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

Programa de Pós-Graduação em Informática, UFRJ

Summary

- Context/Problem
- Initial Research Questions
- Methodology
- 1st Cycle
  - Systematic Literature Review
  - Literature Gaps
- 2nd Cycle
- Next Steps
- Limitations

Context/Problem

- Depression is one of the most reported mental diseases in the world. Sometimes called century illness a and the third lead cause of disability.
- World Health Organization (WHO) presents that around 300 mi people suffer from some level of depression.
- Health Ministry in Brazil presents that 11.5 million people are affected by depression.

Context/Problem

- Horvitz report infodemiology as the use of digital information to inform the population about health policies, earlier epidemics identification and identify potentially affected individuals.

- 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[22].

Context/Problem

- People share with other users their social interests and preferences on Social Media.

- Plenty of data about behaviour, habits, interests, friendship and so on.

Initial Research Questions

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

- The collected information from social media is sufficiently robust to determine if a user has depression or its symptoms?

- Which computational methods and efforts were created or used to understand emotional behavior from a social media user?

Methodology

Design Science Research (DSR) was selected due to its pragmatical approach for a given context. Based on [15], we can identify the three cycles in order to create at the end, an relevant artifact. *Pimentel et al.* [31] 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.

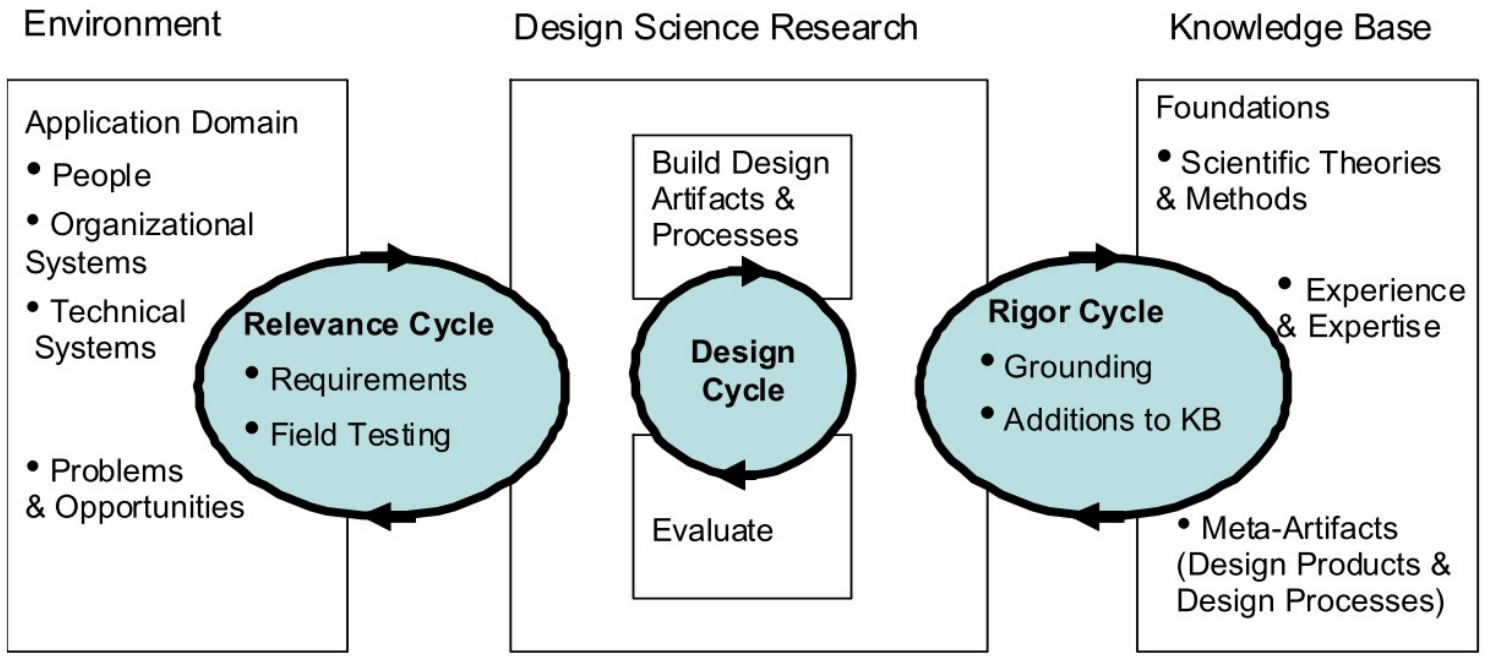

Figure: DSR Cycles by Hevner.

