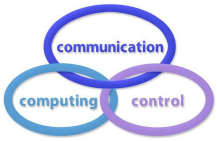
INTERNATIONAL JOURNAL OF COMPUTERS COMMUNICATIONS & CONTROL
Online ISSN 1841-9844, ISSN-L 1841-9836, Volume: 18, Issue: 1, Month: February, Year: 2023
Article Number: 5011, https://doi.org/10.15837/ijccc.2023.1.5011



CCC Publications

Ensemble Learning for Interpretable Concept Drift and Its Application to Drug Recommendation

Yunjuan Peng, Qi Qiu, Dalin Zhang, Tianyu Yang, Hailong Zhang

Yunjuan Peng
School of Software Engineering
Beijing Jiaotong University, Beijing, China
*Corresponding author: 21126333@bjtu.edu.cn

Qi Qiu
1. Department of Pharmacy
Beijing An Zhen Hospital, Beijing, China
2. School of Pharmaceutical Sciences
Capital Medical University, Beijing, China
*Qi Qiu and Yunjuan Peng are co-first authors of the paper: qiuqi8133@163.com

Dalin Zhang
School of Software Engineering
Beijing Jiaotong University, Beijing, China
*Corresponding author: dalin@bjtu.edu.cn

Tianyu Yang
Department of Electrical Engineering
Columbia University, NYC, USA
ty2462@columbia.edu

Hailong Zhang
Pamplin College of Business
Virginia Polytechnic Institute and State University, Blacksburg, VA, USA
hailongzhang@vt.edu

Abstract

During the COVID-19 epidemic, the online prescription pattern of Internet healthcare provides guarantee for the patients with chronic diseases and reduces the risk of cross-infection, but it also raises the burden of decision-making for doctors. Online drug recommendation system can effectively assist doctors by analysing the electronic medical records (EMR) of patients. Unlike commercial recommendations, the accuracy of drug recommendations should be very high due to their relevance to patient health. Besides, concept drift may occur in the drug treatment data streams, handling drift and location drift causes is critical to the accuracy and reliability of the recommended results. This paper proposes a multi-model fusion online drug recommendation system based on the association of drug and pathological features with online-nearline-offline architecture.

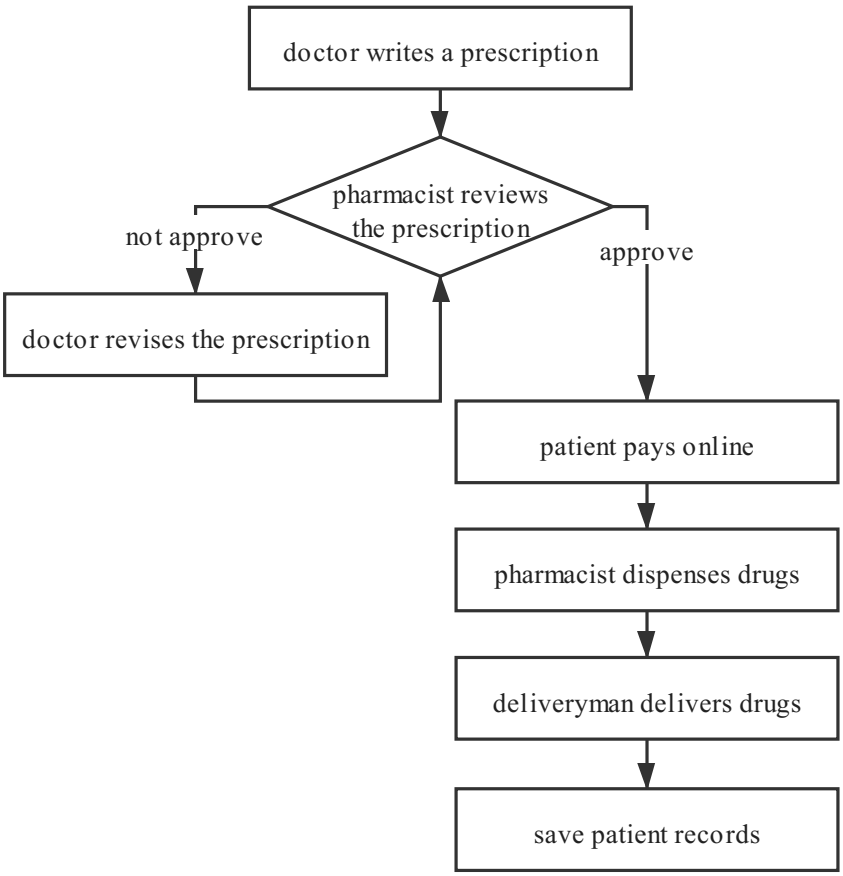


Figure 1: Online prescription process

The system transforms drug recommendation into pattern classification and adopts interpretable concept drift detection and adaptive ensemble classification algorithms. We apply the system to the Percutaneous Coronary Intervention (PCI) treatment process. The experiment results show our system performs nearly as good as doctors, the accuracy is close to 100%.

Keywords: Interpretable Concept Drift, Self-adaptive Ensemble Learning, Drug Recommendation, Pattern Classification.

1 Introduction

Since January 2020, with the fast spread of coronavirus (COVID-19), Internet healthcare is encouraged to release the burden on clinics. Hospital information technology plays an important role in remote consultation. For example, Beijing Anzhen Hospital adds video and audio consultation methods on the mobile phone by using the existing hospital information system (HIS) to complete the follow-up diagnosis of chronic diseases on the Internet. The online prescription pattern [1] is shown in Figure 1. The doctor prescribes according to the data uploaded by the patient and the diagnosis results, then the prescription will be reviewed by the pharmacist. After the online consulting payment is finished, the pharmacist dispenses the drugs according to the prescription and delivers them to the deliveryman who delivers the drugs to the patient. Online prescription service provides guarantee for the patients with chronic diseases and reduces the risk of cross-infection. However, the latest survey found that doctors often need to spend much time communicating with patients and reviewing historical medication records for accurate prescriptions due to the lack of face-to-face consultations in the fight against COVID-19.

The current study of Internet healthcare has done less work at the algorithm for doctors' prescription decision-making. Internet healthcare reduces gathering risks and travel costs and provides convenience for patients [2], but increases the burden for doctors, especially decision-making. The existing AI-assisted consultation system can independently diagnose some diseases, but for complex

common diseases, it still needs to entirely rely on the doctors' knowledge and experience. EMR records the historical medical data of patients. By analyzing the key attributes and relationships in prescriptions, scientific and rational decision support for doctors can be provided [3, 4]. Drug recommendation system has great application value by analysing EMR to assist doctors in diagnosis, without being limited by the complexity of diseases. Therefore, this paper combines the needs of Internet healthcare under COVID-19 to study the drug recommendation system, so as to relieve the pressure of doctors and improve the efficiency of diagnosis.

