

Chapter 9

A New SVM Method for Recognizing Polarity of Sentiments in Twitter

Sanjiban Sekhar Roy
VIT University, India

Rohan Kumar
VIT University, India

Marenglen Biba
University of New York – Tirana, Albania

Rahul Kumar
VIT University, India

Pijush Samui
NIT Patna, India

ABSTRACT

Online social networking platforms, such as Weblogs, micro blogs, and social networks are intensively being utilized daily to express individual's thinking. This permits scientists to collect huge amounts of data and extract significant knowledge regarding the sentiments of a large number of people at a scale that was essentially impractical a couple of years back. Therefore, these days, sentiment analysis has the potential to learn sentiments towards persons, object and occasions. Twitter has increasingly become a significant social networking platform where people post messages of up to 140 characters known as 'Tweets'. Tweets have become the preferred medium for the marketing sector as users can instantly indicate customer success or indicate public relations disaster far more quickly than a web page or traditional media does. In this paper, we have analyzed twitter data and have predicted positive and negative tweets with high accuracy rate using support vector machine (SVM).

INTRODUCTION

There has been a colossal surge in user generated opinion-rich data since the Web 2.0 era. Millions of people post opinions on all aspects of life every day. Social networks such as Facebook, LinkedIn and the microblogging website Twitters are the mainstays of this massive, continuous stream of user-generated content on a wide range of topics. Easy availability of such data via APIs or publicly available data

DOI: 10.4018/978-1-5225-2128-0.ch009

sets has led to new opportunities in the field of Sentiment Analysis and Opinion Mining. Sentiment Analysis is the process of computationally identifying and extracting opinions in natural language texts. Furthermore, known as Opinion Mining, Sentiment analysis uses Natural Language Processing and text data mining (text analysis) to understand opinions expressed in the text. It is not only applied to social media data but also to product reviews, news articles and other public opinion texts. Given the direct applicability to real-world problems such as understanding customer opinion, financial prediction and disaster management, it is imperative that Sentiment Analysis prevails as one of the most researched topics today. There are three different classification levels in sentiment analysis: document level, sentence level and aspect level. The character limitation on Twitter posts may suggest sentence-level classification as the most suited classification; however, given the informal nature of tweets, we expect the sentiment to be compact and explicit. Hence, our paper focuses on the document-level approach to Sentiment Analysis. Twitter is considered one of the best sources of opinion-rich data because a large number of people share and discuss their thoughts and opinions on the platform. That is to say, we focus on Tweets from Twitter in this paper. Polarity classification is the fundamental task in Sentiment Analysis. It can be achieved through multiple approaches. Lexicon-based methods decide the polarity of the document based on polarities of individual words and phrases in the document. Machine Learning based methods aim at building classifiers to resolve polarity and identify which category a document belongs to. We focus on the Machine Learning based approach. Over the years, a large amount of wide-reaching research has been carried out in the area of sentiment classification on large texts as explained by Pang and Lee(2008). There has been research on effects of machine learning techniques such as Naive Bayes, Maximum Entropy, and Support Vector Machines in the specific domains. More recently, similar techniques have been applied to the Twitter microblogging platform as stated by Go and Huang(2009). Machine learning techniques like support vector machine is getting popularity in recent time as stated by Roy and Viswanatham(2016), Roy et al.(2015), Basu et al.(2015), Mittal et al.(2015) and Das et al(2014). Challenges include a wide range of topics to be classified, use of informal language on social media platforms, and the message length limitation of 140 characters per tweet imposed by Twitter. In our experiments, we transact a 2-way classification into positive and negative labels, unlike other approaches such as multi-label or multi-class classification as stated by Liu and Chen(2015). Support Vector Machines are a group of associated learning methods that aim to recognize inherent patterns in data as stated by Boser et al.(1992). They are one of the several kernel-based techniques of machine learning algorithms. SVMs have been successfully applied to varied fields. SVMs can be used for both classification and regression analysis. However, given our focus on sentiment analysis on microblogging data, we limit our investigation to classification techniques.

In this paper, we use Support Vector Machines (with radial kernel) to categorise tweets based on their inherent sentiment, effectively. For the purposes of our research, we define sentiment to be “a personal positive or negative feeling” as stated by Go et al.(2009).

RELATED WORK

Sentiment Analysis began as classification of reviews into two categories: positive and negative as stated by Turney(2002). This was followed by additional work focusing on analysis of movie reviews by overall sentiment of everyone’s document as explained by Pang and Lee(2008). However, these were both early approaches applied to large-sized documents. Sentiment Analysis yielded different results when applied

9 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-new-svm-method-for-recognizing-polarity-of-sentiments-in-twitter/179396

Related Content

Swarm Intelligence for Electromagnetic Problem Solving

Luciano Mescia, Pietro Bia, Diego Caratelli and Johan Gielis (2017). *Handbook of Research on Soft Computing and Nature-Inspired Algorithms* (pp. 69-100).

www.irma-international.org/chapter/swarm-intelligence-for-electromagnetic-problem-solving/179390

Medical Outcome Prediction for Intensive Care Unit Patients

Simone A. Ludwig, Stefanie Roos, Monique Frize and Nicole Yu (2010). *International Journal of Computational Models and Algorithms in Medicine* (pp. 19-30).

www.irma-international.org/article/medical-outcome-prediction-intensive-care/51668

Multi-Objective Big Data View Materialization Using MOGA

Akshay Kumar and T. V. Vijay Kumar (2022). *International Journal of Applied Metaheuristic Computing* (pp. 1-28).

www.irma-international.org/article/multi-objective-big-data-view-materialization-using-moga/292499

A Hybrid Particle Swarm Algorithm for Resource-Constrained Project Scheduling

Jens Czogalla and Andreas Fink (2012). *Hybrid Algorithms for Service, Computing and Manufacturing Systems: Routing and Scheduling Solutions* (pp. 137-157).

www.irma-international.org/chapter/hybrid-particle-swarm-algorithm-resource/58520

Solving Task Scheduling Problem in the Cloud Using a Hybrid Particle Swarm Optimization Approach

Salmi Cheikh and Jessie J. Walker (2022). *International Journal of Applied Metaheuristic Computing* (pp. 1-25).

www.irma-international.org/article/solving-task-scheduling-problem-in-the-cloud-using-a-hybrid-particle-swarm-optimization-approach/284576