Mapping Mental Health In Switzerland

Leonore Guillain **Othman Benchekroun** **Saskia Reiss**
{leonore.guillain, othman.benchekroun, saskia.reiss}@epfl.ch

Abstract

Social media provides a unique look into people's feelings and thoughts. Twitter, due to its relative anonymity, can provide a honest insight into how people deal with taboo topics such mental illness. As such, aggregated data from Twitter has previously been used to identify depression in users, among other mental illnesses, but also to measure on a smaller scale the public's opinion on this subject (Reavley and Pilkington, 2014).

Our project aims to perform a large scale analysis of the expression of mental distress in tweets posted in Switzerland. To achieve this goal, we iteratively build a set of search terms using manual inspection and topic modeling. This dictionary aims to identify mental illnesses and signs of mental distress in tweets. Our dataset is refined using this dictionary, allowing us to analyze the resulting set of tweets.

1 Introduction

Switzerland has one of the best mental health infrastructures, indicating that mental health is a real problem within the population. In fact, an estimated 18% of the Swiss population will experience at least a depression in their lives.

To understand this phenomenon using Data Science, we decided to turn to social media, a very helpful tool providing an 'in situ' approach into people's feelings and thoughts in almost perfect laboratory conditions, as it is used increasingly more by researchers trying to understand public health issues (cf Related Work section).

In addition, self-deprecation 'jokes' are becoming mainstream on social media because of the many people who relate to them. One of the many examples of this trend is a Buzzfeed article we stumbled upon during our research (Kopsky, 2017). This comforted us in our idea that, even though mental health issues are still taboo in society, social media would provide a better insight into this subject.

Through this project, we transpose past research into mental illness issues to Switzerland (which has never been done before) and broaden this subject past simple clinical depression detection by looking for various signs of mental distress in tweets.

2 Related work

Tweets are a popular subject of study in computational public health. We can identify two branches of research in this domain: one aiming at identifying the affected population with the greatest accuracy possible and one aiming at getting insights into how health topics are discussed on Twitter (Ghosh and Guha, 2013).

Mental health has been the subject of intense studies following the first branch of research but was never the subject of any research on its discussion on social media. Our project looks into ways of applying methods used in the second branch of research to topics of mental health, a feat that has not been documented before.

One of the challenges encountered when studying health topics on Twitter is the sparsity of health-related keywords compared to other topics (Reavley and Pilkington, 2014).

3 Data Collection

Our first dataset, *Spinn3r*, was retrieved from the cluster. The original corpus of files weighted about 30GB but after basic cleaning operations (which will be explained below), the filtered set only weighted 1GB, which meant it was small enough to handle further computations locally.

Our second dataset, *twitter-swisscom*, was provided as a '.zip' file. The original corpus weighted

about 5.6GB and later only 1.2GB after local filtering. We saved the cleaned file to '.csv' because it provides the lowest serialization cost from all commonly used formats. Once again, this allowed us to run all computations locally.

4 Methodology

The aim of our project is to use data from Twitter, primarily the textual content, in order to look at the expression of mental distress in tweets. Looking at the metadata, we want to see if any specific regional, temporal or gendered patterns can be found.

4.1 Data Cleaning

To do so, we start by cleaning the data from the Spinn3r dataset, which contains more metadata: we remove tweets containing websites and retweets, those containing punctuation marks and then set all tweets to lowercase. After that, we construct a Natural Language Processing (NLP) pipeline to extract information from the tweets and match the provided keywords by tokenizing the tweets and stemming the resulting words. These cleaned tokens are then matched against a search-term dictionary.

4.2 Dictionary Construction

The dictionary is constructed for the three most widely used languages on Twitter in Switzerland: French, German and English. Building this dictionary in a robust way, that is minimizing the number of false positives, is essential to our analysis as a large number of false positives would greatly impact our results. For this purpose, we build the dictionary iteratively.

We start by using a keyword dictionary defined in a previous research (Zaydman, 2017) to narrow down the tweets and identify those mentioning mental distress (cf Appendix A). After this, we look at a more extensive dictionary (Kale, 2015) serving a similar purpose (cf Appendix B).

We translate these dictionaries in order to extend our dictionary to French and German. Note that the translation was done by native speakers of each language respectively.

4.3 Tweet Labeling

Using these dictionaries, we construct an algorithm to tag all tweets. We then look at a sample of the tweets containing at least one keyword to identify potential issues with the tagging. This prompted us to change our dictionary but also to construct a *"negative"* dictionary used to exclude all tweets containing at least a blacklisted keyword. This negative keyword dictionary is built by using insights gained from manually annotating a sample of 1000 tweets and from an LDA model identifying topics that should be excluded, such as *assad, syria* and *trump*.

We note that for German and French, these translated dictionaries give us underwhelming results given the number of false positives the tagged-tweets set includes due to issues with NLP (it does not treat equivalent words in different languages the same way, leading us to refrain from analyzing French and German tweets). The new keywords set, that is now more accurate, will be addresses as MD tweets in the remaining of the paper, indicating that these tweets contain keywords referencing mental distress.

