


Chapter 32

Sentiment Analysis Using Machine Learning Algorithms and Text Mining to Detect Symptoms of Mental Difficulties Over Social Media

Hadj Ahmed Bouarara

 <https://orcid.org/0000-0002-4973-4385>

GeCoDe Laboratory, Algeria

ABSTRACT

A recent British study of people between the ages of 14 and 35 has shown that social media has a negative impact on mental health. The purpose of the paper is to detect people with mental disorders' behavior in social media in order to help Twitter users in overcoming their mental health problems such as anxiety, phobia, depression, paranoia, etc. For this, the author used text mining and machine learning algorithms (naïve Bayes, k-nearest neighbours) to analyse tweets. The obtained results were validated using different evaluation measures such as f-measure, recall, precision, entropy, etc.

INTRODUCTION AND PROBLEMATIC

Instagram, Twitter, Facebook and Snapchat: these platforms attract the attention of 91% of 16–24-year-olds. Between narcissism and harassment, creativity and self-expression, social networks are at the origin of a social revolution, especially among “millennials” (born between 95 and the early 2000s). Unfortunately, the conclusion of StatusOfMind is that: social networks are, for the most part, bad for the morale of its young users. Thus, twitter is considered to be the most harmful followed closely by Snapchat, then Facebook and finally Instagram (Koenig & McLaughlin, 2018).

DOI: 10.4018/978-1-6684-6303-1.ch032

According to a study by the public health foundation (Mental Health Foundation, 2015), the rate of anxiety and depression has jumped 70% among young people in the last 25 years when it correlated these figures with the increased use of social media. The study established a list of the negative consequences of social networks: cyber-harassment, Addiction (or the feeling of anxiety about missing something), Anxiety, depression, the feeling of loneliness, lack of sleep, physical ill-being. more than one in two (55%) say they have been embarrassed in their daily life by “symptoms of mental difficulty” (anxiety, phobia, depression, paranoia). Even more worrying: one in five young people (22%) say they have felt this discomfort significantly (Barr et al., 2015).

In recent years, we are in a digital world where information is available in large quantities and in various forms. 80% of this mass of information was in textual form. For this reason, we need specific tools to access sentiments and meanings hidden in these data, in order to reduce human intervention. Opinion Mining groups methods and techniques for identifying opinions in textual data.

In this paper we are interested in a subdomain of opinion mining which is the contextual analysis of social networks which attracted many researchers, which gave birth to many works. In our work we are based on a corpus (Corpus-based Approach) principle of assigning data to a classifier for learning in a supervised way, which generates a model that is used for the test part. We used twitter as a social network which is a networking site allows users to write short articles, called “tweets”. The content of this paper discusses the different steps of our proposed approach to solve the problem of detecting depressed people through a decision analysis of tweets. Then we will define the tools used for the realization of the practical part of our work with a general presentation of the results obtained by discussing the different comparisons applied between the different techniques used and proposed during our work.

The general structure of the paper will be as follows: we start with a state of the art for presenting the essential works in this topic, after we go on with a section detailing our approach and proposed components then an experimental and comparative study will be carried out for presenting the best results obtained. Finally, we will finish with a conclusion and describing some lines of thought that remain open and that we want to share them with you.

LITERATURE REVIEW (RELATED WORK)

The work of Hatzivassiloglou and McKeown in 1997 (1997) consists in using the coordinating conjunctions present between a word already classified and an unclassified word, followed by the contributions of researcher Nasukawa and his team in 2003 (2003) who proposed a new method for extracting associated concepts from segments and summing the orientations of the opinion vocabulary present in the same segment.

In the same year, researchers Yu and Hatzivassiloglou (2003) used the probability of ranking a word to measure the strength of the orientation of the named entities. In 2006, researchers Kanayama and Nasukawa (2006) as well as Ding and Liu (2008) in 2008 proposed, for their part, a learning-based approach that uses the coordination conjunctions present between a word already classified and a word unclassified.

The approaches of Pang et al introduced in 2002 (2002), and that of Charton and Acuna-Agost published in 2007 (2007) consist of classifying the texts according to a global polarity (positive, negative and neutral). These methods were optimized by Wilson and his research team in 2005 (2005). However, the difficulty lies in the constitution of these corpora of learning, which is a manual process to per-

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage:
www.igi-global.com/chapter/sentiment-analysis-using-machine-learning-algorithms-and-text-mining-to-detect-symptoms-of-mental-difficulties-over-social-media/308509

Related Content

Productivity Analysis of Public Services: An Application of Data Mining

Aki Jääskeläinen, Paula Kujansivu and Jaani Väisänen (2010). *Data Mining in Public and Private Sectors: Organizational and Government Applications* (pp. 83-105).

www.irma-international.org/chapter/productivity-analysis-public-services/44284

Analytical Processing Over XML and XLink

Paulo Caetano da Silva, Valéria Cesário Times, Ricardo Rodrigues Ciferri and Cristina Dutra de Aguiar Ciferri (2012). *International Journal of Data Warehousing and Mining* (pp. 52-92).

www.irma-international.org/article/analytical-processing-over-xml-xlink/61424

Skeleton Network Extraction and Analysis on Bicycle Sharing Networks

Kanokwan Malang, Shuliang Wang, Yuanyuan Lv and Aniwat Phaphuangwittayakul (2020). *International Journal of Data Warehousing and Mining* (pp. 146-167).

www.irma-international.org/article/skeleton-network-extraction-and-analysis-on-bicycle-sharing-networks/256167

A Comprehensive Workflow for Enhancing Business Bankruptcy Prediction

Rui Sarmiento, Luís Trigo and Liliana Fonseca (2015). *Integration of Data Mining in Business Intelligence Systems* (pp. 216-238).

www.irma-international.org/chapter/a-comprehensive-workflow-for-enhancing-business-bankruptcy-prediction/116817

Vertical Fragmentation in Databases Using Data-Mining Technique

Gorla Narasimhaiah and Wing Yan Betty Pang (2010). *Strategic Advancements in Utilizing Data Mining and Warehousing Technologies: New Concepts and Developments* (pp. 178-197).

www.irma-international.org/chapter/vertical-fragmentation-databases-using-data/40404