International Core Journal of Engineering Volume 8 Issue 5, 2022
ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

A Survey of Evaluating Credibility based on Deep Learning in Online Social Networks

Jinglong Yao, and Ling Xing

College of Information Engineering, Henan University of Science and Technology, Luoyang 471023, China

Abstract

The rapid development of online social networks makes the dissemination of information more rapid and convenient, but the dissemination of false information may have a negative impact on the majority of Internet users, resulting in difficulties in platform management, social unrest, and even the development of the country in serious cases. Therefore, it is significant to evaluate the credibility of the content in online social networks. In the era full of data, models such as recurrent neural networks and convolutional neural networks in deep learning technology have excellent data mining ability. Through feature mining of existing data, we can predict the credibility of new data. To solve the challenge of credibility evaluation, this paper first defines the credibility evaluation of online social networks, then systematically summarizes the credibility evaluation methods in recent ten years, and expounds on three kinds of credibility evaluation methods. Secondly, the performance of the classified reliability evaluation technology is evaluated by using the relevant evaluation indexes. Finally, through the analysis and summary of the existing work, this paper further puts forward the possible research direction of online social network credibility evaluation technology in the future.

Keywords

Online Social Networks; Credibility Evaluation; Deep Learning; Data Type.

1. Introduction

With the rapid development of the Internet, OSN has developed from a small-scale exploratory behavior in the last century to a global explosive growth of users[1]. Web 3.0 technology has changed the simple way of using the Internet through user interaction, with more emphasis on taking users as the main body. The popularity of mobile terminals enables users to participate in online social activities anytime and anywhere and allows users to transfer value or property in the form of data, so as to obtain more wealth and higher popularity[2]. Online social networks provide users with opportunities to share and exchange information to help users meet their own needs, but they are also abused by malicious users to perform some bad behaviors, which play a negative role, such as spreading rumors, spreading false emails, and so on. Social platforms have become one of the main ways to spread rumors. During the covid-19 epidemic, malicious users spread false information about the epidemic. Such information has a high degree of attention and dissemination rate[3], causing people to panic and causing a serious impact on society and the country. Therefore, how to evaluate the credibility of data in OSN and avoid the negative impact of malicious data has become the focus of current research and attracted extensive attention of researchers.

The essence of OSN credibility evaluation is to find out the false content or the users who send the false content. The solution of this problem is of great significance to many fields, mainly reflected in the following aspects:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

1) Information security

With the convenience of social networks, malicious users gradually spread to major social platforms[4]. Research on credibility evaluation can help the platforms or other users identify and mark malicious users as soon as possible. In serious cases, it can limit their IP address, mark false content in time and remind normal users, which greatly promotes network information security.

2) Public opinion analysis

OSN has become the main place for the generation and dissemination of public opinion. At the same time, the role of network public opinion in social development has become more and more important, which is an important part of social public opinion[5]. Therefore, by studying the credibility-related content in OSN, a network public opinion information analysis system can be constructed, and social public opinion trends can be grasped in time, which plays an important role for government departments to make scientific and democratic decisions and prevent emergencies.

3) User authentication

At present, popular social platforms support authentication of accounts. These accounts authenticated by the platform are regarded as having public interests and more credibility than the behavior of other unauthenticated users[6]. The message released by the verified user is more authoritative and can provide reference for other users, which can effectively reduce the negative impact of false information on users[7].

4) Increase user stickiness

After credibility evaluation, the platform can identify or delete the untrusted content. Good platform order can effectively reduce users' concerns about the security of online social behavior, so as to attract more users to participate in online social networking on their platform, increase user viscosity, drive the innovation and development of online social platform and strengthen the stability of the platform.

In the era of big data, related methods of deep learning are very important, especially in the field of data mining, and the related research on credibility evaluation is based on the analysis of various data in OSN. The credibility evaluation based on deep learning uses data to intelligently solve the credibility evaluation problem. Through the analysis and mining of a large number of data by computers with high-precision and intensive computing power, various fine-grained features of the data are obtained. On this basis, the unknown data are predicted to obtain the credibility of the content, This complete process can be called a model. OSN is full of data, and it is growing at a crazy rate. Applying deep learning related algorithms to OSN data can effectively provide solutions to problems related to credibility evaluation.

At present, a series of important achievements have been made in the research on the credibility evaluation of online social networks based on deep learning. From the perspective of data types, this paper introduces the concepts, models, classification and other technologies of online social network credibility evaluation, analyzes the performance evaluation of existing methods in online social network credibility evaluation in detail, and discusses the future research direction of online social network credibility evaluation by comparing and analyzing the advantages and disadvantages of existing methods.

