

Chapter 20

A Probabilistic Deep Learning Approach for Twitter Sentiment Analysis

Mostefai Abdelkader

Dr. Tahar Moulay University of Saida, Algeria

ABSTRACT

In recent years, increasing attention is being paid to sentiment analysis on microblogging platforms such as Twitter. Sentiment analysis refers to the task of detecting whether a textual item (e.g., a tweet) contains an opinion about a topic. This paper proposes a probabilistic deep learning approach for sentiments analysis. The deep learning model used is a convolutional neural network (CNN). The main contribution of this approach is a new probabilistic representation of the text to be fed as input to the CNN. This representation is a matrix that stores for each word composing the message the probability that it belongs to a positive class and the probability that it belongs to a negative class. The proposed approach is evaluated on four well-known datasets HCR, OMD, STS-gold, and a dataset provided by the SemEval-2017 Workshop. The results of the experiments show that the proposed approach competes with the state-of-the-art sentiment analyzers and has the potential to detect sentiments from textual data in an effective manner.

1. INTRODUCTION

It is very important to business industries and organizations to know opinions, thoughts, emotions, or sentiments expressed in texts by peoples using microblogging platforms such as twitter since it allows better decision making (Giachanou et al. 2016). This activity is commonly referenced as sentiment analysis (i.e., opinion mining) and it is an active area of research in recent years.

Sentiment analysis is a research field in the area of text mining and natural language processing that aims at analyzing texts that express opinions, sentiments and emotions of people's about a topic (e.g., services, individuals, products, issues) (Giachanou et al. 2016; Abbasi et al. 2008; Nasukawa & Yi, 2003).

DOI: 10.4018/978-1-6684-6303-1.ch020

If the analyzed text holds a sentiment, then it is viewed as polar (positive or negative), otherwise, it is viewed neutral (Zimbra, 2018). The analysis process consists of automatically predicting such polarity.

Nowadays, large textual data expressing people's opinions and thoughts on products, services, or any topic can be found on social media platforms such as twitter which is a widely known microblogging platform. Thus it is not necessary to get them using traditional approaches such as surveys that are costly processes.

This situation constitutes a valuable opportunity for organizations to know about the quality of their services and products by analyzing user's texts to infer their opinions (Zimbra et al. 2015, Forman et al. 2008).

To organizations and business industries, tweets are viewed as the medium that contains evaluations of business and society works (Jansen et al. 2009; Gleason, 2013). For example, Organizations conduct sentiment analysis processes to study product sales (Rui et al. 2013), stock market movements (Bollen et al. 2011), and so on.

Unfortunately, While it is very easy to collect textual data that express user's opinions on a topic on twitter, It still is very difficult to analyze them to get information about people's sentiments. The achieved accuracies by the proposed techniques were found under 70%, which limits their applicability in real situations (Zimbra, 2018; Minghui et al. 2020; Kiritchenko et al. 2014; Santos et al.2014; Hassan et al. 2017). The difficulty Sources are tweet length, the used abbreviations, the problem of spelling errors, and the existence of special characters(Hassan et al. 2017; Kiritchenko et al. 2014).

So far, many approaches have been proposed to infer sentiments from textual data. These approaches are classified into three classes machine-learning approaches that include deep learning one, Lexicon-based and hybrid (Gupta & Joshi, 2019; Musto et al. 2014; Zimbra, 2018; Giachanou et al. 2016; Turney, 2002; Kim & Hovy, 2004; Tang et al. 2015).

Deep learning approaches which are class of neural networks with multi-hidden layers such as a convolutional neural network (CNN) have been found effective in many areas such as computer vision, natural language processing(Graves et al. 2013; Kalchbrenner et al. 2014). Due to this success, many deep learning approaches have been proposed to replicate this success in the sentiment analysis domain (Zhang et al. 2018). The performance of these deep learning models is dependent on many hyperparameters and an important one of them is the choice of the word embeddings model which is a vector representation of data (Bengio et al. 2003; Zhang et al. 2018; Kim, 2014).

This paper proposes a probabilistic deep learning approach for sentiments analysis. The deep learning model used is a Convolutional Neural Network (CNN). The main contribution of this work is a new probabilistic representation of the textual data (tweet) to be fed to the CNN. This representation is a matrix composed of two vectors. The first vector stores for each word composing the message the probability that it belongs to a positive class. The second one stores for each word composing the message, the probability that it belongs to a negative class. The system is evaluated on four well-known datasets: HCR OMD and STS-gold, and a dataset provided by the SemEval-2017 Workshop(subtask B). The results of the experiments show that the proposed approach can compete with the state of the art tools for sentiment analysis.

The rest of the paper is organized as follow: section2 introduces the deep learning and convolutional neural networks, section 3 presents Related Works, section 4 presents the approach proposed for SA, section 5 presents the empirical study, presents and analyzes the obtained results and section 6 concludes and presents future works.

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/a-probabilistic-deep-learning-approach-for-twitter-sentiment-analysis/308497

Related Content

Data Stream Mining Using Ensemble Classifier: A Collaborative Approach of Classifiers

Snehlata Sewakdas Dongre and Latesh G. Malik (2017). *Collaborative Filtering Using Data Mining and Analysis* (pp. 236-249).

www.irma-international.org/chapter/data-stream-mining-using-ensemble-classifier/159506

The Dynamics of Content Popularity in Social Media

Symeon Papadopoulos, Athena Vakali and Ioannis Kompatsiaris (2010). *International Journal of Data Warehousing and Mining* (pp. 20-37).

www.irma-international.org/article/dynamics-content-popularity-social-media/38952

Lossless Reduction of Datacubes using Partitions

Alain Casali, Sébastien Nedjar, Rosine Cicchetti, Lotfi Lakhal and Noël Novelli (2009). *International Journal of Data Warehousing and Mining* (pp. 18-35).

www.irma-international.org/article/lossless-reduction-datacubes-using-partitions/1821

Updating the Built Prelarge Fast Updated Sequential Pattern Trees with Sequence Modification

Jerry Chun-Wei Lin, Wensheng Gan, Tzung-Pei Hong and Jingliang Zhang (2015). *International Journal of Data Warehousing and Mining* (pp. 1-22).

www.irma-international.org/article/updating-the-built-prelarge-fast-updated-sequential-pattern-trees-with-sequence-modification/122513

A Multi-Agent Neural Network System for Web Text Mining

Lean Yu, Shouyang Wang and Kin Keung Lai (2008). *Emerging Technologies of Text Mining: Techniques and Applications* (pp. 162-183).

www.irma-international.org/chapter/multi-agent-neural-network-system/10181