UFRJ

cores
Laboratório de
Computação Social e
Análise de Redes Sociais

Detection of Depression Symptoms using Social Media Data

Silas P. Lima Filho
silasfilho@ufrj.br

Jonice Oliveira
jonice@dcc.ufrj.br

Monica Ferreira da Silva
monica.silva@ppgi.ufrj.br

IBM Research

Programa Pós Graduação em Informática
Universidade Federal do Rio de Janeiro

Summary

Context

- Depression is one of the most reported mental diseases in the world. Sometimes called century illness[a] and the third lead cause of disability [5].
- Depression afflicts more certain groups than others, e.g. women to men, Europeans to Africans, people with more income.
- World Health Organization (WHO) presents that around 300 million people suffer from some level of depression[b].
- Health Ministry in Brazil presents that 11.5 million people are affected by depression.[c]

---

[a] www.theguardian.com/news/2018/jun/04/what-is-depression-and-why-is-it-rising
[b] www.who.int/en/news-room/fact-sheets/detail/mental-disorders
[c] www.blog.saude.gov.br/index.php/materias-especiais/52516-mais-de-onze-milhoes-de-brasileiros-tem-depressao

Context

- *Horvitz* report *infodemiology* as the use of digital information to inform the population about health policies, earlier epidemics identification and identify potentially affected individuals[4].
- *Lech & Eds* present two main challenges in *mental health informatics*. Provide health care services to remote and non-assisted populations, and turn health services more effective on cost[6].

Applications of Mental Health Informatics

- Telemedicine
- Automatizes Evaluation Systems
- Online Support and Information Management

Motivation

- People share with other users their social interests and preferences on Social Media Platforms.
- Plenty of data about behaviour, habits, interestings, friendship and so on.

Initial Research Question

Is it possible to identify psychological diseases symptoms, more specifically depression symptoms, from social media users content?

Justification

Academic Relevance

This work intend to contribute mainly on social network area, and somewhere between machine learning and recommendation systems. For studying if social media data is adequate to comprehend a person behaviour related to depression.

Practical Relevance

Prior contribution(academic) can improve how data is consumed by professionals in health service as psychologists and physicians, and also can auxiliate in order to construct more accessible health services to groups of people with less resources.

Methodology

Design Science Research (DSR) was selected due to its pragmatical approach for a given context. Based on [3], we can identify the three cycles in order to create at the end, an relevant artifact. Pimentel et al. [7] has presented an overview of different aspects of DSR from many authors. Moreover the authors suggest a framework to implement DSR in a research topic.

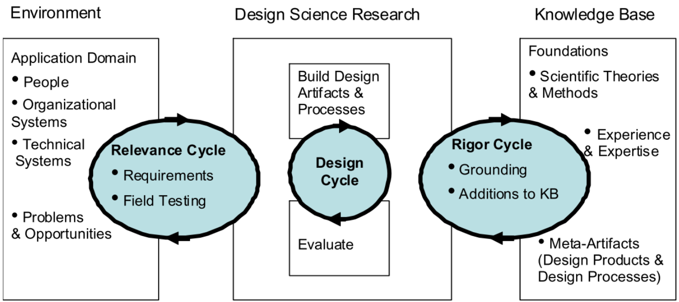

Figure: DSR three cycles Hevner[3].

