

Chapter 28

Feature Optimization in Sentiment Analysis by Term Co-occurrence Fitness Evolution (TCFE)

Sudarshan S. Sonawane

Department of Computer Engineering, Shri Gulabrao Deokar College of Engineering, Jalgaon, India

Satish R. Kolhe

School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University,
Jalgaon, India

ABSTRACT

The opinion of a target audience is a major objective for the assessing state of efficacy pertaining to reviews, business decisions surveys, and such factors that require decision making. Feature selection turns out to be a critical task for developing robust and high levels of classification while decreasing training time. Models are required for stating the scope for depicting optimal feature selection for escalating feature selection strategies to escalate maximal accuracy in opinion mining. Considering the scope for improvement, an n-gram feature selection approach is proposed where optimal features based on term co-occurrence fitness is proposed in this article. Genetic algorithms focus on determining the evolution and solution to attain deterministic and maximal accuracy having a minimal level of computational process for reflecting on the sentiment scope for sentiment. Evaluations reflect that the proposed solution is capable, which outperforms the separate filter-oriented feature selection models of sentiment classification.

DOI: 10.4018/978-1-6684-6303-1.ch028

INTRODUCTION

User-generated content analysis has become very vital in terms of analyzing the data insights that are generated by the organization. Sentiment analysis as a subject has gained momentum and globally there are many companies that are focusing on using the user-centric data to analyze the issues and identify the insights that they could use for enhancing the user experience.

Analyzing user opinions are very critical in the decision making and globally many companies have reaped potential benefits of same. Profoundly, the sentiment analysis enables users to have an opinion and the consumer behavior patterns. Opinion of one single individual might be ignorable but considering the viewpoints of the macro set of consumers is very essential to evaluate and adjudge the brand perception among the customers, as divergent views of customers can be envisaged from such process (Hu & Liu, 2004).

In NLP (Natural Language Processing) opinion mining and the process of sentiment analysis has gained profound importance (Liu, 2012) and profoundly in the case of web mining, analytics of social media messages data mining has become an important part.

The study (Cambria, Das, Bandyopadhyay, & Feraco, 2017) emphasizes the importance of opinion mining as:

- Entity (E) (Deng & Wiebe, 2015), the ones that focus on the entity, event, a service or an individual task
- Entity Aspects (A) are the ones related to a product, an organization, resulting outcome of an event or service.
- Opinion Articulated Time (T) which reflects the event of time wherein the opinion is expressed.
- The sentiment (S) like the positive, or opinion oriented like negative or neutral or the ones that relate to aspect of entity.

It is imperative from the aforesaid objectives that the sentiment analysis are being resourceful in various dimensions that are stated above, that focus on the overall opinion of target set representatives. Majority of the sentiment analysis solutions contributed are machine-learning solutions that focus on the opinion aspects.

The process of feature representation usually impacts the efficacy of machine learning approaches (LeCun, Bengio & Hinton, 2015) and there is need for focused attempt in terms of using the feature selection strategies that can deliver better outcome (Mohammad, Kiritchenko & Zhu, 2013). Learning mechanisms are very effective in terms of choosing the discriminative features (Bengio, Courville & Vincent, 2013), learners ensemble (LeCun, Bengio & Hinton, 2015) which leads to more opportunities for learning from the chosen data at distinct levels and can it can lead to deep learning conditions. However, the majority of contemporary contributions related to opinion mining are portraying the optimal features from individual elements such as terms (subject, predicates, or sentiment lexicons) of the learning data selection.

In this paper, an improved n-gram feature selection approach and optimal features based on term co-occurrence fitness evolution are proposed. The proposed model is compared with genetic rank aggregation model. Performance analysis shows that the proposed model is outstanding and robust with high classification accuracy as compared with genetic rank aggregation model.

21 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/feature-optimization-in-sentiment-analysis-by-term-co-occurrence-fitness-evolution-tcfe/308505

Related Content

Modeling and Querying Continuous Fields with OLAP Cubes

Leticia Irene Gómez, Silvia Alicia Gómez and Alejandro Vaisman (2013). *International Journal of Data Warehousing and Mining* (pp. 22-45).

www.irma-international.org/article/modeling-querying-continuous-fields-olap/78374

Selection of High Quality Rules in Associative Classification

Silvia Chiusano and Paolo Garza (2009). *Post-Mining of Association Rules: Techniques for Effective Knowledge Extraction* (pp. 173-198).

www.irma-international.org/chapter/selection-high-quality-rules-associative/8443

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

Mining Customer Knowledge for a Recommendation System in Convenience Stores

Shu-Hsien Liao, Chih-Hao Wen, Pei-Yuan Hsian, Chien-Wen Li and Che-Wei Hsu (2014). *International Journal of Data Warehousing and Mining* (pp. 55-86).

www.irma-international.org/article/mining-customer-knowledge-for-a-recommendation-system-in-convenience-stores/110386

Adaptation of Error Adjusted Bagging Method for Prediction

Selen Yilmaz Isikhan, Erdem Karabulut, Afshin Samadi and Saadettin Kılçık (2019). *International Journal of Data Warehousing and Mining* (pp. 28-45).

www.irma-international.org/article/adaptation-of-error-adjusted-bagging-method-for-prediction/228936