CARACTERIZAÇÃO DE TRANSTORNOS MENTAIS A PARTIR DE MÍDIAS SOCIAIS UTILIZANDO APRENDIZADO PROFUNDO

ANDRÉ HERMENEGILDO COSTA SILVA

CARACTERIZAÇÃO DE TRANSTORNOS MENTAIS A PARTIR DE MÍDIAS SOCIAIS UTILIZANDO APRENDIZADO PROFUNDO

Dissertação 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 Mestre em Ciência da Computação.

ORIENTADOR: ADRIANO ALONSO VELOSO

Belo Horizonte

Abril de 2016

ANDRÉ HERMENEGILDO COSTA SILVA

MENTAL DISORDER CHARACTERIZATION FROM SOCIAL MEDIA THROUGH DEEP LEARNING

Dissertation presented to the Graduate Program in Computer Science of the Universidade Federal de Minas Gerais in partial fulfillment of the requirements for the degree of Master in Computer Science.

Advisor: Adriano Alonso Veloso

Belo Horizonte

April 2016

Silva, André Hermenegildo Costa
S586m Mental Disorder Characterization from Social Media through Deep Learning / André Hermenegildo Costa Silva. — Belo Horizonte, 2016
xxviii, 45 f. : il. ; 29cm

Dissertação (mestrado) — Universidade Federal de Minas Gerais – Departamento de Ciência da Computação.
Orientador: Adriano Alonso Veloso

1. Computação - Teses. 2. Aprendizado profundo. 3. Aprendizado de máquina. 4. Transtornos mentais. 5. Mídias sociais. I. Orientador. II. Título.

CDU 519.6*82(043)

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

FOLHA DE APROVAÇÃO

Mental disorder characterization from social media through deep learning

ANDRÉ HERMENEGILDO COSTA SILVA

Dissertação defendida e aprovada pela banca examinadora constituída pelos Senhores:

PROF. ADRIANO ALONSO VELOSO - Orientador
Departamento de Ciência da Computação - UFMG

PROF. CILENE APARECIDA NUNES RODRIGUES
Departamento de Letras - PUCRJ

PROF. NIVIO ZIVIANI
Departamento de Ciência da Computação - UFMG

PROF. PEDRO OLMO STANCIOLI VAZ DE MELO
Departamento de Ciência da Computação - UFMG

Belo Horizonte, 05 de abril de 2016.

Eu dedico este trabalho aos meus pais, Celeste e João, pelo carinho, amor e suporte durante toda minha trajetória.

Agradecimentos

Mais um ciclo da minha vida está se fechando. Lembro como se fosse ontem, em Dezembro de 2013, eu tendo que decidir em qual universidade cursar o mestrado. E eu escolhi a UFMG. Esta escolha me proporcionou a melhor fase da minha vida. Pode parecer que dois anos é um período curto de tempo, mas para mim foram anos bastante intensos, onde eu aprendi muito, conheci pessoas incríveis e fiz verdadeiros amigos que levarei para sempre em meu coração.

Quanto aos agradecimentos, agradeço primeiramente a Deus, por sempre auxiliar nas minhas decisões, por me proteger e por nunca ter me deixado desistir nos momentos de dificuldade.

Agradeço profundamente aos meus pais, Celeste e João, por, mesmo que distante, sempre preocuparem comigo, me aconselharem e me tranquilizarem. Vocês são a razão do meu viver. Sem vocês eu não teria conseguido nada nessa vida. Tudo que eu faço é por vocês.

Agradeço também aos meus grandes amigos que fiz no mestrado − Artur Jordão, Clebson Cardoso, Leandro Araújo, Luis Pedraza, Rafael Lima, Thiago Rodrigues. O mestrado teria sido bem mais difícil se eu não tivesse conhecido vocês. Obrigado pela amizade, pelos conselhos, pelas prosas no almoço, pelas idas ao Cabral, pelas festas em BH. Foram momentos memoráveis que eu jamais esquecerei.

Gostaria de agradecer aos colegas do laboratório e-Speed − Alex, Alison, Dan, Denise, Derick, Felipe, Luiz Fernando, Luiz Otávio, Osvaldo, Paulo, Rubens, Vinicius − por várias vezes me ajudarem de alguma forma.

Agradeço também aos colegas que conheci e que participaram do meu dia a dia durante o mestrado − Alloma, Antônio, Edgard, Erik, Evandro, Fabrício, Gabriel, Keiller, Kevin, Natália, Iago, Susana.

Meus sinceros agradecimentos ao professor Adriano Veloso, por me orientar ao longo desta jornada e, também, pelos conhecimentos transmitidos e pelos conselhos valiosos, que serão de grande importância para o meu futuro, tanto acadêmico quanto profissional. Gostaria de agradecer também ao Michel, Roberto e ao professor Wagner

Meira, pela convivência, quase que diária, nos projetos com a Dataprev. Foram dias difíceis, porém de muito aprendizado.

"I lie down and sleep; I wake again, because the Lord sustains me."
(Psalm 3:3-5)

Resumo

Transtornos mentais têm sido motivo de preocupação ao redor do mundo. Estima-se que 54 milhões de americanos sofram de algum tipo de transtorno mental em um determinado ano. Hoje em dia, as pessoas discutem e falam sobre os mais diversos assuntos nas mídias sociais, inclusive sobre saúde, resultando em uma massiva quantidade de dados a respeito. Com isso, psiquiatras, médicos e empresas de planos de saúde têm se interessado em explorar tais tipos de dados. Recentes estudos têm focado na caracterização de saúde mental em mídias sociais e no desenvolvimento de modelos estatísticos de previsão utilizando atributos derivados de forma manual para diagnosticar o estado de saúde mental de um determinado indivíduo. No entanto, na maioria dos casos, tais atributos não são capazes de capturar informações sobre transtornos mentais em dados textuais. O objetivo de trabalho é propor algoritmos para identificar problemas de transtorno mental por meio dos textos publicados por usuários de mídias sociais. Foram desenvolvidas arquiteturas de redes neurais convolutivas para aprender representações vetoriais de textos, considerando as informações de transtorno mental presente nesses textos, levando a um método chamado *Disorder-Specific Embedding* (DSE). Foram realizados vários experimentos e concluiu-se que as representações vetoriais fornecidas pelo DSE superam os *baselines* considerados. Outro algoritmo proposto neste trabalho chama-se *Hidden Subject Discovery* (HSD), o qual consiste em um método para descobrir comunidades e, consequentemente, assuntos implícitos dentro dessas comunidades, considerando um grupo de usuários com o mesmo transtorno mental. Por meio do HSD, foi possível encontrar padrões ocultos em dados textuais de um determinado transtorno mental, bem como descobrir assuntos e temas implícitos em cada comunidade.

Abstract

Mental disorder problems have been cause for concern around the world. An estimated 54 million Americans suffer from some form of mental disorder in a given year. Nowadays, people discuss and talk about the most diverse topics in social media platforms, including their health. This results in a stream of health-related data. Psychiatrists, doctors, and health insurance companies, are increasingly interested in exploring this kind of data. Recent studies are focused on the characterization of mental health in online social media and development of statistical models from hand-crafted features to properly diagnose the mental health condition. However, such features, in most cases, are not able to capture information about mental disorder in textual data. In this work, we devise algorithms to identify mental disorder problems by examining text posted by users of online social medias. We developed convolutional neural networks architectures to learn high-quality text embeddings by taking into consideration the mental disorder information, leading to a method named Disorder-Specific Embedding (DSE). We performed several experiments and conclude that text embeddings provided by DSE outperform the considered baselines. Another algorithm proposed in this work is called Hidden Subject Discovery (HSD), which is a method to discover communities and hidden subjects considering a given mental disorder within these communities. Through HSD, we found implicit patterns in its textual data as well as hidden subjects and themes in each community.

List of Figures

List of Tables

List of Algorithms

List of Acronyms

ADHD: Attention Deficit Hyperactivity Disorder
AvgVec: Average Vector
BD: Bipolar Disorder
BoW: Bag-of-Words
BPD: Borderline Personality Disorder
CG: Control Group
CNN: Convolutional Neural Networks
DSE: Disorder-Specific Embedding
HSD: Hidden Subject Discovery
MD: Mental Disorder
LIWC: Linguistic Inquiry and Word Count
OCD: Obsessive-Compulsive Disorder
ParVec: Paragraph Vector
PTSD: Post-Traumatic Stress Disorder
t-**SNE**: *t*-Distributed Stochastic Neighbor Embedding

Contents

Chapter 1

Introduction

Mental disorder is a term that may refer to a wide range of mental health conditions. These mental disorders can affect mood, reasoning, and people behavior. Specifically, there are more than 200 classified forms of problems related to mental disorder, including anxiety and depression disorders, bipolar disorder, neurodevelopmental disorders (e.g., attention deficit hyperactivity disorder), personality disorders (e.g., borderline personality disorder, obsessive-compulsive disorder), post-traumatic stress disorder, and psychotic disorders (e.g., schizophrenia).

Mental disorders are common and widespread. In 2013, mental disorders reach about 23 million Brazilians [Empresa Brasil de Comunicação, 2016]. In 2014, about 18% of American adults developed some mental disorder. Considering adolescents, about 20% of them may experience a mental disorder problem in any given year, while for the children the estimate is 13%[1]. An estimated 54 million Americans suffer from some form of mental disorder in a given year [Vos et al., 2015].

1.1 Challenges and Hypothesis

Physical and psychological exams are usually necessary to determine an accurate diagnosis of mental disorder. In these exams, the patients report about their own experiences. However, some of them may feel uncomfortable to report some facts, omitting information and influencing the quality of diagnosis. Moreover, it can difficult to the medical community to track the progress of their patients and to diagnose certain types of mental disorder. There is also a lack of investment and resources in some countries.

[1] `http://www.nimh.nih.gov/health/statistics/index.shtml`

A recent alternative to traditional exams is to observe how psychiatric patients behave in online social media platforms, such as Facebook, Twitter, and Reddit[2]. A particular hypothesis is that it would be possible to identify a specific mental disorder by examining patterns in texts posted by psychiatric patients in such online social media platforms. In fact, psychiatrists, doctors, and health insurance companies, are increasingly interested in exploring this kind of data.

1.2 Objectives of This Work

The objective of this work is to devise new algorithms to identify mental disorder problems associated with users of online social media platforms by examining text posted by them.

We developed Convolutional Neural Networks (CNN) architectures to learn high-quality text embeddings by taking into consideration the mental disorder information (i.e., labeled training data). This leads to an algorithm named as Disorder-Specific Embeddings (DSE).

Another algorithm proposed in this work is called Hidden Subject Discovery (HSD), and it is devised to discover hidden subjects and themes inside a given mental disorder community.

1.3 Contributions

In practice, we claim the following benefits and contributions:

- An effective text embedding algorithm which identifies Reddit users that are associated with specific mental disorders. The algorithm operates by examining text posted by users on pre-defined communities addressed to specific topics (aka. subreddits).

- We performed a systematic set of experiments in order to evaluate our text embedding, outperforming the considered baselines.

- Using this text embedding, we designed an algorithm to uncover implicit patterns in textual data related to mental disorders. The algorithm uses semantic word embeddings and clustering methods to discover hidden subjects in subreddits.

[2] `https://www.reddit.com/`

- We collected and analyzed data from 8 subreddits related to specific mental disorder problems, and also from 12 subreddits which are related to different subjects, in order to represent a control group (i.e., subreddits with subjects not related to any mental disorder problem).

1.4 Text Organization

The remainder of this thesis is organized as follows. Chapter 2 presents the background and relevant related work. In Chapter 3 we describe the datasets used in the experiments, its characterization, as well as the data collection process. In Chapter 4 we present our approach, named DSE and its experimental results. In Chapter 5 we present out method to discover communities and hidden subjects within these communities, and we discuss and analyse empirical and subjectively the results of it. And finally, Chapter 6 shows conclusions of this work.

