Automatic Classification of Document Resources Based on Naive Bayesian Classification Algorithm

Rong Wang[1*]
Email: rongwang9@126.com

Keywords: Literature Resources; Naive Bayes Data Discretization; Automatic Classification; Ontology Integration Module

Received: February 3, 2022

World Wide Web has become big as the amount of documents collection is increasing rapidly. The automatic classification of document resources based on Naive Bayesian classification algorithm is detailed in this paper. Firstly, this paper introduces the relevant theories of naive Bayes classification and the automatic document classification system. Then, a massive network academic document automatic classification system is designed and implemented. The system uses modular design, including academic document automatic capture module, academic document word document matrix processing module, ontology integration module and semantic driven classification module. Finally, based on the Naive Bayesian classification algorithm, the training set of 12 categories preset is utilized in the professional classification directory of the Ministry of education.. Experiments show that the naive Bayesian classification algorithm can effectively complete the automatic capture, processing and classification of massive academic documents, which can not only improve the classification accuracy, but also reduce the running time of automatic classification. It solves the problems of the integration of two heterogeneous ontology libraries and also the problem that the traditional word vector space cannot meet people's needs for semantic classification.

Povzetek: Za avtomatsko klasifikacijo dokumentov s spleta je implementiran naivni Bayesov algoritem.

1 Introduction

The Internet is an information resource of text, images, audio and video. There is rapid increase in the amount of information available on the World Wide Web (WWW) at an exponential rate. This rich textual information is contained in the Web documents but the growth of the internet has made it difficult for users for location of relevant information quickly on the Web. At present, network academic resources show an upward trend in both breadth and depth, and have attracted more and more attention from the academic community. The massive network academic literature has a huge scale and fast update speed. Fully mining it has important academic value. However, these characteristics have also become a stumbling block for scientific researchers to make use of it. How to acquire and process a large amount of academic literature is a severe test for computer processing and throughput. Whether in terms of processing speed, storage space, fault tolerance or access speed, it is difficult for single computer platform architecture and processing capacity to successfully complete this task. Due to the huge number of network academic documents, it is difficult to make effective use of them, so it is of practical significance to classify them automatically based on disciplines. Automatic document classification is widely used in the fields of information retrieval, data mining, spam filtering, digital library and so on. There are two common classification methods: one is rule-based, which usually requires a large number of domain experts to extract the rules of the text, which is time-consuming and laborious, and the classification effect is poor; Another kind of method is machine learning method based on statistics, including nearest neighbor method, support vector machine, naive Bayes, decision tree, neural network, etc. this kind of method usually uses feature vector space to train document classification model. However, word feature vectors ignore the semantic relationship between words and cannot reflect synonyms, polysemy and the upper and lower relationship between words, resulting in too high vector space dimension. When automatically classifying massive documents, there will be problems such as insufficient memory, slow classification speed and low classification performance, Automatic document classification technology and method cannot be more widely applied to the practice of specific fields [2]. In order to solve the problems existing in the traditional automatic document classification based on word vector space, a series of semantic driven automatic document classification methods are proposed, such as latent semantic analysis method, ontology semantic mapping method, concept lattice construction method, standardized concept analysis method and so on. Although the semantic driven automatic text classification method can greatly reduce the dimension of document vector space, it also has many defects, such as high requirements for semantic reasoning ability, high computational complexity, and unable to classify web documents quickly and effectively.

Bayesian classification (Figure 1) is proposed on the solid theoretical basis of Bayesian theorem. For a given sample, the posterior probability of belonging to each category is calculated according to the distribution of each category sample in the training set, and then the category of the sample is judged as the category corresponding to the maximum posterior probability. The principle of this method is simple, but when the number of attributes is large, training and learning a classification model completely according to Bayesian theorem will have a huge computational overhead and will be greatly limited in practical application [3]. Therefore, scholars simplified a hypothesis of attribute conditional independence, and proposed a practical naive Bayesian classification algorithm, which greatly reduced the computational overhead in the process of model training. At the same time, the research also shows that naive Bayesian classification method still has good performance in many practical applications.

