第 34 卷第 6 期
2017 年 6 月

控 制 理 论 与 应 用
Control Theory & Applications

Vol. 34 No. 6
Jun. 2017

DOI: 10.7641/CTA.2017.60545

基于聚类和支持向量机的胃癌患者住院费用建模

周 涛, 陆惠玲†, 王文文, 王惠群

(宁夏医科大学 理学院, 宁夏 银川 750004)

摘要: 针对胃癌患者住院费用分类标签设定的复杂性以及传统费用建模算法的局限性, 本文提出了一种基于聚类和支持向量机的住院费用建模算法, 为胃癌患者住院费用的控制和预测提供方法基础. 搜集整理宁夏某三甲医院2009–2011年间1583例胃癌患者为样本, 采用K-means对总住院费用逐年聚类得到分类标签, 最后通过支持向量机对住院费用进行建模预测以及影响因素分析, 用分类准确率作为预测效果的评价指标. 实验结果表明胃癌患者住院费用呈逐年增加趋势, 其中以西药费为主, 占总费用的53.74%. 通过K-Means以年份对费用聚类比单纯以费用分布特征聚类的分类准确率提高了13.13%, 当核函数选用高斯核函数, 且惩罚因子$C=10$和核参数$\gamma=1$时建立的支持向量机模型最稳定, 分类准确率为92.11%. 实验结果表明根据年份聚类得到类别标签更合理, 结合聚类的SVM来预测住院费用更有效.

关键词: 胃癌; 住院费用; 支持向量机; 聚类; 分类标签

中图分类号: TP242 **文献标识码:** A

A new model for hospitalization expenses of Gastric cancer based on clustering and support vector machine

ZHOU Tao, LU Hui-ling†, WANG Wen-wen, WANG Hui-qun

(School of Science, Ningxia Medical University, Yinchuan Ningxia 750004, China)

Abstract: A new modeling method based on clustering and support vector machine (SVM) is proposed to simplify category labels complexity for the hospitalization expenses of gastric cancer patients and overcome the limitation of traditional cost modeling techniques, thereby providing some theoretical evidence to control and predict hospitalization expenses of gastric cancer patients. 1583 cases of gastric cancer patients in a certain tertiary general hospital of Ningxia from 2009 to 2011 were collected as samples. Total hospitalization expenses were clustered by years using K-means to obtain category labels, SVM was used to forecast and analyze the influencing factors of hospitalization expenses. The classification accuracy was used as indexes to evaluate the predicting effect. The experiment result show that hospitalization expenses of gastric cancer patients were increased year by year, and western drugs accounted for most of the hospital expenses(53.74%). The influencing factors of the cost of hospitalization were treatment outcome, surgery, admission situation, hospitalization time, ages and marital status, in which prognosis and surgery were the most important influences. The experimental results showed that the clustering accuracy of K-means by year was increased by 13.13% compared to only by distribution characteristics. The gauss kernel function-based SVM was the most stable model, with a classification accuracy rate of 92.11% when the penalty factor C and parameter γ were set to be 10 and 1, respectively. The method clustered by year was more reasonable to get category labels, and it was effective to combine clustering and SVM to forecast the hospitalization expenses.

Key words: gastric cancer; hospitalization expense; support vector machine; clustering; category label

1 Introduction

Gastric cancer is one of the most common gastrointestinal tract tumors in the People's Republic of China. According to the 2014 World Cancer Report[1], China has the world's largest number of new cases and deaths of gastric cancer. The morbidity of gastric cancer exhibits marked geographical variation, with high-risk areas in Japan, China, Eastern Europe and certain countries in Latin America. Ningxia is a high risk area of gastric cancer in China. The etiology of gastric cancer is not clear and is easy to relapse. The good news is gastric cancer is a disease of long duration thanks to the advances in technology, research and science. At the same time, however, as one of the chronic and long-lasting diseases, gastric cancer is costly and debilitating not only to the individuals, the families, but also to the medical insurance company, the community and the nation. In addition, diversification of diagnostic and therapeutic

