

Chapter 2

Sentiment Mining Approaches for Big Data Classification and Clustering

Ashok Kumar J
Anna University, India

Abirami S
Anna University, India

Tina Esther Trueman
Anna University, India

ABSTRACT

Sentiment analysis is one of the most important applications in the field of text mining. It computes people's opinions, comments, posts, reviews, evaluations, and emotions which are expressed on products, sales, services, individuals, organizations, etc. Nowadays, large amounts of structured and unstructured data are being produced on the web. The categorizing and grouping of these data become a real-world problem. In this chapter, the authors address the current research in this field, issues and the problem of sentiment analysis on Big Data for classification and clustering. It suggests new methods, applications, algorithm extensions of classification and clustering and software tools in the field of sentiment analysis.

DOI: 10.4018/978-1-5225-2805-0.ch002

INTRODUCTION

Sentiment analysis (also called opinion mining) is one of the most important research fields in Information Processing, Text mining, and Linguistics. Information shared in online social networks (forums, blogs, etc.) contain useful information about products, sales, service, individuals, organizations, Governments, policies, politics, etc. This information is expressed in two ways namely, an objective statement and a subjective statement. The objective statement refers to some factual information and subjective statement refers to personal feelings such as love, joy, surprise, anger, sadness, and fear. The amount of information collected in terms of volume, variety, and velocity is called Big Data. It can be treated as software as a service, platform as a service and infrastructure as a service. Based on these services, big data creates opportunities and challenges to extract valuable information.

This chapter addresses the sentiment classification and clustering techniques, such as Machine learning based approach, lexicon-based approach, and the hybrid approach. These in turn include techniques like Symmetric Matrix-based predictive classifier, Comparative experiments using supervised learning, Classifier ensembles, Ranked Word Graph for Sentiment polarity Classification, Sentiment Classification via Social context regularization, Context-Based Sentiment analysis, Semi-supervised subjective feature weighting and intelligent modelling, Cross-lingual sentiment classification: Similarity discovery, Supervised study weighting scheme, SVM with the linear kernel as the classification model, unsupervised dependency parsing-based text classification method, SACI (Sentiment Analysis by Collective Inspection), and a lexicon-based unsupervised method.

PROBLEM OF SENTIMENT ANALYSIS ON BIG DATA

Big data creates new challenges on data processing, data storage, data representation, pattern mining, visualization, etc (Gema Bello-Orgaz et al., 2016) in the field of data mining, machine learning, natural language processing, text mining, social networks, and sentiment analysis. The rapid growth of unstructured data in social networking, blogs, reviews, posts, comments, and tweets are the most important source for sentiment analysis (Changli Zhang et al., 2008). Further, sentiment analysis problems focus on different levels namely Document level, Sentence level, and aspect level (R Feldman, 2013). The document-level sentiment analyzes a piece of information. It represents multiple opinions, but not a single opinion view. Document level sentiment analysis hides insights and useful information by reducing

28 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/sentiment-mining-approaches-for-big-data-classification-and-clustering/185978

Related Content

Mining and Analysis of the Traffic Information Situation in the South China Sea Based on Satellite AIS Data

Tianyu Pu (2023). *International Journal of Data Warehousing and Mining* (pp. 1-25). www.irma-international.org/article/mining-and-analysis-of-the-traffic-information-situation-in-the-south-china-sea-based-on-satellite-ais-data/332864

Adaptation of Error Adjusted Bagging Method for Prediction

Selen Yilmaz Isikhan, Erdem Karabulut, Afshin Samadian 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

A Directed Acyclic Graph (DAG) Ensemble Classification Model: An Alternative Architecture for Hierarchical Classification

Esra'a Alshdaifat, Frans Coenen and Keith Dures (2017). *International Journal of Data Warehousing and Mining* (pp. 73-90). www.irma-international.org/article/a-directed-acyclic-graph-dag-ensemble-classification-model/185659

Business Collaboration by Privacy-Preserving Clustering

Stanley R.M. Oliveira and Osmar R. Zaiane (2009). *Social Implications of Data Mining and Information Privacy: Interdisciplinary Frameworks and Solutions* (pp. 113-133). www.irma-international.org/chapter/business-collaboration-privacy-preserving-clustering/29147

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