2. Basic Concept

2.1 Definition of Credibility Evaluation

At the initial stage of research in the field of credibility evaluation, Castillo et al. first proposed to use the classifier method in machine learning to evaluate the credibility of news content in twitter[8], and selected fifteen appropriate features for classification according to the data type. Most of the subsequent studies are based on this, but due to the different evaluation objects or evaluation types, the research naming methods of credibility evaluation are often different. This paper integrates and analyzes the relevant research literature, and defines the concept of credibility evaluation in OSN:

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

OSN exists dynamically based on content or users and using a variety of connection or interaction modes. Credibility evaluation technology is to analyze or verify the content or users in OSN and identify the parts that may cause adverse social impact. This technology can be simplified into a mathematical model, $T=\{t_1,t_2,t_3,\cdots,t_i,\cdots,t_n\}$ is used to represent the topic-level in the social platform, $E=\{e_1,e_2,e_3,\cdots,e_i,\cdots,e_n\}$ is used to represent the event-level in the social platform, $M=\{m_1,m_2,m_3,\cdots,m_i,\cdots,m_n\}$ is used to represent the message-level in the social platform, and $U=\{u_1,u_2,u_3,\cdots,u_i,\cdots,u_n\}$ is used to represent the user-level in the social platform. The inclusion relationship between different levels is shown in Figure 1. Define a collection of content that may cause adverse effects $\{t^x|t^x\in False\ theme\}$, $\{e^x|e^x\in False\ event\}$, $\{m^x|m^x\in False\ message\}$ and $\{u^x|u^x\in False\ user\}$, mining the content of the collection and the relationship between them. Common credibility evaluation tasks include content credibility evaluation, user credibility evaluation, content credibility ranking, rumor detection, public opinion analysis, Navy detection, spam detection, online bullying detection, etc. The relationship between the credibility assessment tasks is shown in Figure 2.

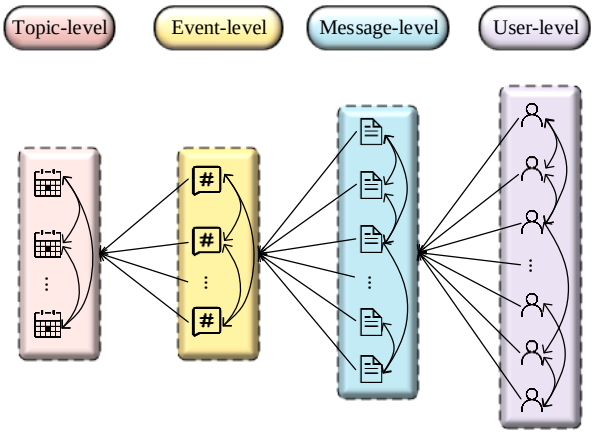

Figure 1. The hierarchy of credibility assessment

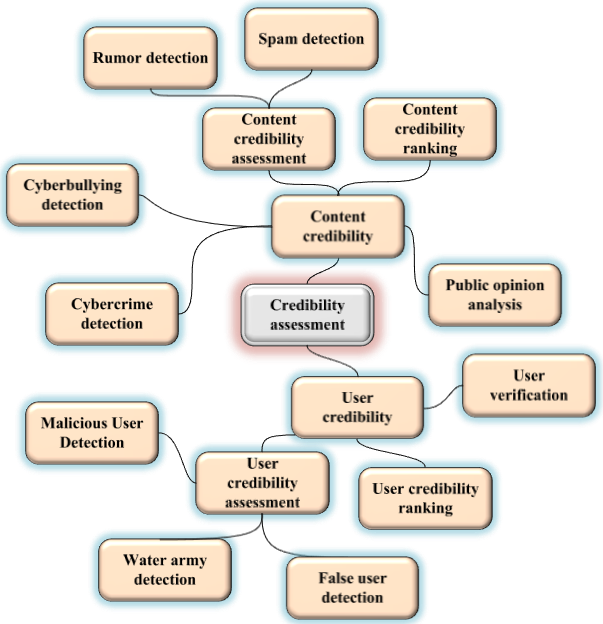

Figure 2. The relationship between credibility assessment tasks

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

2.2 Related Models of Deep Learning

The goal of deep learning is to learn its deeper representation from the original data, to find the feature representation required by the credibility evaluation model. This is achieved by using deep neural networks(DNN), which is composed of multiple layers of trainable layers. The original data is input into DNN in a specific data format[9]. As the data flow passes through each layer, it gets more and more abstract representations. In this way, the original input data is finally represented and output by high-dimensional feature vectors, which is very effective for distinguishing different modes.

Researchers have developed a series of deep neural networks for various application scenarios, such as recurrent neural networks(RNN), convolutional neural networks(CNN), automatic encoder and so on. Although the architecture of DNNs is diverse, most DNNS are composed of basic components, such as input layer, output layer, hidden layer, pool layer, full connection layer, convolution layer, etc.

2.2.1 Recurrent Neural Networks

Because the data in OSN is full of serialization, and the loop structure of RNN is just suitable for processing the serialized data information. Independent analysis of a word in a sentence is not enough, but needs to deal with the whole text sequence connected by these words. In addition, the content in OSN is a kind of time series data closely related to time. In addition to the inherent attributes of various features, the dynamic changes of features over time should also be considered in the analysis. Therefore, RNN is used to model and analyze the text series data in the time dimension to obtain the implicit features of their context information over time, So as to better deal with different tasks. It can be seen that the data with sequence characteristics conforming to time sequence, logical sequence or some other specific sequence is suitable for processing based on RNN.

