Journal of Organizational and End User Computing
Volume 34 • Issue 4

Predicting Patients' Satisfaction With Doctors in Online Medical Communities:
An Approach Based on XGBoost Algorithm

Yunhong Xu, Kunming University of Science and Technology, China
Guangyu Wu, Kunming University of Science and Technology, China
Yu Chen, Kunming University of Science and Technology, China

ABSTRACT

Online medical communities have revolutionized the way patients obtain medical-related information and services. Investigating what factors might influence patients' satisfaction with doctors and predicting their satisfaction can help patients narrow their choices and increase their loyalty towards online medical communities. Considering the imbalanced feature of dataset collected from Good Doctor, the authors integrated XGBoost and SMOTE algorithms to examine what factors can be used to predict patient satisfaction. SMOTE algorithm addresses the imbalanced issue by oversampling imbalanced classification datasets. And XGBoost algorithm is an ensemble of decision trees algorithm where new trees fix errors of existing trees. The experimental results demonstrate that SMOTE and XGBoost algorithms can achieve better performance. The authors further analyzed the role of features played in satisfaction prediction from two levels: individual feature level and feature combination level.

KEYWORDS

Machine Learning, Online Medical Community, Satisfaction Prediction, SMOTE, XGBoost

INTRODUCTION

With the large-scale popularization of the Internet, the concept of "Internet +" has gradually penetrated into the medical industry, and the online medical communities have emerged (Li et al., 2019). Online medical communities have revolutionized the way patients obtain medical-related information and services. The online medical community enables patients to contact doctors without time and space constraints, thereby making it more convenient to receive medical treatment at a lower cost (Sims, 2016). In addition to convenience and cost, patients can also obtain diverse medical suggestions from different doctors, which allow them to make better medical decisions. Doctors can obtain social and economic benefits from engagement in online medical communities, e.g. social recognition of patients and financial returns.

Considering the benefits patients and doctors can obtain from online medical communities, more and more patients and doctors participate in online medical communities (Li et al., 2018). Under the traditional medical mode, due to lack of professional medical knowledge, patients habitually choose high-level hospitals or doctors with strong medical background (Liu et al., 2018). Medical treatment is a multi-dimensional and complex process. In addition to doctors' medical background,

DOI: 10.4018/JOEUC.287571

the way doctors communicate with their patients, how they respond to patients, and other patients' experiences can also influence patients' satisfaction with doctors. However, it is difficult to trace these variables under the traditional medical mode. Online medical community can record several features that occur during or after medical treatment, which provides a holistic framework to predict patient satisfaction with doctors.

Patients' satisfaction with doctors can be defined as the extent to which patients are content with the medical services which they receive from doctors. Understanding what factors that drive patient satisfaction and how these factors can be used to predict patient satisfaction can facilitate patients to make better decisions and promote doctors to provide better medical services. Previous studies have shown that patients' satisfaction with their medical service providers tends to be more positive, that is, patients are more willing to give doctors satisfactory evaluation (Marcinowicz et al., 2009). And data evaluated as dissatisfied may be more important than the satisfaction evaluation data when studying patients' satisfaction (Vaida & Osmo, 2003). Therefore, we should consider both satisfaction and dissatisfaction data when predicting patients' satisfaction. When both kinds of data are considered, there are 6933 satisfactory samples and 77 unsatisfactory samples in our dataset collected from Good Doctor. The data is imbalanced in that the samples classified as positive (which means patients who are not satisfied with their doctors) and negative (which means patients who are satisfied with their doctors) are not equally distributed. Particularly, positive samples account for about 1.1% of the total sample. Data imbalance would cause the performance of machine learning algorithms to degrade, because most machine learning algorithms will ignore or have poor performance on the minority class (Wang et al., 2020). There are two ways to deal with imbalanced dataset: data resampling and algorithms based on ensemble learning. For data resampling, the Synthetic Minority Over-sampling Technique (SMOTE) (Chawla et al., 2002) is used in this research to synthesize new samples from the minority class. XGBoost is a large-scale ensemble algorithm proposed by Chen et al. (2016), which performs a second-order Taylor expansion on the cost function, where both the first-order derivative and the second-order derivative are used. Previous research has demonstrated that SMOTE and XGBoost algorithm can achieve better results when confronted with imbalanced datasets (He et al., 2021; Meng et al., 2020; Wang et al., 2020). In this research, we integrate XGBoost and SMOTE to predict patients' satisfaction with doctors.

LITERATURE REVIEW

Patient satisfaction with doctors is a multi-dimensional concept, which should be operationalized and considered under particular contexts (Naidu, 2009). In offline medical context, questionnaire was used to measure patient satisfaction with doctors. Hussain et al. (2019) found that doctor services and waiting time can influence patient satisfaction. Alharbi et al. (2018) reviewed relevant research and concluded that availability and accessibility, relational conduct, technical skills/knowledge, and personal qualities would affect patient satisfaction. Prado-Galbarro et al. (2020) found that a worse perception of health status and a longer time spent in the waiting room were negatively associated with patient satisfaction. Sl et al. (2020) found that while waiting time negatively influences patient satisfaction, positive communication experiences with health care providers could potentially alleviate negative effects of long waiting time. Boquiren et al. (2015) conducted a literature review and identified five dimensions that can determine patient satisfaction with doctors, including (a) communication attribute, (b) related conduct, (c) technical skill, (d) personal qualities, and (e) availability and accessibility. In comparison to research in offline medical context, the research on patient satisfaction with doctors in online medical community is still preliminary. The depth of an interaction and service content during a first consultation influence patients' subsequent consultation behaviors and satisfaction (Yang et al., 2019).

