Demystifying Topology of Autopilot Thoughts: A Computational Analysis of Linguistic Patterns of Psychological Aspects in Mental Health

Bibekananda Kundu and Sanjay Kumar Choudhury
Language Technology, ICT and Services
Centre for Development of Advanced Computing, Kolkata
E-mail: {bibekananda.kundu,sanjay.choudhury}@cdac.in

Abstract

The paper investigates topology of uncontrolled dynamic depressive thoughts which is popularly known as *"autopilot"* in the psychology domain. Persistent homology, a mathematical tool from algebraic topological has been applied on Vector Space representation of tweets generated by users having neurotic personality for determining the topological structure of autopilot thoughts. State-of-the-art machine learning techniques leveraging linguistic resources akin to LIWC, WordNet-Affect and SentiWordNet have been applied for identifying neurotic personality from different Twitter users. An initiative has been taken for empowering Neuro Linguistic Programming (Bandler and Grinder, 1975; Bandler and Grinder, 1979; Bandler and Andreas, 1985) and other psychotherapy techniques using Natural Language Processing in the domain of Mental Health.

1 Introduction

"Wherever there are sensations, ideas, emotions, there must be words"

— Swami Vivekananda.

We use language for thinking, experiencing, expressing, communicating and problem solving. So, to analyze one's thought process, language is a symbolic medium. In psychotherapy, language is considered as a primary tool to understand patients' experiences and express therapeutic interventions (Pennebaker et al., 2003). All psychological interventions rely on the power of language. Psychotherapists rarely intervene directly in their client's lives, they create changes in the thought process through conversation (Villatte et al., 2015). According to *Relational Frame Theory (RFT)* (Greenway et al., 2010), people use linguistic frames to understand the world around them, and subsequently solve problems. *RFT* has been suggested as an approach to understanding natural language systems. The theory lends itself well to assessment with Natural Language Processing (NLP) precisely because it relies on understanding interaction between sensation, affect, language, and behaviour. When someone uses language, they are labelling their experience. For example, someone might tweet *"I need to escape this world before I get crushed."* indicating a fear based affective response. We are planning to use NLP to assess this label (Pennebaker et al., 2015). Simply labelling events and their attributes as 'positive' or 'negative' increases associated memories and emotional salience. This type of relational network can be evoked with any number of internal or external stimuli, triggering the aforementioned internal feedback loop, and leading to psychological distress. For example, describing a 'negative' event such as a trauma can evoke intense fear and sadness and subsequent sobbing (Miner et al., 2016; Althoff et al., 2016). The person suffering from distress actually using a model of world which is very limited and in this world he/she find no appropriate choice from the options available to their model of world (Bandler and Grinder, 1975; Bandler and Grinder, 1979; Bandler and An-

Proc. of ICON-2017, Kolkata, India. December 2017 ©2017 NLPAI, pages 435–446

dreas, 1985). Therefore, there is a requirement of expanding the model of world i.e. improvise the model to a better model which has more options. Therefore, the therapeutic technique would be somehow transforming the existing model to a better model using a meta-model and transformational grammar. The linguistic theory plays a vital role to understand the client model and transform it using transformational grammar (Bandler and Grinder, 1975). Therefore, one of the key concerns of psychotherapy is to understand topology of the maladaptive autopilot thoughts and changing the topology of thought process using Mindfulness (Collins et al., 2009), Collaborative Empiricism (Beck and Emery, 1979; Kazantzis et al., 2013) and other talk therapy techniques (Pawelczyk, 2011; Ebert et al., 2015; Mayo-Wilson and Montgomery, 2013; Mohr et al., 2013). In this regards, understanding of topology of uncontrolled dynamic depressive thought (known as *"autopilot thought"* in psychology) is important for evaluating mental health of patients. After a brief discussion on psychological background and motivation behind the work, we will understand how we can represent topology of thought in the next section.

2 How to Represent Topology of Thought

We are considering written text as a symbolic representation of thoughts. To understand the topology of *"autopilot thoughts"*, we have collected tweets of neurotic personality from Twitter applying a hybrid approach combining Deep Learning based classification, KL-Divergence (Manning and Schütze, 1999) based Timeline Similarity Analysis and Rule-based sentiment analysis technique leveraging WordNet-Affect[1] , SentiWordNet[2] and psycholinguistic resource akin to LIWC[3]. Detailed data collection procedure has been discussed in the section 3 and 4 .We are interested to study the representation of words used by neurotic persons in the Vector Space using Word Embedding (Mikolov et al., 2013; Mesnil et al., 2013) and topology (Sizemore et al., 2016) of these semantically embedded words using persistent homology (Zhu, 2013; Kaczynski et al., 2004). We have intuition that timeline of neurotic person contains different topological structure than timeline of user having other personality. Studies say that neurotic person uses more first person pronoun, less social words, more negative emotion words (Pennebaker, 2011). An introvert person uses single topic, discusses more regarding problem, uses few self-references, many tentative words, many negation as compare to extrovert person (Mairesse and Walker, 2007). Topological data analysis using persistent homology has been discussed in the section 5. In the next section, we will discuss our data collection procedure from Twitter.