Chapter 2

Background and Related Work

In this chapter we describe the main mental disorder problems and present existing approaches to classify and analyse such disorder problems. In Section 2.1, we provide a brief summary of each considered mental disorder. In Section 2.2, we introduce convolutional neural networks (CNN). In Section 2.3, text embeddings and baselines algorithms are introduced. In Section 2.4, we present relevant recent work related to the identification of mental disorder problems, online social network data, and feature learning.

2.1 Mental Disorder

Mental Disorders (MDs) are changes in the mind and brain function, as well as changes in behavioral patterns which is negatively affect people's lives and how they live, leading to suffering. There are different types of mental disorders and they are commonly characterized by a combination of abnormal thoughts, perceptions, emotions, behavior and relationships with others[1].

In our work, the main mental disorder problems are analyzed and discussed: Attention Deficit Hyperactivity Disorder, Anxiety, Bipolar Disorder, Borderline Personality Disorder, Depression, Obsessive-Compulsive Disorder, Post-Traumatic Stress Disorder, and Schizophrenia. According to the National Institute of Mental Health[2], a brief description of the mental disorder problems, considering this work, are presented as follows:

- Attention Deficit Hyperactivity Disorder (ADHD): defined as a neurobiological

[1] http://www.who.int/mediacentre/factsheets/fs396/en/
[2] http://www.nimh.nih.gov/health/topics/index.shtml

disorder that appears in childhood and, in most cases, persist through adolescence and adulthood. ADHD is characterized by symptoms such as inattention or lack of focus, hyperactivity, impulsive behaviors, and difficulty controlling behavior. Although being more frequently observed in children, teenagers and adults can also present this disorder.

- Anxiety: it disorder involves more than temporary feelings of anxiety, worry or fear. Such feelings do not disappear and can adversely affect a person in many tasks of her daily lives like social interactions, relationships, job performance, and school projects and works.

- Bipolar Disorder (BD): it is associated with depression and mood swings, and can cause loss of friendship, problems in relationships, weakening of social linkages, and even suicide.

- Borderline Personality Disorder (BPD): it is characterized by depression, unstable moods, emotional instability, impulsive behavior, and uncontrollable anger.

- Depression: it is the most common and serious mental disorder problem, and can cause symptoms that influence how we feel, think and deal with our daily activities such as sleeping, eating or working.

- Obsessive-Compulsive Disorder (OCD): it is a common, never-ending and resilient kind of personality disorder where a person has compulsive behavior and recurring obsessive thoughts uncontrollably.

- Post-Traumatic Stress Disorder (PTSD): it is a serious mental disorder problem that can develop in people who have experienced a traumatic event such as an accident, cruel assault, natural disaster, or other traumatic events.

- Schizophrenia: it is a serious and never-ending mental disorder problem that affects thoughts, feelings and behavior of people. It can be characterized by atypical social behavior and difficulty in distinguishing the real from the imaginary world. People with this condition frequently have depression and anxiety as secondary disorder.

2.2 Convolutional Neural Networks

Convolutional neural networks (CNN), as presented by LeCun and Bengio [1995], have been used in a wide range of applications like: image and video recognition

[Karpathy et al., 2014; Ji et al., 2013; Krizhevsky et al., 2012; Ciresan et al., 2012; Lawrence et al., 1997], natural language processing [Zhang et al., 2015; Poria et al., 2015; Kim, 2014; Dos Santos and Gatti, 2014], and recommender systems [Wang et al., 2015; Van den Oord et al., 2013].

CNN are a particular type of neural networks for processing data that has a grid-like topology. Standard neural networks, like Multilayer Perceptron (MLP), apply matrix multiplication in order to characterize the connectivities between input and output layers. This means that every neuron in a output layer interacts with every one in a input layer. On the other hand, CNN normally have sparse interaction between layers, which reduces the memory requirements to build a model and improves its statistical efficiency [Goodfellow et al., 2016].

In this work, to build the CNN architecture, we use three types of layers: 1D Convolutional Layer, 1D Max Pooling Layer, and, a Fully-Connected Layer. We chose 1D Convolution due to the fact that textual data present a temporal sequence and they are continuous over time.

Explaining simply, 1D Convolution works as follows. Given a one-dimensional text $T \in \mathbb{R}^n$ (sequence of words) as our input and a one-dimensional kernel $K \in \mathbb{R}^m$, the convolution $T * K$ of T and K is defined as:

$$h(y) = (T * K)(y) = \sum_{x=1}^{m} K(x) \cdot T(y - x + 1). \tag{2.1}$$

The outputs of $h_j(y)$ are obtained by a sum over i of the convolutions between $K_i(x)$ and $T_{ij}(x)$. After that, we apply the max pooling function. This function takes few units, depending on the pool length (for example, pool length equals to 2 will halve the convolutional output), from the convolutional layer and chooses the unit that provides the greatest value. After all these operations, the high-level learning in the neural network is done through the fully connected layers, likewise a MLP.

2.3 Text Embeddings

Text (word, sentence, or document) embedding is a function that maps a given text in some language to a high-dimensional vector representation[3]. These vector representations can be used in many tasks such as natural language processing (e.g., document classification, sentiment analysis) and information retrieval (e.g., learning to

[3] http://colah.github.io/posts/2014-07-NLP-RNNs-Representations/

rank, query analysis). In this section we describe the main text embeddings known in the literature.

2.3.1 Linguistic Inquiry and Word Count

Linguistic Inquiry and Word Count (LIWC) is a lexicon dictionary composed by 64 psychologically meaningful categories and 4,484 words. Some experimental results using LIWC have shown its capacity to detect meaning in textual data [Tausczik and Pennebaker, 2010].

To build embedding from LIWC, that is, to extract its features, the following steps are needed. Firstly, given a sentence (or document) S, we have to count the number of occurrences for each LIWC category c in S, or S_c. After that, we have to normalize, dividing the count of each category by the highest value, according to Equation (2.2)

$$\hat{S}_c = S_c \cdot (\max_c S_c)^{-1}, \tag{2.2}$$

where $\hat{S}_c$ is the normalized value to the range [0,1].

2.3.2 Bag-of-Words

The most popular text embedding is the Bag-of-Words (BoW) approach with TF-IDF (term frequency − inverse document frequency) weighting scheme [Baeza-Yates and Ribeiro-Neto, 2011]. This embedding represents each document using frequency count of words in a basic vocabulary times the inverse of the word frequence in the collection.

Formally, let $V = \{w_1, w_2, w_3, \ldots, w_n\}$ be a vocabulary, that is, the set of n words that can occur in a document. Let $tf(w_i, d)$ be the term frequency count of word w_i in document d. Let $idf(w_i)$ be inverse document frequency, that is, the logarithmically scaled inverse fraction of the documents that contain the word w_i. Let $\phi(w_i, d) = tf(w_1, d) \times idf(w_1)$ be the weight of the word w_i in document d. Then, each document d will be represented by the embedding $d = \{\phi(w_1, d), \phi(w_2, d), \phi(w_3, d), \ldots, \phi(w_n, d)\}$ [Pang et al., 2002].

2.3.3 Average & Paragraph Vector

Despite their popularity, the BoW approach have some weaknesses: they need many features to perform well, they lose the ordering of the words in a document,

and they also ignore semantics. Mikolov et al. [2013] proposed a technique, named *Word2Vec*[4], that can be used for learning high-quality word embeddings from huge textual data which preserves the semantic and syntactic meaning of the words. Since the word embeddings are learnt, an naive approach, named here as Average Vector (AvgVec), is used to represent a text by averaging the embeddings of the words that appear in that text.

In other words, let $d = \{w_1, w_2, w_3, \ldots, w_k\}$ be a document composed by k words and w_i a word embedding vector. Then, the text embedding can be obtained as:

$$d_i = \frac{1}{k} \sum_{i=1}^{k} w_i. \tag{2.3}$$

A further alternative approach, called Paragraph Vector (ParVec), was proposed by Le and Mikolov [2014]. The process of learning paragraph vectors is inspired by the process of learning word embeddings. Specifically, this approach learns fixed-length feature embeddings from textual data and represents each document by a vector which is trained to predict words in the document.

2.4 Related Work

Nowadays, more and more people use social media, like Facebook, Reddit, and Twitter, to share ideas, opinions and thoughts. In this, people tend to share how they are feeling (happiness and sadness, for example) and even express about their own mental health. Thence, many studies related to mental health using social media have been discussed in recent years.

Characterization of social media is very important because it contains useful content and implicit information about mental health of users. For instance, De Choudhury et al. [2013a,b] developed a statistical model, using a SVM classifier, that is able to predict whether a given text, from Twitter posts, has depressive content or not. To build this model, the input was heavily based on hand-crafted (engineered) features extracted from the textual data such as time, linguistic style, and emotion. For this, they adopted a manual method, using responses from crowd-workers on Amazon's Mechanical Turk (AMT) to derive subjects and themes discussed by the users.

In contrast to the aforementioned works, Coppersmith et al. [2014a] proposed a fast and low-cost method for gathering data about mental disorders from Twitter, instead of using AMT, not requiring manual intervention. Through their method, they

[4] `https://code.google.com/archive/p/word2vec/`

employed statistical models to distinguish users with some mental health disorders (e.g., depression, BD, PTSD) − from a control group. In other similar studies, they focused their analysis in PTSD [Coppersmith et al., 2014b] and Schizophrenia [Mitchell et al., 2015].

Rather than data from Twitter and Facebook, De Choudhury and De [2014] characterized mental health discourse on the Reddit social media platform. They built a model in order to discover factors that influence in communities on Reddit related to mental health. They used as independent variables features derived from LIWC and as response variables the difference between the number of up-votes and down-votes and the number of comments in a given post. Pavalanathan and De Choudhury [2015] explored specific discourse on mental health communities written by anonymous accounts on Reddit. They observed that mental health discourse from anonymous accounts is more negative and the posts content indicate low self-esteem. Balani and De Choudhury [2015] built a model in order to detect levels of self-disclosure in posts on mental health communities on Reddit, which was able to detect it with reasonable accuracy. Thus, they found that discourse in these communities is characterize by high self-disclosure.

The main aspects of the aforementioned studies are based on the characterization of social media and construction of statistical models from hand-crafted features. However, to extract such features is required intense effort and, in most cases, a domain expert where the application will run. Therefore, other dimension of our study is about *feature learning*. Through supervised (deep neural networks) and unsupervised (autoencoders) techniques, it became possible to learn features in an automatic way [Bengio et al., 2013]. Many studies have applied feature learning in NLP (natural language processing) task, especially in sentiment analysis. In our work, we used CNN to learn embedding representation at sentence-, and user-level in order to discriminate mental disorders.

Maas et al. [2011] presented a model that uses supervised and unsupervised techniques to learn word embeddings, taking into account semantic information and sentiment content. Nevertheless, although these embeddings have been very useful, they are not able to express the meaning of long sentences. Therefore, to solve this problem Socher et al. [2013] used the Stanford Sentiment Treebank and a Recursive Neural Tensor Network, achieving excellent results. Tang et al. [2014a,b] developed a set of deep neural networks to combine semantic features with hand-crafted features. Other studies [Dos Santos and Gatti, 2014; Kim, 2014] have used CNN to solve sentiment analysis problems. Recently, Zhang et al. [2015] have explored the use of CNN at character-level for text classification and they have achieved state-of-the art results.

Chapter 3

The Reddit Dataset

In this chapter we present the dataset used in the experiments. In Section 3.1 we describe the way we collected and gathered data from Reddit. Next, in Section 3.2 we describe some details of each subreddit and present a characterization in order to better understand the datasets.

3.1 Data Collection Process

The Reddit[1] platform was released in 2005. Reddit is a social news networking service where the users can submit content, such as text posts. By voting (as a button "*like*"), users themselves decide what has relevant content and what has not. Posts that are supported by the community, that is, receiving many votes and comments are highlighted. Reddit platform is organized by topics of interest, named "*subreddits*". In our work, we mainly consider subreddits related to mental disorders, mentioned in the previous section.

Reddit provides an API (Application Programming Interface)[2] that is well-defined, well-documented, and provides many features and methods to collect and gather data. We used a wrapper, named PRAW (Python Reddit Api Wrapper)[3] that allows easy access to Reddit's API.

