Text Classification Based on Machine Learning: A Review

Yang Sun[1]

Software College, Shenyang Normal University
110034 Shenyang, China
Received Feb. 28, 2025; Revised and Accepted Mar. 11, 2025

Abstract. With the rapid development of the Internet, text information has shown a blowout growth. Massive text data such as news, social media posts, academic literature, etc. are constantly emerging, and manual classification and management of these texts has become time-consuming and inefficient, which is difficult to meet the actual needs. The continuous progress of natural language processing technology, especially the rise of deep learning methods, provides strong technical support for automatic text classification. Deep learning models can automatically mine the essential features of text from massive samples, capture deep semantic representation information, and avoid the tedious process of manual design rules and features. In practical applications, text data often co-exists with data of other modes (such as images, audio, etc.). Through the feature learning of multimodal data, the information of multiple modes can be mapped to the joint vector space, and the unified representation of data can be obtained, so that the text classification can be more accurate. In recent years, pre-trained language models such as BERT and GPT have achieved remarkable results. These models learn a common language representation through unsupervised pre-training on large-scale corpus, and then fine-tune on specific text classification tasks, which can significantly improve the classification performance and further promote the research of automatic text classification. Automatic text classification can classify massive text data into different categories quickly and accurately, which is convenient for information storage, retrieval and management. For example, in the fields of library document management and enterprise document management, automatic classification can greatly improve work efficiency and save labor costs. In social media and online public opinion monitoring, automatic text classification can quickly identify text information with different themes and emotional tendencies. This helps to timely understand the dynamics of public opinion, and provides a basis for the government, enterprises and other institutions to formulate corresponding coping strategies. In the field of customer service, such as online customer service, customer feedback processing, etc., automatic text classification can automatically identify the types of questions and emotional tendencies of customers. Thus, automated customer consultation and problem classification can be realized to improve the efficiency and quality of customer service. Automatic text classification is an important task in the field of natural language processing, and its research progress can provide reference for other natural language processing tasks. For example, in tasks such as sentiment analysis, machine translation, question answering system, etc., the techniques and methods of text classification can be applied and expanded. Automatic text classification technology can be widely used in many fields, such as financial risk assessment, medical text analysis, legal document classification and so on. In these fields, automatic text classification can help professionals quickly sift and process a large amount of text information, improve work efficiency and decision-making accuracy.

Keywords: Automatic text classification, Machine learning, Pre-trained language model.

1. Introduction

With the development of information technology, Internet data and resources present massive characteristics In order to effectively manage and utilize these massive distributed information, content-based information retrieval and data mining have gradually become a field of concern Among them, text classification (TC) technology is an important foundation of information retrieval and text mining. Its main task is to determine its category according to the text content under the preset label set Text classification is widely used in natural language processing and understanding, information organization and management, content information filtering and other fields. The text classification method based on machine learning, which gradually matured in the 1990s, pays more attention to the automatic mining, generation and dynamic optimization ability of the classifier model, and has a breakthrough in classification effect and flexibility compared with the previous text classification model based on knowledge engineering and expert system, It has become a classic in research and application in related fields [1].

It is composed of text representation and effect based classification Sebastiani summarized the development process of text classification and the technology at that time in reference [2]. The main contents include: (1) text about items or feature vector space representation model (VSM) and two dimensionality reduction strategies of feature selection and feature extraction are discussed χ^2. Significance statistics such as IG, MI and OR for feature filtering and feature extraction methods such as item clustering and implicit semantic index (LSI); (2) At that time, the more mature classification model method, that is, the inductive construction of classifier or the construction of model. Mining the learning process; (3) Classification effect evaluation indicators, such as accuracy, recall, BEP and F_β (commonly used F1) and accuracy, as well as the previously reported effect reference comparison on benchmark corpora such as Reuters.