Received 25 July 2016; accepted 13 December 2016.
†Corresponding author. E-mail: Lu_huiling@163.com; Tel.: +86 951-6980131.
Recommended by Associate Editor YU Zhu-liang.
Supported by National Natural Science Foundation of China (61561040), Natural Science Foundation of Ningxia (NZ16067) and Natural Science Foundation of Ningxia Education Department (NGY2016084).

technologies, in combination with the commonly seen unnecessary tests, procedures and treatment, further aggravate the economic burden on the patients. Therefore, it is of considerable importance to understand and model the structure of hospitalization expenses of patients with gastric cancer in order to make a reasonable forecast of anticipated costs of hospitalization of patients.

In general, modeling hospitalization expenses includes both data mining and statistics, and the mostly used statistical tools are multiple linear regression, Cox proportional hazard model and others. The requirements for these traditional methods are usually strict, including large enough data sample size, normal distribution and nonlinear relationship. As the distribution of data on hospitalization expenses is skewed and can be affected by many complicated factors which are related to each other, there are some limitations for traditional statistical methods to study hospitalization expenses. And the neural network method in data mining is too dependent on the sample and it is easy to fall into local minimum value and lead to the disadvantages of slow convergence speed, thereby limiting the development of the method in the hospitalization cost modeling. The advantage of support vector machine (SVM) is its ability to work with small sample data sets, use kernels to solve nonlinear problems, efficiently work in high dimensional space and avoid the local extremum problem. It is based on Vapnik-Chervonenkis (VC) theory and the principle of structural risk minimization to deal with complex multidimensional nonlinear data by nonlinear mapping. The final decision function is determined by only a few support vectors, which is independent of the sample dimension and distribution features, thus avoiding the "Curse of dimensionality" and "over learning". The global optimal solution is obtained by SVM, which is a powerful classification tool with the minimum classification error rate and the maximum generalization ability[2].

In the present study, a model of hospitalization expenses was built based on K-means clustering and SVM with the hospitalization expenses of gastric cancer patients as the samples, in order to explore the appropriate methods suitable for analyzing and evaluating the structure and characteristics of hospitalization expenses, and ultimately provide a basis for taking target-oriented measures to use the medical resources rationally and control the growth of hospitalization expenses.

2 Data sources and methods

2.1 Data sources

Data in our study were collected from the medical record of patients with gastric cancer in a tertiary hospital in Yinchuan city from 2009 to 2011. The total number of cases was 1583.

2.2 Methods

2.2.1 Data preprocessing

The quality problems of all the target samples were addressed by data cleaning by which the missing values were filled (the individual missing values for data were filled by the neighboring values), cases with incomplete information (such as unknown age, hospitalization date and payment information), duplicate cases and illogical cases (such as age over 120 years old, hospitalization time more than 365 days or less than 1 day) were excluded from the data set. A total of 18 disqualified cases were excluded and the total number of valid cases was 1583, representing 98.88% of total samples collected. The basic information, clinical diagnosis and other properties of the patients were dispersed and the discrete results are shown in Table 1.

Table 1 Study variables and discrete values

Serial No.	Variable	Value
1	Years	1=2009; 2=2010; 3=2011
2	Mode of payment	1=Social basic medical insurance; 2=Commercial insurance; 3=Own medical expense; 4=Socialized medicine; 5=Comprehensive arrangement for serious disease; 6=Others
3	Times	Time(actual value)
4	Sex	1=Male; 2=Female
5	Age	Years(actual value)
6	Marital status	1=Unmarried; 2=Married; 3=Divorce
7	Days	Days(actual value)
8	Admission situation	1=Dangerous; 2=Urgent; 3=General
9	Region of disease	1=Cardia; 2=Fundus ventriculi; 3=Corpora ventriculi; 5=Antrum pylori; 6=Pylorus; 7=Lesser curvature; 8=Greater curvature; 9=Trans boundary of the stomach; 10=Others
10	Result	1=Cure; 2=Improved; 3=Uncured; 4=Dead; 5=Others
11	Complication	1=Yes; 0=No
12	Hospital infection	1=Yes; 0=No
13	Surgery	1=Yes; 0=No

In order to better reflect the distribution characteristics of each attribute value, the main attribute values in the hospitalization expenses were shown in the form of the box diagram, as shown in Fig.1.