To efficiently gather data, a total of 9 machines with different IP addresses were used. The data collection process started on January 16th of 2016 and finished on January 30th of 2016. During this period, we collected data from 8 subreddits related to Mental Disorder. We also collected data from 12 subreddits that talks about different

[1] `https://www.reddit.com/`
[2] `https://www.reddit.com/dev/api`
[3] `https://praw.readthedocs.org/en/stable/`

subjects, in order to represent a Control Group (CG)[4], that is, subreddits with subjects not related to MDs. The collected datasets are shown in Table 3.1. They are are dated from January 1st of 2014 until January 1th of 2016.

Table 3.1: Volume of collected datasets.

Dataset	**#users**	**#posts**	**#comments**
ADHD	23,433	17,692	221,761
Anxiety	36,513	26,707	162,696
BD	10,654	13,341	133,255
BPD	6,097	6,569	53,151
Depression	97,756	84,407	387,775
OCD	4,388	3,323	21,445
PTSD	3,059	2,366	16,977
Schizophrenia	2,763	2,886	28,956
CG	301,929	99,562	1,332,126

3.2 Data Characterization

We were able to find out a lot of information about each MDs dataset. In our collected datasets, we used a regular expression[5] to extract the age of some users from the text dataset. In Figure 3.1 we can see the average age of users for each MD, as well as the confidence interval. The users age generally varies between 18 and 30 years.

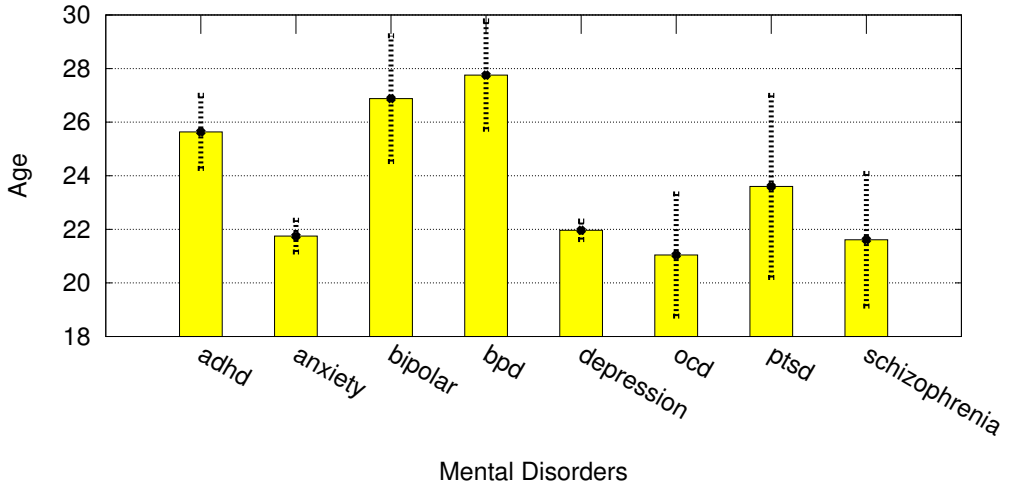


Figure 3.1: Average age of users for each disorder.

[4]**Control Group** consists of the following subreddits: *askreddit, books, fitness, food, funny, gaming, movies, music, religion, space, video, world news.*

[5]For example, *"I am (\d+) years old"*

Considering the number of text messages in the MDs datasets, the whole dataset contains 157,291 posts and 1,026,016 comments. Of these, 140,430 users wrote at least one post, while 72,551 users wrote at least one comment. On the other hand, the CG datasets contains 99,562 posts and 1,332,126 messages. Of these, 248,936 users wrote at least one post, while 150,734 users wrote at least one comment. Table 3.2 presents some descriptive statistics about the datasets. As we can see, posts and comments on the MD datasets are richer (i.e. they have more words) than the CG datasets, showing that users belonging to groups related to MD tend to write posts or comments with more details.

Table 3.2: Descriptive statistics of MD and CG datasets.

	MD	**CG**
Valid users	184,663	301,929
Posts	157,291	99,562
Comments	1,026,016	1,332,126
Posts per user*	1.74 (2.97)	1.39 (4.09)
Words per post*	259.08 (279.99)	73.04 (110.64)
Likes per post*	7.87 (26.68)	16.35 (192.99)
Comments per user*	7.56 (226.19)	5.02 (119.97)
Words per comment*	98.95 (105.45)	66.22 (74.58)
Likes per comment*	2.28 (4.72)	27.32 (221.96)

* Mean and standard deviation.

Also, Figure 3.2 depicts the empirical cumulative distribution (ECDF) of the number of posts and comments in each datasets group. We observe that approximately 99% of the users, in both datasets, wrote 10 or less posts. Moreover, almost 88% of the users wrote 10 or less comments on the MD datasets, while approximately 92% of the users wrote 10 or less comments on the CG datasets.

Figure 3.3 show word clouds[6] (unigrams, bigrams, and trigrams) that were built from posts of users associated with different MDs. From these, we can get some insight about the datasets. Furthermore, we can also understand clear differences between the vocabulary[7] used by each MD. We can observe that most of these n-grams are related to drugs and advices, indicating that users like to ask or help each other about their treatments, as well as feelings and mood about themselves. Feelings like *upset* and *guilt* are very common. With the intention of understand what people usually share in their posts, Table 3.3 shows a few fragments of posts.

[6] `https://www.jasondavies.com/wordcloud/`

[7] n-grams with *IDF* (Inverse document frequency) value less than the first *percentile* were removed (`https://en.wikipedia.org/wiki/Percentile`).

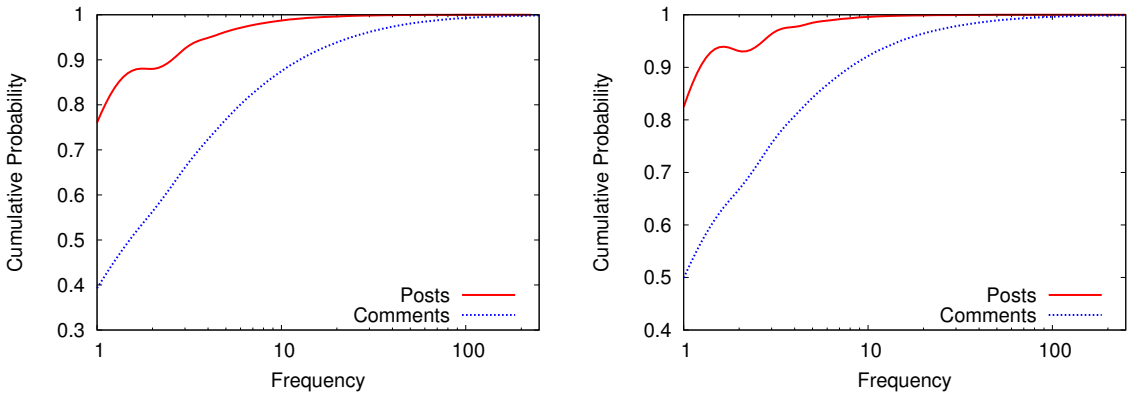


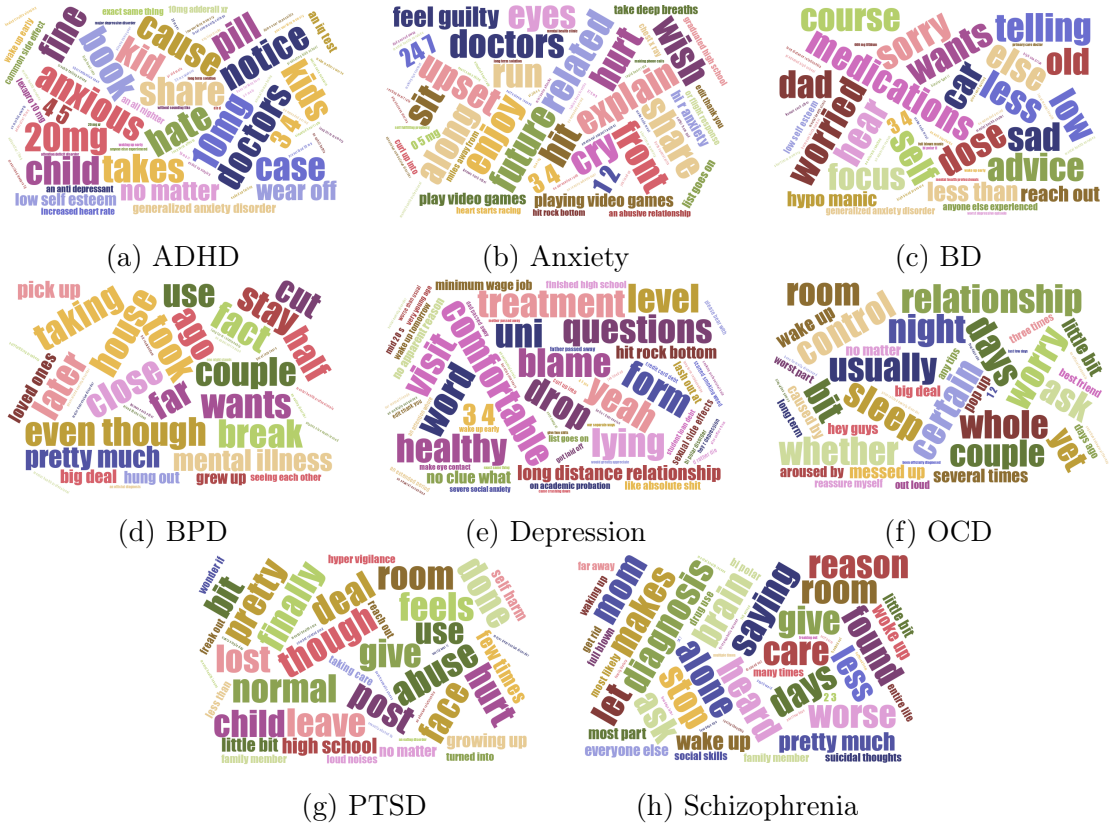
Figure 3.2: Empirical Cumulative Distribution (ECDF) of number of posts/comments by users. **Left** − MD Datasets; **Right** − CG datasets.

Table 3.3: Text fragment of a few posts.

ADHD	*"After years of denial I have finally relented and realized that I have adhd"*
Anxiety	*"It feels like I'm on the verge of a panic attack"*
BD	*"I am having a hard time and would like someone to talk to"*
BPD	*"I'm tired to be faking happiness at home to my young kid and husband"*
Depression	*"I have depression and anxiety and the remnants of an eating disorder"*
OCD	*"I have contamination ocd and recently it has gotten way out of control"*
PTSD	*"I have ptsd from an abusive relationship with my ex-husband"*
Schizophrenia	*"I'm aware that the voices I hear stem from schizophrenia"*

In order to understand the difference between the textual content from MDs and CG datasets, we used some categories provided by LIWC. We considered 23 categories from LIWC that are more related to MDs. For each post, we compute the number of occurrences for each category. After, we calculate the average of each category and transform it in probability. Figure 3.4 shows how different categories may be distributed over the datasets. As we can see, words related to *work*[8], *positive emotion*[9], and *leisure*[10] are more common in the CG than in the MDs datasets. On the other hand, words related to *negative emotion*[11], *health*[12], *feel*[13], and *sadness*[14] are more common in the MDs datasets.

[8]**work:** projects, boss, staff.
[9]**posemo:** favor, neat, fantastic.
[10]**leisure:** playful, bands, party.
[11]**negemo:** assault, hurt, heartbreak.
[12]**health:** headache, bipolar, flu.
[13]**feel:** tight, touch, smooth.
[14]**sad:** defeat, sadly, unsuccessful.



(a) ADHD (b) Anxiety (c) BD

(d) BPD (e) Depression (f) OCD

(g) PTSD (h) Schizophrenia

Figure 3.3: Word clouds generated from unigrams, bigrams, and trigrams of each personality disorder.