Current drug recommendation research is focused on expert systems. For example, a method for mining clinical pathways was proposed based on LDA and PST, which produces daily document of medication effect after fusing drug effects recorded in Drug Bank with prescriptions [5]; a decision support system was developed that helps doctors select appropriate first-line drugs, which classifies patients' abilities to protect themselves from infectious diseases as a risk level for infection [6]. Applications of machine learning and data mining can change the available data to valuable information that can be used for recommending appropriate drugs by analyzing symptoms of the disease. A machine learning approach for multi-disease with drug recommendation is proposed to provide drug recommendations for the patients suffering from various diseases [7]. Implicit feedback and crossing recommendation method was put forward [8], which builds up the relationship between patients' symptoms and doctors' medication scheme by analyzing medical history. However, the above methods have been limited mainly for the following reasons: (1) over-reliant on expert knowledge; (2) the accuracy is not high enough; (3) they are all offline methods that do not adapt well to new patients and symptoms.

Medical data is a continuous, real-time and high-dimensional data stream. Since the concepts implied in the data stream may change in some way over time, i.e., concept drift [9], it increases the difficulty of data mining. In addition, interpretability and predictive capabilities are greatly reduced due to the causes of drift reason cannot be accurately identified, posing a threat to the reliability of diagnostic results. Specifically, when prescribing for a patient, if the recommendation is just rely on the historical static EMR without dynamically identifying changes in drug properties, the system will not be able to make the correct decision; if it does not provide an accurate explanation for the reasons of the changes, the decision result provided by the system will not be trusted. A good learning model not only needs to process the incoming data in real time, but also adapt to the constant changes in concepts and get the reasons of the changes, such as changes in patient pathological characteristics and the introduction of new drugs.

When making recommendations for one patient, in addition to the standard clinical pathways and healthcare process models, the physical status of the patient should also be considered [10]. Machine learning outperforms in mining valid attributes and their correlations without requiring any prior knowledge [11] and has been widely used in the medical field. Ensemble learning performs well in dealing with concept drift, and it forms the final prediction results by maintaining sub models and using combination strategies [12], which can improve the prediction performance effectively.

Based on the above analysis, this paper proposes an online drug recommendation system for Internet healthcare, which adopts the interpretable concept drift detection and adaptive ensemble classification algorithms. At first, multiple classifiers are trained separately as basic classifiers, which produce the results using combined voting strategy; then, new classifiers are built up based on the medical treatment data collected by fixed time step, concept drift and the reason is identified, and the basic classifier collection is updated by self-adaptive ensemble strategy. The main contributions of this paper are conducted as follows:

- A multi-model fusion online drug recommendation system for Internet healthcare based on the association of drug and pathological features is proposed for the first time, which transforms prescription recommendation into pattern classification. Pattern-based classification achieves higher accuracy and can solve the problem of missing values well, which can remove redundant information from data and be not affected by noise.

- Our design comes from the real requirements of the Internet healthcare. The system is designed to adopt an online-nearline-offline architecture that can quickly provide drug recommendation service.

- Interpretable concept drift detection algorithm and online adaptive ensemble learning strategy are proposed for the characteristics of medical data, which improves the recommendation accuracy and reliability.
- In terms of specific application, we apply the system to the PCI treatment process for recommending statins that are used for preoperative pretreatment and postoperative lipid-lowering. The experiment results show our system performs nearly as good as doctors, the accuracy is close to 100%, which shows the online drug recommendation system can effectively help the doctors in fighting against COVID-19.

Subsequent sections are organized as follows: Section 2 systematically reviews recommendation systems, medical recommendation systems and concept drift in recommendation systems; Section 3 introduces the proposed method in detail; Section 4 describes the experiments; Section 5 presents detailed experiment results; lastly, a summary and future work are discussed in Section 6.

2 Related Work

2.1 Recommendation System

Recommendation methods can be divided into content-based, collaborative filtering, and hybrid methods [13]. Content-based methods use the content of items to create features and attributes to match user profiles. Most of the methods draw on the experience of information retrieval and filtering methods to some extent. In addition to the combination of the traditional vector space model and the k-nearest neighbor, there are also some methods based on machine learning such as Naive Bayes, Decision Tree, Linear Classification and Neural Network, which establish preference model and then use the model to predict the probability of users' future behavior to complete the recommendation. Collaborate filtering methods are based on group knowledge and users or items with similar interests. They do not need to analyze the content of the project. On the contrary, they rely on the coding of the relationship between the user and the project and is reflected in the rating feedback matrix. Each element represents the user rating of one user for one project. User-based collaborative filtering methods have certain advantages in the novelty of recommended results, such as the study [14], which combines word frequency statistics and similarity calculation to analyze users' preferences, however, the recommendation results are easily influenced by trends. At the same time, new users or low-active users will also encounter the thorny problem of user cold start. The above two methods can be combined to obtain a hybrid recommendation system. There are two general hybrid ideas: (1) hybrid of recommendation results: the results of two or more recommendation systems are directly combined by a mixing mechanism. Common mixing methods include crossover, weighting, switching and so on; (2) hybrid of recommendation algorithms: a new hybrid algorithm is designed based on multiple algorithms. Generally speaking, the models of various algorithms are connected in series, such as the output of the first algorithm is used as the input of the second algorithm.

2.2 Medical Recommendation System

Medical recommendation systems mainly include knowledge-driven expert and data-driven drug recommendation systems. A diabetes drug recommendation system was proposed based on the domain ontology [15]. Zhang et al. [16] proposed a hybrid recommendation framework that integrates artificial neural networks and case-based reasoning. A Knowledge graph-based drug recommendation system [17] and a mutual information clustering-based recommendation method [18] were proposed to assist in TCM diagnosis. Considering the impact of incomplete knowledge graphs on the robustness of recommendation systems, Gong et al. [19] constructed a heterogeneous graph containing diseases, drugs, patients and their correspondence based on EMR and medical knowledge graphs to decompose drug recommendation into linkage prediction problem and demonstrated their effectiveness.

In Internet hospital, doctors know the patients' physical condition through pictures, words and so on. Prescriptions are usually completely dependent on doctors' knowledge and working experience, resulting in the high decision-making costs. Online drug recommendations are used to recommend the

most suitable drugs and corresponding dosage by mining EMR and combining new data on patient pathological characteristics to relieve doctors' pressure and reduce the risk of medical accidents. However, there are few researches on online prescription recommendation. The current work of Internet healthcare is mainly focused on medical insurance payment service, medical quality supervision and AI chronic disease management, with little work on assisting doctors in prescription decision-making due to the lack of algorithm support.

