no mesmo é extremamente importante para obtenção de um produto final de melhor qualidade. Além disso, na indústria siderúrgica o número de especialistas é pequeno e o conhecimento sobre o processo precisa ser passado para os operadores menos experientes. Desta forma, uma representação dos princípios do processo de forma simples, no sentido de representar o comportamento de uma equação não linear de difícil entendimento, é apropriada.

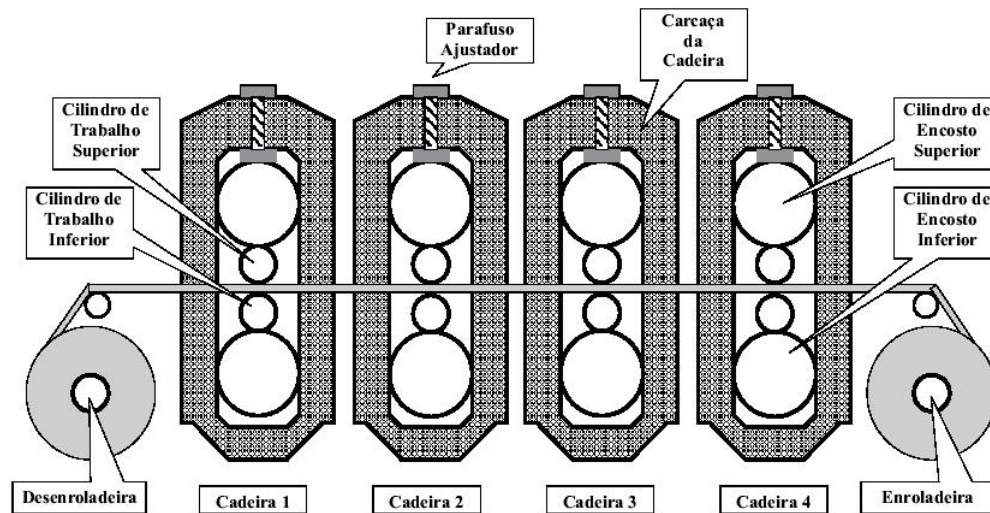


Figura 4.2. Exemplo de laminador utilizado no processo siderúrgico de laminação a frio. Fonte: [178]

O modelo teórico do processo de laminação calcula a carga P , pressão exercida pelos cilindros, através de uma equação $P = f(h_i, h_o, \mu, t_b, t_f, \bar{y}, E, R)$, em que h_i é a espessura de entrada, h_o é a espessura de saída, μ é o coeficiente de atrito, t_b é a tensão à ré, t_f é a tensão à frente, \bar{y} é a tensão de escoamento, E é o módulo de Young da tira e R_c é o raio do cilindro. Esta equação envolve relações não lineares que conduzem geralmente a uma equação não linear de difícil análise ou solução numérica [178]. É o caso do Modelo de Alexander [2], que é considerado um dos mais completos modelos teóricos para o processo de laminação, que requer, todavia, um esforço computacional significativo. Na análise fundamentada na AFC, que se segue, consideraram-se os parâmetros P , h_i , h_o , μ , t_b , t_f e \bar{y} . No contexto de redes neurais e AFC, os referidos parâmetros foram empregados e resultados satisfatórios foram obtidos [46, 181, 182].

4.3.2 Dados para Análise do Processo de Laminação

Uma base de dados representando o processo de laminação, contendo dez milhões de registros, foi disponibilizada pelo *Data Science Research Group* (DSRgroup)³. Ela foi produzida através de simulações do Modelo de Alexander [178].

³<http://www.icei.pucminas.br/projetos/dsrgroup/>

Dez milhões de registros é uma base de dados de referência que apresenta um número elevado de registros, o que dificulta a aplicação da AFC. Assim, faz-se necessária a seleção de amostras significativas dos dados. Em Zárte et al. [183] são propostas técnicas estatísticas para construção de conjuntos de dados representativos, mantendo-se a capacidade de generalização do modelo construído utilizando os dados. A partir dessas sugestões foi utilizada a equação $n = (z/e)^2 \times (f \times (1 - f))$, em que n é o tamanho da amostra, z é o nível de confiança, e é o erro em torno da média e f é a proporção da população.

Fixando-se z em 90% (1.645), f em 0.5 e e em 2%, tem-se $n \cong 1691$ registros. Foram selecionadas 10 amostras disjuntas de 1691 registros utilizando o seguinte procedimento:

1. Para cada parâmetro foram selecionados registros que continham os valores máximo, mínimo e próximo da média.
2. Foi escolhido o ponto de operação utilizado pelo Modelo de Alexander para gerar os dados.
3. Os elementos restantes foram selecionados aleatoriamente com a finalidade de se alcançar o total de 1691 registros.

No contexto de redes neurais artificiais empregou-se esse procedimento, visando selecionar amostras representativas para o processo de treinamento, validação e teste da rede neural, e resultados satisfatórios foram obtidos [182]. Aqui, almeja-se representar, através da AFC, todo o espaço amostral dos dados e demonstrar as potencialidades da AFC para representar o processo. A Tabela 4.2 apresenta os valores máximos e mínimos de cada parâmetro para o conjunto total de dados selecionados.

Tabela 4.2. Valores máximos e mínimos para cada parâmetro do processo de laminação.

	h_i (mm)	h_o (mm)	μ	tf (Kgf/mm ²)	tb (Kgf/mm ²)	\bar{y} (Kgf/mm ²)	P (Kgf/mm ²)
Mínimo	4.6	3.492	0.096	6.3686	0.3087	40.8793	10207.049
Máximo	5.4	3.708	0.144	11.8274	0.5733	50.3415	28722.925

Dentre as amostras selecionadas, metade foi empregada para *modelagem* através da AFC (contexto formal, reticulado conceitual e extração de implicações) e a outra metade utilizada para *validação*.

4.3.3 Representando o Processo de Laminação através da AFC

Os valores dos parâmetros do processo de laminação são valores reais, conforme pode se observar na Tabela 4.2. Contudo, a AFC trabalha com valores discretos. Nesse caso, algumas soluções são a utilização de *pattern structures* [85], AFC *fuzzy* [9] ou transformar os valores contínuos em discretos, por meio de um processo de discretização [173]. Dentre essas alternativas, optou-se por utilizar um processo de discretização, o qual permite aplicar a AFC clássica. Realizou-se uma discretização equidistante nos dados, aplicando-se um total de 15 pontos de corte em cada parâmetro, o que resultou em 105 atributos (7 parâmetros divididos em 15 intervalos cada um). Dias et al. [46] e Zárate et al. [182] utilizaram um total de 15 intervalos de discretização por parâmetro e alcançaram resultados satisfatórios. A Tabela 4.3 apresenta os intervalos de discretização relativos ao atributo carga considerando os valores mínimos e máximos descritos na Tabela 4.2. Por exemplo, o intervalo $[10207.049, 11441.440)$ foi mapeado para o atributo $P = 0$. O limite inferior é sempre fechado e o superior é sempre aberto, exceto o último, isto é, aquele mapeado no atributo $P = 14$. Um procedimento similar foi empregado para os demais parâmetros.

Aplicando-se o processo de discretização nas 5 amostras selecionadas para construção dos modelos, produziu-se um conjunto de contextos formais (G, M, I) , em que G representa os registros da mostra, M os intervalos dos parâmetros e I a existência de um registro g ($g \in G$), que possui um parâmetro no intervalo m ($m \in M$). No caso, 5 amostras são suficientes para reportar os resultados com uma confiança de 90% [80]. Observe-se, ainda, que $|G| = 1961$, $|M| = 105$ e I apresenta uma densidade de 6.6% em cada contexto formal construído. Obviamente, cada objeto g possui 7 atributos incidentes ($|g'| = 7$).

Os contextos formais construídos serão denominados ao longo do texto de contextos #1, #2, #3, #4 e #5. Os reticulados respectivos possuem 21027, 21051, 21054, 20872 e 20999 conceitos formais, respectivamente.

Tabela 4.3. Discretização aplicada no parâmetro carga.

Atributos	Intervalo	
	Limite inferior	Limite superior
$P = 0$	10207.049	11441.440
$P = 1$	11441.440	12675.832
$P = 2$	12675.832	13910.224
$P = 3$	13910.224	15144.616
$P = 4$	15144.616	16379.008
$P = 5$	16379.008	17613.399
$P = 6$	17613.399	18847.791
$P = 7$	18847.791	20082.183
$P = 8$	20082.183	21316.575
$P = 9$	21316.575	22550.966
$P = 10$	22550.966	23785.358
$P = 11$	23785.358	25019.750
$P = 12$	25019.750	26254.142
$P = 13$	26254.142	27488.534
$P = 14$	27488.534	28722.925

4.3.3.1 Extraíndo Conhecimento através da AFC

No processo de laminação, a expressão do comportamento do processo mediante implicações pode ser apropriada por ser de fácil entendimento. Representações dessa natureza são indicadas em cenários industriais, por proporcionarem um fácil entendimento do comportamento das variáveis envolvidas no processo, notadamente para operadores menos experientes [181]. O conjunto de implicações a ser extraído será o de *implicações próprias* (veja a Definição 18). Na presente aplicação, uma implicação $P \rightarrow m$ expressa que uma conjunção de faixas de valores para determinados parâmetros (P), implicam em uma faixa de valores para determinado parâmetro (m). O conhecimento assim expresso pode ser empregado também na construção de sistemas especialistas [56, 180, 181].

No contexto da aplicação, obviamente, as implicações próprias apropriadas são aquelas com suporte maior que zero. Os contextos formais de #1 a #5 possuem, respectivamente, 87650, 86186, 86005, 86360 e 87050 implicações próprias. Um exemplo é a implicação $ho = 0 \rightarrow P = 14$, que expressa o fato de baixa espessura de saída

($h_0 = 0$) resultar em um alto valor de carga ($P = 14$). Mais especificamente, *se h_0 está no intervalo $[3.492, 3.5064)$, então P está no intervalo $[27488.533, 28722.925]$* . Isso é condizente com o que seria esperado para o processo de laminação [178].

Suponha-se que r seja um registro de uma base de dados de validação. Em tal registro deve haver pelo menos um valor real para cada parâmetro. Para um parâmetro p , tal valor é mapeado no atributo respectivo, $p = i$, $0 \leq i \leq 14$, segundo a discretização efetuada. Por exemplo, se o registro r apresenta o valor 20000.111 para o parâmetro P , então tal valor é mapeado para o atributo $P = 7$ (veja a Tabela 4.3). Designando-se por $\mathcal{M}(r)$ o conjunto dos 7 atributos mapeados dos valores de um registro r , dada uma implicação $P \rightarrow m$, tem-se os seguintes casos:

- (a) se $P \not\subseteq \mathcal{M}(r)$, então r não dá suporte a $P \rightarrow m$;
- (b) se $P \subseteq \mathcal{M}(r)$ e $m \notin \mathcal{M}(r)$, então r dá suporte negativo a $P \rightarrow m$; e
- (c) se $P \subseteq \mathcal{M}(r)$ e $m \in \mathcal{M}(r)$, então r dá suporte positivo a $P \rightarrow m$.

No caso (a), embora a implicação não falhe, ela não contribui positiva nem negativamente para a confiança na implicação. Se r não der suporte a nenhuma implicação, diz-se que r é um *registro sem implicações*. No caso (b), a implicação *falha*, contribuindo negativamente para o suporte. Tal falha provoca um decréscimo na confiança na implicação. Já no caso (c), o registro r contribui positivamente para a confiança na implicação.

As implicações próprias obtidas dos contextos formais #1 a #5 foram analisadas utilizando-se as cinco amostras de dados previamente selecionadas para validação. O contexto formal #1 foi avaliado utilizando-se a amostra #1, o contexto #2 utilizando-se a amostra #2, e assim por diante. A Tabela 4.4 apresenta uma síntese dos resultados. Por exemplo, para o contexto formal #1, as mais de 87 mil implicações próprias apresentaram um suporte positivo médio e um desvio padrão de 0.143 e 0.409, respectivamente. Note-se que a média é baixa, pois a grande maioria das implicações próprias não possui um suporte (76789 implicações próprias). Em relação ao suporte negativo, tem-se uma média e um desvio padrão de 0.789 e 1.125. Esse resultado é esperado, uma vez que o espaço amostral dos dados é muito amplo (10 milhões de registros) e foram empregadas amostras distintas de 1961 registros para a construção do contexto formal #1 e validação das implicações próprias utilizando a amostra #1. Para os demais contextos

formais, o comportamento é similar. Destaca-se ainda que, para as cinco amostras de 1961 registros utilizadas para validação, uma quantidade pequena de registros não forneceu suporte a nenhuma implicação (última linha da Tabela 4.4). Por exemplo, para o contexto formal #1, um total de 12 registros não deram suporte a nenhuma implicação. Os resultados apresentados demonstram que as implicações próprias descrevem parte do comportamento do processo de laminação. Tais resultados são uma linha base para comparação entre a aplicação da AFC com e sem redução.

Tabela 4.4. Resumo dos resultados relativos às implicações próprias dos contextos de validação originais.

	Contexto #1	Contexto #2	Contexto #3	Contexto #4	Contexto #5
# Conceitos formais	21027	21051	21054	20872	20999
# Implicações	87650	86186	86005	86360	87050
Suporte positivo - Média	0.143	0.144	0.142	0.143	0.142
Suporte positivo - Desv. P.	0.409	0.418	0.413	0.412	0.413
Suporte negativo - Média	0.789	0.797	0.794	0.797	0.789
Suporte negativo - Desv. P.	1.125	1.129	1.126	1.122	1.118
# Implicações sem suporte	76789	75508	75436	75606	76385
# Registros que não dá suporte	12	25	20	20	24

Os resultados sintetizados na Tabela 4.4 serão empregados para uma análise comparativa e descritiva. O objetivo é identificar se as características do processo siderúrgico de laminação a frio, observadas ao se utilizar a AFC, serão preservadas, em algum grau, pelo processo de redução levado a efeito pelas diferentes técnicas de redução.

O processo de laminação envolve relações não lineares de difícil análise. Felizmente, alguns comportamentos do processo são conhecidos, como, por exemplo, a relação entre os parâmetros carga P e espessura de saída ho e entre carga P e coeficiente de atrito μ . Carga e espessura de saída são, obviamente, inversamente proporcionais; i.e, quanto maior a carga exercida, menor será a espessura de saída. Por outro lado, quanto maior a carga exercida, maior será o atrito obtido. Esse comportamento pode ser observado através da análise formal de conceitos [181]. Com esse intuito, selecionaram-se e contabilizaram-se todas as implicações próprias $P \rightarrow m$, cuja premissa P possui ao menos um atributo representando os parâmetros espessura de saída ($ho = 0, ho = 1, \dots, ho = 14$) e/ou coeficiente de atrito ($\mu = 0, \mu = 1, \dots, \mu = 14$) e cujo consequente m é um atributo referente ao parâmetro carga ($P = 0, P = 1, \dots, P = 14$). Por exemplo,

suponha-se que as seguintes implicações próprias estejam entre as selecionadas:

$$i: ho = 1, tf = 9, \bar{y} = 3 \rightarrow P = 14;$$

$$j: ho = 5, \mu = 13 \rightarrow P = 13; e$$

$$k: ho = 1, tb = 2, hi = 13 \rightarrow P = 14.$$

Nesse caso, 2 implicações, i e k , contribuem com ocorrências de ambos $ho = 1$ e $P = 14$, e uma implicação, j , contribui com ocorrência de ambos $ho = 5$ e $P = 13$.

As Figuras 4.3(a) – (e) apresentam o comportamento dos parâmetros carga P e espessura de saída ho por meio da quantidade de implicações próprias obtidas dos contextos formais #1, #2, #3, #4 e #5. A Figura 4.3(a) corresponde ao contexto #1, a Figura 4.3(b) ao contexto #2 e, assim por diante. Já a Figura 4.3(f) apresenta a média para os cinco contextos formais. A tonalidade da cor em cada posição representa a quantidade de implicações próprias. Tonalidades claras representam poucas ou nenhuma implicação e tonalidades escuras representam uma grande quantidade de implicações próprias. Por exemplo, a posição ($ho = 0, P = 14$) na Figura 4.3(a) indica uma quantidade grande de implicações próprias (cor escura), enquanto que a posição ($ho = 0, P = 0$) apresenta poucas implicações próprias (cor clara). Vê-se que as figuras relativas aos 5 contextos formais e a relativa à média apontam comportamentos similares do processo de laminação com relação aos valores dos dois parâmetros considerados: quando a carga está no nível baixo, a espessura de saída encontra-se no nível alto; e, vice-versa, quando a carga se encontra no nível alto, a espessura de saída encontra-se no nível baixo. Esse comportamento é condizente ao esperado para o processo em análise [178].

De forma similar, as Figuras 4.4(a) – (f) analisam a relação entre carga P e coeficiente de atrito μ . Note-se que altos valores de carga resultam em altos valores de atrito e baixos valores de carga levam a baixos valores de atrito. Novamente, a característica apontada pelo mapeamento é condizente ao comportamento esperado para o processo [178].

Parte do comportamento do processo de laminação a frio foi modelado através da AFC. Nas próximas seções, será feito o mesmo procedimento, após o uso de técnicas de redução. Cabe ressaltar que o objetivo da aplicação é a extração de conhecimento sobre o processo de laminação, sendo a avaliação do conhecimento extraído, uma tarefa de difícil análise qualitativa, que demanda uma análise (subjetiva) de um especialista. Na ausência de tal especialista, concebe-se uma linha base, um conjunto de respostas

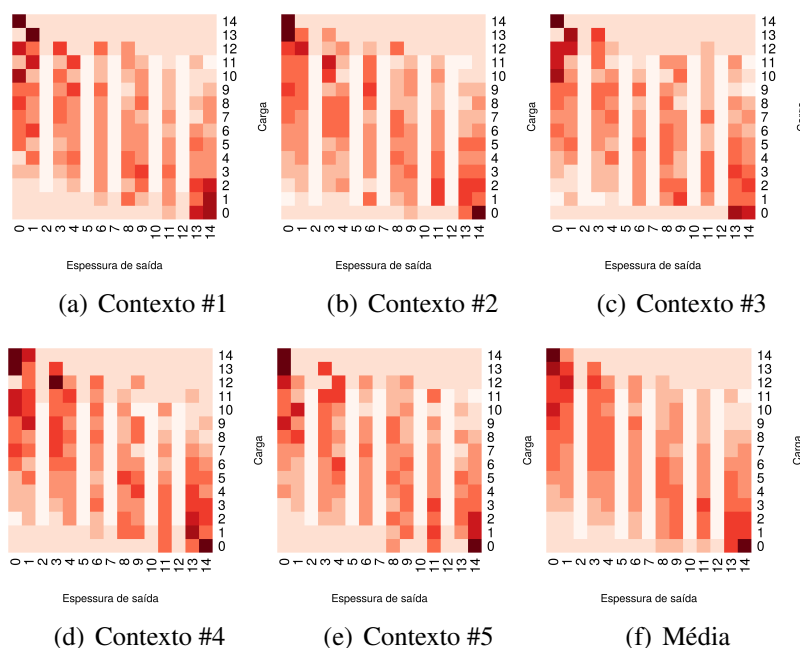


Figura 4.3. Mapeando a relação entre carga (P) e espessura de saída (h_o) através de implicações próprias.

conhecidas, e realiza-se uma comparação entre as respostas obtidas através do reticulado original e reduzido. Aqui, o conhecimento obtido através da AFC antes da aplicação de técnica de redução será considerado uma linha base. Pretende-se avaliar os resultados obtidos, em comparação aos anteriores, utilizando-se reticulados reduzidos obtidos por meio de diferentes técnicas.

4.3.4 Aplicando Técnicas de Redução

O objetivo agora é verificar a efetividade de diferentes técnicas de redução para o problema da laminação e proporcionar uma comparação experimental entre elas. Dentre as diversas técnicas de redução discutidas e classificadas no Capítulo 3, selecionou-se, além da técnica JBOS, as técnicas de estabilidade [101], conceitos frequentes [149] e geradores mínimos [13]. A técnica JBOS foi selecionada com o objetivo de analisá-la frente às demais técnicas da literatura. A técnica de estabilidade, sem dúvida, é a principal técnica de redução observada na literatura e apresenta os melhores resultados quando analisada frente a determinadas técnicas [89, 101, 138]. Conceitos formais frequentes,

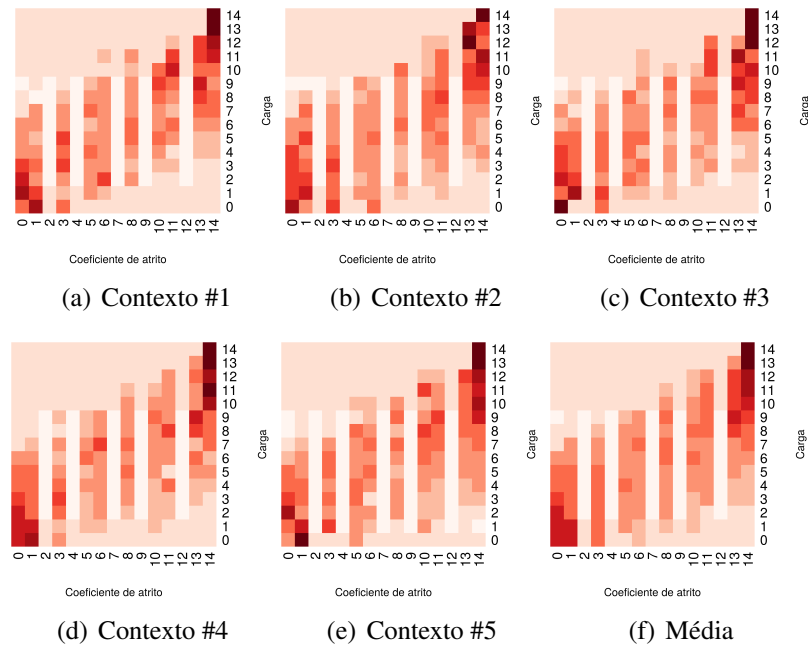


Figura 4.4. Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias.

por outro lado, são relevantes pela capacidade de selecionar conceitos sem ter acesso ao conjunto completo de conceitos formais. Por último, seleção de conceitos formais baseada em geradores mínimos tem sua importância em função de selecionar os conceitos utilizando geradores mínimos. Geradores e geradores mínimos têm ganhado destaque na literatura recente da AFC, pela sua capacidade de representar, de forma compacta, sem perdas, o conhecimento presente nos conceitos formais.

4.3.4.1 Agrupamento Baseado na Técnica - JBOS

Analisando-se os contextos formais que representam o processo de laminação, observa-se que o número de objetos é muito superior ao de atributos. Essa característica é comumente observada em aplicações reais [52, 54, 117]. Uma técnica de redução interessante, nesse cenário, é a técnica JBOS, que realiza agrupamento de objetos. Sob a perspectiva do processo de laminação, o agrupamento levado a efeito pela técnica JBOS resulta na simplificação de um conjunto de objetos do contexto formal, os quais representam configurações de uma operação do processo de laminação.

A aplicação da técnica JBOS demanda um conhecimento prévio da importância relativa dos atributos do contexto formal. A partir desse conhecimento, o usuário deve determinar um peso para cada atributo. No processo de laminação, Zárata & Bittencout [179] propõem a seguinte relevância relativa para os parâmetros: $P \succ \mu \succ ho \succ hi \succ \bar{y} \succ tb \succ tf$. Aqui, $p_2 \succ p_1$ significa que p_2 é mais relevante (maior) que p_1 . Com base nessa relevância relativa, atribuem-se os seguintes pesos para os atributos correspondentes aos parâmetros (para $i = 0, \dots, 14$ os pesos são idênticos para o mesmo parâmetro): 0.7 para os atributos $P = i$, 0.6 para $\mu = i$, 0.5 para $ho = i$, 0.4 para $hi = i$, 0.3 para $\bar{y} = i$, 0.2 para $tb = i$ e 0.1 para $tf = i$. Por exemplo, todo atributo referente ao parâmetro carga (P) possui o peso 0.7 na escala proposta. A relevância, assim como o conjunto de pesos propostos, foi empregada em outra aplicação, tendo sido observados resultados satisfatórios [46].

Além da relevância dos atributos, representada através de pesos, a técnica JBOS necessita de um índice de similaridade, ε , e um número máximo de objetos similares por grupo, α , ambos definidos pelo usuário. Nos contextos formais de #1 a #5, diversas configurações de redução foram experimentadas. A Figura 4.5 resume as configurações testadas para o contexto formal #1, apresentando as reduções do número de conceitos formais em função dos parâmetros ε e α . Observe-se que o número de conceitos formais sofre maior influência do índice de similaridade. Dentre as configurações avaliadas, considerou-se satisfatória aquela empregando $\varepsilon = 0.96$ e $\alpha = 5$. Essa configuração resulta em uma redução do número de conceitos formais para os contextos #1, #2, #3, #4 e #5 de 42.91%, 41.61%, 41.43%, 43.06% e 42.97%, respectivamente. Em todas as reduções, ajustaram-se os parâmetros de forma a se obter uma redução de $40\% \pm 10\%$ no número de conceitos formais.

Os contextos formais #1, #2, #3, #4 e #5 reduzidos pela técnica JBOS possuem 57953, 57898, 57756, 57378 e 56680 implicações próprias, respectivamente. A Tabela 4.5 resume os resultados relativos de cada contexto formal para uma análise comparativa aos resultados da Tabela 4.4 obtidos sem redução. Com uma confiança de 90%, observou-se, para os cinco contextos avaliados, uma redução de aproximadamente 40% no número de conceitos formais, que, por sua vez, proporcionou uma redução no número de implicações próprias. O processo de redução levado a efeito pela técnica JBOS altera o conjunto de conceitos formais e, conseqüentemente, o conjunto de implicações próprias. Essa modificação proporcionou um incremento no suporte positivo. Esse re-

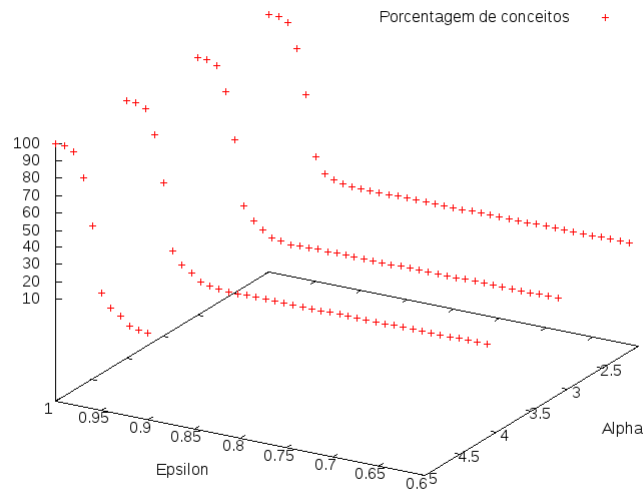


Figura 4.5. Reduções do número de conceitos formais em função dos parâmetros ϵ e α para a técnica JBOS utilizando o contexto formal #1.

sultado pode ser explicado pelo fato de que a técnica JBOS provoca generalizações no conjunto de implicações próprias, conforme será descrito na Seção 5.2.1. De forma semelhante, houve um incremento no suporte negativo. Esse incremento deve-se a inexatidões inseridas pela técnica, representadas por implicações próprias que apresentam falhas. Observe-se ainda que o número de implicações próprias sem suporte decresceu significativamente. Tal decremento acompanha o descréscimo do número total de implicações próprias. Por último, a redução do número de implicações próprias e as inexatidões inseridas proporcionam um incremento no número de registros que não fornecem suporte a nenhuma implicação, conforme mostra a última linha das Tabelas 4.3 e 4.5.

