A Distributed Procurement Cost Control Scheme of Medical Materials for Regional Medical Consortiums

YunXia Feng[1], LongKai Chai[2], Xu Li[3], ShuJun Zhang[4], Bo Song[5]

cloudy_feng@163.com[1], 1243707486@qq.com[2], lixv117@163.com[3] ,
lindazsj@163.com[4], 13789888999@163.com[5]

College of information science and technology, Qingdao university of science and
technology[1,2,4,5], College of Electromechanical Engineering, Qingdao University of
Science & Technology[3], No. 99 songling road, LaoShan district, Qingdao[1,2,3,4,5]

Abstract. Medical materials are the base of hospitals for carring out medical activities. Related research shows that the procurement cost of medical materials accounts for 20%-43.5% of the total cost of the hospital[1-2]. However, the hospital's material procurement plan is still based on the experience of the planner.This procurement method is prone to overpurchase. Overpurchasing easily leads to backlog of inventory, which will seriously occupy or even waste hospital costs. In this paper, we present a precise medical material demand forecast and procurement cost control scheme, called DMPCCS (Distributed Material Procurement Cost Control Scheme) based on distributed data mining. DMPCCS assumes that the regional medical data is stored and managed on a distributed big data platform, such as Hadoop. DMPCCS uses an improved distributed ARIMA (Autoregressive Integrated Moving Average) model to predict the expected demand of each medical material for every medical institution in the region. Based on the predicted results, DMPCCS makes a corresponding procurement plan.We evaluate performance of the proposed scheme use real data from three hospitals. Results show that the proposed scheme greatly reduces the purchasing redundancy simultaneously and avoids the purchasing inefficiency. DMPCCS can be used as a reference for the purchasing plan of the medical institutions to realize more reasonable control on the purchasing cost and the cost of warehouse management.

Keywords: Material Demand Forecast; Procurement Cost Control; Hadoop; ARIMA

1   Introduction

Under the environment of increasingly fierce competition in medical service, cost accounting and control should be strengthened to reduce the purchase cost of medical institutions and enhance the competitiveness of hospitals. In addition, in the context of regional health information connectivity, medical groups need to invest a lot of money to realize the integration and sharing of information resources. Hence, in a certain period, the hospital spending will increase, the income will reduce, hospital financial will face a serious situation.

On the other hand,the hospital consumables procurement mainly depends on experience of procurement staffs to make material procurement plan. And because the hospital is worried about the shortage of inventory, buyers usually purchase between 1.2 and 1.5 times as planned, so as to avoid the shortage of inventory. But on the other hand, this way leads to overpurchasing, overstocking and even the expiration of consumables.Therefore, to reduce the procurement cost of medical consumables from the overall level of the hospital is particularly prominent in the cost control of the whole hospital.

In the 1960s, based on the concept of JIT, the hospital adopted the method of stocking goods with high value stock to achieve the goal of material control of "zero inventory".The supplier keeps the hospital's high-value supplies that may be used by the hospital next month to be kept in the hospital, and the hospital will pay the supplier after the hospital runs out. This stocking method has basically realized the management of "zero inventory" of high value materials and reduced the purchase cost by 30% -50%. However, for ordinary materials not supplied by suppliers, their purchases and actual usage still have large discrepancies.

Today, Information technology play a central role in the medical service system. After more than 30 years of research and development, computer information technology has become an important support of modern medical service system.Computer information technology application in the field of medicine, has become a new edge in the modern medical science, called the medical informatics.Hospital information system (HIS) is the foundation of the discipline. In China, HIS has become an essential part of hospital information construction and is one of the important symbols of hospital modernization.Most of the hospitals in China have established HIS. Of course,the HIS plays a certain role in the cost control of hospital.For example,the hospital can realize the calculation, restriction and supervision of the consumables of the departments through the HIS.Although this is convenient for cost accounting to some extent, but, there is no reasonable budget for the purchase quantity and purchase cost of materials before purchasing.Confined to the cost accounting, the cost control work is placed behind the procurement activities, so the cost has lost control significance. It is a variety of lagging control. Hospitals still assume ... ces of cost.If the cost control of the hospital can not be adjusted in ... be effectively carried out. As a result, problems such as inventory ... cient inventory still plague the hospital's warehouse management. The

goal of material control of "zero inventory" did not really achieve.

