Comprehensive Contemplation of Probabilistic Aspects in Intelligent Analytics

Neeti Sangwan, USICT, GGS Indraprastha University and MSIT, New Delhi, India

Vishal Bhatnagar, Ambedkar Institute of Advanced Communication Technologies and Research, New Delhi, India

ABSTRACT

In Big Data analysis, the application of machine learning has proven to be a revolutionary. The systematic review of literature shows that research has been carried out on the domain of big data analytics particularly text analytics with the inclusion of machine learning approaches. This extensive survey deals with the data at hand that provides different ways and issues while combining the machine learning approaches with the text. During the course of the survey, various publications in the field of synchronous application of machine learning in text analytics were searched and studied. Classification framework is proposed as the contribution of machine learning in text analytics. A classification framework represented the various application areas to motivate researchers for future research on the application of two emerging technologies.

KEYWORDS

1. INTRODUCTION

Big Data is a vast collection of a structured, semi-structured and unstructured data. Unlike the traditional data, requirements of big data applications are complex and advanced algorithms. Big data requires more storage, processing and computing capacity than conventional databases. Various definitions of Big data given by different data scientists include 5Vs namely Volume, Velocity, Variety, Value and Veracity. Big data seeks its applications in almost all the fields that includes healthcare (Azam et al., 2017), banking sector, e-commerce, software industries (Krimpmann and Stühmeier, 2017), transportation, manufacturing (Rahpeyma and Zarei, 2018) and fraud detection (Kayyali et al., 2013). Internet of Things (IoT) is a big data application used to track the vehicle positions with sensors (Chen et al. 2014). Supervision of employees and optimization of delivery routes in e-commerce companies makes use of IoT based big data applications. Concept of big data analytics is also applied on the complex water supply networks to find leakages and unauthorized connections, community detection in social networks (Babers and Hassanien, 2017). While working with big data, several issues may be encountered. Big data management, cleaning or preprocessing Big data, large volume of data, high rate of data streaming, different formats of data, uncertain data, fast retrieval of data, diversity of meaning and security of data are some of the significant challenges that makes the

analysis of big data an area of research. These challenges are big barriers for extracting information out of such a large amount of structured, semi-structured and unstructured data.

Thus, some advanced technologies, algorithms and analytical techniques are required to get efficient results. Big data Analytics deals with the advanced analysis of big data to find relationship among the explored data and to predict future observations. Big data analytics is a field that came across to solve the challenges related to complex nature of big data with real time heterogeneous datasets. Heterogeneous datasets constitutes of structured, semi-structured and unstructured data present in audio, video, image and text formats.

Text analytics is a field that deals with the analysis conducted on textual form of big data. Textual data sources include social network feeds, email, blog and online reviews. Text analytics helps in conversion of large volume of text into summarized form for more efficient multi-attribute-based decision making constitutes the significant part of decision science (Azadfallah, 2018). Machine learning has found its relevance in analyzing the textual data. It is used to mitigate the various challenges faced in text analytics and provide the improved analysis.

In the subsequent section, briefly explains the literature review that motivated the authors to conduct the research followed by a roadmap on the basis of which this research review is carried out. Afterwards, a framework is provided that explains the classification of the research review. Exploration and implications of different research queries is done in which findings from the different papers included in the survey are discussed. Finally, conclusion along with some future perspectives in the area of text analytics using machine learning is provided.

2. LITERATURE REVIEW

Huge amount of work has been done on text analytics with the usage of machine learning. The significance of using machine learning techniques was realized in the field of text analytics as it provides the better analysis of text by uncovering the hidden information. Deep learning became an integral part of many software applications and services. Scalability of the neural network with large data can be improved using deep learning.

Zhang et al. (2015) and Lai et al. (2015) provided the exploitation of deep learning based models for better text classification. Korde and Mahender (2012) discussed different algorithms used for text classification. Zhang and Wallace (2015) experimented for classification using various deep neural networks. Hughes et al. (2017), Islam and Zhang (2017) and Islam and Zhang (2017b) applied the deep networks for classification in healthcare.

Liu et al. (2003) proposed a feature selection method to improve clustering of the text. Shah and Mahajan (2012) discussed the different clustering methods. Ramanujam and Kaliappan (2016) discussed about summarization of documents using time stamping. Alguliev et al. (2013) proposed an evolutionary method for document summarization. Eidelman et al. (2012) distinguished the global and local topic modeling using the machine learning models. Mansouri et al. (2008) proposed an approach for Named Entity Relationship (NER) based on support vector machine. Steinberger and Jezek (2004) and Liu et al. (2004) introduced the machine learning based approaches for finding the semantic relevance in the sentences. Wu et al. (2010) exploited machine learning approaches for fetching keywords from the text for tagging. Huang et al. (2015) discussed the performance of different variants of deep learning framework in data tagging. Al Sallab et al. (2015), Li and Chen (2014), Shen et al. (2014) and Rosenthal et al. (2017) provided the classification of sentiments using deep learning models.

The work carried out in the field of text analytics and machine learning inspired the authors to provide this survey to elaborate all the tools, techniques and machine learning algorithms for various types of analytics on the textual data. In this comprehensive study, authors provide the broad view of the research done in the field and simplify the large data to get the relevant information. A

24 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/article/comprehensive-contemplation-of-probabilistic-aspects-in-intelligent-analytics/240617

Related Content

Cloud Computing Diagnosis: A Comprehensive Study
N. Krishnadasand R. Radhakrishna Pillai (2013). *Principles, Methodologies, and Service-Oriented Approaches for Cloud Computing (pp. 1-18).*
www.irma-international.org/chapter/cloud-computing-diagnosis/74223

E-Business in Education: The Case of Delta State University
Edwin Iroroeavwo Achugbue (2014). *Handbook of Research on Demand-Driven Web Services: Theory, Technologies, and Applications (pp. 356-375).*
www.irma-international.org/chapter/e-business-in-education/103679

Hydrological Drought Forecasting Using Modified Surface Water Supply Index (SWSI) and Streamflow Drought Index (SDI) in Conjunction with Artificial Neural Networks (ANNs)
Raphael M. Wambua (2019). *International Journal of Service Science, Management, Engineering, and Technology (pp. 39-57).*
www.irma-international.org/article/hydrological-drought-forecasting-using-modified-surface-water-supply-index-swsi-and-streamflow-drought-index-sdi-in-conjunction-with-artificial-neural-networks-anns/233606

A Novel Hybrid Optimization-Based Approach for Efficient Development of Business-Applications in Cloud
Soumia Zertal, Mohamed Batoucheand Zakaria Laboudi (2020). *International Journal of Information Systems in the Service Sector (pp. 14-35).*
www.irma-international.org/article/a-novel-hybrid-optimization-based-approach-for-efficient-development-of-business-applications-in-cloud/262142

A Case Study of Cloud Computing Service Models for the General Computing Environment in a University
Ho Yeon Kang, Jong Yun Leeand Seo-Young Noh (2022). *International Journal of Cloud Applications and Computing (pp. 1-17).*
www.irma-international.org/article/a-case-study-of-cloud-computing-service-models-for-the-general-computing-environment-in-a-university/297096