

Feature Sorting Algorithm Based on XGBoost and MIC Combination Model

Gao Xiang

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: 971104101@qq.com

Hu Zhiyi

Engineering Design Institute
Army Research Laboratory
Beijing, 100042, China
E-mail: 18992899862@163.com

Yu Jun

School of Computer Science and Engineering
Xi'an Technological University
Xi'an, 710021, Shaanxi, China
E-mail: yujun@xatu.edu.cn

Hu Yuzhe

Jinan University-University of Birmingham Joint
Institute
Jinan University
Guangzhou, 511400, Guangdong, China
mail:18137910896@163.com

Abstract—Feature ranking can not only help the data analysis system improve efficiency, but also reduce the interference of redundant features and irrelevant features to the results. At present, feature ranking of massive data is an important and difficult problem. In order to solve the above problems, this paper proposes a feature importance ranking algorithm based on XGBoost and MIC model by analyzing the existing algorithm models. Firstly, XGBoost model and MIC model are established respectively; Then, the results of the above two models are weighted and combined by the error reciprocal method. XGBoost model has the advantages of high efficiency, flexibility and portability, while MIC model has universality and easy parameter adjustment. The resulting XGBoost MIC combination model has both advantages; Finally, the first mock exam is used as a sample set of data for anticancer drug candidates. After preprocessing the data set, the XGBoost-MIC combination model is used to analyze the case. At the same time, the calculation results of a single model are calculated, and the model is optimized by adjusting the parameters of the model. The results show that the error of the first mock exam is obviously lower than that of the single calculation model, and the accuracy of the XGBoost-MIC is 0.75, which is 0.02 higher than that of the single model.

Keywords-Feature Sorting; MIC Arithmetic; Xgboost Arithmetic; Combination Model

I. INTRODUCTION

With the rapid development of the Internet, the amount of data generated by human activities is increasing exponentially. Big data is generally considered as PB-level data, including structured, semi-structured and unstructured data, whose scale and complexity greatly exceed the storage and computing capacity of existing hardware. However, due to the variety of irregular or incomplete data, such data brings great difficulties to data processing. Moreover, the traditional GBDT has defects in design, can not be processed in parallel, has high computational complexity, and is not suitable for high-dimensional sparse features.

In order to tackle a series of difficulties and challenges brought by massive data to server storage and operation, researchers have been trying to find out an efficient data analysis system that can extract valuable information from massive data. In recent years, many Internet companies have launched various big data processing systems, such as Google's MapReduce system, Microsoft's Cosmos system developed for parallel Blockchain, Haystack system proposed by Facebook to solve

the massive small files, etc. Under the background of massive data, it is more important to grasp the key information of data quickly. The big data application, which is named "ScholarSpace", developed by the Network and Mobile Data Management Laboratory (WAMDM) of Renmin University of China fully reflects the process of big data processing, and combines modern big data technology with traditional economics, law, literature and other disciplines to obtain key information. However, because the big data is not structured and its relationship is complex, it is not easy to establish association between the data object's structure and attributes. Moreover, because the error value of the single model for calculating the importance is high, the algorithm for extracting key information from big data always has certain limitations. In addition, other researchers also began to pay attention to the importance of feature selection system in massive data. Based on LDA model, Yang Guijun quantified the key information as subject feature vector as explanatory variable, and integrated multiple models with sample disturbance and attribute disturbance by using XGBoost algorithm to build prediction model; According to the film review data, Zhang Hongli extracted features from user reviews as auxiliary prediction indicators, and combined with other factor indicators as independent variables to build a prediction model; Based on the thought process of constructing tree in XGBoost algorithm, Li Zhanshan proposes a new wrapped feature selection algorithm *xgbsfs*, which avoids the limitation of single importance measurement by measuring three important indexes; Ye Qianyi uses the XGBoost model to mine the attribute data of shopping malls, and extracts the features through the cleaning and visual analysis of the original data, so as to predict the user behavior information and the sales of shopping malls. Although the above algorithms have used the existing algorithms to optimize the feature ranking algorithm, the feature ranking algorithm model still has the problems of low accuracy or too complex model.

To solve the above problems, there is an urgent need for an effective data analysis algorithm to extract the key information of massive data. The

first mock exam shows that compared with logistic regression and decision tree, the integration of different models with a certain strategy has higher accuracy and better stability. In recent years, XGBoost algorithm has been widely used in the field of data analysis and provides a method to calculate feature ranking, but the model is complex and it is difficult to adjust parameters. MIC algorithm is often used in feature selection of machine learning. The algorithm is simple and its effect is obviously better than other similar algorithms. Therefore, this paper combines XGBoost and MIC model to design an algorithm to extract the key information of data. The results of the first mock exam of the anti breast cancer candidate drug dataset show that the algorithm can effectively reduce the error value caused by a single model and improve the generalization ability of the model.