This paper aims to improve the efficiency of medical consultations by reducing the prescription time and improving the accuracy of prescriptions. To do this, we propose a three-layer intelligent model based on adaptive learning that recommends the most suitable prescription based on the parameters of patients' pathological characteristics.

2.3 Concept Drift In Recommendation System

Concept drift in recommendation systems refers to the fact that the recommendation model can not accurately grasp the change of concept when the data accumulates over time, resulting in the results deviating from the actual demand [20]. In Internet healthcare, the online drug recommendation system is often affected by various factors, such as the addition of new drugs and changes in patient pathological status. The recommendation performance will be affected when the algorithm cannot adapt to these factors in time.

The current solutions to concept drift of recommendation system mainly focus on the behavior records between temporal features and user items. For example, a time weight collaborative filtering algorithm proposes an exponential decay function to calculate the weight of scoring prediction [21]. A method to calculate the distance between two preference data stream distributions in adjacent time windows is proposed to model the degree of user preference change [22]. These methods use the same half-life or similar time windows for all users and do not consider the pattern of interest changes over time for different users. There is not much theoretical and experimental basis for the size of time window and half-life, which can only be adjusted based on experimental results.

In machine learning, methods to solve concept drift can be divided into sample selection, sample weighting and sliding window [23]. Sample selection defines a correlation index to select the most relevant samples to train the model; sample weighting considers that samples at different periods are not equally important to the model; sliding window sets a time window to select a sample that is relatively new. Although these methods have extensions in recommendation systems, they are not the best choices at present: recommendation systems need to be built on as much data as possible, sample selection will increase data sparsity; artificial weighting and windows lack theoretical basis.

Methods based on other models have also been applied to solve the concept drift problem of recommendation systems. For example, Viniski et al. [24] proposed incremental learning to update the user-item relationship established in a streaming recommendation system; multitransition factor and a forgetting time function were introduced to analyze the evolution of user preferences in order to accurately recommend new items or services to users [25].

Through our analysis, these methods can not solve concept drift in online drug recommendation systems well. The problems of them can be summarized as follows:

- The important parameters in the algorithm are determined subjectively and lack of theoretical basis.
- Increases data sparsity, which affects the recommendation effect.
- Unable to adapt to the incoming samples. Adapting the model to different sample windows with the arrival of data flow is not easy.
- The identified concept drift is lack of interpretability. It's important to explain and investigate the causes of drift, so as to ensure the reliability of drug recommendation system.
- The accuracy of the method is not high. The drug recommendation system is related to the safety of medication and health of patients so the high accuracy is required.

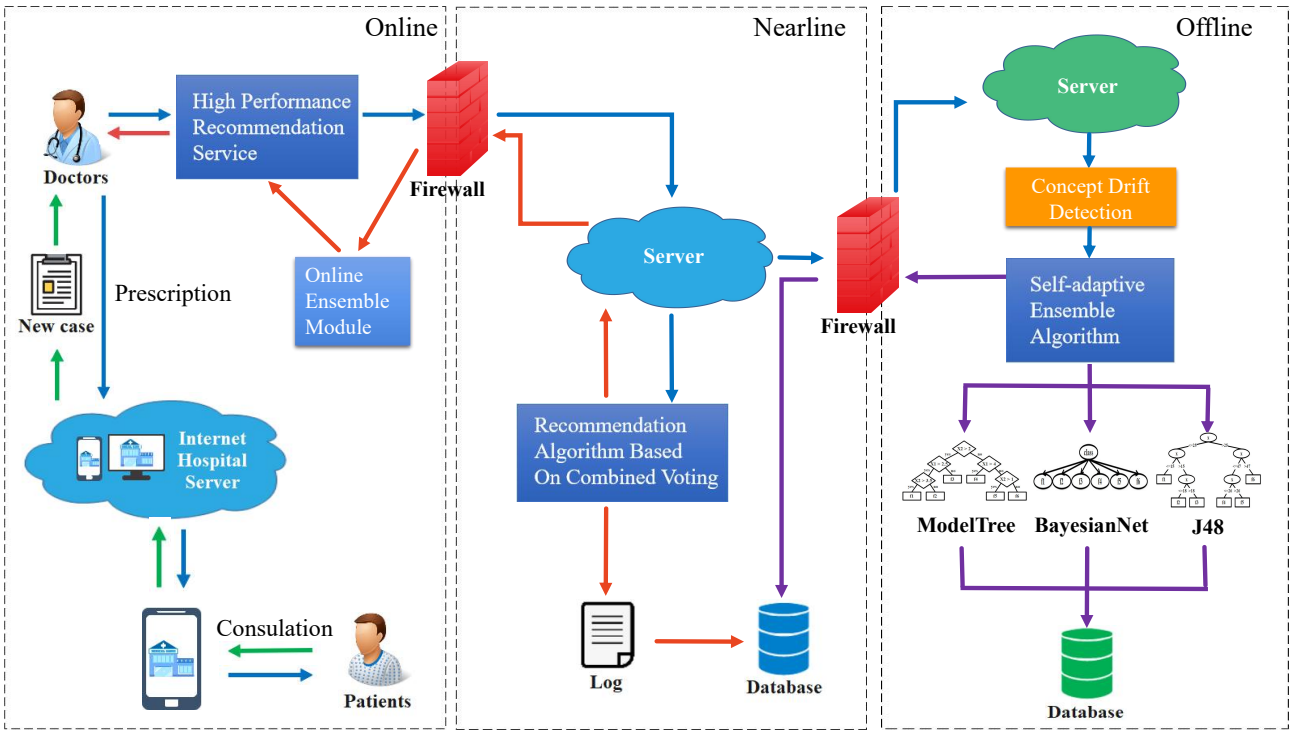


Figure 2: System framework of our online drug recommendation system

In addition to that, the interpretability of concept drift is a major issue ignored by current recommendation systems. Investigating and explaining the causes of drift is very important to provide reliability of drug recommendation systems and improve the ability to predict future drift. Research on the interpretability of conceptual drift is divided into sequential analysis, statistical and window based methods [26]. Among them, statistical methods analyze the changes in the mean and standard deviation of the results to be predicted, and typical methods include EDDM [27], EWMA [28], and RDDM [29]. Recently, a method SDDM [26] by detecting changes in the data distribution has been proposed to quantify drift, which is very intuitive for identifying the reason of drift, having great application value.

Therefore, we construct a recommendation model with online-nearline-offline structure based on the interpretable concept drift detection and multi-model fusion adaptive classification algorithms. Based on drug medical data from an Internet hospital that has implemented the online prescription pattern, the model is used to drug recommendation, which can catch concept drift in time and adaptively update the base classifiers, achieving a higher accuracy rate.

