Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 6, pp. 4960–4966, DOI 10.12694/scpe.v25i6.3332

ALGORITHM IDENTIFICATION AND INTEGRATED WITH PUSH SERVICE FOR TELEMEDICINE SYSTEM

CAI YAN-LING*, LI XIN-YU†, AND KUMAR KANNAN‡

Abstract. Telemedicine systems, while overcoming physical space constraints, often lack personalized interactions. By incorporating a push service and leveraging prediction-oriented algorithms, these systems can offer an improved user experience. Such enhancements enable timely treatment options and reduce unnecessary resource usage in on-site outpatient clinics. This research work starts by creating a robust algorithm using data mining techniques. Next, it establishes the foundation for a telemedicine push service. The service includes essential modules for disease differentiation, doctor recommendations, and diagnosis predictions. To optimize these modules, a merged algorithm combining k-nearest neighbor classification, nearest neighbor recommendation, and FP-growth is needed. This work aims to enhance treatment options for patients and streamline resource usage in on-site outpatient clinics. Moreover, this work has carried out empirical research for identification of algorithm by using available data at a public Chinese telemedicine system. The results of data analysis show the follows: 1. For disease diagnosis, the KNN model (k=1) is more accurate but less efficient, SVM and LibSVM are more efficient but less accurate than the KNN model; 2. In terms of doctor recommendation, nearest neighbor recommendation performs better but is not as efficient as matrix factorization; 3. in diagnostic prediction, the combination of introducing association mining and data segmentation can play a better role. The developed algorithm and its conclusions from this study could make easier and more efficient to provide treatment options for undecided-condition patients.

Key words: sign-nonsingular matrix, LU-factorization, indicator polynomial

1. Introduction. According to the World Bank, about 10 percent of gross domestic product is spent on healthcare every year. Telemedicine has been a treatment mode with the development of information communications technology, which provides more access to medical resources [1, 2]. At present, there are many telemedicine systems supported by various organizations [3-5], but most of them lack the function of active interaction with users and fail to provide personalized recommendation. Therefore, by adding push service to the telemedicine system, it can provide personalized service, which improves user experience and enhances user stickiness.

For an old-mode telemedicine system, it has been accumulating abundant data. Mining user information, doctor information and diagnostic records can realize a push service. On one hand, according to the user's historical data, the department and doctor when he reserves for a return visit and suggestions for daily life are recommended. On the other hand, for some unanswered questions, this study will provide users with information about proper department and doctors, and users can directly consult them. In addition, valuable advice about the question is listed.

For a new telemedicine system, it can build its push service by directly using model trained by other systems, or by adjusting some parameters on the model. This paper aims at introducing personalized recommendation into telemedicine system, and designing integrated algorithms to adapt to different conditions and corresponding results based on real data sets through data mining.

2. Literature Review. The algorithm design for push services in telemedicine systems involves creating efficient mechanisms to deliver timely notifications and updates to users, ensuring seamless communication and information flow [6].

In mainland China, a qualitative study explored family caregivers' perspectives on telemedicine-based services for patients with end-of-life cancer. Key findings highlighted motivations for using telemedicine, supportive

*School of Management, Zhengzhou University, Zhengzhou, PR China 450001.
† International Business Program, Leeds University Business School University of Leeds, Leeds, United Kingdom.
‡School of Computer Science and Engineering, VIT, Vellore, India 632014

Table 3.1: Algorithm Comparison.

Algorithm	Strength	Weakness
ANN	Non-linearity modelling	Block box in nature
Decision Tree	Intuitive result, understand easily	complexity in node selection
Neural network	Solve problems with omplex internal mechanisms	Hard to determine the number of hidden layers
Bayesian classification	Solid mathematical foundations and stability	Assumption isn't applicable and large computation
Rule-based	Obtained rule has great value	Rule is rare and hard to obtain
SVM	Reduce computational complexity	relatively large computation
KNN	Principal fit data situation	Accuracy depends on data size

care needs, and functional expectations of telemedicine platforms. The study underscores the importance of addressing caregivers' unique needs in end-of-life care programs. [7]

In a qualitative study conducted in Germany, researchers interviewed (tele-)medical experts to gain insights into their experiences with adopting telemedicine within the healthcare system. The study uncovered essential themes related to persuasion, knowledge acquisition, implementation processes, decision-making, and confirmation, offering valuable guidance for ongoing telemedicine implementation strategies [8].