Since the research on patient satisfaction in online medical community is limited, we reviewed literature on doctor recommendation. Although doctor recommendation and patient satisfaction are two different research questions, they are relevant to some extent in that both need to consider the features of doctors and patients. Reviewing doctor recommendation literature enables us to understand what features have been considered when recommending doctors. After analyzing 400 questions from Yahoo Answers, Chiu and Cheng (2016) suggested that four types of attributes should be considered when making doctor recommendation, e.g., physician-related, patient-related, illness and disease-related, and institution and procedural-related. Based on doctors' information and their review information, Yan et al. (2020) integrated probabilistic matrix factorization and convolutional neural network to recommend doctors. Gayen and Chakraborty (2020) used content based recommendation to recommend doctors to potential patients, where various kinds of information were involved, e.g., user comments, symptoms, medical records, geographical location, medical department, experience, publications of doctors. Based on collected patient and doctor data, Han et al. (2018) proposed a hybrid recommender system to present each patient a list of doctor recommendations. Narducci et al. (2017) presented a recommendation system which used patient conditions and patient treatments to compute a similarity score between two patients, and then recommendations were made based on the similarity score. Yang et al. (2018) collected data from haodf.com and used information such as name of doctor, name of patient, patient disease, effect of doctor diagnosis, doctor attitude, reason for choosing the doctor, and the current condition of the patient to recommend satisfactory doctors.

Previous research on user satisfaction prediction in other contexts enables us to know what approaches have been used to predict satisfaction. And we could choose an appropriate approach to predict satisfaction based on these approaches. Satisfaction prediction has been implemented in various fields, e.g., new product development, product and service evaluation. Bekiros et al. (2019) used rough sets, neural networks, support vector machine and multi-criteria analysis to forecast customer satisfaction in the shipping industry. Based on past and present customer review data, Chan et al. (2020) integrated semantic analysis and genetic programming algorithm to predict customer satisfaction in the context of new product development. Hew et al. (2020) employed supervised machine learning algorithm, sentiment analysis and hierarchical linear modelling to predict students' satisfaction. Sánchez-Franco, et al. (2018) used tf-idf, bag of words and Bayes classifier to identify relevant features to classify customer satisfaction on Yelp. Isha et al. (2020) used ordered logit model and neural network model to identify factors impacting rental car services' overall satisfaction using the data collected in a survey.

We reviewed literature on the factors that determine patients' satisfaction with doctors in both offline and online medical contexts. Due to the limited literature on patient satisfaction with doctors in the online medical community, we resort to the doctor recommendation literature and find out what kind of information was used to make recommendation. We found that dimensions such as technical skills, doctors' behaviors, treatment process, and treatment outcome are all addressed in the previous literature. However, how to systematically integrate these dimensions of information to predict patients' satisfaction with doctors in online medical communities needs further investigation. Satisfaction prediction has been extensively studied in the previous literature, and many machine learning algorithms have been proposed. However, predicting patients' satisfaction with doctors in the online medical communities has received little attention. Since the machine learning algorithms used for satisfaction prediction are contextual and data aware, directly applying previous research results to predict patients' satisfaction with doctors in online medical communities may lead to ineffective performance. Moreover, the exiting literature lacks a mechanism to explain the roles of different factors played in satisfaction prediction. In summary, we attempt to address the research gaps by systematically integrated XGBoost and SMOTE algorithm to predict patient satisfaction. Moreover, we further analyzed the role of features played in satisfaction prediction from two levels: individual feature level and feature combination level.

THE RESEARCH FRAMEWORK

The research framework can be described as follows. Guided by the relevant dimensions from current literature, we categorized the data collected from Good Doctor into four types: competence characteristics of doctors, online efforts of doctors, service evaluation of doctors, and patient medical treatment process. Then we presented how to preprocess features and used mutual information to determine which features will be remained in the data preprocessing part. Considering the imbalanced dataset, we integrated XGBoost and SMOTE algorithm to examine what factors and how to use them to predict patient satisfaction. Then we compared the performance of the proposed approach with other baseline approaches. We further analyzed the role of features played in satisfaction prediction from two levels: individual feature level and feature combination level.

DATA COLLECTION AND KEY FEATURES

Data collection

An online medical community named Good Doctor (Haodf.com) was chosen to collect data. Good Doctor is the largest online medical community in China. Currently, there are 9,640 hospitals and 783,737 doctors on Good Doctor. The large amount of information about hospitals, doctors and patients contained by this platform provides us a rich dataset to investigate patients' satisfaction with doctors. Figure 1 shows the homepage of a particular doctor on Good Doctor. Python was used to crawl the webpage information of all doctors' personal home pages in all endocrinology departments on Good Doctor before July 25, 2019. The crawled data was composed of 12,928 samples.

According to our literature review, we found that dimensions such as technical skill, doctors' behavior, treatment process, and treatment outcome are all addressed in previous literature, although they have different meanings in online and offline medical contexts. Based on these dimensions and data collected from Good Doctor, we classified these features into four categories: competence characteristics of doctors, online efforts of doctors, service evaluation of doctors and patient treatment process. Competence characteristics of doctors are consistent with technical skills. Online efforts of doctors are related to doctor behavior. Treatment process and outcome features such as patient evaluation are also addressed. Data samples that do not contain these characteristics were removed and the final data was composed of 7,010 samples. Next, we will introduce these key features one by one.

Key Features

Competence Characteristics of Doctors

In China's medical hierarchy, doctors are granted different titles based on distinct evaluation systems, e.g., academic title and clinical title. Particularly, academic title (e.g., professor, associate professor, assistant professor) reflects doctors' academic competence. Clinical title (e.g., chief physicians, associate chief physician, and physician) reflects doctors' medical and clinical ability. Besides academic and clinical title, Good Doctor also contains the educational information of doctors, e.g., their degree. Degree (e.g., Ph.D., master and bachelor) reflects the training and educational level of doctors. Furthermore, the working environment of a doctor may also reflect his/her competence. In China, medical resources are not equally distributed. Big cities have more medical resources and doctors there might have more opportunities to experience new medical knowledge and techniques. Therefore, we use region and hospital department to reflect this aspect.

