


Chapter 39

Challenges of Text Analytics in Opinion Mining

Vaishali Kalra

Manav Rachna International Institute of Research and Studies, India

Rashmi Agrawal

 <https://orcid.org/0000-0003-2095-5069>

Manav Rachna International Institute of Research and Studies, India

ABSTRACT

Text analysis is the task of knowledge distillation from unstructured text. Due to increase in sharing of information over the web in text format, users required tools and techniques for the analysis of the text. These techniques can be used in two ways: One, this can be used for clustering, classification, and visualization of the data. Two, this can be used for predicting the future aspects, for example, in share market. But all these tasks are not easy to perform, as there are lots of challenges in converting the text into the format onto which various actions can be taken. In this chapter, the authors have discussed the framework of text analysis, followed by the background where they have discussed the steps for transforming the text into the structured form. They have shed light on its industry application along with the technological and non-technological challenges in text analysis.

INTRODUCTION TO TEXT ANALYTICS

In real world, organization needs to take decision every other minute, in order to ensure organizational success. Such decisions may include but not limited to introduction of new product and its potential demand, profitability, market share, competitor's benchmarking etc. In the past, such decisions were taken by the top management based on their experience only but such decisions may not be always good for the organization, however, with the advent of technology the amount of data has increased massively, so now decision cannot be taken easily and requires analytics technique. Using Analytics techniques businesses can take strategic decision which are more reliable compared to decision taken on judgment of single person.

DOI: 10.4018/978-1-6684-6303-1.ch039

Challenges of Text Analytics in Opinion Mining

Text Analytics is one of the methods to analyze the textual data available on web. Text analytics can be defined as a way for computers to analyze the text using natural language processing to derive the certain facts from raw text. Such analysis can be in form of retrieval of customer opinion regarding any product, hotel reviews, movie reviews, categorization of documents based on the given information, can be used in market analysis and prediction and so many other similar tasks (Irfan et al., 2015).

In order to achieve the above purpose, the users need to follow the text mining process. Generally, Text mining and text analytics are alternatively used but there is thin line difference between the two. Text Mining can be understood as the process to retrieve information from data; however, information can be retrieved from data using text analytics techniques (Agrawal & Batra, 2013)

But both cannot be used in isolation as the end objective is same to take informed decision and taking an informed decision is not possible without using both. To further elaborate, Text analytics can be applied to any text data, which can be in any native language like Japanese, Chinese, English, Hindi etc. and same is available on the web. Such web data is not only the textual content always and it can have images, audios, and videos, which make it completely unstructured data. So the task of text analytics is to extract the text from retrieved real-world information from the web and application of text mining to visualize the text data only.

On the above retrieved text in the past, the text analytics only plays with bag of words, the word frequencies and are used for summarization, clustering the documents and classification of document topic wise (Agrawal, 2014). It does not have the capability of knowing the meaning of the text; it has difficulty in handling the problem of polysemy, homonymy, synonyms and deriving the hidden information that is called semantic analysis or qualitative analysis (Hu & Liu, 2012). Textual data comes with additional challenges such as incorrect spellings, incorrect syntax of the sentences and it leads to challenges for the extraction of the correct information out of that and its processing also. Therefore, researchers are focusing more on handling such data because of above issues (Knoblock, Lopresti, Roy, & Subramaniam, 2007). They are investing quality time in handling the complexities and its high dimensionality of the large corpus of data.

Although the researchers are applying statistical methods and techniques like using singular value decomposition and support vector machine for handling the high dimensionality issues, word sense disambiguation for handling synonyms problems, however, the challenges has not completely resolved and work is still in progress. The intent of this chapter is to study some of the challenges faced by researchers today, let's make a deep dive.

BACKGROUND

Framework of Text Analytics

The text analytics is performed on text documents, so the first step of text analytics should be extraction of text data from the web which is followed by conversion of this text data into an actionable form is shown in Figure 1.

The data source is any repository available for the analysis, for market analysis it can be data available on social media sites like Twitter, Facebook, blogs, etc, it can be the sensors data, academic records, patient history, like this there are various resources are available from where the data can be extracted. Text content available in these repositories is in unstructured format, as there is not only the text data.

11 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/challenges-of-text-analytics-in-opinion-mining/308517

Related Content

Text Mining and Patient Severity Clusters

Patricia Cerrito (2010). *Text Mining Techniques for Healthcare Provider Quality Determination: Methods for Rank Comparisons* (pp. 287-340).

www.irma-international.org/chapter/text-mining-patient-severity-clusters/36639

Computing Join Aggregates Over Private Tables

Rong She, Ke Want, Ada Waichee Fuand Xu Yabo (2008). *International Journal of Data Warehousing and Mining* (pp. 22-41).

www.irma-international.org/article/computing-join-aggregates-over-private/1816

Detecting Trends in Social Bookmarking Systems: A del.icio.us Endeavor

Robert Wetzker, Carsten Zimmermannand Christian Bauckhage (2010). *International Journal of Data Warehousing and Mining* (pp. 38-57).

www.irma-international.org/article/detecting-trends-social-bookmarking-systems/38953

Automatic Human Emotion Classification in Web Document Using Fuzzy Inference System (FIS): Human Emotion Classification

P Mohamed Shakeeland S Baskar (2022). *Research Anthology on Implementing Sentiment Analysis Across Multiple Disciplines* (pp. 1064-1075).

www.irma-international.org/chapter/automatic-human-emotion-classification-in-web-document-using-fuzzy-inference-system-fis/308533

Domain-Driven Data Mining: A Practical Methodology

Longbing Caoand Chengqi Zhang (2006). *International Journal of Data Warehousing and Mining* (pp. 49-65).

www.irma-international.org/article/domain-driven-data-mining/1774