4.4 Final Set Analysis

We first analyze the distribution of keywords from our dictionary in these MD tweets along with their cooccurrences. We then answer our research questions using our newly-gained insights. Given the *"small"* amount of tweets contained in the *Spinn3r* dataset, we apply this methodology to the *swisscom-twitter* dataset and gain more insights.

5 Spinn3r Dataset Description

In order to have a better understanding of our data and make sure we provide sound answers to our research questions, we need to analyze our entire dataset to compare it to the final results.

5.1 General Information

We have a total of 3936084 tweets, but less than half of them have *geo_point* values and 25 do not provide gender information. However, all of them have an attached sentiment and language and sentiment attached to them.

5.2 Distribution of Categorical Data

First, we observe that almost half of the cleaned tweets are in English (49.29%) with French the next most popular language (29.69%). This is interesting as German is the the most widely used official language with 66% of the country living in German-speaking Switzerland.

We also see that most accounts (63.5%) do not contain information about the user's gender,

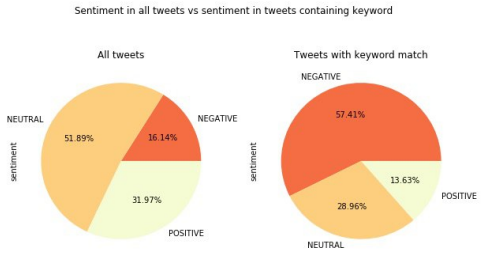

Figure 1: Sentiment distribution in tweets.

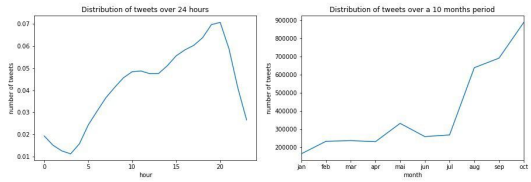

Figure 2: Daily and monthly tweets distribution.

meaning we cannot know if our set is unbiased. 21.53% of the remaining profiles belong to men while only 14.92% belong to women. Throughout the rest of the report, we will assume that the set is unbiased in order to look for any possible insights in the difference of mental distress expression between the genders.

We find that 73.51% of tweet are labeled as *'NEUTRAL'* by the Spinn3r algorithm while only 8.37% are labeled as *NEGATIVE* (cf Figure 1). This shows a clear bias on Twitter where negative topics are not discussed. This shows a possible hurdle for our analysis of mental distress expression in tweets, a clearly negative topic.

5.3 Distribution of Spatio-Temporal Data

Looking at the daily distribution of tweets (cf Figure 2.1), we see $8p.m.$ represents the highest number of posted tweets. This number gradually decreases through the night until $3a.m.$, the lowest posting time. Another low point is found around lunch time.

When looking at the number of tweets posted throughout the $10-$month period (cf Figure 2.2), we see that the dataset is not balanced, with the plausible explanation that the number of Twitter users is rapidly-increasing. This could also be linked to the *Spinn3r*'s tweets retrieval method.

As previously noted, only $\sim$ 40% of tweets are geo-localised; we will only focus on them in our geographic analysis. We see that some tweets are not located in Switzerland but we do not filter them out, as we make the assumption that the dataset is correct. By displaying the location data on a map (which can be found on the Notebook), we can see that tweets are strongly concentrated in urban centers, an expected result as this is where most of the Swiss population is located.

6 Main Methods and Algorithms

In this section, we explicit some of our algorithms that we felt were not detailed enough in the *Methodology* section, specifically our cleaning pipeline and our graph construction.

6.1 Data Processing

One of the reasons the Spinn3r dataset's weight was reduced so drastically is because of the cleaning first choice we made selecting only 11 of the columns provided in the dataset as most of them were useless for our project (such as *main_checksum* or *source_date_found*).

After working with the data, we reduce this number further as some variables do not provide any information (e.g. *source_spam_probability*, which is always equal to 0). We then filtered the tweets by language(only keeping *English*, *French* and *German*). Finally, we removed tweets containing URLS (highly linked to spam), lowered our strings and removed all non-alphanumerical characters.

Our NLP pipeline was also updated throughout the project. It first included tokenization, stop words removal and stemming (a basic NLP operation). However, stop words removal was dropped because of issues with our keyword matching and the analysis following it . For example, the word *myself* was removed, preventing us from including the expression *killing myself* in the dictionary.

6.2 Cooccurrence Graph

The implementation of cooccurrence graphs is here to prove that our findings constitute a real signal and not simply noisy random data. To construct our graphs, we used a circular plot layout where the nodes (cooccurring words) were linked between them if they were present in the same tweet (the number of cooccurrences being represented by the edge's width).

