Big Data Text Mining Techniques in Journalism and Communication

Miao Yuxin[1,a*]

[1] *Yunnan University Of Finance And Economics, Kunming, China*
[a*] *2727158656@qq.com*

ABSTRACT

With the rapid development of big data, data mining, artificial intelligence and other information technologies, text mining technology is increasingly widely applied in the field of humanities and social sciences. This paper firstly studies the value of data in news communication and the significance of text mining. Secondly, it constructs the news big data mining process based on deep learning. Finally, it illustrates the specific application of text mining technology in semantic network analysis, subject model and emotion analysis through cases.

Keywords: *Big data, Text mining, Deep learning, News communication*

1. INTRODUCTION

With the rapid development of information technology such as big data, Internet of Things, machine learning and artificial intelligence, human beings have entered the era of digital existence, and the various behaviours of individuals constitute massive and real big data[1-2]. With the technical assistance of computers, text mining technology based on big data has become a popular research in the field of natural science and humanities and social science[3-4]. In the news production process, data mining technology can deeply participate in news discovery, information collection, editing and processing[5]; in the news distribution platform, data deep mining, through machine learning algorithms to recommend relevant content to users, bringing a new pattern of news industry[6]; in the user platform and news terminal In the user platform and news terminal, all kinds of using devices and scenes will be the user's media entrance, massive data will be stored in the cloud, and the era of Internet of everything will come[7].

At present, the data mining technology is applied to more of the natural sciences, and related research in the field of journalism and communication is less, therefore, this paper first puts forward the data in the news dissemination value and the significance of text mining, second, build the news big data mining based on the deep learning process, and finally, by case discusses the concrete application of data mining technology in journalism and communication, namely, semantic network analysis, theme model, sentiment analysis.

2. THE VALUE AND SIGNFICANCE OF DATA IN NEWS COMMUNICATION

2.1. Data Can Be Presented Directly To Information

Data is traditionally referred to as "educated numbers", and information itself has the ability to eliminate uncertainty and therefore convey information directly and persuasively. Especially during COVID-19, by following up the confirmed cases and suspected cases in different regions every day, the audience can intuitively understand the epidemic situation and trend in different regions according to the data, and make good prevention and control measures and travel plans. At this time, news is data, and through these data, the description is more accurate than words and words.

2.2. Data Can Enhance The Persuasiveness Of Content

When presenting important ideas, "speaking with facts" is the basic principle, while speaking with data is an important support for facts. Real data can make news more convincing. During COVID-19, people can get a clearer understanding by changing the number of confirmed cases every day, so that they can consciously pay attention to the epidemic and actively cooperate with the epidemic prevention work.

2.3. *Mining Data Can Reveal Deep Meaning*

The value of isolated data may be small, and the real value of data often needs to be mined. The mining of news data can reveal the deep meaning. Data mining is the analysis of news and information, traditional content analysis was conducted by means of news text manual coding, analysis the text content, then reasoning, judgment, and enter the era of big data, with the traditional artificial content analysis is not only time-consuming, and inefficient, if the use of data mining, machine learning algorithm for data pre-processing, the core of steps, such as mining operation, statistical analysis, visualization can effectively extract, depicting information, visualization, the content of the spread of potential information and dig deep core information. Table 1 compares the differences in content analysis and text mining reapplication. Through comparison, it is found that text mining has more advantages in the processing of massive network data.

Table 1. Comparison of content analysis and text mining in terms of application differences

Contrasting items	Content analysis	Text mining
Sample size	Small sample	Large sample
Sampling method	Random sampling	Convenience sampling or random sampling
Main measurements	Attitude	Attitudes and behaviours
Unit of analysis	Articles	Phrase or symbol
Coding	Artificial	Man, machine

3. NEWS BIG DATA MINING BASED ON DEEP LEARNING

3.1. *Typical Deep Learning Models*

Deep learning was proposed by Professor Hinton of the University of Toronto in 2006. The emergence of deep learning has facilitated the further development of neural networks, which include input, hidden and output layers, to approximate complex functions as closely as possible through this multi-layer neural network structure. Deep learning models mainly include deep feed-forward networks, convolutional neural networks, recurrent neural networks, recursive neural networks, etc. Table 1 shows the characteristics and applicability of deep learning.