However, the massive electronic texts distributed in the Internet show new characteristics such as variety, skew distribution, complex relationship, frequent updating and difficult labeling, which brings great challenges to the text classification facing the needs of massive information processing does not discuss the difficulties and problems encountered by classification technology in solving the above problems to varying degrees, such as poor scalability, lack of corpus and reduced accuracy, nor can it involve the development of technology in recent years and the important problems and achievements discussed in authoritative academic conferences and journals in the fields of information retrieval, machine learning and data mining.

This paper introduces the latest research on text classification technology based on machine learning, focuses on the problems and progress of text classification in practical applications such as Internet information processing, and summarizes and comments on relevant problems, current situation and trend Section 1 introduces the research trends of basic technology Section 2 discusses the main research problems and the latest progress of text classification facing the challenge of practical application Finally, the summary of the full text and the prospect of related technologies are given.

2. Basic Technology of Text Classification

2.1. Text Representation

VSM is still the main method of text representation, and the relevant research is still focused on the semantic unit as the item and the weight of the calculated item. Most of the work still uses words (or n-grams) as terms and calculates weights based on the frequency of terms, such as $tf \times idf$, etc. It is worth noting that Debole proposed a supervised weight STW, which uses the significance statistics of terms (such as χ^2, etc.) to balance its weights [3]. References [4,5] also use a similar method compared with $tf \times idf$ weights. The introduction of some statistics has improved the classification effect of SVM and linear classification methods to varying degrees.

In addition to VSM, other models based on item probability distribution and two-dimensional view have been proposed. Bigi believes that any text d and class c can be regarded as a probability distribution of all terms $P(t_i, d)$ and $P(t_i, c), i = 1, \cdots, |T|$ (T is the set of all terms or features), called the term distribution probability representation. By measuring the Kullback-Leibler distance (KLD) similarity between distributions, the classification method is better than the linear method under VSM representation. The term distribution probability model is essentially different from VSM only in terms of weight calculation and normalization of terms. Nunzio uses a visual two-dimensional representation method to compress the information of all items into a two-dimensional plane composed of local energy and global energy. After further calculation with a heuristic algorithm, high accuracy is obtained on some test sets [6]. However, the method is only tested on a small data set, and the actual application effect needs to be further verified.

Some other work hopes to consider the relationship between semantic units ignored by bow by referring to the technology of natural language processing. Language and other complex items are applied to the text representation of classification methods However, so far, these representation methods have no obvious advantages in classification effect, and often require complex language preprocessing, which affects the throughput speed of the classifier [7] So far, the rationality of non VSM representation in theory and its scalability in practical application need further verification. The classification method suitable for them is relatively single and has not been widely used.

2.2. Dimensionality Reduction of Representation Space

The related research mainly focuses on three aspects: the model algorithm and comparison of dimensionality reduction, the relationship between feature set and classification effect, and the range of dimensionality reduction.

With regard to the models and algorithms of dimensionality reduction, many studies still follow the traditional ideas: (1) measure and compare the significance of terms on category distribution with probability and statistics methods, such as BNS (BI normal separation) [8]; (2) An item clustering method to study the similarity of item distribution from the perspective of information entropy, such as based on global information (GI); (3) Implicit semantic analysis approach is to obtain the linear mapping that compresses the vector semantics or statistical information to the low dimensional space through different decomposition and simplification of the matrix, such as differential LSI Some novel research ideas include: (1) multi-step or combined selection method, that is, first use the basic feature selection method to determine the initial feature set, and then use some standard (such as considering the co occurrence of other items and the features of the initial set Etc.), or combine other factors (such as according to the second significance selection criteria or considering the value of linear classifier coefficient) to delete redundant features; (2) The research attempts to learn from linguistic technology include learning feature information from manually input features and feature extraction based on WordNet], but the effect of these methods is not ideal.