Online Efforts of Doctors

Online efforts mean the time and energy a doctor might take to treating patients, which reflect a doctor's enthusiasm and dedication. In this research, three variables were used to measure online efforts of doctors, tenure, active degree and the number of published articles per month. Tenure measures the

Figure 1. The personal homepage of a particular doctor on Good Doctor

time interval from registration to the time we crawled the data. The tenure of doctors varies quite differently. Several doctors registered Good Doctor as early as 2008, while others registered a few days before July 25, 2019. Active degree measures doctors' active level on Good Doctor. Active degree was calculated as follows. We first calculated the time interval between the last time when doctors used Good Doctor and the time we crawled the data. Then the time interval was divided into different time boxes. Based on time boxes and standard of 5 days, we assigned 1-7 to each doctor, while 1 means this doctor was active 31 days ago and was least active. And 7 means this doctor was active in recent 5 days. The number of published articles on doctors' home pages was accumulated over time. Considering the variety of tenure for different doctors, we used the published medical articles per month to measure doctors' online efforts.

Service Evaluation of Doctors

Users' evaluation reflects their attitude towards particular product or service. Compared with information provided by product/service providers, online evaluations provided by individual users are more useful in that they can assist other users to make better decisions (Zhang et al., 2020). Particularly, positive evaluation might lead to users' attachment towards a product/service, while negative evaluation would lead to their avoidance behavior. In online medical communities, patients will evaluate doctor's service after receiving treatment or consultation. In this research, three variables were used to measure service quality of doctors: total number of treated patients, number of positive feedback per month and medical expenses. Total number of treated patients indicates the popularity of doctors, and the larger number means that more patients are treated by doctors. Total number of treated patients indirectly reflects the service quality of doctors in that patients may perceive that doctors treated more patients tend to provide higher quality medical services. In this research, positive feedback includes virtual gifts and thank-you letters, which are two ways for patients to express their gratitude. Since virtual gifts should be purchased by patients and thank-you letters should be written by patients to express their gratitude, virtual gifts and thank-you letters reflect the patient's recognition of doctors' medical services to a certain extent.

Patient Medical Treatment Process

Patient medical treatment is a complex process where various steps are involved, e.g., doctor selection, registration of selected doctors, receiving treatment and consultation, and paying treatment fee. In this research, following features related to the medical treatment process were extracted: the reasons why patients choose doctors, registration way, the treatment method and medical expenses. Patients may choose doctors for different reasons, e.g., recommended by other doctors, recommended by hospitals, recommended by acquaintances, doctor selection based on online evaluations, and chosen by patients themselves. Several patients claim that they choose doctors based on both online evaluations and other doctors' recommendations. We preserve this kind of information to truly record the real reasons of patient to select doctors. Patients can register in a variety of ways, e.g., online appointments, queuing appointments, self-service registration, doctor appointment numbers, ticket scalpers, telephone appointments, follow-up appointments, and number replacement. Treatment methods include inspection, re-examination, no treatment required, no plan, medication, surgery, diet exercise, and hospitalization. Medical expenses reflect the costs of patients receiving medical treatment.

DATA PREPROCESSING

To ensure the quality of data, the data was pre-processed and normalized first. In order to verify whether all the features have an impact on the predicted label (satisfaction), this research uses mutual information to filter the features and initially determine the importance of the features. Mutual information can be used to measure the correlation of nonlinear features, and the result is the correlation between every feature and satisfaction. The calculation results demonstrate that the mutual information between all features and classification tags is larger than zero, indicating that all features have a certain correlation with classification. In order to further explore the influence of these features on patient satisfaction, all features were remained in this research.

PREDICTION MODELING

The purpose of this research is to predict patient satisfaction with doctors (satisfied or dissatisfied), which is a binary classification problem. Considering the imbalanced characteristic of data, ensemble learning algorithm based on tree models was used to develop predictive models. Ensemble learning algorithm based on tree models integrates multiple weak classifiers (single tree) to form a new strong classifier through a certain strategy, which can achieve better performance than a single weak classifier. This paper used three ensemble learning algorithms: Random Forest, GBDT and XGBoost. The experimental results demonstrated that XGBoost can achieve the best results, therefore XGBoost was chosen as the final prediction model. XGBoost is a special boosting tree model, which belongs to an additive model in ensemble learning algorithm. It accumulates the weights of all leaves of all subtrees to get the predictive value of each sample. The main idea of the spanning tree is to fit the *t*-th tree according to the residual of the previous *t-1*-th training model iteration, so that the residual sum of the final model can be minimized to obtain the optimal value. The doctor satisfaction prediction based on XGBoost algorithm is described as follows.

The first step is to determine the composition of the XGBoost ensemble algorithm. For the *i*-th sample, the prediction score of this sample can be calculated as follows:

$$\hat{y}_i = f_1(x_i) + f_2(x_i) + \ldots + f_K(x_i) = \sum_{k=1}^{K} f_k(x_i) \tag{1}$$

$$f_k(x_i) = \omega_{q(x)} \tag{2}$$

Where x_i represents the *i*-th patient sample, $\hat{y}_i$ is the sum of the prediction scores of all *k* trees for the *i*-th sample. $f_k(x_i)$ denotes the prediction score of the *k*-th tree for the *i*-th sample. *q* denotes the distribution of leaf nodes of the *k*-th tree, and ω represents the one-dimensional vector of the weight set of leaf nodes.