II. XGBOOST-MIC COMBINATION MODEL

XGBoost model calculates the importance score of each attribute by gradient lifting algorithm, which has the advantages of preventing overfitting, simple models and strong flexibility. But it is not suitable for processing high-dimensional feature data, and there are many parameters, so it is relatively difficult to adjust parameters. The MIC algorithm model just needs a large data samples, which can calculate the high-dimensional features well. MIC algorithm can cover all functional relations evenly when the sample size is large enough, rather than limited to some functional expressions. In order to combine the advantages of the two models, a new algorithm XGBoost-MIC combination model, is obtained by combining and weighting XGBoost and MIC models.

A. XGBoost model

XGBoost, the full name of eXtreme Gradient Boosting, is an optimized distributed gradient boosting library, which is efficient, flexible and portable. XGBoost is a tool for large-scale parallel Boosting tree. XGBoost is an optimization of boosting algorithm, which integrates weak classifier into a strong classifier. XGBoost algorithm generates a new tree through continuous iteration to fit the residual of the previous tree.

With the increase of iteration times, the accuracy continues to improve. It is the fastest and the best open source Boosting tree toolkit at present, which is more than 10 times faster than common toolkits. In terms of data science, XGBoost is a necessary weapon for major data science competitions, and a large number of contestants like to choose XGBoost for data mining competitions. In terms of industrial large-scale data, the distributed version of XGBoost is widely portable. And it is supported in various distributed environments such as Kubernetes, Hadoop, SGE, MPI, Dask, etc, which makes it a good solution to the problem of industrial large-scale data.

The definition function of XGBoost is shown in formula (1).

$$\hat{y}_i = \phi(x_i) = \sum_{k=1}^K f_k(X_i) \quad (1)$$

The goal of formula (1) is to learn such K tree models $f(x)$. In order to learn the model $f(x)$, the objective function shown in formula (2) is defined.

$$L(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

where $\Omega(f) = \gamma T + \frac{1}{2} \lambda \|\omega\|^2$

Among them, k is the total number of trees, f_k is the model of the k th tree, \hat{y}_i represents the predicted value of the model and y_i represents the category label of the i th sample, T represents the number of leaf nodes of each tree, and ω represents the set of scores of leaf nodes of each tree.

The tree model of XGBoost and the method of calculating feature importance are introduced below.

1) Tree model

Decision tree is a basic classification and regression method. It is a decision analysis method to judge the project risk and feasibility by constructing a decision tree based on the known

probability of various situations. As a supervised learning model, decision tree does not need any prior assumptions about data, so as to quickly find decision rules according to the characteristics of data. XGBoost adopted the gradient lifting algorithm to continuously reduce the loss caused by the last decision tree and generate new models to ensure the reliability of the final decision tree.

In the process of building a tree model with a given data set, greedy algorithm is used to select a feature segmentation point in each layer as a leaf node. If the gain value of the whole tree is the largest after segmentation, it means that the more times the feature is segmented, the better the effect of the whole tree is gained. That means the feature is more important. The weight of leaf nodes in the process of feature segmentation is recorded as $w(g_i, h_i)$, which g_i and h_i are expressed by formula (3) and formula (4) respectively.

$$g_i = \delta_{\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \quad (3)$$

$$h_i = \delta_y^{2\hat{y}^{(t-1)}} l(y_i, \hat{y}^{(t-1)}) \quad (4)$$

l indicates the gap between the target value y_i and the predicted value \hat{y} . Combining the weights of all leaf nodes, in order to maximize the gain value of the segmented tree, each feature can be used as the gain of the separation point, and its gain is the difference between the total weight after segmentation and the total weight of the leaf nodes before segmentation.

2) Feature importance model

Importance is a measure to evaluate the importance of each feature in the feature sets to which it belongs. XGBoost uses three attributes to measure the importance of features, including Freq, Gain and Cover. Here, Freq is the percentage of times that a specific feature occurs in the model tree, Gain is the relative contribution value obtained by calculating the contribution of features

in the model to each tree, and Cover is the coverage index of all functions related to a certain function.