As Figuras 4.6(a) – (f) mostram as relações entre carga e espessura de saída e as Figuras 4.7(a) – (f) as relações entre carga e coeficiente de atrito, relações estas obtidas a partir dos conjuntos de implicações próprias dos 5 contextos formais reduzidos. Analisando-se as informações lá presentes, vê-se que, mesmo após a redução média de 40% no número de conceitos formais, tem-se um comportamento similar ao do observado nas Figuras 4.3(a) – (f) e 4.4(a) – (f): altos valores de carga levam a baixos valores de espessura de saída e valores baixos de carga resultam em altos valores de espessura

de saída; baixos valores de atrito são resultantes de baixa carga e altos valores de carga levam a altos valores de atrito. Note-se, nas Figuras 4.6(a) e (e) e também nas Figuras 4.7(a) e (e), que as implicações próprias com o parâmetro carga mapeado no intervalo 14 (atributo $P = 14$) foram removidas pela redução. As poucas implicações próprias com referência ao atributo respectivo foram eliminadas, no entanto a essência do comportamento do processo foi preservada. Em particular, uma análise da média (Figuras 4.6(f) e 4.7(f)) mostra claramente que o comportamento é o esperado.

4.3.4.2 Estabilidade

Uma das principais técnicas para seleção de conceitos formais existentes é a que se baseia no índice de estabilidade [101]. Tal índice objetiva mensurar o quanto a intenção de um conceito depende de sua extensão. Utilizando-se esse índice, pode-se assumir um limiar a partir do qual todos os conceitos formais com valor inferior são removidos.

O índice de estabilidade é agora aplicado na seleção de conceitos formais, que representam o processo siderúrgico de laminação a frio. Para tanto, foi utilizado o algoritmo proposto por Roth et al. [138]. A ideia é selecionar os conceitos formais mais estáveis, i.e., aqueles que melhor descrevem o processo em análise. De posse dos conceitos formais selecionados, implicações próprias são extraídas e a qualidade das mesmas é aferida.

A Figura 4.8 apresenta as reduções do número de conceitos formais em função do índice de estabilidade quando se utiliza o reticulado conceitual resultante do

Tabela 4.5. Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica JBOS.

	Contexto #1	Contexto #2	Contexto #3	Contexto #4	Contexto #5
# Conceitos formais	12003	12289	12330	11883	11975
Redução (%)	42.91	41.61	41.43	43.06	42.97
# Implicações	57953	57898	57756	57378	56680
Suporte positivo - Média	0.211	0.214	0.21	0.214	0.217
Suporte positivo - Desv. P.	0.503	0.505	0.503	0.503	0.514
Suporte negativo - Média	1.274	1.257	1.263	1.277	1.282
Suporte negativo - Desv. P.	1.367	1.359	1.361	1.381	1.378
# Implicações sem suporte	47635	47403	47559	47022	46429
# Registros que não dá suporte	39	48	32	45	32

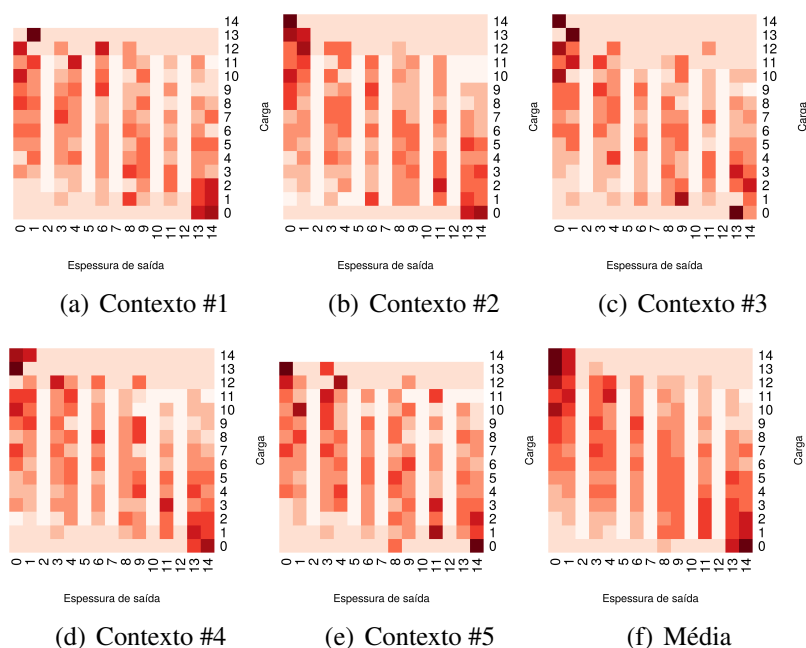


Figura 4.6. Mapeando a relação entre carga (P) e espessura de saída (h_o) através de implicações próprias obtidas após redução pela técnica JBOS.

contexto formal #1. Com um índice de estabilidade de 0.1 nenhum conceito formal é removido. Entre 0.13 e 0.24 o número de conceitos formais é reduzido para 20918 (aproximadamente 1%). Já entre 0.25 e 0.37 o número de conceitos formais é reduzido para 11417 (45.7%). A seleção do índice de estabilidade mínimo é dependente da aplicação intencionada. Dentre as reduções apresentadas na Figura 4.8, optou-se por um índice de estabilidade igual a 0.37. Valores inferiores a 0.37 apresentam uma redução do número de conceitos formais muito baixa, e valores superiores uma redução superior a 50%. Esse valor será usado com todos os contextos formais, de #1 a #5. A primeira linha da Tabela 4.6 apresenta o número de conceitos formais obtidos em cada contexto formal, e a terceira o percentual de redução correspondente. Tem-se em média uma redução de 45.5%.

A Tabela 4.6 apresenta o resumo dos resultados com relação às implicações próprias para os 5 contextos formais, ao se aplicar o corte acima referido. Aqui, e para as outras duas técnicas a serem apresentadas, as implicações próprias são obtidas utilizando o Algoritmo 2 (Seção 5.5), pois os conceitos formais selecionados não formam um reticulado conceitual. Os reticulados conceituais provenientes dos contextos for-

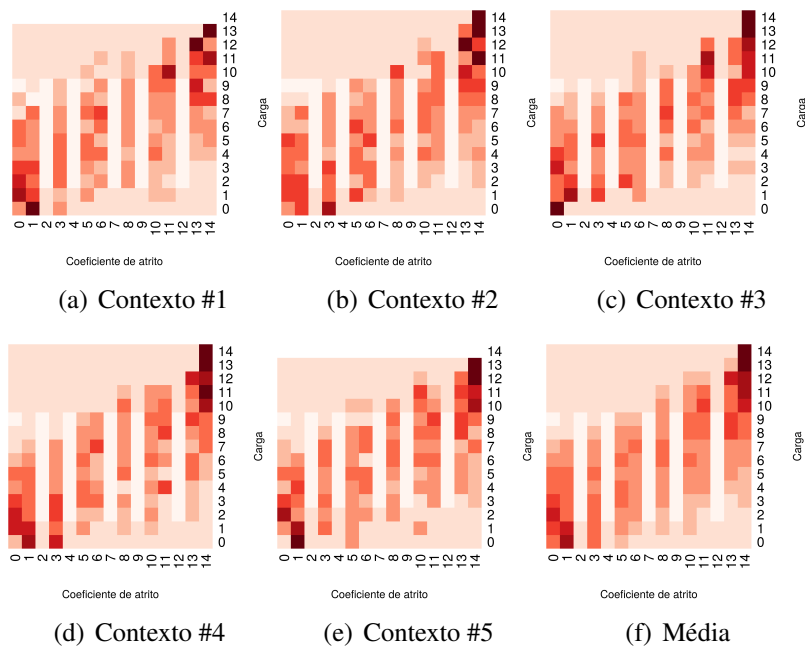


Figura 4.7. Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica JBOS.

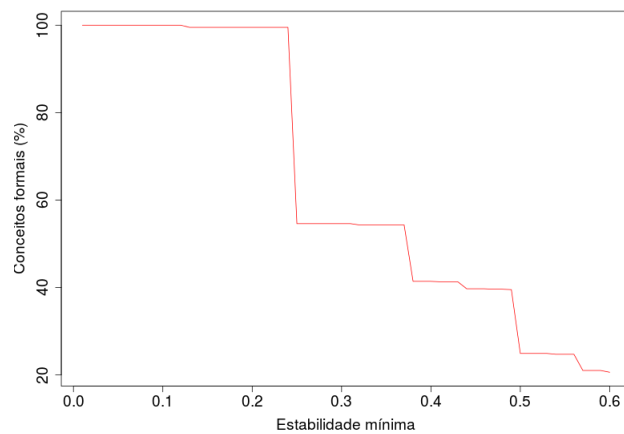


Figura 4.8. Reduções do número de conceitos formais em função de uma estabilidade mínima para o contexto #1.

mais #1, #2, #3, #4 e #5 possuem mais implicações próprias do que os reticulados provenientes dos contextos formais originais (Tabela 4.4). Conforme será visto na Se-

ção 5.5, quando um conceito gerador de implicações próprias é removido, seus vizinhos superiores podem gerar implicações próprias mais específicas e, em maior quantidade. Consequentemente, o número total de implicações próprias pode aumentar. Em relação ao suporte positivo, houve um pequeno decremento na média e no desvio padrão. Observa-se uma pequena variação na média e no desvio padrão do suporte negativo. Assim como o número total de implicações próprias, o de implicações próprias sem suporte também cresceu um pouco. E o incremento no número de registros que não fornecem suporte é esperado, uma vez que parte dos conceitos formais é eliminado.

Tabela 4.6. Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica baseada no índice de estabilidade.

	Contexto #1	Contexto #2	Contexto #3	Contexto #4	Contexto #5
# Conceitos formais	11417	11451	11505	11371	11414
Redução (%)	45.7	45.6	45.35	45.52	45.64
# Implicações	91925	90486	90208	90238	91279
Suporte positivo - Média	0.123	0.126	0.124	0.126	0.124
Suporte positivo - Desv. P.	0.377	0.39	0.382	0.384	0.382
Suporte negativo - Média	0.702	0.707	0.706	0.711	0.703
Suporte negativo - Desv. P.	1.06	1.073	1.071	1.07	1.064
# Implicações sem suporte	81880	80548	80347	80221	81355
# Registros que não dá suporte	36	37	36	27	34

As Figuras 4.9(e) e (f) e 4.10(e) e (f) apresentam as relações entre carga e espessura de saída e carga e coeficiente de atrito para as implicações próprias dos contextos formais reduzidos. Comparando-se às relações antes da aplicação da técnica de redução, mostradas nas Figuras 4.3(e) e (f) e 4.4(e) e (f), nota-se que o comportamento do processo foi preservado: quanto menor a espessura, maior a carga e quanto maior o coeficiente de atrito, maior a carga.

Usando-se a técnica de estabilidade, reduziu-se o reticulado conceitual por meio da seleção dos conceitos formais mais estáveis. Apesar da redução do número de conceitos, o número de implicações próprias aumentou, como já enfatizado. Entretanto, tal incremento não alterou a qualidade das implicações próprias, no sentido de que o comportamento do processo, com relação aos parâmetros carga, espessura de saída e coeficiente de atrito, foi preservado.

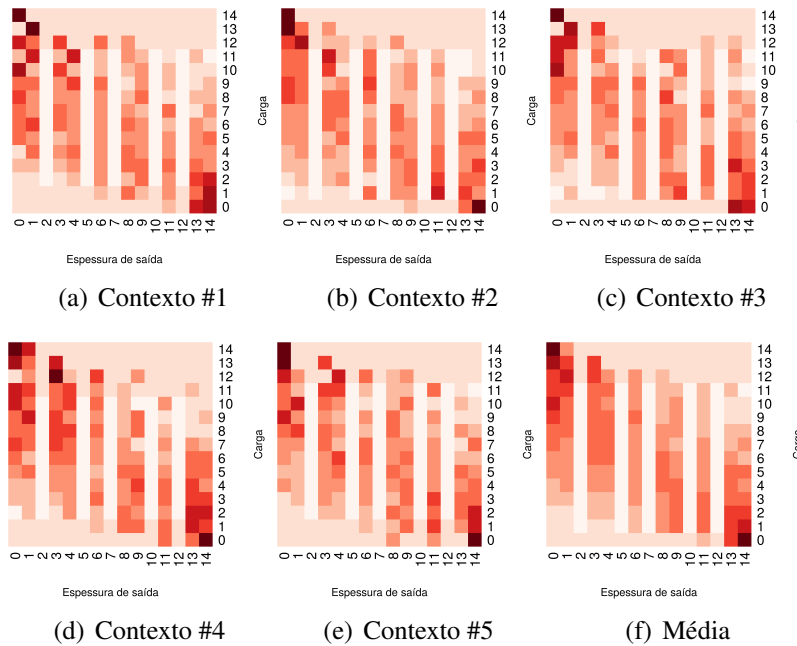


Figura 4.9. Mapeando a relação entre carga (P) e espessura de saída (h_o) através de implicações próprias obtidas após redução pela técnica de estabilidade.

4.3.4.3 Conceitos Formais Frequentes

Reticulados formados por conceitos formais frequentes são reticulados parciais constituídos apenas com conceitos cujas intenções são frequentes [123]. Nessa técnica, conceitos formais cujas intenções não são frequentes, ou seja têm suporte abaixo de um suporte mínimo (veja a Definição 16), são removidos do reticulado conceitual. A técnica produz um reticulado conceitual parcial constituído apenas de conceitos frequentes.

Para a seleção de conceitos formais frequentes dos reticulados conceituais que representam o processo siderúrgico de laminação a frio, empregou-se o algoritmo *frequent next neighbours* apresentado em Carpineto & Romano [29]. A Figura 4.11 mostra as reduções do número de conceitos formais em função do suporte mínimo para o contexto formal #1. Vê-se que baixos valores de suporte e pequenas variações levam a grandes reduções. Assim como na técnica de estabilidade, em algumas configurações nota-se uma grande redução. Obviamente, tal característica é dependente dos dados em análise.

Dentre as configurações avaliadas, optou-se por aquela de suporte mínimo igual a 0.0017, que resulta em uma redução de 47.28%, 46.89%, 46.76%, 46.81% e 46.78% no

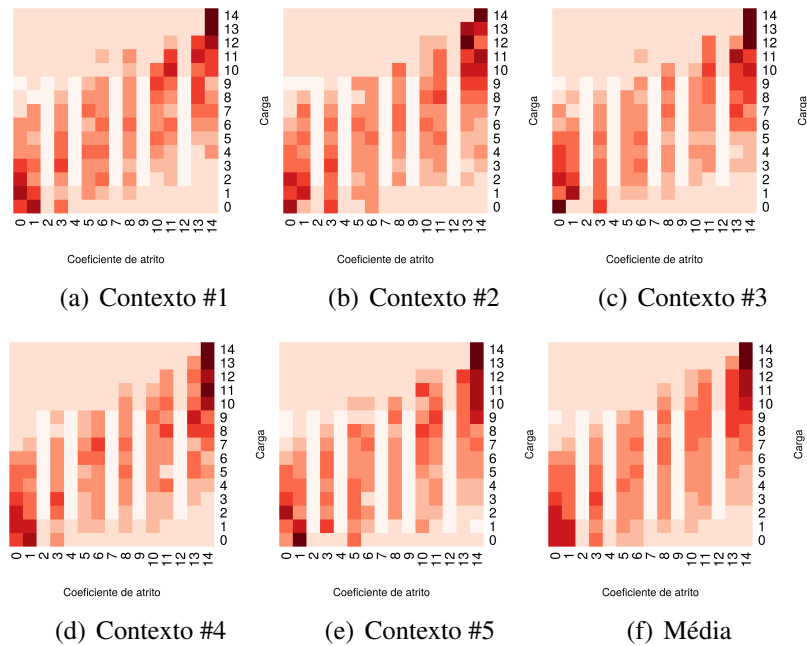


Figura 4.10. Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica de estabilidade.

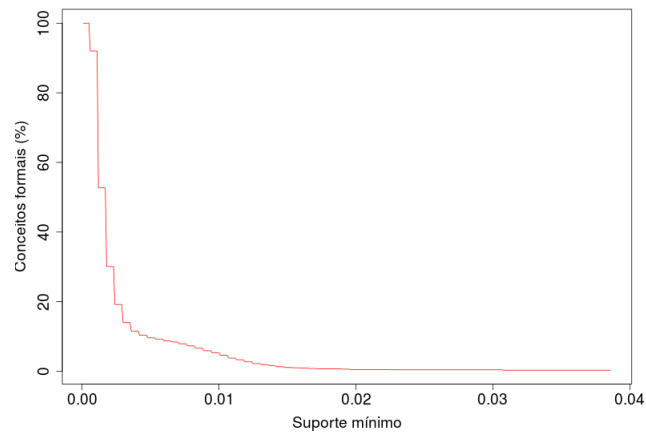


Figura 4.11. Reduções do número de conceitos formais em função do suporte mínimo para a técnica baseada em conceitos formais frequentes para o contexto formal #1.

número de conceitos formais relativos aos contextos #1, #2, #3, #4 e #5.

Como mostrado na Tabela 4.7, tem-se em média 548 implicações próprias por

contexto formal reduzido. Observe-se que o número de implicações próprias é muito inferior ao do contexto formal original. Todos os valores mensurados e reportados na Tabela 4.7 foram prejudicados pelo baixo número de implicações próprias. Em especial, o número de registros que não fornecem suporte a nenhuma implicação é superior a 85%. O número reduzido de implicações próprias é resultado da forma como os conceitos formais considerados relevantes são selecionados pela técnica.

Tabela 4.7. Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica de conceitos formais frequentes

	Contexto #1	Contexto #2	Contexto #3	Contexto #4	Contexto #5
# Conceitos formais	11086	11180	11210	11101	11175
Redução (%)	47.28	46.89	46.76	46.81	46.78
# Implicações	548	521	566	646	572
Suporte positivo - Média	0.693	0.696	0.747	0.674	0.704
Suporte positivo - Desv. P.	1.068	1.241	1.105	1.022	1.109
Suporte negativo - Média	1.337	1.391	1.459	1.399	1.431
Suporte negativo - Desv. P.	1.443	1.356	1.484	1.472	1.36
# Implicações sem suporte	313	322	316	375	334
# Registros que não dá suporte	1449	1487	1432	1448	1446

As Figuras 4.12(a) – (f) e 4.13(a) – (f) apresentam as relações entre carga (P) e espessura de saída (ho) e entre carga (P) e coeficiente de atrito (μ) obtidas das implicações próprias dos contextos formais reduzidos. Não é mais possível identificar o mesmo comportamento que anteriormente à redução. Muitos atributos (intervalos de discretização) não estão presentes nos conjuntos de implicações próprias. Por exemplo, na Figura 4.12(e) não há implicações próprias com valores de carga para os atributos que representam os intervalos de 10 a 15. A média mostrada nas Figuras 4.12(e) e 4.13 (e) também não caracterizam um comportamento dos parâmetros compatível com o original.

O nível de redução utilizado foi similar ao empregado para a técnica de seleção baseada em estabilidade. Entretanto, a redução resultou em um número relativamente pequeno de implicações próprias, que não levaram a um comportamento adequado das variáveis envolvidas no processo de laminação. O baixo número de implicações próprias é resultado do processo de seleção dos conceitos, que seleciona apenas os conceitos frequentes. Portanto, a seleção de conceitos frequentes não é adequada ao contexto da presente aplicação.

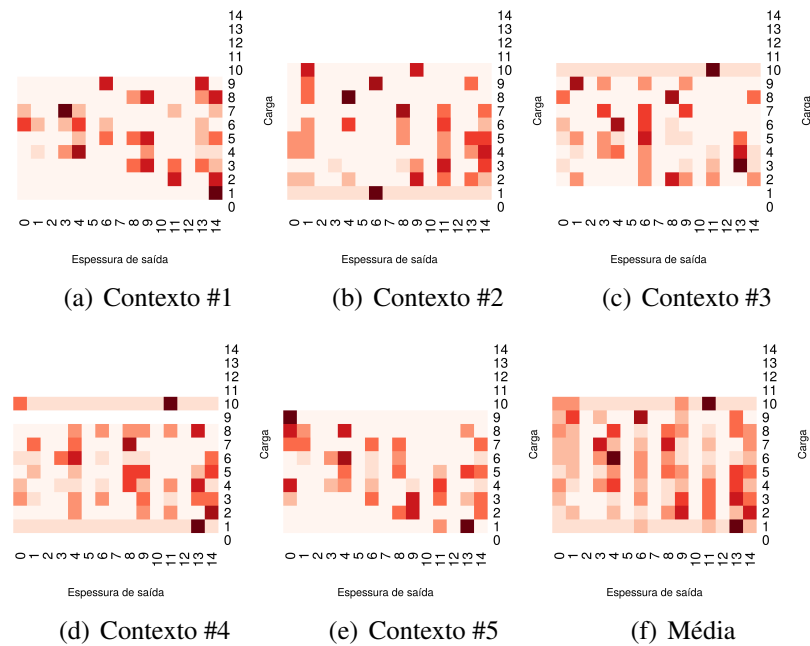


Figura 4.12. Mapeando a relação entre carga (P) e espessura de saída (h_o) através de implicações próprias obtidas após redução pela técnica de conceitos formais frequentes.

4.3.4.4 Geradores Mínimos

A seleção de conceitos formais baseada em geradores mínimos foi proposta por Belohlavek & Macko [13]. Para cada atributo do contexto formal, eles atribuem um peso no intuito de descrever sua relevância. Essa relevância dos atributos é transportada para relevâncias dos conceitos formais, por intermédio da análise dos geradores e geradores mínimos de cada conceito formal (Seção 3.1.3). Um conceito formal é considerado relevante se possui pelo menos um gerador mínimo cuja soma do peso dos atributos do mesmo dividida pelo seu número de atributos é superior a uma relevância mínima definida pelo usuário. Os conceitos mais relevantes são então selecionados.

Empregou-se o mesmo conjunto de pesos usado com a técnica JBOS, a saber, $P = 0.7$, $\mu = 0.6$, $h_o = 0.5$, $h_i = 0.4$, $\bar{y} = 0.3$, $tb = 0.2$ e $tf = 0.1$. Novamente, cada atributo resultante de um parâmetro recebeu o mesmo peso. Para a seleção de geradores mínimos utilizou-se o arcabouço *Coron system* [87].

A Figura 4.14 apresenta as reduções do número de conceitos formais em função da relevância mínima para o contexto formal #1 e o conjunto de pesos já mencionado.

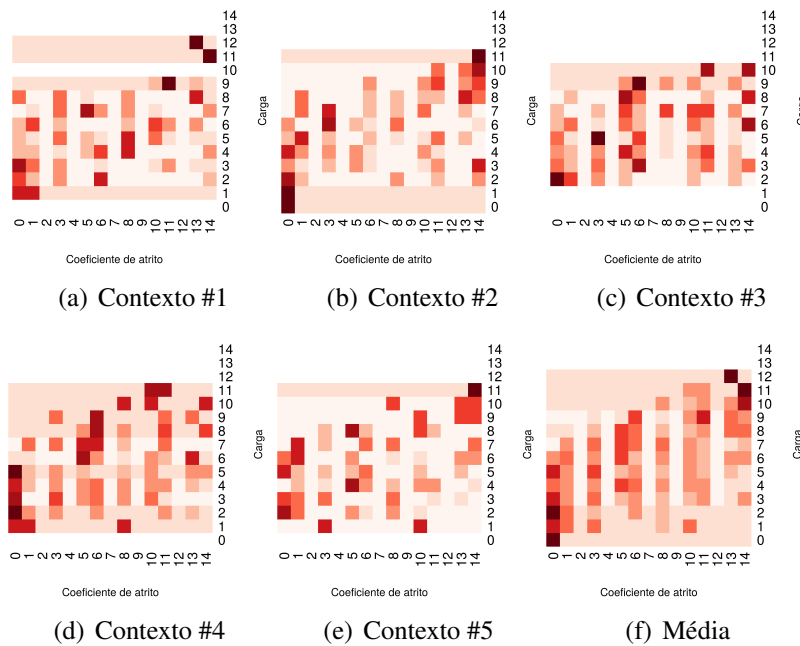


Figura 4.13. Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica de conceitos formais frequentes.

Até uma relevância mínima de aproximadamente 0.1 nenhuma redução é obtida. Entre 0.11 e 0.6 tem-se reduções que variam de 1% a 95%. Observa-se que, a exemplo do ocorrido com as técnicas de estabilidade e conceitos formais frequentes (Figuras 4.8 e 4.11), não são encontradas grandes reduções quando há uma pequena variação na relevância mínima. Dentre as configurações de redução avaliadas, optou-se por uma relevância mínima igual a 0.4. Esse valor resulta numa redução de 37.69%, 37.73%, 37.66%, 37.39% e 37.89% no número de conceitos formais para os contextos #1, #2, #3, #4 e #5, respectivamente.

Uma redução média de 37% no número de conceitos formais resultou em média de 88381 implicações próprias. A Tabela 4.8 sintetiza as quantidades de conceitos formais, implicações próprias, redução obtida, implicações próprias sem suporte e registros que não fornecem suporte a nenhuma implicação. Comparando-se à Tabela 4.4 observa-se, com uma confiança de 90%, que a média e o desvio padrão apresentam pequenas variações. E, o número de implicações próprias sem suporte e o número de registros que não fornecem suporte as implicações próprias apresentam um incremento.

As Figuras 4.15(a) – (f) e Figuras 4.16(a) – (f) apresentam as relações entre carga e

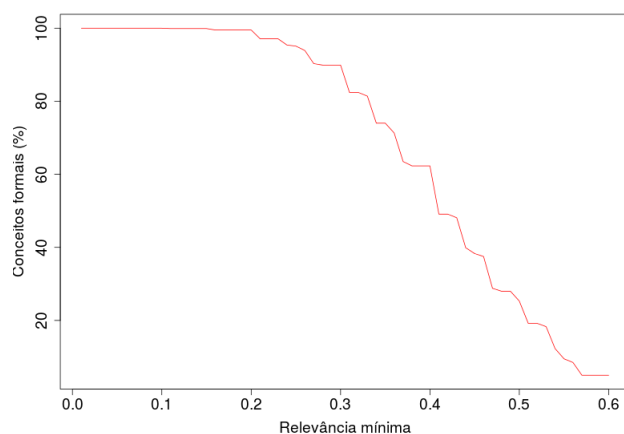


Figura 4.14. Reduções do número de conceitos formais em função da relevância mínima para a técnica de redução baseada em geradores mínimos utilizando o contexto formal #1.

Tabela 4.8. Resumo dos resultados relativos às implicações próprias dos contextos reduzidos pela técnica de geradores mínimos.

	Contexto #1	Contexto #2	Contexto #3	Contexto #4	Contexto #5
# Conceitos formais	13101	13108	13126	13067	13043
Redução (%)	37.69	37.73	37.66	37.39	37.89
# Implicações	89485	87850	87770	88034	88769
Suporte positivo - Média	0.137	0.1396	0.137	0.139	0.138
Suporte positivo - Desv. P.	0.399	0.41	0.402	0.403	0.404
Suporte negativo - Média	0.763	0.771	0.768	0.77	0.763
Suporte negativo - Desv. P.	1.106	1.112	1.109	1.106	1.099
# Implicações sem suporte	78767	77283	77313	77369	78169
# Registros que não dá suporte	18	28	22	20	26

espessura de saída e entre carga e coeficiente de atrito, respectivamente, obtidas das implicações próprias do reticulado conceitual reduzido. Mesmo ante uma redução média de 37% no número de conceitos formais, o comportamento do processo de laminação foi preservado. Observa-se nas Figuras 4.15(a) – (f) que maior carga leva a uma menor espessura de saída, e nas Figuras 4.16(a) – (f) que maior carga gera maior atrito.