However, there is a fatal flaw in the native RNN. During data analysis and processing, it is reasonable that the hidden characteristics of data at each time depend not only on the input at that time, but also on the hidden characteristics fed back at the previous time. For example, the sequence of data is very long, and the calculation at time t depends on the deep calculation results at time t-1, The result of t-1 time depends on the deep calculation result of t-2 time. It can only be processed word by word, and parallel operation cannot be performed. This phenomenon is called gradient disappearance or gradient explosion[10]. To solve this problem, researchers added three control units to the structure of RNN: input control, output control and forgetting control. In this way, the neural network will selectively input, output or forget some data during operation, and only select the data content that has a positive effect on the whole model. This improved model is called Long Short-Term Memory.

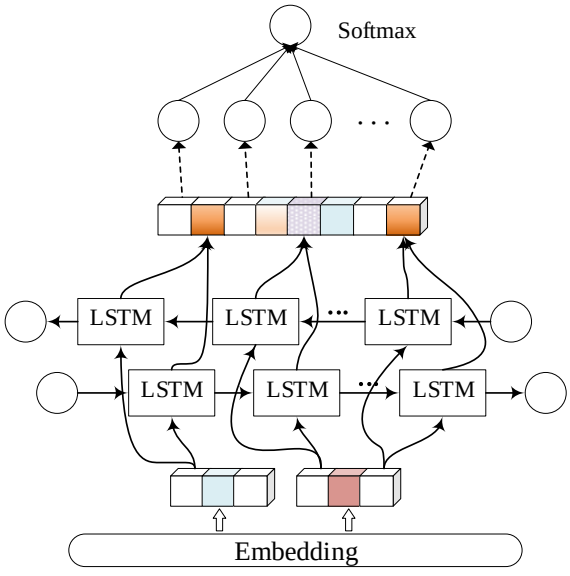

Figure 3. A framework for credibility evaluation of Bi-LSTM

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

LSTM can effectively obtain the semantic features with long distance between words in the text, but in one-way LSTM, LSTM unit only considers the information of the former unit and ignores the information of the latter unit. In order to solve this problem, Bi-directional Long Short-Term Memory (Bi-LSTM) can be used to process text coding, obtain the dependency between context words and obtain more effective information, it contains a forward LSTM layer and a reverse LSTM layer. Figure 3 shows the credibility evaluation task framework based on Bi-LSTM.

2.2.2 Convolutional Neural Networks

CNN was first applied to computer vision, and then began to be applied to various tasks in natural language processing, including related research on credibility evaluation. Researchers have applied the N-Gram model to NLP because its feature is to analyze the local statistics of the text, and the CNN is more like an enhanced version of the N-Gram model, which handles such a task in a low-dimensional space. The difference between them is that the N-Gram model directly counts the probability that different N words are combined together, while CNN obtains the weight of the combination between different words by learning, and achieves the final purpose of the entire model through the weighted sum.

The classic architecture of CNN consists of convolutional layers, pooling layers and fully connected layers. Taking the text content of the credibility evaluation application as an example, when the text passes through each convolutional layer, a set of functions are applied to obtain a new feature map.

When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. When receiving the paper, we assume that the corresponding authors grant us the copyright to use the paper for the book or journal in question. When receiving the paper, we assume that the corresponding authors grant us the copyright to use. As shown in Figure 4, We can express K filters and their corresponding deviations in the m-th convolution layer as $W=\{W_1,W_2,W_3,\cdots,W_k\}$ and $B=\{b_1,b_2,b_3,\cdots,b_k\}$, Then, the k-th text feature of layer m is obtained, which is X_k^m, $X_k^m=f\left(W_k*X^{m-1}+b_k\right)$, Where X^{m-1} is the output of the previous layer, $*$ represents convolution operation, $f(\bullet)$ is the activation function. Using pooling operation can reduce the dimension of feature mapping. With the deepening of training level, learn advanced text representation. Finally, several full connection layers are used for the credibility evaluation task and output through the softmax function.