The calculation formulas of the above three importance metrics are shown below (5).

$$\begin{aligned}
 Freq &= |X| \\
 Gain &= \frac{\sum Gain_x}{FScore} \\
 Cover &= \frac{\sum Cover_x}{FScore}
 \end{aligned}
 \tag{5}$$

X is a set of feature classification into leaf nodes, Gain is the gain value of each leaf node in X when it is segmented, and Cover is the number of samples falling on each node in X .

In order to get the best model, we construct $L = f(\theta)$, and get θ by “gradient descent”. With the support of the ascension tree, what we are looking for θ is a parameter for all trees(tree structure, scores of leaf nodes). In order to optimize the parameters and reduce the amount of calculation, we regard a tree as θ , and build a relationship between the loss function and a certain tree, and then take the derivative of loss for θ , and the

process of finding the tree is the solution process of XGBoost.

In this paper, XGBoost model needs to determine three parameters: general parameters, auxiliary parameters and task parameters. General parameters are used to control the function macroscopically; The auxiliary function controls the iterative model, select tree model or linear model; Task parameters control the performance of learning tasks and learning objectives.

Among the above three parameters, the auxiliary parameters have the greatest impact on the algorithm performance, and the maximum height of the tree will affect the final result. Therefore, first tune the maximum height of the tree. In the tuning process, first give other parameters an initial value, and set the important parameter lines to common typical values or default values. Compare the changes of test data by changing the height of the tree, So as to get better parameter selection. After determining the maximum height of the tree, the best combination of other parameters is obtained by traversal method.

Figure 1 shows the process of building decision tree and optimizing parameters. Figure 2 shows the XGBoost algorithm flow.

