WHICEB 2017 Proceedings

Wuhan International Conference on e-Business

Summer 5-26-2017

# Court Judgment Decision Support System Based on Medical Text Mining

Qing Zhu

*International Business School of Shaanxi Normal University, Xi'an, 710000, China; Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an,710000, China,* zhuqing@snnu.edu.cn

Kezhen Wei

*International Business School of Shaanxi Normal University, Xi'an, 710000, China; Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an,710000, China*

Lanlin Ding

*Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an,710000, China;School of Economics and Finance, Xi'an Jiao Tong University, Xi'an,710000, China*

Kin Keung Kin Keung Lai

*Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an,710000, China; Department of Management Sciences, City University of Hong Kong, HK, China*

# Court Judgment Decision Support System Based on Medical Text Mining

*Qing Zhu*[12*], *Kezhen Wei*[12], *Lanlin Ding*[24], *Kin Keung Lai*[23]
[1]International Business School of Shaanxi Normal University, Xi'an,710000, China
[2]Institute of Cross-Process Perception and Control, Shaanxi Normal University, Xi'an,710000, China
[3]Department of Management Sciences, City University of Hong Kong, HK, China
[4]School of Economics and Finance, Xi'an JiaoTong University, Xi'an,710000, China

**Abstract:** Medical damage is a common problem faced by hospitals around the world and is widely watched by countries and the World Health Organization. As the number of medical damage dispute lawsuit cases rapidly grows, many countries in the world face the problem how to improve the efficiency of the judicial system under the premise of guaranteeing the quality of the trial. Therefore, in addition to reforming the system, the decision support system will effectively improve judicial decisions. This paper takes medical damage judgment documents in China as example, and proposes a court judgment decision support system (CJ-DSS) based on medical text mining and the automatic classification technology. The system can predict the trail results of the new lawsuit documents according to the previous cases verdict - rejected and non-rejected. Combined with the cases, the study in this paper found that combined feature extraction method does improve the performance of three kinds of classifiers - Support Value Machine (SVM), Artificial Neural Network (ANN) and K-Nearest Neighbor (KNN), the degree of improved performance is different from using DF-CHI combined feature extraction method. In addition, integrated learning algorithm also improves the classification performance of the overall system.

**Keywords**: medical damage,;text-mining,;automatic text classification,;decision support system,;CJ-DSS

## 1. INTRODUCTION

In many other countries with common law system, new legal relations are constantly produced, making the problem of untimely formulation and revision of the statute law more visible. In practice, the appropriate addition of case-based laws can increase the flexibility of the Chinese legal system. The long-term accumulation of cases during trial practices includes a variety of cases, providing specific, vivid example for the application of law. These cases, along with the legal norms, provides more assistance to strengthen the authority and stability of law compared to the abstract legislative provisions [1]. Therefore, it is important to fully exploit the value of the court's system.

The case of medical damages is recognized as a difficult point in the case of tort damages. Medical damage is a common problem faced by hospitals around the world and is widely watched by countries and the World Health Organization. According to World Health Organization Patient Safety Solutions Collaboration Center data, millions of patients around the world are exposed to unnecessary injuries each year due to preventable medical errors [2].

With the rapidly rise in the number of medical damage dispute lawsuit cases, how to improve the trial efficiency of the court system while guaranteeing the judicial quality is a serious problem facing many countries. In addition to reforming the system, the decision support system will effectively improve judicial decisions. The Count Judgment Decision Support System (CJ-DSS) designed in this paper is an effort and attempt in this field. The system takes medical damage judgment documents in China as example, and it can predict the trail results of the new lawsuit cases according to the previous cases verdict. It also provides decision support for the court and people.

* Corresponding author. Email: zhuqing@snnu.edu.cn

Since the 1970s, Decision Support System (DSS) has been the focus in academia. Among the current studies, DSS is very popular in the medical and health research field. Clinical Decision Support System (CDSS) has shown great advantage in the auxiliary diagnosis and treatment of diseases [3,4]. However, some scholars believe that CDSS's application in medical practice is still limited [5], especially since medical workers often have difficulty understanding the internal processes of the computer system. In addition, in the field of enterprise management, DSS applications are also very common. Yinghong W et al. [6] design the DSS on the choice of virtual partners. Shufen F and Wenyuan L [7] used DSS in equipment maintenance management to help equipment managers make scientific decisions. In short, DSS design ideas have been widely used in various fields.

For CJ-DSS proposed in this paper, automatic text classification technology is the core of the system design. Tseng Y H et al. [8] believes that the patent text contains many important research results. However, its length and vocabularies make it more time and energy consuming to analyze. Kim J H et al. [9] believes that as the number of patent text increases, the automatic classification system can replace manual classification. At the same time, he stressed that compared to other texts, the patent texts contain some structured information including claims, purposes and embodiments of invention, and this feature should be concerned in the process of classification. Therefore, Kim used this feature and first selected semantic elements as the basic feature of text classification, and then used the k-Nearest Neighbor method to conduct classification. The experiment showed a 74% increase on the classification results compared to classification without the use of structured information.