Dash et al. investigated the factors influencing telemedicine adoption. They employed multiple regression and artificial neural network (ANN) approaches to identify key motivators for accepting telemedicine during the pandemic [9].

3. Algorithm Identification. There are many algorithms about classification, recommendation and prediction, and whether the algorithm is appropriate or not depends on actual data. Then, although telemedicine system is quite different, the data will be similar [10]. Therefore, the first step is to rule out some algorithms by general law, and then carry out empirical research to compare the remaining algorithms.

Information in the telemedicine system mainly includes user information, doctor information and diagnostic records. User information generally includes user name, gender, age, and region and so on. Doctor information generally includes doctor name, title, and skilled field and so on. Diagnosis records generally includes symptoms, doctor advice, and time and so on. Therefore, there will be dozens of attributes that can be extracted from a telemedicine system. In addition, considering that remote interrogation is the most common system, it is chosen in this research. In this system, diagnostic records typically consist of unstructured text data.

3.1. Disease Differentiation. In the push service, it is the primary job to help patients determine the department according to their symptoms. In order to express conveniently, disease differentiation is used to describe the process. The symptoms in the system are described in words, so the word vector model should be constructed by word segmentation, and then the number of attributes will increase, generally over 100.

There are many classification algorithms, including artificial neural networks decision tree, neural network, Bayesian classification, support vector machine (SVM), rule-based classification and k-nearest neighbor classification (KNN) [11]. In view of the general situation of data, the paper selects support vector machine and k-nearest neighbor classification. The comparison of algorithms in this study is shown in table 3.1

Although artificial neural networks have some self-learning ability, the model may cause overfitting problems when the samples are too small. In addition, ANN has poor interpretability and is not able to explain well the reasons for the results.

In decision tree algorithm, internal node selection will be quite complex and need large calculation because of too many attributes. Some previous literature has used decision trees to make a diagnosis for COVID-19. At runtime, this algorithm is not dynamic enough because of the absence of circular references and feedback loops in the decision tree.[12] In addition, numerous leaves and great depth make it lose the advantage of intuitive result. More importantly, for many word vectors such as wound, inflammation and bleeding can't clearly point to a department. Therefore, error of using the algorithm will be larger.

Although neural network algorithm can simulate the human brain to classify the problem, the number of the hidden layers will be very difficult to determine. Few words like rhinitis can point to ENT department, but most of them need feedbacks between different layers of neurons to reach the department, so two coexisting cases will greatly decrease the effectiveness and efficiency of the algorithm.

Some studies have argued that some keywords do not reflect all the information regarding BNs in healthcare.[13] Furthermore, naïve Bayesian classification algorithm assumes that an attribute value is independent on a given class. In this study, there are apparently dependencies between attributes, and conditional probability isn't 0. Bayesian belief networks eliminates the assumption, and uses conditional probabilities table and a direct acyclic graph. Because of numerous attributes, the table and graph will be complex, and frequently used greedy algorithm has a large error when class labels are many.

Rule-based classification algorithm introduces association mining into classification. Although the frequent itemsets which are directly found between a class label and attributes have a certain value, they are rare in the data and have a small applicable scope. Therefore, it is difficult to achieve high accuracy.

SVM uses a kernel function to transform the data into a higher dimension, and searches for the optimal separation hyperplane by solving

$$\begin{gathered} \max \tfrac{1}{\|\omega\|} \\ s.t.yi\left(\omega^T xi + b\right) \geq 1, i = 1 \cdots n \end{gathered} \tag{3.1}$$

A small number of support vectors are used for classification, so to a certain extent, it can avoid exponential increases in calculation owing to the increase of dimensionality. However, because of involving matrix calculations, it needs a relatively large calculation and long time.