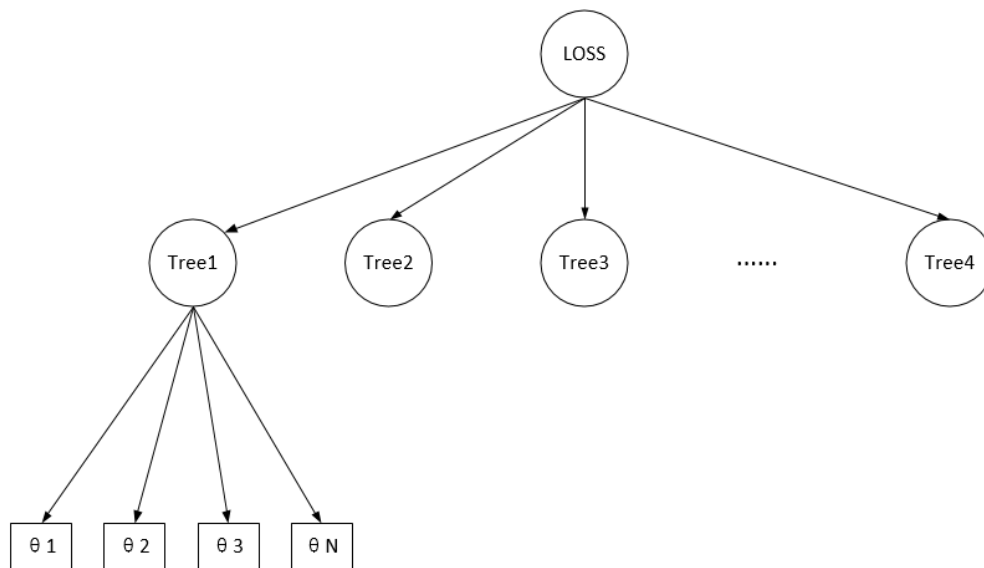


Figure 1. Building decision tree

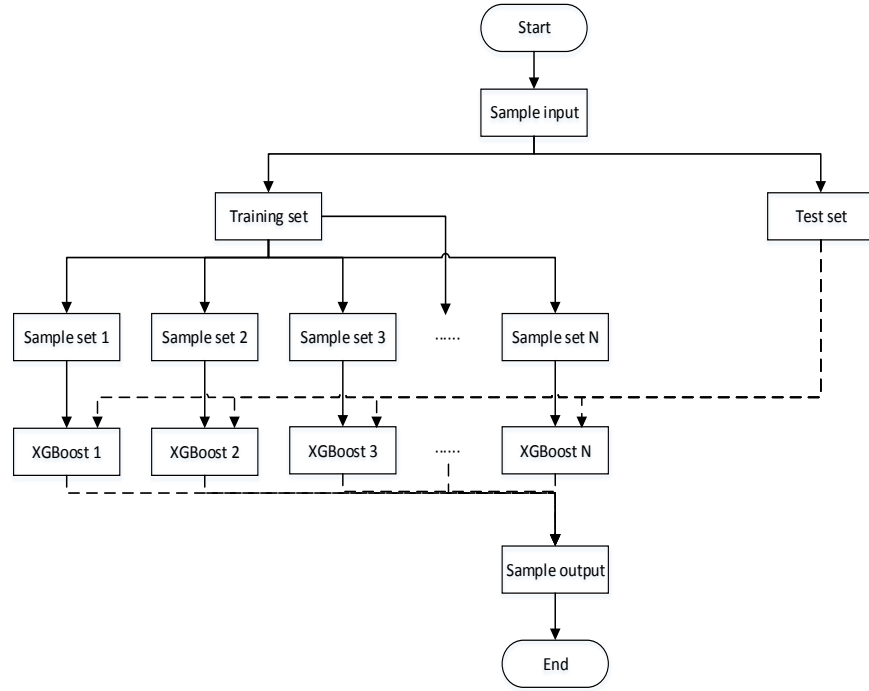


Figure 2. XGBoost algorithm flow

B. MIC correlation coefficient model

MIC (Maximum Mutual Information Coefficient) can be used to measure the degree of correlation between two variables X and Y, which is often used for feature selection of machine learning. Compared with Mutual Information(MI), MIC has a higher accuracy and is an excellent data correlation calculation method.

According to the nature of MIC, MIC is universal, fair and symmetrical. The so-called universality means that when the sample size contains most of the information of the total sample, it can capture all kinds of associations, but not limited to specific function types (such as linear function, exponential function or periodic function), or it can cover all functional relationships in a balanced way. The complex relationship between general variables can be modeled not only by a single function, but also by superposition functions. The so-called fairness means that when the sample size is large enough, similar coefficients can be given for the correlation of different types of single noise with similar degree. Symmetry means that the amount of

information obtained from different footholds is the same.

The basic principle of MIC will make use of the concept of mutual information. Mutual information is a measure of the degree of interdependence between random variables, which can be regarded as the information of another variable contained in one random variable. Mutual information can be defined by formula (6). P(x,y) is the joint probability between variables x and y.

$$I(x, y) = \int p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)} dx dy \tag{6}$$

$$I[x; y] \approx I[X; Y] = \sum_{xy} \log_2 \frac{p(X, Y)}{p(X)p(Y)}$$

The following formula (7) of MIC model is given as follows.

$$MIC(x; y) = \max_{a*b < B} \frac{I(x; y)}{\log_2 \min(a, b)} \tag{7}$$

$$MIC[x; y] = \max_{|X||Y| < B} \frac{I[X; Y]}{\log_2 (\min(|X|, |Y|))}$$

MIC model calculates the relationship between two attributes in turn, discretizes them in two-dimensional space, and uses scatter plot to represent them. The model divides the current two-dimensional space into a certain number of intervals in the X and Y directions, and then checks the situation that the current scatters fall into each grid, thus solving the problem that the joint probability in mutual information is hard to find.

The calculation of MIC value is generally divided into the following three steps: first normalize the data volume through different meshing methods, then calculate the maximum mutual information value, then normalize the maximum mutual information value, and finally select the maximum mutual information value under different scales as the mic value.

C. Combination model

In XGBoost algorithm, the steps of constructing decision tree are as follows:

- Traverse all feature nodes from the root node. For a certain feature, according to the sample value, sort and then determine the segmentation point with the best gain.
- Select the feature with the highest gain from all the selected segmentation points for segmentation.
- Divide to the maximum depth and build the next tree.
- Integrate all trees to complete the model construction.
- The feature importance measure is calculated according to the feature importance index.

According to the above algorithm ideas, the results of the two algorithms are weighted and combined by the error reciprocal method which are shown in formula (8).

$$f_t = \alpha_1 f_{1t} + \alpha_2 f_{2t}, \quad t = 1, 2, 3, \dots, n$$

$$\alpha_1 = \frac{\varepsilon_2}{\varepsilon_1 + \varepsilon_3}, \quad \alpha_2 = \frac{\varepsilon_1}{\varepsilon_1 + \varepsilon_2} \quad (8)$$

Respectively, ε_1 and ε_2 is the error between the calculated result and the true value by XGBoost and MIC algorithms. It can be seen from formula (8) that this method all error, thus reducing the error of the whole combination model and obtaining the predicted value closer to the real situation, thus improving the overall prediction accuracy. The error reciprocal method can not only give greater weight to the method with higher prediction accuracy, but also ensure that the error reciprocal method can give greater weight to the method with smaller absolute error value (i.e. the method with high prediction accuracy) at any time. If the number of positive and negative errors of each single prediction error is equal, the error reciprocal method can effectively reduce the combined prediction error and achieve better combination effect.