Methodology

Peffers states the research process in six stages. The author specifically proposes a Design Science Research Methodology (DSRM).

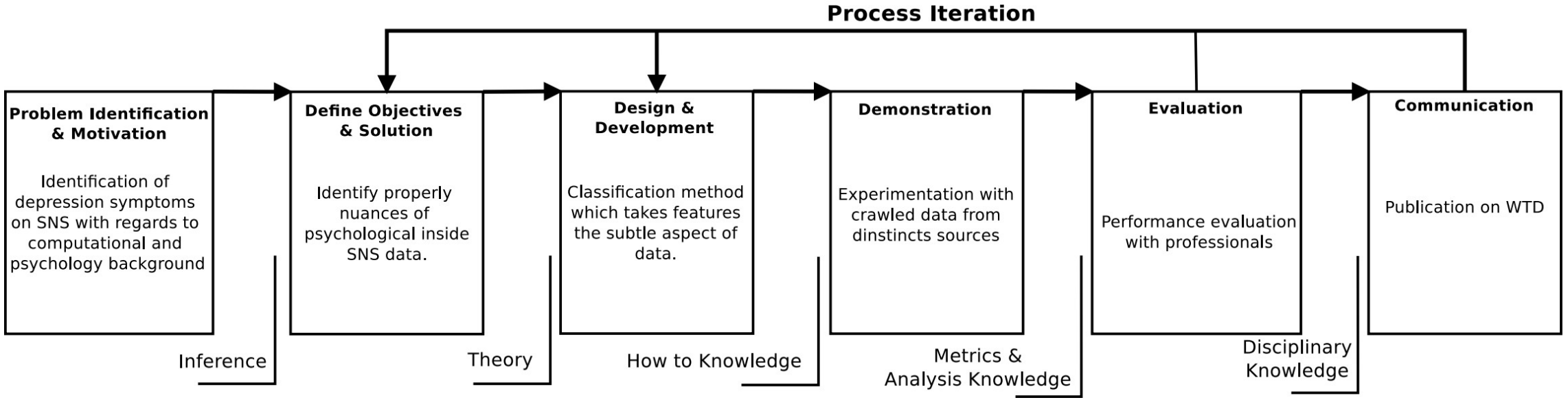

Figure: Research stages based on DSRM from *Peffers*.

Methodology

We establish which cycles will comprehend each component to fullfill DSR requirements. Taking into account Wieringa approach [40]:

- Systematic Literature Review (SLR) - Related to Empirical Cycle
- Experiments to test and validate initial classification models - Related to Design/Engineering Cycle.

1st Cycle
Systematic Literature Review

Still in progress.

- (“Social Media” OR “Social Network” OR “Complex Network”) AND (Depression OR “Major Depressive Disorder”)
- 2013 ≥ published ≤ 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 CS contribution
-	It has less than 4 pages

	Textual				Content Metadata	User Profile	Interactions Data			Group Charac.	Long. Variation
Reference	**Emotion /Sent.**	**Linguis.**	**Topic**	**Other**			**Platform**	**Other Users**	**Both**		
Fang, 2014	x										
Nguyen, 2014	x	x	x								
Larsen, 2015	x	x									
Nambisan, 2015							X				
Chomutare, 2015			x	x							
Dao, 2015	x		x								
Dao, 2016	x						x				
Dao, 2016	x	x	x								
Saha, 2016	x	x	x								
Akay, 2016	x	x						x		x	
Simms, 2017	x	x									
Oyong, 2018		x		x							
Katchapakirin, 2018	x	x			x	x			x		
Silveira, 2018				x					x	x	
Wongkoblap, 2018	x	x			x	x			x		
Trotzek, 2018	x	x			x						

	Textual				Content Metadata	User Profile	Interactions Data			Group Charac.	Long. Variation
Reference	**Emotion /Sent.**	**Linguis.**	**Topic**	**Other**			**Platform**	**Other Users**	**Both**		
DeChoudhury, 2013	X	X							X		
DeChoudhury, 2013	X	X		X	X				X		
DeChoudhury, 2013	X	X					X				X
Homan, 2014								X		X	
Wilson, 2014		X									
Tsugawa, 2015	X		X	X	X				X		
Kavuluru, 2016	X			X							
DeChoudhury, 2017	X	X	X								
Vedula, 2017	X	X							X	X	
Yazdavar, 2017	X	X	X								X
Bagroy, 2017				X							X
Nobles, 2018	X	X		X							
Chen, 2018	X										X
Sadeque, 2018		X		X							
Zhao, 2018				X	X						

Questions from SLR

- Not all the analyzed researches take into account the psychology point of view. (11 papers)
- 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?
- Few amount of papers which analyse the value of features variation over time.