Name	Advantage	Disadvantage	Scope of application
Deep feedforward network	Good at extracting features	Not suitable for labeled training samples	Feature recognition; Data classification; Capture data relevance
Convolutional Neural Network	Good at processing image data	A large number of class markers are required	Text classification; Image Data Extraction
Recurrent Neural Network	Good at handling problems related to sequence data	Training difficulties	Regression; Text data processing
Recursive Neural Network	Easy to realize the functions of optimizing calculation	Need a lot of manual tagging	Prediction analysis

Figure 1 Characteristics and scope of application of deep learning model

3.2. *Text Mining Process Of News Big Data Based On Deep Learning*

The deep learning-based news big data mining process includes the acquisition and processing of news big data, which is divided into unsupervised deep learning and supervised deep learning according to the objectives or problems to be solved, and then selecting a suitable deep learning model, evaluating the model to reduce the number of iterations and iteration time as much as possible, and determining the best model through the corresponding evaluation criteria. The process of deep learning based news big data mining is shown in Figure 2.
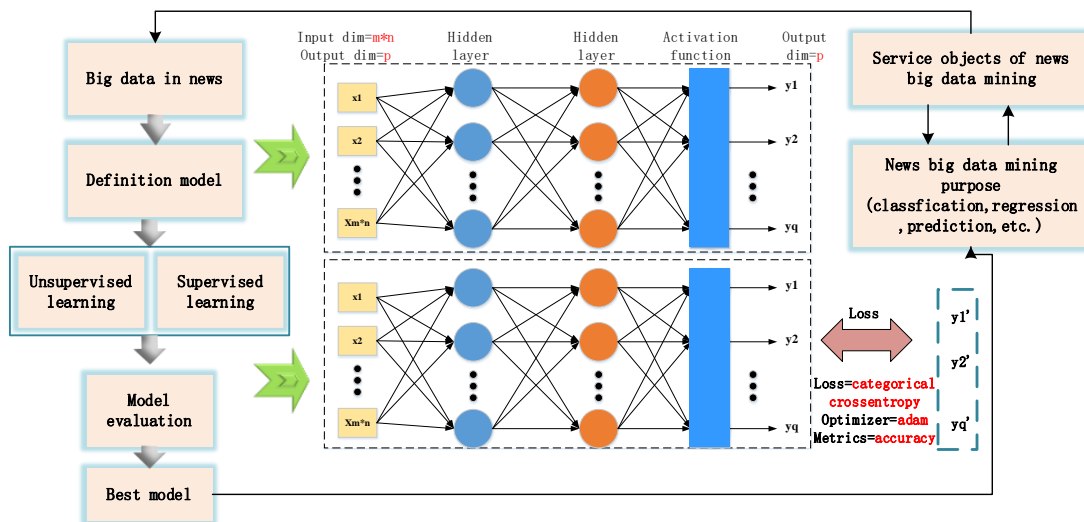

Figure 2 News big data mining process based on deep learning

3.2.1. Unsupervised Deep Learning

Unsupervised deep learning does not have training samples in advance. The purpose of unsupervised deep learning is to seek hidden structures in big news data, which requires the machine to search for potential category rules hidden in the data through clustering algorithm. Unsupervised deep learning is widely used in subject models. Topic model algorithm the disorder of the document as a series of words, by putting the word frequency attach weights, and then to feature extraction, feature extraction module as shown in figure 3, the main body for the encoder, decoder network, in the encoder network, convolution layer C1 extract rough, RC2 and deep RC3 convolution are extracted, and decoder in the network to join attention mechanism, by adding SE the convolution layer module, to a constant weight for each output channel prediction, emphasizes the effective information, inhibit invalid information, combined with latent dirichlet distribution algorithm or other algorithms, Unsupervised learning is carried out to obtain the topic probability of each document. For example, in a document, feature extraction, training, convergence and so on frequently occur, and the topic model will calculate the existence of topics such as machine learning and data analysis in the document.
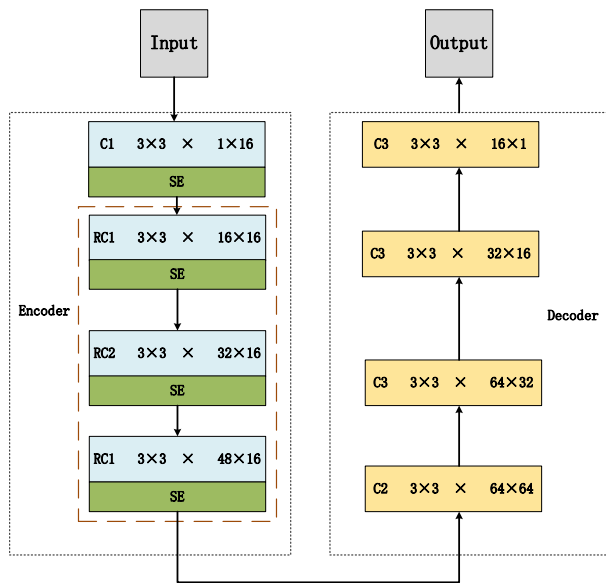