III. EXPERIMENTAL PROCESS AND RESULT ANALYSIS

This experiment runs under Linux Ubuntu16.04 system, and the framework version adopted is XGBoost1.5.0. In terms of hardware environment, the CPU used in the experiment is Intel(R) XEON W-2133 and the GPU is NVIDIA TITAN XP 12G.

The experimental data comes from the effect of 729 molecular descriptors of 1974 compounds provided by China Postgraduate Mathematical Modeling Competition in 2021 on inhibiting the activity of breast cancer cells, including the biological activity value of compounds on ER α and pIC50 obtained by converting IC50 value. In this data set, the biological activities of various compounds are constant and do not react with each other.

The flow of this experiment is shown in Figure 3. Firstly, preprocess the source data. Secondly, the processed data are calculated by XGBoost algorithm and MIC algorithm, and then weighted and combined by the reciprocal error method to obtain a combination model (namely XGBoost-MIC combination model). Finally, the algorithm results of XGBoost-MIC combination model are compared with those of single model (XGBoost or MIC model), and it is decided whether to adjust the parameters. If the results of the combination model are better than those of the single model,

the experimental results will be analyzed. Otherwise, adjust the parameters and retrain the model.

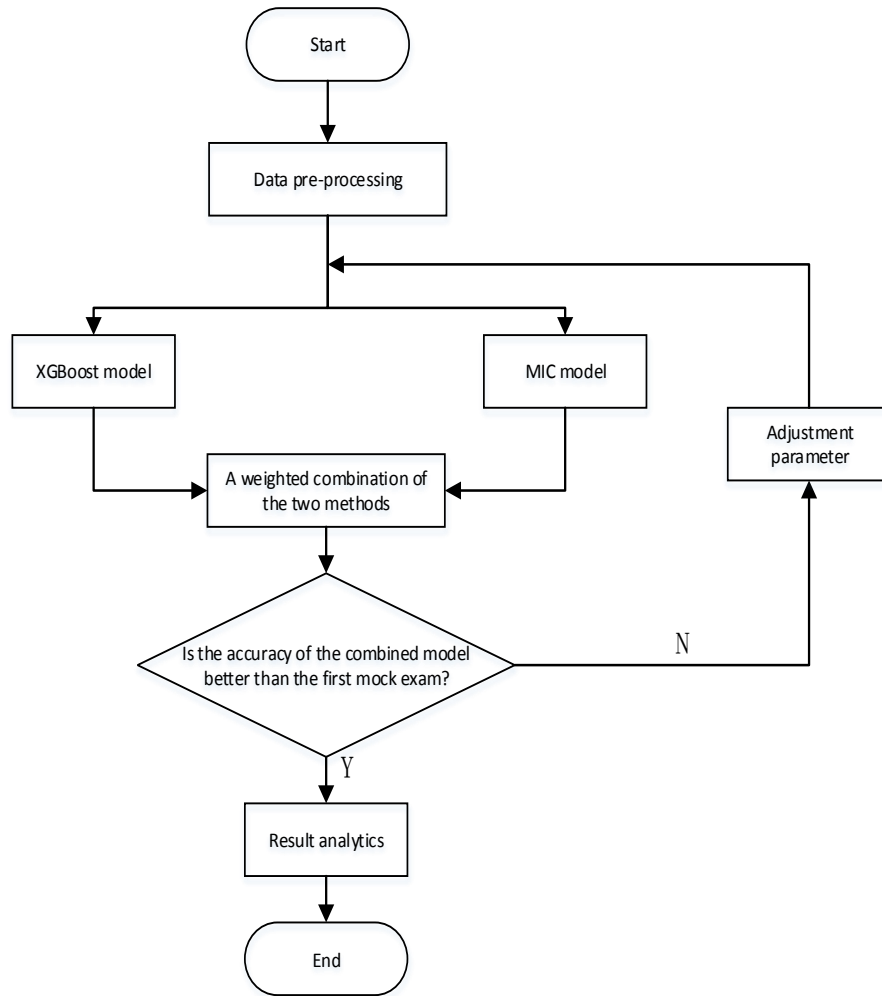


Figure 3. Experimental flow chart

A. Data preprocessing

In this data set, there are a certain number of unique attributes and empty attributes, so before using this data set, data processing should be carried out first.