3 How to Collect Tweets of Neurotic Persons

The proposed approach utilizes an ensemble of state-of-the-art machine learning techniques based on psycholinguistic features to detect distress users (having neurotic personality) from their social media text. We have used Twitter API to search in the Twitter using some seed words/phrases like *"awful"*, *"terrible"*, *"lousy"*, *"hate"*, *"lonely"*, *"hopeless"*, *"helpless"*, *"crap"*, *"sad"*, *"miserable"*, *"tired"*, *"sleep"*, *"hurt"*, *"pain"*, *"kill"*, *"die"*, *"dying"*, *"stressed"*, *"frustrated"*, *"irritated"*, *"depressed"* etc. and name of some antidepression drugs like *"Sertraline"*, *"Citalopram"*, *"Clonazepam"*, *"Propanol"*, *"Prozac"*, *"Zopiclone"*, *"Fluoxetine"*, *"Quetiapine"*, *"Hydroxyzine"* etc. Next, we have filtered out the tweets starting with RT to avoid considering retweets. We have also removed tweets containing url. Thereafter, these tweets are sent to (in house developed) Psychological Annotation Interface for manual annotation. Figure 1 shows screenshot of the interface along with some examples of negative tweets. An annotator can label a tweet considering three aspects Viz:

(a) Personal/Impersonal Emotion Labelling

(b) Polarity Labelling

(c) Psychological Annotation

Individual words in a tweet are annotated according to Psychological Process as discussed

[1] http://wndomains.fbk.eu/wnaffect.html
[2] http://sentiwordnet.isti.cnr.it
[3] http://liwc.wpengine.com/

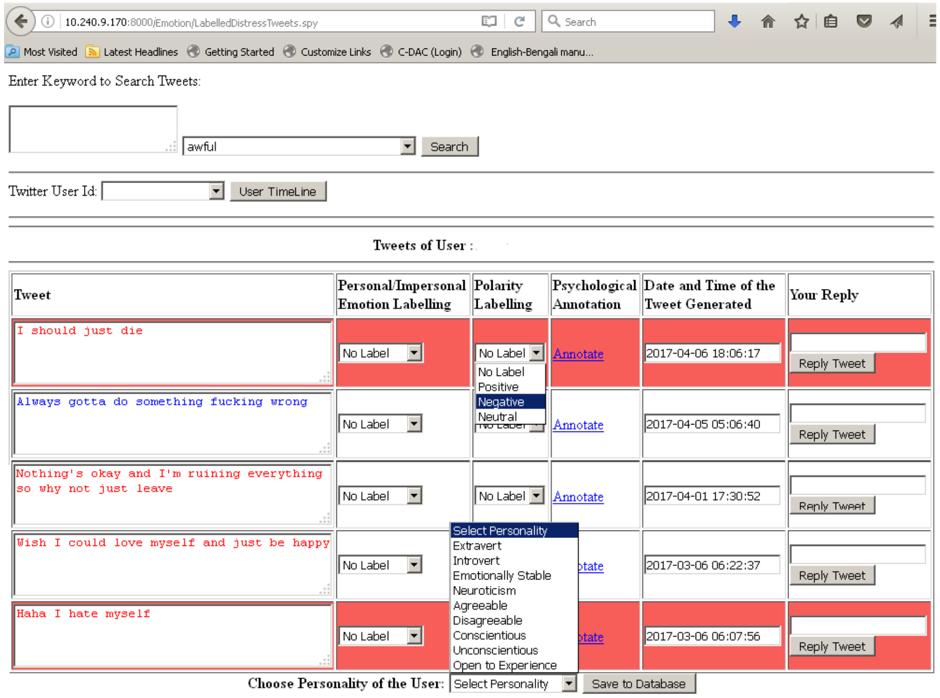

Figure 1: Psychological Annotation Interface

in (Pennebaker et al., 2015). Special care has been taken during annotation to find out "Linguistic Marker of Depression" (Bucci and Freedman, 1981; Pyszcynski and Greenberg, 1987) in the tweet. Pronouns tell us where people focus their attention. If someone uses the pronoun *"I"*, it's a sign of self-focus. Depressed people use the word *"I"* much more often than emotionally stable people (Pennebaker, 2011; Ramirezesparza et al., 2008; Nguyen et al., 2014). Researchers have found that people who frequently use first-person singular words like *"I"*, *"me"* and *"myself"* are more likely to be depressed and have more interpersonal problems than people who often say *"we"* and *"us"*. Using LIWC2001, (Stirman and Pennebaker, 2001) found that suicidal poets were more likely to use first person pronouns (e.g., *"I"*, *"me"*, *"mine"*) and less first plural pronouns (e.g., *"we"*, *"ours"*) throughout their writing careers than were non-suicidal poets. These findings supported the social engagement/disengagement model of depression, which states that suicidal individuals have failed to integrate into society in some way, and are therefore detached from social life (Durkheim, 1951). Similarly, (Rude et al., 2004) found that currently depressed students used more first person singular pronouns, more negative emotional words, and slightly fewer positive emotion words in their essays about coming to college, relative to students who had never experienced a depressive episode. These results are in line with (Pyszcynski and Greenberg, 1987) self-awareness theory. Therefore, in our Psycho-logical Annotation Interface, we have implemented a feature to highlight tweets with red back ground containing First Person Personal Pronoun i.e. *"I"*. Focus on temporal orientation of people that is how often they emphasize the past, present and future is necessary because it affects their health and happiness (Zimbardo and Boyd, 2008). We are interested the proportion of a user's tweets that the analytic finds evidence in: Insomnia and Sleep Disturbance which is often a symptom of mental health disorders (Weissman et al., 1996; De Choudhury et al., 2013), so we have calculated the proportion of tweets that a user makes between midnight and 4 am according to their local time-zone. Therefore, during annotation, special attention should be given on the time perspective of the tweets. Moreover, in the interface we have highlighted tweets in grey colour that have been generated midnight assuming that the user have sleeping problem or insomnia. From all labelled tweets, unique users name are automatically extracted for analysis of their timelines. Thereafter tweets are extracted from unique users' timeline and annotated using the aforementioned procedure. After annotation of all tweets from a user timeline, the personality of tweet user is labelled in one of the nine classes viz.*Extravert*, *Introvert*, *Emotionally Stable*, *Neuroticism*, *Agreeable*, *Disagreeable*,*Conscientious*,*Unconscientious* and *Open to Experience*. This personality classes are basically extended form of *"Big Five Personality"* (John and Srivastava, 1999) classes. We are more interested for the users those are labelled as *"Neuroticism"* after analyzing the tweets from their timeline. Following the above mentioned procedure, we have created our training data. Manual analysis of tweets collected from users' timeline reveals that information contains in two neurotic users is similar. Moreover, people discuss similar problem among their friend circles, therefore automatic searching *"friend"* and *"followers"* of a neurotic users increases the chance of getting more data of similar nature automatically.