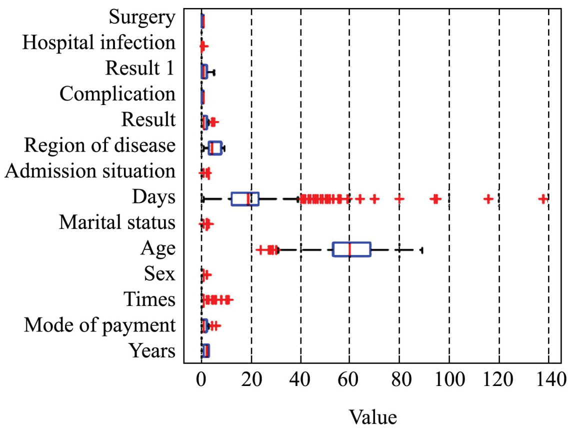

Fig. 1 The data attribute values of hospitalization expenses for gastric cancer patients

2.2.2 Basic theory

K-means is used to cluster the samples to K clusters. First, the K object is randomly selected, each object is represented by an average or center of a cluster. The remaining object is assigned to the nearest cluster according to its distance from the center of each cluster. Finally, the average value of each cluster is recalculated. This process is repeated until the criterion function converges.

SVM is a pattern recognition method developed from statistical learning theory based on the idea of structural risk minimization principle. In the case of ensuring classification accuracy, SVM can improve the generalization ability of the learning machine by maximizing the classification interval. The biggest advantage of SVM is that it overcomes the over learning and high dimension both of which lead to computational complexity and local extremum. A reliable classification model based on SVM is urgently needed for the study of hospitalization expenses of patients with gastric cancer.

At present, the commonly used kernel functions are shown in Table 2.

Table 2 Common kernel functions

Serial No.	Kernel functions	Formula
1	Lineal kernel	$k(x,y)=x\cdot y$
2	Polynomial kernel	$k(x,y)=[\gamma(x\cdot y)+C]^{\mathrm{d}}$
3	Gauss kernel	$k(x,y)=\exp(-\gamma\|x-y\|^2)$
4	Sigmoid	$k(x,y)=\tanh(\gamma(x\cdot y)+C)$

2.2.3 Process of SVM model

In this paper, the hospitalization expenses of patients with gastric cancer were used as samples, a modeling method based on clustering and SVM was proposed. All the experiments were performed on the MATLABR2012b platform. Firstly, the standard experimental data were obtained by preprocessing, including data cleaning, discretization, and normalization according to

$$\hat{x}=\frac{x-x_{\min}}{x_{\max}-x_{\min}}.$$

Secondly, two classes of classification labels (low as 1, high as 2) of hospitalization expenses were obtained by two kinds of clustering methods based on distribution characteristics and years, respectively. Thirdly, 80% of the data were used as training set for training SVM model, which was conducted using 13 dimensional features as the input variables and the one dimensional classification label as the output variable. The classification results of SVM model using four kernel functions were compared with different parameters. Finally, the remaining 20% of the data were used as testing set to validate SVM model. The classification method and the optimal SVM classification model were obtained by comparing the classification accuracy of the experimental results. The algorithm flow chart is shown in Fig.2.