3 Methods

Figure 2 shows the framework of our online drug recommendation system. First of all, collaborative recommendation based on combined voting improves the accuracy of the system; secondly, with the passage of time, new user characteristic information may appear and the concept drift may occur. Therefore, an interpretable concept drift detection algorithm and an adaptive ensemble classification algorithm is proposed. In this section, we first introduce the framework of our proposed system and the workflow based on the data interaction between each layer, then we introduce the concept drift detection algorithm, collaborative recommendation strategy based on combined voting and the classifier ensemble strategy in detail.

3.1 System Framework

3.1.1 Online System

Directly facing users, this system contains the high performance and availability recommendation service. The online ensemble module will fuse the recommendation results calculated by Nearline and content-based recommendations. Online system needs to return results in a short period of time, so the simplest priority fusion algorithm is often used here.

3.1.2 Nearline System

This system is deployed on server. On the one hand, it will receive the prescription recommendation requests from doctors and invoke the combined voting algorithm to generate the results according to the latest pathological characteristics parameters of patients. On the other hand, the data flow is collected in fixed time step and stored in the data flow database. The time series data flow will be detected by the interpretable concept drift detection algorithm and fused with the original basic classifiers by Offline's self-adaptive ensemble algorithm, which will be described in detail in the following sections.

3.1.3 Offline System

The task of this system is to mine long-term patient drug treatment process data. Our original base classifiers include linear regression based on Model Tree, Bayesian Network and J48 based on decision tree, which are selected from ten classifiers by the experiments. In the testing stage, the accuracy of each classifier is about 99%. The ensemble training process is a bit more complicated because the classifiers need to be trained separately in the Nearline and Offline systems.

3.2 Collaborative Recommendation Based On Combined Voting

The base classifiers based on the adaptive ensemble strategy provide the local classification results of test instances. All the class labels $\{c_1, c_2, \ldots, c_T\}$ of test instances are loaded into the voting. All classifiers will classify the current test instance x_j, and the class of x_j is determined by the voting weight and classification probability of each base classifier. Assuming that the output function of each base classifier is $h_i(x_j)$, then the ensemble output is:

$$H(x_j) = c_{\underset{k}{argmax} \sum_{i=1}^{N_E} s_i h_i^k(x_j)} \tag{1}$$

where, c_k is the predicted class label of the current test instance, s_i is the voting weight of the base classifier, $\sum_{i=1}^{N_E} s_i = 1$. When the performance of each base classifier is not equal, the stronger classifiers can be given more voting weights to make classification results more reasonable. In this paper, the voting weight of each classifier is the same, that is, $s_i = 1/N_E$.

Suppose that e_1, e_2 and e_3 are three base classifiers, B_1 and B_2 are class labels. For the test instance x_j, the probability that x_j is classified as B_1 are respectively 80%, 70% and 40%, the probability that the test instance are classified as B_2 are 20%, 30% and 60%. Then the classification probability of the ensemble classifier for B_1 is $1/3 \times 80\% + 1/3 \times 70\% + 1/3 \times 40\% = 63\%$, the classification probability for B_2 is $1/3 \times 20\% + 1/3 \times 30\% + 1/3 \times 60\% = 37\%$. Finally, B_1 is the prediction class label of x_j.

3.3 Interpretable Concept Drift Detection

The development of classification decisions in recommendation systems is based on the posterior probability distribution $P(C|X)$, where X is the sample, C is the target. The posterior probability distributions may be similar or different for the incoming data streams X_t and $X_{t'}$ at different moments. The posterior probability distribution is indicated as "real concept drift" [23]. Once the concept drift occurs, the posterior probability distributions at different moments will exhibit large differences. In order to enhance the interpretability of concept drift, the change of posterior probability distribution

needs to be quantified. In this paper, we use Kullback-Leibler divergence to quantify this difference, and the quantified difference value is called the concept drift magnitude [30].

$$dm\left(X_{t'}\right) = KL\left(P_t \| P_{t'}\right) = \sum_{x_t, x_t \in X} P\left(C_t \mid X_t\right) \log\left(\frac{P\left(C_t \mid X_t\right)}{P\left(C_{t'} \mid X_{t'}\right)}\right) \tag{2}$$

$$P\left(C \mid X\right) = \frac{P\left(X \mid C\right) P\left(C\right)}{P\left(X\right)} \tag{3}$$

where, $dm\left(X_{t'}\right)$ denotes the conceptual drift magnitude at $t^{'}$, $P_{t'}$ and P_t are the posterior probability distributions for $t^{'}$ and t.

In practical medical scenarios, patient medical data are often high-dimensional, involving basic patient information (e.g., gender, age), pathological features (e.g., family medical history, adverse hobbies), and historical drug use, etc. When calculating the drift magnitude on the entire high-dimensional data stream, due to the monotonicity of the distance measures, the drift of a few features is difficult to be identified and then it's hard to detect whether the conceptual drift has really occurred [26, 30]. Therefore, in order to capture the conceptual drift in the medical data stream more accurately, we calculate the drift amplitude $dm\left(X_{t'}^{f_i}\right)$ for each feature f_i and use the maximum of them as the conceptual drift amplitude of the whole data stream, and determine whether the conceptual drift occurs in the whole data stream by comparing it with a predefined threshold ε. When $\max\left(dm\left(X_{t'}^{f_1}\right), dm\left(X_{t'}^{f_2}\right), \ldots, \left(X_{t'}^{f_F}\right)\right) \geq \varepsilon$, a higher drift amplitude indicates that concept drift occurs, otherwise no concept drift occurs.

3.4 Adaptive Ensemble Algorithm

Online medical data is a kind of streaming data that continuously increases. As time passes, an infinite number of base classifiers may need to be constructed. However, due to the limitation of time, memory and performance, it is not necessary to combine all base classifiers for prediction. Therefore, in the integration, one of the problems to be solved is how to choose the useful base classifiers to build the optimal integration. A base classifier associated with the current concept is useful for predicting instances. When no concept drift occurs, the number of base classifiers increases. However, when concept drift occurs, most of the old base classifiers are not representative of the latest concepts and only a few useful base classifiers can participate in the classification.