Through empirical research, Pong Y H et al. [10] showed that the KNN algorithm is an important method to constructing an effective text classification system, which is an important method to enhancing the current library information intelligence. Coussement K and Poel D V D [11] used the semantic feature of Email as an index and designed an automatic mail classification system which can distinguish between complaints and non-complaints. The system achieved 83% accuracy. Al Qady M and Kandil A [12] classified project documents based on text contents and the performance of the classifier was tested under different conditions. It was found that the classifier with the highest accuracy is the Rocchio and KNN classifier that applied the dimension reduction techniques and TF-IDF weighting method. Moreover, the performance of the classifier can be improved by using the voting method.

Although the Decision Support System and text automatic classification technology has been widely used in many fields in recent years, its application in the field of law and justice has not received much attention. One of the reasons is due to the law of precedents; another reason is due to the difficulty in obtaining data and text analysis technology. However, with the establishment and development of judicial models, database and text mining techniques, Court Judgment Decision Support System (CJ-DSS) not only can improve the efficiency of the judicial system, but also can enhance the reference value of previous court cases. More specifically, with the accumulation of a large number of jurisprudence in China's judicial practice, automatic classification technology can discover valuable information to assist the judicial decision, effectively ensuring the trial quality under the premise of improving the efficiency of the trial court system.

## 2. METHODOLOGY

### 2.1 Preprocessing

#### 2.1.1 Chinese word segmentation and basic dimension reduction

Since there are differences between Chinese and English texts, there are different methods of word segmentation. In this paper, we use the Rwordseg package in R software as a tool for segmentation. The package is proposed by Jian Li, it is a Chinese word segmentation tool under the R environment that uses rjava to call for the Java segmentation tool Ansj. Ansj is a ICTCLAS Chinese segmentation algorithm developed by the Chinese

Academy of Sciences, which uses hidden Markov model (HMM). The existing Chinese word segmentation method mainly includes dictionary matching method. The Rwordseg can become more flexible and recognizable through the addition of professional vocabulary. In this paper, we can identify professional vocabularies in the sentence by loading the legal vocabulary, the legal text word and the medical word dictionary in the Sougou Cell Bank, so as to avoid the interference of the unidentified words.

In preliminary dimension reduction stage, firstly we need to delete the Chinese stop words, such as "I" " you "and other nonsense words. Then we will select more valuable words, like nouns, verbs and professional vocabulary words, by part-of-speech tagging to reduce interference from meaningless terms and ready for later stages.

**2.1.2 Text representation model and words weighting**

Feature representation of text contents mainly include the Boolean model, Vector space model, probability model and knowledge based representation model, and Vector space model is a relatively better method used in the past studies. Vector space model was proposed by Salton et al. [13] (Salton, Yang & Wang, 1975). Vector space model (VSM) says that: given a document $D = D(\mathrm{T}_1, \mathrm{W}_1; \mathrm{T}_2, \mathrm{W}_2; \ldots; \mathrm{T}_n, \mathrm{W}_n)$, where $T_i\,(\mathrm{i} = 1, 2, \cdots, \mathrm{n})$ denotes the terms in the document, which are different from each other. And $W_i\,(\mathrm{i} = 1, 2, \cdots, \mathrm{n})$ represent the corresponding number values for the terms. Now we can see $T_1, T_2, \ldots, T_n$ as an n-dimensional coordinate, and $W_1, W_2, \ldots, W_n$ is the value corresponding to n-dimensional coordinates. Therefore, the document can be seen as an n-dimensional vector.

In order to obtain higher accuracy, this paper uses vector space model. The vector space model can translate texts into electronic form, and each column of the electronic form is associated with a feature. Each row represents a text document; the word weighting represents the appearance frequency of a term in a document. The term weights typically use 0 (The term does not appear) and 1 (The term appears) to fill. Presently, the most commonly used method is the TF-IDF weighting, it is a frequency that is corrected by the word importance factor. This weighting factor is also known as the inverse document frequency (IDF) [14] (Rocchio, J. J., 1971). And then Salton et al. [15] proposed the weight calculation formula of TF-IDF:

$$
\begin{aligned}
&tf\text{ - }idf(j) = tf(j) \times idf(j) \\
&idf(j) = log\left(\frac{N}{df(j)}\right)
\end{aligned}
\tag{1}
$$

In which *tf(j)* represents the actual frequency of term $j$, N represents the number of documents, *df(j)* represents the number of documents that contains term $j$.

