Chapter 17
Text Mining:
Current Trends and Applications

Kijpokin Kasemsap
Suan Sunandha Rajabhat University, Thailand

ABSTRACT

This chapter reveals the overview of text mining; text mining, patent analysis, and keyword selection; text mining and sentiment analysis in modern marketing; text mining applications in the biomedical sciences; and the multifaceted applications of text mining. Text mining is an advanced technology utilized in business, marketing, biomedical sciences, education, and operations. Text mining offers a solution to many problems, drawing on techniques concerning information retrieval, natural language processing, information extraction, and knowledge management. Through text mining, information can be extracted to derive summaries for the words contained in the documents. Text mining has the potential to increase the research base available to business and society and to enable business to utilize the research base more effectively. Economic and societal benefits of text mining include cost savings, productivity gains, innovative new service development, new business models, and new medical treatments.

INTRODUCTION

There is a tremendous growth in the volume of online text documents from networked resources, such as the Internet, digital libraries, and company-wide intranets (Kim, 2009). Many text mining approaches are based on words in the texts (Loh, Wives, Lichtnow, & de Oliveira, 2009). Text mining deals with how to extract the latent knowledge from the unstructured textual descriptions (Yoon, Park, & Coh, 2014). Text mining requires the highly scalable algorithms to meet the overall performance demands (Indurkhya, 2015) and facilitates the identification of relevant literature, its rapid categorization, and its summarization (Thomas, McNaught, & Ananiadou, 2011). Text mining has adopted certain techniques from the more general field of data analysis, including sophisticated methods for analyzing relationships among highly formatted data, such as numerical data or data with a relatively small fixed number of possible values (Kao et al., 2012).

DOI: 10.4018/978-1-5225-1877-8.ch017

Text mining can be broadly defined as a knowledge-intensive process in which a user interacts with a document collection over time by using a combination of analysis tools (Cheney, 2015). The good text knowledge representation model should contain rich text semantics and should automatically construct with a lower complexity (Zhang, Luo, He, & Cai, 2013). Text mining involves six main phases (i.e., business understanding, data understanding, data preparation, modeling, evaluation, and deployment). Perovšek et al. (2016) indicated that text mining can be distinguished from general data mining by special procedures applied in the data preparation phase, where unstructured or poorly structured text needs to be converted into organized data, structured as a table of instances (rows) described by attributes (columns).

This chapter aims to bridge the gap in the literature on the thorough literature consolidation of text mining. The extensive literature of text mining provides a contribution to practitioners and researchers by describing the trends and applications of text mining in order to maximize the technological impact of text mining in the digital age.

Background

Texts are written in natural language, carrying out implicit knowledge, and ambiguities (Cherfi, Napoli, & Toussaint, 2009). Text mining, also known as text data mining or text analytics, is considered as a subfield of data mining research (Perovšek et al., 2016) and continues to expand as mass volumes of unstructured data (Karl, Wisnowski, & Rushing, 2015). Data mining is the process of applying these computational methods in showing unknown data formats in large data sets (Kasemsap, 2015) and provides a set of techniques of artificial intelligence which can be used to increase the efficiency of data mining methods (Guerrero et al., 2014). Data mining brings the new direction on business planning from the last decades (Klepac & Berg, 2015) and is utilized to discover patterns and relationships in the data in order to help make better business decisions (Kasemsap, 2016a).

Text mining tasks include the activities of search engines, such as assigning texts to one or more categories (i.e., text categorization), grouping similar texts together (i.e., text clustering), finding the subject of discussions (i.e., concept/entity extraction), finding the tone of a text (i.e., sentiment analysis), summarizing documents, and learning relations between entities described in a text (i.e., entity relation modeling) (Truyens & van Eecke, 2014). Text mining techniques include text categorization, summarization, topic detection, concept extraction, search and retrieval, and document clustering (Hashimi, Hafez, & Mathkour, 2015). Text categorization concerns of classifying documents into some categories according to their contents, characteristics, and properties (Yang, Lee, & Hsiao, 2015). When documents are properly categorized, documents in a cluster should have a common theme (Yang et al., 2015).

The applications of text mining include various disciplines, ranging from biomedicine to legal, business intelligence, and security (Truyens & van Eecke, 2014). Text mining applications enhance novel thinking (Segers & de Vries, 2003), develop artificial intelligence (Falkenhainer, Forbus, & Gentner, 1986), and create valuable knowledge for more effective knowledge management (Xu & Luo, 2009). Text mining seeks to extract the useful information from data sources through the identification and exploration of interesting patterns (Cheney, 2015) and helps individual efficiently analyze a large number of texts (Yamada, Kato, & Hirokawa, 2013). As more reliable tools are developed for text analysis, it is important to capture the text information for an analysis that was unavailable within conventional text analysis approaches (Fujii, Iwayama, & Kando, 2007).

Related Content