In this paper, an adaptive ensemble strategy is adopted. When data block arrives, a new base classifier is firstly constructed on the data block, and the accuracy of each basic classifier in the existing integration is calculated, so as to determine the accuracy weight of them. Every time a new classifier is added, the base classifier with contribution less than 0 will be deleted. The adaptive strategy adopts the above two steps to keep a better base classifier and make it better adapt to concept drift.

$$w_e = \frac{\phi_i - \phi_\theta}{\sum_{i=1}^{N_E}\left(\phi_i - \phi_\theta\right)}. \tag{4}$$

where, w_e is the accuracy weight of the classifier on the data block, ϕ_i is the classification accuracy of each base classifier in the new data block; N_E is the number of classifiers in current integration E; ϕ_θ is a custom threshold, which is used to determine whether a classifier should be discarded. The calculation equation is:

$$\phi_\theta = \begin{cases} \max\left(\min\left(E_{cor}\right) - \tau, \frac{1}{2}\right), & concept \quad drift \quad occurs \\ \max\left(\bar{\phi}, \frac{1}{2}\right), & no \quad concept \quad drift \quad occurs \end{cases} \tag{5}$$

where, E_{cor} is the classification accuracy set of the base classifier in the current integration E; $\bar{\phi}$ is the average classification accuracy of the base classifiers. When no concept drift is detected at this time, all base classifiers in the current integration should be useful, τ is a smaller value greater than 0, ensure $w_e > 0$, therefore, all selected base classifiers can be used for classification; otherwise, when concept drift is detected at this time, only some useful base classifiers can participate in the prediction.

A classifier can not only improve the classification accuracy of the current observed data, but also improve the classification accuracy of the entire infinite data stream. After the new classifier is created, the original base classifiers in the integration is evaluated. The new data block is used to calculate the accuracy rate of the existing ensemble classifier:

$$p\left(E_{new}, B_k\right) = \frac{\sum_{i=1}^{d} Corr_{new}\left(x_i\right)}{d}. \tag{6}$$

where, x_i is the i-th real example of data block B_k;

$$Corr_{new}\left(x_i\right) = \begin{cases} 1, \text{ correct} \\ 0, \text{ wrong} \end{cases} \tag{7}$$

When deleting a classifier e, calculate classification accuracy:

$$p\left(E_{\text{new-e}}, B_k\right) = \frac{\sum_{i=1}^{d} \text{Corr}_{\text{new-e}}\left(x_i\right)}{d}. \tag{8}$$

The contribution of the classifier to the new data block B_k in the integration:

$$contribution\left(B_k\right) = p\left(E_{\text{new}}, B_k\right) - p\left(E_{\text{new-e}}, B_k\right) \tag{9}$$

The contribution can be positive or negative. Negative value means that the base classifier reduces the overall classification accuracy; otherwise, it improves the overall classification accuracy. Once a new classifier is added, discard the base classifier with contribution less than 0. In Algorithm 1, new training data block $B_k = \{x(0), x(2), \ldots, x(n-1)\}$, E is current integration classifier collection, N_E is number of ensemble classifiers, N_{max} is maximum number of classifiers, E' is optimized ensemble classifier, e is a new classifier, E_{new} is new integration classifier collection, e_i is each base classifier in E_{new}, ϕ_i is the correct classification rate of integrated classifier, E'_{new} is the integration after accuracy weights are updated and e'_i is each base classifier in it, $w_{e'_i}$ is the accuracy weights of e'_i on B_k, $N_{E'}$ is the current number of classifiers, e_{lw} is a classifier with a small weight according to the weight sort.

4 Experiments

4.1 Data Analysis And Processing

The experiments are based on patient drug medical data from an Internet hospital that has implemented the online prescription pattern. A total of 56 variables describe the 5690 instances in the data set, which contains 13 attributes for basic patients' information, 38 attributes for describing patient pathological characteristics, and 3 attributes for patient medication. The purpose of the experiments is to use the classifier to predict which type of drugs are most suitable for the patient, so as to achieve a recommendation.

We use 60% of the data as historical static data block of patients, and get an initial ensemble recommendation model. Then, we use the three-tier recommendation framework proposed in this paper to test the robustness of our online self-adaptive ensemble learning (OSEL) algorithm. The rest of the data is divided into the new data block 1 and 2.

4.2 Experimental Steps

In the experiments, we achieve OSEL and nine other algorithms, and compare their medication grouping prediction results to explore the classification effect of OSEL proposed in this paper. Table 1 shows these nine algorithms.

We achieve OSEL and the above nine algorithms on WEKA 3.8.5 platform of Windows 10 system, the running environment is jdk9. In the experiments, the batch size is 40, the seed is 10, and the other parameters of the classifiers keep default. The ensemble method of OSEL is $Vote$ and the voting strategy is $Average\,of\,Probabilities$.

In this paper, we carry out two experiments. In the first experiment, we compare and evaluate the classification effects of the ten algorithms on Offline system, and in the second experiment we evaluate the classification effects on Online system. The specific experimental steps are as follows:

Algorithm 1 adaptive ensemble learning algorithm

Input: B_k, E, N_E, N_{max}
Output: E'
build e on B_k
add e to E to form E_{new}
test the performance of E_{new} on B_k
for each $e_i \in E_{new}$ **do**
    calculate ϕ_i according to equation(6)(7)(8)
    if contribution(B_k) < 0 **then**
        drop e_i
    else
        keep e_i
    end if
end for
get E'_{new}
for $e'_i \in E'_{new}$ **do**
    calculate $w_{e'_i}$ on B_k according to equation(4)(5)
    if $w_{e'_i}<0$ **then** delete e'_i
    else
        retain e'_i
    end if
end for
while $N_{E'} > N_{max}$ **do**
    delete e_{lw}
end while
return E'

Table 1: Nine algorithms

algorithm	type
Classification Via Regression (CVR)[31]	meta
MultiClass Classifier (MCC)[32]	meta
MultiClass Classifier Updateable (MCCU)[33]	meta
Bayes Network (BN)[34]	bayes
Naive Bayes(NB)[35]	bayes
Naive Bayes Updateable (NBU)[35]	bayes
J48[36]	trees
Logic Model Trees (LMT)[37]	trees
Random Forest (RF)[38]	trees

- The first experiment: firstly, nine classifiers are trained on historical static data block, and the classifiers with best performance are selected as the base classifiers of OSEL. It is worth noticing that we only consider the new data block 1 as a normal test set, i.e., no concept drift is considered, so as to demonstrate the performance of OSEL. Then, the trained nine classifiers and OSEL classifier are tested on data block 1 to obtain the classification results on Offline system.
- The second experiment: in this experiment, concept drift is simulated by exchanging labels on new data block 1 and 2. First, new OSEL base classifiers are built on data block 1 and added to the OSEL base classifier set. The drift magnitude of data block 1 relative to the historical static data block will be calculated to confirm whether concept drift occurs. According to Algorithm1, the OSEL base classifiers are added or removed to form the new integrated classifier OSEL'. Then, we evaluate the performance of the nine classifiers trained in last experiment as well as OSEL' on data block 2.