**2.2 Feature selection**

The difficulty encountered during the study is the high-dimensional feature space and document vector representation sparseness. In recent years, the often-used feature extraction methods of automatic Chinese text classification include DF [16] (document frequency) (Yang & Pedersen, 1997), MI (mutual information), IG [17] (information gain) (Lee & Lee, 2006) and the chi-square(CHI) test. At the same time, Dai Liu-ling et al. [18] comparatively studied the influence of feature selection method for classification results in the Chinese text classification, and the results showed that, these feature selection methods (IG, MI and CHI) outperformed in the English text classification is not suitable for Chinese text categorization without correction. He believes that in addition to increasing the training set, but also through a combination of feature extraction methods to improve performance. Therefore, this paper will select DF and CHI tested method and the combination of both (DF-CHI) to extract features. The basic idea and the advantages and disadvantages as follows:

(1) Document Frequency(DF) refers to the number of documents which contains a certain word in the dictionary. It is the simplest feature extraction technology. By setting a frequency threshold, this technology can eliminate the words with low frequency and selects words with high document frequency. The basic

assumption for the low frequency words is that it has a low level of contribution and no significant effect on the classification result. Therefore, it can reduce the feature dimension and improve the classification accuracy.

(2) The Chi-Square(CHI) statistics principle determines the correctness of a theory by observing the deviation between the actual value and the theoretical value. In the process of feature selection, it can be used to measure the degree of correlation between features and class labels, and assumes it has a distribution with a degree of freedom of 1. The original hypothesis is not related to the feature and the class label, the higher the statistical value, the greater the correlation between the entry and the category, and more valuable to the prediction results. Since Chi-Square Statistics is an extraction method that relies on categorical features and low frequency words, therefore, it contains a “low frequency word defect”. The calculation formula is as follows:

$$\chi^2(t,c) = \frac{N \times \left(AD - CB\right)^2}{(A+C)(B+D)(A+B)(C+D)} \quad (2)$$

Where N denotes the total number of documents contained in the training corpus; $c$ denotes a certain category; $t$ denotes a certain word; A denotes the number of documents that belongs to category $c$ and includes word $t$ ; B denotes the number of documents that does not belong to category $c$ but includes word $t$ ; C denotes the number of documents that belongs to category $c$ but does not include word $t$ . D denotes the number of documents that neither belongs to category $c$ nor includes word $t$ .

(3) Feature combination extraction method refers to combining two or more selection algorithms and selecting text features in order to find more valuable classification results and more predictive features. DF-CHI feature combination extraction method is created based on the fundamental difference between the DF method and the Chi-square statistics method in terms of low frequency words. We think when it is difficult to determine which hypothesis is more reasonable, these two methods can be combined. The DF method is first used to filter out the low frequency words. Then the Chi-square statistics method can be used to obtain a feature set that is more relevant to the classification results. This combination of two methods can complement each other. It can be used to extract entries with more categorized information, which can improve the performance of the classifier in theory.

**2.3 Classification technique**

**2.3.1 Support Value Machine（SVM）**

SVM was proposed by Vapnik V [19] in 2000, it is a relatively new machine learning technology that has been widely used in mode detection in many areas in recent years [20]. At the same time, this algorithm is also very effective and efficient for text classification problems. In Geometry, a two-value SVM classifier can be viewed as the feature space of hyperplanes, representing the positive and negative examples. The classification of hyper plane is the larger of the two kinds of boundary intervals [21]. Through the learning algorithm, SVM is looking for the best identification ability of the sample point set in the training sample called the Support Vectors. In the classification stage, SVM uses these support vectors to predict the class properties of the unknown class sample [22]. The SVM classifier is not related to the dimension of the feature space, therefore, in theory, it can solve simulation problems appropriately.

**2.3.2 Artificial Neural Network（ANN）**

Artificial Neural Network was proposed by Neurobiologist MeCulloch and young Mathematician Pitts. The ANN classifier can use a three-layer feed-forward B-P network, including the input layer, the output layer and the hidden layer. The input node of the network receives the characteristic value, the output node generates the class value, and the connection weight represents their dependent relationship. Neural networks can correct classification errors through inverse propagation training and improve the accuracy of the classifier. ANN has a wide range of self-purification, so it can realize the function of fuzzy reasoning, and it can also keep the high speed of operation in the case of a large amount of data load [23].

### 2.3.3 K-Nearest Neighbor（KNN）

K-Nearest Neighbor [24] is a traditional mode identification method, it is very common among search engines and also widely applied in text classification researches [9,10,12]. It has an outstanding performance in accuracy and recall. KNN selects K documents that are most similar to the new document by calculating the similarity between the new document and the known class of documents, and the label with the highest frequency among the K documents is the label of the new document.

## 2.4 Evaluation criteria

In order to evaluate the performance of different feature selection methods and classifiers, the most common performance evaluation method is adopted: The recall rate R (Recall) and the accuracy rate P (Precision) and the $F_1$ evaluation. For a particular category, the recall rate is defined as the ratio of the number of documents correctly classified and the total number of test documents, that is, the probability of the correct recognition of the class of samples. The accuracy rate is defined as the ratio of the number of documents and the number of documents to be classified, that is, the probability that the classifier will make the correct decision [25]. The $F_1$ calculation formula is as follows:

$$F_1 = \frac{2RP}{R+P} \tag{3}$$

## 2.5 Integrated learning model

By combining different feature extraction methods and classifiers, we can find a combination of multiple groups to achieve our expected performance. But when faced with new cases, which combination can be used to ensure the reliability of the classification results? How do we combine different combinations of decisions? How much weight should be given to the prediction of each combination? Based on the theory of Delphi Method, this paper uses integrated learning to solve these problems. Integrated learning refers to constructing a new model and using the prediction result of base learning machines that have met expectations as input after proper training, and finally outputting a prediction result with maximum probability through linear or non-linear calculations. Compared to simple models, integrated learning is more time-saving and its performance is more universal [22].