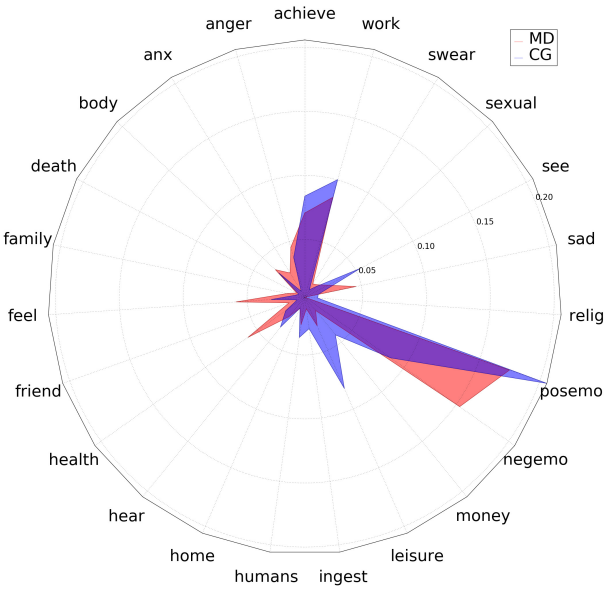


Figure 3.4: Distribution of 23 categories, provided by LIWC, obtained from MDs (in red) and CG (in blue) datasets.

Chapter 4

The Disorder-Specific Embeddings Algorithm and Its Evaluation

In this chapter, we present the DSE algorithm and show how to learn embeddings taking into account the mental disorder information and how to evaluate empirically our text embedding approach against the baselines. In Section 4.1, we introduce our proposed approach, named Disorder-Specific Embedding (DSE). In Section 4.2, we present the settings and parameters used in the CNN and baselines, and describe the procedures for evaluating the quality of text embeddings in classification tasks. Finally, in Section 4.3, we show the performance of the considered text embeddings approaches on binary classification and multiclass classifications tasks.

4.1 DSE Approach

The text embeddings methods described in the Section 2.3 are not able to capture information about mental disorder in textual data. Considering our DSE approach, we introduce the disorder information from textual data in the learning phase in order to obtain embedding (or continuous) representations, for sentences and users, and consequently being able to discriminate mental disorder in textual data. To do this, we developed two CNN architectures to learn DSE, described in the following.

4.1.1 Capturing Mental Disorder Information at Sentence-level

Considering disorder-specific sentence-level embedding, we employed the architecture depicted in Figure 4.1. As we can see, for each sentence we have a ordered set of words. Each word is mapped to a low-dimensional embedding (vectors initialized

by using a random uniform distribution). Thus, we have a set of word embeddings as input for our model.

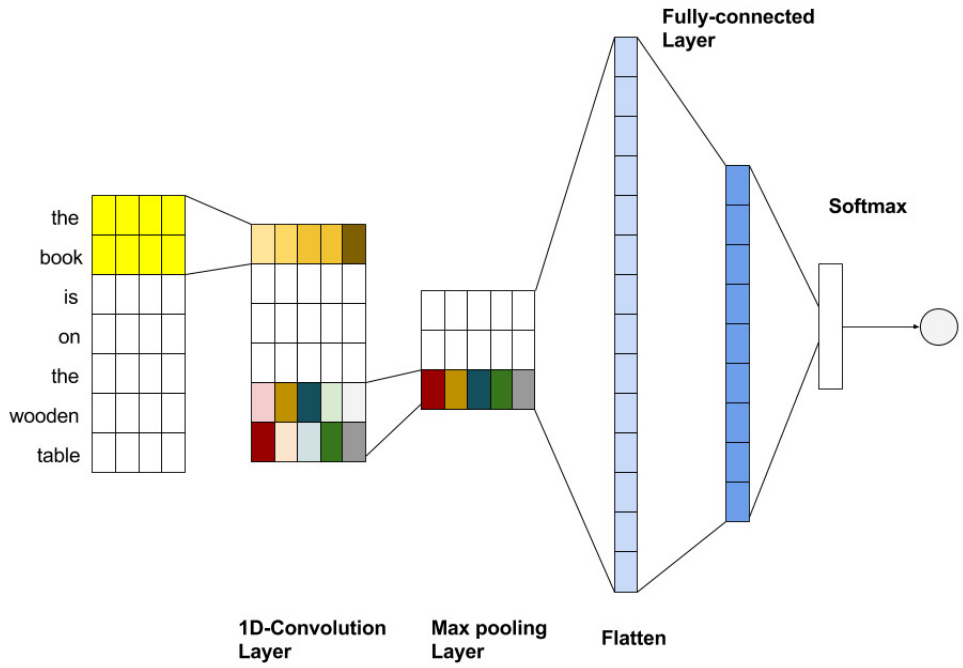


Figure 4.1: CNN architecture at sentence-level. The dark blue squares indicate the sentence embedding.

To capture mental disorder information, this set of word embeddings should be able to predict the mental disorder label through a model (classifier) yielded by this CNN architecture. This model is learnt during the training phase, where the prediction error is computed, by using cross-entropy loss, and weights are updated to minimize this error. At the end, we have a model that, given a set of words from a sentence, can predict the mental disorder label, as follows,

$$\hat{m} = C(s_i|\theta). \tag{4.1}$$

Let $C(\cdot)$ be a classifier yielded by the architecture mentioned above. Let $s_j = \{w_1, w_2, w_3, \ldots, w_k\}$ be a set of word embeddings (a padded sequence with k words, due to the fact that the size of the sentences may vary) and $w_i \in \mathbb{R}^n$ be a n-dimensional word embedding. Let θ be the weights of the classifier $C(\cdot)$. Using this model we can map a set of word embeddings s_j to a predicted mental disorder label $\hat{m}$.

However, what we want is a low-dimensional sentence-level embedding. If we consider the set of word embeddings, by concatenating them we would have a sentence embedding of dimension $k \times n$, where k is the number of words and n is the word

embedding dimension. As the convolution and fully-connected layers extract relevant features from each set of words and these features are compressed across the layers, we pick up the hidden layer immediately after the flatten operation as our low-dimensional sentence-level embedding, or disorder-specific sentence-level embedding, represented as dark blue in Figure 4.1. In short, this sentence embedding takes into account the mental disorder information.

4.1.2 Capturing Mental Disorder Information at User-level

In a similar fashion, we can get disorder-specific user-level embeddings by using the architecture illustrated in Figure 4.2. As we can see, for each user we have a set of documents. Each document has a set of sentences. And, as explained above, each sentence has a ordered set of words. Thus, we have a set of documents (or a family of independent sets of word embeddings) from a given user as input for our model.

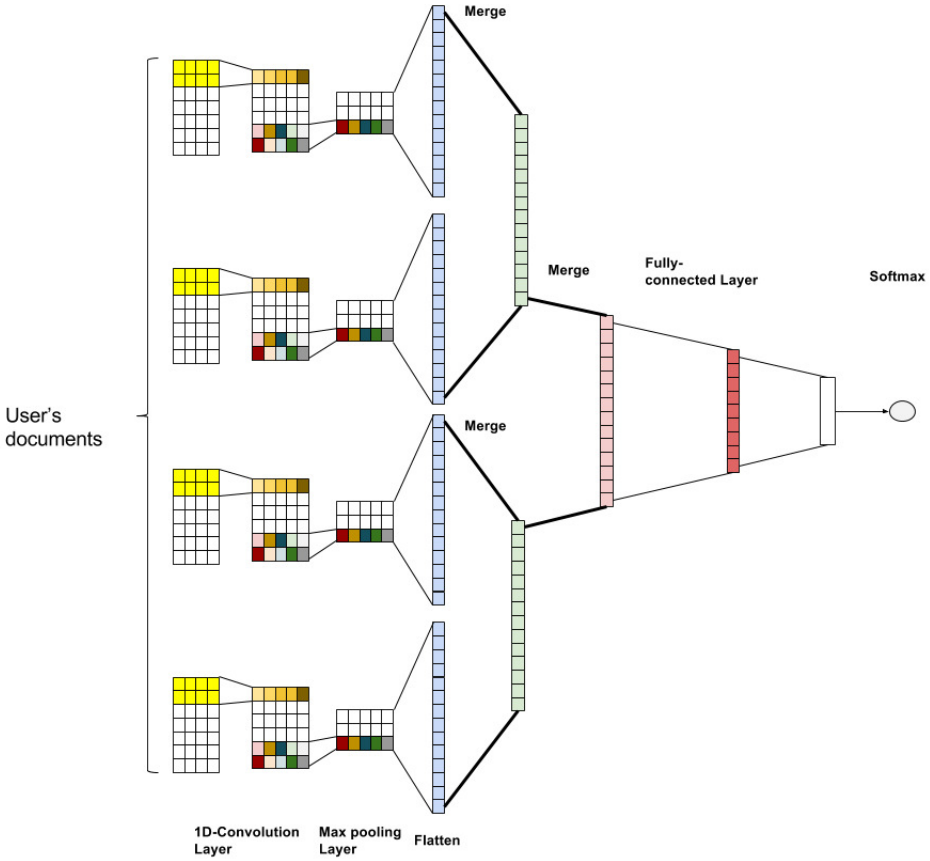


Figure 4.2: CNN architecture at user-level. The dark red squares indicate the user embedding.

In this case, to capture mental disorder information, this family of independent

word embeddings should be able to predict the mental disorder label through a model yielded by this CNN architecture. The learning phase is similar to the previously described, but the input format and architecture are different. Here, each sentence or word embeddings has its own convolutional layer. After each convolutional layers, sentences of the same document are combined via a *average merging operation*[1] producing a document embedding. On a deeper level, these document embeddings of the same user are also combined via *average merging operation*, producing a user embedding.

At the end, we have a model that, given a set of documents from a given user, can predict the mental disorder label, as follows,

$$\hat{m} = E(u_p|\theta) = E(\{d_1, d_2, d_3, \ldots, d_l\}|\theta), \tag{4.2}$$

where $E(\cdot)$ is a classifier yielded by the architecture mentioned above, θ is the weights of the classifier $E(\cdot)$, $u_p = \{d_1, d_2, d_3, \ldots, d_l\}$ is the set of documents of a given user p (padded with l documents), and $d_q = \{s_1, s_2, s_3, \ldots, s_m\}$ be the set of sentences of the document q (padded with m sentences).

As we want a low-dimensional user-level embedding, we consider the hidden layer immediately after the last merged layer as our disorder-specific user-level embedding, represented as dark red in Figure 4.2. In short, this user embedding takes into account the mental disorder information.

4.2 Experimental Setup

In this section, we present the experimental setup that we used in the CNN and baseline algorithms. All experiments described in this chapter were run on an Intel Xeon CPU E5620 2.40GHz with 36GB main memory, equipped with a Titan Black GPU accelerator with 6GB memory and ultra-fast 336 GB/s throughput.[2]

4.2.1 CNN – Parameter Settings

The construction of a CNN require many hyperparameters to choose from. In this work, the parameters were defined empirically. We designed a CNN with:

- Sentence-level: one embedding layer, two 1D-convolution layer, and two fully-connected layers.

[1] **Average Merging Operation** combines a list of embeddings into a single embedding by their average.

[2] We gratefully acknowledge the support of NVIDIA Corporation with the donation of the GPU accelerator for this research.

- User-level: one embedding layer, two 1D-convolution layer, merge layers (for combine sentences, documents, and user representations), and two fully-connected layers.

Taking into account the embedding layer, the input receives a padded sequence with 50 words. Each word is represented as a low-dimensional embedding vector of size 100. The size of the vocabulary was defined in 80,000 words. At document-level, we limited the number of sentences per document in 25. At user-level, we limited the number of documents per user in 10.

In the first 1D-convolution layer, we used a filter length of size 5 with 512 convolution kernels. While, in the second 1D-convolution layer, we used a filter length of size 3 with 256 convolution kernels.

In all layers we applied the activation function ReLU (rectified linear unit). We used the standard max pooling with pool length equals to 2, halving the output of the previous 1D-convolution layers. We also insert dropout [Srivastava et al., 2014] modules to regularize: one between the embedding and first convolution layer and the other between the fully-connected layers. They have dropout probability of 0.5. At the end of the CNN there is a softmax layer, where each neuron is interpreted as a probability value, allowing binary or multiclass classification.