4.3 Evaluation Metrics

In this paper, in order to evaluate the classification prediction effect of these ten classifiers, we use Accuracy, Kappa, Precision, Recall, F-Measure and AUC as evaluation metrics. Here are the brief introductions to them:

- Accuracy: The proportion of correctly classified samples to the total number of samples for a given data.
- Kappa: It is a method of evaluating consistency in statistics and can be used to evaluate the accuracy of multi-classification models. The value range is [-1, 1].
- Precision: In all the samples that are predicted to be positive, the proportion of the samples that are actual positive.
- Recall: In all the samples that are actually positive, the proportion of the samples that are predicted to be positive.
- F-Measure: The harmonic mean of Precision and Recall.
- AUC: The area under ROC curve within the value range of [0.1, 1].

5 Result Analysis

In order to evaluate the classification prediction effect of OSEL, we achieve nine other algorithms on patient medication data set. We finish two experiments on Offline and Online system respectively and discuss the classification results. In the experiments, we set the maximum number of base classifiers is 3, $\varepsilon = 0.003$, $\tau = 0.000001$.

5.1 Experiment Result On Offline System

We trained nine classifiers on training set 1. Table 2 shows the training results. As can be seen from Table 2, CVR1 has 6 metrics, BN1 has 6 metrics, J481 has 5 metrics and LMT1 has 1 metric in the top three. Therefore, CVR1, BN1 and J481 perform better than other classifiers on training set 1.

According to the above analysis, we choose CVR1, BN1 and J481 as the base classifiers of OSEL classifier, that is, the ensemble classifier set $E = \{$CVR1, BN1, J481$\}$. We test the nine classifiers that trained on training set 1 and OSEL classifier on testing set 1 and get their accuracy. Table 3 shows the results. It can be clearly seen that OSEL classifier gets the highest classification accuracy on new data block 1 when there is no concept drift occurs (the data block 1 is only used to evaluate the performance of OSEL classifier with other benchmark classifiers). Besides, OSEL also has the highest

Table 2: Evaluation metric scores for nine classifiers on historical static data block

Classifier	Accuracy(%)	Kappa	Precision	Recall	F-Measure	AUC
CVR1	**99.7364**	**0.9962**	**0.997**	**0.997**	**0.997**	**0.999**
MCC1	69.2736	0.507	0.667	0.693	0.633	0.881
MCCU1	68.0726	0.4641	-	0.681	-	0.935
BN1	**99.7364**	**0.9962**	**0.997**	**0.997**	**0.997**	**0.999**
NB1	54.1886	0.3638	0.602	0.542	0.561	0.778
NBU1	54.1886	0.3638	0.602	0.542	0.561	0.778
J481	**99.6192**	**0.9944**	**0.996**	**0.996**	**0.996**	0.998
LMT1	99.4728	0.9923	0.995	0.995	0.995	**0.999**
RF1	83.8606	0.749	0.862	0.839	0.814	0.984

Table 3: Evaluation metric scores for each classifier on new data block 1

classifier	Accuracy(%)	Kappa	Precision	Recall	F-Measure	AUC
CVR1	99.8243	0.9974	0.998	0.998	0.998	0.999
MCC1	69.9473	0.5169	0.674	0.699	0.639	0.887
MCCU1	68.8094	0.4818	-	0.688	-	0.832
BN1	99.6485	0.9949	0.997	0.996	0.996	0.999
NB1	54.3058	0.3665	0.600	0.543	0.559	0.778
NBU1	54.3058	0.3665	0.600	0.543	0.559	0.778
J481	99.3849	0.9910	0.994	0.994	0.994	0.996
LMT1	99.7364	0.9962	0.997	0.997	0.997	0.999
RF1	83.5677	0.7443	0.863	0.836	0.816	0.985
OSEL	**99.9121**	**0.9987**	**0.999**	**0.999**	**0.999**	**1.000**

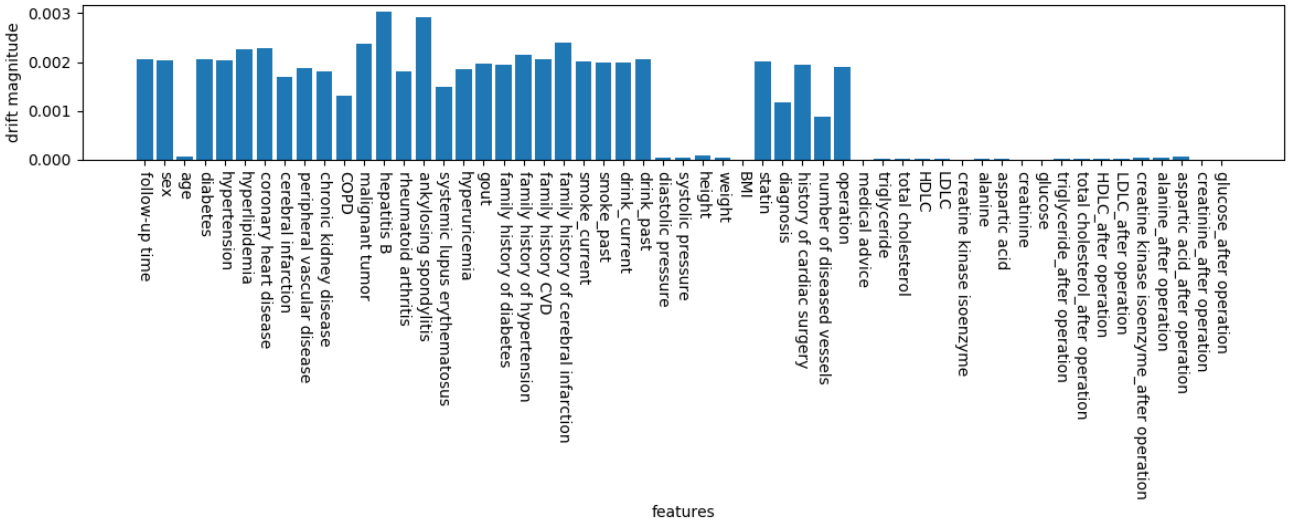


Figure 3: Drift magnitude of each feature on new data block1

scores in the other five metrics, which is better than any of its base classifiers: CVR1, BN1 and J481. Therefore, the performance of OSEL classifier is the best on testing set 1.

When the drift magnitude is greater than 0.2, it shows that concept drift is detected at this time. The classification accuracy set of the base classifiers is E_{cor} = {99.8243, 99.6485, 99.3849}, $\bar{\phi}$ = 99.6192%. Since 99.6192 - 99.3849 > 0.2, it can be seen that concept drift is detected at this time, and only part of the base classifiers can participate in the prediction of the instance to be predicted. We calculate ϕ_θ = 99.6888% and $w_{J481} < 0$ according to equation (3) and (4). Due to Algorithm 1, we delete the base classifier J481.