Assuming the number of base learning machine that have met expectations is n, each of which is represented as $D_i (where\ i = 1,2,\cdots,n)$. Each machine's prediction category is denoted as $T_i$, and $T_i \in \{0,1\}$. We can use the prediction results $T_i$ of n machines on a certain document as input and use their potential classification results as output to construct a B-P neural network to determine the weight of each base learning machine. And thus, constructing the integrated learning model of the automatic decision system. This model can be used to predict and classify unlabeled documentations. The integrated learning model is shown in figure 1.

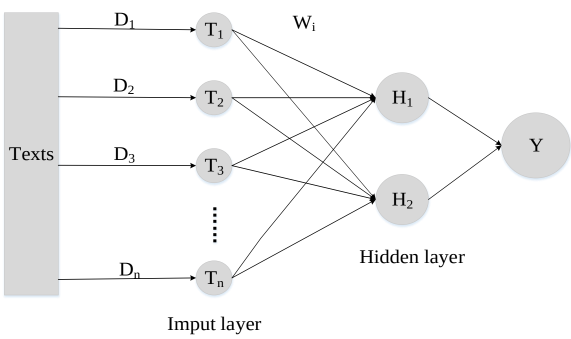


**Figure 1. Integrated learning model**

# 3. SYSTEM OVERALL DESIGN

Using the research method from the second part, the overall design of the court decision support system (CJ-DSS) in this paper is shown in Figure 2.

The system is mainly divided into two parts: first part uses unstructured text as input and forms structured document matrix through preprocessing; and then filters out the base learning machines that have met expectations by replacing feature extraction method and classifier, while outputting the decision obtained from the selected base learning machines; The second part uses base learning machines' classification results on the test set document as input and outputs the final decision on the test set document through integrated learning.

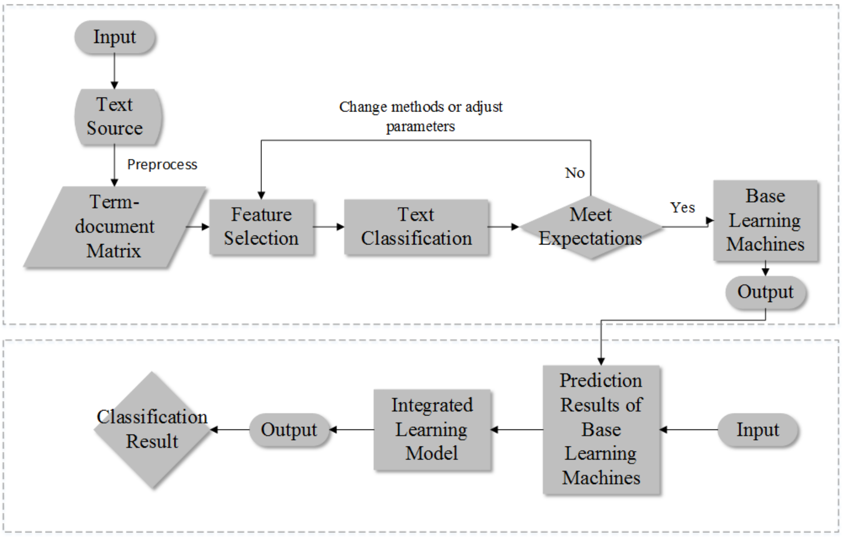


**Figure 2. CJ-DSS overall design**

## 4. EMPIRICAL ANALYSIS AND RESULTS

### 4.1 Data source and structure

This paper's data source is the "BeiDaFaBao" legal database. This paper used "medical malpractice" as the keyword and retrieved more than 300 court verdict and mediation documents from 2013. Due to the short format of mediation documents and its brief case explanations, they were eliminated from the study. The rest of the documents were trained and tested after preprocessing. The total number of database documents used in the experiment is 220. Among them, a total of 100 documents were selected as the training set for the base learning machines; 60 documents were selected as B-P neural network model training, the remaining 60 documents are used as a test set. In order to facilitate the performance comparison between integrated learning model and single model, the test set of new text data remains unchanged.

Regarding the rejected and non-rejected structural problem in the training set, through the statistical empirical distribution of rejected and non-rejected texts in 2011-2013 Chinese medical dispute cases, it was found that the ratio between rejected and non-rejected cases is approximately 1:4. In the experiment, 1:1 and 1:4 training set data ratio were tested. The results showed that the performance of all kinds of classifiers under 1:1 data structure was lower than that under 1:4 data structure. Therefore, we concluded that conforming to the actual empirical distribution ratio can guarantee better classification results. Therefore, we set the distribution ratio of rejected and non-rejected cases in the training set to 1:4. Since the classifier classifies new texts individually, the structure of the new text will not affect the classification results. The data structure is as shown in table 1.