For the unique attributes in the data set which can't describe the distribution law of the sample itself, we can delete these attributes directly.

In order to avoid dealing with high-dimensional data and reduce the difficulty of learning tasks, feature coding is needed. The feature coding method adopted in this paper is One-Hot Encoding. Only One-Hot Encoding uses N-bit status registers

to code N possible values, and each state is represented by an independent register, and only one of them is valid at any time.

For data with different magnitudes in the data set, the magnitude difference may greatly affect the calculation results. In order to reduce its effect on the results, the data must be standardized. In this paper, the data are normalized. Its conversion function is shown in formula (9).

$$x = \frac{x - \min}{\max - \min} \tag{9}$$

Finally, we divide the data set into two parts: the training set (80%) and the verification set (20%). The verification set is used to record the accuracy of the algorithm to find out the best model parameters.

B. Experimental results and analysis

As shown in Figure 3, the XGBoost and MIC models are realized respectively. The two models

are optimized by grid search method to get the optimal parameters, and then the combination model is calculated by the reciprocal error method. The score and ranking of feature importance calculated by this model are shown in Figure 4.

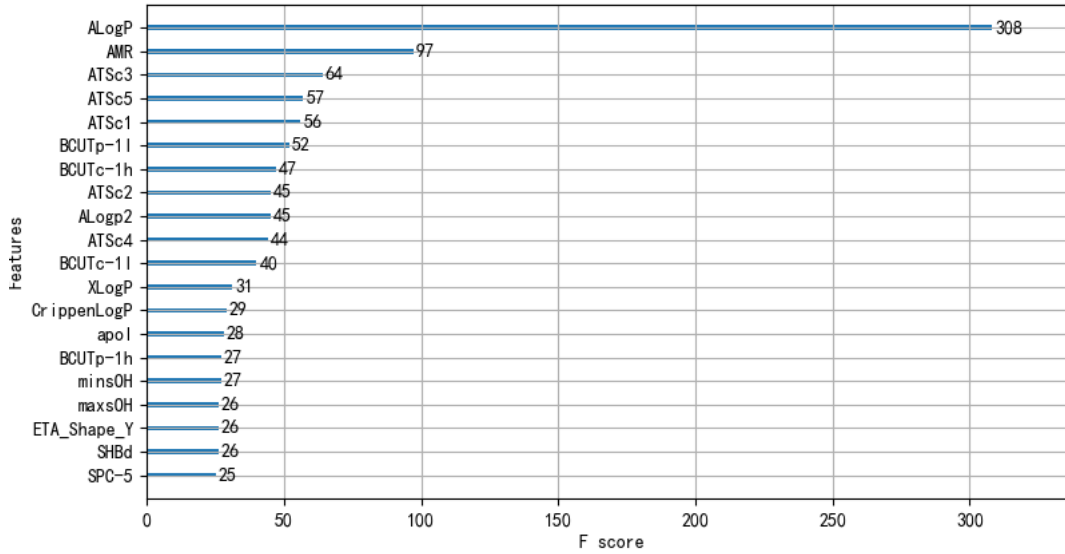


Figure 4. Order of feature importance

In this paper, R^2 (the decisive coefficient of the relationship between a random variable and multiple random variables) is selected to reflect the regression model, and is used as the main evaluation index of the prediction performance of each model. Root mean square error (RMSE) is selected as the auxiliary evaluation index. The calculation of R^2 and RMSE is shown in formula (10) and (11) respectively.

$$R^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum_i (y_i - f_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (10)$$

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2} \quad (11)$$

y_i is the predicted value of XGBoost-MIC combination model for the I-th data set, \hat{y}_i is the actual measured value in the I-th data set, and m is the number of samples.

The XGBoost-MIC combination model was tested with the testing-set data divided in the pre-processing, and the R^2 and RMSE values were compared with those of the single model. The results are shown in Table 1.

TABLE I. COMPARISON OF R^2 AND RMSE VALUES BETWEEN XGBOOST-MIC COMBINED MODEL AND SINGLE MODEL.

Method	R^2	RMSE
composite pattern	0.759	8.705
XGBoost model	0.723	9.453
MIC model	0.689	8.994

As can be seen from Table 1, the value of XGBoost-MIC combination model is higher(the

closer the R^2 value is to the 1, the closer the predicted value is to the real value), The prediction accuracy of the combination model is nearly 0.02 higher than that of the single model. By analyzing RMSE values, it can be concluded that the combination model has lower error than the single model. In general, the combination model has higher accuracy and less error, which has obvious advantages over the single model.

