Utilizing Crowd Intelligence for Online Detection of Emotional Distress

Master's Thesis Presentation

Siddhant Goel
Advisor: Han Xiao
Supervisor: Prof. Dr. Claudia Eckert

Chair for IT Security
Technische Universität München

March 11, 2013

Outline

- Introduction
- Problem Definition
- Theoretical Background
  - Text Classification
  - Support Vector Machines
  - Ensemble Learning methods
- Experiments
- Results
- Conclusion

Introduction

- Millions of people die every year because of suicide
- Most people are between 15 to 29 years old
- Rise of social media - Twitter, Facebook, Reddit, Wordpress
- Reddit - “/r/happy” and “/r/suicidewatch”
- People are not afraid of posting their inner feelings on the web

Introduction

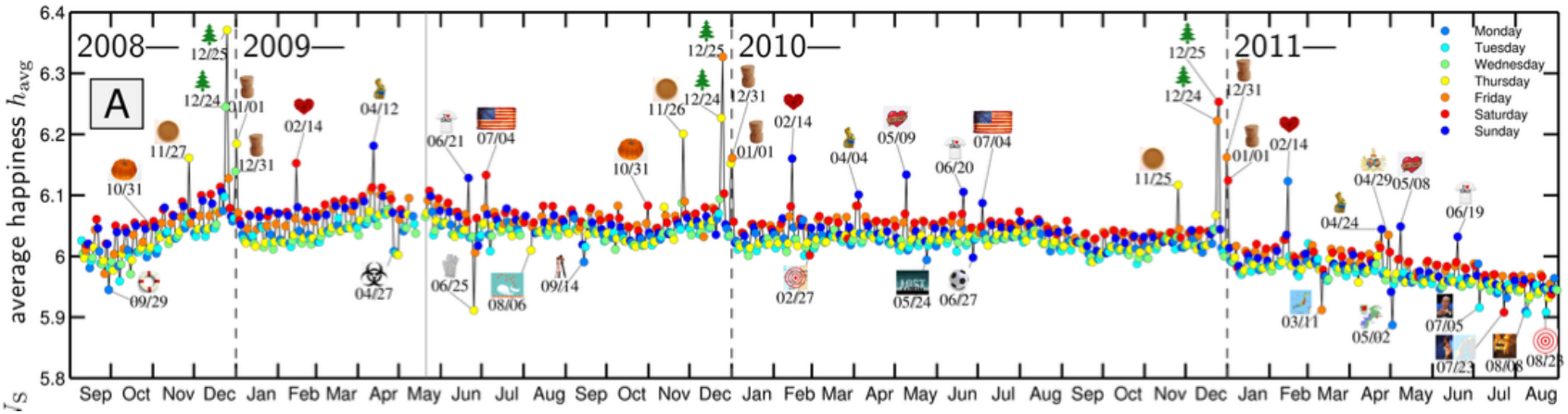

Figure : Happiness on Twitter as a function of time

- Study conducted in 2011
- 46 billion words collected over 33 months
- Negativity on Twitter has been on the rise
- Words include *death*, *hate*, and even *suicide*

Motivation

- Last tweet of Twitter user "@CapitalSTEEZ_" [1]
- Some accounts have lots of followers, some don't
- Lives can be saved if there is a surveillance system of suicide
- Public sentiment information available on the web + No analysis possible = Disconnect

[1] http://twitter.com/CapitalSteez_

Problem Definition

- Evaluate machine learning algorithms (including support vector machines and ensemble learning methods) that can be used for text classification
- Build a web based system that can
  - tap into crowd intelligence to incrementally improve the classifiers
  - detect content on the web that indicates that its author is depressed or suicidal