**Table 1. Text data size and composition**

| Train set | Data size | Proportion | Rejected cases | Non-rejected cases |
|---|---|---|---|---|
| Single model | 100 | 1:4 | 20 | 80 |
| Integrated learning model | 60 | 1:4 | 12 | 48 |

### 4.2 Automatic judgment results of single and combination feature extraction methods

By adjusting the parameters of DF and CHI method, we selected feature sets under different conditions and classified them according to the different conditions. In order to make the text concise and easy to understand, a detailed description of the symbols that appears in the text is listed in Table 2 below.

**Table 2. Symbolic representation and description**

| Feature extraction methods | Examples | Symbolic representation | Description |
|---|---|---|---|
| DF | DF（Type=0）>5 | $D_0$>5 | If a term appears in five or more documents, then the term is selected as the feature word to compose the feature set. Where Type=0 denotes rejected texts, and Type=1 denotes non-rejected texts. |
| CHI | CHI>1 | C >1 | If the chi-square (which is used to measure the degree of correlation between features and class labels) of a term is greater than 1, then the term is selected as the feature word to compose the feature set. |
| DF-CHI | DF(Type=0)>2& DF(Type=1)>3&CHI>1 | $D_0$>2& $D_1$>3&C >1 | The term is selected if the term appears in two or more rejected documents and in three or more non-rejected documents, then if its chi-square is also greater than 1, the term is selected as feature word to compose the feature set. |

#### 4.2.1 Automatic decision result of single Chi-square feature extraction method

Figure 3 shows the performance of three classifiers under different Chi-Square values. As the Chi-square value increases, the number of words that passes the related texts successfully decreases. Since the support vector machine relies on support vector for fast classification, it is not sensitive to features' dimensions. Therefore, it maintained a relatively stable trend. The performance of artificial neural network has fluctuated significantly as the features value decreased. The performance of K nearest neighbor was poor, fluctuating around 0.700. It is also not sensitive to the number of features.

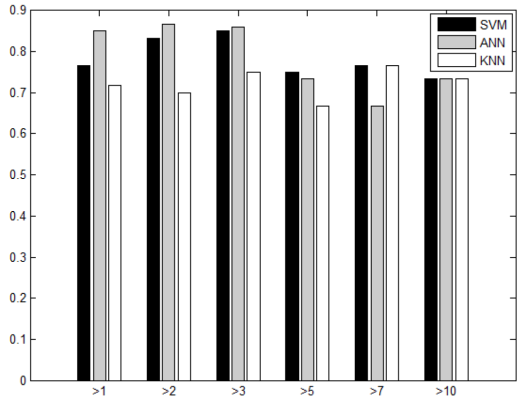


**Figure 3. $F_1$-value for Different CHI Parameter Setting**

#### 4.2.2 Automatic decision results of single DF and combination feature extraction methods

In order to represent the decision result difference between the DF method and DF-CHI combination method more directly, we rewrite the DF method parameters to coordinates in the following way: Take the combination of DF and SVM method as an example, we use $D_0$ for the X-axis, $D_1$ as the Y-axis and $F_1$ value as the Z-axis, as shown in Table 3.

The following three figures 4-1, 4-2 and 4-3 are the performances of SVM, ANN and KNN classifier, respectively. Red represents the single DF method, and green represents DF-CHI combination method. As shown in the diagram, the DF-CHI combination method performance improved differently according to different classifiers (Otherwise the top view will be full red): SVM improved about 45%, ANN improved about 80% and KNN improved about 40%.

**Table 3. the $F_1$–value of DF/SVM combination**

| Z: $F_1$-SVM/DF | X | $D_1>3$ | $D_1>5$ | $D_1>10$ | $D_1>15$ | $D_1>20$ | $D_1>30$ | $D_1>40$ | $D_1>50$ | $D_1>60$ | $D_1>70$ | $D_1=80$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Y | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
| $D_0>2$ | 1 | 0.767 | 0.783 | 0.783 | 0.783 | 0.850 | 0.750 | 0.800 | 0.750 | 0.800 | 0.833 | 0.767 |
| $D_0>5$ | 2 | 0.767 | 0.800 | 0.800 | 0.767 | 0.733 | 0.767 | 0.767 | 0.767 | 0.783 | 0.800 | 0.783 |
| $D_0>7$ | 3 | 0.767 | 0.783 | 0.783 | 0.783 | 0.800 | 0.717 | 0.750 | 0.733 | 0.717 | 0.717 | 0.717 |
| $D_0>10$ | 4 | 0.783 | 0.783 | 0.783 | 0.783 | 0.767 | 0.733 | 0.767 | 0.683 | 0.617 | 0.800 | 0.783 |
| $D_0>15$ | 5 | 0.783 | 0.800 | 0.767 | 0.783 | 0.767 | 0.733 | 0.733 | 0.750 | 0.717 | 0.800 | 0.800 |
| $D_0=20$ | 6 | 0.783 | 0.783 | 0.783 | 0.767 | 0.767 | 0.733 | 0.767 | 0.733 | 0.700 | 0.800 | 0.800 |