The second step is to determine the objective function of XGBoost algorithm. The objective function is calculated using the loss function. As shown in equation (3), the goal of the objective function is to find the minimum loss between the targets and predictions for all samples.

$$Obj^{(t)} = L(\varnothing)^{(t)} \approx \sum_{i=1}^{m} l[(y_i, \hat{y}_i^{(t-1)} + f^{(t)}(x_i)] + \Omega\left(f^{(t)}\right) \tag{3}$$

$$\Omega\left(f^{(t)}\right) = \gamma T + \frac{1}{2}\lambda\omega^2 \tag{4}$$

In equation (3), *t* denotes the *t*-th iteration of the training process. y_i is the target value of the sample *i*, $\hat{y}_i^{(t-1)}$ represents the sum of prediction scores obtained from the *t-1*-th iteration of sample *i*, $f^{(t)}(x_i)$ is the prediction score of *t*-th iteration of the sample *i*. $l\left[y_i, \hat{y}_i^{(t-1)} + f^{(t)}(x_i)\right]$ denotes the loss between the targets y_i and predictions $\hat{y}_i^{(t-1)} + f^{(t)}(x_i)$ for sample *i*. In order to prevent over-fitting, the regular term $\Omega\left(f^{(t)}\right)$ is added. As shown in equation (4), γT represents the numerical complexity of leaf nodes, γ is a constant between 0 and 1, and *T* represents the number of leaf nodes of the tree generated through the *t*-th iteration. $\frac{1}{2}\lambda\omega^2$ represents the *L2* regular term of leaf nodes, where ω is the weight of leaf nodes. Adding this term can avoid excessive weight of some leaves.

The third step is to find the optimal solution of the objective function, that is, to find the minimum value of the loss function. To find the optimal solution, a second-order Taylor expansion was applied. The second-order Taylor expansion was shown in equation (5).

$$f\left(x+\Delta x\right) \cong f\left(x\right)+f'\left(x\right)\Delta x+\frac{1}{2}f''\left(x\right)\Delta x^2 \text{ (5).}$$

Then the second-order Taylor expansion was applied on equation (3), and we get:

$$L\left(\varnothing\right)^{(t)} \approx \sum_{i=1}^{m}[l(y_i,\hat{y}_i^{(t-1)})+g_i f^{(t)}\left(x_i\right)+\frac{1}{2}h_i f^{(t)2}\left(x_i\right)]+\Omega\left(f^{(t)}\right) \tag{6}$$

In this equation,

$$g_i=\partial_{\hat{y}_i^{(t-1)}} l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{7}$$

Equation 7 represents first partial derivative of the loss function.

$$h_i=\partial^2_{\hat{y}_i^{(t-1)}} l\left(y_i,\hat{y}_i^{(t-1)}\right) \tag{8}$$

Equation 8 represents second partial derivative of the loss function.

Since the term $l\left(y_i,\hat{y}_i^{(t-1)}\right)$ is the loss value obtained through the *t-1*-th iteration, which is not related to the fitting of the model in the *t*-th iteration. Therefore, the term is removed from the objective function. And we have equation (9) shown as follows:

$$L\left(\varnothing\right)^{(t)} \approx \sum_{i=1}^{m}[g_i f^{(t)}\left(x_i\right)+\frac{1}{2}h_i f^{(t)2}\left(x_i\right)]+\Omega\left(f^{(t)}\left(x_i\right)\right) \tag{9}$$

$\mathrm{L}\left(\varnothing\right)^{(\mathrm{t})}$ is the sum of the loss of all samples in the *t*-th iteration. Multiple samples can be distributed on a leaf node, and the samples mapped to the same leaf node have the same score. In order to facilitate the calculation, we can traverse the leaf node first, and then traverse the sample set on the leaf node. If we use I_j to represents all the samples belonging to the *j*-th leaf node, then formula (9) can be rewritten as:

$$L\left(\varnothing\right)^{(t)} \approx \sum_{j=1}^{T}[\frac{1}{2}\left(\sum_{i\in I_j} h_i\right)\omega_j^2 + \left(\sum_{i\in I_j} g_i\right)\omega_j] + \gamma T + \frac{1}{2}\lambda\omega_j^{\ 2} \tag{10}$$

$$= \sum_{j=1}^{T}[\frac{1}{2}\left(H_j + \lambda\right)\omega_j^2 + G_j\omega_j] + \gamma T \tag{11}$$

Obviously, equation (11) can be considered as a quadratic equation of variable ω_i. The critical point of the equation can be obtained by setting the first derivative of ω_j to zero. Then the optimal weight corresponding to leaf node *j* can be obtained.

$$\omega_j^* = -\frac{G_j}{H_j + \lambda} \tag{12}$$

Replacing ω_j^* with $-\frac{G_j}{H_j + \lambda}$, then equation (11) can be rewritten as:

$$L\left(\varnothing\right)^{(t)} = -\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \gamma} + \gamma T \tag{13}$$

From objective function as shown in equation (13), we can find that the subformula of each leaf node ($-\frac{1}{2}\sum_{j=1}^{T}\frac{G_j^2}{H_j + \gamma}$) is independent of each other. In other words, when the subformula of each leaf node reaches the minimum value, then the whole objective function ($\sum_{j=1}^{T}\frac{G_j^2}{H_j + \gamma}$) will reach the minimum value.