The influence of dimensionality reduction on classification must be considered, that is, pay attention to the change trend of classifier effect index with the increase of feature number The consistent phenomenon in reference [9] is that the reasonable dimensionality reduction method will make most classifiers show that the effect will improve rapidly with the increase of the number of features, and can quickly approach stability; However, if the number of features is too large, the performance may slow down This shows that dimensionality reduction can not only greatly reduce the processing overhead, but also improve the effect of classifier in many cases Forman and Yang made extensive comparison on different feature selection methods from the aspects of effectiveness, discrimination ability and the opportunity to obtain the best effect From the results: BNS, IG and other statistics and combination methods have certain advantages; In addition, different classifiers tend to accept different specific dimensionality reduction methods [10] The effects of commonly used feature extraction and feature selection algorithms are high or equal to each other in different cases Although the selection method is more widely used because of its low complexity, the extracted features are closer to the semantic description of the text, so it has great research value.

The determination of dimension reduction scale is commonly used in empirical estimation methods, such as empirical value (PFC) or proportion (THR) of a given characteristic number; Or consider factors such as statistics threshold (MVS) or vector space sparsity (SPA) Soucy gives the method that the feature number is proportional to the text number (PCS), compares it with the other four methods under the accuracy standard, and comes to the conclusion that $MVS > PCS > SPA > PFC > THR$. The traditional standard is worth reviewing.

2.3. Machine Learning Classification Method

The main goal of classification method research is to improve the classification effect. A practical system must also take into account the scalability of the learning process and the throughput (speed) of the classification process under the conditions of limited storage and computing power [11] In recent years, the method of ensemble learning using multiple classifiers has been widely accepted; Support vector machine (SVM) still represents the development level of single method.

The application of SVM is one of the most important advances in text classification in recent years Although it has the ability of separating a large number of samples and effectively fitting the data, it needs a large number of redundant data sets, such as the ability of fitting The comparison of relevant literature shows that compared with all other methods, SVM has advantages in effect and stability [12] In recent years, many models or methods not involved in literature [13] have been proposed or applied, and some have achieved good results, such as maximum entropy model, fuzzy theory, KLD similarity of term probability distribution, two-dimensional text model, and method based on equivalent radius (section), but they are still limited to the commonly used classification mode of similarity measurement Bayes, linear classification, decision tree and k-NN are relatively weak, but their models are simple and efficient. The modification and improvement of these methods have attracted continuous attention Wu pointed out that the assumption of classifier about data distribution is an important factor affecting the classification effect. When the model is not suitable for the characteristics of data set, the performance may become very bad This model deviation is particularly prominent in weak classification methods. He gave a flexible heuristic improvement strategy based on error correction; GIS method gathers samples into different instance sets, and the centroid of each instance set is called extended real. For example (GI), after replacing the sample set with the set of GI, the number of instances is reduced, which greatly improves the online speed of k-NN method and the classification effect; Tsay uses the idea opposite to GIS. He increases the number of categories and essentially selects multiple centroids for the original category, partially overcoming the weakness that a single centroid is difficult to adapt to sparse samples; Tan improved Bayes and centroid based methods using drag-pull strategy; Chakrabarti's simpl method uses Fisher linear discriminant analysis to project the text representation into the low dimensional space, and then construct the decision tree It can be seen that most of the research on classification models and methods focus on obtaining the advantage of computational overhead over SVM when the effect on a specific test set is basically the same.

Ensemble learning, also known as multiple learning or classifier combination, mainly synthesizes the capabilities of several weak classifiers through decision optimization or coverage optimization, so as to optimize the overall performance of the classification system Decision optimization for different classifiers, complete sample sets are used for training. During the test, the decisions of all classifiers are voted or evaluated (such as MV (majority voting), W(weighted) MV and WLC (weighted linear combination)) to determine the output category of the whole system; Bennett looks at specific classifiers.