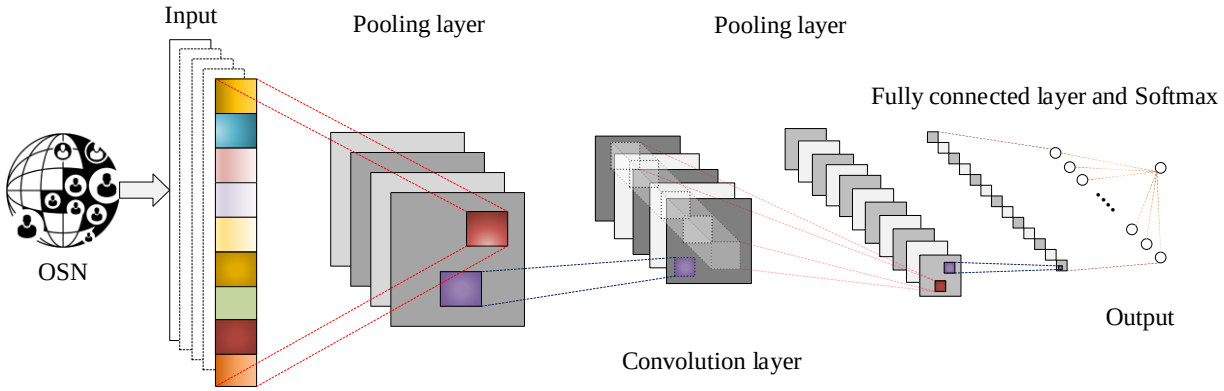

Figure 4. A framework for credibility evaluation of CNN

2.2.3 Attention Mechanism

At present, the attention mechanism has become one of the most important concepts in the field of deep learning research. It draws on the specific brain signal processing mechanism in human vision. When there is often something you want to see in a certain area, if a similar scene appears again, humans tend to focus on this part. This is a way for humans to quickly select high-value information

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

from massive amounts of information using limited processing resources. The core of the attention mechanism is to select the information that is more critical to the current task objective from a large number of information, which improves the efficiency and accuracy of perceptual information processing.

When applying the attention mechanism, it is necessary to distinguish the features of different data, for example the source data is marked as k_i, the incoming query data is marked as q_i, the correlation is calculated through the score function f to obtain the energy score e, then, the attention weight α is obtained by mapping the attention distribution function g. When we calculate the text context vector, we need to introduce a new data feature v_i. after weighting them, we can get the weighted value z_i, and then we can get the context vector c. The architecture of the attention computing model described is shown in Figure 5. Reasonable application of attention mechanism can greatly improve the performance of the model.

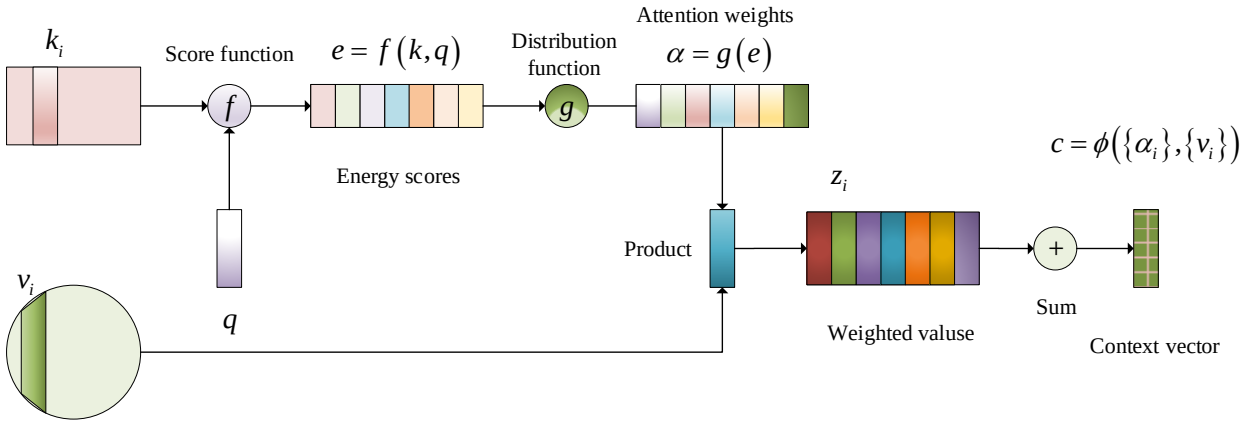

Figure 5. The architecture of Attention mechanism model

2.3 Performance Evaluation Index.

At present, most of the relevant researches on credibility evaluation focus on datasets with a certain amount of data. During learning, the training set and the validation set are used for feature mining and parameter learning, and then the data obtained by the model is used to evaluate the data in the test set. The model ability is reflected by the evaluation index. The higher the accuracy, precision, recall and F1 in the evaluation index, the better the model ability for credibility evaluation. The evaluation indexes used are calculated according to the confusion matrix in table 1.

Table 1. Confusion matrix

Confusion matrix		Prediction category	
		1(True)	-1(False)
Actual category	1(True)	TP	FN
	-1(False)	FP	TN

The calculation formulas of accuracy, precision, recall and F1 are as follows:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{1}$$

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$\mathrm{Re}call = \frac{TP}{TP + FN} \tag{3}$$

$$F1 = \frac{2 \times precision \times recall}{precision + recall} \tag{4}$$

In the research of credibility evaluation task, the proposed method hopes to identify all malicious data as much as possible, and what is reflected in the evaluation index is that the recall rate should be high. In some cases, there will be some contradictions between accuracy and recall rate. When one is higher, the other will be lower. Of course, the real data can not be identified as false, which will increase the later investigation work, and may also miss the investigation and cause impact. Therefore, the accuracy of credibility evaluation should also reach a certain level. Considering these problems, most researchers usually use the harmonic average F1 of accuracy and recall rate to measure the relationship between the two values in credibility research. The higher the F1 value, the better the overall effect of the evaluation method or model.