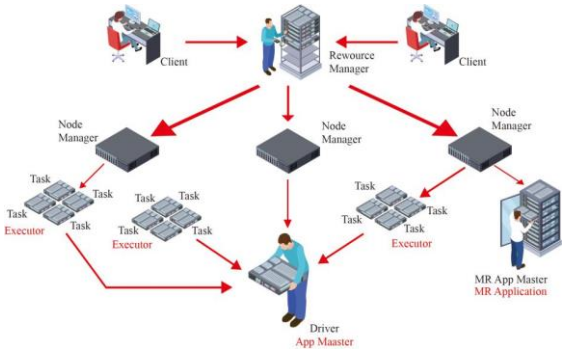

Figure 1: Bayesian classification

Contribution: This paper introduces the relevant theories of naive Bayes classification and the automatic document classification system. Then, a massive network academic document automatic classification system is designed and implemented.

The organization of the paper is as follows. Section 2 provides an overview of the exhaustive literature survey followed by the Automatic classification of massive network academic documents adopted in section 3. The experimental analysis is in section 4. Finally, Section 5 concludes the paper.

2 Literature review

Du, J. h. and others also proposed a network extended naive Bayesian classification model (BAN). This method extends the structure of naive Bayesian classifier to a greater extent. It is the same as the improved model of TAN. Its fundamental starting point is to weaken the assumption that attributes are independent to a greater extent. The ban model is the same as the TAN model in many aspects. The BAN model also stipulates that the class node is the root. At the same time, all other attribute nodes take its parent node and the BAN classifier uses Bayesian network as the expression structure, which is the only difference [4]. Y Kumar's Bayesian augmented naive Bayesian classifier GBAN is based on genetic algorithm. GBAN model can meet the limitations of the network extended naive Bayesian classification model on the network structure, that is, any attribute node has at most M parent nodes (generally m < 4), but the category variables are not included [5]. The hybrid tree augmented naive Bayesian classification model proposed by DIAS, K. L. is based on rough set theory.

The composition process of augmented naive Bayesian classification model is as follows: Based on the attribute reduction theory of rough set, under the condition of keeping the classification ability unchanged, it is divided into two categories according to the impact of attribute variables on the classification results. It is assumed that the attribute variables that have no or little impact on the classification results are independent of each other, and these nodes can only have one parent node, The attribute variables that affect the classification results are not independent of each other, and these nodes can have two parent nodes [6]. Tajanpur proposed a hybrid model (nbtree) combining decision tree and naive Bayes. The process of learning nbtree by the algorithm is similar to that of decision tree (C4.5), but it is different in the selection of attribute splitting evaluation score function [7]. Gaber, A. and others proposed an average naive Bayesian tree model [8]. Lopes, F. and others proposed an improved naive Bayes model (LBR) combining lazy technology and naive Bayes, which can obtain high classification accuracy, but the classification efficiency of this method is not very high [9]. In terms of automatic document classification, the classification method based on coverage coefficient by an, Y. and others is a classification method based on the inherent attributes of document set. This method borrows mathematical tools to derive a classification step with rigorous reasoning.

The premise is that (under certain general assumptions) the class and number of classes of each document in the document set have been determined by the inherent attributes of the document set itself [10]. Rueda and others proposed an automatic acquisition and parallel processing model of massive network academic documents. The rules specified by the heritrix platform are used to capture the data of the seed site. For the captured file resources, they are judged according to the set academic literature feature rules, and then some of them are selected to invite domain experts for category indexing, train the machine learning classification algorithm, and finally realize the classification of all documents [11]. In previous years, many researchers have worked on this particular field, some of the relevant articles are tabulated in Table 1.

