Text classification using deep learning techniques: a bibliometric analysis and future research directions

Gaurav Sarin
Delhi School of Business, Delhi, India
Pradeep Kumar
Indian Institute of Management Lucknow, Lucknow, India, and
M. Mukund
VDX.tv, New Delhi, India

Abstract

Purpose – Text classification is a widely accepted and adopted technique in organizations to mine and analyze unstructured and semi-structured data. With advancement of technological computing, deep learning has become more popular among academicians and professionals to perform mining and analytical operations. In this work, the authors study the research carried out in field of text classification using deep learning techniques to identify gaps and opportunities for doing research.

Design/methodology/approach – The authors adopted bibliometric-based approach in conjunction with visualization techniques to uncover new insights and findings. The authors collected data of two decades from Scopus global database to perform this study. The authors discuss business applications of deep learning techniques for text classification.

Findings – The study provides overview of various publication sources in field of text classification and deep learning together. The study also presents list of prominent authors and their countries working in this field. The authors also presented list of most cited articles based on citations and country of research. Various visualization techniques such as word cloud, network diagram and thematic map were used to identify collaboration network.

Originality/value – The study performed in this paper helped to understand research gaps that is original contribution to body of literature. To best of the authors' knowledge, in-depth study in the field of text classification and deep learning has not been performed in detail. The study provides high value to scholars and professionals by providing them opportunities of research in this area.

Keywords Data mining, Text analytics, Classification, Deep learning, Bibliometric analysis

Paper type Research paper

1. Introduction

Text classification (TC) is the process of dividing a specific text into organized groups from an unstructured data set by assigning labels to various text units. It is one of the classical problems in the natural language processing (NLP) domain. Based on the content, text classifiers use NLP to analyze the text automatically and subsequently determine a suitable set of predefined labels. TC has become a crucial part of businesses due to its ability to get remarkable insight from unstructured data and gradually assist businesses in making rational and comprehensive decisions about their future strategies for a product or a service. Some of the most prominent use cases of TC include sentiment analysis, which is a handy tool for understanding the polarity of a text. It is helpful to businesses in familiarizing themselves with the perception of customers about their brand or a specific product. Language detection is another application of TC that helps in detecting the language of a given text. Businesses that have a user base across the globe use language detection to help them with their requests.