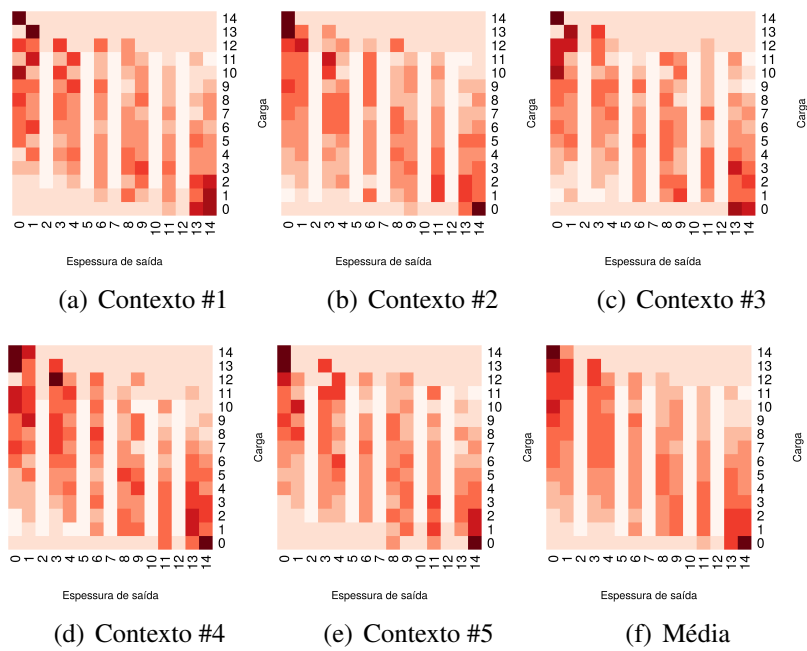


Figura 4.15. Mapeando a relação entre carga (P) e espessura de saída (h_o) através de implicações próprias obtidas após redução pela técnica de seleção baseada em geradores mínimos.

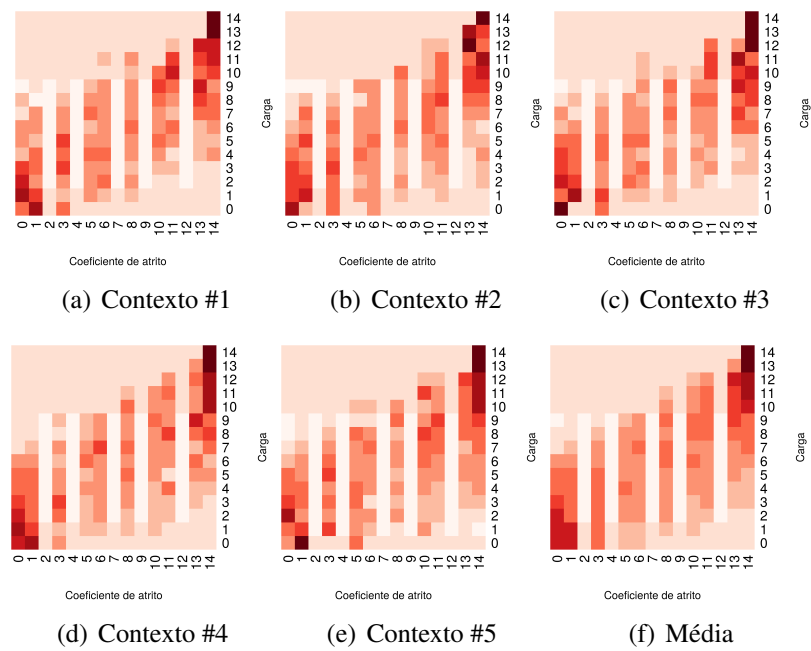


Figura 4.16. Mapeando a relação entre carga (P) e coeficiente de atrito (μ) através de implicações próprias obtidas após redução pela técnica de seleção baseada em geradores mínimos.

4.4 Conclusões

Este capítulo apresentou a técnica JBOS, uma técnica da classe de simplificação, para redução de reticulados conceituais. Apresentou também um exemplo de uso da técnica, juntamente com 3 outras, para redução em uma aplicação real, a modelagem de parte de um processo siderúrgico de laminação a frio.

Como ressaltado ao longo do capítulo, os requisitos estabelecidos na Seção 4.1 são atendidos. Em especial, o nível de satisfação do requisito 3, chegar a um reticulado que exponha o essencial, é dependente de uma escolha criteriosa dos pesos para os atributos e escolha do parâmetro ϵ . De qualquer forma, tal grau de satisfação pode ser ajustado por meio de experimentos com diferentes atribuições de pesos aos atributos e valores para ϵ (e, eventualmente, para α).

Destaca-se, ainda, o fato de JBOS determinar a similaridade entre objetos baseando-se, não apenas na presença e/ou ausência dos atributos, mas na relevância dos mesmos. Embora existam outros tantos métodos que também valorizam a importância dos atributos, JBOS traz uma característica que pode torná-la mais adequada para muitas aplicações: o agrupamento de objetos semelhantes, com base em uma medida de relevância diretamente apresentada pelo usuário. Ademais, a técnica apresenta um desempenho computacional satisfatório, uma vez que não necessita ter acesso aos conceitos formais originais. Por último, produz, efetivamente, um contexto formal e, em consequência, um reticulado conceitual reduzido.

Os resultados relativos ao estudo de caso demonstram que, mesmo se utilizando uma redução média de 40% no número de conceitos formais, as técnicas são capazes de preservar parte do conhecimento do processo de laminação representado através de implicações. Aplicando-se um teste pareado [80], com uma confiança de 90%, no quantitativo de registros que não contribuem para o suporte, que contribuem de forma negativa e de forma positiva, observou-se que as implicações extraídas do reticulado reduzido pela técnica de geradores mínimos descrevem de forma mais concisa o processo de laminação. Seguido pelas extraídas dos reticulados reduzidos pelas técnicas de estabilidade, JBOS e conceitos frequentes.

As técnicas baseadas em geradores mínimos e estabilidade têm acesso a todos os conceitos formais para seleção daqueles considerados apropriados. Consequentemente, tais técnicas são capazes de selecionar os conceitos formais considerados ideais. Entre-

tanto, nem sempre se faz possível o acesso ao conjunto completo de conceitos formais, e, por outro lado, sem acesso ao conjunto completo de conceitos, fazendo-se uso tão somente do contexto formal, a técnica JBOS foi capaz de preservar parte do reticulado conceitual. As implicações próprias oriundas de tal técnica representaram, de forma satisfatória, a parte analisada do processo de laminação. De forma semelhante, a técnica de seleção baseada em conceitos frequentes não precisa do acesso ao conjunto completo de conceitos. Entretanto, as implicações próprias resultantes de tal técnica não foram capazes de descrever de forma satisfatória o processo de laminação. O critério utilizado pela técnica leva à obtenção de um conjunto muito restrito de implicações próprias.

Capítulo 5

Análise da Qualidade de Reduções

Conforme visto no Capítulo 3, são muitas as técnicas de redução propostas, com características distintas e, em muitos casos, elas não vêm acompanhadas de uma análise que ateste os efeitos e a qualidade das reduções efetuadas. Na realidade, normalmente há uma alta dose de subjetividade quando se analisa os resultados das reduções.

Neste capítulo, apresenta-se uma metodologia para análise da qualidade de reduções, através da qual é possível identificar as transformações levadas a efeito no conhecimento representado. Ela utiliza, como formalismo de representação do conhecimento, o conjunto de implicações do contexto formal, e apresenta uma forma de identificar o que é preservado, eliminado, transformado ou inserido. Além disso, quatro índices complementares são sugeridos, por meio dos quais se pode mensurar quatro aspectos qualitativos diferentes do conhecimento representado.

5.1 Qualidade das Reduções

O ato de avaliar a qualidade de uma redução apresenta uma série de desafios. Um deles é que a noção de qualidade é, muitas vezes, subjetiva e dependente da aplicação que tenha como base o contexto formal ou reticulado conceitual em questão.

As técnicas da classe de *eliminação de informação redundante* (como definida no Capítulo 3), em geral não vêm acompanhadas de análises da qualidade do reticulado obtido. Embora obtenham um reticulado conceitual isomorfo ao original, os objetos e/ou atributos removidos podem ser importantes em determinados cenários. Por exemplo, a

eliminação de objetos implica na remoção dos mesmos das extensões de todos os conceitos, mas extensões ou suas cardinalidades são importantes em algumas aplicações. Logo, a remoção de elementos “redundantes” pode proporcionar uma degradação em termos de qualidade, dependendo da aplicação intencionada.

Perdas ocasionadas pelas técnicas de *simplificação* e *seleção* são mais comumente observadas. Naturalmente, ao agrupar conceitos formais, objetos ou atributos considerados semelhantes, mesmo que mediante critérios bem definidos, perde-se a capacidade de discriminar os elementos que formam um grupo. De forma semelhante, a seleção de conceitos formais significativos pode provocar a perda de alguns conceitos e/ou relacionamentos relevantes. Tratando-se de técnicas que produzem ou selecionam reticulados conceituais menores que o original, o conhecimento representado resultante pode vir a conter generalizações, especializações, cortes e inserções de porções de conhecimento. Em nenhuma das muitas técnicas estudadas fica muito claro que transformações são efetivamente proporcionadas.

Algumas técnicas de redução, dentre aquelas descritas no Capítulo 3, vêm acompanhadas de *medidas de interesse* ou critérios que visam garantir a qualidade do reticulado conceitual produzido. Entretanto, essas medidas são raras, específicas e aplicadas, em geral, apenas no contexto da técnica. Medidas para análise de reticulados conceituais devem ser capazes de mensurar o conhecimento expresso e seu potencial de aplicação. Segue uma definição de medida de interesse.

Definição 17. *Uma medida de interesse é uma estimativa da qualidade de um reticulado conceitual com relação a certo aspecto.* □

Medidas de interesse podem ser classificadas em *objetivas* e *subjetivas* [115]. Medidas objetivas são as fortemente baseadas em dados estatísticos ou propriedades existentes nos dados, que possam ser mensuradas numericamente. Medidas subjetivas são aquelas derivadas de crenças ou expectativas relacionadas com o domínio do problema. Um exemplo de medida objetiva é o da cardinalidade da extensão de um conceito, que pode ser utilizada para selecionar conceitos formais com tal cardinalidade superior a um limiar fornecido. Agora seja o problema de avaliar uma regra tomada de determinado contexto formal. Ela pode ser considerada proveitosa, desde que apresente características tais como “surpresa”, “utilidade” e “novidade”, características estas subjetivas e dependentes do domínio da aplicação.

Uma medida de interesse objetiva, bastante utilizada é a estabilidade [101]. Ela apresenta um índice para os conceitos que indica o quanto a intenção do conceito depende do conjunto de objetos. Outras medidas, tais como restrições baseadas em frequência [149], baseadas em área [25] e medidas de similaridade ou distância [137], também são empregadas. Entretanto, essas são medidas específicas, aplicáveis apenas ao contexto de cada técnica.

As medidas objetivas existentes são utilizadas, frequentemente, na seleção de conceitos formais. Algumas dessas medidas são inspiradas em medidas aplicadas em mineração de dados, as quais procuram selecionar padrões relevantes em grandes volumes de dados [115]. Entretanto, no contexto da AFC essas medidas podem não ser apropriadas, visto que não levam em consideração a estrutura do reticulado conceitual.

Além disso, medidas de interesse propostas em associação com técnicas específicas muitas vezes não são apropriadas para analisar outras técnicas. Medidas como suporte (conceitos formais frequentes) e estabilidade utilizadas em técnicas da classe de seleção não são adequadas para mensurar, por exemplo, a qualidade de um reticulado reduzido obtido por uma técnica de simplificação. Apesar de as técnicas terem o propósito comum de reduzir o reticulado conceitual, em geral elas apresentam características distintas e propósitos específicos.

Por outro lado, algumas técnicas analisam a qualidade do reticulado conceitual resultante de forma indireta, de forma um tanto subjetiva, através de aplicações que façam uso do reticulado. Isso acontece, por exemplo, em algumas aplicações relativas a análise de reticulado para visualização de consultas [34], recuperação de documentos [88], análise de redes sociais [81], sistemas de recomendação [61], recomendações médicas [96], etc.

Em conclusão, há carência de uma metodologia que permita identificar, com alguma clareza, o que ocorre quando se aplica uma técnica de redução: quais porções de conhecimento são preservadas, eliminadas, inseridas ou transformadas. Neste último caso, em particular, seria importante identificar se a técnica faz especializações e/ou generalizações. Como medidas de interesse objetivas ou subjetivas não se mostram suficientes para avaliar todos os aspectos de importância em um reticulado conceitual, diferentes índices devem ser usados para avaliar diferentes aspectos.

5.2 Análise de Reduções através de Implicações

A seguir, é desenvolvida uma metodologia de análise de reduções baseada no uso de implicações. O conhecimento expresso em um contexto formal (G, M, I) , ou no correspondente reticulado conceitual $\mathcal{B}(G, M, I)$, pode ser utilizado para derivar implicações $P \rightarrow Q$, em que P e Q são conjuntos de atributos, conforme visto na Seção 2.3.4. O conjunto de atributos P , na implicação $P \rightarrow Q$, é denominado *lado esquerdo* (LE) e Q *lado direito* (LD) da implicação. Uma implicação $P \rightarrow Q$ é *supérflua* sse $P \cap Q \neq \emptyset$. O *fecho não trivial* de um conjunto de implicações \mathcal{I} é $\mathcal{I}^* = \{i \mid i \text{ segue de } \mathcal{I} \text{ e } i \text{ não é uma implicação superflua}\}$. Um conjunto de implicações \mathcal{I} é dito *fechado não trivial* se $\mathcal{I}^* = \mathcal{I}$. Note-se que se \mathcal{I} é um conjunto de implicações fechado não trivial, então \mathcal{I} complementado com todas as implicações superfluas que sigam de implicações em \mathcal{I} é fechado (simplesmente). Por outro lado, se \mathcal{I} é fechado, então retirando-se todas as implicações superfluas de \mathcal{I} , obtém-se um conjunto fechado não trivial.

Cada implicação não superflua de um contexto formal será vista como a expressão de um pedaço de conhecimento. Desta forma, se \mathcal{I} e \mathcal{I}_r são conjuntos *completos* de implicações do contexto formal (ou reticulado conceitual) original e reduzido, então \mathcal{I}^* e \mathcal{I}_r^* possuem todas as expressões de conhecimento obtidas a partir desse contexto formal, no sentido de que implicações superfluas, que não estão presentes, são realmente superfluas:

- se $P \cap Q \neq \emptyset$ e $Q \setminus P = \emptyset$ (i.e., $Q \subseteq P$), então $P \rightarrow Q$ é implicação de qualquer contexto formal, e, portanto, não expressa realmente conhecimento (do ponto de vista lógico, esse tipo de implicação é *tautológico*); e
- se $P \cap Q \neq \emptyset$ e $Q \setminus P \neq \emptyset$ e $P \rightarrow Q$ é implicação de um contexto formal, então $P \rightarrow (Q \setminus P)$ também é implicação do mesmo contexto formal.

Os seguintes tipos de conhecimento podem então ser identificados:

- **preservado:** $Pr = \mathcal{I}^* \cap \mathcal{I}_r^*$;
- **eliminado:** $El = \mathcal{I}^* \setminus \mathcal{I}_r^*$;
- **inserido:** $In = \mathcal{I}_r^* \setminus \mathcal{I}^*$;

- **transformado:**

- generalizado: $Ge = \{P \rightarrow Q \in In \mid \exists R \rightarrow S \in Pr: (P \subset R \text{ e } S \subseteq Q) \text{ ou } (P \subseteq R \text{ e } S \subset Q)\}$; e
- especializado: $Sp = \{P \rightarrow Q \in Pr \mid \exists R \rightarrow S \in El: (R \subset P \text{ e } Q \subseteq S) \text{ ou } (R \subseteq P \text{ e } Q \subset S)\}$.

Note-se que o conhecimento generalizado é um conhecimento inserido (parte de In), que é a generalização de um conhecimento presente em \mathcal{I}^* , o qual, obviamente, segue do conhecimento generalizado. Por conseguinte, trata-se de um conhecimento de In que generaliza um conhecimento de Pr . Por outro lado, o conhecimento especializado é um conhecimento preservado (parte de Pr) que é a especialização de um conhecimento presente em \mathcal{I}^* , o qual não está presente em \mathcal{I}_r^* . Portanto, é uma especialização do conhecimento presente em El . Como definido anteriormente, apenas implicações individuais podem generalizar ou especializar implicações individuais.

Como os conjuntos Pr , In e El são usualmente muito grandes, faz-se altamente conveniente representá-los por meio de conjuntos muito menores \mathcal{I}_p , \mathcal{I}_n e \mathcal{I}_e tais que $\mathcal{I}_p^* = Pr$, $\mathcal{I}_n^* = In$ e $\mathcal{I}_e^* = El$. Por outro lado, para o tipo de análise que se pretende fazer, vai ser altamente conveniente, quando se referir aos elementos dos conjuntos \mathcal{I}_p , \mathcal{I}_n , \mathcal{I}_e , Ge e Sp , mencionar apenas os *elementos atuais de \mathcal{I} e \mathcal{I}_r* que se refiram ao conhecimento preservado, novo, eliminado, generalizado e especializado. Em particular, as generalizações e especializações identificadas devem ser implicações dos elementos atuais do conjunto \mathcal{I}_r , e não qualquer elemento de \mathcal{I}_r^* .

Adicionalmente, em alguns contextos será conveniente que cada implicação represente uma unidade de informação. Por esse motivo, exige-se que o conjunto completo de implicações \mathcal{I} do contexto formal (G, M, I) tenha as seguintes características para serem utilizadas como representantes do conhecimento expresso no contexto:

- o LD de cada implicação é unitário: se $P \rightarrow m \in \mathcal{I}$, então $m \in M$;
- implicações supérfluas não são permitidas: se $P \rightarrow m \in \mathcal{I}$, então $m \notin P$; e
- especializações não são permitidas, i.e., os lados esquerdos são mínimos: se $P \rightarrow m \in \mathcal{I}$, então não existe $Q \rightarrow m \in \mathcal{I}$ tal que $Q \subset P$.

Um conjunto completo de implicações de um contexto formal com tais propriedades é o chamado *conjunto de implicações próprias* do mesmo [152].

Definição 18. *Seja \mathcal{J} o conjunto completo e fechado não trivial de implicações de um contexto formal (G, M, I) . Então o conjunto de implicações próprias de (G, M, I) é:*

$$\{P \rightarrow m \in \mathcal{J} \mid P \subseteq M \text{ e } m \in M \setminus P \text{ e } \forall Z \subset P : Z \rightarrow m \notin \mathcal{J}\}.$$

□

A partir de agora, os conceitos de generalização e especialização serão usados restritos a implicações com as duas características descritas. Segue-se que:

- $P \rightarrow m$ *generaliza* $Q \rightarrow n$ se e somente se $P \subset Q$ e $m = n$; e
- a implicação i *especializa* j se e somente se j generaliza i .

Além das duas características descritas anteriormente, um conjunto de implicações próprias \mathcal{I} possui a propriedade bem conhecida de que o conjunto de implicações que possuem um atributo m como LD é o fecho do subconjunto de implicações próprias que possuem o atributo m como LD; assim, esse subconjunto, em certo sentido, possui toda a informação relacionada a m . De forma mais geral, o conjunto de implicações que referenciam um conjunto de atributos X é o fecho do subconjunto de implicações próprias que possuem referências restritas aos atributos em X [152]; assim, tal conjunto contém toda a informação relacionada aos atributos em X . Uma consequência importante, que será chamada *propriedade do conjunto de implicações próprias (pip)* é:

se \mathcal{I} é um conjunto de implicações próprias e $i \in \mathcal{I}^$, então $i \in \mathcal{I}$ ou existe $j \in \mathcal{I}$ tal que j generaliza i .* (pip)

Essa propriedade é utilizada no Anexo A para provar algumas proposições importantes sobre a metodologia a ser proposta.

Um problema na utilização do conjunto de implicações próprias é que ele pode ser muito maior (na verdade, exponencialmente maior) do que um conjunto não redundante obtido do mesmo. Caso seu tamanho seja proibitivamente grande, pode-se adotar uma das seguintes atitudes:

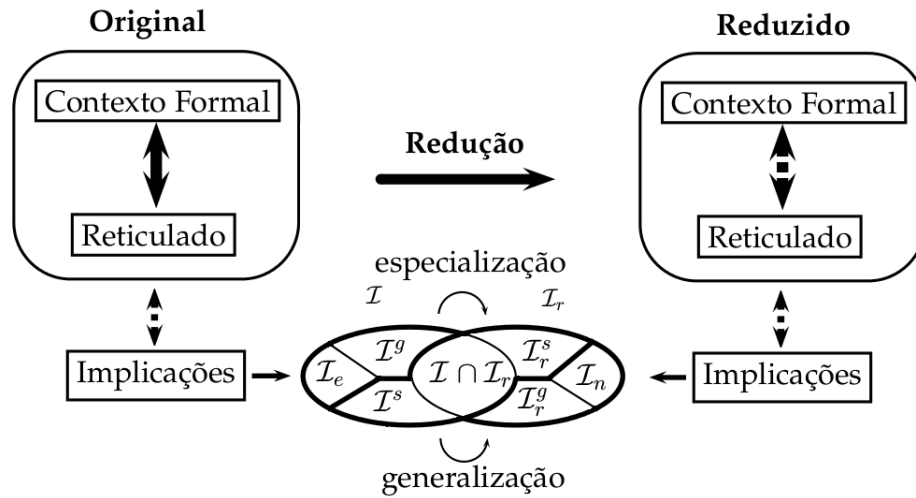
- Utilizar uma base derivada. Neste caso, nem todas as implicações estarão diretamente acessíveis. Para recuperar o conjunto de implicações próprias para um conjunto de atributos X (ou sua cardinalidade), pode-se ter que empregar pseudo-transitividade.
- Utilizar apenas implicações com suporte maior que zero (implicações $P \rightarrow m$ tais que $P' \neq \emptyset$). Isso é justificável em aplicações que julgam importantes apenas tais implicações.

Para utilizar conjuntos de implicações próprias na análise de reduções, assume-se as seguintes hipóteses:

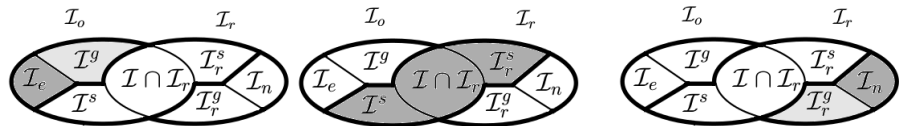
- *Nomes idênticos*: um atributo (objeto) de um contexto formal original é considerado o mesmo atributo (objeto) de um contexto formal reduzido se e somente se ambos possuem o mesmo nome em ambos os contextos formais (assim como em I e I_r). Assim, supõe-se que a técnica de redução não modifica o nome de um atributo ou objeto. Se, por exemplo, a é utilizado como atributo em I ($a \in M$), mas não é utilizado em I_r ($a \notin M_r$), considera-se que a técnica de redução *eliminou* o atributo a .
- *Redução real*: o reticulado reduzido tem que ser uma “versão reduzida” do reticulado original, de forma a ser possível, graças à primeira hipótese, distinguir elementos em comum, generalizações e especializações. Espera-se também que $M_r \setminus M$ e $G_r \setminus G$ não possuam muitos elementos (idealmente eles não terão elementos) e $\mathcal{I}_n \setminus \mathcal{I}_g$ também não possuam muitos elementos. Por fim, quanto maior a redução, maior o número de elementos em \mathcal{I}_e .

Segue-se uma metodologia de análise independente de técnica de redução, que é baseada no uso de implicações próprias.

A Figura 5.1 (a) mostra a relação entre os conjuntos de implicações próprias antes e depois da aplicação de uma técnica de redução. No diagrama, as setas simples expressam que a entidade na ponta pode ser obtida da entidade da base; as setas duplas, que as entidades em ambas as pontas podem ser obtidas uma a partir da outra; e as setas tracejadas expressam que, em alguns casos, uma entidade não é obtida a partir



(a) Relação entre implicações antes e depois da redução.



(b) Conhecimento eliminado. (c) Conhecimento preservado. (d) Conhecimento inserido.

Figura 5.1. Estrutura de reduções através de implicações.

da outra. Originando-se do contexto formal ou reticulado conceitual, a técnica de redução obtém um contexto formal e/ou reticulado conceitual reduzido. Os conjuntos de implicações próprias do contexto formal original e reduzido, \mathcal{I} e \mathcal{I}_r , são compostos de vários subconjuntos (disjuntos) que evidenciam o conhecimento eliminado (Figura 5.1 (b)), preservado (Figura 5.1 (c)) ou inserido (Figura 5.1 (d)):

- **Preservado:** $\mathcal{I}_p = \{i \in \mathcal{I}_r \mid i \in \mathcal{I}^*\} \cup \{i \in \mathcal{I} \mid i \in \mathcal{I}_r^*\}$.
- **Eliminado:**
 - parcialmente eliminado (fonte de especializações): $\mathcal{I}^s = \{j \in \mathcal{I} \mid \exists i \in \mathcal{I}_r : j \text{ generaliza } i\}$; e
 - totalmente eliminado: $\mathcal{I}_e = \mathcal{I} \setminus (\mathcal{I}_p \cup \mathcal{I}^s)$.
- **Inserido:**

- parcialmente novo (generalizado): $\mathcal{I}_r^g = \{i \in \mathcal{I}_r \mid \exists j \in \mathcal{I} : i \text{ generaliza } j\}$; e
- totalmente novo: $\mathcal{I}_n = \mathcal{I}_r \setminus (\mathcal{I}_p \cup \mathcal{I}_r^g)$.

- **Transformado:**

- especialização: de \mathcal{I}^s para $\mathcal{I}_r^s = \{i \in \mathcal{I}_r \mid \exists j \in \mathcal{I} : j \text{ generaliza } i\}$; e
- generalização: de $\mathcal{I}^s = \{j \in \mathcal{I} \mid \exists i \in \mathcal{I}_r : i \text{ generaliza } j\}$ para \mathcal{I}_r^g .

Como provado no Anexo A:

- $\mathcal{I} \cap \mathcal{I}_r$, \mathcal{I}^s e \mathcal{I}_r^s são conjuntos mutuamente disjuntos cuja união é \mathcal{I}_p (Proposição 1);
- \mathcal{I}_e , \mathcal{I}^g , \mathcal{I}^s , $\mathcal{I} \cap \mathcal{I}_r$, \mathcal{I}_r^s , \mathcal{I}_r^g e \mathcal{I}_n são conjuntos mutuamente disjuntos e a união deles é igual a $\mathcal{I} \cup \mathcal{I}_r$ (Proposição 2);
- Pr e \mathcal{I}_p^* diferem, eventualmente, apenas em algumas especializações em comum obtidas de implicações em $\mathcal{I}_e \cup \mathcal{I}^g$ e implicações em $\mathcal{I}_n \cup \mathcal{I}_r^g$, que estariam presentes em Pr , mas não em \mathcal{I}_p^* (Proposição 3);
- $El \subseteq (\mathcal{I}_e \cup \mathcal{I}^g)^*$ (Proposição 4); e
- $In \subseteq (\mathcal{I}_n \cup \mathcal{I}_r^g)^*$ (Proposição 5).