In addition, availability and computational overhead are also important metrics for evaluating models.

3. Research Classification

After nearly ten years of continuous development and improvement, the social platform has generated a large amount of data and various forms. Different types of data correspond to different credibility evaluation processes and evaluation results. Therefore, it is necessary to evaluate the operational capabilities and evaluation results according to the actual situation. The needs determine the appropriate evaluation model, and then use the appropriate data to optimize the model accordingly. According to the common deep learning-based credibility evaluation model, the data types are divided into three categories, namely content data, user data and dissemination data. Next, this paper will introduce the research status of multi-content credibility assessment at home and abroad one by one according to different data types.

3.1 Content Data

Content is the main embodiment of data dissemination in OSN, including text content, audio and video and other multimedia content, which can help us more intuitively carry out credibility evaluation tasks.

Ratkiewicz et al. created an extensible "truth" framework using meme (tags, links and reference features) in the content to realize the real-time analysis of meme propagation in social media through data mining and mapping[11], so as to detect malicious content. However, this method will be affected by the repeated words in a section of content, resulting in fluctuations in the results. Therefore, Qazvinian et al. uses relevant corpora to extract phrases in the text, and puts forward four content-based features according to the vocabulary mode and part of speech mode of the text content, which solves the problem of the influence of repeated words[12]. However, if there is no meme in the false information, the recall rate of this method will be low.

To mine more new features for credibility evaluation, Setiawan et al. added 17 new features to Twitter and 49 new features to Facebook for credibility evaluation through the investigation of previous relevant research literature[13]. Its research confirmed that among all the excavated features, some can improve the evaluation results, the other can not improve and even have a negative impact. Different feature combinations will also have a positive or negative impact on the evaluation results.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

Blind or subjective selection of features for combination is accidental and wastes time and manpower, Therefore, Singh J et al. used CNN and LSTM to automatically extract the features of the text content of tweets, thus creating a hybrid feature set[14]. On this basis, particle swarm optimization algorithm is used to find the best features from the established hybrid feature set. Particle swarm optimization algorithm is a population-based optimization algorithm, which selects the best feature subset from the established mixed feature set to obtain an optimal feature set. Singh A et al. focuses on the selection and optimization of classification models and parameters[15]. They propose a framework for the integration of Rhododendron search and machine learning models to automatically find the best combination of classification technology and its optimization parameters.

Most studies are based on feature selection or model optimization. Unlike other studies, Han et al. focuses on data sets[16]. They propose a data expansion paradigm, which effectively expands the existing rumor data sets by using the public large-scale unmarked data related to real-world events. Based on the limited labeled rumor source tweets, the unlabeled data under weak supervision are labeled by using semantic relevance. In fact, it is verified that enhancing semantic data has certain efficiency and effectiveness to overcome the scarcity and category imbalance of marked data in the existing public rumor data set.

The credibility evaluation technology based on content data is mainly evaluated through the content ontology spread in OSN, which is often more intuitive. However, with the continuous development of malicious counterfeiters, it will hide the falsity in the content through a variety of methods, resulting in the performance of the evaluation method based on content data being affected. Therefore, more and more methods of combining user data and disseminating data are used more.

3.2 User Data

User data mainly refers to the characteristics of users' personal information or the characteristics generated by users through some online social behaviors such as forwarding, comments, etc. users with real and perfect personal information have a higher degree of credibility when socializing. Through the investigation of real users, Gearhart et al. found that comments will bias the existing content information, resulting in the impact of content credibility[17]. Mendoza et al. analyzed the user information and behavior of sending earthquake related content after the Chilean earthquake[18]. It was found that the propagation mode of rumors in OSN was different from news, and OSN questioned rumors much more than news. It can be seen that user data plays an important role in the credibility evaluation task.

Qazvinian et al. constructs positive and negative user models to capture the historical behavior of users posting or forwarding. According to the user behavior in twitter, two features are proposed[12]. The user model is used to calculate the log likelihood ratio of the determined messages as the eigenvalues of the two features to identify rumors, which shows that the user's social characteristics have a positive impact on the evaluation of content credibility, but the method they proposed does not take into account the user's interactive behavior when socializing. Liang et al. extracts features according to the user's behavior after reading microblog, including the number of forwarding and comments, the proportion of questioning comments and the number of corrections, and jointly detects rumors in combination with other behavior features[19]. However, the imbalance between true and false data in training data leads to limited evaluation performance. Subsequently, Zhang et al. proposed to use the resampling mechanism to evaluate the credibility of commentators according to the user's behavior, the user's social relationship and the quality of user generated comments[20]. They evaluated the data set obtained by resampling and the data set obtained without resampling respectively, which confirmed the role of resampling mechanism in solving the problem of data set imbalance.