Make reliability indicator; The system uses the probability method to synthesize the outputs of different classifiers to determine the final decision; Xu et al. [14] put forward an idea of serial integration of SVM and Rocchio algorithm, that is, after Rocchio algorithm quickly processes all text vectors, SVM corrects the error of some interesting categories and exchanges the accuracy of important categories with low calculation cost; Coverage optimization uses different training subsets for the same learning to form single classifiers with different parameters. Some synthesis of these single classifier decisions determines the score of each test sample. Classes, such as bagging and boosting methods; In the iterative process of boosting method, each round pays attention to the classification error of the previous round, which is used to improve the weak classification method and obtain better results than SVM, AdaBoost MH and AdaBoost MR and other specific algorithms are widely used].

2.4. Evaluation Method

ROC (receiver operating characteristics) curve in the field of signal detection has been involved in the effect evaluation and optimization of classifiers in recent years For category C, table 2 is the adjacency table of its test results Let $TPR = TP/(TP + FN)$, $FRP = FP/(FP + TN)$. With the adjustment of classifier threshold parameters, the curve in ROC space (TPR, FPR) can not only directly reflect the performance of classifier, but also the area under curve (AUC) can quantify the tendency of classifier to accept positive examples In addition, ROC space is not sensitive to the distribution of samples among categories, and can reflect the changes of error cost and other indicators, which has special advantages. It has become a hot spot to effectively use ROC curve in the evaluation, comparison and optimization of classifiers.

In theory, Li and Yang believe that the error and complexity penalty of training data make the comparison between classifier capabilities clearer Through the formal analysis of common classification methods, they divided the loss function equivalent to the condition and standard for the classifier to obtain the optimal effect into two parts: training loss and model complexity, and gave a method of mutual comparison between classifiers from the perspective of optimization [15].

The experimental comparison between methods is often carried out on the benchmark corpus Reuters is an important benchmark corpus, which is tested on Reuters.

Most tests Common corpora also include OHSUMED, 20 newsgroups, webkb and AP. Reference [16] gives the relative difficulty analysis and reference of Reuters subset Rcv1 (Reuters corpus Volume I) is a relatively complete ”official” corpus compiled and released recently, which is improved.

Some shortcomings of the previous corpus are to meet the needs of multi-layer classification, data skew and scalability of classification methods The construction of corpus plays a very important role in promoting and referencing the research of text classification. Reference [17] gives the corpus processing technology of rcv1 and the reference performance of some methods Most of the public corpora of Chinese classification are under construction, especially the processed benchmark corpora are relatively scarce. Tan has disclosed a relatively new processed Chinese classification corpus tancorp and the reference performance of some classification methods].

3. Key Challenges and Research Progress

The text classification technology based on machine learning has developed continuously for more than 20 years, especially directly drawing on the latest research achievements in the field of machine learning.

As a result, it can better solve most problems and applications with the characteristics of relatively small amount of data, relatively complete annotation and relatively uniform data distribution However, the large-scale application of automatic text classification technology is still plagued by many problems. For example, the problem of depicting the (nonlinear) semantic relationship between texts is not considered to be well solved The main challenges in recent years come from the processing of massive information such as the web on the Internet. Its main characteristics are: (1) the large-scale classification system brings scalability difficulties to classifier training; (2) The samples obtained when establishing the classifier are very limited relative to the massive unknown data, so it becomes difficult to simulate the spatial distribution of samples, which may lead to the problems of over-fitting and data skew; (3) Text and categories are updated frequently, and there is a labeling bottleneck when trying to obtain more samples for each category; (4) The relationship between categories is also more complex, and a better category organization method is needed; (5) Web text is a kind of semi-structured data. Its structural information (such as link relationship, topic, etc.) may provide some help for classification On the whole, we believe that the text classification technology mainly faces several key problems at this stage, such as nonlinearity, data set skew, annotation bottleneck, multi-layer classification, algorithm scale scalability and web page classification The following mainly discusses the possible methods to solve these key problems.