Portanto, quando \mathcal{I} e \mathcal{I}_r são *conjuntos de implicações próprias*, a caracterização dos diferentes tipos de informação descrita anteriormente não é exata, caso alguma especialização de uma implicação em $\mathcal{I}_e \cup \mathcal{I}^g$ é também uma especialização de uma implicação em $\mathcal{I}_n \cup \mathcal{I}_r^g$. Como já visto, as especializações que seguem de ambos os conjuntos representam um conhecimento que não é eliminado ou novo, mas preservado. Entretanto, sob as hipóteses de nomes idênticos e redução real, espera-se, e os experimentos comprovam a expectativa, que o conjunto de especializações em comum de $\mathcal{I}_e \cup \mathcal{I}^g$ e $\mathcal{I}_n \cup \mathcal{I}_r^g$ é vazio. Isso sendo verdade, as implicações efetivamente eliminadas são aquelas em $(\mathcal{I}_e \cup \mathcal{I}^g)^* \setminus \mathcal{I}_r^{s*}$ e as implicações efetivamente inseridas são aquelas em $(\mathcal{I}_n \cup \mathcal{I}_r^g)^* \setminus \mathcal{I}^{s*}$.

A composição dos subconjuntos de $\mathcal{I} \cup \mathcal{I}_r$ depende da técnica de redução aplicada. A próxima seção apresenta características gerais desses subconjuntos para cada classe de técnicas de redução.

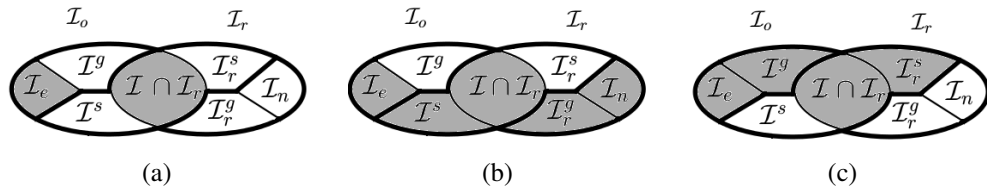


Figura 5.2. Modelos básicos de redução.

5.2.1 Características de cada Classe de Redução

A grande maioria das técnicas de redução revisadas em cada uma das classes definidas no Capítulo 3 têm características em comum quando analisadas conforme descrito na última seção. A seguir, apresentam-se as principais características de cada classe. Exemplos específicos serão apresentados em um estudo de caso na Seção 5.6.

Primeiramente, considere-se a classe de eliminação de informação redundante. Seja $X \subset M$ um conjunto tal que $\mathcal{B}(G, X, I')$ seja isomorfo a $\mathcal{B}(G, M, I)$, em que $I' = I \cap (G \times X)$. Como o reticulado gerado é isomorfo ao reticulado original, o número de conceitos permanece o mesmo. O resultado é que os atributos de $M \setminus X$ são eliminados, assim como o conhecimento associado, em conformidade com a hipótese de nomes idênticos. Como consequência, todos os subconjuntos de $\mathcal{I} \cup \mathcal{I}_r$ em destaque na Figura 5.1(a) são vazios, exceto $\mathcal{I} \cap \mathcal{I}_r$ e \mathcal{I}_e , consoante mostrado na Figura 5.2(a). Em particular:

- todo conhecimento expresso por \mathcal{I}_r é preservado: como $\mathcal{I}_r \subset \mathcal{I}$, segue-se que $\mathcal{I}_p = \mathcal{I} \cap \mathcal{I}_r = \mathcal{I}_r$;
- não existem especializações nem generalizações: $\mathcal{I}_r^s = \mathcal{I}_r^g = \emptyset$;
- todas as implicações em \mathcal{I} , exceto as presentes em \mathcal{I}_r , são eliminadas: $\mathcal{I}_e = \mathcal{I} \setminus \mathcal{I}_r$;
e
- não existe conhecimento novo inserido: $\mathcal{I}_n = \emptyset$.

Algumas técnicas com essas características são descritas por Li et al. [107], Pei & Mi [126], Zhang et al. [185] e Medina [116].

No caso de técnicas de simplificação, todas as três quantidades (números de objetos, atributos ou conceitos) podem ser reduzidas. As características compartilhadas

por muitos membros da classe de técnicas de simplificação, resumidas na Figura 5.2(b), são:

- além de $\mathcal{I} \cap \mathcal{I}_r$, implicações generalizadas (\mathcal{I}^s) expressam conhecimento preservado;
- existem generalizações, mas não especializações: $\mathcal{I}_r^s \neq \emptyset$ e $\mathcal{I}_r^e = \emptyset$;
- parte do conhecimento expresso por \mathcal{I} é eliminado: $\mathcal{I}_e \neq \emptyset$; e
- conhecimento novo, diferente de generalizações, é introduzido: $\mathcal{I}_n \neq \emptyset$.

Algumas técnicas com tais características são as apresentadas em Belohlavek et al. [10], Cheung & Vogel [34], Snásel et al. [145] e Dias & Vieira [40].

Dentre as técnicas da classe de seleção existem aquelas que têm acesso explícito a todo o conhecimento do reticulado conceitual e aquelas que não têm. Se a técnica possui acesso a todos os conceitos formais, sua caracterização se dá pela Figura 5.2(a), de forma similar à da classe de eliminação de informação redundante. Isso se deve ao fato de que, mesmo que apenas alguns conceitos formais sejam selecionados, todo o conhecimento está disponível para a seleção ser feita. Exemplos são as técnicas descritas por Babin & Kuznetsov [7], Belohlavek & Macko [13], Kuznetsov [101], Pernelle et al. [127] e Zhang et al. [184]. As técnicas de seleção sem acesso a todo o conhecimento efetivamente reduzem o número de conceitos e, eventualmente, o número de atributos e objetos. A caracterização das técnicas de seleção sem acesso a todo o conhecimento está mostrada na Figura 5.2(c):

- \mathcal{I}_r contém apenas conhecimento preservado: $\mathcal{I}_r = (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}_r^s = \mathcal{I}_p$;
- existem especializações, mas não generalizações: $\mathcal{I}_r^e \neq \emptyset$ e $\mathcal{I}_r^s = \emptyset$;
- parte do conhecimento expresso por \mathcal{I} é eliminado: $\mathcal{I}_e \neq \emptyset$; e
- nenhum conhecimento novo é introduzido: $\mathcal{I}_n = \emptyset$.

Algumas técnicas com as características descritas são aquelas apresentadas por Belohlavek & Vychodil [18, 19], Boulicaut & Besson [25], Rice & Siff [137], Stumme et al. [149] e Belohlavek & Vychodil [20].

5.3 Qualidade de um Conjunto de Implicações

Na última seção, mostrou-se quais implicações do conjunto original de implicações próprias são preservadas (\mathcal{I}_p), eliminadas (\mathcal{I}_e), especializadas (de \mathcal{I}^s para \mathcal{I}_r^s), generalizadas (de \mathcal{I}^s para \mathcal{I}_r^g) e adicionadas (\mathcal{I}_n). Para se avaliar a qualidade de uma técnica de redução ou comparar o desempenho de várias técnicas, tal caracterização pode não ser suficiente. Pode ser importante também quantificar o nível de qualidade de uma ou mais partes com respeito a um ou mais critérios, dependendo da aplicação que tenha em mente.

Três índices, com características e objetivos distintos, são apresentados aqui: conteúdo de informação, fidelidade e representatividade. Um quarto índice, sem relação direta com implicações, será apresentado na próxima seção.

O *conteúdo de informação* é um índice que visa medir a quantidade de informação associada a uma expressão lógica com base na razão entre o número de atribuições de valores-verdade que fazem a expressão lógica falsa (ou verdadeira) e o número total de possíveis atribuições de valores-verdade [8]; em geral, assume-se que as atribuições sejam igualmente prováveis. Esse índice deve informar, por exemplo, que um conjunto de implicações \mathcal{I} acrescido de uma implicação que não siga de \mathcal{I} é mais informativo que \mathcal{I} . Da mesma forma, a generalização de uma implicação i deve ser considerada mais informativa que i .

Quando implicações de um contexto formal reduzido são verificadas utilizando-se os objetos do contexto formal original, elas podem eventualmente falhar, pois pode-se esperar que tal contexto menor contenha algumas imprecisões. Assim, uma possível medida de qualidade de uma redução seria a taxa de sucesso das implicações do contexto formal reduzido (ou reticulado correspondente) quando elas são aplicadas aos objetos do contexto formal original. Essa taxa é chamada *fidelidade* [40].

Um terceiro índice, *representatividade* [44], é a taxa de sucesso das implicações considerando-se apenas implicações com LE satisfeito (implicações com suporte). O cálculo dos três índices é discutido a seguir.

5.3.1 Conteúdo de Informação

O *conteúdo de informação* (de acordo com sua conceituação pela filosofia da informação [53]) de ambos os conjuntos \mathcal{I} e \mathcal{I}_r pode ser mensurado como na teoria da informação semântica clássica proposta por Bar-Hillel & Carnap [8]. Nessa teoria, uma possibilidade é considerar o conteúdo de informação de uma sentença (aqui, uma implicação) como a probabilidade de ela ser falsa: o número de seus *contra-modelos* (atribuições de valores-verdade que a tornem falsa) dividido pelo número de atribuições de valores-verdade possíveis.

Considerando-se implicações como sentenças lógicas e atributos como símbolos proposicionais, o conjunto de contra-modelos de uma implicação $P \rightarrow m$, $cm(P \rightarrow m)$, consiste de todo $X \subseteq M$ que *não* respeita a implicação, ou seja, tal que $P \subseteq X$ e $m \notin X$, i.e., $cm(P \rightarrow m) = \{X \mid P \subseteq X \subseteq M \setminus \{m\}\}$. Intuitivamente, ele consiste de todas as $(2^{|M|-(|P|+1)})$ possibilidades de atribuições de valores-verdade para os atributos em M que fazem $P \rightarrow m$ falso. Considerando cada uma das $2^{|M|}$ possibilidades como igualmente prováveis, o conteúdo de informação de $P \rightarrow m$ é $cont(P \rightarrow m) = 2^{|M|-(|P|+1)}/2^{|M|} = 1/2^{|P|+1}$. Veja-se que uma implicação *supérflua* $P \rightarrow m$ (ou seja, uma implicação em que $m \in P$) teria $cont(P \rightarrow m) = 0$, visto que tal implicação não tem contra-modelos. Além disso, como qualquer $\{m\} \cup X \subseteq M$ respeita uma implicação $P \rightarrow m$, $cont(P \rightarrow m)$ é sempre menor que 1.

Um conjunto de atributos não respeita um conjunto \mathcal{I} de implicações se e somente se ele não respeita *alguma* implicação em \mathcal{I} . Consequentemente, o conjunto de contra-modelos de $\mathcal{I} = \{P_1 \rightarrow m_1, \dots, P_n \rightarrow m_n\}$ é $cm(\mathcal{I}) = \bigcup_{i=1}^n cm(P_i \rightarrow m_i)$. Para determinar o número de contra-modelos de \mathcal{I} pode-se aplicar o princípio de inclusão e exclusão (PIE). Fazendo-se $S_1 = \sum_{i=1}^n |cm(P_i \rightarrow m_i)|$ e, para $2 \leq k \leq n$, $S_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} |cm(P_{i_1} \rightarrow m_{i_1}) \cap \dots \cap cm(P_{i_k} \rightarrow m_{i_k})|$, segue-se, pelo PIE, que o número de contra-modelos de \mathcal{I} é $|cm(\mathcal{I})| = S_1 - S_2 + \dots + (-1)^{n+1} S_n$.

Exemplo 10. Se $M = \{a, b, c\}$, existem $2^3 = 8$ possibilidades de valores-verdade para os 3 atributos. Então $cont(a \rightarrow b) = 2^{3-2}/2^3 = 2/8$, pois $a \rightarrow b$ é falsa somente se a é verdadeira e b é falsa; e c pode ser verdadeira ou falsa. A implicação $a \rightarrow c$ também tem duas possibilidades de ser falsa, mas uma delas é idêntica a uma de $a \rightarrow b$ ser falsa: quando a é verdadeira e ambos b e c são falsos; logo, $cont(\{a \rightarrow b, a \rightarrow c\}) = (2 + 2 - 1)/2^3 = 3/8$. \square

Note-se que o conjunto de atributos $\{m \mid P \rightarrow m \in \mathcal{I} \text{ para algum } P \subseteq M\}$ respeita \mathcal{I} , ou seja tal conjunto respeita todas as implicações em \mathcal{I} , ou ainda, ele não é contra-modelo de nenhuma das implicações em \mathcal{I} . Logo, $\text{cont}(\mathcal{I})$ é sempre menor que 1 para qualquer conjunto de implicações \mathcal{I} ; e o paradoxo semântico de Bar-Hillel-Carnap [57] não causa problema aqui (um conjunto de implicações como as provenientes de um contexto formal são sempre consistentes)¹.

Uma maneira concisa de representar o conjunto de todos os contra-modelos é listar apenas os atributos que, necessariamente, pertencem a cada contra-modelo (os atributos do LE) e os atributos que não pertencem a nenhum contra-modelo (o atributo do LD). A seguir, tal representação será realizada mediante um conjunto de atributos barrados e/ou não barrados; um atributo não barrado significa que o atributo ocorre em todo contra-modelo, enquanto um atributo barrado significa que o atributo não ocorre em nenhum dos contra-modelos. Por exemplo, o conjunto $\{a, b, \bar{e}\}$ representa o conjunto de todos os contra-modelos X tais que $a, b \in X$ e $e \notin X$; neste caso, se $M = \{a, b, c, d, e\}$, então $\{a, b, \bar{e}\}$ representa um total de $2^{5-3} = 4$ contra-modelos (c e d podem estar presentes ou não em cada contra-modelo do conjunto representado). Dessa forma, os contra-modelos de $\mathcal{I} = \{P_1 \rightarrow m_1, \dots, P_n \rightarrow m_n\}$ seriam representados pelo conjunto $\mathcal{N}(\mathcal{I}) = \{P_1 \cup \{\bar{m}_1\}, \dots, P_n \cup \{\bar{m}_n\}\}$. Assim, como X e Y pertencentes a $\mathcal{N}(\mathcal{I})$ são conjuntos de atributos representando, cada qual, um conjunto de contra-modelos, então se um deles contém um atributo não barrado e o outro contém o mesmo atributo, porém barrado, eles são ditos ser *contraditórios* (não têm contra-modelos em comum). Se X e Y não são contraditórios, então a *interseção de seus contra-modelos* é representada por $X \cup Y$. Por exemplo, se $X = \{a, b, \bar{c}\}$ e $Y = \{a, \bar{e}\}$, então $X \cup Y = \{a, b, \bar{c}, \bar{e}\}$ representa a interseção dos contra-modelos representados por X e Y (no caso, $\{\{a, b\}, \{a, b, d\}\}$). Assim, em preparação para aplicar o PIE, define-se:

$$X \cap Y = \begin{cases} \emptyset, & \text{se } X \text{ e } Y \text{ são contraditórios;} \\ X \cup Y, & \text{caso contrário.} \end{cases}$$

Agora, fazendo-se $X_1 = P_1 \cup \{\bar{m}_1\}$, ..., $X_n = P_n \cup \{\bar{m}_n\}$, os termos do PIE são: $S_1 = \sum_{i=1}^n N(X_i)$, e $S_k = \sum_{1 \leq i_1 < \dots < i_k \leq n} N(X_{i_1} \cap \dots \cap X_{i_k})$, em que se X representa um conjunto de contra-modelos, então $N(X) = 2^{|M|-|X|}$. E o número de contra-modelos de

¹O paradoxo semântico de Bar-Hillel-Carnap é o de se atribuir um conteúdo de informação máximo a uma contradição.

$$\mathcal{I} \text{ é } \mathcal{N}(\mathcal{I}) = \sum_{i=1}^n (-1)^{n+1} S_n.$$

Exemplo 11. Suponha $M = \{a, b, c, d, e\}$ e $\mathcal{I} = \{a \rightarrow d, ab \rightarrow c, ace \rightarrow b, bd \rightarrow e\}$. Tem-se:

- $\mathcal{N}(\mathcal{I}) = \{\{a, \bar{d}\}, \{a, b, \bar{c}\}, \{a, \bar{b}, c, e\}, \{b, d, \bar{e}\}\}.$

- Os contra-modelos para \mathcal{I} são:

- $N(\{a, \bar{d}\}) = 2^{5-2} = 8$ contra-modelos:

$$\{a\}, \{a, b\}, \{a, c\}, \{a, e\}, \{a, b, c\}, \{a, b, e\}, \{a, c, e\}, \{a, b, c, e\}.$$

- $N(\{a, b, \bar{c}\}) = 2^{5-3} = 4$ contra-modelos:

$$\{a, b\}, \{a, b, d\}, \{a, b, e\}, \{a, b, d, e\}.$$

- $N(\{a, \bar{b}, c, e\}) = 2^{4-3} = 2$ contra-modelos:

$$\{a, c, e\}, \{a, c, d, e\}.$$

- $N(\{b, d, \bar{e}\}) = 2^{5-3} = 4$ contra-modelos:

$$\{b, d\}, \{a, b, d\}, \{b, c, d\}, \{a, b, c, d\}.$$

Os contra-modelos para \mathcal{I} , listados acima, são em número de 14.

- As interseções dois a dois:

- $\{a, \bar{d}\} \cap \{a, b, \bar{c}\} = \{a, b, \bar{c}, \bar{d}\}$. Logo, $N(\{a, \bar{d}\} \cap \{a, b, \bar{c}\}) = 2^{5-4} = 2$.

- $\{a, \bar{d}\} \cap \{a, \bar{b}, c, e\} = \{a, \bar{b}, c, \bar{d}, e\}$. Logo, $N(\{a, \bar{d}\} \cap \{a, \bar{b}, c, e\}) = 2^{5-5} = 1$.

- $\{a, \bar{d}\} \cap \{b, d, \bar{e}\} = \emptyset$. Logo, $N(\{a, \bar{d}\} \cap \{b, d, \bar{e}\}) = 0$.

- $\{a, b, \bar{c}\} \cap \{a, \bar{b}, c, e\} = \emptyset$. Logo, $N(\{a, b, \bar{c}\} \cap \{a, \bar{b}, c, e\}) = 0$.

- $\{a, b, \bar{c}\} \cap \{b, d, \bar{e}\} = \{a, b, \bar{c}, d, \bar{e}\}$. Logo, $N(\{a, b, \bar{c}\} \cap \{b, d, \bar{e}\}) = 2^{5-5} = 1$.

- $\{a, \bar{b}, c, e\} \cap \{b, d, \bar{e}\} = \emptyset$. Logo, $N(\{a, \bar{b}, c, e\} \cap \{b, d, \bar{e}\}) = 0$.

- As interseções três a três são vazias, assim como a interseção dos quatro. Logo, o número de contra-modelos é $8 + 4 + 2 + 4 - (2 + 1 + 0 + 0 + 1 + 0) = 14$. \square

Obviamente, espera-se que um método de redução leve a um decréscimo do conteúdo de informação. Quanto maior a redução, maior o decréscimo com respeito ao conhecimento expresso pelo conjunto de implicações. Apenas em situações muito especiais o conteúdo de informação pode aumentar, como, por exemplo, em situações nas quais hajam poucas eliminações e muitas generalizações.

No pior caso, o cálculo do conteúdo de informação tem custo $O(2^{|X|})$, em que $X \subseteq M$ é o conjunto de atributos referenciados no conjunto de implicações, e, portanto, tem aplicabilidade limitada.

5.3.2 Fidelidade e Representatividade

As modificações feitas por técnicas de redução podem produzir imprecisões no conjunto de implicações, que são observadas quando se considera objetos em uma aplicação, ou mesmo objetos do contexto formal original. Em outras palavras, quando se considera um objeto $g \in G$, em que G é o conjunto de objetos de uma aplicação ou do contexto formal original, uma implicação $P \rightarrow m \in \mathcal{I}_r$ pode *falhar*, i.e., pode ser o caso que $g \in P'$ e $g \notin m'$ (o operador de derivação aqui é o do contexto formal reduzido). Um índice que mede a taxa de sucesso dos objetos a implicações em geral é o que Dias & Vieira [40] denominam *fidelidade* (F), que pode ser assim definida:

$$F = \frac{\sum_{g \in G} |\overline{LE^{\mathcal{I}}(g)} \cup LD^{\mathcal{I}}(g)|}{|\mathcal{I}| |G|}$$

em que \mathcal{I} é o conjunto de implicações em consideração (possivelmente após redução), $LE^{\mathcal{I}}(g) = \{P \rightarrow m \in \mathcal{I} | g \in P'\}$ e $LD^{\mathcal{I}}(g) = \{P \rightarrow m \in \mathcal{I} | g \in m'\}$. Fidelidade nada mais é do que a taxa de sucesso das implicações em \mathcal{I} com respeito aos objetos em G , considerando-se que uma implicação $P \rightarrow m$ tem sucesso para um objeto g se $g \notin P'$ or $g \in m'$.

Como apenas técnicas de simplificação podem criar conhecimento novo (principalmente generalizações), apenas tais técnicas tendem a apresentar reduções no índice de fidelidade quando se considera o conjunto de objetos do contexto formal original, uma vez que as implicações que expressam conhecimentos novos podem falhar.

Em algumas aplicações, a capacidade de uma redução em preservar o que é relevante pode ser medida pela taxa de sucesso das implicações considerando apenas obje-

tos que satisfaçam o LE das implicações do contexto formal reduzido ($LE^{\mathcal{I}}$). Com isto, objetos g que satisfaçam uma implicação $P \rightarrow m$ apenas porque $g \notin P'$ não são levadas em conta ao se calcular o índice de *representatividade*, assim definido [44]:

$$R = \frac{\sum_{g \in G} \frac{|LE^{\mathcal{I}}(g) \cap LD^{\mathcal{I}}(g)|}{|LE^{\mathcal{I}}(g)|}}{|G|}$$

supondo-se que no caso em que $LE^{\mathcal{I}}(g) = \emptyset$, $|LE^{\mathcal{I}}(g) \cap LD^{\mathcal{I}}(g)|/|LE^{\mathcal{I}}(g)|$ deve tomar o valor 0. Note-se que $|LE^{\mathcal{I}}(g) \cap LD^{\mathcal{I}}(g)|/|LE^{\mathcal{I}}(g)|$, no caso em que o contexto formal seja efetivamente representativo do universo em consideração, pode ser visto como a probabilidade de que, para o objeto g , implicações $P \rightarrow m$ tenham seu lado direito satisfeito (ou seja, $g \in m'$), dado que tenham seu LE satisfeito (ou seja, $g \in P'$). Assim, a representatividade é a média de tais probabilidades.

Exemplo 12. Seja g um objeto tal que $g' = \{a, b, c\}$. Considere as implicações $i = a \rightarrow c$, $j = a \rightarrow d$, $k = e \rightarrow c$ e $l = d \rightarrow e$. As implicações i , k e l têm sucesso para o objeto g , e a implicação j falha para o objeto g . Assim, para o objeto g três implicações, i , k e l , contribuem para incrementar o índice de fidelidade, no universo das quatro implicações; mas apenas uma implicação, i , contribui para aumentar o índice de representatividade, no universo de duas implicações, i e j . \square

A falha de uma implicação para um dado objeto causa uma redução em ambos os índices, de fidelidade e de representatividade. Uma redução adicional no índice de representatividade ocorre quando um objeto falha em dar suporte a alguma implicação. Como resultado, a representatividade é, usualmente, menor que a fidelidade. Particularmente, se uma implicação deixa de ter suporte, a representatividade decresce, mas a fidelidade não.

Com relação à complexidade, os cálculos da fidelidade e da representatividade podem ser efetuados eficientemente. Para cada objeto, deve-se acessar todas as implicações. Por conseguinte, as complexidades dos cálculos de ambos os índices é $O(|\mathcal{I}||G|)$.

5.4 Perda Descritiva

Quando o resultado de uma redução for um contexto formal modificado, pode ser apropriado mensurar a perda da capacidade de caracterizar objetos devida à perda de atribuições.

tos. Com esse intuito, Dias & Vieira [40] propõem o índice de *perda descritiva*.

Seja (G, M, I) o contexto formal original e (G_r, M_r, I_r) o contexto formal reduzido. Para definição do índice perda descritiva, PD , usa-se uma função sobrejetora $v : G \rightarrow G_r$ que mapeia cada objeto g de G em um objeto $v(g)$ de G_r como resultado de uma possível simplificação de g propiciada pela técnica de redução. Supõe-se que o objeto $v(g)$ tem como atributos um subconjunto dos atributos de g , i.e., $v(g)' \subseteq g'$. Em particular, se o objeto g é “eliminado” no processo de redução, $v(g)' = \emptyset$. Note-se que pode ocorrer que $v(g) = v(h)$ para dois objetos distintos g e h , o que pode ser interpretado como dizendo que g e h permaneceram com os mesmos atributos após o processo de redução, tornando-se indistinguíveis. Em particular, todos os objetos eliminados tornam-se indistinguíveis. A perda descritiva é dada por:

$$PD = 1 - \frac{\sum_{g \in G} \left(\frac{|v(g)'|}{|g'|} \right)}{|G|} \quad (5.1)$$

em que $|v(g)'|$ é o número de atributos do objeto reduzido $v(g)$, e $|g'|$ é o número de atributos do objeto original g . Ressalte-se que o índice PD, como definido acima, registra perdas descritivas quando atributos considerados “redundantes” na aplicação são eliminados.

Exemplo 13. Seja o contexto formal original apresentado na Tabela 2.1 e o contexto formal reduzido, apresentado na Tabela 4.1. No contexto formal reduzido, os objetos 1, 2 e 5 foram agrupados formando um objeto $H = \{1, 2, 5\}$ com atributos $H' = \{a\}$. Tem-se que $v(1) = v(2) = v(5) = H$, $v(3) = 3$ e $v(4) = 4$. O atributo b do objeto 1 é eliminado ao se formar o objeto H , o que leva a uma perda descritiva de 0.1. \square

A complexidade do cálculo da perda descritiva é $O(|G||M|)$.

5.5 Extração de Implicações de um Conjunto de Conceitos Formais

A metodologia de análise de reduções é baseada no conjunto de implicações próprias, que pode ser obtido de um contexto formal ou reticulado conceitual [22–24, 139, 152]. Entretanto, em alguns casos é necessário extrair essas implicações de um subconjunto

de conceitos formais que formam uma seção de um reticulado conceitual, e tal seção pode não formar um reticulado completo [138]. Esse é o caso para técnicas da classe de seleção que não tenham acesso a todo o reticulado original. Dessa forma, pode não ser possível obter-se um conjunto completo de implicações próprias. A seguir, propõe-se um algoritmo para extração de um subconjunto de implicações próprias com suporte maior que zero a partir de um subconjunto de conceitos formais.

Considerando um conceito formal isoladamente, o conjunto de implicações a ser obtido compartilhará as seguintes propriedades com o conjunto de implicações próprias original: 1) ele não terá especializações de suas implicações; 2) o LD de suas implicações será unitário; 3) ele será consistente com os operadores de derivação originais, quando aplicados apenas aos elementos do conceito. Adicionalmente, 4) ela terá apenas implicações com suporte maior que zero. Já ao se considerar um conjunto de conceitos, é necessário levar em conta a ordem dos conceitos formais, ainda que o conjunto não leve a um reticulado completo. Desta forma, 5) o conjunto de implicações deve refletir o reticulado formado pelo conjunto de conceitos formais selecionados.