Authors	Presented Work	Key points	Benefits	References
Mohamed EL KOURDI et al.,	"Naive Bayes (NB) is a statistical machine learning algorithm utilized for the classification of non-vocalized Arabic web documents which is presented in this paper."	"The data set utilized during the experiments consists of 300 web documents per category."	High classification accuracy	[12]
Huaixin Chen et al., 2018	""Improved Naïve Bayes classifiers are presented utilizing multinomial model."	"The proposed method is able to improve the accuracy of Naïve Bayes classifiers dramatically."	Good scalability	[13]
Yong Wang et al., 2003	"An automatic document classification system, WebDoc, which classifies Web documents according to the Library of congress is presented."	"Performance of each method in terms of recall, precision, and F-measures is reported."	Highly effective and efficient.	[14]
A. B. Adetunji et al., 2018	"A University web site is used as a case study and a machine learning workbench called WEKA is discussed."	"General-purpose environment for automatic classification, clustering and feature selection are provided."	Naïve Bayes algorithm ability is to accurately classify the web document vast amount.	[15]
Yugang Dai a et al., 2014	"Naïve bayesian classification algorithm is presented by the author which is further combining with the rough set theory."	"This algorithm is implemented on a cloud platform utilizing map-reduce programming mode."	High recall rate	[16]

Table 1: Some existing and relevant articles in previous years

3 Automatic classifications of massive network academic documents

With the goal of automatically acquiring massive documents and automatically classifying documents, its framework is shown in Figure 2:

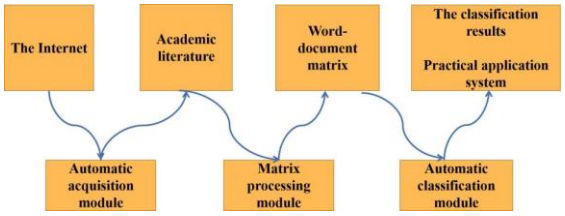

Figure 2: Framework of automatic classification system for massive network academic documents

The automatic document acquisition module first captures and determines academic documents from the Internet according to predetermined rules and conditions, so as to filter irrelevant documents; Then, through the matrix processing module, the academic literature is transformed into a word document matrix for subsequent processing; Finally, the word document matrix is imported into the automatic classification module after training and ontology integration to obtain the classification results [17, 18].

(1) Automatic acquisition of massive network academic documents

In the automatic classification system of massive network academic documents, it is necessary to obtain massive academic documents. First, use heritrix to grab all PDF files under the domain name from a specific website, read all PDF files with checkpdf, and identify academic literature through rule-based judgment method, as shown in Figure 3:

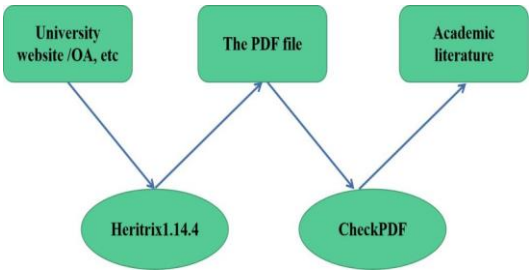

Figure 3: Automatic acquisition of massive network academic documents

In the selection of capture tools, the author studies and analyzes the network resource capture platforms such as nutch, heritrix, jspider and web harvest from the aspects of capture efficiency and scalability, and finally selects heritrix as the capture platform. Heritrix has high scalability [19, 20], can retain the original file structure and directory, and has a web user interface. It runs on Linux system and can ensure high capture speed. In terms of file format, considering the convenience of subsequent processing and the proportion of various file types, PDF is selected as the main capture file type. After the PDF file is captured, it needs to be screened to retain the academic literature. The rule-based decision method is used, that is, the decision is made through keywords. By analyzing a large number of academic documents, it is found that its unique characteristic words include abstract, keywords, introduction, discussion, conclusion and recognition. Different documents may contain several words respectively. Therefore, a threshold can be set to judge according to the number of the above words [21-22].