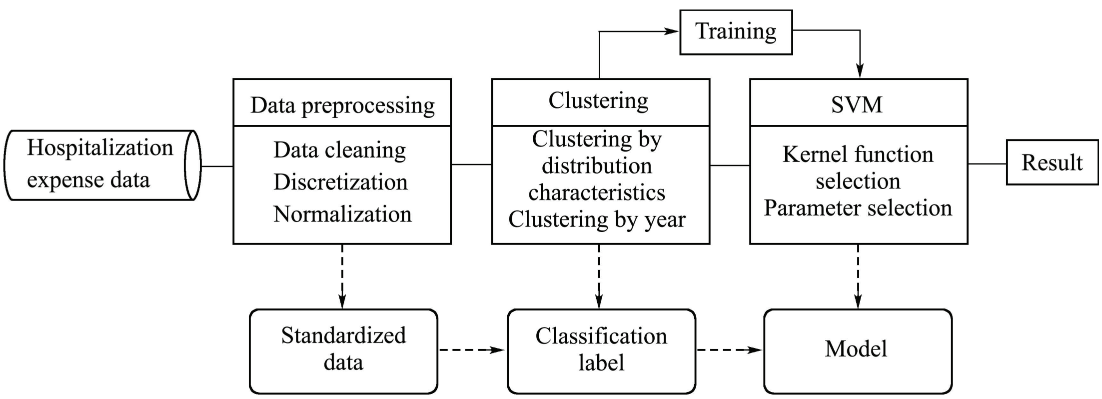

Fig. 2 Schematic diagram of hospitalization expense modeling by clustering and SVM

3 Result

3.1 General results

In our data set of 1583 gastric cancer patients, there were 1201 male patients and 382 female patients, accounting for 75.87% and 24.13% of the total cases, respectively. The fact that men were significantly more than women in in our data samples was consistent with the results reported in [3]. The hospitalization time varied from 1 to 138 days, with an average value of 19 days. The age of patients ranged from 24 to 89 years with an average age of 60 years. The majority of payment was received from social basic medical insurance (72.71% of all hospitalizations), followed by commercial insurance (17.06%), own medical expense (17.56%), socialized medicine (1.07%) and other (0.13%). The number of comprehensive arrangements for serious disease was 0.

3.2 The composition of hospitalization expenses

The average cost of three years of gastric cancer patients was 83653.33 RMB (Table 3). Extensive analyses of hospitalization expenses derived from three years of data indicated that among the composition of hospitalization expenses, western medicine expense was the number one, accounting for 53.74% of total expenses, followed by other expense, which mainly included the one-time cost of materials and others, accounting for 21.38% of total expenses. Comparing the growth of different costs for three years as shown in Fig.3 indicated that the per capita hospitalization expenses increased year by year. The growth rate of drug expense and other expense was more obvious, while the growth of nursing expenses and operating expenses was relatively stable compared with the others.

Table 3 The composition of hospitalization expenses in patients with gastric cancer (RMB)

Cost type	Years 2009	Years 2010	Years 2011	Sum	Composition/%
Number of cases	497	544	542	1583	/
Per capita hospitalization expenses	23330.22	28907.26	31415.85	83653.33	/
Bed fee	446.57	427.38	371.64	1245.59	1.49
Nursing fee	269.52	234.96	223.13	727.61	0.87
Western medicine fee	12217.25	15829.14	16905.13	44951.52	53.74
Chinese medicine fee	104.8	63.33	80.03	248.16	0.3
Laboratory fee	1419.04	1579.34	1667.19	4665.57	5.58
Examination fee	1064.8	2081	1254.89	4400.69	5.26
Operation fee	1784.35	1895.06	1873.96	5553.37	6.64
Inspection fee	1304.5	1330.32	1337.35	3972.17	4.75
Others	4716.92	5466.79	7702.61	17886.32	21.38

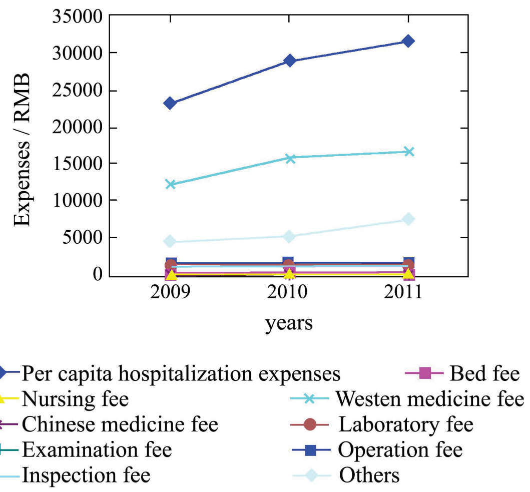

Fig. 3 Change in different costs of patients with gastric cancer during three years

3.3 The process and results of K-means

As shown in Fig.4, the distribution histogram of hospitalization expenses of patients with gastric cancer was skewed. This was further verified by a normal distribution Kolmogorov–Smirnov test in which a Z statistic was calculated by using nonparametric test method. From the test, it was found that the Kolmogorov Smirnov $Z = 3.09$ and $P < 0.01$, indicating that the data set of hospitalization expenses was not modeled by a normal but a skewed distribution.