IV. CONCLUSION

This paper proposes a feature importance selection algorithm based on the combination of XGBoost and MIC. In the first mock exam, the algorithm uses XGBoost to rank the feature importance. The XGBoost-MIC model combines XGBoost and MIC models by weighted reciprocal method. The method automatically improves the model and gives more weight to the smaller error models, which can correct the larger errors caused by the single model. A combination of the first and the first mock exam models was used to compare the data set of the anti breast cancer candidate drugs. The results show that the first mock exam is better than the single model, and the XGBoost-MIC combination model can effectively improve the accuracy and efficiency of the feature ranking, and has a better generalization ability, which is suitable for dealing with large-scale data.

In the future work, we will continue to improve the performance of the algorithm proposed in this paper on a large data set, continue to explore more suitable parameter optimization methods and model fusion methods, continue to process data sets in other fields, and further obtain more robust models combined with scenarios.

REFERENCES

- [1] Zhang Yu, Zhang Yansong, Chen Hong, Susan Wang. OLAP foreign key join algorithm for MIC coprocessor [J]. Journal of Software, 2017, 28(03):490-501.
- [2] Yang Guijun, Xu Xue, Zhao Fuqiang. User score prediction model based on XGBoost algorithm and its application [J]. Data Analysis and Knowledge Discovery, 2019, 3(01):118-126.
- [3] Ye Qianyi, Rao Hong, Ji Mingshu. Commercial sales forecast based on Xgboost [J]. Journal of Nanchang University (Science Edition), 2017, 41(03):275-281.
- [4] Chen Zhenyu, Liu Jinbo, Jerry Lee, Ji Xiaohui, Li Dapeng, Huang Yunhao, Di Fangchun, Gao Xingyu, Xu Lizhong. Ultra-short term power load forecasting based on LSTM and XGBoost combined model [J]. Power Grid Technology, 2020, 44(02):614-620.
- [5] Zhang Chengchang, Zhang Huayu, Luo Jianchang, He Feng. Analysis method of massive electricity consumption data based on cloud computing and improved K-means algorithm [J]. Computer Applications, 2018, 38(01):159-164.
- [6] Liu Nian, Liu Yu. Research on visualization technology of massive relational data based on clustering analysis algorithm [J]. Electronic Design Engineering, 2018, 26(10):92-95.
- [7] Cheng Xueqi, Jin Xiaolong, Wang Yuanzhuo, Guo Jiafeng, Zhang Tieying, Li Guojie. Overview of Big Data System and Analysis Technology [J]. Journal of Software, 2014, 25(09):1889-1908.
- [8] Zhou Yanjun, Wang Shuangcheng, Wang Hui. Research on classifier based on Bayesian network [J]. Journal of Northeast Normal University (Natural Science Edition), 2003(02):21-27.
- [9] Xuanxuan Lin. Research on Enterprise Bankruptcy Prediction Method Based on XGBOOST Model [A]. Wuhan Zhicheng Times Cultural Development Co., Ltd. proceedings of 4th international conference on e-education, e-business and information management (EEIM 2021) [c]. Wuhan Zhicheng Times Cultural Development Co., Ltd.: Wuhan Zhicheng Times Cultural Development Co., Ltd., 2021:8.
- [10] Shenglong Li, Xiaoqing Zhang. Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm [J]. Neural Computing and Applications, 2020, 32(prepublish):
- [11] Feng Chen, Chen Zhidong. Application of xgboost and LSTM weighted combination model in sales forecast [J]. Computer system application, 2019, 28 (10): 226-232. Doi: 10.15888/j.cnki.csa.007091
- [12] Shenglong Li, Xiaoqing Zhang. Research on orthopedic auxiliary classification and prediction model based on XGBoost algorithm [J]. Neural Computing and Applications, 2020, 32(prepublish):
- [13] Wei Dong, Yimiao Huang, Barry Lehane, Guowei Ma. XGBoost algorithm-based prediction of concrete electrical resistivity for structural health monitoring [J]. Automation in Construction, 2020, 114(C):
- [14] Chixiang Wang, Junqi Guo. A data-driven framework for learners' cognitive load detection using ECG-PPG physiological feature fusion and XGBoost classification [J]. Procedia Computer Science, 2019, 147:
- [15] Gao Yifan, Yu Wenzhe, Chao Pingfu, et al. Score prediction and recommendation based on comment analysis [J]. Journal of East China Normal University: Natural Science Edition, 2015 (3): 80-90. (Gao Yifan, Yu Wenzhe, Chao Pingfu, et al. Analyzing reviews for rating prediction and item recommendation [J]. Journal of East China Normal University: Natural Science, 2015 (3): 80-90.)
- [16] Li V., Costantino H., Rowland J., Yue L., Gupta S.. ML3 LASSO (Least Absolute Shrinkage and Selection Operator) and XGBoost (eXtreme Gradient Boosting) Models for Predicting Depression-Related Work Impairment in US Working Adults [J]. Value in Health, 2021, 24(S1).
- [17] Li V., Costantino H., Rowland J., Yue L., Gupta S.. ML3 LASSO (Least Absolute Shrinkage and Selection Operator) and XGBoost (eXtreme Gradient Boosting) Models for Predicting Depression-Related Work Impairment in US Working Adults [J]. Value in Health, 2021, 24(S1).