The algorithm used to perform weight update was an adaptive learning rate method called Adam [Kingma and Ba, 2015] with a batch size of 300 and learning rate 0.001. During the training, 20% of the training set was used as validation set. If no improvement occurred (increased accuracy and error reduction) after 5 epochs, the training was stopped. The code implementation was done using Keras [Chollet, 2015] and TensorFlow [Abadi et al., 2015].

4.2.2 Baselines – Parameter Settings

We got the LIWC embeddings following the steps shown in Section 2.3.1. We obtained AvgVec[3] and ParVec[4] embeddings using the default parameters provided by the tools. The embeddings were built using a skip-gram architecture and hierarchical softmax as training algorithm. We developed our own BoW approach using *scikit-learn* [Pedregosa et al., 2011]. Here, we considered all words of the vocabulary to build this embedding.

[3]https://code.google.com/p/word2vec
[4]`https://github.com/mesnilgr/iclr15`

4.2.3 Evaluation Procedure

To evaluate the quality of our embeddings (obtained from DSE) against the baselines (described in Section 2.3), we used them as input to a SVM (Support Vector Machine) classifier with kernel linear[5].

We split our evaluation process in two classification tasks: binary and multiclass classification.

- Binary Classification: in this task we have two groups and we want to determine if a given sentence (sentence-level embedding) refers to a some mental disorder (MD) or a control group (CG). To evaluate the prediction performance of this task, we employed the following evaluation metrics: accuracy (ACC), F-measure (F1), AUC (Area Under the Curve) and ROC (Receiver operating characteristic) curve.

- Multiclass Classification: in this other task, given a user (user-level embedding), we want to find out what type of mental disorder that user fits in. We evaluate the prediction performance of this task using the micro-F_1 and macro-F_1 metrics.

We conducted ten-fold cross-validation (CV). Thus, the labeled dataset was splited into ten folds, including training and test. At each run, nine folds are used as training set, and the remaining fold as test set. The results reported are the average (mean and standard deviation) of the ten runs. For the AUC and ROC evaluation metrics, we employed hold-out validation, splitting the labeled dataset into two equal parts: training (50%) and test (50%) set.

4.3 Classification Performance

Our first experiment is concerned with classifying different types of mental disorders against a control group (CG), in order to compare the accuracy and F1 results of the text embeddings techniques described in Section 2.3.

In Tables 4.1, 4.2 and 4.3 we can see the performance of each embedding technique on sentence-level. The experimental results show that the DSE embedding, provided from a CNN, outperform the baseline embeddings in all scenarios. That is, our approach provides better accuracy and F1 results in all datasets.

The LIWC embedding was the worst performer in all classification experiments. In Tables 4.2 and 4.3, taking into a count the Depression and Schizophrenia scenarios,

[5] `https://www.csie.ntu.edu.tw/~cjlin/liblinear/`

LIWC achieves high accuracy (on average 91.55% and 80.30%, respectively), however very low F1 values (on average 50.48% and 59.29%,respectively). Due to the fact that this scenarios presents very unbalanced data compared to other scenarios, the LIWC was biased in favor of the majority class. Hence the very low F1 values.

BoW and ParVec achieved similar performance numbers, while AvgVec was a little worse than them. BoW approach provided better accuracy and F1 results than ParVec in five out the eight scenarios − ADHD, Depression, OCD, PTSD and Schizophrenia. While ParVec approach achieved better performance than BoW in two out the eight scenarios − Anxiety and BPD. In the Bipolar scenario, these two approaches tied.

Table 4.1: Accuracy and F1 results on the sentence-level binary classification: {ADHD, Anxiety, Bipolar} *vs.* CG.

Embedding	**ADHD**		**Anxiety**		**Bipolar**	
	Accuracy (%)	**F1 (%)**	**Accuracy (%)**	**F1 (%)**	**Accuracy (%)**	**F1 (%)**
LIWC	74.96 (±0.288)	73.97 (±0.311)	80.70 (±0.224)	69.96 (±0.341)	75.59 (±0.299)	75.10 (±0.304)
BoW	83.50 (±0.249)	83.43 (±0.243)	85.61 (±0.137)	79.76 (±0.194)	81.71 (±0.341)	81.46 (±0.341)
AvgVec	82.53 (±0.235)	82.34 (±0.236)	85.12 (±0.151)	78.62 (±0.204)	80.92 (±0.273)	80.64 (±0.276)
ParVec	82.83 (±0.161)	82.69 (±0.158)	85.94 (±0.197)	80.20 (±0.299)	81.73 (±0.343)	81.46 (±0.344)
DSE	85.56 (±0.406)	85.52 (±0.401)	88.21 (±0.185)	83.68 (±0.239)	84.65 (±0.282)	84.44 (±0.292)

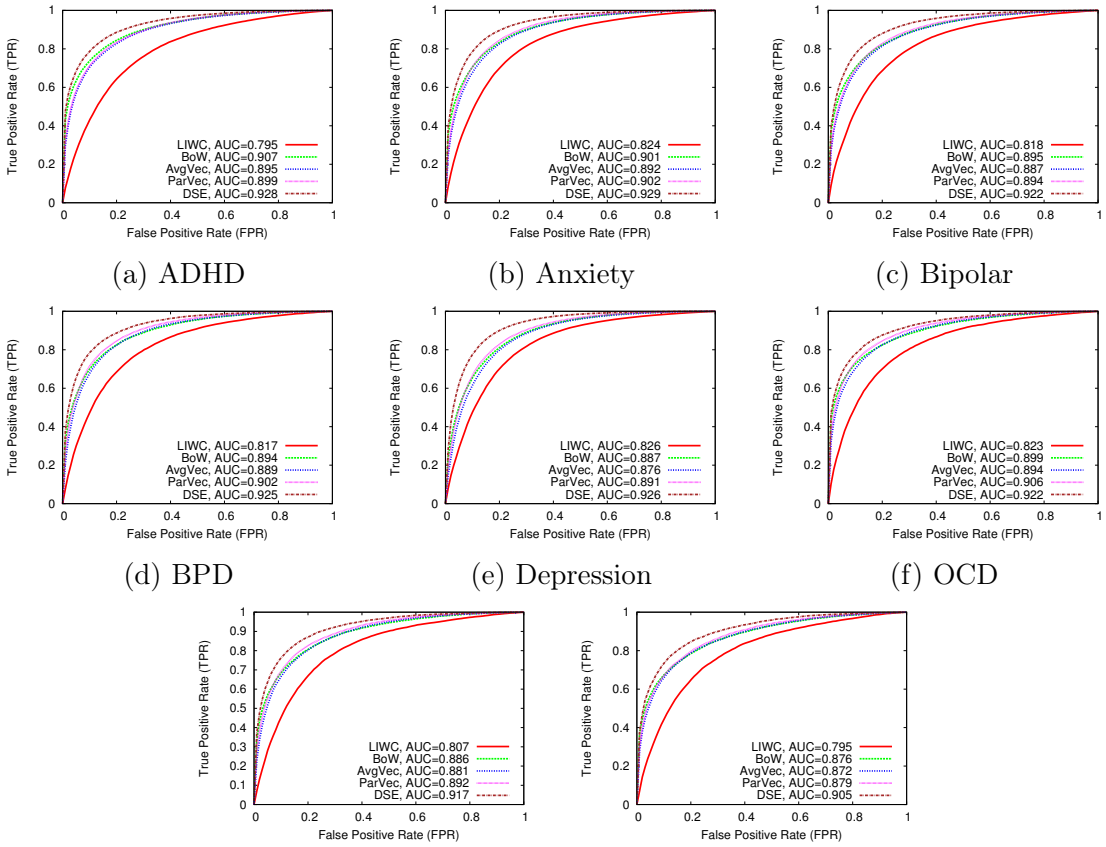
Table 4.2: Accuracy and F1 results on the sentence-level binary classification: {BPD, Depression, OCD} *vs.* CG.

Embedding	**BPD**		**Depression**		**OCD**	
	Accuracy (%)	**F1 (%)**	**Accuracy (%)**	**F1 (%)**	**Accuracy (%)**	**F1 (%)**
LIWC	75.05 (±0.258)	74.18 (±0.282)	91.55 (±0.023)	50.48 (±0.196)	78.93 (±0.247)	67.76 (±0.501)
BoW	82.23 (±0.200)	81.77 (±0.210)	93.21 (±0.078)	70.30 (±0.377)	86.74 (±0.301)	81.78 (±0.451)
AvgVec	81.38 (±0.205)	80.91 (±0.214)	92.79 (±0.054)	65.65 (±0.364)	85.22 (±0.184)	79.45 (±0.274)
ParVec	82.69 (±0.289)	82.26 (±0.299)	93.09 (±0.050)	68.07 (±0.356)	86.19 (±0.399)	81.17 (±0.542)
DSE	85.23 (±0.455)	84.86 (±0.458)	94.39 (±0.068)	77.35 (±0.296)	87.68 (±0.386)	82.93 (±0.535)

Table 4.3: Accuracy and F1 results on the sentence-level binary classification: {PTSD, Schizophrenia} *vs.* CG.

Embedding	**PTSD**		**Schizophrenia**	
	Accuracy (%)	**F1 (%)**	**Accuracy (%)**	**F1 (%)**
LIWC	79.08 (±0.236)	62.81 (±0.504)	80.30 (±0.164)	59.29 (±0.581)
BoW	86.30 (±0.293)	79.39 (±0.465)	87.28 (±0.250)	78.77 (±0.434)
AvgVec	84.80 (±0.334)	76.31 (±0.523)	85.87 (±0.339)	75.28 (±0.659)
ParVec	85.76 (±0.232)	78.44 (±0.428)	86.43 (±0.217)	77.03 (±0.389)
DSE	87.43 (±0.243)	81.19 (±0.423)	87.63 (±0.383)	79.21 (±0.622)

In the next experiment, we used ROC (Receiver Operating Characteristic) with the aim of understanding how well a Linear SVM classifier on top of sentences embeddings can discriminate sentences from a mental disorder and sentences from a control group, and to find the best threshold for discriminating them.



(a) ADHD (b) Anxiety (c) Bipolar

(d) BPD (e) Depression (f) OCD

(g) PTSD (h) Schizophrenia

Figure 4.3: ROC curves for sentence-level binary classification and the corresponding AUC values.

In Figures 4.3 we show the ROC curves for each scenario. Again, we can see that the DSE embedding outperforms the baselines. The results report that our approach achieves high level performance over all 8 scenarios. Conversely, using the LIWC embeddings produces the worst results.

The last experiment is concerned with constructing a user-level representation and trying to find out what type of mental disorder each user belongs to. Table 4.4 shows the performance of each embedding technique on the multiclass classification task. The DSE embedding clearly achieves the highest values of Micro-F_1 and Macro-F_1. Micro-F_1 is like accuracy, it does not take label imbalance into account, while Macro-F_1 globally counts the total true positives, false negatives and false positives. As we can observe, for the LIWC embedding, the discrepancy between Micro-F_1 and Macro-F_1 is very high, whereas for others this difference is smoother. This indicates that LIWC embedding do not work very well in multiclass tasks.

Finally, we show a interesting visualization of DSE user-level embeddings, us-

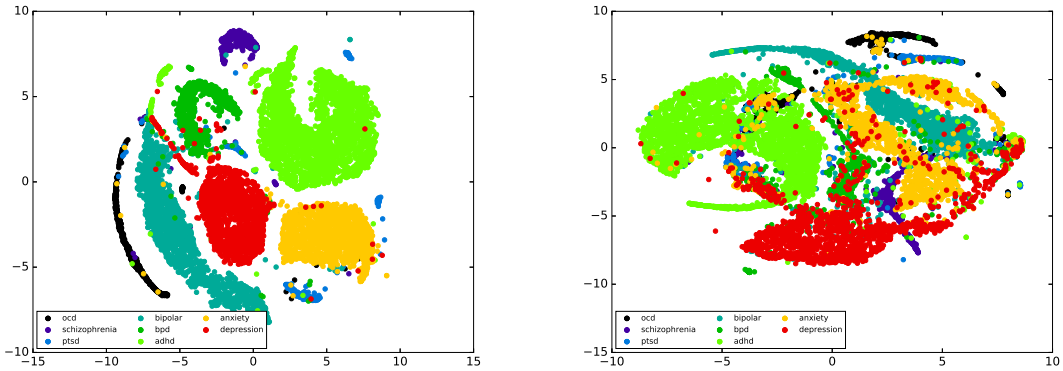


Figure 4.4: t-SNE Visualization of users. **Left** − User-level embeddings from training set; **Right** − User-level embeddings from test set.