At present, big data technology has greatly promoted the information development of medical institutions.The development of new technologies such as data mining and machine learning and its successful application in other fields[3-8] have provided a possible solution to the realization of the target of "zero inventory". In addition,With the development of the construction of regional medical and health information platform, it has solved the current situation that a large number of medical units in a medical group store large numbers of data fragmented and do not communicate with each other.So this can facilitate the application of big data technologies.

Based on the above situation, this paper proposes a precise medical material demand forecast and procurement cost control scheme based on distributed data mining(DMPCCS). The scheme can provide reference for the hospital to make procurement plans and help hospital to solve the problem of over-procurement caused by over-estimation and avoid the problem of insufficient inventory.By those,the hospital can reasonably reduce the purchase quantity and effectively save the purchase cost.

2   Premises

2.1 Regional medical and health information platform

The regional health information platform is a data exchange and sharing platform that connects the basic business information systems of various agencies within the planning area. The regional health information platform is the basis and carrier for effective information integration among various information systems in the region, and it is an integrated service platform integrating multiple subsystems.

One of the objectives of the construction of regional health information platform is to realize interoperability of information of medical institutions in the region. Regional health information platform can collect information and exchange information for medical and health service institutions,which enable medical information to share with the sub platforms of public health information system platform, and achieve information transmission and sharing between systems. The construction of regional health information platform has solved the status quo of information isolated islands formed by the independence of each system , so as to achieve data sharing.

2.2 The differences between the traditional ARIMA model and the improved ARIMA model

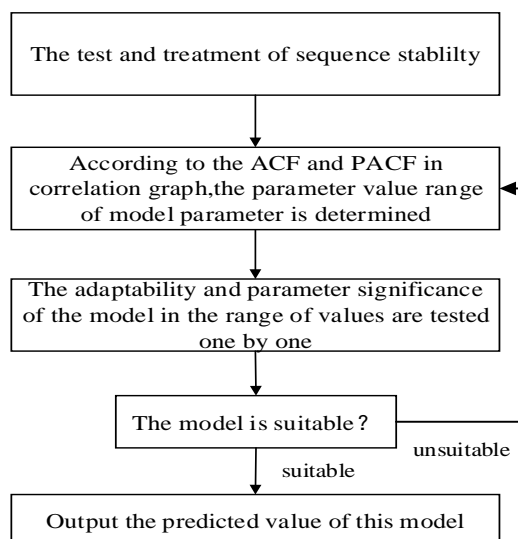The traditional ARIMA modeling process is shown in Fig.1.

Fig. 1. Traditional ARIMA Modeling Flowchart

The ARIMA (p, d, q) model is developed from the ARMA (p, q) model. The idea of the ARMA model is to fit higher order models by gradually increasing the order of the model. The model order stops increasing when the remaining residual variance no longer significantly decreases. However, since the ARMA(p,q) model is suitable for stationary time series.But most of the time series are nonstationary in practical problems. Therefore, the ARIMA(p,d,q) model is introduced. The ARIMA (p, d, q) model is called the Autoregressive Integrated Moving Average Model and is a time series prediction method proposed by Box and Jenkins in the early 1970s[9], where AR is an autoregressive model, p is an autoregressive item; MA is a moving average model, q is the moving average number of items, and d is the number of differences made when the time order is stationary.

The basic idea of the ARIMA model is that the data sequence formed by the prediction of the object over time is regarded as a random sequence, and a certain mathematical model is used to approximate the sequence. Once the model is identified, it can predict future values from the past and the present values of time series.