No Algoritmo 2, as propriedades 1, 2, 3 e 4 são garantidas pelo uso adequado de geradores mínimos (Definição 8), e a propriedade 5 é garantida assegurando-se tais propriedades globalmente.

Algoritmo 2 Determina um conjunto de implicações com suporte maior que zero.

Entrada: Lista L de conceitos formais em ordenamento topológico, e conjunto de geradores mínimos $mger(A, B)$ para cada conceito formal (A, B) em L .

Saída: Conjunto de implicações \mathcal{I} , para conceitos em L , com suporte maior que zero.

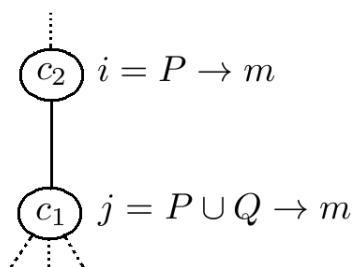
```

1:  $\mathcal{I} = \emptyset$ 
2: para cada  $(A, B)$  em  $L$  com  $A \neq \emptyset$  e  $B \neq \emptyset$  faça
3:   para cada  $P \in mger(A, B)$  faça
4:     se  $P \neq \emptyset$  então
5:       para cada  $m \in B \setminus P$  faça
6:         se  $\nexists Q \rightarrow m \in \mathcal{I} : Q \subset P$  então
7:            $\mathcal{I} = \mathcal{I} \cup \{P \rightarrow m\}$ 
8:         fim se
9:       fim para
10:    fim se
11:  fim para
12: fim para
13: retorne  $\mathcal{I}$ 

```

As entradas para o Algoritmo 2 são um ordenamento topológico (dado pela relação de cobertura do reticulado, menores intenções primeiro) dos conceitos formais selecionados e os geradores mínimos de cada conceito. Para se compreender o que resulta do processamento do “próximo conceito” (no ordenamento topológico dado), sejam c_1 e c_2 dois conceitos formais e i e j duas implicações, j uma especialização de i , obtidas dos geradores mínimos dos conceitos, como mostrado na Figura 5.3. Suponha que o algoritmo processa c_2 , inserindo a implicação i em \mathcal{I} . Se o algoritmo processa c_1 , que deve vir depois de c_2 , pois a lista de entrada está em ordenamento topológico, então a implicação j não é inserida em \mathcal{I} . Por outro lado, se c_2 não está na lista de entrada, o processamento de c_1 causa a inserção de j em \mathcal{I} ; esse é o caso em que uma especialização é produzida pelo algoritmo.

Figura 5.3. Conceitos c_1 e c_2 , tais que $c_1 \prec c_2$.



A ordem na qual os conceitos formais são explorados é importante para garantir que o LE de uma implicação seja o menor possível. No Algoritmo 2, os conceitos da lista dada são explorados na repetição externa (linha 2). Os geradores mínimos de cada conceito são processados na próxima repetição (linha 3). Os atributos da intenção que não ocorram no gerador mínimo atual são selecionados (linha 5) como LD das implicações. Uma implicação é inserida em \mathcal{I} se ela é consistente com o conjunto \mathcal{I} construído até o momento (linhas 6 e 7).

O número de iterações proporcionado pela repetição na linha 2 é $|L|$ e o proporcionado pela linha 3 é limitado pelo número máximo de geradores mínimos dos conceitos em L . Seja m tal número. Já o número de iterações propiciado pela repetição na linha 5 é limitado pelo número de atributos $|M|$. Finalmente, antes de uma implicação ser inserida em \mathcal{I} na linha 7, ela é comparada com aquelas já presentes em \mathcal{I} . Portanto, a complexidade do algoritmo é $O(|L|m|M||\mathcal{I}|)$. Para extração dos geradores mínimos

podem ser empregados, dentre outros, os algoritmos apresentados em TekayaK et al. [153] e Nehmé et al. [119].

5.6 Estudo de Caso

Nesta seção, a metodologia de análise de reduções é aplicada em um estudo de caso. O estudo consta de exemplos que atendem às hipóteses explicitadas na Subseção 5.2. Tais exemplos mostram a utilidade da metodologia na análise de reduções. Em particular, as características das classes de redução descritas na Subseção 5.2.1 são evidenciadas.

Quatro técnicas, representativas das classes de redução identificadas no Capítulo 3, foram selecionadas para análise: *seleção do conjunto minimal de atributos* [185], *agrupamento pela técnica JBOS* [40], *seleção de conceitos baseada no índice de estabilidade* [101] e *restrições determinadas a partir dos dados* [20].

Cada uma das 4 técnicas foi aplicada a três reticulados, aqueles apresentados na Figura 5.4. A Figura 5.4(a) apresenta a descrição taxonômica de um conjunto de *animais*. O reticulado possui 24 conceitos resultantes de um contexto formado por 12 objetos e 10 atributos com uma densidade de 37.5%.² O conjunto de implicações próprias é composto de 126 implicações. O contexto formal, assim como o respectivo reticulado, é um exemplo de dimensões reduzidas, utilizado para ilustrar com clareza as características da redução levada a efeito pelas técnicas em análise.

O segundo reticulado conceitual, apresentado na Figura 5.4(b), é denominado *tuberculose*, e descreve dados relativos a pacientes e seus respectivos sintomas. Esses dados foram obtidos de Horner [78], tendo sido utilizados na construção de um sistema especialista utilizando AFC. Posteriormente, foram utilizados com objetivos similares por Kumar & Srinivas [98] e Kumar [95]. O reticulado possui 137 conceitos formais originados de um contexto formal com 31 objetos, representando pacientes, e 13 atributos, representando sintomas, e tem uma densidade de 52.1%. O conjunto de implicações próprias correspondente tem 102 implicações.

Por último, tem-se o reticulado *laminação*, apresentado na Figura 5.4(c). O reticulado conceitual reflete dados relativos ao processo siderúrgico de laminação a frio e foi empregado em diversos trabalhos relacionados à AFC [45, 180–182]. O reticulado

²Densidade de um contexto formal (G, M, I) : percentagem do número de elementos em $|I|$ com relação a $|G| \times |M|$.

possui 1714 conceitos formais derivados de um contexto formal com 1691 objetos e 14 atributos, com uma densidade de 50%. O conjunto de implicações próprias é constituído de 150 implicações.

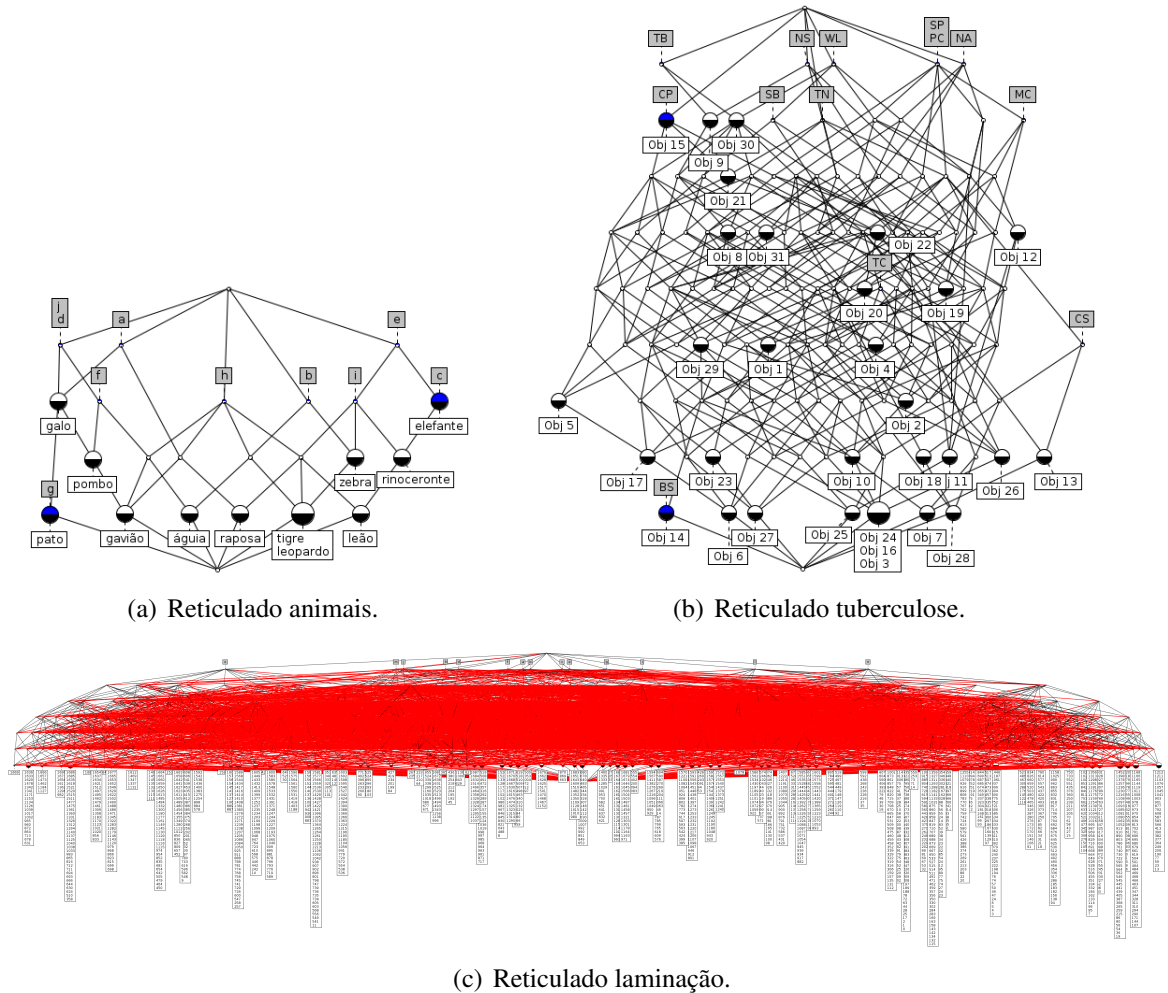


Figura 5.4. Reticulados conceituais utilizados para análise das técnicas de redução.

As características relatadas anteriormente para as três bases de dados estão resumidas na Tabela 5.1. A seguir, vem informações sobre as implicações próprias especificamente, assim como sobre os 3 índices (conteúdo de informação, fidelidade e representatividade) aplicados sobre as mesmas.

A Tabela 5.2 apresenta as informações relativas ao reticulado sobre animais. O conjunto de implicações próprias original, \mathcal{I} , contém 126 implicações e um conteúdo

Tabela 5.1. Características das três bases de dados.

Base de dados	$ G $	$ M $	Densidade (%)	#conceitos	$ \mathcal{I} $
Animais	12	10	37.5	24	126
Tuberculose	31	13	52.1	137	102
Laminação	1691	14	50	1714	150

de informação igual a 0.976. Para se permitir uma análise visual posterior da transformação do conhecimento, propiciada por uma redução, exhibe-se apenas o subconjunto formado pelas implicações com suporte maior que zero, aqui denominado $sup(\mathcal{I})$. Tal subconjunto tem 22 implicações e 0.947 de conteúdo de informação. Obviamente, tal conteúdo de informação é inferior àquele proporcionado pelo conjunto completo \mathcal{I} . Ambos os conjuntos, \mathcal{I} e $sup(\mathcal{I})$, possuem fidelidade e representatividade iguais a 1, quando se considera os objetos do contexto formal original.

Tabela 5.2. Implicações próprias para o reticulado animais.

Conjunto completo de implicações próprias				
$ \mathcal{I} = 126$				
$cont(\mathcal{I}) = 0.976$				
Conjunto de implicações próprias com suporte				
$g \rightarrow a$	$c \rightarrow e$	$a, e \rightarrow h$	$e, h \rightarrow i$	$g \rightarrow j$
$j \rightarrow d$	$d, h \rightarrow f$	$b, j \rightarrow h$	$b, e \rightarrow i$	$d \rightarrow j$
$g \rightarrow d$	$h, j \rightarrow f$	$b, d \rightarrow h$	$a, e \rightarrow i$	
$f \rightarrow d$	$b, j \rightarrow f$	$b, f \rightarrow h$	$c, h \rightarrow i$	
$i \rightarrow e$	$b, d \rightarrow f$	$a, i \rightarrow h$	$f \rightarrow j$	
$ sup(\mathcal{I}) = 22$				
$cont(sup(\mathcal{I})) = 0.947$				
$F = 1$				
$R = 1$				

As Tabelas 5.3 e 5.4 apresentam as características dos conjuntos de implicações próprias dos reticulados tuberculose e laminação, respectivamente. Como o número de implicações é grande nos dois casos, elas não são apresentadas. O conteúdo de informação das implicações com suporte maior que zero é menor que o do conjunto completo

de implicações, como esperado. A fidelidade é 1 nos dois casos, i.e., as implicações não falham. Já a representatividade para o reticulado tuberculose é 0.81, e, para o reticulado laminação, é 0.59. A representatividade é menor que 1 nos dois casos porque alguns objetos não fornecem suporte a nenhuma implicação. Um fator que também contribui para o baixo índice de representatividade para o reticulado laminação é que existe um quantitativo elevado de objetos com os mesmos atributos; o contexto formal não é clarificado.

Tabela 5.3. Conjunto de implicações próprias para o reticulado tuberculose.

Conjunto completo de implicações próprias
$ \mathcal{I} = 102$
$cont(\mathcal{I}) = 0.983$
Conjunto de implicações próprias com suporte
$ sup(\mathcal{I}) = 65$
$cont(sup(\mathcal{I})) = 0.980$
$F = 1$
$R = 0.87$

Tabela 5.4. Conjunto de implicações próprias para o reticulado laminação.

Conjunto completo de implicações próprias
$ \mathcal{I} = 150$
$cont(\mathcal{I}) = 0.895$
Conjunto de implicações próprias com suporte
$ sup(\mathcal{I}) = 39$
$cont(sup(\mathcal{I})) = 0.492$
$F = 1$
$R = 0.59$

Nas próximas seções, \mathcal{I} e \mathcal{I}_r significam, exceto quando explicitado de forma diferente, $sup(\mathcal{I})$ e $sup(\mathcal{I}_r)$, respectivamente.

5.6.1 Seleção do Conjunto Minimal de Atributos

Na classe de *eliminação de informação redundante*, conjuntos minimais podem ser obtidos a partir do uso de uma matriz de discriminação [185] (ver Definição 13). O reticulado conceitual da Figura 5.4(a), sobre animais, possui dois conjuntos minimais: $\{a, b, c, d, e, f, g, h, i\}$ e $\{a, b, c, e, f, g, h, i, j\}$. Dos reticulados correspondentes, isomorfos ao original, obtém-se um conjunto de implicações próprias com 97 implicações. Supondo-se que o reticulado seja reduzido a partir do conjunto minimal $\{a, b, c, d, e, f, g, h, i\}$, a Tabela 5.5 mostra como os conjuntos de implicações próprias original e reduzido obedecem ao esquema apresentado na Figura 5.2(a). O conjunto minimal utilizado não possui o atributo j , o que leva a uma redução no número de atributos em 10%. O número de objetos não diminui e, como o reticulado gerado é isomorfo ao original, o número de conceitos também não se altera. Na Tabela 5.5, as primeiras quatro colunas apresentam os subconjuntos das implicações originais \mathcal{I} , e as últimas quatro colunas, os subconjuntos das implicações \mathcal{I}_r . Como esperado, \mathcal{I}_r contém as implicações de \mathcal{I} , as quais referenciam apenas atributos do conjunto minimal utilizado. Em outras palavras, $\mathcal{I}_r \subset \mathcal{I}$ e algumas implicações são, simplesmente, eliminadas: $\mathcal{I}_e = \mathcal{I} \setminus \mathcal{I}_r$. De fato, não existe conhecimento transformado ($\mathcal{I}^s = \mathcal{I}_r^s = \mathcal{I}_e^s = \mathcal{I}^s = \emptyset$) e nenhum conhecimento criado ($\mathcal{I}_n = \emptyset$). Em geral, toda técnica de redução que apenas elimine atributos, mantendo a forma do reticulado inalterada, apresenta esse comportamento: nenhum conhecimento é criado e especializações ou generalizações não são produzidas; as implicações que referenciam atributos removidos são simplesmente eliminadas, como exemplificado na primeira coluna da Tabela 5.5 (todas as implicações que referenciam o atributo j foram eliminadas).

A Tabela 5.2 mostra que o conteúdo de informação das 22 implicações com suporte maior que zero é 0.947, enquanto a Tabela 5.5 mostra que o conteúdo de informação das 15 implicações retidas após a redução é 0.894. Conforme esperado, o conteúdo de informação decresceu após a redução, mesmo o reticulado reduzido sendo isomorfo ao reticulado original. A redução é consequência do corte de todas as implicações que referenciam o atributo excluído.

O índice de fidelidade manteve o valor 1 para o conjunto de implicações reduzido, visto que se trata de um subconjunto das implicações originais. Por outro lado, a representatividade decresceu de 1 para 0.91. A perda de representatividade é devida ao

Tabela 5.5. Transformação do conhecimento utilizando seleção do conjunto minimal de atributos para o reticulado animais.

\mathcal{I}_e	\mathcal{I}^g	\mathcal{I}^s	$\mathcal{I} \cap \mathcal{I}_r$	\mathcal{I}_r^s	\mathcal{I}_r^g	\mathcal{I}_n
$j \rightarrow d$ $h, j \rightarrow f$ $b, j \rightarrow f$ $b, j \rightarrow h$ $d \rightarrow j$ $f \rightarrow j$ $g \rightarrow j$	\emptyset	\emptyset	$g \rightarrow a$ $b, d \rightarrow h$ $g \rightarrow d$ $b, f \rightarrow h$ $f \rightarrow d$ $a, e \rightarrow h$ $i \rightarrow e$ $a, e \rightarrow i$ $c \rightarrow e$ $e, h \rightarrow i$ $d, h \rightarrow f$ $b, e \rightarrow i$ $b, d \rightarrow f$ $c, h \rightarrow i$ $a, i \rightarrow h$	\emptyset	\emptyset	\emptyset
$ \mathcal{I}_r = 15$ $cont(\mathcal{I}_r) = 0.894$ $F = 1$ $R = 0.91$ $PD = 0.1$				Tipos de redução Objetos: 0% Atributos: 10% Conceitos: 0%		

fato de nenhuma implicação em \mathcal{I}_r ser suportada pelo objeto *galo*, i.e. $LE^{\mathcal{I}}(\textit{galo}) = \emptyset$. As implicações $d \rightarrow j$ e $j \rightarrow d$ em \mathcal{I}_e , que eram suportadas pelo objeto referido, foram eliminadas. As implicações não falham; agora, entretanto, existe um objeto que não suporta implicação alguma. A perda descritiva foi de 0.1, uma vez que o conjunto minimal utilizado não possui o atributo *j* e, com isto, nenhum objeto do contexto formal reduzido possui tal atributo.

A Tabela 5.6 resume a transformação do conhecimento levada a efeito utilizando-se seleção do conjunto minimal de atributos para os reticulados tuberculose e laminação. Novamente, os conjuntos de implicações próprias são divididos de acordo com a Figura 5.2(a). Devido à alta quantidade de implicações, apresenta-se apenas a cardinalidade de cada conjunto. No reticulado tuberculose, o conteúdo de informação foi reduzido de 0.980 para 0.960; a fidelidade continuou máxima; a representatividade foi reduzida para 0.8 e a perda descritiva foi de 0.1. A redução do número de atributos alcançou 7%. Por outro lado, o reticulado laminação não apresentou redução nos índices, os valores são idênticos aos apresentados na Tabela 5.4. Isso ocorre porque seu conjunto de atributos é minimal e, portanto, não é possível remover qualquer atributo e obter um

reticulado isomorfo ao original.

Tabela 5.6. Transformação do conhecimento utilizando seleção do conjunto minimal de atributos para os reticulados tuberculose e laminação.

Reticulado tuberculose						
$ \mathcal{I}_e $	$ \mathcal{I}^g $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
7	0	0	58	0	0	0
$ \mathcal{I}_r = 58$			Tipos de redução Objetos: 0% Atributos: 7% Conceitos: 0%			
$cont(\mathcal{I}_r) = 0.960$						
$F = 1$						
$R = 0.8$						
$PD = 0.1$						
Reticulado laminação						
$ \mathcal{I}_e $	$ \mathcal{I}^g $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
0	0	0	39	0	0	0
$ \mathcal{I}_r = 39$			Tipos de redução Objetos: 0% Atributos: 0% Conceitos: 0%			
$cont(\mathcal{I}_r) = 0.492$						
$F = 1$						
$R = 0.58$						
$PD = 0$						

5.6.2 Agrupamento pela Técnica JBOS

JBOS (*junction based on objects similarity*), um representante da classe de técnicas de simplificação, procura substituir grupos de objetos considerados similares por objetos que os representem, baseado na importância relativa dos atributos [40], como mostrado no Capítulo 4.

Assumindo-se que o peso (relevância) de cada atributo é proporcional à frequência de sua ocorrência nos objetos, um grau de similaridade $\varepsilon = 0.6$ e um limite no número de objetos para cada grupo $\alpha = 3$, o conjunto de objetos do contexto formal do reticulado animais da Figura 5.4(a) é particionado em cinco grupos: $\{pombo, galo, pato\}$, $\{gavião, águia\}$, $\{elefante, rinoceronte, leão\}$, $\{leopardo, tigre, zebra\}$, $\{raposa\}$. O reticulado reduzido está mostrado na Figura 5.5. Comparado ao reticulado original, o número de

objetos foi reduzido em 58%, e o número de conceitos formais, pela metade. Apesar da substancial redução dos objetos e conceitos, o número de atributos remanescentes mostra-se o mesmo. Além disso, a redução lida com perdas, generalizações e inserção de novos conhecimentos como apresentado na Tabela 5.7.

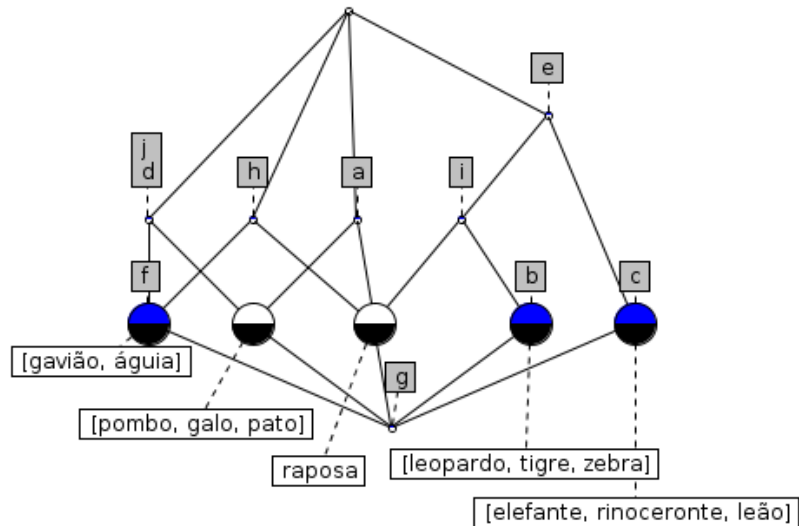


Figura 5.5. Reticulado conceitual animais reduzido pela técnica JBOS.

Quase metade das implicações em \mathcal{I} foram preservadas em \mathcal{I}_r (coluna $\mathcal{I} \cap \mathcal{I}_r$) e duas implicações em \mathcal{I}^s foram generalizadas para se tornarem novas implicações em \mathcal{I}_r^s . Além dessas duas implicações de \mathcal{I}_r^s , que expressam parcialmente novos conhecimentos, \mathcal{I}_n possui cinco novas implicações não presentes em \mathcal{I} . As implicações em \mathcal{I}_n , como são implicações que não são obtíveis do reticulado da Figura 5.4(a), representam conhecimento novo. Por outro lado, quase metade das implicações em \mathcal{I} , que formam \mathcal{I}_e , foram eliminadas e não seguem de \mathcal{I}_r . Especializações não foram observadas e, dessa forma, $\mathcal{I}_r^s = \mathcal{I}^s = \emptyset$.

O conteúdo de informação diminuiu de 0.947 para 0.923, indicando que, mesmo com a introdução de informação nova, através de generalizações e novas implicações, existe uma grande redução devido à eliminação de implicações.

O agrupamento de objetos levado a efeito pela técnica JBOS resulta no agrupamento de conceitos formais, com os objetos de cada grupo preservando apenas os atributos em comum, como pode ser visto analisando-se os reticulados das Figuras 5.4(a)

Tabela 5.7. Transformação do conhecimento utilizando a técnica JBOS para o reticulado animais.

\mathcal{I}_e	\mathcal{I}^s	\mathcal{I}^s	$\mathcal{I} \cap \mathcal{I}_r$	\mathcal{I}_r^s	\mathcal{I}_r^g	\mathcal{I}_n
$g \rightarrow a$	\emptyset	$b, f \rightarrow h$	$f \rightarrow d$	\emptyset	$f \rightarrow h$	$e, h \rightarrow a$
$g \rightarrow d$		$b, e \rightarrow i$	$j \rightarrow d$		$b \rightarrow i$	$h, i \rightarrow a$
$b, d \rightarrow f$			$c \rightarrow e$			$a, h \rightarrow e$
$b, j \rightarrow f$			$i \rightarrow e$			$b \rightarrow e$
$b, d \rightarrow h$			$d, h \rightarrow f$			$a, h \rightarrow i$
$b, j \rightarrow h$			$h, j \rightarrow f$			
$d \rightarrow j$			$a, e \rightarrow h$			
$f \rightarrow j$			$a, i \rightarrow h$			
$g \rightarrow j$			$e, h \rightarrow i$			
			$a, e \rightarrow i$			
			$a, e \rightarrow i$			
$ \mathcal{I}_r = 17$			Tipos de redução			
$cont(\mathcal{I}_r) = 0.923$						
$F = 0.94$						
$R = 0.83$						
$PD = 0.19$						
			Objetos: 58%			
			Atributos: 0%			
			Conceitos: 50%			

e 5.5. Por exemplo, o reticulado original, apresentado na Figura 5.4(a), possui os conceitos formais de extensão $\{gavião\}$ e $\{águia\}$. Esses conceitos foram agrupados, resultando em um novo conceito formal de extensão $\{[gavião, águia]\}$, cuja intenção é a interseção do conjunto de atributos de ambos $gavião$ e $águia$, como pode se vê no reticulado reduzido da Figura 5.4(a).³ A partir do gerador mínimo desse novo conceito, implicações como $a, h \rightarrow i \in \mathcal{I}_n$ e $f \rightarrow h \in \mathcal{I}_r^g$ foram geradas. Essas implicações não são obteníveis do contexto formal original. As reduções da fidelidade, de 1 para 0.94, e da representatividade, de 1 para 0.83, devem-se às implicações em \mathcal{I}_n , que falham, para alguns objetos do contexto formal original. Por exemplo, a implicação $a, h \rightarrow i \in \mathcal{I}_n$ falha para o objeto $gavião$ (o objeto $gavião$ possui o atributo a e h , mas não possui o atributo i). A diferença nos índices reflete a forma como ambos são calculados, como visto na Seção 5.3.2. A redução levou a uma perda descritiva de 0.19. Essa perda foi

³Note-se que $[gavião, águia]$ é um rótulo para um novo objeto.

resultado da eliminação de atributos não comuns a alguns objetos que formavam grupos de objetos similares. Veja-se que nenhum objeto contém o atributo g no contexto formal reduzido.