We call the sum of subformulas of all leaf nodes in a subtree as structure score, and whether the tree continues to split can be determined by the gain before and after node splitting. The tree structure is realized by splitting nodes to form left and right subtrees. And the gain before and after node splitting can be used to determine whether the tree continues to split. In other words, the gain is the difference between the sum of the structure scores of the left and right subtrees and the structure scores of the leaf node before node splitting. The node splitting gain of XGBoost is calculated as follows:

$$Gain = \frac{1}{2}\left(\frac{G_L^2}{H_L + \gamma} + \frac{G_R^2}{H_R + \gamma} - \frac{(G_L + G_R)^2}{H_L + H_R + \gamma}\right) - \gamma \tag{14}$$

Based on gain information, we can decide whether the tree continues to split.

EXPERIMENT AND RESULT ANALYSIS

Baseline Approaches

To evaluate the performance of the proposed XGBoost algorithm, we compared the XGBoost algorithm with other baseline approaches which are based on tree learning algorithm, including decision trees, random forest, and GBDT. Decision tree (Quinlan, 1986) is a hierarchical structure composed of nodes and directed edges, where the paths from root to leaf node represent classification rules. Single decision tree may lead to bias and variety. Ensemble methods which combines several decision trees can achieve better performance, the basic principle is that weak learners can work together to form a strong learner. Bagging and Boosting are two widely used ensemble methods. The idea of Bagging is to create subsets of data and using each collection of subset data to train prediction models. The final results are generated by aggregating the results of each prediction model. Random Forest (Lim et al., 2014) is an extension over bagging, which uses random subset and selection of data to grow trees. Boosting works by combining weak prediction models into a single strong model in an iterative manner. Common Boosting algorithms include GBDT (Friedman, 2000) and XGBoost algorithms. The reason we choose these baseline approaches is that although they are based on decision trees to generate prediction models, the mechanisms to generate final results are quite different.

Evaluation Metrics

As mentioned before, there were two types of samples in this paper (satisfied samples and dissatisfied samples). Therefore, the research problem can be considered as a binary classification problem. The data used in this research is quite imbalanced in that the positive samples (samples labeled as dissatisfied) account for less than 1% of the total samples. Generally, we take the key identified samples as positive examples. Because unsatisfactory samples are the key identification targets, they are set as positive samples. For binary classification problem, widely used evaluation metrics include accuracy, precision and recall, etc. In this context, the commonly used accuracy metric is not appropriate. In extreme situation where all predicted samples are positive, the precision can still reach to 99%. Moreover, machine learning algorithms will also tend to predict samples with larger proportion, which will fail to identify positive samples. Therefore, it is important to choose the appropriate metric to evaluate the performance of classification approach when the data is imbalanced. Two metrics were used to evaluate the performance of the proposed approach: AUC and kappa value.

Results

Prediction Results of Four Approaches

As mentioned before, the original data is imbalanced in that a large difference exists between the number of satisfied and dissatisfied samples. To address the issue of imbalanced samples, SMOTE (Synthetic Minority Over-sampling Technique) was used to create synthetic samples so that the categories in the original data (satisfied and dissatisfied) were not seriously imbalanced. Since parameters can influence the performance of prediction approaches, we adjusted the parameters to achieve best results. For every approach, their performance was evaluated in terms of two metrics: AUC and Kappa. For every approach, we compare their performance in different contexts: with or without SMOTE, using training or test set.

The evaluation results of four approaches are shown in table 1. From table 1, we can see that XGBoost (without SMOTE) or XGBoost (with SMOTE) can achieve the best prediction performance in three contexts. The only exception is using training set and AUC metric to evaluate these approaches. As shown in the second column, Decision Tree (with SMOTE) can achieve the best performance in that the AUC value (0.896) of this approach is the highest among all prediction approaches. However, when the performance is extended to test set, the AUC value of Decision Tree (with SMOTE) is not as good as XGBoost (with SMOTE). As shown in the third column of table 1, the AUC value

Table 1. Evaluation results of four approaches

Prediction approaches	AUC of Training Set	AUC of Test Set	Kappa of Training Set	Kappa of Test Set
Decision Tree	0.756	0.680	0.572	0.396
Decision Tree (SMOTE)	**0.896**	0.718	0.598	0.235
Random Forest	0.644	0.636	0.444	0.427
Random Forest (SMOTE)	0.824	0.724	0.678	0.395
GBDT	0.750	0.682	0.664	**0.531**
GBDT(SMOTE)	0.847	0.725	0.777	0.450
XGBoost	0.697	0.682	0.55	**0.531**
XGBoost(SMOTE)	0.885	**0.726**	**0.827**	0.496

of Decision Tree (with SMOTE) is 0.718, which is smaller than the AUC value of XGBoost (with SMOTE). This could be explained by over-fitting phenomenon, which means the prediction model performs well on the training set, but the result is not good when it comes to the test set.

We also notice that four approaches (with SMOTE) perform better than their original approaches in most contexts. The reason is that when using SMOTE, the proportion of positive samples is enlarged, which enables the prediction model to learn the features of positive samples. On the other hand, SMOTE may lead to the results that the training set has too much repeated information, which will decrease the performance of prediction model (as shown in the last column of table 1). And when using SMOTE, the AUC value increases and Kappa value decreases on the test set. Therefore, we should balance between the advantages and disadvantages of SMOTE and decide whether SMOTE will be combined with other prediction models.