Although it is hard to determine the class directly through few word vectors, there is no doubt that word vectors between the same department are more similar than different departments, which fits the principal of KNN. By setting k neighbors, the test data are classified by the classes of them. In addition, it is important to note that KNN doesn't construct a complete model, and accuracy depends on data size.

3.2. Doctor Recommendation. Common recommendation methods include collaborative filtering, content-based, interaction-based and hybrid recommendation [14]. In view of the general situation of data, the paper selects collaborative filtering including nearest neighbor recommendation and matrix factorization [15].

Content-based recommendation recommends doctors who are similar to those in user's history, which is very effective in marketing. However, in telemedicine system, the next question has a big probability that it is different from the last one. In addition, it is a problem whether historical questions are enough to work. Therefore, mean and variance of the data should be carefully evaluated before using the method.

Interaction-based recommendation recommends doctors through several questions. However, there are hundreds of doctors even in a small telemedicine system, so it is hard to finish personalized recommendation merely by several questions which are often used to filter data.

Collaborative filtering recommends doctors by collecting similar users' preferences, and includes nearest neighbor recommendation and matrix factorization. The former is similar to KNN, and the latter is especially suitable for numerous attributes and uses the feature vector of user and item to effectively improve speed.

$$\hat{\mathrm{x}}ui = \langle \omega u, hi \rangle = \sum_{f=1}^{k} \omega uf \cdot hif \tag{3.2}$$

3.3. Diagnosis Prediction. Problem descriptions and doctor advice both consist of unstructured text data. To predict diagnosis, this research firstly segments words to construct word vector model, and then mine association rules between the two types. [16]

Common algorithms in association mining include Apriori and FP-growth. Apriori needs to frequently construct candidate itemsets, and then select frequent itemsets. To finish the process, it is necessary to scan original data for many times, which makes efficiency low. However, FP-growth algorithm only requires two scanning in the construction of FP-tree. In addition, the same prefix can be shared in the mining process, which also greatly improves efficiency.

Fig. 4.1: Disease differentiation.

4. Empirical Research. By fetching the real data of a telemedicine system, the paper carries out an e empirical research. 120ask (www.120ask.com) has been in operation for over a decade and owns huge information resources. This research grabs 65862 data of surgery, internal medicine and ENT department from 2005 to 2017. Considering the differences between children and adults, and removing invalid data, finally 61646 data are used for research. Each data has eleven attributes including user name, age, gender, region, time, problem description, doctor name, title, help number, like number and doctor advice. In addition, data preprocessing is completed by using ICTCLAS2013 to construct word vector model.

4.1. Disease Differentiation. Disease differentiation is to divide questions to department, and the study uses three first-level departments and nineteen second-level departments respectively. With a preliminary screening job by the authors, the employed algorithms include support vector machine (SVM) and k-nearest neighbor classification (KNN), and LibSVM which is developed by professor Lin Chin-Jen is also used as a reference [17-20].

The operation time is converted to the index EFF, and higher EFF indicates higher efficiency and less time.

$$EFF = 0.5 + 0.5 \times \frac{time\min}{time} \tag{4.1}$$

In figure 4.1, it can be observed that KNN performs best in terms of accuracy. LibSVM shows its advantage in speed, but fails to improve performance. Compared with first-class departments, the accuracy in the classification of second-level departments decreases, but the basic trend remains unchanged. In addition, the number of class labels greatly affects KNN, and the increase of number can lead to a sharp increase in time.

Figure 4.2 shows that K and kernel function are important parameters in KNN and SVM, respectively. Considering the operation time, more appropriate approach is to use KNN and LibSVM to study the classification of first-class departments. K performs best when K is equal to 1 and needs the shortest time. Linear kernel function has the best performance and polynomial kernel function performs the worst.