(2) Massive network academic literature words -- document matrix processing

In view of the large number of documents to be processed, the word frequency matrix is generated by distributed processing. This part is implemented using Hadoop, including Hadoop namenode and Hadoop datanode. Namenode is responsible for the scheduling of parallel processing, and datanode is responsible for the actual parallel processing. Academic documents are first read into the Hadoop platform, and an index of all documents is saved on the namenode. The actual documents are saved on at least two datanodes in the form of redundancy, and finally passed Namenode calls the parallel processing program to generate the word document matrix of academic literature [23-25], as shown in Figure 4:

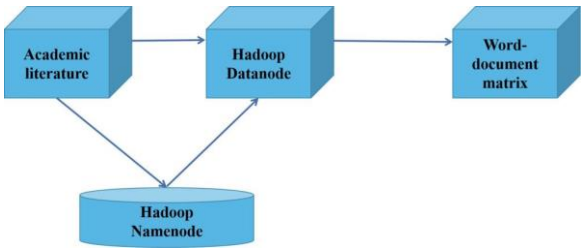

Figure 4: Massive network academic literature words - document matrix processing

In the map phase of Hadoop, stringtokenizer is used to extract the words in the literature in turn and generate a key \ value pair < word, document ID >. In the reduce phase of Hadoop, a reducer is used to process the same word, create an array with the length of documents, save the word frequency of the current word in the corresponding documents, and then accept the key \ value pair in turn and update the array. Output the matrix after all reducer work is completed. Since this matrix is sparse, you can delete 0 bits and output sparse matrix to reduce storage space [26, 27].

(3) Ontology integration

In order to understand natural language, the common method is to use ontology library to annotate and integrate text. This part mainly uses prompt. Prompt first reads the ontology, then analyzes the relationship between concepts, maps the same concepts, retains the special concepts in an ontology library, and finally generates an integrated integrated ontology, as shown in Figure 5.

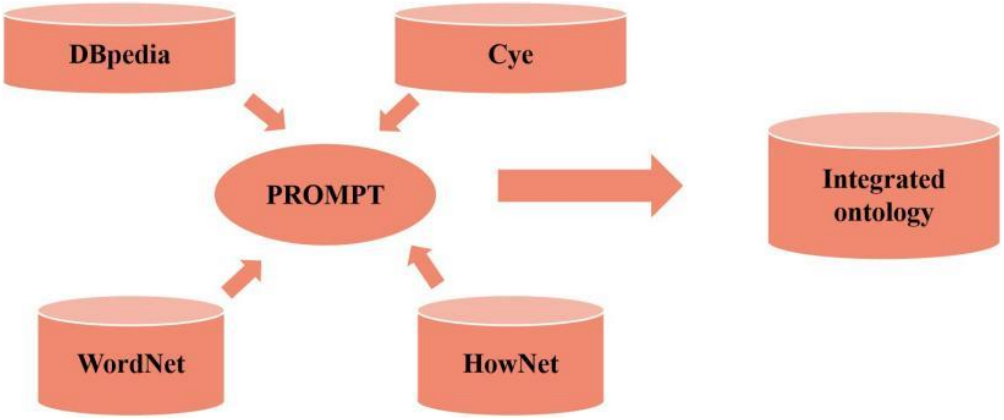

Figure 5: Ontology integration

3.1 Naive Bayesian algorithm

Before describing the naive Bayesian classification algorithm, the classification problem is formalized from the perspective of statistics. Let X represent the attribute set of the system data set, $X=(A_1,A_2,\cdots A_m)$, Y represent the class label set of the system data set, and $Y=(C_1,C_2,\cdots C_t)$. Because the relationship between class variables and attributes is uncertain, X and Y can be regarded as random variables, and $P(Y'|X')$ can be used to capture the relationship between them in a probabilistic manner. $P(Y'|X')$ is also called a posteriori probability of class Y. Correspondingly, $P(Y)$ is called a priori probability of Class Y [28, 29].

In the training stage of naive Bayesian classification algorithm, firstly, the information statistics of the training data set is carried out, and the a posteriori probability $P(Y|X)$ of each combination of attribute sets X and Y is calculated. After calculating these probabilities, the test sample X 'can be classified by finding the class Y ' that maximizes the delay probability $P(Y'|X')$. However, it is very difficult to accurately estimate the a posteriori probability of each possible combination of Class Y and attribute values, because even if the number of attributes is not many, a large training set is still required. At this time, the Bayesian theorem plays an important role, because the posterior probability can be expressed by the prior probability $P(Y)$, the class conditional probability $P(X|Y)$ and the evidence $P(X)$ through the Bayesian theorem. The formula for calculating the posterior probability $P(Y|X)$ by the Bayesian theorem is formula (1).