3.1. Dataset Skew

Through many researches in the field of machine learning, it is found that the distribution of data sets about categories is often skewed or unbalanced, that is, there may be an order of magnitude gap in the number of samples between categories, which is an important factor leading to the unsatisfactory classification effect When the data is skewed, the sample can not accurately reflect the data distribution of the whole space, and the classifier is easy to be submerged by large classes and ignore small classes In text classification, especially the classification of Internet information, there are a lot of data skew Especially when using binary classification strategy, for a certain class, the samples of positive examples may only account for a small proportion of all samples. Yang compared the robustness of SVM, Nb, k-NN and other methods in the case of controlled sample distribution and the relationship between classification effect and data distribution [18]. The results show that the robustness of SVM and k-NN to sample distribution is better than that of Nb and other methods, which confirms the generalization performance of SVM and the dependence of Nb on class a priori probability, but the accuracy of all methods in rare categories is very low.

The main countermeasures to solve the problem of data skew are: (1) resampling, which can appropriately shield the amount of information of large classes or increase the cost of classification errors of small classes; (2) A new classification strategy is adopted, such as one class SVM, which takes the origin as the center of the unknown category and constructs the separation surface surrounding the training samples, so as to transform the problem into an equivalent two kinds of problems that are not affected by the category distribution; Reference [19] discusses the training of SVM with only a few positive examples; The nknn method proposed in reference [20] improves the effect of k-NN on skewed data sets; (3) Using better effect evaluation methods, such as ROC curve or cost curve, can more accurately evaluate the overall performance of the classifier in the case of data skew; (4) In the case of data skew, features are also very important. The attention of the classifier to small category features can be obtained by optimizing the feature selection framework or improving the feature selection method At present, all methods can not improve the recognition level of rare categories (BEP of about 0.5 or lower) to an acceptable level, and the relevant research still needs to be further deepened.

3.2. Marking Bottlenecks

The learning algorithm needs a large number of labeled samples, but the labeled samples can provide limited information; On the other hand, the number of unlabeled samples that are easy to obtain (such as through the Internet) is larger than that of labeled samples, and is closer to the data distribution in the whole sample space Providing as many labeling samples as possible requires hard and slow manual labor, which restricts the construction of the whole system, which leads to a labeling bottleneck problem Therefore, how to train a good classifier with a small number of labeled samples and a large number of unlabeled samples has gradually attracted people's attention Nigam first uses the method based on expectation maximization (EM) to learn from unlabeled samples, and uses test samples to improve the classification effect of Bayes classifier; Another method used for unlabeled text learning is direct inference, which makes the classifier predict only a small number of current unknown samples with the smallest error through the learning of labeled samples, without considering the optimality of the expected performance of all future instances Then, these samples are added to the learning process to improve the effect of the classifier; Jaochims uses direct push support vector machine TSVM for text classification, which has been improved; Direct push boosting text classification is discussed in literature [21]; Literature [22] adopts the method of CO training and uses unlabeled samples to classify e-mail and text. Its idea is to divide the characteristics of samples into two subsets with sufficient information from two perspectives, establish classifiers on the two subsets respectively, and use labeled samples for cooperative learning In addition, literature [23] only uses positive samples and unmarked samples for learning; In reference [24], SVM is used for active learning The above method is of great help to improve the performance of the classifier when there are few labeled samples. Although it partially alleviates the labeling bottleneck, it also comes at the cost of a large number of iterations In addition, there is no comparative work under the same standard between different unlabeled sample learning methods.

3.3. Multi-layer Classification

In the classification problems usually discussed, the categories are isolated, and they are not related to each other, which is called flat classification In the case of many categories and complex relationships, such as the management of rich web information on the Internet, a large class of applications need a better multi-layer information organization Hierarchical classification refers to the classification problem under the multi-layer category relationship. There is a multi-layer hierarchical category structure similar to tree or directed acyclic graph between the facing categories, which can better support browsing and query, and also make some large-scale classification problems better solved by over-treatment methods.