Table 4.4: Micro-F_1 and Macro-F_1 results on the user-level multiclass classification.

Embedding	**Micro-F_1**	**Macro-F_1**
LIWC	42.61 (±0.540)	25.25 (±0.795)
BoW	80.30 (±0.756)	78.40 (±1.081)
AvgVec	70.79 (±0.725)	66.70 (±1.031)
ParVec	75.62 (±0.656)	72.09 (±1.009)
DSE	83.19 (±0.780)	81.94 (±1.066)

ing the t-Stochastic Neighbor Embedding (t-SNE). t-SNE is a dimensionality reduction technique that provides a high-quality low-dimensional representation in a two-dimensional space. To perform dimensionality reduction, this technique try to minimize the Kullback-Leibler divergence between the high-dimensional data and the low-dimensional embedding obtained [Van der Maaten and Hinton, 2008]. We used the default parameters from *scikit-learn*.

As we can see in Figure 4.4, the DSE embeddings can discriminate very well these mental disorders classes. Considering the right figure, we can note the generalizability of the CNN in distinguish mental disorders. Moreover, we can note some indications of comorbidity, the presence of one or more mental disorders co-occurring with a primary disorder. For instance, we can observe that depression points appear in all the other disorders clusters, indicating that depression acts as a secondary mental disorder in some individuals. Anxiety is another disorder that occurs as secondary, for example, in individuals with OCD.

Chapter 5

The Hidden Subject Discovery Algorithm and Its Evaluation

In this chapter, we present a method to discover communities and hidden subjects within these communities, taking into account mental disorders. In Section 5.1, we introduce our proposed approach, named Hidden Subject Discovery (HSD). In Section 5.2, we present the algorithm, how it works, and its parameters. Finally, in Section 5.3, we discuss and analyse empirical and subjectively the results of the HSD algorithm.

5.1 HSD Description

In order to identify topics in Reddit communities, De Choudhury and De [2014] used Latent Dirichlet Allocation (LDA), by making use of unigrams and bigrams present in the textual data. However, using LDA becomes difficult to identify what users have discussed about a given subject. Moreover, it becomes more difficult to assign what the main subject discussed in a given community, because LDA does not provide a importance score for its topics (or subjects).

To find out hidden subjects in each mental disorder community, we propose an algorithm the HSD algorithm which is able to detect communities, where each one deals with a specific subject. Furthermore, we found in a *qualitative* way that the subjects provided by HSD are easier to interpret and understand than those provided by LDA, as we can observe in the examples presented in Table 5.1. Looking at this table, we can see that HSD words are more related to the subjects than the LDA words.

Initially, given a mental disorder, we want to find groups of users (communities) within this disorder, taking into account the textual data from each user. To obtain

Table 5.1: HSD *vs.* LDA words

Subject	**HSD Words**	**LDA Words**
Drugs	doses, sedating, atypical, vistaril, mirtazapine, sedative, drowsiness, cymbalta, miracle, buspar, trazadone, dosage, clonazepam, lifesaver, 300mg, withdrawal, ativan, neurontin	prazosin, accident, began, alcohol, woke, chest, sexually, episode, waking, psych, girlfriend, cbt, yesterday, anybody, screaming, peace, hello, eating
Trauma	consent, rapes, sketchy, relations, harassment, orientation, sensed, violence, trafficking, abuse, encounters, stealing, initiating, favors, assaults, accomplishing, preference, encounter, screws, assault	abusers, respect, survivors, dissociate, society, strength, responsibility, abuser, yoga, recover, boundaries, power, addiction, notice, steps, choose, choice, actions, op, victims
Militarism	troops, horses, fellow, served, elderly, volunteered, sufferer, generation, civilians, former, vietnam, veterans, marines, soldiers, australia, active, serving, battlefield	veterans, dawareness, anybody, dwi, study, army, stigma, vet, link, accident, video, advance, regular, stories, mdma, film, weed, dog

these communities, we need some clustering algorithm (e.g. K-Means, Spectral Clustering, Ward Hierarchical Clustering, Agglomerative Clustering). Once communities have been found, we want to identify the main subjects discussed in them. To identify the main subjects, we need semantic clusters. The semantic clusters are obtained by clustering semantic word embeddings. Thus, for each community, we compute the occurrences of each semantic cluster and consider the most frequent. However, it may occur a scenario where a given semantic cluster appears in many communities. In order to handle this, we remove these semantic clusters, as will be explained in the next section.

5.2 HSD Algorithm

The pseudocode of our method is assembled in Algorithm 1. Let A be a clustering algorithm. Let U be user embeddings (e.g. BoW, ParVec, DSE). Let W be a semantic word embeddings (e.g. *word2vec*, *glove*[1]). Let T be the textual data, where $T(u)$ is all text written by user u. Let k be the number of semantic clusters. Let c be the number of communities. Let s be the maximum number of subjects. Let α^{th} be the percentile

[1] *Glove* by Pennington et al. [2014]. `http://nlp.stanford.edu/projects/glove/`

parameter used to filter semantic clusters that occurs in many communities. That is, semantic clusters with *IF* (inverse frequency) value less than the first *percentile* were removed.

Initially, k semantic clusters G are obtained, where $G(i)$ is the set of words in the cluster i and $G'(w)$ is the semantic cluster where the word w belongs to. After that, we obtain c community clusters C: $C(j)$ is the set of users who belong to the community j and $C'(u)$ is the community where the user u belongs to. Then, a mapping function ϕ is employed over C and T in order to obtain M, where $M(j)$ is all text written by users of community j. Next, F is initialized. F represents the frequency of each semantic cluster over the communities. Inside the *for* loop, H represents the frequency of each semantic cluster over the current community j. Moreover, we have a function *top* which returns the s semantic clusters that most occur in community j. After, the inverse frequency (IF) of each semantic cluster is computed. The ϑ function transforms IF in a sorted vector. Finally, semantic clusters with IF values lower than the α^{th} percentile (in this case, σ) will be removed (filtered). The subjects discovered R, semantic clusters G, and community clusters C will be returned.

The HSD algorithm requires some inputs and parameters. Considering the clustering algorithm, we employed the Spectral Clustering [Von Luxburg, 2007] because it works well for a small number of clusters and is able to find clusters with non-flat geometry. Taking into account the user embeddings, we used the algorithm DSE presented in Chapter 4. Regarding the semantic word embeddings, we employed the embeddings provided by *word2vec*. The number of semantic clusters was defined as 500. Considering the number of communities and the maximum number of subjects, we set as 10 and 20, respectively. The percentile parameter α^{th} was set as 10 (10^{th} percentile).

5.3 Clustering Analysis

Now, in order to show the intuition behind algorithm HSD, we make clustering analysis of four mental disorders − ADHD, Anxiety, BD and PTSD. Through it, we find out implicit patterns in its textual data, as well as discover hidden subjects and themes in each cluster. Some details should be highlighted considering findings from this algorithm and are described in the following sections.

5.3.1 Attention Deficit Hyperactivity Disorder

Firstly, we made analysis of ADHD dataset. In Figure 5.1 we can visualize their user embeddings and the found clusters, using t-SNE. In Table 5.2 are described

Data: clustering algorithm A; user embeddings U; semantic word embeddings W; textual data T; number of semantic clusters k; number of communities c; maximum number of subjects s; α^{th} percentile
Result: subjects discovered R; semantic clusters G; community clusters C

$G, G' \leftarrow A(W, k)$
$C, C' \leftarrow A(U, c)$
$M \leftarrow \phi(C, T)$
initialize(F)
for *each community* $j \in M$ **do**
  initialize(H)
  for *each word* $w \in M(j)$ **do**
    $i \leftarrow G'(w)$
    $H(i) \leftarrow H(i) + 1$
  end
  $R(j) \leftarrow top(s, H)$
  for *each semantic cluster* $i \in R(j)$ **do**
    $F(i) \leftarrow F(i) + 1$
  end
end
initialize(IF)
for *each semantic cluster* $i \in G$ **do**
  $IF(i) \leftarrow \log \frac{c}{F(i)}$
end
$v \leftarrow \vartheta(IF)$
$\sigma \leftarrow v(|v| \cdot \alpha)$
for *each community* $j \in M$ **do**
  for *each semantic cluster* $i \in R(j)$ **do**
    if $IF(i) < \sigma$ **then**
      $R(j) \leftarrow R(j) - \{i\}$
    end
  end
end

Algorithm 1: Hidden Subject Discovery Algorithm.

information about the subjects of each cluster.

For instance, in cluster #2 we can note that their words indicate details about *drugs* and *treatment*. This means that users of this cluster generally comment, discuss, help and share experiences and information on these subjects. In the following case – *"**vyvanse** and **adderall** have by far been the best for me"* – the user states that the two drugs have been effective in her ADHD treatment. Another example is about the beneficial influence of a drug in her social life – *"the **adderall** helps my speaking skills a ton but it's just not affecting me anymore late at night, when I finally do go out with friends"*. Even on treatment, users share information about what the best drugs for certain symptoms – *"... those two **anticonvulsants** I mention can help with*

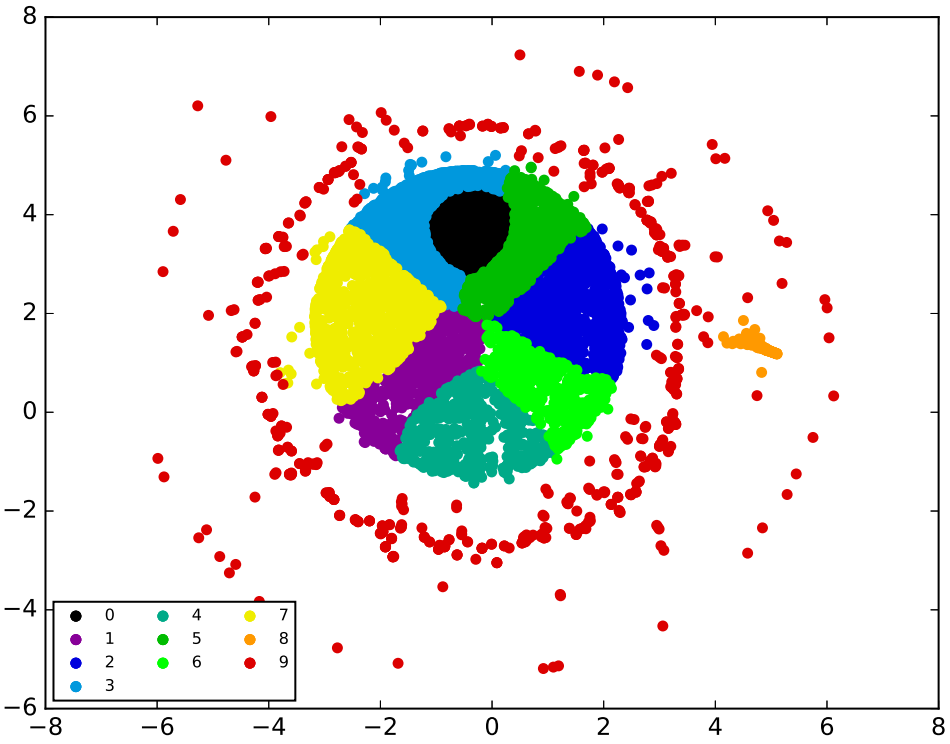


Figure 5.1: *t*-SNE Visualization of users with ADHD.

some problems, but there are other ***anticonvulsants*** *that work better for symptoms of irritability, anger, sudden, frustration".*

Similarly, another subject shared by users with ADHD is about drugs and treament *effects* (generally side-effects). For example, a user describes the harm caused by antipsychotics – *"...* ***antipsychotics*** *have very nasty side-effects like:* ***weight gain***, ***akathisia***, ***tardive dyskinesia***, *increased risk of type 2 diabetes and potentially even* ***brain damage****".* Considering the word *akathisia*, from National Center for Biotechnology Information (NCBI), this is a side-effect of antidepressant and atipsychotic drugs that causes agitation, extreme anxiety, and fidgetiness[2].