5.2 Experiment Result On Online System

When data block 1 is considered as a new data stream, the concept drift magnitude calculated for each feature on it is shown in Figure 3. Among them, the largest concept drift amplitude is 0.003 on feature "hepatitis B". This is taken as the concept drift magnitude of data block 1, i.e., $dm(data\ block1) = 0.003$. Since $dm(data\ block1) \geq \varepsilon$, it can be determined that concept drift has occurred at this point, and only some of the base classifiers can participate in the next prediction.

Table 4: Accuracy of the classifiers on data block 1

Classifier	E_{new}	$E_{new-CVR1}$	$E_{new-BN1}$	$E_{new-J481}$	$E_{new-CVR2}$	$E_{new-BN2}$	$E_{new-J482}$
Accuracy(%)	58.5237	99.8243	99.8243	99.8243	56.9420	56.9420	56.9420

Table 5: Accuracy of the classifiers on data block 1

Classifier	$CVR2$	$BN2$	$J482$
Accuracy(%)	98.4183	99.7364	98.7698
w_e	-0.3654	0.5000	-0.1346

According to Algorithm 1, we construct new base classifiers CVR2, BN2, and J482 on the new data block 1 and add them to the existing integrated classifier to get $E_{new} = CVR1, BN1, J481, CVR2, BN2, J482$. The classification accuracy are shown in Table 4. The contribution of CVR1, J481, and BN1 to the integration is -41.3006, and the contribution of BN1 and BN2 is 1.5817. Therefore, CVR1, BN1, and J481 are removed to obtain the new integrated classifier $E'_{\text{new}} = \{CVR2, BN2, J482\}$.

The accuracy of the base classifiers of E'_{new} is show in Table 5, the average accuracy $\bar{\varphi} = 98.9748\%$. Due to the concept drift is detected now, $\varphi_\theta = 98.9748\%$. According to equation (3), the precision weight of the base classifiers are also shown in Table 5. We remove the base classfiers with precision weights less than 0, and get integrated classifier $OSEL' = \{BN2\}$.

In order to evaluate the classification effect of OSEL', we test the nine classifiers trained in the first experiment and OSEL' classifier on data block 2. Table 6 shows the scores of each metrics. As can be seen from Table 6, OSEL' gets the highest scores in all evaluation metrics, indicating that it performs best on new data block 2.

According to the results of the above experiments, ensemble classification algorithms based on combined voting are better than independent classification algorithms; besides, OSEL algorithm performs best in processing the data flow with concept drift, and ability to adaptively update the base classifiers makes it more advantageous than other fixed classifiers.

6 Conclusion

To relieve pressure of doctors and reduce the risk of medical accidents under the online prescription pattern of Internet healthcare, this paper proposes a multi-model fusion online drug recommendation system based on the characteristics of medical data stream. The system adopts an online-nearline-offline architecture based on interpretable concept drift detection and adaptive ensemble algorithms, which can effectively identify concept drift and the reasons, and improve the accuracy and reliability of recommendation results. In the experiments, we apply it to the PCI treatment process. The results show that the proposed online drug recommendation system is highly accurate, the accuracy is close to 100%. Our system performs nearly as good as doctors. We believe that our online drug recommendation system can effectively help the doctors in fighting against COVID-19.

In the future work, we will focus on increasing the practicability of the system so as to meet more actual medical recommendation scenes. For example, drug recommendations for multiple courses of treatment.

Table 6: Evaluation metric scores for each classifier on data block 2

classifier	Accuracy(%)	Kappa	Precision	Recall	F-Measure	AUC
CVR1	56.8541	0.4135	0.9920	0.5690	0.6930	0.6410
MCC1	30.8436	0.0402	0.6970	0.3080	0.4190	0.5340
MCCU1	26.3620	0.0310	-	0.2640	-	0.5010
BN1	56.7663	0.4123	0.9920	0.5680	0.6930	0.8930
NB1	31.0193	0.1086	0.6730	0.3100	0.3790	0.6050
NBU1	56.7663	0.4117	0.9910	0.5680	0.6920	0.7820
J481	56.8541	0.4129	0.9910	0.5690	0.6930	0.7740
LMT1	42.2671	0.2365	-	0.4230	-	0.6040
RF1	85.3251	0.7741	0.875	0.853	0.832	0.985
OSEL'	**99.0334**	**0.9768**	**0.9880**	**0.9900**	**0.9890**	**0.9950**

Funding

This work was supported in part by the Fundamental Research Funds for the Central Universities [grant number 2021QY010].

Author contributions

The authors contributed equally to this work.

Conflict of interest

The authors declare no conflict of interest.

References

[1] LIU Bo, GUO Youyan, LIN Yang, WU Xinghai, and WEI Yongxiang. A brief analysis on the development of public hospitals assisted by internet medical services under the covid-19 epidemic. *Chinese Hospitals*, 24(09):62–64, 2020.

[2] Gheorghe Zaman, Anamaria-Cătălina RADU, Ivona RĂPAN, and Florian Berghea. New wave of disruptive technologies in the healthcare system. *Economic Computation & Economic Cybernetics Studies & Research*, 55(1), 2021.

[3] Yu Wang, Peng-Fei Li, Yu Tian, Jing-Jing Ren, and Jing-Song Li. A shared decision-making system for diabetes medication choice utilizing electronic health record data. *IEEE Journal of Biomedical and Health Informatics*, 21(5):1280–1287, 2017.

[4] Jinsung Yoon, Camelia Davtyan, and Mihaela van der Schaar. Discovery and clinical decision support for personalized healthcare. *IEEE Journal of Biomedical and Health Informatics*, 21(4):1133–1145, 2017.

[5] Lin Liu, Lin Tang, Wen Dong, Shaowen Yao, and Wei Zhou. An overview of topic modeling and its current applications in bioinformatics. *SpringerPlus*, 5(1):1–22, 2016.

[6] Corville O Allen, Timothy A Bishop, Michael T Payne, Sue S Schmidt, and Leah R Smutzer. Identifying drug-to-drug interactions in medical content and applying interactions to treatment recommendations, November 17 2020. US Patent 10,839,961.

[7] N. Komal Kumar and D. Vigneswari. A drug recommendation system for multi-disease in health care using machine learning. In Gurdeep Singh Hura, Ashutosh Kumar Singh, and Lau Siong Hoe, editors, *Advances in Communication and Computational Technology*, pages 1–12, Singapore, 2021. Springer Singapore.

[8] Chun Chen, Lu Zhang, Xiaopeng Fan, Yang Wang, Chengzhong Xu, and Renkai Liu. A epilepsy drug recommendation system by implicit feedback and crossing recommendation. In *2018 IEEE SmartWorld, Ubiquitous Intelligence Computing, Advanced Trusted Computing, Scalable Computing Communications, Cloud Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*, pages 1134–1139, 2018.