In this way collected training data has been used to train our ensemble learner for detecting depression from social media text in order to collect more tweets from neurotic users. In the next session, we will discuss the depression detection technique using an ensemble classifier.

4 Depression Detection from Social Media Text

4.1 Why Social Media

Currently, depression is primarily assessed through surveys. The standard approach to diagnosing psychological health disorders is through a series of clinically administered diagnostic interviews and tests (Weathers and Davidson, 2001). However, assessment of patients using these tests is expensive and time-consuming. Furthermore, the stigma associated with mental illnesses motivates inaccurate self-reporting by affected individuals and their family members, thus making the tests unreliable. Commonly, the evaluation of a patient is typically performed through the use of standardized questionnaires like Beck Depression Inventory (BDI)[4] , Big Five Inventary (BFI) (John and Srivastava, 1999) etc. A patient's answers are then compiled and compared with disease classification guidelines, such as the International Classification of Diseases or the Diagnostic and Statistical Manual (DSM), to guide the patient's diagnosis. However, these diagnostic methods are not precise and have high rates of false positives and false negatives. In addition, societal and financial barriers prevent many people from seeking medical attention (Michels et al., 2006). Many societies around the world stigmatize and discriminate against people with mental disorders, contributing to the unwillingness of individuals to acknowledge the problem and seek help (Fabrega, 1991). While psychological treatments for depression can be effective (Cuijpers et al., 2008), they are often plagued by access barriers and high rates of attrition (Mohr et al., 2010). Internet interventions have been touted as an antidote to access barriers, but they appear to produce more modest outcomes (Andersson and Cuijpers, 2009), in part also due to high attrition (Christensen et al., 2009). In recent years, there has been a tremendous growth in social interactions on the Internet via social networking sites and online discussion forums. In contrast to clinical tests, the Internet is an ideal, anonymous medium for distressed individuals to relate their experiences, seek knowledge, and reach out for help. Social media is an emerging tool that may assist research in this area, as there exists the possibility of passively surveying and then subsequently influencing large groups of people in real time. (Ruder et al., 2011) have shown that some Facebook users do, in fact, post suicide notes on their profiles, exposing the potential for suicide related research in social media. The amount of publicly available information spread across the realm of social media is extensive. We prefer Twitter because of its greater public availability of data, larger user base, and it being a platform of personal expression. Users generate over 400 million tweets per day (Bennett, 2012). This large reservoir of information regarding

[4] `http://www.hr.ucdavis.edu/asap/pdf_files/Beck_Depression_Inventory.pdf`

people's daily lives and behaviours, if handled correctly, can be used to study depression, suicide and possibly intervene. Twitter is also used for keeping in touch with friends and colleagues, sharing interesting information within one's network, seeking help and opinions, and releasing emotional stress (Johnston and Hauman, 2013). Therefore, Twitter can be identified as an important surveillance tool for detecting depression and suicidal patterns.

4.2 Methodology

We have applied an ensemble classifier to classify distress and non-distress user based on their social media text collected from Twitter. The ensemble classifier has been built using linear combination of Document Similarity and Emotional Intensity Estimator. Weights in this linear combination are estimated empirically to achieve higher accuracy in this classification task.

We have followed two approaches for detecting depressive tweets viz.

(a) **Document Similarity Measurement:** We have used KL-Divergence (Manning and Schütze, 1999) to measure similarity between searched user's timeline and labelled tweets from all users' timeline and used the similarity score for final scoring of negativity of the user. Applying Latent Dirichlet Allocation (LDA) (Blei et al., 2003), we have estimated topic distribution on the labelled negative[5] and positive data extracted from users' timeline. We have used `sklearn.lda.LDA` library to estimate topic distribution. Then we have estimated the topic similarity of a query user's tweets (user whose personality needs to be estimated) from their timeline with these labelled negative and positive tweets. Then we have estimated the overall similarity score using the following equation:

$$SimilarityScore = 0.6 * \beta + 0.4 * \gamma \quad (1)$$

where β and γ are the similarity scores estimated using LDA and KL-Divergence based topic distribution. Study shows that theme-based retrieval does a better job of finding relevant and effective documents (tweets in user timeline in our case) for this application than conventional approaches (Dinakar et al., 2012b; Dinakar et al., 2012a). All the weights used in the above equations are empirically determined.

(b) **Emotional Intensity Measurement:** We have used following resources for measuring emotional intensity of individual tweet:

(a) SentiWordNet

(b) Manually classified lexicon based on psychological process akin to LIWC.

(c) WordNet Affect

We have calculated $NegetivityScore$ combining LSTM (Hochreiter and Schmidhuber, 1997; Gers et al., 2000; Graves, 2012), a deep learning model for detecting polarity from tweets and rule-based approach using the above mentioned resources and psycholinguistic features.Features used in this classification task are mainly psycholinguistic types, other than that "Pattern of Life Analytics"[6] (Greetham et al., 2011; Berkman et al., 2000; De Choudhury et al., 2013), *"Capitalized Text"*, *"SpecialHashTag"*, *"Probability of Personal Pronoun"*, *"UserName Conatining Special Keywords"* etc. have also been used. Count of some common phrases like *"why me"*, *"I hate myself"* etc. have also been considered as important feature. Examples of "SpecialHashTag" feature are *#depressionprobs*, *#thisiswhatdepressionlike*, *#depression*, *#suicide* etc.. It have been seen that if userid of the users contains some clue substrings like *"depressing"*, *"depression"*, *"hell"*,

[5] In this paper we have used the term "negative tweet" and "distressed tweet" interchangeably to represent the tweet generate by user having neurotic personality.

