Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 113 (2017) 65–72

Procedia Computer Science

www.elsevier.com/locate/procedia

The 8th International Conference on Emerging Ubiquitous Systems and Pervasive Networks (EUSPN 2017)

Machine Learning and Semantic Sentiment Analysis based Algorithms for Suicide Sentiment Prediction in Social Networks

Marouane Birjali[a,*], Abderrahim Beni-Hssane[a], Mohammed Erritali[b]

[a]*LAROSERI Laboratory, Department of Computer Sciences, University of Chouaib Doukkali, Faculty of Sciences, El Jadida, Morocco*
[b]*TIAD Laboratory, University of Sultan Moulay Slimane, Faculty of Sciences and Technologies, Béni Mellal, Morocco*

Abstract

Sentiment analysis is one of the new challenges appeared in automatic language processing with the advent of social networks. Taking advantage of the amount of information is now available, research and industry have sought ways to automatically analyze sentiments and user opinions expressed in social networks. In this paper, we place ourselves in a difficult context, on the sentiments that could thinking of suicide. In particular, we propose to address the lack of terminological resources related to suicide by a method of constructing a vocabulary associated with suicide. We then propose, for a better analysis, to investigate Weka as a tool of data mining based on machine learning algorithms that can extract useful information from Twitter data collected by Twitter4J. Therefore, an algorithm of computing semantic analysis between tweets in training set and tweets in data set based on WordNet is proposed. Experimental results demonstrate that our method based on machine learning algorithms and semantic sentiment analysis can extract predictions of suicidal ideation using Twitter Data. In addition, this work verify the effectiveness of performance in term of accuracy and precision on semantic sentiment analysis that could thinking of suicide.

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Sentiment Analysis; Machine Learning; Suicide; Social Networks; Tweets; Semantic Sentiment Analysis.

1. Introduction

Nowadays, social networks have changed the way that people express their opinions and points of view[1]. This opportunity is given through textual publications, online discussion sites, product evaluation websites etc. People rely

* Corresponding author.
E-mail address: birjali.marouane@gmail.com

1877-0509
Peer-review under responsibility of the Conference Program Chairs.
10.1016/j.procs.2017.08.290

heavily on this user-generated content. Social networks offer considerable amount of content generated by the user, it is important content for analysis and offer more services adapted to the needs of users[1,2]. In recent years, the majority of developments in the field of information and opinion exchange have launched the research work for the analysis of feelings expressed on these social networks, presented in the literature as *sentiment analysis*[3].

The analysis of sentiments uses, among other things, the detection of opinions on social networks, clarifying consumer behavior, recommending products and explaining the outcome of the elections. It consists of searching for evaluative texts on the Internet such as criticisms, recommendations and analyzing the feelings expressed therein in an automatic or manual way, in order to understand public opinion[3]. Millions of people are using Twitter and ranked as one of the most visited sites with the average of 58 million tweets per day[2].

However, social networks such as Facebook, Twitter and Google+ are increasingly associated with many social phenomena such as harassment, intimidation, depression[4] or even suicide[5]. This is why it is essential to try to detect potential victims as soon as possible in order to reinforce the prevention of suicide using social networks[5]. In fact, we can cite the case of two American rappers Capital Steez[7] and Freddy E[6] who receive death in the course of commenting live their actions on their social networks accounts.

This research focuses on suicide. Because suicide is simply one of the 20 leading causes of death in the world[8]. Already the pronunciation of suicide word should not be taken with simplicity and lightly. It can be the last cry of someone's help, and yet if the signs and clues are recognized at the beginning, lives could be saved. Suicide is preventable, it is an act of those who have not been able to accomplish others and the prevention of suicide should be the responsibility of everyone.

Recent research studies have demonstrated that social networks can be used to extract feelings of depression[4], which makes it easier for us to prevent suicide online[5]. The content addressed and the phrases used by depression, discouragement and suicides are well known and available[4]. In case of example, these users of social networks are often victims of cyberbullying or sexual harassment. In the case of Twitter, it would set up a real time monitoring in relation to several risk factors[2,9].

On the other hand, sentiment analysis could lead to several challenges like semantic sentiment analysis which is evaluate and implement a new semantic similarity metric to determine the affective content of a word in different dimensions[10]. We have proposed a distance metric based on WordNet to determine the semantic orientation of social networks data[11]. The purpose of this paper is to propose a method of predicting suicidal ideas, to predict suicidal acts and ideas using data collected from social. In this work, we use Weka as a data mining tool to extract all useful information for the classification of this data according to the machine learning algorithms implemented in Weka. Therefore, we present our algorithm to calculate the semantic similarity between the tweets collected from Twitter in the training set based on a semantic analysis resource using WordNet.

The paper is organized as follows: Section 2 presents related works. Proposed method of Suicidal acts analysis is in section 3, while presenting the collection of data from social networks based on the vocabulary associated with suicide and presenting the proposed algorithm. Section 4 gives experimental results, conclusion and future works are given in Section 5.

2. Related Works

Social networks have attracted the attention of researchers, which attempts to understand and analyze, among others, the structure of interconnection and the interaction of users in social networks. Internet users tend to express their opinions and feelings and talk about their lives and activities of everyday life via Twitter[3,9,11]. The application of machine learning methods for the identification of suicide has increased in recent years. The LIWC2007 (Linguistic Inquiry and Word Count Version 2007) application judges evaluate emotional and cognitive words and expressions in written and oral sentences of individuals. Noting that the use of LIWC2007 could identify and trend an emotional post[12]. The work of Ramirez-Esparza, et al.[13] focused on the idea of language markers and on the discussion of feeling depression by collecting information from both already depressed people and undepressed forums while using Bulletin Board Systems (BBS). They also explain that depressed people who have written in English are more likely to report medical problems.

Sentiment analysis is treated as a task of natural language processing at several levels of granularity. There has been a significant amount of research on feeling analysis, rule-based approaches, from bag-of-words to machine learning