$$P(Y|X)=\frac{P(X|Y)P(Y)}{P(X)} \quad (1)$$

When comparing the posterior probabilities of different Y values, the denominator $P(X)$ is always constant and can be ignored. The prior probability $P(Y)$ can be easily estimated by calculating the proportion of training samples belonging to each class in the total training samples in the training data set. However, for the training data with m attributes [30, 31], the calculation of class conditional probability $P(X|Y)$ is time-consuming. In order to improve the efficiency of calculating $P(X|Y)$, naive Bayesian classification algorithm assumes that the attributes are conditionally independent when estimating the conditional probability of classes. The assumption of attribute conditional independence can be expressed by formula (2):

$$P(X|Y=y)=\prod_{i=1}^{m}P(X_i|Y=y) \quad (2)$$

Through the conditional independence assumption, it is not necessary to calculate the class conditional probability of each value group sum of X, but to mark Y for a given class and calculate the conditional probability of each X_i. In contrast, the latter method is more practical. Because through the assumption of conditional independence, better probability statistics can be obtained without a large training data set [32-34].

In the classification test stage, naive Bayesian classification algorithm calculates a posteriori probability for each X, as shown in formula (3):

$$P(Y|X)=\frac{P(Y)\prod_{i=1}^{m}P(X_i)}{P(X)} \quad (3)$$

Because $P(Y)$ and $P(X)$ are fixed for fixed training data sets and determined test data. Therefore, it is sufficient to find the class that maximizes the molecular $P(Y)\prod_{i=1}^{m} P(X_i)$. For naive Bayesian classification algorithm, the biggest disadvantage is that naive Bayesian classification algorithm can only deal with discrete attributes [35, 36].

4 Experimental Analysis

The experimental classification standard selects 12 categories preset in the professional classification catalogue of the Ministry of education of the people's Republic of China, namely philosophy, economics, law, pedagogy, literature, history, science, engineering, agronomy, medicine, management and military science. The literature data sets used in the experiment include isolet, covtype and census_. The specific description of the data set is shown in Table 2.

Experiment No	Number of documents	Number of matrix rows	Number of matrix columns	Matrix size	Computing time
1	72	72	29876	5.7	3
2	728	728	175897	332.6	14
3	7159	7159	746239	17.8	27 minutes and 100 seconds
4	108026	108026	903452	198.6	7 hours 34 minutes 20 seconds

Table 2: Description of algorithm experimental data

In terms of data sources, after analyzing different target sources, it is found that famous university websites, some discipline portals and OA warehouses contain a large number of publicly published academic documents, which can be captured without restrictions. Therefore, it is determined to take university websites, OA warehouses and discipline portals as target sources. In order to make the results more representatives, the conference website and the researcher's home page were also added. The target sites selected in this experiment are shown in Table 3.

No.	Site	Brief Introduction	Type
1	https: //www. stanford.edu	Stanford University website	University website
2	https://www.omicsonline.org	Omnics group website	OA warehousing
3	https://www.acm.org	American Computer Society website	Subject Portal
4	https://webis.de	International Conference pan website	Conference website

Table 3: Document capture target sites

It can be seen from the experimental results that the classification accuracy of naive Bayes has been slightly improved after discretization. The reason is that after discretization, the continuous attributes are mapped into discrete classification attributes, which makes the system more complete, and avoids a potential problem in estimating a posteriori probability from training data to a certain extent: the class conditional probability of attributes is equal to zero, The extreme case that the posterior probability of the whole class is equal to zero, resulting in classification error or inability to classify. The experimental results show that the classification accuracy of the algorithm can be greatly improved by discretizing the continuous data through the parallel attribute discretization algorithm based on direct.

In the aspect of algorithm execution efficiency, the running time of the two algorithms to deal with data classification tasks of different scales under the

environment of different number of nodes is recorded respectively. The specific running time is shown in Figure 6.