[6] Social engagement has been correlated with positive mental health outcomes. Tweet rate measures how often a Twitter user posts and pro-portion of tweets with @mentions measures how often a user posts 'in conversation' with other users. Number of @mentions is a measure of how often the user in question engages other users, while Number of self @mentions is a measure of how often the user responds to mentions of themselves.

"depressed", *"sad"*, *"cry"*, *"suicidal"*, *"anxious"*, *"anxiety"*, *"lonely"*, *"die"*, *"broken"*, *"stress"*, *"worthless"*, *"lost"* etc. the timeline of these users contains depressive tweets. Therefore the users having such userid have been considered as an important feature. Tweets written in Upper Case, are considered as important assuming that these are written in Upper Case for providing more importance/intensifying the emotion involved in the tweet.We have used `Theano`, a python based deep leaning library for implementing our LSTM classifier[7].

Final score for selecting neurotic persons has been calculated as follows :

$$FinalScore = \alpha * SimilarityScore + (1-\alpha) * NegetivityScore \quad (2)$$

Value of alpha ($0 \leq \alpha \leq 1$) can be set experimentally to achieve highest accuracy. We have seen empirically that better result is found when the value of α is 0.8. It has been observed that when the *FinalScore* is greater than 0.14 then the user can be accepted as neurotic person. Following the procedure discussed in section 3 and 4 we have collected 2500 negative tweets from the timeline of 12 Twitter users having neurotic personality. Same numbers of positive tweets have been collected from the timeline of users having tweets with hashtag *#motivationaltweet*, *#positivethinking*, *#motivationalquotes* etc.

4.3 Vector Space Representation of Distressed Tweets

After manual verification, we have converted the negative and positive tweets into the multidimensional Vector Space. Thereafter, using "t-Distributed Stochastic Neighbor Embedding" (t-SNE) technique (van der Maaten and Hinton, 2008), the higher dimensional data points are projected into a 2d plane. We have used `gensim`[8] python library to convert neurotic persons' tweets into Vector Space. We have considered 2000 dimensions and ±5 context window during Word Embedding. Words that are appeared at least 10 times in the corpus have been selected for vector representation. We have used t-SNE api available in the `sklearn.manifold` library[9] for dimensionality reduction of these higher dimensional points and visualization in the two dimension space. Figure 2 shows representation of embedding words in 2d space using t-SNE. We can see semantically closer words are forming clusters in the Vector Space. *"Kill me"*, *"Suicidal"*, *"destroy"*, *"Cutting"* are appearing closer to each other and *"rejected"*, *"unloved"*, *"worthless"* are forming separate cluster. Separate cluster represent the different topic of the thoughts those are having in the mind of neurotic persons. Conversely, analysing the tweets of positive minded people, we have seen that *"adorable"*, *"comfortable"*, *"eager"*, *"hopeful"*,*"satisfied"* etc. words are frequently used in their timeline. Persistence homology has been applied to the point clouds of positive and negative tweets separately. In the next section we will discuss the topological data analysis of negative and positive tweets based on their vector representation.

5 Topological Data Analysis of Tweets

Persistent homology (Zhu, 2013), a mathematical tool from topological data analysis has been applied on the collected tweets for multiscale analysis on a set of points and identifies clusters, holes, and voids therein. Persistent homology can identify clusters (0-th order holes), holes (1st order, as in our loopy curve), voids (2nd order holes, the inside of a balloon), and so on in a point cloud. It finds "holes" by identifying equivalent cycles. Detailed discussion on Persistent homology[10] and Algebraic Topology is out of scope of the paper. Interested readers can follow work of (Zhu, 2013; Singh et al., 2008; Giblin, 2010; Freedman and Chen, 2011; Zomorodian, 2001; Carlsson, 2008; Edelsbrunner and Harer, 2010; Hatcher, 2002). After representing the words in Vector Space, we have used these data points

[7] http://deeplearning.net/tutorial/lstm.html
[8] https://radimrehurek.com/gensim/
[9] http://scikit-learn.org/
[10] http://outlace.com/