Methodology

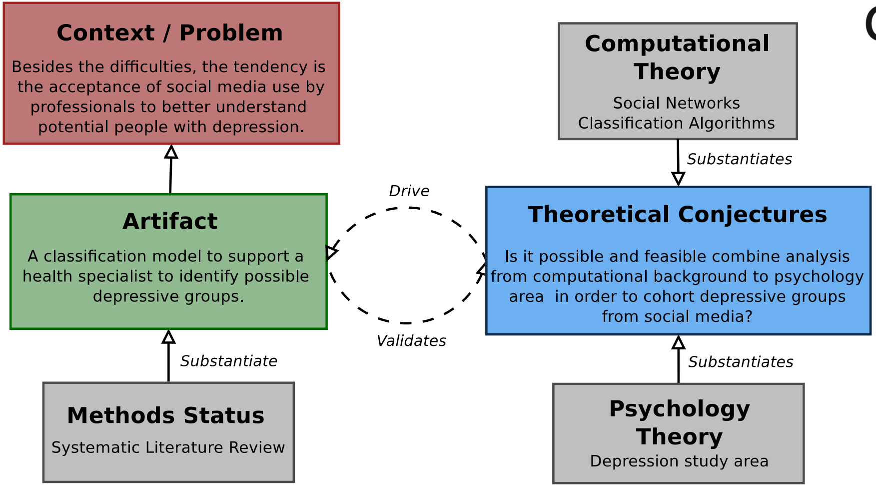

Figure: DSR instantiation.

Methodology

In order to fullfill the components and create the proposed artifact in DSR instatiation, we establish which cycles will comprehend each component.

- Systematic Literature Review
- Experiments to test and validate initial classification models.

Systematic Literature Review

Still in progress.

- (“Social Media” OR “Social Network” OR “Complex Network”) AND (Depression OR “Major Depressive Disorder”)
- $2013 \geq \textit{published} \leq 2018$
- 22 from ACM and 25 from IEEE bases

Inclusion	**Exclusion**
Directly tackles depression	Out of 2013-2018 scope
Have computational approach	Not written in english or portuguese
Attend both approaches	It is not a primary study
-	It does not have abstract
-	It does not have computing contribution
-	It has less than 4 pages

Table: SLR Criterias for inclusion and exclusion.

Systematic Literature Review

Chen et al. have detected eight basic emotions and calculated the overall intensity (strength score) of the emotions extracted from all past tweets of each user[1].

Approach:

- "...identify users with or at risk of depression by incorporating measures of eight basic emotions as features from Twitter posts over time, including a temporal analysis of these features."
- Data obtained from expressions: "I was/have been diagnosed with depression."
- Contributions:
  - Emotions are the analyzed features
  - Data analysis over time
  - Explore how time series analysis can help on task of identify depressive users
- Features:
  - Non-temporal: 9 entries emotion feature vector (emotions + Emotion Overall Score)
  - Temporal: Timestamp for each emotion score

Systematic Literature Review

Vedula and Parthasarathy conduct an observational study to understand the interactions between clinically depressed users and their ego-network when contrasted with a group of users without depression. They identify relevant linguistic and emotional signals from social media exchanges to detect symptomatic cues of depression[9].

Approach:

Examine network effects related to:

- participation (passive: tweets a user is exposed to, retweets or mentions a user receives; active: mentions, retweets and conversations made by the user)
- engagement (content (e.g., linguistic cues, emotion) and relational dynamics (e.g., conflict/support, influence))
- Ego-neighborhood (size, centrality and affinity to form clusters or communities)

Systematic Literature Review

De Choudhury et al. have used crowdsourcing to obtain data from twitter by people who were clinically diagnosed with depression. They also have constructed a corpus and developed a probabilistic model. The trained model classifies if a post indicates depression[2]. Similarly, *Tsugawa et al* have applied the same analysis to replicate the results in a group of japanese users[8].

- Problem:
  - Characterize levels of depression in populations.
- Approach:
  - Crowdsourcing to build a large corpus of postings on Twitter that have been shared by individuals diagnosed with clinical depression.
  - Probabilistic model trained on this corpus to determine if posts could indicate depression
  - Social media depression index that may serve to characterize levels of depression in populations

Indagations

- Not all the analyzed researches take into account the psychology point of view.
- Use of Psychometrics
- Would the screening of depressive users be more robust and reliable if they take others psychology approaches?
- Is it possible to identify the same symptoms from clinical using computational techniques?
- Aggregated Data Analysis vs Individual Data Analysis