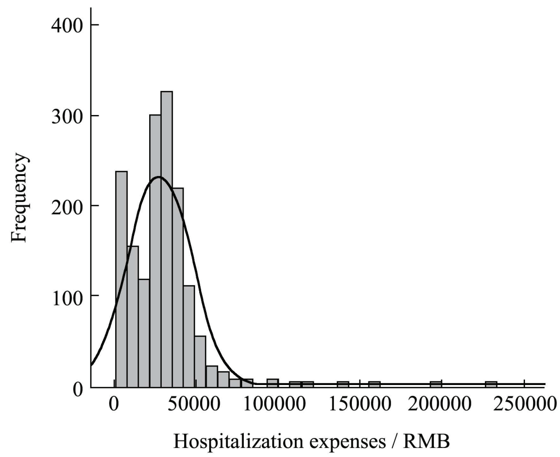

Fig. 4 Distribution histogram of total hospitalization expenses in patients with gastric cancer

It is a controversial issue whether the total cost of hospitalization should be divided into different categories. Since there are no scientific bases and specific classification criteria for the high and low level of hospitalization expenses, many scholars simply classify them only by their subjective experience and others use median method to divide hospitalization expenses into two categories. It was reported recently that the clustering method is better than the median method for hospitalization cost classification and that the demarcation point between the two types of data is completely unrelated to the median[5]. However, it will lead to unreasonable distribution if hospitalization expenses are divided into three and more categories by clustering. Therefore, in this paper, K-means method was used to classify the total hospitalization expenses of patients with gastric cancer for binary classification. Since the hospitalization expenses of patients with gastric cancer were time series data of three years, the effect of the year on the expenses of the patients should be considered. In literature [5] the category labels were obtained only by the distribution of hospitalization expenses. In this paper the year of hospitalization was registered as a variable and it was also taken into consideration when the composition of hospitalization expenses was analyzed by the clustering method and the rationality of two kinds of cluster methods were compared. The results of clustering are shown in Tables 4 and 5.

The total hospitalization expenses were clustered by year into two categories, as shown in Table 5.

As shown in Table 5, the clustering center of hospitalization cost was increasing by years, so did the high and low level of hospitalization expenses, indicating that the year had a great impact on the cost of hospitalization. In accordance with the statistical classification standard shown in Table 5, if the time factor was ignored and only the total hospitalization expenses were clustered, the overall dividing line separating high cost from low cost would be shifted to a higher level, resulting in more misclassification samples every year, especially in 2009. These results again demonstrated that it is reasonable and necessary to cluster the hospitalization expenses by year.

Table 4 Results of clustering based on the distribution characteristics of the cost of total hospitalization expenses of patients with gastric cancer

Classification	Cases	Clustering center/RMB	Range of total expenses/RMB
First category	910	16681.49	210~30577
Second category	673	43338.38	30600~234726

Table 5 Results of clustering by year for the cost of total hospitalization expenses of patients with gastric cancer

Year	Classification	Cases	Clustering center/RMB	Range of total expenses/RMB	No. of misclassification samples
2009	First category	184	9462.03	275~20187	207
	Second category	313	31478.84	20547~234725	
2010	First category	168	9511.98	120~23512	125
	Second category	326	37573.31	23640~164136	
2011	First category	187	10835.39	807~26064	39
	Second category	355	42256.94	26841~150238	

3.4 The process and results of support vector machine modeling

The support vector machine constructed by Gauss kernel function only has two user determined parameters, the kennel function parameter γ and the error penalty parameter C. Here γ significantly influences the classification accuracy, and C controls the complexity and approximate error of SVM. Therefore, the parameter settings are of great significance[6]. In order to get a robust SVM, the first thing first is to choose an appropriate kernel function for nonlinear conversion, then configure some core kernel parameters and the penalty factor to obtain an appropriate feature space in order that an optimal ratio of the confidence intervals to the empirical risk can be obtained for SVM model. Therefore, it is necessary to select proper parameter values in order to get a perfect performance of SVM.