Different from the study of user behavior characteristics, Chen et al. focuses on the comment content, and analyzes public opinion by using the comment content generated by users as characteristics[21]. They divided comments into spam comments, subjective comments and objective comments, combined feature extraction methods with classification algorithms, and classified online comments

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

according to certain constraints. Because it involves semantic analysis, the results rely on a well-trained emotional dictionary.

The social information of users can effectively assist the credibility evaluation of data, but the social information is relatively superficial. If it can reflect the real thoughts and emotions of users, especially commentators, the relevant research characteristics of emotion analysis begin to appear in the credibility evaluation. In the research of credibility evaluation, the dissemination of content is driven by psychological and emotional factors, especially when a large number of users have a burst of negative emotions, it will aggravate the dissemination of false information. Therefore, emotion analysis is one of the important research points in credibility evaluation.

Ghanem et al. believes that there are certain differences in emotion between online news and social media. Therefore, they proposed an LSTM network with attention mechanism. The attention mechanism assigns a weight to each word vector result from the LSTM layer. The neural network only uses emotional features. Their purpose is to study whether emotional features can independently detect false news[22]. Experiments show that different types of false information and news have different emotional patterns, and emotion plays a key role in cheating readers.

To obtain dynamic emotional features, Li et al. puts the content in a time step and tries to apply the change of user emotion over time to the credibility evaluation task. However, due to the large gap between the representation space of content data and the representation space of highly abstract features such as semantics and emotion of information, the traditional classification effect is not good[23]. To solve this problem, they proposed a rumor detection method based on deep bidirectional gated loop unit. To capture the evolution of emotion of microblog event group over time, the model considers the forward and reverse sequences of information microblog flow along the timeline at the same time. The evolutionary representation of deep potential space including semantics and emotion learned by multilayer neural network is used to detect rumors.

Wang et al. added emoticons on the basis of text analysis and used automatic methods to build an emotion dictionary to capture users' fine-grained emotional responses to different events[24]. They proposed a two-step dynamic time series algorithm, which introduced emotional information in the segmentation process and retained the time span distribution information of microblog events, to obtain the characteristics of emotion changing with time. A two-layer cascade threshold recursive unit rumor event detection model based on emotion dictionary and two-step dynamic time series algorithm is proposed, which is an advanced algorithm in the field of reliability evaluation based on emotion analysis.

The credibility evaluation method based on user data can get the credibility of users or content according to the actual social situation of users. However, due to the privacy protection, it is difficult to obtain some data of users, so most of them should be evaluated according to the actual situation and the results of the method based on content data.

3.3 Dissemination Data

The information released by social platforms is a kind of time series data closely related to time. Ma et al. believes that previous studies only considered the inherent attributes of various features and ignored the dynamic changes of these features over time with the dissemination of content[25]. Therefore, they took the lead in establishing a dynamic time series structure to obtain the changes of features with time series in the process of propagation.

RNN has powerful information about the direction of energy modeling in time series. Ma et al. first introduced RNN into the credibility evaluation problem. Through the modeling and analysis of text sequence data in the time dimension, the implicit characteristics of the context information of false content changing with time are obtained[26]. Chen et al. believed that learning the time series representation of rumors is very important in early detection. They proposed a deep attention model CallAtRumors based on RNN. Different from the previous methods, this model can selectively learn the time representation of continuous tweet sequences and penetrate Soft-Attention into the cycle

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

process to collect different features with specific concerns at the same time, and generate hidden representations to capture the context changes of relevant posts over time. Based on these changes, we can predict the changes of false content over time, so as to point it out as soon as possible[27].

The traditional feature-based credibility evaluation method can capture rich user and text features, but it can not obtain the change information over time like deep neural network. Deep neural network has the advantage of processing serialized text and can obtain deep-seated change information, but some basic and important characteristics of false information, such as user characteristics, can not be used as the input of neural network. At the same time, training a robust neural network requires a large number of data sets and time, which is an important challenge for researchers in this field. Meng et al. combined the advantages of traditional feature detection and neural network detection, proposed a new feature change extraction framework. The framework uses sliding windows of different sizes to extract the change information of sensitive features over time[28]. They extracted three dynamic features and used the framework to capture the change information of these features over time. Finally, the evaluation effect of the framework is confirmed on two microblog data sets, and the framework can also be applied to early evaluation.

Feature-based credibility evaluation methods often only focus on static or planar features from content or users, while ignoring the impact of information dissemination structure. Xu et al. was the first to apply structural features to credibility evaluation. They proposed a rumor detection model that comprehensively considered the original tweet content, forwarding diffusion and user information. They used LSTM model to learn the representation of the original tweet, and introduced content attention mechanism to aggregate the keywords in the original tweet[29]. The LSTM model with time interval level is used to extract the dynamic characteristics of forwarding, and a propagation attention mechanism is proposed to pay attention to the forwarding with large amount of information in the propagation process. It is confirmed that the characteristics produced in the process of content dissemination can play a role in the research of credibility evaluation.