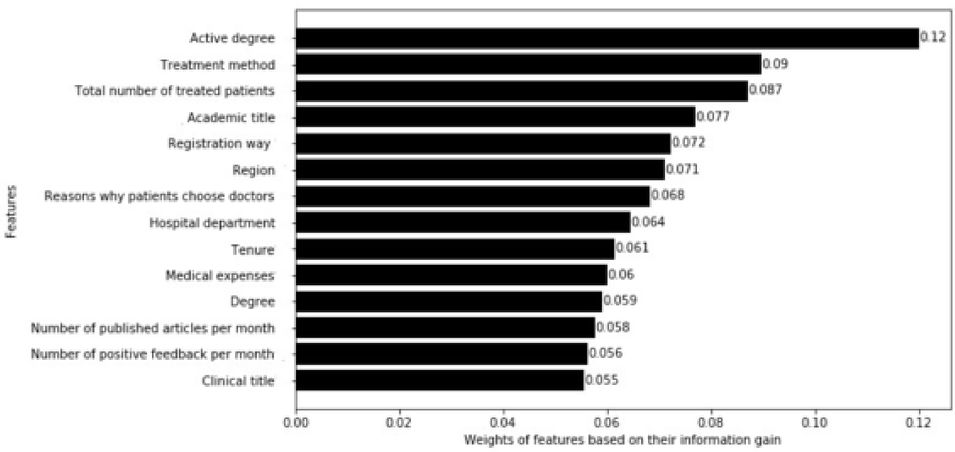
Analyzing the Role of Factors in Satisfaction Prediction at Individual Feature Level

From the experimental results, we know that XGBoost can achieve the best performance. Next, we will analyze the roles of features played in XGBoost. We rank the features based on the average information gain of all tree nodes after a certain feature is split in XGBoost. Figure 2 demonstrates the ranking results of features based on their average information gain. In this part, we focus on the first four features and explore their roles in satisfaction prediction.

We first examine the role of active degree played in satisfaction prediction. As shown in figure 2, the average information gain of active degree ranks the highest among all features. According to the statistics of the original data, the data samples with the active degree of 1 and 7 account for the majority. Among them, data samples with the active degree of 1 (least active) account for about 12% of the total samples and data samples with the active degree of 7 (most active) account for about 83% of the total samples. We further analyze the data samples labeled as "satisfied" whose active degree are 1 and 7 respectively. The satisfied samples with active degree 1 account for 96% of the total samples with active degree 1, while the satisfied samples with active degree 7 account for 99.4% of the total samples with active degree 7. From these figures, we can infer that doctors with higher active degree tend to generate satisfaction evaluation results. Higher active degree of doctors means that they are always there and care for patient concerns.

Next, we will investigate the role of medical treatment method played in satisfaction prediction. Among total samples marked as "dissatisfied", 49.3% of samples are treated with drugs only, and 32.5% of the samples are without solution. These figures suggest that patients are more satisfied with doctors who can provide them with long-term treatment method, which is consistent with current over treatment phenomenon in China.

Figure 2. Ranking of features based on their average information gain



When the total number of treated patients is concerned, we first investigate the maximum number of total treated patients. The maximum number of total treated patients was 3,815. Based on this number, we divided the samples into three categories, whose total number of treated patients was located in different intervals: [0, 1271], [1272, 2543] and [2544, 3815]. The corresponding proportions of the total dissatisfied samples in each interval were 1.7%, 0.3%, 0.2%. It can be inferred that patients are more satisfied with doctors who treat more patients. More treatment experiences enable doctors to enrich their medical knowledge and communication skills. Therefore, those doctors tend to be highly recognized by patients.

As far as academic tile is concerned, we found that doctors without academic titles accounted for the highest proportion of dissatisfied samples. Patients may be more critical of doctors without academic titles. Different from common belief, professors who have the highest academic titles account for a high proportion of dissatisfied samples. The possible explanation is that patients have high expectation of doctors with higher academic titles. When doctors fail to meet patients' expectations, they will express their dissatisfaction attitude.

Analyzing the role of feature combination played in satisfaction prediction

When gradually adding features to predict patient satisfaction, the experimental results demonstrate that XGBoost (with SMOTE) can achieve better performance than XGBoost (without SMOTE). Therefore, to further investigate the influence of the number of features and feature combinations, the data was processed using SMOTE. Then XGBoost was used to predict patient satisfaction. Features were gradually added to the prediction model based on their weight information using information gain as demonstrated in previous section. For ease of presentation, the top 6 features were used for stepwise prediction. And the results are shown in table 2.

The first column of table 2 represents the combination of features. For example, (1, 2) represents that the top 2 features (active degree and treatment method) are used to predict patient satisfaction. From the second row, we can see that the performance is not good in that the Kappa value is very small, especially for the test set. Gradually adding features to the prediction model, we can see that the overall performance on the training set has improved (both in terms of AUC and Kappa), which shows that the model can make better use of information for prediction. The increase also has been demonstrated via the Kappa value on the test set. As shown in the last column, the Kappa value shows an upward trend. However, there is no such upward trend in terms of AUC of test set. As far as the AUC of the test set is concerned, more features are not guaranteed to improve the results. As shown in the third column of table 2, the AUC value drops first and then remains at 0.636. The above analysis demonstrates that more features may generate better and more stable results, especially in terms of Kappa value. However, whether more features should be involved in the prediction model depends on the domain context and the real data.

Table 2. Analysis of prediction results for different feature combinations

Feature combinations	*AUC of Training Set*	*AUC of Test Set*	*Kappa of Training Set*	*Kappa of Test Set*
(1,2)	0.646	**0.698**	0.096	0.091
(1,2,3)	0.598	0.636	0.313	0.426
(1,2,3,4)	0.613	0.636	0.358	0.397
(1,2,3,4,5)	**0.651**	0.636	**0.451**	**0.427**
(1,2,3,4,5,6)	0.636	0.636	0.415	**0.427**

CONCLUSIONS AND FUTURE RESEARCH

This research integrates XGBoost and SMOTE to predict patients' satisfaction with doctors. This research has several theoretical contributions. First, this research enriches current literature on patients' satisfaction with doctors in online medical community. Although extensive research has been conducted to examine patients' satisfaction with doctors in offline medical contexts, literature on this topic in online contexts is limited. Following relevant literature and using the real data from an online medical community, we demonstrate that competence characteristics of doctors, online efforts of doctors, patient medical treatment process, and service evaluation of doctors could influence users' satisfaction with doctors in online medical community. Second, we propose the approach based on XGBoost to predict patient satisfaction with doctors. The crawled data is imbalanced, therefore XGBoost was combined with SMOTE to make prediction. The experimental results demonstrated that the proposed approach outperforms other baseline approaches in terms of AUC and Kappa. Third, we analyze the roles of features played in satisfaction prediction at individual level and at feature combination level. To investigate the role of factors played in satisfaction prediction at individual level, features were ranked based on their information gain. And the weights of features represent their importance in satisfaction prediction to some extent. To investigate how feature combination influences satisfaction prediction, different feature combinations were used in the stepwise prediction.