In this paper, we built a SVM model with four common kernel functions and chose the values of penalty factor C and kernel parameter γ empirically. The classification labels used as output variables were clustered by two kinds of methods. Of the total sample data, 80% of them were selected as the training set and 20% the testing set and the accuracy of the two clustering methods of classification labels were compared by using different kernel functions and parameter settings of the SVM. The classification results of SVM with different values of γ and C are given in Table 6.

Table 6 The classification results of SVM with different values of C and γ/%

Kernel functions*	C	γ	Clustering by distribution characteristics		Clustering by year	
			Training accuracy	Testing accuracy	Training accuracy	Testing accuracy
1-7 1	10	/	79.91	76.28	87.28	89.27
2	10	1	87.60	77.92	92.02	88.12
	1	10	92.73	76.03	95.42	84.23
3	10	1	88.15	77.92	92.34	92.11
	1	10	92.84	76.39	93.21	87.38
4	10	1	43.60	41.01	70.46	67.82
	1	10	57.50	57.41	65.96	65.93
1-7 Average accuracy	—	—	77.49	68.99	85.24	82.12

* Kernel functions 1 to 4 are listed in Table 2.

As shown in Table 6, SVM with different kernel functions and parameter settings achieved different training accuracy, testing accuracy and the number of support vectors. The kennel function parameter γ and the error penalty parameter C significantly influenced the complexity and classification performance of SVM. The results showed that different kernel functions had a great influence on the performance of SVM. For example, the classification results based on Gauss kernel function were the best, and the Sigmoid kernel function were the worst. At the same time, when the penalty factor $C = 10$ and kernel parameter $\gamma = 1$, an optimal overall classification accuracy was obtained. The classification accuracy of four kernel functions could be seen more intuitively in the histogram as shown in Fig.5.

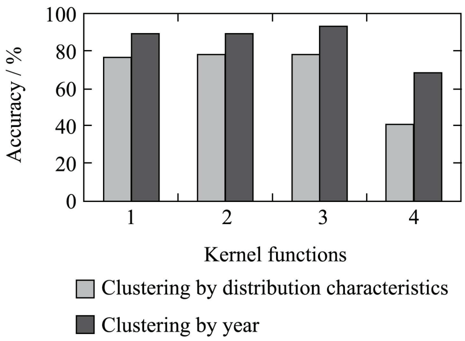

Fig. 5 Classification accuracy of support vector machine with different clustering methods

3.5 The influencing factors of hospitalization expenses

As described above, our experimental results indicated that the classification labels of hospitalization expenses obtained by clustering by year in combination with SVM using Gaussian radial basis kernel function produced the best results. Therefore, SVM model based on Gaussian radial basis kernel function was further employed to analyze the impact factors of hospital costs.

Hospital cost data were substituted into the trained SVM model, the classification labels were used as output variables and all influencing factors were used as input variables. With the Gaussian radial basis function and parameters $C = 10$ and $\gamma = 1$, the main influencing factors were obtained as shown in Fig.6.