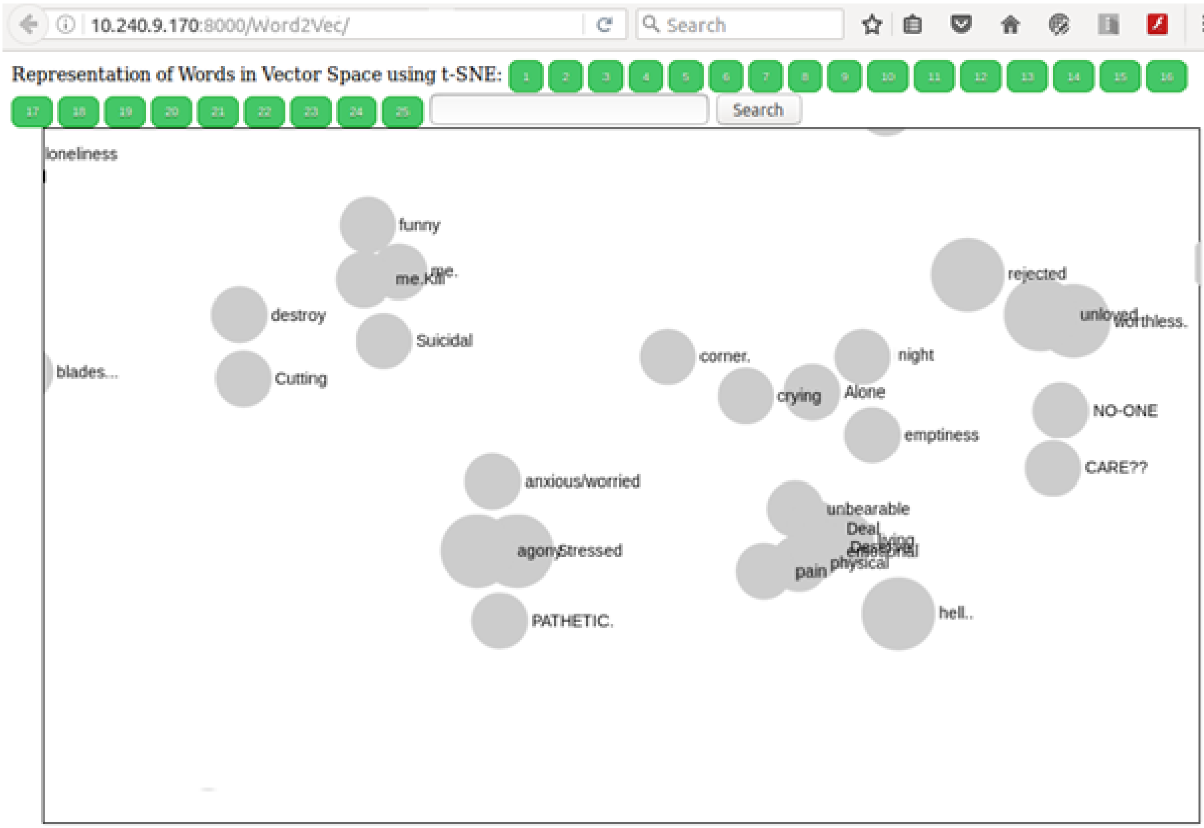

Figure 2: Representation of Words in Vector Space using t-SNE

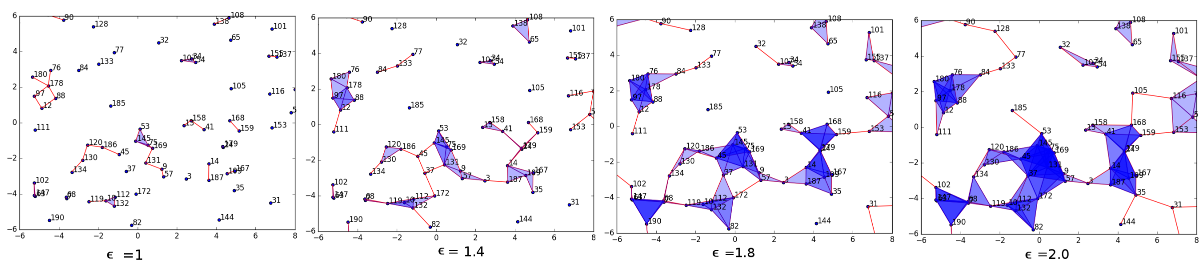

Figure 3: Generated Vietoris-Rips Complexes on Negative Point Cloud with Incremental Values of ϵ.

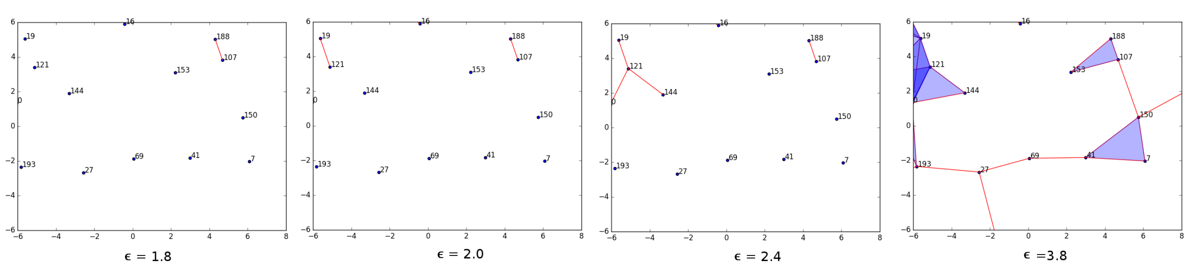

Figure 4: Generated Vietoris-Rips Complexes on Positive Point Cloud with Incremental Values of ϵ.

to build *Vietoris-Rips*[11] complexes of diameter ϵ which are simplicial complexes $VR(\epsilon) = \{\sigma | diam(\sigma) < \epsilon\}$. Here $diam(\sigma)$ represents the largest distance between two points in σ. Distance measures varies according to different contexts. Here we have used euclidean distance for our purpose. Figure 3 and figure 4 show generated *Vietoris-Rips* complexes on negative and positive point cloud respectively. Here we can see, if we set ϵ too small, then generated complexes may just consist of the original point cloud, or only a few edges between the points. If we set ϵ too big, then the point cloud will just become one massive ultradimensional simplex. Our intention is to discover meaningful patterns in a simplicial complex by continuously varying the ϵ parameter (and continually re-build complexes) from 0 to a maximum that results in a single massive simplex. Then we generate a diagram that shows what topological features are born and die as ϵ continuously increases. We assume that features that persist for long intervals over ϵ are meaningful features whereas features that are very short-lived are likely noise. This procedure is called persistent homology computation as it finds the homological features of a topological space that persist while we vary ϵ. Persistent homology examines all ϵ's to see how the system of hole change (also known as *"Birth and Death process"*). An increasing sequence of ϵ produces a filtration. Persistent homology tracks homology classes along the filtration to know for what value of ϵ does a hole appear and how long the hole persists. We have followed the methodology as reported in (Carlsson, 2008) to study the homology of the complexes constructed. The steps involve in this methodology are as follows:

- Construct the $\mathbb{R}$ persistence simplicial complex $\{C_\epsilon\}$ using *Vietoris-Rips* method.
- Select a partial order preserving map $f : \mathbb{N} \rightarrow \mathbb{R}$
- Construct the associated $\mathbb{N}$-persistence simplicial complex.
- Construct the associated $\mathbb{N}$-persistence chain complex $\{C_*(n)\}_n$ with co-efficients in F.
- Compute the barcodes associated to the $\mathbb{N}$-persistence F-vector spaces $\{H_i(C_*(n), F)\}_n$

Please refer (Carlsson, 2008) for detail explanation.