Proposal

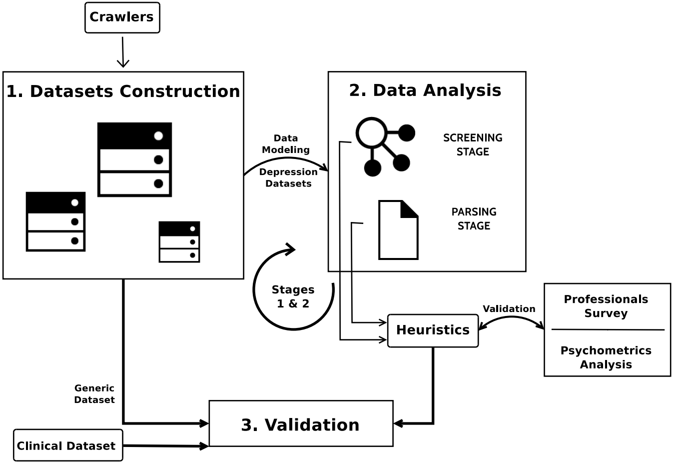

Figure: Proposal Conceptual Model

Problems

- Validation with clinical data
- Datasets to train classification model
- Reliable data, taking into account people consent
- Improve Computer Science area

Schedule

- Submission to WTD on SBBD (08/07)
- Proposal writing - In progress - First issue 15/07
- Qualification - End of August
- Ending of 1st Cycle -
- Ending of 2nd Cycle

Backup
Abstraction of Studies

1) Classificação dos trabalhos encontrados no mapeamento, ressaltando: Objetivo do trabalho, Abordagem utilizada, Técnicas utilizadas (algoritmos e métricas), Principais desafios levantados (ou trabalhos futuros).
2) A partir do material que vc. leu, destacar o conjunto de teorias/referências que vão embasar o seu problema e a sua motivação (Conjecturas Teóricas)
3) Quadro explicativo, com as métricas/features que os trabalhos relacionados usam e as métricas/features que vc. pensou em incluir na sua proposta. Este quadro deve ter: o nome da métrica/feature, explicação sucinta sobre a métrica/feature, referência (ao trabalho que a usou), justificativa do seu uso.

Reference

[1] CHEN, X., SYKORA, M. D., JACKSON, T. W., AND ELAYAN, S.
What about Mood Swings.
In *Companion of the The Web Conference 2018 on The Web Conference 2018 - WWW '18* (New York, New York, USA, 2018), ACM Press, pp. 1653–1660.

[2] DE CHOUDHURY, M., COUNTS, S., AND HORVITZ, E.
Major Life Changes and Behavioral Markers in Social Media: Case of Childbirth.
2013.

[3] HEVNER ALAN, R.
A Three Cycle View of Design Science Research.
Scandinavian Journal of Information Systems 19, 2 (2007), 87–92.

[4] HORVITZ, E., AND MULLIGAN, D.
Data, privacy, and the greater good.
Science 349, 6245 (jul 2015), 253–255.

[5] JAMES, S. L., ABATE, D., ABATE, K. H., ABAY, S. M., ABBAFATI, C., ABBASI, N., ABBASTABAR, H., ABD-ALLAH, F., ABDELA, J., ABDELALIM, A., ABDOLLAHPOUR, I., ABDULKADER, R. S., ABEBE, Z., ABERA, S. F., ABIL, O. Z., ABRAHA, H. N., ABU-RADDAD, L. J., ABU-RMEILEH, N. M. E., ACCROMBESSI, M. M. K., AND ACHARYA, D.
Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the global burden of disease study 2017.
The Lancet 392, 10159 (2018), 1789 – 1858.

[6] LECH, M., LOW, L.-S., AND OOI, K. E.
Detection and prediction of clinical depression.
In *Mental Health Informatics*. Springer, 2014, pp. 185–199.

[7] PIMENTEL, M., FILIPPO, D., AND MARIA SANTORO, F.
Design Science Research: fazendo pesquisas científicas rigorosas atreladas ao desenvolvimento de artefatos computacionais projetados para a educação.
29–43.

[8] TSUGAWA, S., KIKUCHI, Y., KISHINO, F., NAKAJIMA, K., ITOH, Y., AND OHSAKI, H.
Recognizing Depression from Twitter Activity.
Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems - CHI '15 (2015), 3187–3196.

[9] VEDULA, N., AND PARTHASARATHY, S.
Emotional and Linguistic Cues of Depression from Social Media.
Proceedings of the 2017 International Conference on Digital Health - DH '17 (2017), 127–136.