- [18] Deng Xiaoyi, Jin Chun, Han Qingping, et al. Collaborative filtering recommendation model based on situational clustering and user rating [J]. *System engineering theory and practice*, 2013, 33 (11):2945-2953. (Deng Xiaoyi, Jin Chun, Han Jim C, et al. Improved Collaborative Filtering Model Based on Context Clustering and User Ranking [J]. *Systems Engineering –Theory & Practice*, 2013, 33(11): 2945-2953.)
- [19] Zhang Hongli, Liu Jiyang, Yang Sinan, et al. Research on scoring prediction model based on Internet user comments [J]. *Data analysis and knowledge discovery*, 2017, 1 (8): 48-58
- [20] McLachlan P, Munzner T, Koutsofios E, et al. LiveRAC: interactive visual exploration of system management time-series data[C] //Proceeding of the 26th International Conference on Human Factors in Computing Systems. New York: ACM Press, 2008: 1483-1492.
- [21] Agryzkov T, Oliver J L, Tortosa L, et al. Analyzing the commercial activities of a street network by ranking their nodes: a case study in Murcia, Spain [J]. *International Journal of Geographical Information Science*, 2014, 28(3/4): 479-495.
- [22] Qiu X, Suganthan P N, Amaratunga G A J. Ensemble incremental learning random vector functional link network for short-term electric load forecasting [J]. *Knowledge-Based Systems*, 2018, 145(4):182-196.
- [23] Wang J, Lou C, Yu R, et al. Research on hot micro-blog forecast based on XGBOOST and random forest [M]. *Knowledge Science, Engineering and Management, KSEM 2018, Lecture Notes in Computer Science*, Springer.
- [24] Li C, Chen Z.Y, Liu J.B, et al. Power load forecasting based on the combined model of LSTM and XGBoost [C]//PRAI '19:Proceedings of the 2019 the International Conference on Pattern Recognition and Artificial Intelligence. Wenzhou, China: ACM, 2019: 46-51.
- [25] Gómez-Ríos A, Luengo J, Herrera F. A study on the noise label influence in boosting algorithms : Adaboost, GBM and XGBoost [C]//International Conference on Hybrid Artificial Intelligence Systems (HAIS), 2017:268-280.
- [26] Yue Yanchun, Huang Tingzhu. Error reciprocal variable weight combination prediction method [J]. *Journal of University of Electronic Science and technology*, 2007 (S1): 349-351.
- [27] Cao y B, Xu J, Liu T Y, et al. Adapting ranking SVM to document retrieval[C] //Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. New York: ACM Press, 2006: 186-193.
- [28] Xu P P, Mei H H, Ren L, et al. ViDX: visual diagnostics of assembly line performance in smart factories [J]. *IEEE Transactions on Visualization and Computer Graphics*, 2017, 23(1):291-300.
- [29] Sun Guanglu, song Zhichao, Liu Jinlai, Zhu Suxia, he Yongjun. Feature selection method based on maximum information coefficient and approximate Markov blanket [J]. *Journal of automation*, 2017, 43 (05): 795-805. Doi: 10.16383/j.aas.2017.c150851.
- [30] Zhang Li, Yuan Yuyu, Wang Zong. Fcbf feature selection algorithm based on maximum correlation information coefficient [J]. *Journal of Beijing University of Posts and telecommunications*, 2018, 41 (04): 86-90. Doi: 10.13190/j.jbupt.2017-229.