The "barcode plot" is a convenient way to visualize persistent homology (Zhu, 2013; Ghrist, 2007). Barcode plot shown in figure 5 and figure 6 are drawn based on increasing sequence of ϵ and zeroth Betti number (β_0) calculated from positive and distressed tweets respectively.

We have selected 500 data points randomly from the word to vector representation of positive and distressed tweets for the filtration process. The word to vector representation using Word Embedding ensures that words that share common contexts (semantics) in the negative and positive tweets are located in close proximity to one another in the Vector Space. Using persistent homology we are trying to examine the topology of these semantically oriented data points (words). The number of connected components is an important topological invariant of a graph. In topological graph theory, it can be interpreted as the zeroth Betti number of the graph. From figure 5, we can see that positive tweets have less connected components (142 disconnected components out of 500 data points) whereas figure 6 shows that negative tweets have much more connected component compare to positive tweets (only 7 disconnected components). Less number of connected components in users' timeline represents wide variation of topics. Conversely, less number of disconnected components indicate that tweets are much more focused towards some specific topics. Manually investigating the contents of the positive and negative tweets, we have seen that users having neurotic personality discuss more regarding their pain and problems. Hence, topic discussed in their timeline more focused to their problem area. On the other hand, positive minded users discuss on different topics and also share ideas and thoughts among their friends and followers. Therefore, tweets generated by them have wide variation of topics.

Barcode plot shown in figure 7 and figure 8 are

[11] http://outlace.com/

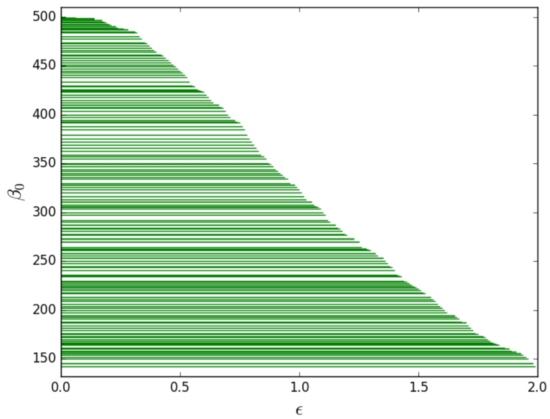

Figure 5: Barcode Plot of Positive Tweets at Betti Dimension 0 (β_0)

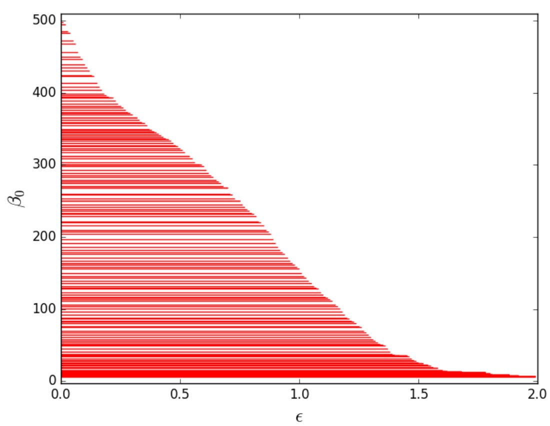

Figure 6: Barcode Plot of Distressed Tweets at Betti Dimension 0 (β_0)

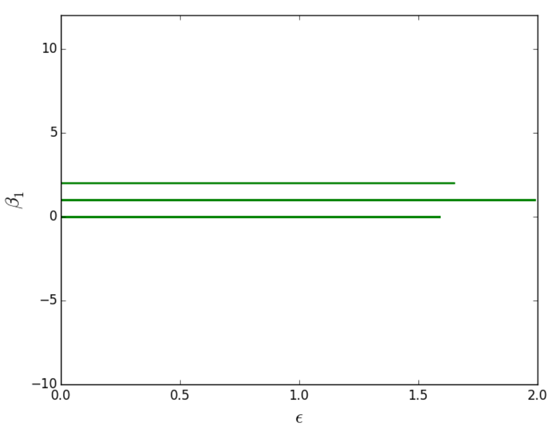

Figure 7: Barcode Plot of Positive Tweets at Betti Dimension 1 (β_1)

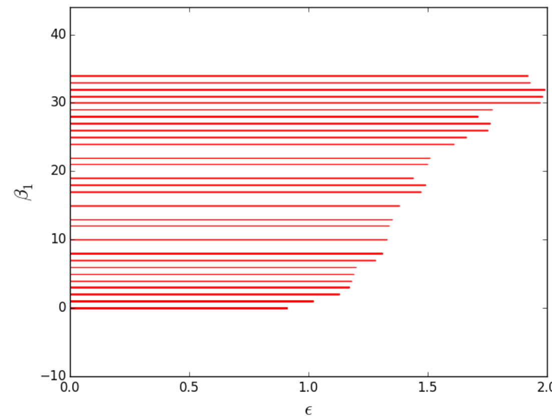

Figure 8: Barcode Plot of Distressed Tweets at Betti Dimension 1 (β_1)

drawn based on increasing sequence of ϵ and 1st Betti number (β_1) calculated from positive and distressed tweets respectively. We can see that figure 7 has very less number of holes and number of holes in the figure 8 are much more compare to figure 7. As the number of disconnected components are much more in the positive tweets, the chance of appearance of one dimensional holes are less. Conversely, number of one dimensional holes are much more in negative tweets because of less number of disconnected components. We have found that number of one dimensional holes in positive tweets is 1 and for negative tweets, it is 34. This observation corroborates the first observations that the people with negative mindset has more oriented set of thoughts (focused to their problem domain) than people having positive mindset. The higher order homology groups produces Betti numbers having values zeros, as expected.

6 Conclusion

In this paper, we have proposed a novel approach for collecting tweets of neurotic persons. Then these tweets are represented in the Vector Space using Word Embedding and dimensionality has been reduced using t-SNE. Persistent homology has been applied to analyse the topology of tweets resembling autopilot thoughts. Psychological features in term of linguistic pattern has been discussed.

