Chapter 8

Detection of Economy-Related Turkish Tweets Based on Machine Learning Approaches

Jale Bektaş
Mersin University, Turkey

ABSTRACT

Conducting NLP for Turkish is a lot harder than other Latin-based languages such as English. In this study, by using text mining techniques, a pre-processing frame is conducted in which TF-IDF values are calculated in accordance with a linguistic approach on 7,731 tweets shared by 13 famous economists in Turkey, retrieved from Twitter. Then, the classification results are compared with four common machine learning methods (SVM, Naive Bayes, LR, and integration LR with SVM). The features represented by the TF-IDF are experimented in different N-grams. The findings show the success of a text classification problem is relative with the feature representation methods, and the performance superiority of SVM is better compared to other ML methods with unigram feature representation. The best results are obtained via the integration method of SVM with LR with the Acc of 82.9%. These results show that these methodologies are satisfying for the Turkish language.

1. INTRODUCTION

Social media such as Twitter makes interpersonal communications more effective with the aid of virtual platforms. Today, the internet has become a global forum where people can freely express and share their ideas. However, this situation has brought along some problems. It may be difficult for users to get through this

excessive information as anyone can upload exorbitant amounts of information, be it consciously or unconsciously.

Social media makes the communication of people with each other more effective with the use of virtual platforms. Microblogs such as Twitter, a type of social network, are used by professionals to share information and news. Twitter is a popular Microblog service that quickly disseminates information about any incident happening anywhere on the world. Jost et al. (2018) discusses that such personal records on social networks provide valuable and useful information for social psychology, marketing intelligence, and opinion mining research. Prabowo & Thelwall (2009) discuss that it is important to examine the content of the data collected from social media to derive meaningful information through classification research.

In social media studies, the techniques of data mining and text mining are used in the analysis of textual information (Surjandari et al., 2015). The core of these statistical methods consist of text mining techniques, which use such parameters as: author recognition, text classification, sentiment analysis, and opinion mining (Hemmatian et al., 2019).

The process of text analysis can be summarised as follows. First, the informational content of the document is transformed into a structure form in a vector space model. In fact, most text mining techniques are based on the idea that any document can be represented with bag-of-words projection in accordance with its term group (Karthikeyan et al., 2019). According to this projection, each *i* document of the collection is represented as a *N*-dimensional vector and the document is shown as $V_i = \left\{ w\left(t_{i1}\right), \ldots, w\left(t_{ij}\right), \ldots, w\left(t_{iN}\right) \right\}$. According to this projection, *N* determines the number of terms, and $w\left(t_{ij}\right)$ determines the weight of t_j term in document *i*. This way of thought can improve upon by tweaking the parameters by introducing new operations so that the text mining may specialise in specific goals. The introduction of operations, such as: accommodating terms for machine learning (ML) techniques by transforming them into numeric features (i.e. weights), linguistic analysis techniques, indexing, statistical techniques, filtering terms via particular keywords, feature extraction, future selection (this subject will scale down in high-dimension datasets and it is highly important for the selection of attributions that will contribute to the study the most) and text summarization are some of the examples that help with the optimization the process (Sharma & Jain, 2019). Next, data mining and ML algorithms are applied to classify documents in vector space projections, grouping or building regression models.

23 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/detection-of-economy-related-turkish-tweets-based-on-machine-learning-approaches/293155

Related Content

A Survey on Database Performance in Virtualized Cloud Environments
Todor Ivanov, Ilia Petrovand Alejandro Buchmann (2012). *International Journal of Data Warehousing and Mining (pp. 1-26).*
www.irma-international.org/article/survey-database-performance-virtualized-cloud/67571

Computational Intelligence Clustering for Dynamic Video Watermarking
Daw-Tung Linand Guan-Jhih Liao (2010). *Intelligent Soft Computation and Evolving Data Mining: Integrating Advanced Technologies (pp. 298-318).*
www.irma-international.org/chapter/computational-intelligence-clustering-dynamic-video/42366

From Change Mining to Relevance Feedback: A Unified View on Assessing Rule Interestingness
Mirko Boettcher, Georg Ruß, Detlef Nauckand Rudolf Kruse (2009). *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction (pp. 12-37).*
www.irma-international.org/chapter/change-mining-relevance-feedback/8435

Data Mining in the Social Sciences and Iterative Attribute Elimination
Anthony Scime, Gregg R. Murray, Wan Huangand Carol Brownstein-Evans (2008). *Data Mining and Knowledge Discovery Technologies (pp. 308-332).*
www.irma-international.org/chapter/data-mining-social-sciences-iterative/7522

Preference-Based Frequent Pattern Mining
Moonjung Cho, Jian Pei, Haixun Wangand Wei Wang (2005). *International Journal of Data Warehousing and Mining (pp. 56-77).*
www.irma-international.org/article/preference-based-frequent-pattern-mining/1759