A Tabela 5.8 apresenta os resultados da transformação do conhecimento utilizando a técnica JBOS para os reticulados tuberculose e laminação. Em ambos os reticulados, foi assumido que o peso (relevância) de cada atributo é proporcional à frequência de sua ocorrência nos objetos. Para o reticulado tuberculose, foram utilizados $\varepsilon = 0.9$ e $\alpha = 4$, e, para o reticulado laminação, $\varepsilon = 0.7$ e $\alpha = 6$. O comportamento para ambos é similar àquele observado para o reticulado animais. Em particular, no reticulado tuberculose o conteúdo de informação reduziu-se de 0.980 para 0.957, a fidelidade para 0.98, e a representatividade para 0.83. Verificou-se uma redução de 32% no conjunto de objetos e 27% no total de conceitos formais. Para o reticulado laminação, o conteúdo de informação foi incrementado de 0.492 para 0.656. Esse incremento é resultado do grande número de implicações que representam os conhecimentos generalizado e novo, apresentados nas duas últimas colunas da tabela. Note-se ainda que a fidelidade foi reduzida para 0.99 e a representatividade para 0.87. O número de objetos foi reduzido em 83% e o número de conceitos, em 16%. A elevada taxa de redução no número de objetos deve-se à ocorrência de muitos objetos idênticos nos dados sobre laminação. Finalmente, a última linha da tabela mostra que os contextos laminação e tuberculose, quando reduzidos apresentaram perda descritiva de 0.04 e 0.03, respectivamente.

5.6.3 Seleção de Conceitos Baseada no Índice de Estabilidade

A seleção de conceitos baseada no índice de estabilidade [101] pressupõe acesso a todos os conceitos formais originais. O índice de estabilidade objetiva mensurar o quanto a intenção de um conceito depende de sua extensão. Tal índice, para um conceito formal (X, Y) , define-se como

$$\alpha(X, Y) = \frac{|\{C \subseteq X | C' = Y\}|}{2^{|X|}}.$$

Para a seleção, define-se um limiar tal que todo conceito com valor inferior a ele é descartado. Com um limiar de 0.5, os seguintes conceitos formais do reticulado da

Tabela 5.8. Transformação do conhecimento utilizando a técnica JBOS para os reticulados tuberculose e laminação.

Reticulado tuberculose						
$ \mathcal{I}_e $	$ \mathcal{I}^s $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
23	0	18	24	0	8	6
$ \mathcal{I}_r = 38$			Tipos de redução Objetos: 32% Atributos: 0% Conceitos: 27%			
$cont(\mathcal{I}_r) = 0.957$						
$F = 0.98$						
$R = 0.82$						
$PD = 0.04$						
Reticulado laminação						
$ \mathcal{I}_e $	$ \mathcal{I}^s $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
6	0	24	9	0	16	22
$ \mathcal{I}_r = 47$			Tipos de redução Objetos: 83% Atributos 0% Conceitos: 16%			
$cont(\mathcal{I}_r) = 0.656$						
$F = 0.99$						
$R = 0.87$						
$PD = 0.03$						

Figura 5.4 são selecionados:⁴

- ($\{\text{leopardo, tigre, raposa, leão}\}, \{e, h, i\}$);
- ($\{\text{gavião, leopardo, tigre, raposa, leão, águia}\}, \{h\}$);
- ($\{\text{leopardo, zebra, tigre, raposa, leão, rinoceronte}\}, \{e, i\}$);
- ($\{\text{pombo, galo, pato, gavião}\}, \{a, d, j\}$); e
- ($\{\text{leopardo, tigre}\}, \{b, e, h, i\}$).

Como a seleção dos conceitos ocorre após sua geração, não existe redução no número de conceitos, atributos ou objetos. Entretanto, se for considerada a porcentagem de conceitos selecionados com respeito aos dados originais, existirá uma redução de

⁴Utilizou-se o algoritmo de Roth et al. [138] para obter tais resultados, sendo que os conceitos supremo e ínfimo não foram selecionados.

79% no número de conceitos formais, 8% no número de objetos e 30% no número de atributos. Adicionalmente a esses dados, a Tabela 5.9 mostra a transformação do conhecimento realizada, implicitamente, pela seleção desses conceitos formais. Como a técnica de seleção de conceitos formais baseada no índice de estabilidade possui acesso a todo reticulado, as implicações correspondentes são aquelas de \mathcal{I} obtidas diretamente através dos geradores mínimos dos conceitos formais selecionados. Como consequência, $\mathcal{I}_r \subset \mathcal{I}$ e assim $\mathcal{I}_p = \mathcal{I} \cap \mathcal{I}_r = \mathcal{I}_r$, $\mathcal{I}_e = \mathcal{I} \setminus \mathcal{I}_r$ e $\mathcal{I}_n = \emptyset$. Generalizações e especializações não são observadas, i.e., $\mathcal{I}^s = \mathcal{I}_r^s = \mathcal{I}^g = \mathcal{I}^s = \emptyset$. Esse tipo de comportamento é comum a todas as técnicas com acesso ao conjunto total de conceitos formais, visando a selecionar um subconjunto de acordo com algum critério de relevância. Como já ressaltado na Seção 5.2.1, tais técnicas, de modo similar às de eliminação de informação redundante, seguem o modelo da Figura 5.2(a).

Tabela 5.9. Transformação do conhecimento utilizando a seleção de conceitos formais baseada no índice de estabilidade para o reticulado animais.

\mathcal{I}_e	\mathcal{I}^g	\mathcal{I}^s	$\mathcal{I} \cap \mathcal{I}_r$	\mathcal{I}_r^s	\mathcal{I}_r^g	\mathcal{I}_n
$g \rightarrow a$ $a, i \rightarrow h$ $f \rightarrow d$ $b, j \rightarrow h$ $g \rightarrow d$ $b, d \rightarrow h$ $c \rightarrow e$ $b, f \rightarrow h$ $d, h \rightarrow f$ $a, e \rightarrow i$ $h, j \rightarrow f$ $c, h \rightarrow i$ $b, d \rightarrow f$ $f \rightarrow j$ $b, j \rightarrow f$ $g \rightarrow j$ $a, e \rightarrow h$	\emptyset	\emptyset	$j \rightarrow d$ $i \rightarrow e$ $d \rightarrow j$ $e, h \rightarrow i$ $b, e \rightarrow i$	\emptyset	\emptyset	\emptyset
$ \mathcal{I}_r = 5$ $cont(\mathcal{I}_r) = 0.718$ $F = 1$ $R = 0.91$			Tipos de redução Objetos: 8% Atributos: 30% Conceitos: 79%			

Devido ao grau relativamente alto de implicações eliminadas, superior ao observado nos exemplos anteriores, o conteúdo de informação reduziu para 0.718. Como $\mathcal{I}_r \subset \mathcal{I}$, $F = 1$. E como alguns objetos não fornecem suporte a nenhuma implicação de \mathcal{I}_r , a representatividade foi reduzida para 0.91.

A Tabela 5.10 apresenta a transformação do conhecimento para os reticulados tuberculose e laminação. No reticulado tuberculose foi aplicado um índice de estabilidade igual a 0.5, e no reticulado laminação, um índice igual a 0.95. Tais índices foram escolhidos por proporcionarem, o primeiro uma redução bem acentuada (75%) e o outro uma redução baixa (18%) no número de conceitos. Os comportamentos resultantes são similares ao observado para o reticulado animais. Para o reticulado tuberculose, o conteúdo de informação foi reduzido de 0.980 para 0.925, sendo que a fidelidade e a representatividade se mantiveram constantes. O número de objetos foi reduzido em 3%, o de atributos em 15% e o de conceitos formais em 75%. Mesmo ante uma expressiva redução no número de conceitos formais, os conceitos selecionados resultaram em implicações com uma representatividade idêntica. Os objetos que não suportam nenhuma implicação são os mesmos para \mathcal{I} e \mathcal{I}_r . Assim como em quaisquer técnicas de seleção, a fidelidade é máxima ($F = 1$). Para o reticulado laminação tem-se uma redução no conteúdo de informação, de 0.491 para 0.440, e da representatividade, de 0.59 para 0.58. Apenas o número de conceitos foi reduzido em 18%. A união das extensões e das intenções dos conceitos formais selecionados resulta no conjunto completo de objetos e atributos. Em ambos os reticulados, parte do conhecimento é eliminado e parte preservado. Generalizações, especializações ou implicações novas não existem.

5.6.4 Restrições Determinadas a partir dos Dados

Uma técnica de seleção que não possui acesso concomitante a todo conhecimento é a seleção de conceitos formais baseada em restrições aplicadas durante a construção do reticulado conceitual. Exemplos de técnicas de seleção baseadas em restrições são aquelas baseadas em frequência, peso mínimo, atributos excluídos, implicações obrigatórias entre atributos, etc. [18–20, 149]. Adicionalmente às restrições, frequentemente se utiliza algum conhecimento prévio para guiar o processo de seleção. Belohlavek & Vychodil [20] apresentam métodos para aplicar restrições na extração de conceitos formais e discutem vários tipos de restrições, dentre elas, *restrições determinadas a partir dos dados*.⁵ Os dados (conhecimento adicional) são fornecidos por intermédio de um contexto formal (G_x, M, I_x) , em que M são os atributos do contexto formal original (G, M, I) . Esse contexto formal pode ser uma amostra do contexto original ou um

⁵Do inglês *constraint learned from data*.

Tabela 5.10. Transformação do conhecimento utilizando a técnica de seleção de conceitos baseada no índice de estabilidade para os reticulados tuberculose e laminação.

Reticulado tuberculose						
$ \mathcal{I}_e $	$ \mathcal{I}^s $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
33	0	0	31	0	0	0
$ \mathcal{I}_r = 31$			Tipos de redução Objetos: 3% Atributos: 15% Conceitos: 75%			
$cont(\mathcal{I}_r) = 0.925$						
$F = 1$						
$R = 0.87$						
Reticulado laminação						
$ \mathcal{I}_e $	$ \mathcal{I}^s $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
15	0	0	24	0	0	0
$ \mathcal{I}_r = 24$			Tipos de redução Objetos: 0% Atributos: 0% Conceitos: 18%			
$cont(\mathcal{I}_r) = 0.440$						
$F = 1$						
$R = 0.58$						

contexto artificial construído por um especialista. Nesse caso, a aplicação intencionada se vê interessada apenas nos conceitos formais de (G, M, I) cujas intenções são também intenções de (G_x, M, I_x) . A Tabela 5.11 apresenta um exemplo de contexto formal construído com quatro objetos artificiais, a ser utilizado com o contexto formal do reticulado animais. Com isto, os seguintes conceitos formais foram selecionados:

- $(\{\text{gavião, leopardo, tigre, raposa, leão, águia}\}, \{h\})$;
- $(\{\text{pombo, galo, pato, gavião, raposa}\}, \{a\})$;
- $(\{\text{gavião, raposa}\}, \{a, h\})$;
- $(\{\text{leopardo, tigre, raposa, leão}\}, \{e, h, i\})$;
- $(\{\text{gavião}\}, \{a, d, f, h, j\})$;
- $(\{\text{raposa}\}, \{a, e, h, i\})$;
- $(\{\text{leopardo, tigre}\}, \{b, e, h, i\})$; e

- $(\{pombo, gavião\}, \{a, d, f, j\})$.

Tabela 5.11. Conhecimento prévio aplicado ao contexto formal do reticulado animais.

	a	b	c	d	e	f	g	h	i	j
g1	x			x		x				x
g2	x			x		x		x		x
g3		x			x			x	x	
g4	x				x			x	x	

Conforme anotado na Tabela 5.12, há uma redução de 63% no número de conceitos formais, 25% no número de objetos e 20%, no de atributos. Como não há acesso ao reticulado completo, \mathcal{I}_r é obtido aplicando-se o Algoritmo 2. O conjunto \mathcal{I}_r contém apenas conhecimento preservado: algumas implicações de \mathcal{I}_r são originadas de \mathcal{I} e algumas são especializações. Por exemplo, $a, f \rightarrow d \in \mathcal{I}_r^s$ é a especialização de $f \rightarrow d \in \mathcal{I}^s$. No reticulado original da Figura 5.4(a), $f \rightarrow d$ é obtido a partir do gerador mínimo do conceito cuja extensão é $\{pombo, gavião, águia\}$. Entretanto, esse conceito formal não foi selecionado. A implicação $a, f \rightarrow d$ tem origem no conceito de extensão $\{pombo, gavião\}$, que foi selecionado pela técnica. Não existem implicações novas em \mathcal{I}_n e nenhuma generalização ($\mathcal{I}^s = \mathcal{I}_r^g = \emptyset$). Muitas implicações foram eliminadas.

Devido à significativa eliminação de implicações, o conteúdo de informação caiu para 0.554. A fidelidade permaneceu igual a 1, mas a representatividade caiu para 0.5. Essa última redução é consequência do fato de os conceitos formais selecionados não apresentarem alguns objetos na extensão, quais sejam, *águia*, *rinoceronte*, *zebra*, *elefante*, *pato* e *galo*. Portanto, nenhuma implicação gerada é suportada por tais objetos.

A Tabela 5.13 apresenta a transformação do conhecimento para os reticulados tuberculose e laminação, utilizando seleção baseada em restrições determinadas a partir dos dados. Em ambos os reticulados foram utilizadas amostras aleatórias de objetos do contexto formal original como conhecimento adicional empregado como restrição. No reticulado tuberculose utilizou-se uma amostra aleatória de 48% dos objetos e no reticulado laminação uma amostra de 3% dos objetos. Os dois reticulados possuem números de objetos muito discrepantes, mas tais quantitativos proporcionam uma redução similar, de aproximadamente 50%, no número de conceitos formais. O reticulado tuberculose apresentou uma redução no conteúdo de informação de 0.980 para 0.946. A

Tabela 5.12. Transformação do conhecimento utilizando seleção de conceitos formais baseada em restrições determinadas a partir dos dados.

\mathcal{I}_e	\mathcal{I}^g	\mathcal{I}^s	$\mathcal{I} \cap \mathcal{I}_r$	\mathcal{I}_r^s	\mathcal{I}_r^g	\mathcal{I}_n	
$g \rightarrow a$ $g \rightarrow d$ $c \rightarrow e$ $b, d \rightarrow f$ $b, j \rightarrow f$ $g \rightarrow j$	$b, d \rightarrow h$ $b, f \rightarrow h$ $b, j \rightarrow h$ $c, h \rightarrow i$ $b, e \rightarrow i$	$f \rightarrow d$ $j \rightarrow d$ $i \rightarrow e$ $d, h \rightarrow f$ $h, j \rightarrow f$ $d \rightarrow j$ $f \rightarrow j$	\emptyset	$a, i \rightarrow h$ $a, e \rightarrow h$ $e, h \rightarrow i$ $a, e \rightarrow i$	$a, f \rightarrow d$ $a, h, j \rightarrow d$ $h, i \rightarrow e$ $a, i \rightarrow e$ $a, d, h \rightarrow f$ $a, h, j \rightarrow f$ $a, f \rightarrow j$ $a, d, h \rightarrow j$	\emptyset	\emptyset
$ sup(\mathcal{I}_r) = 12$			Tipos de redução				
$cont(sup(\mathcal{I}_r)) = 0.554$							
$F = 1$							
$R = 0.5$							
			Objetos 25%				
			Atributos: 20%				
			Conceitos: 62%				

representatividade e a fidelidade não apresentaram reduções. Em particular, os mesmos objetos que não fornecem suporte às implicações em \mathcal{I} não fornecem a \mathcal{I}_r . Apenas o número de conceitos formais foi reduzido em 49%. O reticulado laminação apresentou uma redução no conteúdo de informação de 0.492 para 0.374 e, na representatividade, de 0.59 para 0.45. O número de conceitos foi reduzido para 48%. O comportamento das reduções é similar ao observado para o reticulado animais. Parte do conhecimento é eliminado e parte é preservado, e especializações são observadas. Entretanto, generalizações não ocorrem.

Tabela 5.13. Transformação do conhecimento utilizando restrições determinadas a partir dos dados para os reticulados tuberculose e laminação.

Reticulado tuberculose						
$ \mathcal{I}_e $	$ \mathcal{I}^s $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
6	25	0	32	50	0	0
$ \mathcal{I}_r = 82$			Tipos de redução Objetos: 0% Atributos: 0% Conceitos: 49%			
$cont(\mathcal{I}_r) = 0.946$						
$F = 1$						
$R = 0.87$						
Reticulado laminação						
$ \mathcal{I}_e $	$ \mathcal{I}^s $	$ \mathcal{I}^s $	$ \mathcal{I} \cap \mathcal{I}_r $	$ \mathcal{I}_r^s $	$ \mathcal{I}_r^g $	$ \mathcal{I}_n $
11	16	0	12	16	0	0
$ \mathcal{I}_r = 28$			Tipos de redução Objetos: 0% Atributos: 0% Conceitos: 48%			
$cont(\mathcal{I}_r) = 0.374$						
$F = 1$						
$R = 0.45$						

5.7 Conclusões

Neste capítulo, apresentou-se uma metodologia independente da aplicação para análise de reduções. A adesão às hipóteses de nomes idênticos e reduções reais, além do uso do conjunto de implicações próprias, levaram à identificação de subconjuntos disjuntos dos conjuntos de implicações original e reduzido, a qual fornece uma visão clara dos diferentes tipos de conhecimento resultantes da transformação realizada por uma redução.

A identificação do que é preservado, eliminado, inserido ou transformado (e como) permitiu um melhor entendimento das classes de redução existentes (como definidas no Capítulo 3). De fato, além de preservar e eliminar parte do conhecimento, cada classe tem algumas características próprias. Técnicas da classe de eliminação de informação redundante têm o efeito de selecionar um subconjunto do conjunto de implicações próprias do contexto formal original. As da classe de simplificação produzem generalizações e, eventualmente, implicações que denotam conhecimento novo. Finalmente, as técnicas de seleção sem acesso prévio a todos os conceitos formais produzem

especializações, mas não generalizações, e não introduzem conhecimento novo.

Adicionalmente, quatro índices foram discutidos, cada um deles mensurando diferentes aspectos: conteúdo de informação, fidelidade, representatividade e perda descritiva. Os índices são complementares e, em conjunto à identificação dos diferentes tipos de transformação do conhecimento, auxiliam na análise da qualidade de uma redução.

Quatro técnicas de redução, representantes das três classes, ilustraram num estudo de caso a metodologia proposta. Os resultados comprovaram como diferentes classes de redução lidam com diferentes tipos de transformação do conhecimento. Assim, o método de eliminação de informação redundante baseado no conjunto minimal de atributos apenas elimina as implicações que referenciam atributos removidos. O agrupamento realizado pela técnica JBOS preserva parte das implicações, generaliza outras e cria algumas novas. Já a seleção de conceitos baseada no índice de estabilidade elimina parte do conhecimento e preserva o resto. Finalmente, a redução baseada em restrições determinadas a partir dos dados preserva, elimina e especializa partes do conhecimento.

Consoante esperado, a maioria dos exemplos lida com perdas no conteúdo de informação; apenas exemplos da classe simplificação lidam com perdas no índice de fidelidade, como consequência da produção de algumas implicações que falham; e todos apresentam perdas no índice de representatividade, pois alguns objetos passam a não dar suporte a implicações.

Capítulo 6

Conclusões

Análise formal de conceitos (AFC) propicia um importante formalismo para extração, representação e análise de conhecimento, concebido a partir da reestruturação da teoria de reticulados. Um problema chave da AFC é a relação entre a completude do conhecimento representado no reticulado conceitual e o custo computacional para gerar e manter essa estrutura. De fato, a AFC induz uma complexidade combinatória potencialmente alta e as estruturas obtidas, até mesmo de um pequeno conjunto de dados, podem tornar-se proibitivamente grandes. E, mesmo quando o número de conceitos formais não se apresenta elevado, os aspectos essenciais, aqueles efetivamente necessários, podem estar imersos em um emaranhado de detalhes irrelevantes.

Para fazer frente aos problemas decorrentes do excesso de informação presente nos reticulados conceituais, trabalhos com diferentes características são observados na literatura, poucos deles abordando o problema da qualidade dos reticulados produzidos. Observa-se também que um número significativo de técnicas se concentra exclusivamente na seleção de conceitos formais, o que, frequentemente, não redundando na produção de um reticulado completo.

6.1 Contribuições

O tema central desta tese é a redução de reticulados conceituais, de forma a preservar e/ou expor as informações mais relevantes. Dentro desse tema, três contribuições são apresentadas: análise e classificação de técnicas para redução de reticulados con-

ceituais, uma técnica de simplificação de reticulados conceituais e uma metodologia independente de aplicação para análise de reduções.

A primeira contribuição foi uma *análise e classificação de técnicas para redução de reticulados conceituais*. Quarenta técnicas de redução, selecionadas dentre as mais representativas, foram analisadas com base em sete dimensões. Três classes foram identificadas: eliminação de informação redundante, simplificação e seleção. As técnicas de eliminação de informação redundante produzem um reticulado conceitual isomorfo ao original. Já as de simplificação produzem uma abstração do reticulado conceitual original. E as técnicas de seleção atuam selecionando, em geral, conceitos formais com base em algum critério.

Na análise das técnicas construiu-se um contexto formal no qual objetos representavam técnicas e atributos suas características. A análise foi conduzida utilizando a própria AFC. Dentre os resultados, destacam-se os fatos de todas as técnicas de redução alterarem a relação de incidência e causarem algum grau de perda descritiva. Metade das técnicas de eliminação de informação redundante inicia o processo de redução a partir dos conceitos formais, sendo a outra metade, a partir de alguma extensão da AFC; além disso, três quartos utilizam algoritmos exatos e o restante utiliza métodos heurísticos. Com relação às técnicas de simplificação, a maioria utiliza o contexto formal como ponto de partida, produz um reticulado conceitual não isomorfo ao original, que não é subconjunto do reticulado original, e utiliza algoritmos exatos; adicionalmente, algumas técnicas utilizam conhecimento prévio para auxiliar no processo de redução. Por último, as técnicas de seleção iniciam o processo pelo contexto formal, reticulado ou conceitos formais; algumas utilizam conhecimento prévio e a maioria emprega algoritmos exatos.

As técnicas de simplificação e de seleção que não necessitem, estas últimas, gerar todos os conceitos formais, são aquelas que, efetivamente, podem produzir grandes reduções. Enquanto as primeiras podem ser consideradas “perigosas”, devido ao potencial de modificar substancialmente o conjunto de conceitos formais, as últimas selecionam uma parte do reticulado original, não produzindo, eventualmente, um reticulado completo.

Na literatura há uma concentração na concepção de técnicas da classe de seleção, sejam elas com acesso prévio ao conjunto dos conceitos formais ou não. De fato, tais técnicas são relevantes e apresentam aplicações importantes. Entretanto, ao se selecionar um subconjunto de conceitos formais, não necessariamente se alcança um reticulado

conceitual completo; nesse caso, pode ser necessário um pós-processamento para formar um reticulado conceitual. Além disso, aquelas técnicas que necessitem do acesso a todos os conceitos podem não ser viáveis em termos de gasto de tempo ou espaço.

Tendo em vista o problema assinalado no último parágrafo, e que certos problemas podem se beneficiar de uma efetiva abstração do reticulado original pela obtenção de um reticulado simplificado, surgiu a segunda contribuição desta tese, *uma técnica para redução de reticulados conceituais*, a qual não necessita do acesso aos conceitos formais, sendo capaz de abstrair e generalizar o conhecimento do contexto formal original, com níveis de simplificação e de complexidade controlados pelo usuário. Tal técnica, denominada JBOS (*junction based on objects similarity*), substitui grupos de objetos considerados similares por objetos representativos. A similaridade é mensurada com base, não simplesmente na presença e/ou ausência dos atributos, como desenvolvido em muitos trabalhos, mas na relevância dos atributos definida pelo usuário. A técnica JBOS, diferentemente de outras técnicas, produz efetivamente um reticulado conceitual. Através da definição de relevância dos atributos, o usuário pode controlar o nível de simplificação e de desempenho suficientes para sua aplicação.

Além de preservar e eliminar parte do conhecimento, JBOS faz generalizações e eventualmente introduz algum conhecimento novo. Em geral, esse conhecimento novo reflete algumas imprecisões provenientes das simplificações feitas. Essas imprecisões são detectadas, por exemplo, ao se aplicar o índice de fidelidade utilizando os objetos do contexto formal original. A técnica apresenta um desempenho computacional satisfatório, uma vez que não necessita ter acesso aos conceitos formais originais. Por último, produz efetivamente um contexto formal e, conseqüentemente, um reticulado conceitual reduzido.

Como estudo de caso, JBOS foi comparada experimentalmente com três técnicas clássicas de seleção em uma aplicação real. A extração de parte do conhecimento do processo siderúrgico de laminação a frio foi conduzido antes e após a aplicação da técnica JBOS e das técnicas de seleção baseadas no índice de estabilidade, em conceitos formais frequentes e em geradores mínimos. Assim como as técnicas de seleção baseadas em geradores mínimos e em estabilidade, JBOS foi capaz de preservar o comportamento da parte estudada do processo, mesmo após redução em torno de 40% do número de conceitos, enquanto que a redução baseada em conceitos frequentes se revelou inadequada para a aplicação.

A análise das técnicas de redução existentes mostrou que pouco se sabe a respeito da natureza e qualidade das estruturas reduzidas. A terceira contribuição dessa tese é uma *metodologia para análise de reduções*, capaz de identificar as transformações levadas a efeito pelas diferentes técnicas de redução. Utilizando conjuntos de implicações próprias, a metodologia permite a clara identificação do que é preservado, eliminado, inserido ou transformado (generalizado e especializado). Em particular, ela coloca em evidência os tipos de transformação levados a efeito pelas diferentes classes de técnicas de redução. Assim, ela mostra que as técnicas de eliminação de informação redundante e de seleção com acesso a todo conhecimento original selecionam um subconjunto das implicações próprias; as de simplificação, produzem generalizações e, eventualmente, implicações novas; e as técnicas de seleção sem acesso a todo o conhecimento fazem especializações e não produzem conhecimento novo.

Dependendo da tarefa intencionada, apenas a identificação do conhecimento preservado, eliminado, inserido ou transformado pode não ser suficiente para uma boa caracterização da qualidade do reticulado reduzido. Esse é o caso, por exemplo, em grandes reduções, aquelas para as quais não é possível uma análise simples das implicações que representam os diferentes tipos de conhecimento. Nesse caso, para auxiliar em uma melhor caracterização, quatro índices foram indicados, cada um mensurando um aspecto diferente: conteúdo de informação, fidelidade, representatividade e perda descritiva. Considerando uma implicação como uma expressão lógica, o primeiro índice mensura o conteúdo de informação como sendo a razão entre o número de atribuições de valores-verdade aos atributos que façam pelo menos uma implicação ser falsa e o número total de atribuições. O índice de fidelidade mensura a taxa de sucesso quando as implicações do contexto formal em questão são aplicadas a um certo conjunto de objetos. A representatividade mensura a taxa de sucesso das implicações considerando-se apenas implicações com suporte. Finalmente, a perda descritiva mensura a perda da habilidade de caracterizar o conjunto de objetos devido à eliminação de atributos.

O uso da metodologia foi exemplificado em um estudo de caso com quatro técnicas representantes das três classes de redução. Comprovou-se como diferentes classes de redução levam a diferentes tipos de transformação do conhecimento, que elas todas proporcionam decréscimo no índice de conteúdo de informação, que apenas técnicas de simplificação levam a perdas no índice de fidelidade (como consequência da produção de algumas implicações que falham), e que todas ocasionam alterações no índice de

representatividade.