On the other hand, we also have clusters related to leisure activities such as *games*, *movies*, and *series*. Some users report that such activities help in controlling ADHD. For instance, in the following examples, a user mentions about *video games* – *"I found playing one hour of video* ***games*** *after school to help tremendously"* –, while another describes about *movies* – *"I also need subtitles captions with* ***movies*** *and* ***TV*** *shows for it helps me to focus".* In a similar way, physical sports has helped some users

[2] `http://www.ncbi.nlm.nih.gov/pubmed/10647977`

Table 5.2: ADHD Clusters.

Cluster ID	Cluster Label	Example Words
#0	Social Behavior	aggressive, arrogant, defeatist, egotistical, impolite, malicious, racist
#1	Effects	ache, akathisia, anxiousness, dehydration, diarrhea, disorientation, fever, sleeplessness
#2	Drugs/Treatment	adderall, anticonvulsant, antidepressant, antipsychotics, carbamazepine, flibanserin, milnacipran, vortioxetine
#3	Games	2048, 3ds, animations, consoles, gameplay, rpg, simulation
#4	Achievements	accomplishments, bravery, celebrate, congrats, congratulations
#5	Family	aunt, boyfriend, cousin, dad, girlfriend, husband, mother, wife
#6	School	anatomy, botany, calculus, english, geography, immunology, physics, stats
#7	Drinks	beer, beverage, coca, coffee, gatorade, soda, sprite, vodka, whisky, wine
#8	Fruits/Plants	agave, almond, banana, blackberry, cayenne, honey, lemon, peach, raspberry, strawberry
#9	Movies/Series	avatar, avengers, battlestar, dexter, hannibal, marvel, simpsons, supernatural, titanic

to relax – *"For some reason **golf** and **running** are the two things that help me **relax**".*

5.3.2 Anxiety

Now, the next analysis is made on the Anxiety dataset. Figure 5.2 shows a visualization of the user embeddings and found clusters regarding this dataset. In Table 5.3 are described information about the subjects of each cluster.

Firstly, as we can see in this table, at the cluster #9 users talk about other platforms related to *online support*. Such platforms serve as a complementary therapy to users. One of the most talked about is *7cupsoftea*[3]: a emotional health service that connect anonymously and real people to real listeners in a chat. The users on

[3] `http://www.7cups.com/`

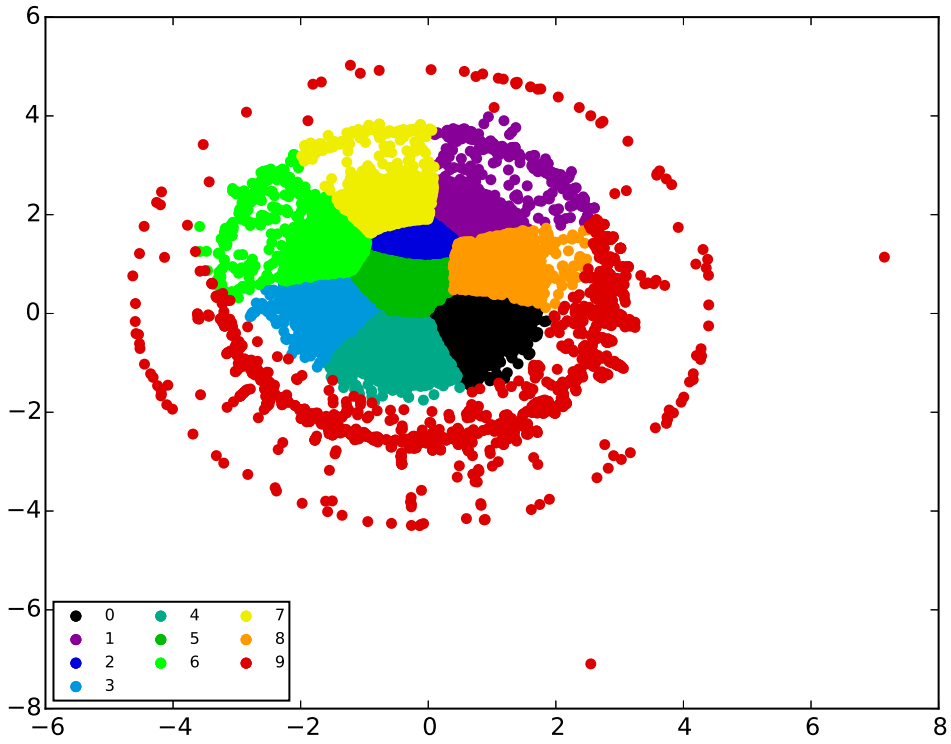


Figure 5.2: *t*-SNE Visualization of users with Anxiety.

Reddit recommend such services to other – *"**7cupsoftea** is great! It's anonymous and confidential I'm a listener myself but before that I was seeking help and 7 cups gave me that!"*. Another similar service mentioned by users is *blahtherapy*[4]. As we can observe in the following sentence by a user of this cluster – *"Another decent online one that I've used myself when having panic attacks is **blahtherapy**. People on there have always been really nice in my experience."* –, these services can help people overcome their problems with anxiety.

Another cluster that deserves to be discussed is about *bullying*. Here, users typically comment about persecution, harassment, or hostilities that occurred in their lives. For instance, the following user talks about his friends' behavior with him – *"I've been made fun of humiliated by people close to me."*. In another example – *"I left school when I was 15, partly because of my anxiety and partly because of intense bullying."* – we can see that this is an evidence that anxiety not properly treated can affect the learning process at school.

[4] `http://blahtherapy.com/`

Table 5.3: Anxiety Clusters.

Cluster ID	Cluster Label	Example Words
#0	Body	asymmetrical, bleeding, cheek, deformity, discoloration, eyebrows, itchy, redness, scaly, skin
#1	Kitchen	basket, blender, container, cupboard, dish, forks, freezer, jars, microwave, stove
#2	Bullying	accused, belittled, betrayed, confronted, criticized, disliked, embarrassed, humiliated, ostracized, ridiculed
#3	Drugs	abilify, accutane, aropax, buspirone, clonazepam, lexapro, prozac, sertralinem, topamax
#4	Music	acoustic, album, amy, beatles, britney, chopin, chorus, classical, instrumental, playlist, ramones
#5	Clothes	clothes, coat, corset, dresses, glasses, hat, hoodie, jacket, leather, pajama, pants, sweater, tshirt
#6	Jobs	accountant, artist, cosmetology, designer, engineer, freelance, journalist, librarian, programmer, writer
#7	Discussion	abducted, arrested, brutality, crime, genocide, homicide, molestation, murder, pedophile, rape, survivor
#8	Environments	appartment, bedroom, corridor, doors, floor, hall, kitchen, neighbor, toilets, yard
#9	Online Support	7cupsoftea, adaa, anxietynomore, anxietyzone, blahtherapy, misdirectedanxiety

5.3.3 Bipolar Disorder

In this subsection, we made analysis of bipolar disorder dataset. In Figure 5.3 we can visualize their user embeddings and the found clusters. In Table 5.4 are described information about the subjects of each cluster.

Taking into account the cluster #4, their words indicate details about *beliefs* and *religions*. In this cluster, users normally comment about their rituals and beliefs that help to overcome bipolar disorder, as in the following example – *"... and when my*

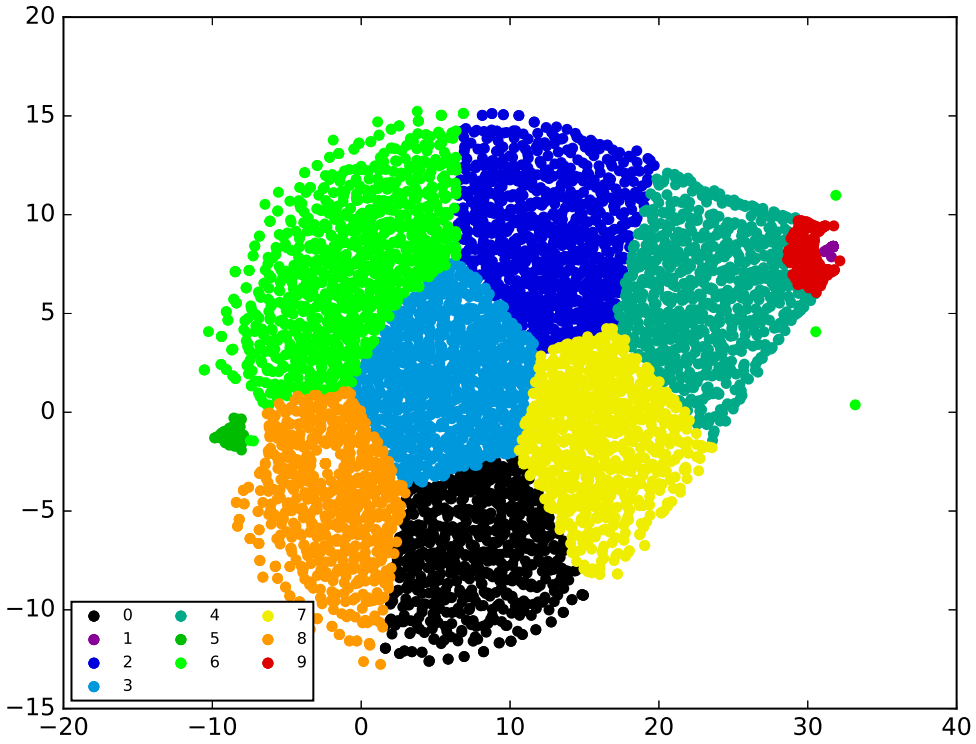


Figure 5.3: *t*-SNE Visualization of users with BD.

hallucinations get the better of me or the bad images come or the paranoia is creepin in, I'm known to say Jesus and it helps or put on my safe music which is worship". This another user found in religion a way to improve his bipolar disorder – *"I've also found learning more about religion and being spiritual helps me"*.

Another important subject discussed by users, in the cluster #7, is about *self-esteem*. A lot of users from this specific cluster relate low self-esteem and problems in their social, like, for instance, the following sentence – *"I actually have low* ***self-esteem****, it really puts people off and makes it difficult for me to* ***make friends****, thus I am extremely* ***lonely***". This another user afirms that self-esteem problems was caused by his bipolar disorder – *"I had major* ***self-esteem*** *issues caused by my* ***bipolarity*** *and purging just came along with it"*.

On the other hand, we also have clusters related to *diet*. Here, users discuss what they eat and what helps to control and treat their BD. For instance, in the following example, a user talks about a diet that was helpful for his treatment – *"Ketogenic diet it's a super low* ***carb diet*** *that puts your body into a state called ketosis which has been proven to have major* ***beneficial effects*** *on bipolar disorder"*.

Table 5.4: Bipolar Clusters.

Cluster ID	Cluster Label	Example Words
#0	Music	acoustic, album, drum, electronic, flute, indie, jazz, melodic, metal, orchestral, rap, symphony
#1	Clothes	clothes, costumes, dress, gloves, jeans, leggings, pants, skirts, socks, tshirt
#2	College	algebra, anthropology, arts, bachelor, biology, engineering, epidemiology, geology, math, philosophy, robotics, science
#3	Diet	breads, calorie, carbs, fiber, grains, greens, meat, peanuts, protein, soy, sugar, vegetables
#4	Beliefs/Religions	afterlife, aliens, angels, biblical, buddha, christ, demons, god, hells, incarnation, jesus, saints, satan
#5	Hobby/Sightseeing	beach, bonsai, buildings, mountains, ocean, skateboarding, traveling, waterfalls
#6	Drugs	anafranil, buproprion, divalproex, latuda, risperidol, seroquel
#7	Self-esteem	anguish, badness, betrayal, crushing, existential, heartbreak, humiliation, loneliness, shame
#8	Miscellaneous	anonymous, browse, forum, hijack, memes, moderator, newbies
#9	Social Behavior	ableist, arrogant, clueless, flippant, foolish, homophobic, hypocritical, racist, sarcastic

5.3.4 Post-Traumatic Stress Disorder

Finnaly, our last analysis is concerned with PTSD. Figure 5.4 shows a visualization of the user embeddings and found clusters regarding this dataset. In Table 5.5 are described information about the subjects of each cluster.