2nd Cycle
Experimental Aproach

- Develop an initial classification model based
- Use of different types of data in identification task
- Analyze variation of data over time

2nd Cycle Experimental Aproach

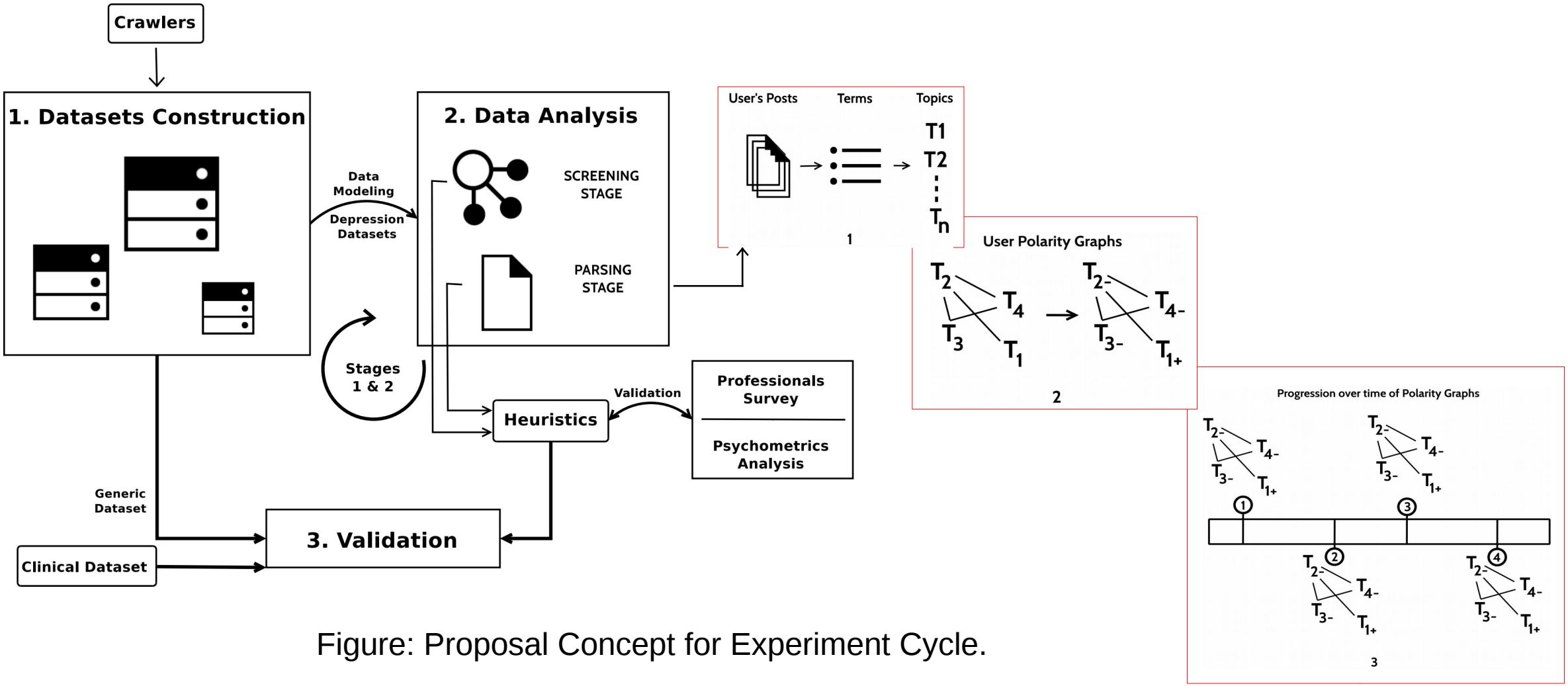

Figure: Proposal Concept for Experiment Cycle.

Next Steps

- Rerun and validate 1st Cycle with one more source.
- Rerun and validate 2nd Cycle.
- Qualification – This year yet.
- Measure implications of COVID-19 in mental health using Social Media Data. CAPES pandemics project.

Limitations – Work in progress

- Metrics and algorithms are not well defined yet.
- Time limitation. Probably will be extended.
- Problem still not clear yet. Although the longitudinal analysis is not well covered by literature.
- Ethical Aspects of Experiments
- It is not a diagnosis solution.

Limitations – Work in progress

We have at the moment a dataset with around 1.5 GB of collected information. The first dataset is related to the depression forum from HealingWell website. We have collected 3075 posted topics andtheir respective posts summing a total of 18450 replying posts.

Datasets from Reddit and HeallingWell forum.

Backup