Em suma, a relevância da metodologia está em que ela independe da técnica e da aplicação, proporciona a identificação do que é preservado, inserido, eliminado e transformado pelo processo de redução, e apresenta quatro índices complementares para mensurar as perdas. Há a limitação de que o conjunto de implicações próprias pode ser exponencial com relação ao conjunto de implicações não redundante obtido do mesmo. Se o tamanho é proibitivamente grande, pode-se utilizar um conjunto de implicações derivadas ou, dependendo da aplicação, apenas implicações com suporte maior que zero.

6.2 Publicações e Prêmios

Esta tese recebeu o seguinte prêmio: selecionada pelos professores avaliadores do programa de pós-graduação em Ciência da Computação (PPGCC), dentre os trabalhos apresentados na Semana de Seminários do PPGCC, em 2015, como um dos três *Melhores Trabalhos de Doutorado*.

Os resultados desta tese foram publicados por Dias & Vieira [40, 41, 42, 44], Dias et al. [46, 47] e Ch. et al. [33]. O trabalho sobre análise de reduções encontra-se em processo de revisão no periódico *Information Sciences*¹ [43].

Outros trabalhos relacionados diretamente aos resultados desta tese foram coorientações na Pontifícia Universidade Católica de Minas Gerais em análise formal de conceitos, direcionadas a alunos da graduação em Ciência da Computação e do mestrado em Informática, com resultados publicados por de Moraes et al. [38], Neto et al. [120, 121] e Jota Resende et al. [82].

6.3 Recomendações para Trabalhos Futuros

Um primeiro trabalho futuro é a inclusão de técnicas de redução que atuem diretamente nos dados originais, aqueles utilizados para gerar o contexto formal, na análise realizada no Capítulo 3. Essas técnicas podem ser colocadas em uma classe diferente das três lá identificadas. Variações não tradicionais da AFC também carecem de uma análise mais

¹<http://www.journals.elsevier.com/information-sciences/>

criteriosa, como, por exemplo, *pattern structures* [65], reticulados *fuzzy* [9], AFC *triadic* [106], etc.

Um estudo que requer aprofundamento é o do uso de conjuntos de implicações não redundantes a partir dos quais se obtenha com facilidade o conjunto das implicações próprias, assim como conjuntos de implicações com outras características, para servir como base na análise dos processos de redução. Um aprimoramento do cálculo do conteúdo de informação deve ser estudado e, eventualmente, novos índices propostos. Pode ser interessante, ainda, um estudo da aplicação de índices nos subconjuntos de $\mathcal{I} \cup \mathcal{I}_r$ para melhor caracterizar o processo de redução levado a efeito.

Por último, existe a intenção de adaptar a ideia da técnica JBOS, de substituir grupos de objetos similares por objetos representativos, para substituir grupos de conceitos similares por conceitos representativos dos mesmos.

Anexo A

Demonstrações das Proposições

Como destacado na Figura 5.1, $\mathcal{I} \cap \mathcal{I}_r$, \mathcal{I}^s e \mathcal{I}_r^s são conjuntos disjuntos cuja união é \mathcal{I}_p , como provado a seguir.

Proposição 1. $(\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s = \mathcal{I}_p$ e $\mathcal{I} \cap \mathcal{I}_r$, \mathcal{I}^s e \mathcal{I}_r^s são mutuamente disjuntos.

Prova

Para provar $\mathcal{I}_p \subseteq (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s$, seja $i \in \mathcal{I}_p$. Por definição de \mathcal{I}_p , (1) $i \in \mathcal{I}_r$ e $i \in \mathcal{I}^*$ ou (2) $i \in \mathcal{I}$ e $i \in \mathcal{I}_r^*$. Considerando os dois casos:

1. $i \in \mathcal{I}_r$ e $i \in \mathcal{I}^*$. Pelo *pip*, $i \in \mathcal{I}$ ou existe $j \in \mathcal{I}$ tal que j generaliza i . Se $i \in \mathcal{I}$, então $i \in \mathcal{I} \cap \mathcal{I}_r$ e assim $i \in \mathcal{I}_p$; caso contrário, existe $j \in \mathcal{I}$ tal que j generaliza i e, nesse caso, $i \in \mathcal{I}_r^s$. Portanto, $i \in (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}_r^s$.
2. $i \in \mathcal{I}$ e $i \in \mathcal{I}_r^*$. Pelo *pip*, $i \in \mathcal{I}_r$ ou existe $j \in \mathcal{I}_r$ tal que j generaliza i . Se $i \in \mathcal{I}_r$, então $i \in \mathcal{I} \cap \mathcal{I}_r$ e assim $i \in \mathcal{I}_p$; caso contrário, existe $j \in \mathcal{I}_r$ tal que j generaliza i e, nesse caso, $i \in \mathcal{I}^s$. Portanto, $i \in (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s$.

Assim, $i \in (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s$ e, portanto, $\mathcal{I}_p \subseteq (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s$. Agora, para provar $(\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s \subseteq \mathcal{I}_p$, seja $i \in (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s$ e considerando os três casos:

1. $i \in \mathcal{I} \cap \mathcal{I}_r$. É imediato que $i \in \mathcal{I}_p$.
2. $i \in \mathcal{I}^s$. Segue que $i \in \mathcal{I}$ e existe $j \in \mathcal{I}_r$ tal que j generaliza i . Portanto, $i \in \mathcal{I}_r^*$. Como $i \in \mathcal{I}$ e $i \in \mathcal{I}_r^*$, então $i \in \mathcal{I}_p$.

3. $i \in \mathcal{I}_r^s$. Segue que $i \in \mathcal{I}_r$ e existe $j \in \mathcal{I}$ tal que j generaliza i . Portanto, $i \in \mathcal{I}^*$. Como $i \in \mathcal{I}_r$ e $i \in \mathcal{I}^*$, então $i \in \mathcal{I}_p$.

Assim, $i \in \mathcal{I}_p$ e, portanto, $(\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s \subseteq \mathcal{I}_p$. Em conclusão, $\mathcal{I}_p = (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s$. Logo $\mathcal{I} \cap \mathcal{I}_r$, \mathcal{I}^s e \mathcal{I}_r^s serem mutuamente disjuntos é uma consequência de \mathcal{I} e \mathcal{I}_r serem conjuntos de implicações próprias:

1. Seja $i \in \mathcal{I} \cap \mathcal{I}_r$. Isso segue de que não existe $j \in \mathcal{I}_r$ tal que j generaliza i e de que não existe $j \in \mathcal{I}$ tal que j generaliza i , pois \mathcal{I} e \mathcal{I}_r são conjuntos de implicações próprias. Portanto, se $i \in \mathcal{I} \cap \mathcal{I}_r$, o $i \notin \mathcal{I}^s$ e $i \notin \mathcal{I}_r^s$.
2. Se $i \in \mathcal{I}^s$, i.e. existe $i \in \mathcal{I}$ e $j \in \mathcal{I}_r$ tal que j generaliza i , então $i \notin \mathcal{I}_r$, pois \mathcal{I}_r é um conjunto de implicações própria; assim, $i \notin \mathcal{I}_r^s$. Portanto, se $i \in \mathcal{I}^s$, então $i \notin \mathcal{I}_r^s$.

□

Proposição 2 mostra que o conjunto $\mathcal{I} \cup \mathcal{I}_r$ é a união dos conjuntos apresentados na Figura 5.1 e todos esse conjuntos são mutuamente disjuntos.

Proposição 2. Os conjuntos \mathcal{I}_e , \mathcal{I}^s , \mathcal{I}^s , $\mathcal{I} \cap \mathcal{I}_r$, \mathcal{I}_r^s , \mathcal{I}_r^s e \mathcal{I}_n são mutuamente disjuntos e a união deles é $\mathcal{I} \cup \mathcal{I}_r$.

Prova

Seja $U = \cup\{\mathcal{I}_e, \mathcal{I}^s, \mathcal{I}^s, \mathcal{I} \cap \mathcal{I}_r, \mathcal{I}_r^s, \mathcal{I}_r^s, \mathcal{I}_n\}$. Primeiro, prova-se que $\mathcal{I} \cup \mathcal{I}_r = U$. Para provar $\mathcal{I} \cup \mathcal{I}_r \subseteq U$, seja $i \in \mathcal{I} \cup \mathcal{I}_r$. Tem-se três casos:

1. $i \in \mathcal{I} \cap \mathcal{I}_r$.
2. $i \in \mathcal{I} \setminus \mathcal{I}_r$. Aqui existem três casos possíveis: se existe $j \in \mathcal{I}_r$ tal que i generaliza j , então $i \in \mathcal{I}^s$; se existe $j \in \mathcal{I}_r$ tal que j generaliza i , então $i \in \mathcal{I}^s$ e assim $i \in \mathcal{I}_p$; caso contrário, $i \in \mathcal{I} \setminus (\mathcal{I}_p \cup \mathcal{I}^s)$, i.e. $i \in \mathcal{I}_e$. Portanto, $i \in \mathcal{I}^s \cup \mathcal{I}^s \cup \mathcal{I}_e$.
3. $i \in \mathcal{I}_r \setminus \mathcal{I}$. Três casos são considerados: se existe $j \in \mathcal{I}$ tal que i generaliza j , então $i \in \mathcal{I}_r^s$; se existe $j \in \mathcal{I}$ tal que j generaliza i , então $i \in \mathcal{I}_r^s$ e assim $i \in \mathcal{I}_p$; caso contrário, $i \in \mathcal{I}_r \setminus (\mathcal{I}_p \cup \mathcal{I}_r^s)$, i.e. $i \in \mathcal{I}_n$. Portanto, $i \in \mathcal{I}_r^s \cup \mathcal{I}_r^s \cup \mathcal{I}_n$.

Portanto, em todo caso $i \in U$. Agora, para provar $U \subseteq \mathcal{I} \cup \mathcal{I}_r$, seja $i \in U$. Considerando os sete casos: se $i \in \mathcal{I} \cap \mathcal{I}_r$, então $i \in \mathcal{I} \cup \mathcal{I}_r$; nos três casos em que $i \in \mathcal{I}^s$, $i \in \mathcal{I}^s$ e

$i \in \mathcal{I}_e$, suas definições dizem que $i \in \mathcal{I}$; e nos três casos remanescentes, em que $i \in \mathcal{I}_r^g$, $i \in \mathcal{I}_r^s$ e $i \in \mathcal{I}_n$, suas definições dizem que $i \in \mathcal{I}_r$. Assim, como requerido, em todo caso $i \in \mathcal{I} \cup \mathcal{I}_r$.

Resta provar que os sete conjuntos são mutuamente disjuntos. Dada a Proposição 1 e o fato de que cada um dos três subconjuntos $\mathcal{I} \setminus \mathcal{I}_r$ (\mathcal{I}^g , \mathcal{I}^s ou \mathcal{I}_e) é disjunto de cada um dos três subconjuntos de \mathcal{I}_r (\mathcal{I}_r^g , \mathcal{I}_r^s ou \mathcal{I}_n), é suficiente mostrar que \mathcal{I}^g , \mathcal{I}^s e \mathcal{I}_e são mutuamente disjuntos e que \mathcal{I}_r^g , \mathcal{I}_r^s e \mathcal{I}_n são mutuamente disjuntos. De fato, a definição de \mathcal{I}_e implica que ele é disjunto de ambos \mathcal{I}^g e \mathcal{I}^s . Que \mathcal{I}^g é disjunto de \mathcal{I}^s segue do fato de que não é possível ter $i \in \mathcal{I}^g$ generalizando $j \in \mathcal{I}_r$ tal que ao mesmo tempo tenha $k \in \mathcal{I}_r$ generalizando i , pois em tal caso k generalizaria j , mas \mathcal{I}_r é um conjunto de implicações própria. De forma semelhante, pode se mostrar que \mathcal{I}_r^g , \mathcal{I}_r^s e \mathcal{I}_n são mutuamente disjuntos. \square

Próxima proposição mostra que Pr e \mathcal{I}_p^* diferem, eventualmente, apenas em especializações de implicações comuns em $\mathcal{I}_e \cup \mathcal{I}^g$ e $\mathcal{I}_n \cup \mathcal{I}_r^g$.

Proposição 3. (a) $\mathcal{I}_p^* \subseteq Pr$ e (b) $Pr \setminus \mathcal{I}_p^* \subseteq \{i \in Pr \mid \exists j \in \mathcal{I}_e \cup \mathcal{I}^g, k \in \mathcal{I}_n \cup \mathcal{I}_r^g : j \text{ generaliza } i \text{ e } k \text{ generaliza } i\}$.

Prova

(a) Seja i uma implicação arbitrária em \mathcal{I}_p^* . Então, pela Proposição 1 e *pip*, existe $j \in (\mathcal{I} \cap \mathcal{I}_r) \cup \mathcal{I}^s \cup \mathcal{I}_r^s$ tal que $j = i$ ou j generaliza i . Considerando os três casos:

1. $j \in \mathcal{I} \cap \mathcal{I}_r$. Então $i \in (\mathcal{I} \cap \mathcal{I}_r)^* \subseteq \mathcal{I}^* \cap \mathcal{I}_r^* = Pr$.
2. $j \in \mathcal{I}^s$. Então existe $k \in \mathcal{I}_r^g$ tal que k generaliza j . Assim $k \in \mathcal{I}_r^*$, e assim $j \in \mathcal{I}_r^*$, $j \in \mathcal{I}^* \cap \mathcal{I}_r^*$ e portanto $i \in \mathcal{I}^* \cap \mathcal{I}_r^* = Pr$.
3. $j \in \mathcal{I}_r^s$. Então existe $k \in \mathcal{I}^g$ tal que k generaliza j . Assim $k \in \mathcal{I}^*$, e assim $j \in \mathcal{I}^*$, $j \in \mathcal{I}^* \cap \mathcal{I}_r^*$ e portanto $i \in \mathcal{I}^* \cap \mathcal{I}_r^* = Pr$.

(b) Suponha $i \in Pr$, i.e. $i \in \mathcal{I}^* \cap \mathcal{I}_r^*$, e $i \notin \mathcal{I}_p^*$. Isso segue que (1) $i \in \mathcal{I}$ ou existe $j \in \mathcal{I}$ tal que j generaliza i , e (2) $i \in \mathcal{I}_r$ ou existe $j \in \mathcal{I}_r$ tal que j generaliza i . Considerando os quatro casos:

1. $i \in \mathcal{I}$ e $i \in \mathcal{I}_r$. Isso segue, pela Proposição 1, que $i \in \mathcal{I}_p$, o que contradiz $i \notin \mathcal{I}_p^*$.

2. $i \in \mathcal{I}$ e $\exists j \in \mathcal{I}_r$: j generaliza i . Isso segue que $i \in \mathcal{I}^s$ e então, pela Proposição 1, $i \in \mathcal{I}_p$, o que contradiz $i \notin \mathcal{I}_p^*$.
3. $\exists j \in \mathcal{I}$: j generaliza i e $i \in \mathcal{I}_r$. Isso segue que $i \in \mathcal{I}_r^s$ e então, pela Proposição 1, $i \in \mathcal{I}_p$, o que contradiz $i \notin \mathcal{I}_p^*$.
4. $\exists j \in \mathcal{I}$: j generaliza i e $\exists k \in \mathcal{I}_r$: k generaliza i . Uma vez que $i \notin \mathcal{I}_p^*$, $j \notin \mathcal{I}_p$ e $k \notin \mathcal{I}_p$ e então, considerando as Proposições 1 e 2, $j \in \mathcal{I}_e \cup \mathcal{I}^s$ e $k \in \mathcal{I}_n \cup \mathcal{I}_r^s$.

Portanto, se $i \in Pr \setminus \mathcal{I}_p^*$, então existe $j \in \mathcal{I} \cup \mathcal{I}^s$ tal que j generaliza i e existe $k \in \mathcal{I}_r \cup \mathcal{I}_r^s$ tal que k generaliza i . \square

Próxima proposição mostra que El está totalmente contido em $(\mathcal{I}_e \cup \mathcal{I}^s)^*$.

Proposição 4. $El \subseteq (\mathcal{I}_e \cup \mathcal{I}^s)^*$.

Prova

Suponha $i \in El$, i.e. $i \in \mathcal{I}^*$ mas $i \notin \mathcal{I}_r^*$. A partir de $i \in \mathcal{I}^*$, pela Proposição 2 e pip , existe $j \in (\mathcal{I}_e \cup \mathcal{I}^s \cup \mathcal{I}^s)$ tal que $j = i$ ou j generaliza i . Mas se $j \in \mathcal{I}^s$, então $i \in \mathcal{I}_r^{s*}$ e assim $i \in \mathcal{I}_r^*$, uma contradição. Dessa forma $j \in (\mathcal{I}_e \cup \mathcal{I}^s)$ e portanto $i \in (\mathcal{I}_e \cup \mathcal{I}^s)^*$. \square

Similiar a Proposição 4, a próxima proposição mostra que In está totalmente contido em $(\mathcal{I}_n \cup \mathcal{I}_r^s)^*$.

Proposição 5. $In \subseteq (\mathcal{I}_n \cup \mathcal{I}_r^s)^*$.

Prova

Similar a prova da Proposição 4. \square

Referências Bibliográficas

- [1] Albano, A. L. J. H. (2011). Reticulados de conceitos. Dissertação de mestrado, Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo.
- [2] Alexander, J. M. (1972). On the theory of rolling. Em *Proc. of the Royal Society of London. Series A, Mathematical and Physical Sciences*, volume 326, pp. 535–563.
- [3] Alqadah, F. & Bhatnagar, R. (2011). Similarity measures in formal concept analysis. *Annals of Mathematics and Artificial Intelligence*, 61(3):245--256. ISSN 1012-2443.
- [4] Arévalo, G.; Berry, A.; Huchard, M.; Perrot, G. & Sigayret, A. (2007). Performances of galois sub-hierarchy-building algorithms. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 166--180, Berlin, Heidelberg. Springer-Verlag.
- [5] Armstrong, W. W. (1974). Dependency Structures of Data Base Relationships. Em *IFIP congress*, volume 74, pp. 580--583. Geneva, Switzerland.
- [6] Axler, S. & Ribet, K. A. (2005). *Lattices and Ordered Algebraic Structures*. Springer-Verlag, London. ISBN 1-85233-905-5.
- [7] Babin, M. A. & Kuznetsov, S. O. (2012). Approximating concept stability. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 7--15. Springer.
- [8] Bar-Hillel, Y. & Carnap, R. (1953). Semantic information. *The British Journal for the Philosophy of Science*, 4(14):147--157.

- [9] Belohlavek, R. (2011). What is a Fuzzy Concept Lattice? II. Em *Proceedings of the 13th International Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing*, pp. 19–26, Berlin, Heidelberg. Springer-Verlag.
- [10] Belohlavek, R.; Baets, B. & Konecny, J. (2012). Zoom-In/Zoom-Out Algorithms for FCA with Attribute Granularity. Em *Computer and Information Sciences II*, pp. 549–555. Springer London.
- [11] Belohlavek, R.; De Baets, B.; Outrata, J. & Vychodil, V. (2009). Inducing decision trees via concept lattices. *International journal of general systems*, 38(4):455--467.
- [12] Belohlavek, R.; Dvorak, J. & Outrata, J. (2007). Fast factorization by similarity in formal concept analysis of data with fuzzy attributes. *Journal of Computer and System Sciences*, 73(6):1012--1022.
- [13] Belohlavek, R. & Macko, J. (2011). Selecting important concepts using weights. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 65--80, Berlin, Heidelberg. Springer-Verlag.
- [14] Belohlavek, R.; Outrata, J. & Trnecka, M. (2014). Impact of boolean factorization as preprocessing methods for classification of boolean data. *Annals of Mathematics and Artificial Intelligence*, 72(1-2):3--22.
- [15] Belohlavek, R. & Sklenar, V. (2005). Formal concept analysis over attributes with levels of granularity. Em *Computational Intelligence for Modelling, Control and Automation and International Conference on Intelligent Agents, Web Technologies and Internet Commerce*, volume 1, pp. 619 –624.
- [16] Belohlavek, R.; Sklenár, V. & Zaczal, J. (2004a). Concept Lattices Constrained by Attribute Dependencies. Em *Proceedings of Dateso, CEUR Workshop Proceedings*, pp. 56--66.
- [17] Belohlavek, R.; Sklenár, V. & Zaczal, J. (2004b). Formal concept analysis with hierarchically ordered attributes. *International Journal of General Systems*, 33:383 – 394. ISSN 0308-1079.

- [18] Belohlavek, R. & Vychodil, V. (2006). Formal Concept Analysis with Constraints by Closure Operators. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, volume 4068, pp. 131–143. Springer.
- [19] Belohlavek, R. & Vychodil, V. (2009). Formal concept analysis with background knowledge: attribute priorities. *Trans. Sys. Man Cyber Part C*, 39(4):399--409. ISSN 1094-6977.
- [20] Belohlavek, R. & Vychodil, V. (2013). Closure-based constraints in formal concept analysis. *Discrete Applied Mathematics*, 161(13–14):1894 – 1911. ISSN 0166-218X.
- [21] Belohlavek, R. & Vychodil, V. (2014). Computing minimal sets of descriptive conditions for binary data. *Int. J. General Systems*, 43(5):521--534.
- [22] Bertet, K. (2006). Some algorithmical aspects using the canonical direct implicational basis. Em *International Conference on Concept Lattices and their Applications - CLA*, pp. 101–114.
- [23] Bertet, K. (2012). The dependence Graph of a Lattice. Em *International Conference on Concept Lattices and their Applications - CLA*, volume 972, pp. 223–232. CEUR-WS.org.
- [24] Bertet, K. & Monjardet, B. (2010). The multiple facets of the canonical direct unit implicational basis. *Theoretical Computer Science*, 411(22–24):2155 – 2166. ISSN 0304-3975.
- [25] Boulicaut, J.-F. & Besson, J. (2008). Actionability and Formal Concepts: A Data Mining Perspective. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 14–31.
- [26] Brito, P. & Polaillon, G. (2011). Homogeneity and Stability in Conceptual Analysis. Em *Proceedings of international Conference on Concept Lattices and their Applications - CLA*, pp. 251–263, Nancy, France.
- [27] Carpineto, C. & Romano, G. (1997). Inferring minimal rule covers from relations. Em *Advances in Artificial Intelligence*, volume 1321, pp. 147–158. Springer Berlin Heidelberg.

- [28] Carpineto, C. & Romano, G. (2003). Mining Short-Rule Covers in Relational Databases. *Computational Intelligence*, 19(3):215–234.
- [29] Carpineto, C. & Romano, G. (2004a). *Concept Data Analysis: Theory and Applications*. John Wiley & Sons, England.
- [30] Carpineto, C. & Romano, G. (2004b). Exploiting the Potential of Concept Lattices for Information Retrieval with CREDO. *J.UCS - The Journal of Universal Computer Science*, 10(8):985–1013.
- [31] Carpineto, C.; Romano, G. & D'adamo, P. (1999). Inferring dependencies from relations: A conceptual clustering approach. *Computational Intelligence*, 15:15--4.
- [32] Cerf, L. (2010). *Constraint-Based Mining of Closed Patterns in Noisy n-ary Relations*. Tese de doutorado, National Institute for Applied Sciences, Lyon.
- [33] Ch., A. K.; Dias, S. M. & Vieira, N. J. (2015). Knowledge reduction in formal contexts using non-negative matrix factorization. *Mathematics and Computers in Simulation*, 109(0):46–63. ISSN 0378-4754.
- [34] Cheung, K. S. K. & Vogel, D. (2005). Complexity Reduction in Lattice-Based Information Retrieval. *Information Retrieval*, 8(2):285--299. ISSN 1386-4564.
- [35] Codocedo, V. & Napoli, A. (2015). Formal Concept Analysis and Information Retrieval - A Survey. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 61--77.
- [36] Codocedo, V.; Taramasco, C. & Astudill, H. (2011). Cheating to achieve Formal Concept Analysis over a large formal context. Em *Proceedings of international Conference on Concept Lattices and their Applications - CLA*, pp. 349–362, Nancy, France.
- [37] Davey, B. & Priestley, H. (1990). *Introduction to lattices and order*. Cambridge University Press, Cambridge, England.
- [38] de Moraes, N. R. M.; Dias, S. M.; Freitas, H. C. & Zárata, L. E. (2016). Parallelization of the NextClosure Algorithm for Generating the Minimum Set of Implication Rules. *Artificial Intelligence Research*, 5(2):40–54.

- [39] Dias, S. M. (2010). Algoritmos para Geração de Reticulados Conceituais. Dissertação de mestrado, Universidade Federal de Minas Gerais (UFMG), Instituto de Ciências Exatas, Departamento de Ciência da Computação, Belo Horizonte, Minas Gerais, Brasil.
- [40] Dias, S. M. & Vieira, N. J. (2010). Reducing the Size of Concept Lattices: The JBOS Approach. Em *Proceeding of the International Conference on Concept Lattices and their Applications - CLA*, volume 672, pp. 80–91, Seville, Spain.
- [41] Dias, S. M. & Vieira, N. J. (2011). Um Arcabouço para Desenvolvimento de Algoritmos da Análise Formal de Conceitos. *Revista de Informática Teórica e Aplicada - RITA*, 18:31–57.
- [42] Dias, S. M. & Vieira, N. J. (2013). Applying the JBOS reduction method for relevant knowledge extraction. *Expert Systems with Applications*, 40(5):1880–1887. ISSN 0957-4174.
- [43] Dias, S. M. & Vieira, N. J. (2015a). A methodology for analysis of concept lattices reductions. Submitted to Information Sciences.
- [44] Dias, S. M. & Vieira, N. J. (2015b). Concept lattices reduction: definition, analysis and classification. *Expert Systems with Applications*, 42(20):7084–7097. ISSN 0957-4174.
- [45] Dias, S. M. & Zárate, L. E. (2006). FCANN: An Approach to Knowledge Representation From ANN Through - FCA Effects of Synthetic Data Base and Discretization Process, Application in the Cold Rolling Process. Em *ICIT*, pp. 2013–2018, Mumbai, India. IEEE express.
- [46] Dias, S. M.; Zárate, L. E. & Vieira, N. J. (2013a). Extracting reducible knowledge from ANN with JBOS and FCANN approaches. *Expert Systems with Applications*, 40(8):3087 – 3095. ISSN 0957-4174.
- [47] Dias, S. M.; Zárate, L. E. & Vieira, N. J. (2013b). Using Iceberg Concept Lattices and Implications Rules to Extract Knowledge from ANN. *Intelligent Automation & Soft Computing*, 19(3):361–372.

- [48] Díaz-Agudo, B. & González-Calero, P. A. (2001). Formal Concept Analysis as a Support Technique for CBR. *Knowledge-based systems*, 14(3):163--171.
- [49] Distel, F. (2011). Some Complexity Results about Essential Closed Sets. Em *International Conference on Formal Concept Analysis*, volume 6628, pp. 81--92.
- [50] Domingos-Silva, J. P. & Vieira, N. J. (2008). A Classification Algorithm based on Concept Similarity. Em *Research and Development in Intelligent Systems XXIV*, pp. 281--291. Springer London.
- [51] Dong, G.; Jiang, C.; Pei, J.; Li, J. & Wong, L. (2005). Mining succinct systems of minimal generators of formal concepts. Em *Proceedings of the 10th international conference on Database Systems for Advanced Applications*, pp. 175--187, Berlin, Heidelberg. Springer-Verlag.
- [52] Du, Y. & Hai, Y. (2013). Semantic ranking of web pages based on formal concept analysis. *Journal of Systems and Software*, 86(1):187 – 197. ISSN 0164-1212.
- [53] D'Alfonso, S. (2011). On Quantifying Semantic Information. *Information*, 2(1):61--101. ISSN 2078-2489.
- [54] Eklund, P.; Wray, T.; Goodall, P. & Lawson, A. (2012). Design, information organisation and the evaluation of the Virtual Museum of the Pacific digital ecosystem. *Journal of Ambient Intelligence and Humanized Computing*, 3(4):265--280. ISSN 1868-5137.
- [55] Emilion, R. & Lévy, G. (2009). Size of random Galois lattices and number of closed frequent itemsets. *Discrete Applied Mathematics*, 157(13):2945 – 2957. ISSN 0166-218X.
- [56] Ferreira, H. C. (2004). Um Método para Análise da Laminação Tandem a Frio. Dissertação de mestrado, Escola Politécnica da Universidade do estado de São Paulo, São Paulo, Brasil.
- [57] Floridi, L. (2004). Outline of a theory of strongly semantic information. *Minds and machines*, 14(2):197--221.