As a future work we are planning to explore, how natural language generation can be applied for therapeutic text generation following RFT and based on topology of patients' thought. We have hypothesized that tweets having psychological features, linguistic markers of depression are indicator of neurotic user's time line as per our understanding of literature. Therefore, as a future work we would like to get expert guidance from psychotherapists for better understanding of the psychological process involved in Mental Health.

The work discussed in this paper is an initiative towards applying NLP in the domain of Mental Health which will motivate researchers for further exploration of linguistic markers and topology involved in psychology.

References

Tim Althoff, Kevin Clark, and Jure Leskovec. 2016. Natural language processing for mental health: Large scale discourse analysis of counseling conversations. *Transactions of the Association for Computational Linguistics*.

G Andersson and P Cuijpers. 2009. Internet-based and other computerized psychological treatments for adult depression: a meta-analysis. *Cognitive Behavioral Therapy*, 38(4):196–205.

R. Bandler and S. Andreas. 1985. *Using Your Brain for a Change*.

R. Bandler and J. Grinder. 1975. *The Structure of Magic I: A Book About Language and Therapy*. Science and Behavior Books.

R. Bandler and J. Grinder. 1979. *Frogs into Princes: Neuro Linguistic Programming*. Real People Press.

Rush A. J. Shaw B. F. Beck, A. T. and G. Emery. 1979. *Cognitive therapy of depression*. Guilford Publications, New york.

Shea Bennett. 2012. Twitter now seeing 400 million tweets per day, increased mobile ad revenue, says ceo@online.

Lisa F. Berkman, Thomas Glass, Ian Brissette, and Teresa E. Seeman. 2000. From social integration to health: Durkheim in the new millennium? *Social Science and Medicine*, 51(6):843—857.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.*, 3:993–1022.

W. Bucci and N. Freedman. 1981. The language of depression. *Bulletin of the Menninger Clinic*, 45(4):334–358.

Gunnar Carlsson. 2008. Topology and data. Technical report.

H Christensen, KM Griffiths, and Farrer L. 2009. Review adherence in internet interventions for anxiety and depression. *Journal of Medical Internet Research*, 11(2).

SE Collins, N Chawla, SH Hsu, J Grow, JM Otto, and Marlatt GA. 2009. Language-based measures of mindfulness: initial validity and clinical utility. *Society of Psychologists in Addictive Behaviors*, 23(4):743–749.

P Cuijpers, A Van Straten, L Warmerdam, and N Smits. 2008. Characteristics of effective psychological treatments of depression: a meta regression analysis. *Psychother Res*, 18(2):225–36.

Munmun De Choudhury, Scott Counts, and Eric Horvitz. 2013. Predicting postpartum changes in emotion and behavior via social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '13, pages 3267–3276, New York, NY, USA. ACM.

Karthik Dinakar, Birago Jones, Catherine Havasi, Henry Lieberman, and Rosalind Picard. 2012a. Common sense reasoning for detection, prevention, and mitigation of cyberbullying. *ACM Trans. Interact. Intell. Syst.*, 2(3):18:1–18:30, September.

Karthik Dinakar, Birago Jones, Henry Lieberman, Rosalind W. Picard, Carolyn Penstein Rosé, Matthew Thoman, and Roi Reichart. 2012b. You too?! mixed-initiative LDA story matching to help teens in distress. In *Proceedings of the Sixth International Conference on Weblogs and Social Media, Dublin, Ireland, June 4-7, 2012*.

E. Durkheim. 1951. *Suicide*. Free Press, New York.

DD Ebert, AC Zarski, H Christensen, Y Stikkelbroek, P Cuijpers, M Berking, and H Riper. 2015. Internet and computer-based cognitive behavioral therapy for anxiety and depression in youth: a meta-analysis of randomized controlled outcome trials. *PLoS One*, 10(3).

Herbert Edelsbrunner and John Harer. 2010. *Computational Topology - an Introduction*. American Mathematical Society.

Horacio Fabrega. 1991. Psychiatric stigma in nonwestern societies. *Comprehensive Psychiatry*, 32:534–551.

Daniel Freedman and Chao Chen. 2011. Algebraic topology for computer vision. *Computer Vision*, 5:239—268.

Felixv Gers, Jurgen Schmidhuber, and Fred Cummins. 2000. Learning to forget continual prediction with lstm. *Neural Comput*, 12(10):2451–2471.

Robert Ghrist. 2007. Barcodes: The persistent topology of data. Technical report.

Peter Giblin. 2010. *Graphs, Surfaces and Homology*. Cambridge University Press.

Alex Graves. 2012. *Supervised Sequence Labelling with Recurrent Neural Networks*, volume 385 of *Studies in Computational Intelligence*. Springer.

David E Greenway, Emily K Sandoz, and David R Perkins. 2010. Potential applications of relational frame theory to natural language systems. In *Proceedings of the Seventh International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*, pages 2955–2958. IEEE.

Danica Vukadinovic Greetham, Robert Hurling, Gab rielle Osborne, and Alex Linley. 2011. Social networks and positive and negative affect. *Procedia-Social and Behavioral Sciences*, 22:4—13.

Allen Hatcher. 2002. *Algebraic topology*. Cambridge University Press, Cambridge.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural Comput.*, 9(8):1735–1780, November.

O. P. John and S. Srivastava. 1999. The big-five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, 2:102–138.

Chan M.M Johnston, K. and M. Hauman. 2013. Use, perception and attitude of university students towards facebook and twitter. *The Electronic Journal Information Systems Evaluation*, 16(3):201–211, November.

T. Kaczynski, K. Mischaikow, and M. Mozek. 2004. *Computational Homology*. Springer.

Nikolaos Kazantzis, John Tee, Frank Dattilio, and Keith Dobson. 2013. Collaborative empiricism as the central therapeutic relationship element in cbt an expert panel discussion. *International Journal of Cognitive Therapy*, 6(4).