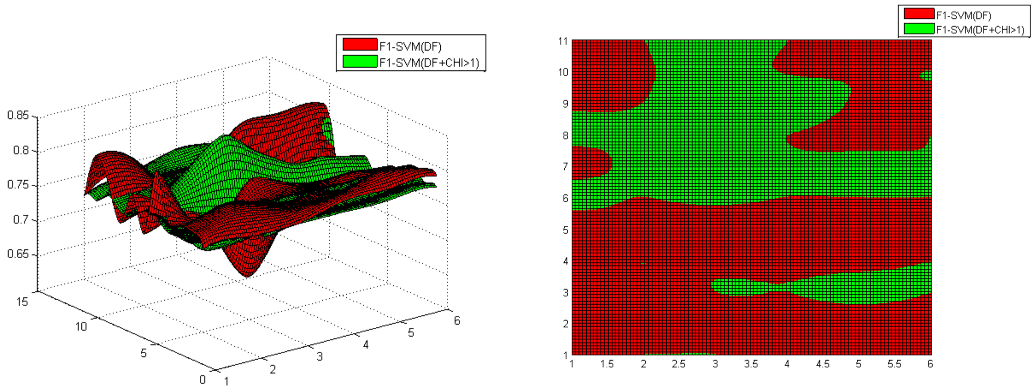


**Figure 4-1. Three-dimensional diagram and plan view of SVM(DF) & SVM(DF-CHI)**

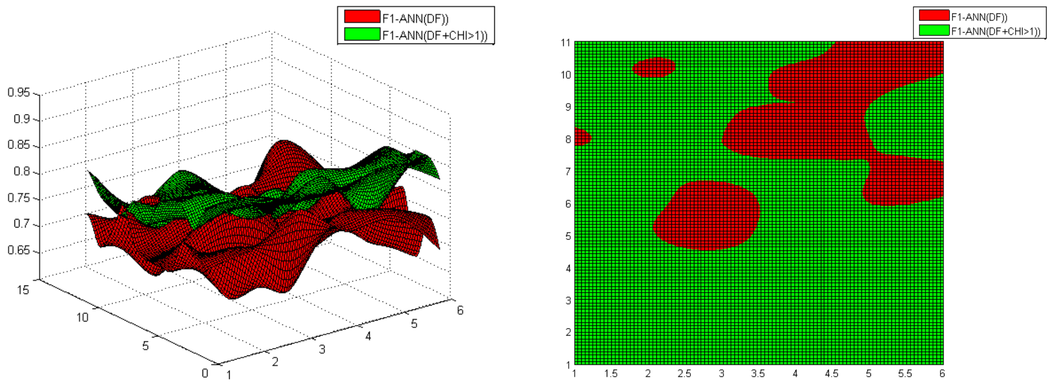


**Figure 4-2. Three-dimensional diagram and plan view of ANN(DF) & ANN (DF-CHI)**

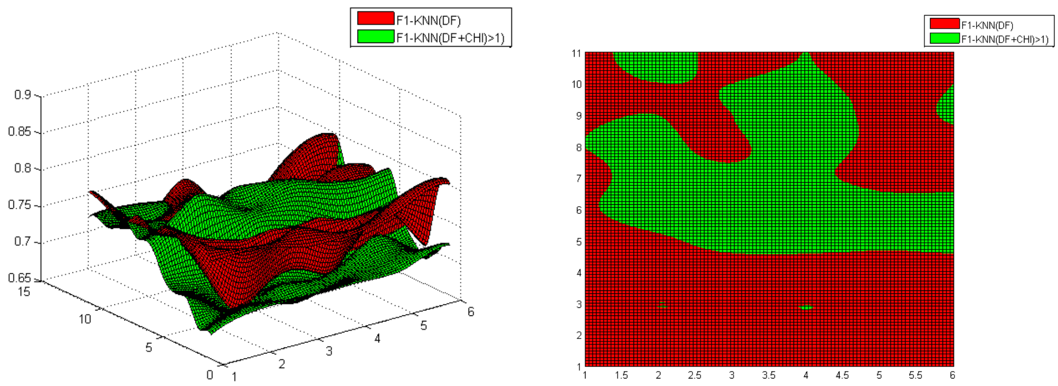


**Figure 4-3. Three-dimensional diagram and plan view of KNN(DF) & KNN(DF-CHI)**

### 4.3 Integrated learning of B-P neural network

We have improved the performance of each classifier in different degrees through the combination feature extraction method. This increased the number of base learning machines that met the expected classification result. We set the expected classifier performance to $F_1$>0.850. According to the above results, we selected the following combinations that met the expected performance as the base learning machines. Table 4 shows the full combination of classifiers with performance $F_1$>0.850. This constitutes the multiple base learning machine in this paper, see Table 4 for detailed explanations.