[9] Ferran Torrent-Fontbona and Beatriz López. Personalized adaptive cbr bolus recommender system for type 1 diabetes. *IEEE Journal of Biomedical and Health Informatics*, 23(1):387–394, 2019.

[10] Wen San Yee, Hu Ng, Timothy Tzen Vun Yap, Vik Tor Goh, Keng Hong Ng, and Dong Theng Cher. An evaluation study on the predictive models of breast cancer risk factor classification. *Journal of Logistics, Informatics and Service Science*, 2022.

[11] BS Kim, Tag-Gon Kim, and SH Choi. Codevs: An extension of devs for integration of simulation and machine learning. *INTERNATIONAL JOURNAL OF SIMULATION MODELLING*, 20(4):661–671, 2021.

[12] Peng Tang, Qiaokang Liang, Xintong Yan, Shao Xiang, and Dan Zhang. Gp-cnn-dtel: Global-part cnn model with data-transformed ensemble learning for skin lesion classification. *IEEE Journal of Biomedical and Health Informatics*, 24(10):2870–2882, 2020.

[13] Z. Cui, X. Xu, F. Xue, X. Cai, Y. Cao, W. Zhang, and J. Chen. Personalized recommendation system based on collaborative filtering for iot scenarios. *IEEE Transactions on Services Computing*, 13(4):685–695, 2020.

[14] Jiyoung Yoon and Soonhee Joung. A big data based cosmetic recommendation algorithm. *Journal of System and Management Sciences*, 10(2):40–52, 2020.

[15] Rung-Ching Chen, Yun-Hou Huang, Cho-Tsan Bau, and Shyi-Ming Chen. A recommendation system based on domain ontology and swrl for anti-diabetic drugs selection. *Expert Systems with Applications*, 39(4):3995–4006, 2012.

[16] Qian Zhang, Guangquan Zhang, Jie Lu, and Dianshuang Wu. A framework of hybrid recommender system for personalized clinical prescription. In *2015 10th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 189–195, 2015.

[17] Yinghui Wang. A novel chinese traditional medicine prescription recommendation system based on knowledge graph. *Journal of Physics: Conference Series*, 1487:012019, mar 2020.

[18] Yao Qin and Zherui Ma. A traditional chinese medicine prescription recommendation method based on mutual information clustering. *Journal of Physics: Conference Series*, 1544:012065, may 2020.

[19] Fan Gong, Meng Wang, Haofen Wang, Sen Wang, and Mengyue Liu. Smr: medical knowledge graph embedding for safe medicine recommendation. *Big Data Research*, 23:100174, 2021.

[20] Tan Ke. *Research on the Concept Drift Oriented Recommender System*. PhD thesis, University of Electronic Science and Technology of China, 2018.

[21] Yi Ding and Xue Li. Time weight collaborative filtering. In *Proceedings of the 14th ACM International Conference on Information and Knowledge Management*, CIKM '05, page 485–492, New York, NY, USA, 2005. Association for Computing Machinery.

[22] Lei Wang, Yunqiu Zhang, and Xiaohu Zhu. Concept drift-aware temporal cloud service apis recommendation for building composite cloud systems. *Journal of Systems and Software*, 174:110902, 2021.

[23] Joo Gama, Indr Liobait, Albert Bifet, Mykola Pechenizkiy, and Abdelhamid Bouchachia. A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 2014.

[24] Antônio David Viniski, Jean Paul Barddal, Alceu de Souza Britto Jr, Fabrício Enembreck, and Humberto Vinicius Aparecido de Campos. A case study of batch and incremental recommender systems in supermarket data under concept drifts and cold start. *Expert Systems with Applications*, 176:114890, 2021.

[25] Charinya Wangwatcharakul and Sartra Wongthanavasu. A novel temporal recommender system based on multiple transitions in user preference drift and topic review evolution. *Expert Systems with Applications*, 185:115626, 2021.

[26] Simona Micevska, Ahmed Awad, and Sherif Sakr. Sddm: an interpretable statistical concept drift detection method for data streams. *Journal of Intelligent Information Systems*, 56(3):459–484, 2021.

[27] Manuel Baena-Garcıa, José del Campo-Ávila, Raúl Fidalgo, Albert Bifet, R Gavalda, and Rafael Morales-Bueno. Early drift detection method. In *Fourth international workshop on knowledge discovery from data streams*, volume 6, pages 77–86, 2006.

[28] Gordon J Ross, Niall M Adams, Dimitris K Tasoulis, and David J Hand. Exponentially weighted moving average charts for detecting concept drift. *Pattern recognition letters*, 33(2):191–198, 2012.

[29] Roberto SM Barros, Danilo RL Cabral, Paulo M Gonçalves Jr, and Silas GTC Santos. Rddm: Reactive drift detection method. *Expert Systems with Applications*, 90:344–355, 2017.

[30] Geoffrey I Webb, Loong Kuan Lee, François Petitjean, and Bart Goethals. Understanding concept drift. *arXiv preprint arXiv:1704.00362*, 2017.

[31] Eibe Frank, Yong Wang, Stuart Inglis, Geoffrey Holmes, and Ian H Witten. Using model trees for classification. *Machine learning*, 32(1):63–76, 1998.

[32] Mehak Naib and Amit Chhabra. Predicting primary tumors using multiclass classifier approach of data mining. *International Journal of Computer Applications*, 96(8), 2014.

[33] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. The weka data mining software: An update. *SIGKDD Explor. Newsl.*, 11(1):10–18, November 2009.

[34] Nir Friedman, Dan Geiger, and Moises Goldszmidt. Bayesian network classifiers. *Machine learning*, 29(2):131–163, 1997.

[35] George H. John and Pat Langley. Estimating continuous distributions in bayesian classifiers, 2013.

[36] Leroy A Gondy, C. Rindflesch B Thomas, and Naïve Bayes. Programs for machine learning. *Advances in Neural Information Processing Systems*, 79(2):937–944, 1993.

[37] Niels Landwehr, Mark Hall, and Eibe Frank. Logistic model trees. *Machine learning*, 59(1-2):161–205, 2005.

[38] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

Journal's webpage: http://univagora.ro/jour/index.php/ijccc/

This journal is a member of, and subscribes to the principles of,
the Committee on Publication Ethics (COPE).
https://publicationethics.org/members/international-journal-computers-communications-and-control

Cite this paper as:

Peng, Y.; Qiu, Q; Zhang, D.; Yang, T.; Zhang H.(2023). Ensemble Learning for Interpretable Concept Drift and Its Application to Drug Recommendation, *International Journal of Computers Communications & Control*, 18(1), 5011, 2023.

https://doi.org/10.15837/ijccc.2023.1.5011