- [58] Formica, A. (2006). Ontology-based concept similarity in Formal Concept Analysis. *Information Sciences*, 176(18):2624 – 2641. ISSN 0020-0255.
- [59] Formica, A. (2008). Concept similarity in Formal Concept Analysis: An information content approach. *Knowledge-Based Systems*, 21(1):80 – 87. ISSN 0950-7051.
- [60] Formica, A. (2012). Similarity reasoning for the semantic web based on fuzzy concept lattices: An informal approach. *Information Systems Frontiers*, pp. 1–10. ISSN 1387-3326.
- [61] Gaillard, E.; Lieber, J. & Nauer, E. (2011). Adaptation knowledge discovery for cooking using closed itemset extraction. Em *Proceedings of international Conference on Concept Lattices and their Applications - CLA*, pp. 87–99, Nancy, France.
- [62] Gajdos, P.; Moravec, P. & Snásel, V. (2004). Concept Lattice Generation by Singular Value Decomposition. Em *Proceedings of international Conference on Concept Lattices and their Applications - CLA*, pp. 102 –110.
- [63] Ganter, B. (1999). Attribute Exploration with Background Knowledge. *Theoretical Computer Science*, 217(2):215–233.
- [64] Ganter, B. & Kuznetsov, S. (2001a). Pattern Structures and Their Projections. Em *Conceptual Structures: Broadening the Base*, volume 2120, pp. 129–142. Springer Berlin / Heidelberg.
- [65] Ganter, B. & Kuznetsov, S. O. (2001b). Pattern structures and their projections. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, pp. 129--142. Springer-Verlag.
- [66] Ganter, B. & Stumme, G. (2003). Formal Concept Analysis: Methods and Applications in Computer Science. Relatório técnico, Otto - von - Guericke - Universität Magdeburg. Technical report.
- [67] Ganter, B. & Wille, R. (1998). *Applied lattice theory: Formal concept analysis*. Springer-Verlag, Germany.
- [68] Ganter, B. & Wille, R. (1999). *Formal Concept Analysis: Mathematical Foundations*. Springer-Verlag, Germany.

- [69] Gély, A. (2011). Links between modular decomposition of concept lattice and bimodular decomposition of a context. Em *Proceedings of international Conference on Concept Lattices and their Applications - CLA*, pp. 393–403, Nancy, France.
- [70] Godin, R. & Mili, H. (1993). Building and maintaining analysis-level class hierarchies using galois lattices. *SIGPLAN Not.*, 28:394--410. ISSN 0362-1340.
- [71] Godin, R. & Missaoui, R. (1994). An incremental concept formation approach for learning from databases. *Theor. Comput. Sci.*, 133(2):387--419. ISSN 0304-3975.
- [72] Godin, R.; Pichet, C. & Gecsei, J. (1989). Design of a browsing interface for information retrieval. *SIGIR Forum*, 23(SI):32--39. ISSN 0163-5840.
- [73] Godin, R.; Saunders, E. & Gecsei, J. (1986). Lattice model of browsable data spaces. *Information Sciences*, 40(2):89--116. ISSN 0020-0255.
- [74] Grand, B.; Soto, M. & Aufaure, M.-A. (2009). Conceptual and Spatial Footprints for Complex Systems Analysis: Application to the Semantic Web. Em *Proceedings of the 20th International Conference on Database and Expert Systems Applications, DEXA '09*, pp. 114--127, Berlin, Heidelberg. Springer-Verlag.
- [75] Grünwald, P. D. (2007). *The minimum description length principle*. MIT press.
- [76] Harzheim, E. (2005). *Ordered Sets*. Springer-Verlag, London. ISBN 0-387-24219-8.
- [77] Hermann, M. & Sertkaya, B. (2008). On the Complexity of Computing Generators of Closed Sets. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 158--168, Berlin, Heidelberg. Springer-Verlag.
- [78] Horner, V. (2007). Developing a consumer health informatics decision support system using formal concept analysis. Dissertação de mestrado, University of Pretoria.
- [79] Jain, A. K.; Murty, M. N. & Flynn, P. J. (1999). Data clustering: a review. *ACM Comput. Surv.*, 31(3):264--323. ISSN 0360-0300.
- [80] Jain, R. (1991). *The Art of Computer Systems Performance Analysis: techniques for experimental design, measurement, simulation, and modeling*. Wiley.

- [81] Jay, N.; Kohler, F. & Napoli, A. (2008). Analysis of social communities with iceberg and stability-based concept lattices. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 258--272, Berlin, Heidelberg. Springer-Verlag.
- [82] Jota Resende, G.; De Moraes, N.; Dias, S.; Marques Neto, H. & Zárata, L. (2015). Canonical Computational Models Based on Formal Concept Analysis for Social Network Analysis and Representation. Em *International Conference on Web Services (ICWS) - IEEE*, pp. 717--720.
- [83] Kardoš, F.; Pócs, J. & Pócsová, J. (2016). On concept reduction based on some graph properties. *Knowledge-Based Systems*, 93:67--74.
- [84] Kauer, M. & Krupka, M. (2015). Removing an incidence from a formal context. Em *International Conference on Concept Lattices and their Applications - CLA*, pp. 11--21.
- [85] Kaytoue, M. (2011). *Mining numerical data with formal concept analysis and pattern structures*. Tese de doutorado, Université Henri Poincaré.
- [86] Kaytoue, M.; Kuznetsov, S. O.; Napoli, A. & Duplessis, S. (2011a). Mining gene expression data with pattern structures in formal concept analysis. *Information Sciences*, 181:1989--2001. ISSN 0020-0255.
- [87] Kaytoue, M.; Marcuola, F.; Napoli, A.; Szathmary, L. & Villerd, J. (2011b). The Coron System. *CoRR*, abs/1111.5690.
- [88] King, C. S. (2004). *Complexity Reduction in Lattice-based Information Retrieval: Theory, Prototype Development, and Evaluation*. Tese de doutorado, City University of Hong Kong.
- [89] Klimushkin, M.; Obiedkov, S. & Roth, C. (2010). Approaches to the Selection of Relevant Concepts in the Case of Noisy Data. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 255--266. Springer Berlin, Heidelberg.
- [90] Koester, B. (2006). *FooCA - Web Information Retrieval with Formal Concept Analysis*. Verlag Allgemeine Wissenschaft.

- [91] Konecny, J. & Krupka, M. (2011). Block relations in fuzzy setting. Em *International Conference on Concept Lattices and their Applications*, pp. 115–130.
- [92] Krupka, M. (2009). Factorization of residuated lattices. *Logic Journal of IGPL*, 17(2):205--223.
- [93] Krupka, M. (2011). On factorization of concept lattices by incompatible tolerances. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, volume 6628, pp. 167–182. Springer Berlin Heidelberg.
- [94] Krupka, M. (2012). On complexity reduction of concept lattices: three counterexamples. *Information retrieval*, 15(2):151--156.
- [95] Kumar, C. A. (2011). Mining Association Rules Using non-negative Matrix Factorization and formal concept analysis. Em *5th International Conference on Information Processing - ICIP*, pp. 31–39.
- [96] Kumar, C. A. (2012). Fuzzy Clustering-Based Formal Concept Analysis for Association Rules Mining. *Applied Artificial Intelligence*, 26(3):274–301.
- [97] Kumar, C. A. & Srinivas, S. (2010a). Concept lattice reduction using fuzzy k-means clustering. *Expert Systems with Applications*, 37(3):2696 – 2704. ISSN 0957-4174.
- [98] Kumar, C. A. & Srinivas, S. (2010b). Mining Associations in Health Care Data Using Formal Concept Analysis and Singular Value decomposition. *Journal of Biological Systems*, 18(4):787–807.
- [99] Kuznetsov, S. (1990). Stability as an estimate of the degree of substantiation of hypotheses derived on the basis of operational similarity. *Nauchn. Tekh. Inf.*, (12):21–29.
- [100] Kuznetsov, S. (2001). On Computing the Size of a Lattice and Related Decision Problems. *Order*, 18:313–321. ISSN 0167-8094.
- [101] Kuznetsov, S. (2007). On stability of a formal concept. *Annals of Mathematics and Artificial Intelligence*, 49:101–115. ISSN 1012-2443.

- [102] Kuznetsov, S.; Obiedkov, S. & Roth, C. (2007). Reducing the Representation Complexity of Lattice-Based Taxonomies. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, volume 4604, pp. 241–254, Berlin, Heidelberg. Springer-Verlag.
- [103] Kuznetsov, S. O. (2004). On the Intractability of Computing the Duquenne-Guigues Base. *J.UCS - Journal of Universal Computer Science*, 10(8):927–933.
- [104] Kuznetsov, S. O. & Obiedkov, S. (2008). Some decision and counting problems of the Duquenne-Guigues basis of implications. *Discrete Applied Mathematics*, 156(11):1994 – 2003. ISSN 0166-218X.
- [105] Kuznetsov, S. O. & Poelmans, J. (2013). Knowledge representation and processing with formal concept analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(3):200–215. ISSN 1942-4795.
- [106] Lehmann, F. & Wille, R. (1995). A triadic approach to formal concept analysis. Em *Conceptual Structures: Applications, Implementation and Theory*, volume 954, pp. 32–43. Springer Berlin Heidelberg.
- [107] Li, J.; Mei, C. & Lv, Y. (2011a). A heuristic knowledge-reduction method for decision formal contexts. *Computers & Mathematics with Applications*, 61(4):1096 – 1106. ISSN 0898-1221.
- [108] Li, J.; Mei, C. & Lv, Y. (2011b). Knowledge reduction in decision formal contexts. *Knowledge-Based Systems*, 24(5):709 – 715. ISSN 0950-7051.
- [109] Li, J.; Mei, C. & Lv, Y. (2012). Knowledge reduction in real decision formal contexts. *Information Sciences*, 189:191--207. ISSN 0020-0255.
- [110] Li, J.; Mei, C. & Lv, Y. (2013). Incomplete decision contexts: Approximate concept construction, rule acquisition and knowledge reduction. *International Journal of Approximate Reasoning*, 54(1):149 – 165. ISSN 0888-613X.
- [111] Li, S.-T. & Tsai, F.-C. (2013). A fuzzy conceptualization model for text mining with application in opinion polarity classification. *Knowledge-Based Systems*, 39(0):23 – 33. ISSN 0950-7051.

- [112] Liu, H.; Liu, L. & Zhang, H. (2011). A fast pruning redundant rule method using galois connection. *Applied Soft Computing*, 11(1):130 – 137. ISSN 1568-4946.
- [113] Liu, J. & Mi, J.-S. (2008). A Novel Approach to Attribute Reduction in Formal Concept Lattices. Em *Rough Sets and Knowledge Technology*, volume 5009, pp. 426–433. Springer Berlin, Heidelberg.
- [114] Maier, D. (1983). *The theory of relational databases*. Computer Science Press.
- [115] McGarry, K. (2005). A survey of interestingness measures for knowledge discovery. *Knowl. Eng. Rev.*, 20(1):39--61. ISSN 0269-8889.
- [116] Medina, J. (2012). Relating attribute reduction in formal, object-oriented and property-oriented concept lattices. *Computers & Mathematics with Applications*, 64(6):1992 – 2002. ISSN 0898-1221.
- [117] Meza, R. & Buchmann, R. (2010). Real-time social networking profile information semantization using pipes and FCA. Em *Automation Quality and Testing Robotics (AQTR), 2010 IEEE International Conference on*, volume 3, pp. 1–5.
- [118] Muangprathub, J.; Boonjing, V. & Pattaraintakorn, P. (2013). A New Case-Based Classification using Incremental Concept Lattice Knowledge. *Data & Knowledge Engineering*, 83(0):39 – 53. ISSN 0169-023X.
- [119] Nehmé, K.; Valtchev, P.; Rouane, M. H. & Godin, R. (2005). On computing the minimal generator family for concept lattices and icebergs. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 192--207, Berlin, Heidelberg. Springer-Verlag.
- [120] Neto, S. M.; Zárate, L. E.; , M. A. J. S. & Dias, S. M. (2015a). Minimal Cover of Implication Rules to Represent Two Mode Networks. Em *IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pp. 211--218.
- [121] Neto, S. M.; Zárate, L. E.; , M. A. J. S. & Dias, S. M. (2015b). Using implications from FCA to represent two mode network data. Em *The 27th International Conference on Software Engineering and Knowledge Engineering (SEKE)*, pp. 253–256, Pittsburgh.

- [122] Pasquier, N.; Bastide, Y.; Taouil, R. & Lakhal, L. (1999a). Discovering Frequent Closed Itemsets for Association Rules. Em *Proceedings of the 7th International Conference on Database Theory*, pp. 398--416, London, UK, UK. Springer-Verlag.
- [123] Pasquier, N.; Bastide, Y.; Taouil, R. & Lakhal, L. (1999b). Discovering frequent closed itemsets for association rules. Em *Database Theory—ICDT'99*, pp. 398--416. Springer.
- [124] Pattaraintakorn, P.; Boonjing, V. & Tadrat, J. (2008). A New Case-Based Classifier System Using Rough Formal Concept Analysis. Em *Convergence and Hybrid Information Technology*, volume 2, pp. 645 –650.
- [125] Pei, D.; Li, M.-Z. & Mi, J.-S. (2011). Attribute reduction in fuzzy decision formal contexts. Em *2011 International Conference on Machine Learning and Cybernetics (ICMLC)*, volume 1, pp. 204 –208. ISSN 2160-133X.
- [126] Pei, D. & Mi, J.-S. (2011). Attribute reduction in decision formal context based on homomorphism. *International Journal of Machine Learning and Cybernetics*, 2:289–293. ISSN 1868-8071.
- [127] Pernelle, N.; Rousset, M.-C.; Soldano, H. & Ventos, V. (2002). ZooM: a nested Galois lattices-based system for conceptual clustering. *J. Exp. Theor. Artif. Intell.*, 14(2-3):157–187.
- [128] Poelmans, J.; Elzinga, P.; Viaene, S. & Dedene, G. (2010). Formal Concept Analysis in Knowledge Discovery: A Survey. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, pp. 139–153.
- [129] Poelmans, J.; Kuznetsov, S. O.; Ignatov, D. I. & Dedene, G. (2013). Formal Concept Analysis in Knowledge Processing: a Survey on Models and Techniques. *Expert Systems with Applications*, 40(16):6601 – 6623. ISSN 0957-4174.
- [130] Poshyvanyk, D. & Marcus, A. (2007). Combining Formal Concept Analysis with Information Retrieval for Concept Location in Source Code. Em *Program Comprehension, 2007. ICPC '07. 15th IEEE International Conference on*, pp. 37 –48. ISSN 1063-6897.

- [131] Priss, U. (2006). Formal concept analysis in information science. *Annual Rev. Info. Sci & Technol.*, 40(1):521--543. ISSN 0066-4200.
- [132] Priss, U. & Old, L. J. (2011). Data weeding techniques applied to roget's thesaurus. Em *Proceedings international conference on Knowledge processing and data analysis*, pp. 150--163, Berlin, Heidelberg. Springer-Verlag.
- [133] Qi, C.; Cui, S. & Sun, Y. (2009). Learning classification rules based on concept semilattice. Em *Computing, Communication, Control, and Management, 2009. CCCM 2009. ISECS International Colloquium on*, volume 3, pp. 221 --224.
- [134] Qi, J.-J. (2009). Attribute reduction in formal contexts based on a new discernibility matrix. *Journal of Applied Mathematics and Computing*, 30:305--314. ISSN 1598-5865.
- [135] Qu, K.-S. & Zhai, Y.-H. (2008). Generating complete set of implications for formal contexts. *Knowledge-Based Systems*, 21(5):429--433. ISSN 0950-7051.
- [136] Riadh, T. M.; Le Grand, B.; Aufaure, M.-A. & Soto, M. (2009). Conceptual and statistical footprints for social networks' characterization. Em *Proceedings of the 3rd Workshop on Social Network Mining and Analysis, SNA-KDD '09*, pp. 1--8, New York, NY, USA. ACM.
- [137] Rice, M. D. & Siff, M. (2001). Clusters, Concepts, and Pseudometrics. *Electronic Notes in Theoretical Computer Science*, 40(0):323 -- 346.
- [138] Roth, C.; Obiedkov, S. & Kourie, D. G. (2006). Towards concise representation for taxonomies of epistemic communities. Em *Proceeding of the International Conference on Concept Lattices and their Applications - CLA*, pp. 205--218.
- [139] Ryssel, U.; Distel, F. & Borchmann, D. (2014). Fast algorithms for implication bases and attribute exploration using proper premises. *Annals of Mathematics and Artificial Intelligence*, 70(1-2):25--53. ISSN 1012-2443.
- [140] Sahami, M. (1995). Learning Classification Rules Using Lattices (Extended Abstract). Em *Proceedings of the 8th European Conference on Machine Learning*, pp. 343--346, London, UK. Springer-Verlag.

- [141] Schutt, D. (1998). Abschätzungen für die Anzahl der Begriffe von Kontexten. Dissertação de mestrado, TH Darmstadt, Germany.
- [142] Sertkaya, B. (2010). A Survey on how Description Logic Ontologies Benefit from Formal Concept Analysis. Em *International Conference on Concept Lattices and their Applications - CLA*, pp. 2--21.
- [143] Shao, M.-W. (2005). The reduction for two kind of generalized concept lattices. Em *International Conference on Machine Learning and Cybernetics*, volume 4, pp. 2217 –2222x.
- [144] Snásel, V.; Dahwa Abdulla, H. & Polovincak, M. (2007). Behavior of the Concept Lattice Reduction to visualizing data after Using Matrix Decompositions. Em *International Conference on Innovations in Information Technology*, pp. 392 –396.
- [145] Snásel, V.; Polovincak, M.; Abdulla, H. M. D. & Horak, Z. (2008). On Concept Lattices and Implication Bases from Reduced Contexts. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, pp. 83–90.
- [146] Soldano, H.; Ventos, V.; Champesme, M. & Forge, D. (2010). Incremental Construction of Alpha Lattices and Association Rules. Em *Knowledge-Based and Intelligent Information and Engineering Systems*, pp. 351–360.
- [147] Stumme, G. (1996). Attribute Exploration with Background Implications and Exceptions. Em Bock, H.-H. & Polasek, W., editores, *Data Analysis and Information Systems. Statistical and Conceptual approaches*, pp. 457–469, Heidelberg. Springer.
- [148] Stumme, G. (2002). Formal Concept Analysis on its Way from Mathematics to Computer Science. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, pp. 2–19, London, UK. Springer-Verlag.
- [149] Stumme, G.; Taouil, R.; Bastide, Y.; Pasquier, N. & Lakhal, L. (2002). Computing iceberg concept lattices with Titanic. *Data and Knowledge Eng.*, 42:189–222(34).
- [150] Tadrat, J.; Boonjing, V. & Pattaraintakorn, P. (2012). A new similarity measure in formal concept analysis for case-based reasoning. *Expert Systems with Applications*, 39(1):967--972. ISSN 0957-4174.

- [151] Tan, P.-N.; Steinbach, M. & Kumar, V. (2005). *Introduction to Data Mining*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA. ISBN 0321321367.
- [152] Taouil, R. & Bastide, Y. (2001). Computing Proper Implications. Em *Proceedings of the International Conference on Conceptual Structures - ICCS*, p. 13 p, Stanford, CA US.
- [153] TekayaK, S. B.; Yahia, S. B. & Slimani, Y. (2005). GenAll Algorithm: Decorating Galois lattice with minimal generators. Em *Proceedings of international Conference on Concept Lattices and their Applications - CLA*, pp. 166–178.
- [154] Tsopze, N.; Nguifo, E. M. & Tindo, G. (2008). M-CLANN: Multi-Class Concept Lattice-based Artificial Neural Network for supervised classification. Em *18th International Conference on Artificial Neural Networks(ICANN-08)*, Prague (RTC). Springer-Verlag LNAI.
- [155] Ventos, V. & Soldano, H. (2005). Alpha galois lattices: an overview. Em *Proceedings of international conference on Formal concept analysis - ICFCA*, pp. 298--313. Springer.
- [156] Vimieiro, R. (2007). Um estudo de algoritmos para a extração de regras baseados em Análise Formal de Conceitos. Dissertação de mestrado, Universidade Federal Minas Gerais (UFMG), Belo Horizonte, Brasil.
- [157] Vimieiro, R. & Moscato, P. (2012). Mining disjunctive minimal generators with TitanicOR. *Expert Systems with Applications*, 39(9):8228 – 8238. ISSN 0957-4174.
- [158] Vo, B.; Coenen, F. & Le, B. (2013a). A new method for mining Frequent Weighted Itemsets based on WIT-trees. *Expert Systems with Applications*, 40(4):1256 – 1264. ISSN 0957-4174.
- [159] Vo, B.; Coenen, F. & Le, B. (2013b). A new method for mining frequent weighted itemsets based on wit-trees. *Expert Systems with Applications*, 40(4):1256 – 1264. ISSN 0957-4174.

- [160] Vychodil, V. (2015). On minimal sets of graded attribute implications. *Information Sciences*, 294(0):478 – 488. ISSN 0020-0255. Innovative Applications of Artificial Neural Networks in Engineering.
- [161] Wang, H. & Zhang, W.-X. (2008a). Approaches to knowledge reduction in generalized consistent decision formal context. *Mathematical and Computer Modelling*, 48(11-12):1677–1684. ISSN 0895-7177.
- [162] Wang, J.; Liang, J. & Qian, Y. (2010). A heuristic method to attribute reduction for concept lattice. Em *International Conference on Machine Learning and Cybernetics, ICMLC*, pp. 483–487.
- [163] Wang, X. & Ma, J. (2006). A Novel Approach to Attribute Reduction in Concept Lattices. Em *Rough Sets and Knowledge Technology*, volume 4062, pp. 522–529. Springer Berlin / Heidelberg.
- [164] Wang, X. & Zhang, W. (2008b). Relations of attribute reduction between object and property oriented concept lattices. *Knowledge-Based System*, 21(5):398–403.
- [165] Wang, X. & Zhang, W. (2009). Concept Lattices of Subcontexts of a Context. *Fundam. Inf.*, 90(1-2):157--169. ISSN 0169-2968.
- [166] Wang, X. & Zhang, W. (2010). Attribute reduction in concept lattices based on deletion transformations. Em *Sixth International Conference on Natural Computation - ICNC*, volume 4, pp. 2065 –2069.
- [167] Wei, L.; Qi, J. & Zhang, W. (2008). Attribute reduction theory of concept lattice based on decision formal contexts. *Science in China Series F: Information Sciences*, 51:910–923. ISSN 1009-2757.
- [168] Wille, R. (1982). Restructuring lattice theory: an approach based on hierarchies of concepts. *I. Rival (Ed.): Ordered Sets*, pp. 445–470.
- [169] Wille, R. (1996). Conceptual landscapes of knowledge: a pragmatic paradigm for knowledge processing. Fb4–preprint, TH Darmstadt.
- [170] Wille, R. (2002). Why can concept lattices support knowledge discovery in databases? *Journal of Experimental & Theoretical Artificial Intelligence*, 14(2-3):81–92.

- [171] Xie, Z.; Hsu, W.; Liu, Z. & Lee, M. L. (2002). Concept lattice based composite classifiers for high predictability. *Journal of Experimental & Theoretical Artificial Intelligence*, 14(2-3):143–156.
- [172] Yan, S.; Yun, L. & Luan, L. (2009). A Concept Similarity Method in Structural and Semantic Levels. Em *Information Science and Engineering (ISISE)*, pp. 620–623.
- [173] Yang, Y.; Webb, G. & Wu, X. (2005). Discretization methods. Em *Data Mining and Knowledge Discovery Handbook*, pp. 113–130. Springer US.
- [174] Yevtushenko, S. A. (2000). System of data analysis “Concept Explorer”. Em *Proceedings of the 7th national conference on Artificial Intelligence KII-2000*, pp. 127–134, Russia.
- [175] Yun, U.; Shin, H.; Ryu, K. H. & Yoon, E. (2012). An efficient mining algorithm for maximal weighted frequent patterns in transactional databases. *Knowledge-Based Systems*, 33(0):53 – 64. ISSN 0950-7051.
- [176] Zaki, M. & Hsiao, C. (2005). Efficient algorithms for mining closed itemsets and their lattice structure. *IEEE Transactions on Knowledge and Data Engineering*, 17(4):462 – 478. ISSN 1041-4347.
- [177] Zaki, M. J. & Hsiao, C.-J. (2002). CHARM: An Efficient Algorithm for Closed Itemset Mining. Em *SDM*, volume 2, pp. 457--473.
- [178] Zárate, L. E. (1998). *Um Método para Análise da Laminação Tandem a Frio*. Tese de doutorado, Universidade Federal de Minas Gerais, Escola de Engenharia, Belo Horizonte, MG, Brasil.
- [179] Zárate, L. E. & Bittencout, F. R. (2008). Representation and control of the cold rolling process through artificial neural networks via sensitivity factors. *Journal of Materials Processing Technology*, 197(1–3):344 – 362. ISSN 0924-0136.
- [180] Zárate, L. E. & Dias, S. M. (2006). FCANN: An Approach to Knowledge Representation From ANN Through Formal Concept Analysis - Application in the Cold Rolling Process. Em *Proc. of the 32nd Annual Conference of the IEEE Industrial Electronics Society*, pp. 3773–3778, França, Paris. IEEE express. ISSN 1553-572X.

- [181] Zárate, L. E. & Dias, S. M. (2009). Qualitative behavior rules for the cold rolling process extracted from trained ANN via the FCANN method. *Engineering Applications of Artificial Intelligence*, 22(4-5):718–731. ISSN 0952-1976.
- [182] Zárate, L. E.; Dias, S. M. & Song, M. A. J. (2008). FCANN: A new approach for extraction and representation of knowledge from ANN trained via Formal Concept Analysis. *Neurocomputing*, 71:2670–2684.
- [183] Zárate, L. E.; Pereira, E. M. D.; Oliveira, L. A. R.; Gil, V. P.; Santos, T. R. A. & Nogueira, B. (2006). Techniques for Training Sets Selection in the Representation of a Thermosiphon System Via ANN . Em *The International Joint Conference on Neural Networks*, pp. 2736– 2741. IEEE express.
- [184] Zhang, S.; Guo, P.; Zhang, J.; Wang, X. & Pedrycz, W. (2012). A completeness analysis of frequent weighted concept lattices and their algebraic properties. *Data & Knowledge Engineering*, 81-82(0):104 – 117. ISSN 0169-023X.
- [185] Zhang, W.; Wei, L. & Qi, J. (2005). Attribute reduction theory and approach to concept lattice. *Science in China Series F: Information Sciences*, 48:713–726. ISSN 1009-2757.