Huang et al. is the first researcher to use Graph Convolutional Network(GCN) to capture the user's behavior to model[30]. The model includes three parts: a user coder that uses graph convolutional network to model the user's attributes and behavior and obtain the user's representation; The structure of the rumor propagation tree is encoded into a vector, the propagation tree encoder bridging the content semantics and propagation clues, and the integrator integrating the output of the above modules to identify the rumor. On the basis of GCN, Bian et al. proposed a Bi-Directional Graph Convolutional Networks(Bi-GCN), which not only considers the mode of content dissemination in online social networks, but also considers its widely dispersed structure, and explores the characteristics of top-down and bottom-up dissemination methods[31]. Using the GCN of rumor propagation with top-down directed graph to learn the propagation mode of rumor. GCN with the opposite direction of rumor diffusion pattern is used to describe the structure of rumor diffusion.

Content dissemination usually includes temporal and spatial structural features. However, the existing methods based on DNN usually model the temporal structure and spatial structure respectively, and do not comprehensively model them as a whole. Huang et al. proposed a spatiotemporal structure neural network for rumor detection[32]. The network regards the spatial structure and temporal structure as a whole, models the propagation of content and detects rumor.

The credibility evaluation method based on communication data can make up for the lack of information in the credibility evaluation of content and users. The existing credibility evaluation research based on communication mainly combines data with communication structure and sequence with appropriate methods, which can often obtain better credibility evaluation performance.

4. Performance Evaluation

According to the evaluation indicators in Section 2.3 of this paper, the three types of methods are compared. Among them, the model performance analysis comparison is shown in Table 2. The analysis and comparison of the advantages and disadvantages of the models are shown in Table 3.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

Table 2. Performance analysis of OSN credibility evaluation models

Model type	Evaluation ability	Computational cost	Missing data	Difficulty in obtaining data
Content data	Medium	Medium	Low	Low
User data	High	High	Medium	High
Dissemination data	High	High	High	Medium

Online social network credibility evaluation has a wide range of applications and is a hot research direction in academic circles in recent years. This paper summarizes the research progress and research status of credibility evaluation, classifies the existing research results in this field according to the data characteristics, and introduces three online social network credibility evaluation technologies. It can be seen that various methods have their own characteristics. According to different characteristics and application requirements, their analysis and comparison results are listed in Table 2. From table 2, it can be seen that their application scope and performance are different. Through comparative analysis, these evaluation technologies have at least a certain evaluation ability, but their performance is different in terms of computational overhead, lack of data and difficulty of user data acquisition. For example, the evaluation ability of the method based on dissemination data and user analysis is relatively strong, but there are some problems of computing overhead and data loss. Table 3 makes a further comparative analysis on the credibility evaluation of online social networks, and gives the advantages and disadvantages of each evaluation technology.

Table 3. Comparison of advantages and disadvantages of OSN credibility evaluation models

Model type	Main advantages	main disadvantages
Content data	Data is easy to obtain and model implementation is simple.	Confusion in feature selection.
User data	Fewer features, easy to implement the model.	Ignore that user characteristics and content characteristics are not in the same dimension.
Dissemination data	Can relate to information other than data and its changing characteristics over time.	Ignore the internal connection between the propagation structure and semantic features.

5. Conclusion

From the perspective of deep learning method, this paper summarizes the research status of OSN credibility evaluation technology in the past ten years. At present, the methods of credibility evaluation are constantly developing. The proposed methods can help social platforms better provide users with network services, reduce the consumption of network resources, and reduce the negative impact on the society to a certain extent.

This paper first defines the credibility evaluation task of online social networks, and then describes in detail the comparison and analysis of existing research work in terms of model, calculation method, evaluation framework, research status and performance evaluation from three aspects. Due to the long-term nature of OSN, the credibility evaluation of online social networks is still a hot research field in the future, and there are still many key related issues that need in-depth and detailed research.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

References

[1] Samanta S, Dubey V K, Sarkar B. Measure of influences in social networks. Applied Soft Computing. Vol. 99(2021), p. 106858.

[2] Wang Y, Wang J, Wang H, et al. Users' mobility enhances information diffusion in online social networks. Information Sciences. Vol. 546 (2021), p. 329-348.

[3] Cinelli M, Quattrociocchi W, Galeazzi A, et al. The COVID-19 social media infodemic. Scientific reports. Vol. 10(2020) No. 1, p. 1-10.

[4] Latah M. Detection of malicious social bots: A survey and a refined taxonomy. Expert Systems with Applications. Vol. 151 (2020), p. 113383.

[5] Jia F, Chen C C. Emotional characteristics and time series analysis of Internet public opinion participants based on emotional feature words. International Journal of Advanced Robotic Systems. Vol. 17 (2020) No. 1, p. 1729881420904213.

[6] Paul I, Khattar A, Chopra S, et al. What sets Verified Users apart? Insights, Analysis and Prediction of Verified Users on Twitter. Proceedings of the 10th ACM Conference on Web Science. Amsterdam, 2019, p. 215-224.