Multi-layer classification generally adopts two strategies: big bang or top-down level based. The former uses the same classifier in the whole classification process, that is, all leaf node categories in the category tree structure are regarded as equal classes. This is essentially a single-layer classification and can not apply the relationship between categories well; The latter can train different classifiers for different levels, and the branch node classifier only cares about the current different branches. Sun et al. [25] discussed the multi-layer classification effect evaluation method based on category similarity and category distance, and gave a specification language for scheduling classifiers at different levels. Ruiz's doctoral thesis introduces several multi-layer classification methods proposed earlier, and gives his own HME (hierarchical mixture of expert) model. Huang et al. introduced liveclassifer for building multi-layer classifiers from Web corpus [26]. In multi-level classification, the complexity and mutual interference of category relations and the propagation of classification errors between different category levels may affect the accuracy evaluation of classifiers. Only sun considered this influence in literature [27] For the same label (category) set, the effect of multi label classification under single-layer classification setting and classification under multi-layer classification setting also need to be compared. These problems have not been deeply studied at present.

3.4. Web Page Classification

Traditionally, the text classification discussed is generally oriented to the text content itself. In the text preprocessing stage, the structural information contained in the text, such as HTML tags, topics and hyperlinks, will be cleared. However, in the classification of Internet oriented information, especially web pages, these structured information contained in the text will provide rich information about the attribution of the text. For example, we can consider the category in which the hyperlink contained in the test sample points to the text, so as to confirm the decision of the content classifier; The anchor word in the hyperlink or the words around it (extended anchor word) are used as features to express the text pointed to by the hyperlink; Using the structural and topological information of hyperlinks and HTML tags to describe the relationship between texts, and using kernel function to express hyperlinks. The classification effect of these works on different corpora is better than that without structural information The work of using structural information is not always effective. For example, the method of treating the words of the linked text as local words reduces the accuracy of the classifier. Yang pointed out that this is due to the improper assumption of the relationship between hyperlinks and categories on the corpus. At present, how to properly represent these structured information and automatically learn their statistical models is still an open problem.

4. Conclusion

Text classification technology is widely used and gradually tends to be practical However, with the development of related applications and the continuous improvement of demand, there are still many problems worthy of study, such as: ways and methods to solve the problem of large-scale classification applications; Reliable, effective and fast online classification; Combined with research in the field of natural language research on data model and classification method based on semantic measurement; Alleviate the bottleneck of sample labeling and the impact of sample data distribution With the deepening of theoretical and technical research in the field of machine learning and data mining, aiming at the characteristics of different practical applications and data, especially the problems of data model, category scale and performance bottleneck in Internet content processing and other large-scale and complex applications, will become the focus and main breakthrough direction of text classification related research and application.

5. Conflict of Interest

The authors declare that there are no conflict of interests, we do not have any possible conflicts of interest.

Acknowledgments. None.

References

1. Hassan S U, Ahamed J, Ahmad K. Analytics of machine learning-based algorithms for text classification[J]. Sustainable Operations and Computers, 2022, 3: 238-248.
2. Li R, Liu M, Xu D, et al. A review of machine learning algorithms for text classification[J]. Cyber Security, 2022, 226.