According to the B-P neural network model in 2.5, we used the test set prediction results of 18 base learning machines as input, rejected (0) and non-rejected (1) as output and used multiple B-P network structure, after repeated trial and error, 1000 iterations, learning rate of 0.1 and 5 hidden layer neurons, we obtained the final integrated prediction result. As shown in Figure 5, the $F_1$ of the integrated learning model reached 93.3%, effectively improving the performance of CJ-DSS system.

**Table 4. Detailed explanations of base learning machine**

| $D_i\,(i=1,2,...,18)$ | Feature Selection | Classifier | $F_1$ |
|---|---|---|---|
| $D_1$ | C>1 | ANN | 0.850 |
| $D_2$ | C>2 | ANN | 0.867 |
| $D_3$ | C>3 | SVM | 0.850 |
| $D_4$ | C>3 | ANN | 0.850 |
| $D_5$ | $D_0$>2 & $D_0$>20 | SVM | 0.850 |
| $D_6$ | $D_0$>2 & $D_0$>3 & C >1 | ANN | 0.883 |
| $D_7$ | $D_0$>2 & $D_0$>10 & C >1 | ANN | 0.867 |
| $D_8$ | $D_0$>2 & $D_0$>15 & C >1 | ANN | 0.867 |
| $D_9$ | $D_0$>2 & $D_0$>20 & C >1 | ANN | 0.850 |
| $D_{10}$ | $D_0$>2 & $D_0$=80 & C >1 | ANN | 0.850 |
| $D_{11}$ | $D_0$>5 & $D_0$>5 & C >1 | ANN | 0.850 |
| $D_{12}$ | $D_0$>5 & $D_0$>10 & C >1 | ANN | 0.867 |
| $D_{13}$ | $D_0$>5 & $D_0$>15 & C >1 | ANN | 0.850 |
| $D_{14}$ | $D_0$>7 & $D_0$>3 & C >1 | ANN | 0.850 |
| $D_{15}$ | $D_0$>10 & $D_0$>3 & C >1 | ANN | 0.850 |
| $D_{16}$ | $D_0$>10 & $D_0$>10 & C >1 | ANN | 0.850 |
| $D_{17}$ | $D_0$>15 & $D_0$>3 & C >1 | ANN | 0.867 |
| $D_{18}$ | $D_0$=20 & $D_0$>5 & C >1 | ANN | 0.850 |

Notes: According to the number of texts in training set, the value of $D_0$ and $D_1$ both are integer, meanwhile the domain of $D_0$ value is [1,20] and the domain of $D_1$ value is [1,80].

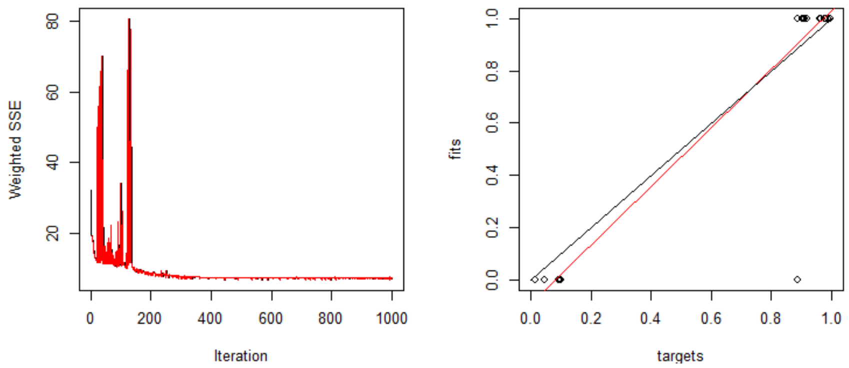


**Figure 5. Iterative result and Fitting chart**

## 5. CONCLUSIONS

This paper used China's medical malpractice litigation judicial documents as example and proposed a court decision support system (CJ-DSS) through text mining and automatic classification technology. This system is able to predict the judicial decision of new documents as rejected or non-rejected. Through different combinations of feature selection methods, classifier models, we used F1 value to evaluate their performance. At the same time, in order to improve the system's practical application ability, we used a combination of feature extraction method to improve the classification performance and chose integrated learning to combine the decision results of multiple classifiers in order to increase the system's consistency. Thus, we were able to build the CJ-DSS that was suitable for the Chinese court litigation documents.

At the same time, by combining with real cases, this study found that the combination feature extraction method can indeed improve the classifier's classification performance, especially for SVM, ANN and KNN classifiers. DH-CHI combined feature extraction method improved the performance of each classifier differently: SVM improved about 45%, ANN improved around 80% and KNN improved around 40%. In addition, the system classification performance became more consistent after integrated learning. The best performance reached 93.3%, which significantly increased system accuracy.