[7] Vaidya T, Votipka D, Mazurek M L, et al. Does being verified make you more credible? Account verification's effect on tweet credibility. Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems. Chicago, 2019, p. 1-13.

[8] Castillo C, Mendoza M, Poblete B. Information credibility on twitter. Proceedings of the 20th international conference on World wide web. Salzburg, 2011, p. 675-684.

[9] Rusk N. Deep learning. Nature Methods. Vol. 13(2016) No. 1, p. 35-35.

[10] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. Vol. 6 (1998) No. 02, p. 107-116.

[11] Ratkiewicz J, Conover M, Meiss M, et al. Detecting and tracking the spread of astroturf memes in microblog streams. arXiv preprint arXiv. Vol. 1011 (2010), p. 3768.

[12] Qazvinian V, Rosengren E, Radev D, et al. Rumor has it: Identifying misinformation in microblogs. Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing. Edinburgh, 2011, p. 1589-1599.

[13] Setiawan E B, Widyantoro D H, Surendro K. Measuring information credibility in social media using combination of user profile and message content dimensions. International Journal of Electrical & Computer Engineering. Vol. 10 (2020) No. 4, p. 3537-3549.

[14] Singh J P, Kumar A, Rana N P, et al. Attention-based LSTM network for rumor veracity estimation of tweets. Information Systems Frontiers. (2020), p. 1-16.

[15] Singh A, Kaur M. Cuckoo inspired stacking ensemble framework for content-based cybercrime detection in online social networks. Transactions on Emerging Telecommunications Technologies. Vol. 32 (2021) No. 6, p. e4074.

[16] Han S, Gao J, Ciravegna F. Neural language model based training data augmentation for weakly supervised early rumor detection. Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. New York, 2019, p. 105-112.

[17] Gearhart S, Moe A, Zhang B. Hostile media bias on social media: Testing the effect of user comments on perceptions of news bias and credibility. Human behavior and emerging technologies. Vol. 2 (2020) No. 2, p. 140-148.

[18] Mendoza M, Poblete B, Castillo C. Twitter under crisis: Can we trust what we RT. Proceedings of the first workshop on social media analytics. New York, 2010, p. 71-79.

[19] Liang G, He W, Xu C, et al. Rumor identification in microblogging systems based on users' behavior. IEEE Transactions on Computational Social Systems. Vol. 2 (2015) No. 3, p. 99-108.

[20] Zhang W, Wang L, Han X, et al. A Method for User Credibility Evaluation on Online Business Review Services. 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA). Chengdu, 2020, p. 199-205.

ISSN: 2414-1895 DOI: 10.6919/ICJE.202205_8(5).0020

[21] Chen X, Fan H, Duan S, et al. Public Opinion Analysis of Big Data Based on Machine Learning. Journal of Physics: Conference Series. IOP Publishing. Vol. 1302 (2019) No. 2, p. 022035.

[22] Ghanem B, Rosso P, Rangel F. An emotional analysis of false information in social media and news articles. ACM Transactions on Internet Technology (TOIT). Vol. 20 (2020) No. 2, p 1-18.

[23] Li L, Cai G, Chen N. A rumor events detection method based on deep bidirectional GRU neural network. 2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC). Chongqing, 2018, p. 755-759.

[24] Wang Z, Guo Y. Rumor events detection enhanced by encoding sentimental information into time series division and word representations. Neurocomputing. Vol. 397 (2020), p. 224-243.

[25] Ma J, Gao W, Wei Z, et al. Detect rumors using time series of social context information on microblogging websites. Proceedings of the 24th ACM international on conference on information and knowledge management. New York, 2015, p. 1751-1754.

[26] Ma J, Gao W, Mitra P, et al. Detecting rumors from microblogs with recurrent neural networks. Proceedings of the 25th International Joint Conference on Artificial Intelligence. New York, 2016, p. 3818-3824.

[27] Chen T, Li X, Yin H, et al. Call attention to rumors: Deep attention based recurrent neural networks for early rumor detection. Pacific-Asia conference on knowledge discovery and data mining. Springer, 2018, p. 40-52.

[28] Meng Z, Yu S, Li R, et al. Dynamic Features Based Rumor Detection Method. 2020 Chinese Control and Decision Conference (CCDC). Hefei, 2020, p. 379-384.

[29] Xu N, Chen G, Mao W. MNRD: A merged neural model for rumor detection in social media. 2018 International Joint Conference on Neural Networks. Rio de Janeiro, 2018, p. 1-7.

[30] Huang Q, Zhou C, Wu J, et al. Deep structure learning for rumor detection on twitter. 2019 International Joint Conference on Neural Networks (IJCNN). Budapest, 2019, p. 1-8.

[31] Bian T, Xiao X, Xu T, et al. Rumor detection on social media with bi-directional graph convolutional networks. Proceedings of the AAAI conference on artificial intelligence. New York, 2020, p. 549-556.

[32] Huang Q, Zhou C, Wu J, et al. Deep spatial–temporal structure learning for rumor detection on Twitter. Neural Computing and Applications, 2020, p. 1-11.