3. Allam H, Makubvure L, Gyamfi B, et al. Text Classification: How Machine Learning Is Revolutionizing Text Categorization[J]. Information, 2025, 16(2): 130.
4. Shyrokykh K, Girnyk M, Dellmuth L. Short text classification with machine learning in the social sciences: The case of climate change on Twitter[J]. Plos one, 2023, 18(9): e0290762.
5. Mohapatra S K, Sarangi P K, Sarangi P K, et al. Text classification using NLP based machine learning approach[C]//AIP conference proceedings. AIP Publishing, 2022, 2463(1).
6. Gasparetto A, Marcuzzo M, Zangari A, et al. A survey on text classification algorithms: From text to predictions[J]. Information, 2022, 13(2): 83.
7. Duarte J M, Berton L. A review of semi-supervised learning for text classification[J]. Artificial intelligence review, 2023, 56(9): 9401-9469.
8. Chen H, Wu L, Chen J, et al. A comparative study of automated legal text classification using random forests and deep learning[J]. Information Processing & Management, 2022, 59(2): 102798.
9. Li Q, Peng H, Li J, et al. A survey on text classification: From traditional to deep learning[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2022, 13(2): 1-41.
10. Hong S, Kim J, Yang E. Automated text classification of maintenance data of higher education buildings using text mining and machine learning techniques[J]. Journal of Architectural Engineering, 2022, 28(1): 04021045.
11. Mohammed A, Kora R. An effective ensemble deep learning framework for text classification[J]. Journal of King Saud University-Computer and Information Sciences, 2022, 34(10): 8825-8837.
12. Dogra V, Verma S, Kavita, et al. A complete process of text classification system using state-of-the-art NLP models[J]. Computational Intelligence and Neuroscience, 2022, 2022(1): 1883698.
13. Allam H, Makubvure L, Gyamfi B, et al. Text Classification: How Machine Learning Is Revolutionizing Text Categorization[J]. Information, 2025, 16(2): 130.
14. Chen H, Zhang H, Yang Y, et al. A text classification network model combining machine learning and deep learning[J]. International Journal of Sensor Networks, 2024, 44(3): 182-192.
15. Khaksar A, Hassani H. Shifting from endangerment to rebirth in the Artificial Intelligence Age: An Ensemble Machine Learning Approach for Hawrami Text Classification[J]. arXiv preprint arXiv:2409.16884, 2024.
16. Alyasiri O M, Cheah Y N, Zhang H, et al. Text classification based on optimization feature selection methods: a review and future directions[J]. Multimedia Tools and Applications, 2024: 1-47.
17. Sun G, Cheng Y, Zhang Z, et al. Text classification with improved word embedding and adaptive segmentation[J]. Expert Systems with Applications, 2024, 238: 121852.
18. Ben Abdennour G, Gasmi K, Ejbali R. An Optimal Model for Medical Text Classification Based on Adaptive Genetic Algorithm[J]. Data Science and Engineering, 2024: 1-15.
19. Behzadidoost R, Mahan F, Izadkhah H. Granular computing-based deep learning for text classification[J]. Information Sciences, 2024, 652: 119746.
20. Xie Y, Li Z, Yin Y, et al. Advancing Legal Citation Text Classification A Conv1D-Based Approach for Multi-Class Classification[J]. Journal of Theory and Practice of Engineering Science, 2024, 4(02): 15-22.
21. Liu B, Chen J, Wang R, et al. Optimizing news text classification with Bi-LSTM and attention mechanism for efficient data processing[C]//2024 5th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI). IEEE, 2024: 281-285.
22. Akhadam I, Ayyad H. Enhancing Arabic Text Classification: A Comparative Study of Machine Learning and Deep Learning Approaches[C]//2024 IEEE 12th International Symposium on Signal, Image, Video and Communications (ISIVC). IEEE, 2024: 1-6.
23. Dehkori M K, Sattari F, Lefsrud L. Creating an Incident Investigation Framework for a Complex Socio-Technical System: Application of Multi-label Text Classification and Bayesian Network Structure Learning[J]. Reliability Engineering & System Safety, 2025: 110971.
24. Jamshidi S, Mohammadi M, Bagheri S, et al. Effective text classification using BERT, MTM LSTM, and DT[J]. Data & Knowledge Engineering, 2024, 151: 102306.
25. Hijazi M M, Althebyan Q. A Review study on Utilizing Deep Learning in Arabic Text Classification[C]//2024 25th International Arab Conference on Information Technology (ACIT). IEEE, 2024: 1-3.
26. Malik S, Jain S. Deep convolutional neural network for knowledge-infused text classification[J]. New Generation Computing, 2024, 42(1): 157-176.
27. Atanasova P. A diagnostic study of explainability techniques for text classification[M]//Accountable and Explainable Methods for Complex Reasoning over Text. Cham: Springer Nature Switzerland, 2024: 155-187.

Biography

Yang Sun is with the Software College, Shenyang Normal University. Research direction is computer application and AI.