François Mairesse and Marilyn Walker. 2007. Personage: Personality generation for dialogue. In *In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 496–503.

Christopher D. Manning and Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. MIT Press, Cambridge, MA, USA.

Evan Mayo-Wilson and Paul Montgomery. 2013. Media-delivered cognitive behavioural therapy and behavioural therapy (self-help) for anxiety disorders in adults. *Cochrane Database Syst Rev*, 9.

Grégoire Mesnil, Xiaodong He, Li Deng, and Yoshua Bengio. 2013. Investigation of recurrent-neural-network architectures and learning methods for spoken language understanding. In Frédéric Bimbot, Christophe Cerisara, Cécile Fougeron, Guillaume Gravier, Lori Lamel, François Pellegrino, and Pascal Perrier, editors, *INTERSPEECH*, pages 3771–3775. ISCA.

Kathleen M. Michels, Karen J. Hofman, Gerald T. Keusch, and Sharon H. Hrynhow. 2006. Stigma and global health: Looking forward. *Lancet*, 367:538–539.

Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient estimation of word representations in vector space. *CoRR*, abs/1301.3781.

Adam Miner, Amanda Chow, Sarah Adler, Ilia Zaitsev, Paul Tero, Alison Darcy, and Andreas Paepcke. 2016. Conversational agents and mental health: Theory-informed assessment of language and affect. In *Proceedings of the Fourth International Conference on Human Agent Interaction*, HAI '16, pages 123–130, New York, NY, USA. ACM.

DC Mohr, J Ho, J Duffecy, KG Baron, KA Lehman, L Jin, and D Reifler. 2010. Perceived barriers to psychological treatments and their relationship to depression. *Journal of Clinical Psychology*, 66(4):394–409.

David C Mohr, Michelle Nicole Burns, Stephen M Schueller, Gregory Clarke, and Michael Klinkman. 2013. Behavioral intervention technologies: evidence review and recommendations for future research in mental health. *General hospital psychiatry*, 35(4):332–338.

Thin Nguyen, Dinh Phung, Bo Dao, Svetha Venkatesh, and Michael Berk. 2014. Affective and content analysis of online depression communities. *IEEE Transactions on Affective Computing*, 5(3).

Joanna Pawelczyk. 2011. *Talk as Therapy: Psychotherapy in a Linguistic Perspective*. Walter de Gruyter.

James W. Pennebaker, Matthias R. Mehl, and Kate G. Niederhoffer. 2003. Psychological aspects of natural language use: Our words, our selves. *Annual review of psychology*, 54(1):547–577.

James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of liwc2015. *UT Faculty/Researcher Works*.

James W. Pennebaker. 2011. *The secret life of pronouns: What our words say about us*. Bloomsbury Press, New York.

T. Pyszczynski and J. Greenberg. 1987. Self-regulatory perseveration and the depressive self-focusing style: A self-awareness theory of depression. *Psychological Bulletin*, 102:122–138.

Nairan Ramirezesparza, Cindy K. Chung, Ewa Kacewicz, and James W. Pennebaker. 2008. The psychology of word use in depression forums in english and in spanish: Testing two text analytic approaches. In *Proceedings of the ICWSM*.

Stephanie Rude, Eva-Maria Gortner, and James Pennebaker. 2004. Language use of depressed and depression-vulnerable college students. *Cognition and Emotion*, 18:1121–1133.

Thomas Ruder, Gary M Hatch, Garyfalia Ampanozi, and Nadja Fischer. 2011. Suicide announcement on facebook. *Crisis*, 32(5):280–282.

Gurjeet Singh, Facundo Memoli, Tigran Ishkhanov, Guillermo Sapiro, Gunnar Carlsson, and Dario L. Ringach. 2008. Topological analysis of population activity in visual cortex. *Journal of Vision*, 8(8):1–18.

Ann Sizemore, Chad Giusti, Ari Kahn, Richard F. Betzel, and Danielle S. Bassett. 2016. Cliques and cavities in the human connectome. *arXiv:1608.03520*.

S. W. Stirman and J. W. Pennebaker. 2001. Word use in the poetry of suicidal and non-suicidal poets. *Psychosomatic Medicine*, 63:517–522.

Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing high-dimensional data using t-sne. *Journal of Machine Learning Research*, 9:2579–2605, November.

M Villatte, JL Villatte, and SC Hayes. 2015. *Mastering the Clinical Conversation: Language as Intervention*. Guilford Publications, Palo Alto, CA, October.

Keane T. M. Weathers, F. W. and J. Davidson. 2001. Clinician-administered ptsd scale: A review of the first ten years of research. *Depression and Anxiety*, 13:132–156.

Myrna M. Weissman, Roger C. Bland, Glorisa J. Can-ino, Carlo Faravelli, Steven Greenwald, HaiGwo Hwu, Peter R. Joyce, Eile G. Karam, Chung-Kyoon Lee, Joseph Lellouch, Jean-Pierre Lepine, Stephen C. Newman, Maritza Rubio-Stipec, J. Elisabeth Wells, Priya J. Wickramaratne, Hans-Ulrich Wittchen, and Eng-Kung Yeh. 1996. Cross-national epidemiology of major depression and bi-polar disorder. *Journal of the American Medical Association (JAMA)*, 276(4):293—299.

Xiaojin Zhu. 2013. Persistent homology: An introduction and a new text representation for natural language processing. In Francesca Rossi, editor, *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, pages 1953–1959. IJCAI/AAAI.

P.G. Zimbardo and J.N. Boyd. 2008. *The Time Paradox: The new psychology of time that will change your life*. Free Press, New York.

Afra Joze Zomorodian. 2001. *Computing and comprehending topology: persistence and hierarchical Morse complexes*. Ph.D. thesis, University of Illinois at Urbana-Champaign.