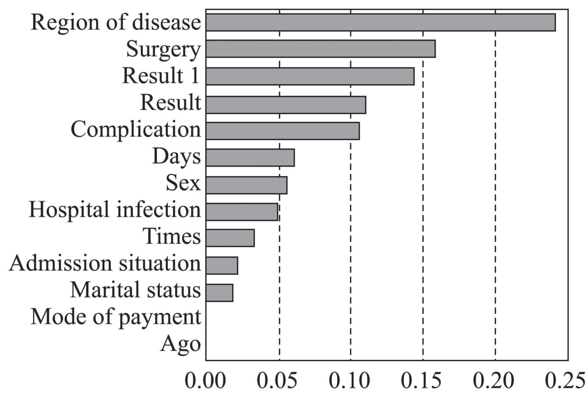

Fig. 6 Influencing factors of hospitalization expenses in patients with gastric cancer

The importance of each variable was given by SVM and the descending order of the factors affecting the hospitalization expenses of patients with gastric cancer were as follows: lesion site (0.242) $>$ surgery (0.16) $>$ treatment outcome 1 (other disease) (0.144) $>$ treatment outcome (0.111) $>$ complication (0.105) $>$ hospitalization time (days) (0.061) $>$ sex (0.056) $>$ hospital infection (0.049) $>$ number of hospitalization (0.033) $>$ admission situation (0.021) $>$ marital status (0.018). Therefore, the three main factors affecting the hospitalization expenses were lesion site, surgery and result1 (other diseases), while mode of payment and age almost had no impact on hospitalization expenses.

4 Discussion

4.1 Basic characteristics of the disease and the impact of the year on the hospitalization expenses

Through the analysis of the cost of hospitalization, it was found that the cost was increasing year by year. In 2010 and 2011, the average hospitalization expenses increased by 23.90% and 8.68% over the previous year, respectively. Gastric cancer is an expensive disease. On the one hand, most of the gastric cancer patients are the elderly and the recovery and rehabilitation of the patients are relatively poor with long-term complications. On the other hand, the improvement of medical technology, especially the new technology and materials leads to an increase in medical cost of patients with gastric cancer. However, based on the hospital cost distribution (Fig.3), most patients spent no more than 5000 RMB. One reason was that some patients gave up treatment because of the seriousness of the disease and unaffordable high medical expenses. Therefore, we should increase the actual compensation ratio of malignant tumor and promote the implementation of the serious illness medical insurance, to improve the reimbursement rate and the level of protection of patients with gastric cancer. Influencing factor analysis of hospitalization expenses demonstrated that in addition to the disease itself, treatment outcomes and surgery were the main factors affecting the cost of hospitalization, Therefore, the reasonable treatment method is the key to control the cost of hospitalization. In addition, in order to effectively reduce the hospitalization expenses of patients, other measures should also be taken, including, but not limit to improvement of the level of medical technology and professional ethics of medical personnel, control over dependence on medical equipment, implementation of standardized processes of diagnosis and treatment of gastric cancer and shortening hospitalization time.

4.2 The problem of the high expense ratio in composition of hospitalization expenses

The recently accumulated studies show that western medicine cost accounts for more than 50% while technical fees accounts for less than 25% of the average income of Chinese general hospitals[7]. In this study, the western medicine cost itself accounted for up to 53.74% of the total hospitalization expenses, while nursing expenses, reflecting the labor value of medical personnel, accounted only for 0.87%, and the diagnostic fees accounted for 5.26%, indicating that the "drug-maintaining-medicine" is still a very serious problem in China. These results are consistent with those reported by others[8–9]. At the same time, other expenses were the second-largest expenses, but the expenditure pattern is not clear. Taken together, hospitals should strengthen drug management, strictly enforce the essential drug-list and speed up the implementation of the medical cost transparency, and eventually implement the separation of dispensing from prescription to let patients rest assured to see doctors.

4.3 The modeling method based on clustering and support vector machine

The level of hospitalization expenses is influenced by many factors. Using K-means to classify the cost is a classification method completely based on statistics, which eliminates the jamming signal and makes the results more reliable. Because the hospitalization expenses showed a trend of increasing year by year, the internal factors such as price increases and other factors were ignored by clustering sample data from three years. By doing so, however, the accuracy of the classification was affected. Therefore, it is reasonable to cluster the data set by year. Indeed, the experiment showed that the results of hospitalization expenses clustered by year were much better than those clustered by distribution characteristics. It should be noted that the appropriate data preprocessing method was also very important for the prediction of the final model since it could provide reference for the classification of hospitalization expenses.