Figure 3 Encoder-decoder network structure

3.2.2. Supervised Deep Learning

In the processing of news media data, the supervised deep learning application is born due to the machine's lack of consideration of context, abbreviations and semantic ambiguity, etc. In supervised deep learning, the training set needs to be manually coded, and the feature set and content label of all documents are trained to form a classifier, and the feature classification is carried out based on the training set. Supervised deep learning has

two advantages. On the one hand, when a classifier training is completed, it can quickly analyze a large number of texts with high efficiency; on the other hand, the accuracy of the manually coded training set can be guaranteed and the results of machine learning can be evaluated at any time.

4. THE SPECIFIC APPLICATION OF TEXT MINING TECHNOLOGY IN JOURNALISM AND COMMUNICATION

4.1. Semantic Network Analysis

Semantic network analysis (SEA) uses words and sentences as nodes in the network, studies the role of different words and sentences in the overall network by quantifying network relations, and then presents the connections between texts. This method takes high-frequency words as nodes, and constructs semantic network analysis high-frequency phrase library to translate the mental map of texts. Such as 1789 to 2013, the American presidential inaugural speech in the high frequency words, the higher the frequency of words, the greater the size, through the establishment of the semantic web, the classification, can be concluded that different periods of hot topics, such as human rights, family, employment, culture, etc., can understand during the presidency, will focus on issue of concern.

4.2. Topic Model

The topic model classifies the relationship among words, text and topics through mining and analysis. Christian Pentzold et al. conducted unsupervised deep learning and subject analysis on more than 7,000 tweets from the 2012 US presidential election, and the results showed that the unsupervised deep learning algorithm could effectively identify the subjects of tweets[8]. In a study on fire conducted by McCaffrey et al., researchers used 14,000 tweets to conduct topic modeling and analyzed more than 20 topics in total[9], so as to fully understand the basic situation of fire, including the area, scope, air quality and disaster situation.

4.3. Sentiment Analysis

Emotion analysis refers to the automatic classification of texts through the natural language processing process to find out the user's attitude towards something, whether it is positive or negative, positive or negative, so as to identify the user's views and attitudes towards something or someone. In the study of public communication, we can judge whether the netizens' views on major issues of the Internet are positive, neutral or negative through emotional analysis. In the research of brand communication, we can make an emotional analysis of the comments to judge the evaluation of the products and services. In political communication, the

political tendency of residents can be judged through the emotional analysis of users' tweets.

5. CONCLUSION

Through the above analysis and research of text mining technology is widely applied in the journalism and communication, this paper first expounds the news data has great value, and compared the content analysis and text mining, shows the text mining technology is of great significance in the field of journalism and communication, second, build the news big data mining based on the deep learning process, and expounds in detail without supervision depth depth of learning and supervised learning, studies show that text mining technology to analysis of the semantic network, themes, models and emotional analysis to obtain the very good application in the news spread.

REFERENCES

[1] Erna Qi & Zongjun Wang(2019). Data mining and visualization of data-driven news in the era of big data. 22(1), 10333-10346.

[2] Nikki Usher(2020). News cartography and epistemic authority in the era of big data: Journalists as map-makers, map-users, and map-subjects. 22(2), 247-263.

[3] Ji-Wan Lee, Chung-Gil Jung & Seong-Joon Kim(2019). The relationship among meteorological, agricultural, and in situ news-generated big data on droughts. 98(2), 765-781.

[4] Kim Tae-Jong(2020). COVID-19 News Analysis Using News Big Data : Focusing on Topic Modeling Analysis. 20(5), 457-466.

[5] Arslan Ali Raza, Asad Habib & Muhammad Javed(2019). Semantic Orientation Based Decision Making Framework for Big Data Analysis of Sporadic News Events. 17(2), 367-383.

[6] W. Ken Redekop(2018). Fake news, big data, and the opportunities and threats of targeted actions. 7(2), 113-114.

[7] Reilly Susan(2017). The Need to Help Journalists with Data and Information Visualization.. 37(2), 8-10.

[8] Christian Pentzold & Lena Fölsche(2019). Imagining big data: Illustrations of "big data" in US news articles, 2010–2016. 21(1), 139-167.

[9] Sonya Sachdeva & Dexter Locke(2016). Social media approaches to modeling wildfire smoke dispersion: spatiotemporal and social scientific investigations. 20(8), 1146-1161.