Chapter X
Current Issues and Future Analysis in Text Mining for Information Security Applications

Shuting Xu
Virginia State University, USA

Xin Luo
Virginia State University, USA

ABSTRACT

Text mining is an instrumental technology that today's organizations can employ to extract information and further evolve and create valuable knowledge for more effective knowledge management. It is also an important tool in the arena of information systems security (ISS). While a plethora of text mining research has been conducted in search of revamped technological developments, relatively limited attention has been paid to the applicable insights of text mining in ISS. In this chapter, we address a variety of technological applications of text mining in security issues. The techniques are categorized according to the types of knowledge to be discovered and the text formats to be analyzed. Privacy issues of text mining as well as future trends are also discussed.

INTRODUCTION

Text mining is an instrumental technology that today's organizations can employ to extract information and further evolve and create valuable knowledge for more effective knowledge management. The deployment of text mining technology is also of vital importance in the arena of information systems security (ISS), especially after the 9/11 tragedy which galvanized the government to increasingly spend resources pursuing hardened homeland security. Furthermore, providing security to computer information systems and communications infrastructures is now one of the national priorities. While a plethora of text mining or data mining research has been conducted in search of revamped technological development, relatively limited attention has been paid to the applicable insights of text mining in ISS.

In an effort to fill the void, this article tends to shed light on the correlations between text mining and ISS and to address a variety of technological applications of text mining in ISS privacy issues as well as future trends related to security. As such, this article is organized as follows: after an introduction section, the second section presents the background of text mining and its application in security; the third section addresses different categories of techniques used for text mining in ISS, including social network analysis, abnormal detection, topic discovery, and identity detection; section four discusses social perspectives of text mining versus privacy violation and section five presents future analysis of text mining for ISS research.

BACKGROUND

Defined as "*the discovery by computer of new, previously unknown, information by automatically extracting information from different written resources*"(Fan et al. 2006), text mining is an emerging technology characterized by a set of technological tools which allow for the extraction of unstructured information from text. With the exponential growth of the internet, it is literally cumbersome for individuals as well as companies to process all the overwhelmed information. Not like some data mining techniques discovering knowledge from only the structured data, such as numeric data, text mining is related to finding knowledge from the unstructured textual data including e-mails, Web pages, business reports, and articles, etc. Leaping from old-fashioned information retrieval to information and knowledge discovery, text mining applies the same analytical functions of data mining to the domain of textual information and replies on sophisticated text analysis techniques that distill information from free-text documents (Dörre et al. 1999).

As voluminous corporate information must be merged and managed and the dynamic business environment pushes decision makers to promptly and effectively locate, read, and analyze relevant documents to produce the most informative decisions, discovering hidden patterns from the structured data plays an important role in business where patterns are paramount for strategic decision making. Text mining pursues knowledge discovery from textual databases by isolating key bits of information from large amounts of text, by identifying relationships among documents, and by inferring new knowledge from them (Durfee 2006). Furthermore, (Fan et al. 2006) indicated that the key to text mining is creating technology that combines a human's linguistic capabilities with the speed and accuracy of a computer. Gluing the generic process model for text-mining application proposed by (Fan et al. 2006) and general text mining framework suggested by (Durfee 2006), we think that the following model can capture the processes involved in text mining from text collection and distillation to knowledge representation (see Figure 1).

Related Content

Adaptation of Error Adjusted Bagging Method for Prediction

Selen Yilmaz Isikhan, Erdem Karabulut, Afshin Samadiand Saadettin Klçkap (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

Efficient Top-k Keyword Search Over Multidimensional Databases

Ziqiang Yu, Xiaohui Yuand Yang Liu (2013). *International Journal of Data Warehousing and Mining (pp. 1-21).*

www.irma-international.org/article/efficient-top-keyword-search-over/78373

An Intelligent Data Mining System Through Integration Of Electromagnetism-Like Mechanism And Fuzzy Neural Network

Peitsang Wuand Yung-Yao Hung (2010). *Intelligent Soft Computation and Evolving Data Mining: Integrating Advanced Technologies (pp. 61-84).*

www.irma-international.org/chapter/intelligent-data-mining-system-through/42356

An Efficient Pruning and Filtering Strategy to Mine Partial Periodic Patterns from a Sequence of Event Sets

Kung-Jiuan Yang, Tzung-Pei Hong, Yuh-Min Chenand Guo-Cheng Lan (2014). *International Journal of Data Warehousing and Mining (pp. 18-38).*

www.irma-international.org/article/an-efficient-pruning-and-filtering-strategy-to-mine-partial-periodic-patterns-from-a-sequence-of-event-sets/110384

Data Management and Big Data Analytics: Data Management in Digital Economy

Vellingiri Jayagopaland Basser K. K. (2019). *Optimizing Big Data Management and Industrial Systems With Intelligent Techniques (pp. 1-23).*

www.irma-international.org/chapter/data-management-and-big-data-analytics/218738