SVM shows a great advantage in the modeling of the hospitalization expenses, which does not require of the data set with large sample size and is very effective to deal with complex high dimensional data with strong generalization ability and learning ability of the classification model. The experimental results showed that the effectiveness of the RBF kernel function is the best among different kernel functions. With the empirical selection of the parameters $C = 10$ and $\gamma = 1$, a high classification accuracy of up to 92.11% was obtained with the SVM. In conclusion, this study demonstrates the validity and applicability of SVM to conduct classification and prediction of hospitalization expenses. It has a strong ability to promote and develop space. This study can effectively improve the prediction accuracy of SVM and provide a new idea for data preprocessing and modeling of hospitalization expenses by combining clustering and SVM.

References:

[1] Cancer[Online]. available: http: //www.who.int/mediacentre/factsheets/fs297/en/2015,02.

[2] SONG X, WANG J K, LI Q M, et al. Robust least squares constant modulus algorithm to signal steering vector mismatches [J]. *Wireless Personal Communications*, 2013, 68(1): 79 – 94.

[3] AIZIMAITI R, AIKEREMU Y, PAERHATI S, et al. Analysis of survival rate of different parts of gastric cancer [J]. *Modern Preventive Medicine*, 2010, 37(20): 3957 – 3960.

(艾孜买提·热合木吐拉, 艾克热木·玉素甫, 帕尔哈提·沙依木, 等. 胃癌不同部位可根治性手术的生存率分析 [J]. 现代预防医学, 2010, 37(20): 3957 – 3960.)

[4] HSU C W, CHANG C C, LIN C J. A practical guide to support vector classification, 2009 [J/OL]. Software available at http: //www.csie.ntu.edu.tw/ cjlin.

[5] SHEN Pei . Influencing factors and control strategies of hepatitis a medical cost based on data mining [D]. Wuhan: Huazhong University of Science and Technology, 2012.

(沈培. 基于数据挖掘的甲肝医疗费用影响因素与控制策略研究 [D]. 武汉: 华中科技大学, 2012.)

[6] ZHANG Yunqing, ZHANG Peilin. Machine training and parameter settings with social emotional optimization algorithm for support vector machine [J]. *Pattern Recognition Letters*, 2015, 54: 36 – 42

[7] DENG Caiyun. Analysis of influencing factors and hospitalization expenses in patients with injury and poisoning [J]. *Chinese Journal of Health Statistics*, 2012, 29(5): 732 – 733.

(邓彩云. 损伤与中毒病人的住院费用及其影响因素分析 [J]. 中国卫生统计, 2012, 29(5): 732 – 733.)

[8] WANG Anjue, JIANG Qicheng. Factor analysis on hospitalization expenses of inpatient intownship hospital [J]. *Chinese Journal of Health Statistics*, 2014, 31(2): 288 – 290.

(王安珏, 江启成. 某乡镇卫生院住院患者的住院费用因子分析 [J]. 中国卫生统计, 2014, 31(2): 288 – 290.)

[9] FAN Yuzheng, LÜ Wei, WANG Yong, et al. Study on hospital expenditure and influence factors of provincial general [J]. *Chinese Journal of Health Statistics*, 2014, 31(1): 61 – 63.

(范玉铮, 吕炜, 王勇, 等. 广西省级综合医院住院费用现状及影响因素分析 [J]. 中国卫生统计, 2014, 31(1): 61 – 63.)

作者简介:

周　涛 (1977–), 男, 博士, 教授, 硕士生导师, 目前研究方向为基于影像的计算机辅助诊断、医学大数据分析与处理、智能算法等, E-mail: zhoutaonxmu@126.com;

陆惠玲 (1976–), 女, 硕士, 副教授, 目前研究方向为医学大数据分析、医学图像分析与处理等, E-mail: lu_huiling@163.com;

王文文 (1990–), 女, 硕士研究生, 目前研究方向为医学大数据分析、医学图像分析与处理等, E-mail: 1043359441@qq.com;

王惠群 (1990–), 女, 硕士研究生, 目前研究方向为医学大数据分析、医学图像分析与处理等, E-mail: 782281527@qq.com.