4.2. Doctor Recommendation. This study uses nearest neighbor recommendation and matrix factorization (MF) to recommend doctors. The former is similar to KNN in classification, but class label becomes the doctor. The number of classes increases greatly, so only depending on several objects will make error large. According to Herlocker's empirical research, it is reasonable to set K between 20 and 50 [21,22]. Bayesian personalized ranking matrix factorization (BPRMF) and weighted regularization matrix factorization (WRMF) are common models in matrix factorization and use them to study.

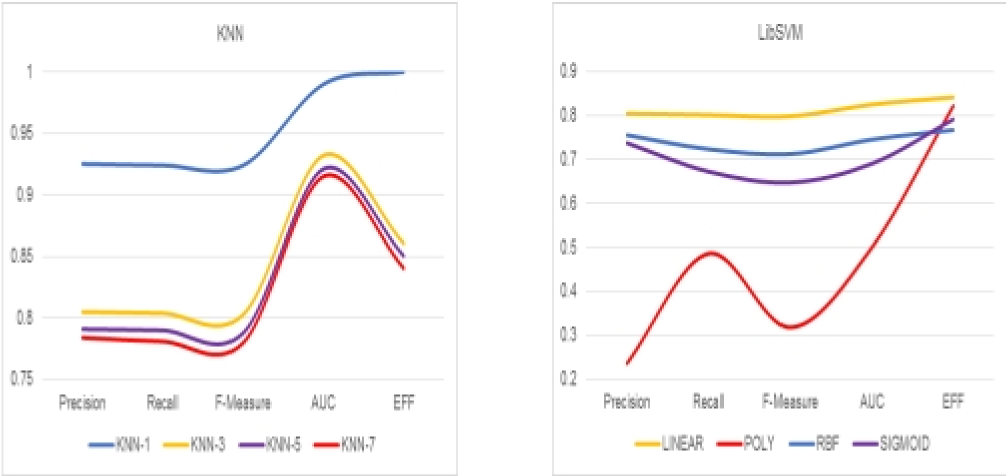

Fig. 4.2: Disease differentiation.

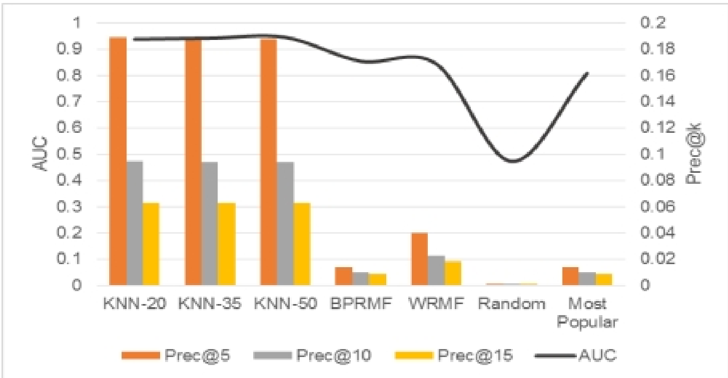

Fig. 4.3: Recommendation comparison.

In terms of accuracy, shown in figure 4.3, nearest neighbor recommendation performs better than MF, and they are both better than non-personalized methods including random and most popular. MF algorithm is particularly suitable for sparse matrix, but it can't show its advantage in the study because of few attributes. Whether K takes 20, 35 or 50, the performance is excellent. In MF, WRMF is better than BPRMF.

In terms of operation time as shown in Table 4.1, the difference is large, and nearest neighbor recommendation is much higher than matrix factorization.

4.3. Diagnosis Prediction. For problem description and doctor advice, word segmentation is used to construct word vector model. To distinguish them, the letter Q and S are respectively added. Therefore, diagnostic prediction is to find the strong association rules between Q and S, and rule antecedent only contains word vector with Q, and rule consequent only contains word vector with S.