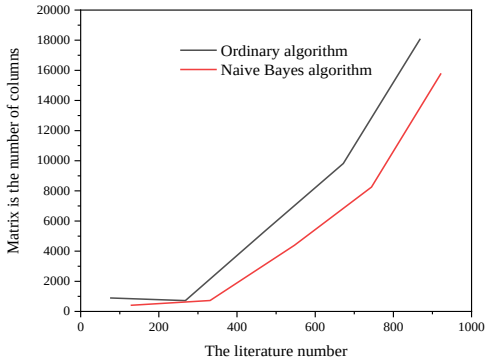

Figure 6: Comparison of algorithm running time

As can be seen from Figure 6, the classification results of all academic documents can be viewed through this system. The year data is not mined at the text level, but directly uses the PDF file metadata (creation date). On the document display page, you can view the title, category, original URL and excerpt of the text of the document. Each interface is equipped with faceted search function to facilitate users' secondary retrieval. The efficiency of these algorithms in terms if run time is calculated and shown in Figure 7.

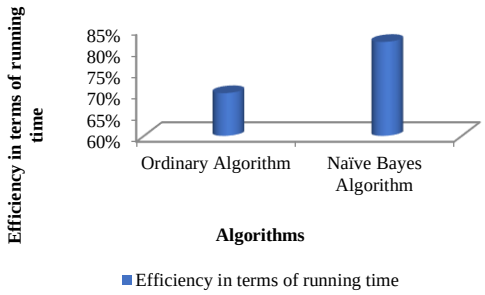

Figure 7: Comparative analysis of the algorithms in terms of efficiency

The Naïve algorithm is much more effective and efficient in terms of complexity. The Naïve Bayes algorithm is 82% efficient and the ordinary algorithm efficiency is 70%.

4 Conclusion

The successful design and implementation of naive Bayesian classification algorithm can not only solve the problems of large memory consumption, slow processing speed and high feature vector dimension in the process of massive document processing, but also enable scientific researchers to effectively obtain and use the documents. At the same time, it also solves the problems of the integration of two heterogeneous ontology libraries and how to apply them in specific fields. The problem is that the traditional word vector space cannot meet people's needs for semantic classification, semantic navigation and semantic retrieval of massive network information resources due to high dimension and lack of semantics. Therefore, it has academic value and practical significance. The design idea and framework of the system can be directly applied to e-government system, portal website, vertical search engine, digital library website and so on. The main strength of the approach lies in its ability to classify the web documents into the right categories correctly and in zero seconds. The future work of this work can be on combining two classification techniques to increase the accuracy of a web page classification.

References

[1] Li, W. Q. , Li, Y. , Chen, J. , & Hou, C. Y. . (2017). Product functional information based automatic patent classification: method and experimental studies. Information Systems, 67(JUL.), 71-82. https://doi.org/10.1016/j.is.2017.03.007

[2] Agnihotri, D. , Verma, K. , & Tripathi, P. . (2018). An automatic classification of text documents based on correlative association of words. Journal of Intelligent Information Systems, 50(3), 549-572.

[3] Pajic, M. S. , Veinovic, M. , Peric, M. , & Orlic, V. D. . (2020). Modulation order reduction method for improving the performance of amc algorithm based on sixth – order cumulants. IEEE Access, PP(99), 1-1. DOI: 10.1109/ACCESS.2020.3000358

[4] Du, J. H. . (2017). Automatic text classification algorithm based on gauss improved convolutional neural network. Journal of Computational Science, 21(jul.), 195-200.

[5] Y Kumar, Sheoran, M. , Jajoo, G. , & Yadav, S. K. . (2020). Automatic modulation classification based on constellation density using deep learning. IEEE Communications Letters, PP(99), 1-1, DOI: 10.1109/LCOMM.2020.2980840

[6] Dias, K. L. , Pongelupe, M. A. , Caminhas, W. M. , & Errico, L. D. . (2019). An innovative approach for real-time network traffic classification. Computer networks, 158(JUL.20), 143-157, https://doi.org/10.1016/j.comnet.2019.04.004

[7] TajanpureRupalirupalidixit@gmail.comMuddanaAkkalakshmiamuddana@gitam.eduGITAM University,Hyderabad,Telangana,India. (2021). Circular convolution-based feature extraction algorithm for classification of high-dimensional datasets. Journal of Intelligent Systems, 30(1), 1026-1039, https://doi.org/10.1515/jisys-2020-0064

[8] Gaber, A. , Hamdy, A. , Abdelaal, H. M. , Elkattan, A. , & Youness, H. A. . (2021). Automatic classification algorithm for diffused liver diseases based on ultrasound images. IEEE Access, PP(99), 1-1, DOI: 10.1109/ACCESS.2021.3049341.