This was prompted by the huge difference between the 3 predominant keywords in the first English dictionary i.e. *depress*, *addict* and *suicide*, each occurring at least a 1000 times and the other

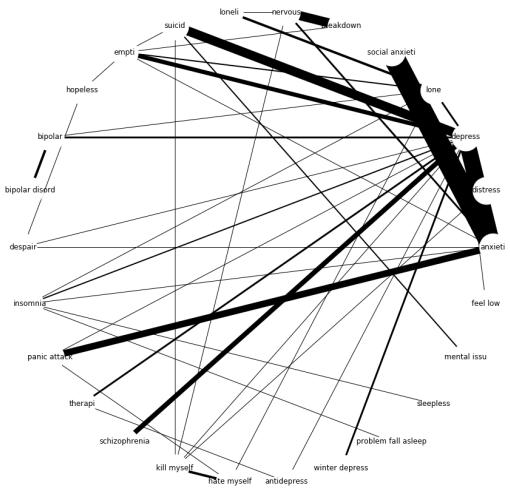

Figure 3: English keywords cooccurence graph.

keywords occurring less than 200 times each. However, using the small dictionary on other languages did not give any conclusive results.

The second dictionary gave us much better results both in English and in French (cf Figure 3). You can refer to the above figure to see the most frequent pairs (which also give us the most interesting insights). However, German cooccurrences were still inconclusive.

7 Results and Findings

7.1 Answering the Research Questions

We find that almost 5% of users tweet about mental distress. However, their discussions only represent $\sim\ 0.4\%$ of the dataset, corresponding to a total of 7152 tweets. We were not able to see clear linguistic patterns in these tweets because of our handicap using NLP methods which were not appropriate for our research. However, we noticed some convergence in the keywords distribution and their cooccurrences between French and English MD tweets.

Unfortunately, we cannot draw any further conclusions due to the overall sparsity of the data. We cannot analyze the gender difference they are too similar to get any statistically relevant results or the spatio-temporal information the time distribution is skewed and the location data only shows significant Twitter activity in 3 municipalities (i.e. *Geneva*, *Lausanne* and *Zurich*).

When trying to compare our results to national census data, we found that these numbers were also too sparse and approximative. This further shows the need for real research on mental health related issues in Switzerland and for the conversation to get going.

7.2 Discussion

As we can see from analysis, tweets showing mental distress are very sparse. For this reason, it very hard to get all the information we need in order to answer our research questions (for example determining seasonal or temporal patterns). Thus, it is not possible to conduct studies in Switzerland with the same scope of those presented in American papers, which is why we consider our results to offer more of a qualitative approach rather than purely quantitative results.

Another aspect of our study needs to be highlighted: the errors in the data we used which made us drop over $300'000$ tweets from the provided *.tsv* file. This prevents us from performing any relevant temporal analysis as the final number of tweets of this set is similar to the *Spinn3r* dataset. Thus, it can be considered even sparser as it spans over a $6-$year period instead of a $10-$months period.

Moreover, we noticed that the *Spinn3r* data labeling is not reliable. The more blatant example of inconsistencies was tweets showing clear mental distress being labeled as ***'POSITIVE'*** (e.g. "i love insomnia" or "i'm not suicidal"). Another disparity in the data was the presence of tweets which were not in Switzerland. Finally, another less problematic issue, was the weakness of the language labeling (presence of *Italian* tweets in the *French* labeled set).

8 Conclusion

It is possible to use dictionaries to study the expression of mental distress in tweets and to analyze their patterns. However, the sparsity of the data in Switzerland prevents us from performing a large scale research, not offering the possibility of drawing any statistically relevant conclusions. This issue can only be addressed by developing new NLP methods which can be applied to other languages than English (namely French, German and Italian) and by defining an original methodology which takes into account Switzerlands special situation.

References

Nicola J. Reavley and Pamela D. Pilkington. 2014. *Use of Twitter to monitor attitudes toward depression and schizophrenia: an exploratory study*. https://peerj.com/articles/647.pdf

Anna Kopsky. 2017. *23 Self-Deprecating Jokes That Are Too F***ing Real*. https://www.buzzfeed.com/annakopsky/by-hating-yourself-most

Debarchana Ghosh and Rajarshi Guha. 2013. *What are we tweeting about obesity? Mapping tweets with Topic Modeling and Geographic Information System*. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4128420/pdf/nihms-574479.pdf

Mikhail Zaydman. 2017. *Tweeting About Mental Health. Big Data Text Analysis of Twitter for Public Policy*. https://www.rand.org/content/dam/rand/pubs/rgs_dissertations/RGSD300/RGSD391/RAND_RGSD391.pdf

Sayali Shashikant Kale. 2015. *Tracking Mental Disorders Across Twitter Users*. https://getd.libs.uga.edu/pdfs/kale_sayali_s_201512_ms.pdf

Appendix

A First Dictionary

This dictionary contains the keywords: *mental health, depression, eating disorder, ptsd, mental illness, suicide, addiction, bipolar*.

B Second Dictionary

This dictionary contains the keywords: *depression, schizophrenia, bipolar, suicide, nervous, depress, distress, dejection, gloomy, cheerless, blue, empty, sad, insomnia, feeling low, hate myself, kill myself, dont want to live anymore, ashamed of myself, ashamed of myself, heat broken, feelings of worthlessness/guilt, lonely, loneliness, winter depression, SAD, seasonal affective disorders, antidepressants, pills for depression, bipolar disorder, pristiq, cymbalta, vilazodone, social anxiety, anxiety, worried, hopeless, despair*.