

## Chapter 53

# Hybrid Approach for Sentiment Analysis of Twitter Posts Using a Dictionary–Based Approach and Fuzzy Logic Methods: Study Case on Cloud Service Providers

**Jamilah Rabeh Alharbi**

*Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia*

**Wadee S. Alhalabi**

*Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia*

### **ABSTRACT**

*Recently, sentiment analysis of social media has become a hot topic because of the huge amount of information that is provided in these networks. Twitter is a popular social media application offers businesses and government the opportunities to share and acquire information. This article proposes a technique that aims at measuring customers' satisfaction with cloud service providers, based on their tweets. Existing techniques focused on classifying sentimental text as either positive or negative, while the proposed technique classifies the tweets into five categories to provide better information. A hybrid approach of dictionary-based and Fuzzy Inference Process (FIP) is developed for this purpose. This direction was selected for its advantages and flexibility in addressing complex problems, using terms that reflect on human behaviors and experiences. The proposed hybrid-based technique used fuzzy systems in order to accurately identify the sentiment of the input text while addressing the challenges that are facing sentiment analysis using various fuzzy parameters.*

## INTRODUCTION

Social media are considered as a huge corpus for extracting information of various types. One of the 'hot topics' in social media usage is the sentiment analysis. A group's opinions and feelings about any mentioned topic, serve as a valuable source of marketing for companies and consumers alike. Companies may utilize this information to measure customers' satisfaction about the product and it helps facilitate the decision-making process of the consumer. Consumers may seek both positive and/or negative feedback prior to make a decision, which is known as 'opinion mining'. Twitter, which has over 600 million users and nearly 330 million active users worldwide (Statista, 2017), rapidly becomes a 'golden astrologer' in the corporate universe, as it allows companies to elicit and analyze the sentiment of users' tweets and employ the result of the analysis to grow their own image and trademark. There are two key parameters that directing any research on sentiment analysis, these are: the target field and the study sample. In this study, the study sample will be the consumer reviews that are collected from Twitter and Cloud Service Providers will serve as the target field in this study.

Several approaches were developed to enhance the accuracy of the sentiment analysis in social networks. The most accurate results were obtained using a hybrid technique of machine learning approaches (e.g. support vector machine, Naïve-Bayes and K-mean Fuzzy) and lexicon-based approaches, like a dictionary-based and corpus-based approaches. A dictionary-based approach is based on utilizing dictionaries that include words and their sentiments, which is a robust and straight-forward approach as most dictionaries list synonyms and antonyms for each term. While there are various such dictionaries available online, according to Islam Zibran (2017) and Jhaveri et al. (2011) SentiWordNet and SentiStrength are the best among other online dictionaries. Different techniques addressed different challenges of sentiment analysis. These challenges have been addressed in the previous work, but separately. Some techniques focused on increasing the accuracy using a hybrid approach, but did not address the challenges of intensity, word sense and comparative words. Accordingly, a new technique that solved the expected forms of sentimental text, these are word sense, comparative word, compound words, negative, intensity, sarcasm and big data, is required.

The aim of this study is to measure user satisfaction about the Cloud services provided by Google, Microsoft and Amazon. Tracking the users' replies will present a general view on the satisfaction level and customers' opinions, which give the providers information about possible improvements. The proposed approach is built using these well-known dictionaries that are mentioned above and using fuzzy logic classification. Fuzzy logic is a computational model that mimics human decision-making process. Fuzzy logic represents the domain of interest using fuzzy terms, similar to how the human brain absorbs information (e.g. a temperature is hot, a speed is slow). The ability of the human brain to reason with uncertainties, inspired researchers to develop a fuzzy logic model (Liu & Cocea, 2017). Fuzzy allows to create general rules for decision making with uncertainty (Mary & Arockiam, 2017; Serguieva et al., 2017). Tweet classification, similar to other classification systems, is implemented based on a set of extracted features and using a classification procedure/algorithm. Fuzzy Inference Process (FIP) is used for the classification process because of its flexibility in addressing complex problems. Moreover, FIP allows for incorporating various feature forms, such as those extracted directly from the input or using a dictionary-based approach.

The proposed approach accurately classifies tweets based on their sentiment content. Accordingly, the tweets will be categorized into five categories: Very Positive, Positive, Neutral, Negative, Very Negative. The proposed model is built to achieve a set of goals: 1) To collect cloud services' relevant tweets

30 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/hybrid-approach-for-sentiment-analysis-of-twitter-posts-using-a-dictionary-based-approach-and-fuzzy-logic-methods/308531](http://www.igi-global.com/chapter/hybrid-approach-for-sentiment-analysis-of-twitter-posts-using-a-dictionary-based-approach-and-fuzzy-logic-methods/308531)

## Related Content

---

### Big Collusion: Corporations, Consumers, and the Digital Surveillance State

Garry Robson and C. M. Olavarria (2016). *Big Data: Concepts, Methodologies, Tools, and Applications* (pp. 1755-1772).

[www.irma-international.org/chapter/big-collusion/150241](http://www.irma-international.org/chapter/big-collusion/150241)

### Contextualized Text OLAP Based on Information Retrieval

Lamia Oukid, Nadja Benblidia, Fadila Bentayeb, Ounas Asfari and Omar Boussaid (2015). *International Journal of Data Warehousing and Mining* (pp. 1-21).

[www.irma-international.org/article/contextualized-text-olap-based-on-information-retrieval/125648](http://www.irma-international.org/article/contextualized-text-olap-based-on-information-retrieval/125648)

### Financial Data Mining Using Flexible ICA-GARCH Models

Philip L.H. Yu, Edmond H.C. Wu and W.K. Li (2010). *Dynamic and Advanced Data Mining for Progressing Technological Development: Innovations and Systemic Approaches* (pp. 255-272).

[www.irma-international.org/chapter/financial-data-mining-using-flexible/39645](http://www.irma-international.org/chapter/financial-data-mining-using-flexible/39645)

### Personal Health and Illness Management and the Future Vision of Biomedical Clothing Based on WSN

Ge Zhang and Zubin Ning (2023). *International Journal of Data Warehousing and Mining* (pp. 1-21).

[www.irma-international.org/article/personal-health-and-illness-management-and-the-future-vision-of-biomedical-clothing-based-on-wsn/316126](http://www.irma-international.org/article/personal-health-and-illness-management-and-the-future-vision-of-biomedical-clothing-based-on-wsn/316126)

### LF-LDA: A Supervised Topic Model for Multi-Label Documents Classification

Yongjun Zhang, Zijian Wang, Yongtao Yu, Bolun Chen, Jialin Ma and Liang Shi (2018). *International Journal of Data Warehousing and Mining* (pp. 18-36).

[www.irma-international.org/article/lf-lda/202996](http://www.irma-international.org/article/lf-lda/202996)