As we can see in Table 5.4, at the cluster #2 users talk about *therapy*. This indicates that users normally share with each other their experiences about this subject. For instance, in the following case – *"Now I'm in intensive* ***therapy*** *that I feel is going to help me get to the root of and eliminate the parasomnias"* – the user is happy about therapy, while in this another case –"*Most of the time in* ***therapy*** *I've felt worst*

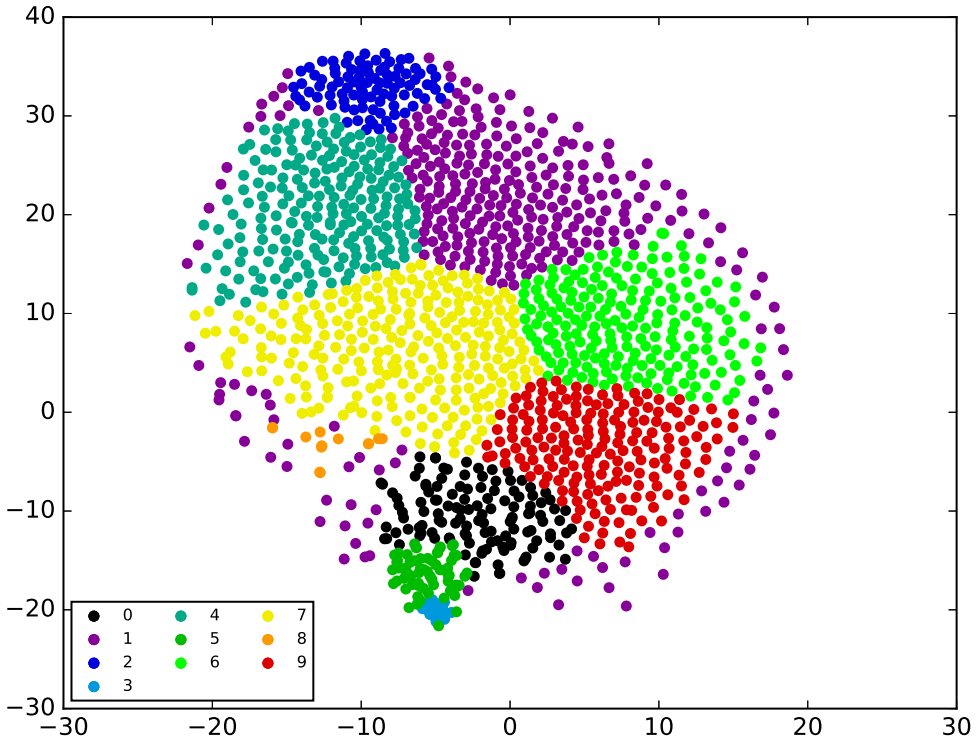


Figure 5.4: *t*-SNE Visualization of users with PTSD.

because all of the questions were about negative things or how I have been feeling bad just reinforcing the sense of helplessness" – the user is shown dissatisfied with therapy.

Another relevant cluster found by the algorithm is about *trauma*. The most discussed are those related to *abuse*. For instance, the following user relates abuse in their childhood the reason that caused PTSD. – *"I was diagnosed with PTSD several months ago stemming from childhood* ***abuse****"*. On the other hand, this one reports assault and rape which caused PTSD – *"I don't know if this affecting my recovery or is just fueling my depression, but I developed my PTSD from a violent sexual assault"*.

Table 5.5: PTSD Clusters.

Cluster ID	**Cluster Label**	**Example Words**
#0	Achievements	admire, challenge, conquer, gifts, journey, resiliency
#1	Drugs	adderall, antidepressant, cymbalta, miracle, paroxetine, ritalin, vistaril
#2	Therapy	analysis, ayahuasca, breakthroughs, clinical, oriented, participants, sections, surveys, therapies, treatments
#3	Trauma	abuse, abusers, assault, bizarre, damages, harassment, molestation, rapes, revenge, sexuality, threats, violate, violence
#4	Combat-related PTSD	battlefield, canadian, civilian, elderly, marine, military, served, soldiers, sufferer, troops, veterans, vietnam
#5	Activities	acupunture, conditioning, exercises, focusing, meditation, mindfullness, relaxation, skill, strategy, tasks, tool, yoga
#6	Family	abort, adopted, aunt, brother, cousin, dad, grandparents, husband, mom, sister, son
#7	College	academically, arts, bachelor, career, courses, deadlines, degree, grades, graduated, master, phd, student, teacher
#8	Internet	accessed, addresses, blogs, browser, download, internet, search, sites, tumblr, wiki
#9	Miscellaneous	air, bake, band, candle, dance, ear, garden, nutrients, pajamas, shelf, taste, water

Chapter 6

Conclusions

In this work we employed deep learning techniques in order to obtain high-quality text embeddings about mental disorders. We summarized mental disorders and present existent and new approaches to classify and analyse such disorders.

To perform the experiments, we collected from 8 subreddits related to mental disorder and from 12 subreddits that talks about different subjects, in order to represent a control group, that is, subreddits with subjects not related to mental disorders.

We developed convolutional neural networks architectures to learn these embeddings at sentence-, and user-level. This approach was named Disorder-Specific Embedding (DSE), due to the fact that, during the learning phase, the embedding vectors are built taking into account the mental disorder information.

We also proposed a new way to find out implicit patterns in textual data. Our approach was named Hidden Subject Discovery (HSD). This algorithm uses semantic word embeddings and clustering methods to discovery hidden subjects and themes in communities.

Considering the experiments concerned with classification, our approach DSE outperformed the considered baselines (LIWC, Bag of Words, Average Vector, and Paragraph Vector). Moreover, in a visualization scenario, via t-SNE, their embeddings provided an excellent discrimination among mental disorders.

Regarding our proposed algorithm HSD, we shown the intuition behind, taking into account four mental disorders for analysis. Through it, we could found out implicit patterns in its textual data, as well as discovered hidden subjects and themes in each cluster.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G. S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., and Zheng, X. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems. Software available from tensorflow.org.

Baeza-Yates, R. and Ribeiro-Neto, B. (2011). *Modern Information Retrieval - the concepts and technology behind search, Second edition.* Pearson Education Ltd., Harlow, England.

Balani, S. and De Choudhury, M. (2015). Detecting and characterizing mental health related self-disclosure in social media. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1373--1378. ACM.

Bengio, Y., Courville, A., and Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(8):1798--1828.

Chollet, F. (2015). Keras. `https://github.com/fchollet/keras`.

Ciresan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 3642--3649. IEEE.

Coppersmith, G., Dredze, M., and Harman, C. (2014a). Quantifying mental health signals in twitter. *ACL 2014*, page 51.

Coppersmith, G., Harman, C., and Dredze, M. (2014b). Measuring post traumatic stress disorder in twitter. In *ICWSM*.

De Choudhury, M., Counts, S., and Horvitz, E. (2013a). Social media as a measurement tool of depression in populations. In *Proceedings of the 5th Annual ACM Web Science Conference*, pages 47--56. ACM.

De Choudhury, M. and De, S. (2014). Mental health discourse on reddit: Self-disclosure, social support, and anonymity. In *ICWSM*. Citeseer.

De Choudhury, M., Gamon, M., Counts, S., and Horvitz, E. (2013b). Predicting depression via social media. In *ICWSM*, page 2.

Dos Santos, C. N. and Gatti, M. (2014). Deep convolutional neural networks for sentiment analysis of short texts. In *COLING*, pages 69--78.

Empresa Brasil de Comunicação (2016). Saúde mental: transtornos atingem cerca de 23 milhões de brasileiros. `http://www.ebc.com.br/noticias/saude/2013/05/saude-mental-em-numeros-cerca-de-23-milhoes-de-brasileiros-passam-por`. Access date: March 15, 2016.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). Deep learning. Book in preparation for MIT Press.

Ji, S., Xu, W., Yang, M., and Yu, K. (2013). 3d convolutional neural networks for human action recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 35(1):221--231.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., and Fei-Fei, L. (2014). Large-scale video classification with convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 1725--1732.

Kim, Y. (2014). Convolutional neural networks for sentence classification. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing, EMNLP 2014, October 25-29, 2014, Doha, Qatar, A meeting of SIGDAT, a Special Interest Group of the ACL*, pages 1746--1751.

Kingma, D. P. and Ba, J. (2015). Adam: A method for stochastic optimization. In *Proceedings of the 3rd International Conference for Learning Representations (ICLR)*.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097--1105.

Lawrence, S., Giles, C. L., Tsoi, A. C., and Back, A. D. (1997). Face recognition: A convolutional neural-network approach. *Neural Networks, IEEE Transactions on*, 8(1):98--113.

Le, Q. and Mikolov, T. (2014). Distributed representations of sentences and documents. In *Proceedings of the 31th International Conference on Machine Learning*, pages 1188--1196.

LeCun, Y. and Bengio, Y. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995.

Maas, A. L., Daly, R. E., Pham, P. T., Huang, D., Ng, A. Y., and Potts, C. (2011). Learning word vectors for sentiment analysis. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies-Volume 1*, pages 142--150. Association for Computational Linguistics.

Mikolov, T., Chen, K., Conrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. In *Proceedings of Workshop at International Conference on Learning Representations*.

Mitchell, M., Hollingshead, K., and Coppersmith, G. (2015). Quantifying the language of schizophrenia in social media. *NAACL HLT 2015*, page 11.

Pang, B., Lee, L., and Vaithyanathan, S. (2002). Thumbs up?: sentiment classification using machine learning techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, pages 79--86. Association for Computational Linguistics.

Pavalanathan, U. and De Choudhury, M. (2015). Identity management and mental health discourse in social media. In *Proceedings of the 24th International Conference on World Wide Web Companion*, pages 315--321. International World Wide Web Conferences Steering Committee.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825--2830.

Pennington, J., Socher, R., and Manning, C. D. (2014). Glove: Global vectors for word representation. In *EMNLP*, volume 14, pages 1532--1543.

Poria, S., Cambria, E., and Gelbukh, A. (2015). Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In *Proceedings of EMNLP*, pages 2539--2544.

Socher, R., Perelygin, A., Wu, J. Y., Chuang, J., Manning, C. D., Ng, A. Y., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the conference on empirical methods in natural language processing (EMNLP)*, volume 1631, page 1642. Citeseer.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *The Journal of Machine Learning Research*, 15(1):1929--1958.

Tang, D., Wei, F., Qin, B., Liu, T., and Zhou, M. (2014a). Coooolll: A deep learning system for twitter sentiment classification. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 208--212.

Tang, D., Wei, F., Yang, N., Zhou, M., Liu, T., and Qin, B. (2014b). Learning sentiment-specific word embedding for twitter sentiment classification. In *ACL (1)*, pages 1555--1565.

Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24--54.

Van den Oord, A., Dieleman, S., and Schrauwen, B. (2013). Deep content-based music recommendation. In *Advances in Neural Information Processing Systems*, pages 2643--2651.

Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85.

Von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4):395--416.

Vos, T., Barber, R. M., Bell, B., Bertozzi-Villa, A., Biryukov, S., Bolliger, I., Charlson, F., Davis, A., Degenhardt, L., Dicker, D., et al. (2015). Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic

diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the global burden of disease study 2013. *The Lancet*, 386(9995):743--800.

Wang, H., Wang, N., and Yeung, D.-Y. (2015). Collaborative deep learning for recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1235--1244. ACM.

Zhang, X., Zhao, J., and LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in Neural Information Processing Systems*, pages 649--657.