[9] Lopes, F. , Agnelo, J. , Teixeira, C. A. , Laranjeiro, N. , & Bernardino, J. . (2020). Automating orthogonal defect classification using machine

learning algorithms. Future generation computer systems, 102(Jan.), 932-947, DOI: 10.1109/ACCESS.2021.3049341

[10] An, Y. , Xu, M. , & Shen, C. . (2019). Classification method of teaching resources based on improved knn algorithm. International Journal of Emerging Technologies in Learning (iJET), 14(4), 73-88, https://doi.org/10.3991/ijet.v14i04.10131

[11] Rueda, C. A. , & Ryan, J. P. . (2020). Humpback whale song analysis based on automatic classification performance. The Journal of the Acoustical Society of America, 148(4), 2597-2597, https://doi.org/10.1121/1.5147215

[12] El Kourdi, M., Bensaid, A., & Rachidi, T. E. (2004). Automatic Arabic document categorization based on the Naïve Bayes algorithm. In *proceedings of the Workshop on Computational Approaches to Arabic Script-based Languages* (pp. 51-58), https://dl.acm.org/doi/10.5555/1621804.1621819

[13] Chen, H., & Fu, D. (2018, March). An improved Naive Bayes classifier for large scale text. In *2018 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)* (pp. 33-36). Atlantis Press, https://doi.org/10.2991/icaita-18.2018.9

[14] Wang, Y., Hodges, J., & Tang, B. (2003, November). Classification of web documents using a naive bayes method. In *Proceedings. 15th IEEE International Conference on Tools with Artificial Intelligence* (pp. 560-564). IEEE, DOI: 10.1109/TAI.2003.1250241

[15] Adetunji, A. B., Oguntoye, J. P., Fenwa, O. D., & Akande, N. O. (2018). Web Document Classification Using Naïve Bayes. *Journal of Advances in Mathematics and Computer Science*, *29*(6), 1-11, DOI: https://doi.org/10.48550/arXiv.2006.01715

[16] Dai, Y., & Sun, H. (2014). The naive Bayes text classification algorithm based on rough set in the cloud platform. *Journal of Chemical and Pharmaceutical Research*, *6*(7), 1636-1643, https://doi.org/10.1007/s00500-020-05410-9

[17] Koopman, B. , Zuccon, G. , Nguyen, A. , Bergheim, A. , & Grayson, N. . (2015). Automatic icd-10 classification of cancers from free-text death certificates. International journal of medical informatics, 84(11), 956-965, DOI: 10.1016/j.ijmedinf.2015.08.004

[18] Li, K. , & Sidorovskaia, N. . (2019). Detection and classification beaked whale vocalization calls based on unsupervised machine learning algorithm. The Journal of the Acoustical Society of America, 145(3), 1855-1856.

[19] Sharma, A., & Kumar, R. (2019). Computation of the reliable and quickest data path for healthcare services by using service-level agreements and energy constraints. *Arabian Journal for Science and Engineering*, *44*(11), 9087-9104, 10.1007/s13369-019-03836

[20] Harakawa, R. , Ogawa, T. , Haseyama, M. , & Akamatsu, T. . (2018). Automatic detection of fish sounds based on multi-stage classification including logistic regression via adaptive feature weighting. The Journal of the Acoustical Society of America, 144(5), 2709-2718, DOI: 10.1121/1.5067373

[21] Hartvigsen, L. , Kongsted, A. , Vach, W. , Salmi, L. R. , & Hestbaek, L. . (2018). Does a diagnostic classification algorithm help to predict the course of low back pain? a study of danish chiropractic patients with one-year follow up. Journal of Orthopaedic and Sports Physical Therapy, 48(11), 1-35, DOI: 10.2519/jospt.2018.8083

[22] Sharma, A., & Kumar, R. (2019). Risk-energy aware service level agreement assessment for computing quickest path in computer networks. *International Journal of Reliability and Safety*, *13*(1-2), 96-124.