This research has several practical implications. For patients in online medical community, providing value-added service such as predicting their satisfaction with doctors can help them to narrow down their choices from a large number of candidate doctors, which will make their medical experience easier and more convenient. For doctors, understanding what factors and how these factors might influence patients' satisfaction with them can provide several guidelines to direct their behavior. Particularly, they can improve patients' satisfaction from following aspects: competence characteristics, online efforts, service evaluation of doctors, and patient medical treatment process. For online medical community, the business and social value of online medical community greatly depends on patients' engagement in such communities. Providing value-added service can improve users' loyalty toward a particular online medical community. Moreover, understanding what factors might influence patients' satisfaction with doctors might help platform design efficient mechanisms to motivate doctors to change their behavior and interactions with patients, which will ultimately increase patients' satisfaction.

This research suggests a few areas that merit further investigation. First, the data was collected from one department of Good Doctor, whether the results can be generalized to other departments or other communities needs further investigation. Second, this research used four dimensions of features to predict patient satisfaction. In future research, other text data like patients' thank-you letters would be used to enrich the features. Third, this research proposed an XGBoost approach based on SMOTE to predict patient satisfaction. Other advanced techniques such as data analytics and machine learning approaches can also be used to make predictions (Shiau et al., 2021; Doganer et al., 2021).

ACKNOWLEDGEMENTS

This research was partially supported by the National Natural Science Foundation of China (71861019, 71361017, 71640021).

REFERENCES

Alharbi, M., Alhaiti, A., & Gillespie, J. (2018). Patient satisfaction and experience of primary care in Saudi Arabia: A systematic review. *International Journal for Quality in Health Care*, *30*(10), 751–759. doi:10.1093/intqhc/mzy104 PMID:29860320

Bekiros, S., Loukeris, N., Matsatsinis, N., & Bezzina, F. (2019). Customer satisfaction prediction in the shipping industry with hybrid meta-heuristic approaches. *Computational Economics*, *54*(2), 647–667. doi:10.1007/s10614-018-9842-5

Boquiren, V., Hack, T., Beaver, K., & Williamson, S. (2015). What do measures of patient satisfaction with the doctor tell us? *Patient Education and Counseling*, *98*(12), 1465–1473. doi:10.1016/j.pec.2015.05.020 PMID:26111500

Chan, K. Y., Kwong, C. K., & Kremer, G. E. (2020). Predicting customer satisfaction based on online reviews and hybrid ensemble genetic programming algorithms. *Engineering Applications of Artificial Intelligence*, *95*, 103902. doi:10.1016/j.engappai.2020.103902

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*(1), 321–357. doi:10.1613/jair.953

Chen, T., He, T., & Benesty, M. (2016, August). xgboost: Extreme Gradient Boosting. *Proceedings of the 22nd ACM SIGKDD International Conference on knowledge discovery and data mining*.

Chiu, M.-H., & Cheng, W.-C. (2016, July). *Building a classification model for physician recommender service based on needs for physician information*. Paper presented at the HCI in Business, Government, and Organizations: Information Systems, Cham, Switzerland. doi:10.1007/978-3-319-39399-5_3

Doganer, A., & Zhang, Z. J. (2021). An infodemiological analysis of google trends in covid-19 outbreak: Predict case numbers and attitudes of different societies. *Journal of Database Management*, *32*(2), 1–19. doi:10.4018/JDM.2021040101

Friedman, J. (2000). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, *29*(5), 1189–1232.

Gayen, A., & Chakraborty, S. (2020). MedicaNet: Trust based recommender system to recommend doctors with the help of symptoms. *Computational Intelligence and Neuroscience*, *2020*(8826557).

Han, Q., Ji, M., Troya, I. M. R. d., Gaur, M., & Zejnilovic, L. (2018, October). A hybrid recommender system for patient-doctor matchmaking in primary care. *2018 IEEE 5th International Conference on Data Science and Advanced Analytics (DSAA)*. doi:10.1109/DSAA.2018.00062

He, J., Hao, Y., & Wang, X. (2021). An interpretable aid decision-making model for flag state control ship detention based on SMOTE and XGBoost. *Journal of Marine Science and Engineering*, *9*(2), 156. doi:10.3390/jmse9020156

Hew, K. F., Hu, X., Qiao, C., & Tang, Y. (2020). What predicts student satisfaction with MOOCs: A gradient boosting trees supervised machine learning and sentiment analysis approach. *Computers & Education*, *145*, 103724. doi:10.1016/j.compedu.2019.103724

Hussain, A., Asif, M., Jameel, A., & Hwang, J. (2019). Measuring OPD patient satisfaction with different service delivery aspects at public hospitals in Pakistan. *International Journal of Environmental Research and Public Health*, *16*(13), 1–9. doi:10.3390/ijerph16132340 PMID:31269722

Isha, N., Meghna, V., & Ashish, V. (2020). Measuring satisfaction of rental car services in India for policy lessons. *Case Studies on Transport Policy*, *8*(3), 832–838. doi:10.1016/j.cstp.2020.05.019

Li, Y., Song, Y., Zhao, W., Guo, X., Ju, X., & Vogel, D. (2019). Exploring the role of online health community information in patients' decisions to switch from online to offline medical services. *International Journal of Medical Informatics*, *130*, 103951. doi:10.1016/j.ijmedinf.2019.08.011 PMID:31473534

Li, Y., & Wang, X. (2018). Seeking health information on social media: A perspective of trust, self-determination, and social support. *Journal of Organizational and End User Computing*, *30*(1), 1–22. doi:10.4018/JOEUC.2018010101

Lim, A., Breiman, L., & Cutler, A. (2014). *bigrf: Big Random Forests: Classification and Regression Forests for Large Data Sets*. Academic Press.