This research constructs a model with 150 word vectors with Q and 140 word vectors with S after it is filtered, use FP-growth to mine strong association rules. The purpose is to discover advice according to the symptom, so confidence is set as an important parameter and reduce the significance of support. Setting confidence threshold to 0.5, it is difficult to obtain effective rules. By using the whole data to research causes the failure, because similar symptoms may correspond to different diseases, and correspond to different doctor advice. To solve the problem, this study proposes two methods including threshold adjustment and data

Table 4.1: Operation time

	KNN-20	KNN-35	KNN-50	BPRMF	WRMF	Random	Most Popular
Time	1h	1,5h	2h	1min	0.5min	10s	10s

segmentation.

Although similar symptoms may correspond to different advice, higher confidence still indicates higher value, so the vital part is to make threshold adaptive to find relatively valuable rules. The process can respectively generate 2 and 9 effective rules when threshold is adjusted to 0.25 and 0.20. However, the fundamental cause is too much data, so the solution is that only use the data belonging to general surgery department. When confidence threshold is set to 0.5, there will be 67 strong association rules, which is a significant improvement in effect and efficiency.

5. Conclusion. Based on the analysis of the present situation of telemedicine, this paper constructs a comprehensive push service system of telemedicine that provides disease differentiation, doctor recommendation and diagnosis prediction; in this push service, an integration of various methods of data mining has been done, and the work of its algorithm design and coding has been completed.

In disease differentiation, KNN has the best accuracy especially when K is equal to 1. However, operation time increases dramatically when class labels increase, so SVM or LibSVM can be considered to improve efficiency at the cost of some accuracy. In doctor recommendation, nearest neighbor recommendation performs better than matrix factorization, but the time consumption is a hundred times. In diagnosis prediction, there is a good result by introducing association mining into it, and this research finds that more valuable association rules can be obtained by data segmentation. In addition, adaptive confidence threshold is also helpful in the research.

The result from this method has a strong meaningful value and can also better provide viable methods for upgrading and enhancing the next-generation healthcare system in the future, which hopefully can then save more resources to improve the efficiency of access to healthcare. Through case comparisons, it can be observed the performance of the push service is varied with the work quality of words segmentation. In the future work, a specific professional segmentation system in medical domain will be integrated into the push service for for higher efficiency by relacing the current general segmentation system.

REFERENCES

[1] Tabaeeian, R.A., Hajrahimi, B. and Khoshfetrat, A. (2022). A systematic review of telemedicine systems use barriers: primary health care providers' perspective. Journal of Science and Technology Policy Management.

[2] Kaur, G., Sharma, D. and Kaur, V. (2016). Telemedicine in Transient Phase: Emergence of M-Health Care Services. Indian Journal of Science and Technology, 9(15).

[3] Zhong, Y., Xu, Z. and Cao, L. (2021). Intelligent IoT-based telemedicine systems implement for smart medical treatment. Personal and Ubiquitous Computing.

[4] Wu, H. (2023). Sharing and Cooperation of Improved Cross-Entropy Optimization Algorithm in Telemedicine Multimedia Information Processing. *International Journal of Telemedicine and Applications*, 2023, pp.1–14.

[5] Ren, X., Wang, Z., Wu, Y., Li, Y., Chen, M., Zhai, Y. and Li, Y. (2019). Design and Implementation of a Message-Based Regional Telemedicine System to Achieve High Availability and Scalability. *Telemedicine and e-Health*, 25(3), pp.243–249.

[6] Cai, Y., Su, Y., Zhao, L., & Cui, Z. (2018, May). Algorithm Design of Push Service for Telemedicine System. In 2018 International Conference on Robots & Intelligent System (ICRIS) (pp. 250-252). IEEE.

[7] Guo, J., Xu, X., Liu, C., Wang, Y., & Chen, Y. (2024). Perspectives of telemedicine-based services among family caregivers of patients with end-of-life cancer: a qualitative study in mainland China. BMC Palliative Care, 23(1), 1-9.

[8] Rauner, Y., & Stummer, H. (2024, January). Adoption Processes of Innovations in Health Systems: The Example of Telemedicine in Germany. In Healthcare (Vol. 12, No. 2, p. 129). MDPI.