[23] M Foroutan, & JR Zimbelman. (2017). Semi-automatic mapping of linear-trending bedforms using 'self-organizing maps' algorithm. Geomorphology, 293(PT.A), 156-166.

Heidari, M. , Lakshmivarahan, S. , Mirniaharikandehei, S. , Danala, G. , & Zheng, B. . (2021). Applying a random projection algorithm to optimize machine learning model for breast lesion classification. IEEE Transactions on Biomedical Engineering, PP(99), 1-1, https://doi.org/10.1109/TBME.2021.3054248

[24] Nardini, A. , & Brierley, G. . (2020). Automatic river planform identification by a logical-heuristic algorithm. Geomorphology, 375(1–2), 107558, https://doi.org/10.1016/j.geomorph.2020.107558

[25] Yan, J. , Lin, S. , Kang, S. B. , & Tang, X. . (2015). Change-based image cropping with exclusion and compositional features. International Journal of Computer Vision, 114(1), 74-87, DOI: https://doi.org/10.1007/s11263-015-0801-5

[26] Elsanadily, S. , Mahran, A. , & Elghandour, O. . (2018). Classification-based algorithm for bit-flipping decoding of gldpc codes over awgn channels. IEEE Communications Letters, PP(99), 1-1, DOI: 10.1109/LCOMM.2018.2840146

[27] Sharma, A., Kumar, R., & Bajaj, R. K. (2021). On Energy-constrained quickest path problem in green communication using intuitionistic trapezoidal fuzzy numbers. *Recent Advances in Computer Science and Communications (Formerly: Recent Patents on Computer Science)*, *14*(1), 192-200, DOI: https://doi.org/10.2174/2213275911666181025125224

[28] Bahadure, N. B. , Ray, A. K. , & Thethi, H. P. . (2018). Comparative approach of mri-based brain tumor segmentation and classification using genetic algorithm. Journal of Digital Imaging, 31(1), 1-13, DOI: DOI: 10.1007/s10278-018-0050-6

[29] Redzic, M. , Laoudias, C. , & Kyriakides, I. . (2019). Image and wlan bimodal integration for indoor user localization. IEEE Transactions on Mobile Computing, 19(99), 1109-1122, DOI: DOI: 10.1109/TMC.2019.2903044

[30] Zhao, D. , Liu, S. , Yang, X. , Ma, Y. , & Chu, W. . (2021). Research on camouflage recognition in simulated operational environment based on hyperspectral imaging technology. Journal of Spectroscopy, 2021(2), 1-9, DOI: https://doi.org/10.1155/2021/6629661

[31] Poongodi, M., Sharma, A., Vijayakumar, V., Bhardwaj, V., Sharma, A. P., Iqbal, R., & Kumar, R. (2020). Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Computers & Electrical Engineering*, *81*, 106527, DOI: https://doi.org/10.1016/j.compeleceng.2019.106527

[32] Ahmed, I. , Ali, R. , D Guan, Lee, Y. K. , Lee, S. , & Chung, T. C. . (2015). Semi-supervised learning using frequent itemset and ensemble learning for sms classification. Expert Systems with Applications, 42(3), 1065-1073, DOI: https://doi.org/10.1016/j.eswa.2014.08.054

[33] Kumar, C., Singh, A. K., Kumar, P., Singh, R., & Singh, S. (2020). SPIHT-based multiple image watermarking in NSCT domain. *Concurrency and Computation: Practice and Experience*, *32*(1), e4912, DOI: https://doi.org/10.1002/cpe.4912

[34] Dadaneh, B. Z. , Markid, H. Y. , & Zakerolhosseini, A. . (2016). Unsupervised probabilistic feature selection using ant colony optimization. Expert Systems with Applications, 53(Jul.), 27-42, https://doi.org/10.1016/j.eswa.2016.01.021