Liu, N., Finkelstein, S. R., Kruk, M. E., & Rosenthal, D. (2018). *When waiting to see a doctor is less irritating: understanding patient preferences and choice behavior in appointment scheduling*. Academic Press.

Marcinowicz, L., Grbowski, R., & Chlabicz, S. (2009). Exploring negative evaluations of healthcare by Polish patients: An attempt at cross-cultural comparison. *Health & Social Care in the Community*, *17*(2), 187–193. doi:10.1111/j.1365-2524.2008.00817.x PMID:19040699

Meng, C., Zhou, L., & Liu, B. (2020). A case study in credit fraud detection with SMOTE and XGBoost. *Journal of Physics: Conference Series*, *1601*, 052016. doi:10.1088/1742-6596/1601/5/052016

Naidu, A. (2009). Factors affecting patient satisfaction and healthcare quality. *International Journal of Health Care Quality Assurance*, *22*(4), 366–381. doi:10.1108/09526860910964834 PMID:19725209

Narducci, F., Musto, C., Polignano, M., Gemmis, M. D., & Semeraro, G. (2017, April). Recommending doctors and health facilities in the HealthNet social network. *Workshop on Social Media for Personalization and Search co-located with 39th European Conference on Information Retrieval (ECIR 2017)*.

Prado-Galbarro, F.-J., Cruz-Cruz, C., Gamiño-Arroyo, A.-E., & Sanchez-Piedra, C. (2020). Satisfaction with healthcare services among patients with diabetes, hypertension, and/or dyslipidemia in Mexico: A cross-sectional study. *Value in Health Regional Issues*, *23*, 19–24. doi:10.1016/j.vhri.2019.11.002 PMID:32062192

Quinlan, R. J. (1986). Induction of Decision Trees. *Machine Learning*, *1*(1), 81–106. doi:10.1007/BF00116251

Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2018). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, *101*, 499–506. doi:10.1016/j.jbusres.2018.12.051

Shiau, W. L., Siau, K., Yu, Y., & Guo, J. (2021). Research commentary on is/it role in emergency and pandemic management: Current and future research. *Journal of Database Management*, *32*(2), 67–75. doi:10.4018/JDM.2021040105

Sims, J. M. (2016). Communities of practice: Telemedicine and online medical communities. *Technological Forecasting and Social Change*, *126*(JAN), 53–63.

Sl, A., Seg, B., Hp, A., & Ad, A. (2020). Waiting time, communication quality, and patient satisfaction: An analysis of moderating influences on the relationship between perceived waiting time and the satisfaction of breast cancer patients during their inpatient stay - ScienceDirect. *Patient Education and Counseling*, *103*(4), 819–825. doi:10.1016/j.pec.2019.11.018 PMID:31767241

Vaida, B., & Osmo, S. (2003). Why are people dissatisfied with medical care services in Lithuania? A qualitative study using responses to open-ended questions. *International Journal for Quality in Health Care Journal of the International Society for Quality in Health Care*, *15*(1), 23–29. doi:10.1093/intqhc/15.1.23 PMID:12630798

Wang, C., Deng, C., & Wang, S. (2020). Imbalance-XGBoost: Leveraging weighted and focal losses for binary label-imbalanced classification with XGBoost. *Pattern Recognition Letters*, *136*, 190–197. doi:10.1016/j.patrec.2020.05.035

Yan, Y., Yu, G., & Yan, X. (2020). Online doctor recommendation with convolutional neural network and sparse inputs. *Computational Intelligence and Neuroscience*, *2020*(3), 1–10. doi:10.1155/2020/8826557 PMID:33123187

Yang, Y., Hu, J., Liu, Y., & Chen, X. (2018). Doctor recommendation based on an intuitionistic normal coud model considering patient preferences. *Cognitive Computation*, *12*(2), 460–478. doi:10.1007/s12559-018-9616-3

Yang, Y., Zhang, X., & Lee, P. K. C. (2019). Improving the effectiveness of online healthcare platforms: An empirical study with multi-period patient-doctor consultation data. *International Journal of Production Economics*, *207*, 70–80. doi:10.1016/j.ijpe.2018.11.009

Zhang, G., Wang, Z., & Zhao, H. (2020). Relationships among perceived value, satisfaction, and e-Trust: An e-CRM view of online restaurant consumption. *Journal of Systems Science and Information*, *8*(5), 458–475. doi:10.21078/JSSI-2020-458-18

Yunhong Xu is a Professor of Information Systems at the Kunming University of Science and Technology. She received her Ph.D. in Information Systems at the City University of Hong Kong and Ph.D. in Management Science at the University of Science and Technology of China. Her research interests include network analysis, knowledge recommendation and business analytics. She has published in journals such as Decision Support Systems, Information Systems Frontiers, Expert Systems with Applications, etc.

Guangyu Wu is a master student at Kunming University of Science and Technology. Her research interests include data mining and knowledge management.

Yu Chen is professor of information System at Kunming University of Science and Technology. His research interests involve IT/IS adopt and use, e-healthcare.