In previous studies, the accuracy of text classification system has been greatly influenced by the training set size: the larger the training set data, the better the performance. This paper was able to achieve the design of a high performance based on a small scale of text set, which has a reference value for constructing structured high-performance system based on a small sample training set in the future. More research should also be conducted for this area. At the same time, in real world applications, since the process of labelling documents is costly, therefore, the study and model construction for unlabeled text should be the focus of future research for data scientists.

## ACKNOWLEDGEMENT

This research was supported by the National Natural Science Foundation of China (Grant 71401093 and 71403158) and Shaanxi Science and Technology Department soft and major scientific and technological innovation project plan (2013(2)KC001).

## REFERENCES

[1] Dong M Y. Code Law, Case Law and China's Codification Road[J]. Study of Comparative Law, 1997, 11(4):1-31 in Chinese.

[2] Heng Li, Lihua Fan,Lei Gao. A Comparative Study on the Current Situation of Medical Injury Compensation Litigation [J]. China Health Law, 2013(1):24-28 in Chinese.

[3] Prevedello L M, Raja A S, Ip I K, et al. Does Clinical Decision Support Reduce Unwarranted Variation in Yield of CT Pulmonary Angiogram? [J]. American Journal of Medicine, 2013, 126(11):975-81.

[4] Park S H, Rha S W, Byun J S, et al. Performance evaluation of the machine learning algorithms used in inference mechanism of a medical decision support system [J]. The scientific world journal,2014,2014(7):137896-137896.

[5] O'Sullivan D, Fraccaro P, Carson E, et al. Decision time for clinical decision support systems[J]. Clinical Medicine, 2014, 14(4):338-41.

[6] Yinghong Wan, Jiang Li, Huaizu Li. Research on Virtual Cooperative Partner Selection Intelligent Decision Support System Framework [J]. System Engineering Theory and Practice, 2001, 21(12):60-65 in Chinese.

[7] Shufen Fang, Wenyuan Lv. Research on Intelligent Decision Support System for Equipment Maintenance Management [J]. System Engineering Theory and Practice, 2001, 21(12):53-59 in Chinese.

[8] Tseng Y H, Lin C J, Lin Y I. Text mining techniques for patent analysis[J]. Information Processing & Management,

2007, 43(5):1216-1247.

[9] Kim J H, Choi K S. Patent document categorization based on semantic structural information[J]. Information Processing & Management International Journal, 2007, 43(5):1200–1215.

[10] Pong Y H, Kwok C W, Lau Y K, et al. A comparative study of two automatic document classification methods in a library setting[J]. Journal of Information Science, 2008, 34(2):213-230.

[11] Coussement K, Poel D V D. Improving customer complaint management by automatic email classification using linguistic style features as predictors[J]. Decision Support Systems, 2008, 44(4):870-882.

[12] Al Qady M, Kandil A. Automatic Classification of Project Documents on the Basis of Text Content[J]. American Society of Civil Engineers, 2014.

[13] Salton, G., Yang, C., & Wang, A. (1975). A vector space model for automatic indexing. Communications of the ACM, 18(11), 613–620.

[14] Rocchio, J. J. (1971). Relevance feedback in information retrieval. In G. Salton (Ed.), The SMART retrieval system: Experiments in automatic document processing (pp. 313–323). Englewood cliffs, NJ: Practice-Hall.

[15] Salton, G., & Buckley, C. (1988). Term weighting approaches in automatic text retrieval. Information Processing and Management, 24,513–523.

[16] Yang, Y., & Pedersen, J. O. (1997). A comparative study on feature selection in text categorization. In Proceedings of ICML-97, 14th international conference on machine learning (pp. 412–420).

[17] Lee, C., & Lee, G. G. (2006). Information gain and divergence-based feature selection for machine learning-based text categorization. Information Processing and Management, 42, 155–165.

[18] DAI Liu-ling, HUANG He-yan, CHEN Zhao-xiong. A Comparative Study on Feature Selection in Chinese Text Categorization[J]. JOURNAL OF CHINESE INFORMATION PROCESSING, 2004, 18(1):26-32 in Chinese.

[19] Vapnik V. The Nature of Statistical Learning Theory. 2nd ed[M]. Springer Berlin, 2000.

[20] Burges C J C. A tutorial on support vector machines for pattern recognition[C]// Data Mining and Knowledge Discovery1998:121--167.

[21] Chen Xianyi, Zhu Qian. Principle of Text Mining[M]// Science Press, 2010 in Chinese.

[22] B Lantz. Machine Learning with R[M]. China Machine Press,2015.

[23] Liu Gang, Hu Siquan, Fan Zhihua, et al. An Application of ANN in Text Categorization[J]. Computer Engineering and Applications, 2003, 39:73-74 in Chinese.

[24] Dasarathy B V. Nearest neighbor (NN) norms: NN pattern classification techniques[J]. Los Alamitos IEEE Computer Society Press, 1991, (100).

[25] Sholom M. Weiss, Nitin I, Tong Z. Fundamentals of Predictive Text Mining[M], Springer, 2012.Bird R B, Stewart W E. (1960). Lightfoot E N. Transport Phenomena. New York: John Wiley & Sons Inc, 75.