[9] Dash M, Shadangi PY, Muduli K, et al. Predicting the motivators of telemedicine acceptance in COVID-19 pandemic using multiple regression and ANN approach. J Stat Manage Syst. 2021;24(2):319–339. doi:10.1080/09720510.2021.1875570

[10] J. Jayapradha, S. Boovaneswari, S. Sabarivadivelan, D. Uvarajan and S. Sarathi, "A Telehealth System Driven by Artificial Intelligence for Effective Patient Consultation and Diagnosis in Hospitals," *2023 International Conference on System, Computation, Automation and Networking (ICSCAN)*, PUDUCHERRY, India, 2023, pp. 1-6, doi: 10.1109/ICSCAN58655.2023.10395119.

[11] Jiawei Han, Micheline Kamber and Jian Pei. 2012. *Data mining concepts and techniques(Third Edition)*. China Machine Press, pp. 243-248.

[12] Chrimes, D. (2022). Using Decision Tree towards Expert System for Decision-Support for COVID-19 (Preprint). Interactive Journal of Medical Research. doi:https://doi.org/10.2196/42540.

[13] McLachlan, S., Dube, K., A Hitman, G., Fenton, N. and Kyrimi, E. (2020). Bayesian Networks in Healthcare: Distribution by Medical Condition. Artificial Intelligence in Medicine, p.101912. doi:https://doi.org/10.1016/j.artmed.2020.101912.

[14] Dietmar Jannach, Markus Zanker, Alexander Felfernig and Gerhard Friedrich. 2010. *Recommender Systems: An Introduction*. Cambridge University Press, pp. 121-123.

[15] Maryam Khanian Najafabadi and Mohd Naz'ri Mahrin. 2016. "A systematic literature review on the state of research and practice of collaborative filtering technique and implicit feedback," *Artificial Intelligence Review.*, 45(2):167-201.

[16] I. G. S. M. Diyasa, A. Prayogi, I. Y. Purbasari, A. Setiawan, Sugiarto and P. A. Riantoko, "Data Classification of Patient Characteristics Based on Nutritional Treatment Using the K-Nearest Neighbors Algorithm," *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, Bandung, Indonesia, 2021, pp. 1-6

[17] Li Kun, Liu Peng, Lv Yajie, Zhang Guopeng and Huang Yihua. 2016. "The parallel algorithms for LIBSVM parameter optimization based on Spark,"*Journal of Nanjing University Natural Sciences*, 52(2):343-352.

[18] Li, S., Yang, L., Chen, P. and Yang, C. (2019). Natural Neighbor Algorithm for Breast Cancer Diagnosis. Journal of Physics: Conference Series, 1395(1), p.012003.

[19] I Ketut Agung Enriko, Muhammad Suryanegara and Gunawan, D. (2018). Heart Disease Diagnosis System with k-Nearest Neighbors Method Using Real Clinical Medical Records.

[20] I. K. A. Enriko, M. Suryanegara and D. Gunawan, "My Kardio: A telemedicine system based on machine-to-machine (M2M) technology for cardiovascular patients in rural areas with auto-diagnosis feature using k-Nearest Neighbor algorithm," *2018 IEEE International Conference on Industrial Technology (ICIT)*, Lyon, France, 2018, pp. 1775-1780

[21] Jon Herlocker, Joseph A. Konstan and John Riedl. 2002. "An empirical analysis of design choices in neighborhood-based collaborative filtering algorithms," Information Retrieval, 5(4):287-310.

[22] Uddin, S., Haque, I., Lu, H., Moni, M.A. and Gide, E. (2022). Comparative performance analysis of K-nearest neighbour (KNN) algorithm and its different variants for disease prediction. Scientific Reports, 12(1).

Edited by: Jingsha He
Special issue on: Efficient Scalable Computing based on IoT and Cloud Computing
Received: Feb 2, 2024
Accepted: Apr 7, 2024