Although the ARIMA model is well-suited for time series predictions, hospital supplies procurement projections differ from those in other areas[10-18]. Because the particularity of the medical unit to save the sick, it is strictly forbidden to appear the situation of insufficient stock of materials. If the predicted medical supplies fluctuate around the actual consumption, which may lead to insufficient stock.

Based on this status quo, the medical supplies procurement forecasting model based on the improved ARIMA model proposed in this paper can reasonably solve the above problems. Firstly, by the ARIMA model,we can predict the actual consumption and obtain the forecasting values of consumption ,then the redundant values are calculated from the difference between the forecasted values and the actual consumption.If there are negative values in the redundant values, the redundant values are used as sample to obtain the forecasting value of the redundant .The absolute value of the minimum negative value in the forecasting value of the redundant is added to the prediction value of the consumption, so as to correct the prediction value of the consumption and avoid shortage of inventory.

The modeling process of the medical supplies procurement forecasting model based on the ARIMA model is shown in Fig.2.
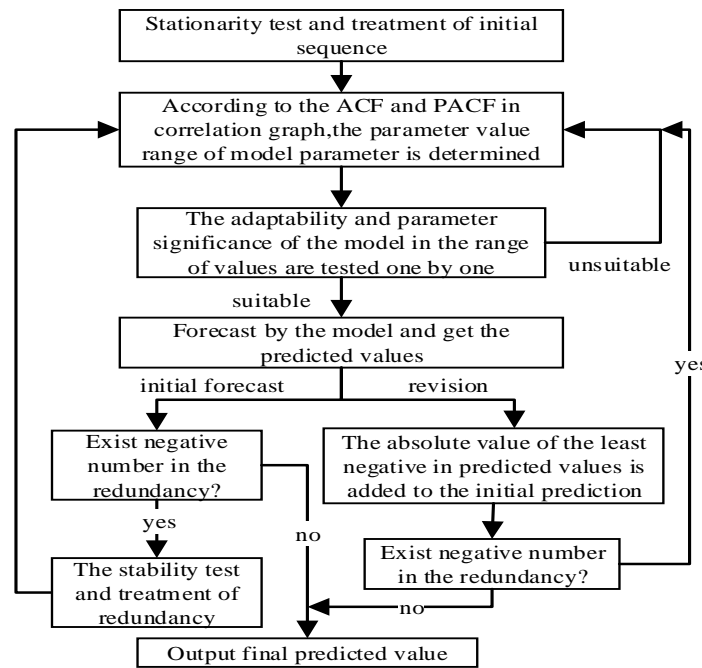

Fig. 2. Flowchart of the estimated model of material consumption based on ARIMA

2.3 Overview of the proposed scheme

In this paper, we study the problem of procurement cost control in Regional Medical Consortium. We present a precise medical material demand forecast and procurement cost control scheme(DMPCCS) based on distributed data mining. DMPCCS assume that the regional medical data is stored and managed on a distributed big data platform, such as Hadoop. DMPCCS uses an improved distributed ARIMA model to predict the expected demand of each medical material for every medical institution in the region. Based on the predicted results, DMPCCS makes a corresponding procurement plan. DMPCCS achieves precise demond of every medical material by improved ARIMA, based on which, make the corresponding procurement plan.

The system selected a server deployment as a controller. After the data is divided into blocks, the Map stage workstations are responsible for sorting the data by the type of materials and months used, and outputting the key pairs of the format of <material name, usage month, usage> to the local disk. After the controller reads the storage locations on the disk, the controller sends the location information to the workstation at the Reduce stage. After these phases are grouped by the reduce stage and sorted by the date, ARIMA model is used to forecast the usage of different materials in each month and output the forecast result. Finally, the predicted values are compared with the actual purchasing data by the average number of redundancies per month and the average costs of redundancies to determine whether the scheme can replace the artificial prediction.

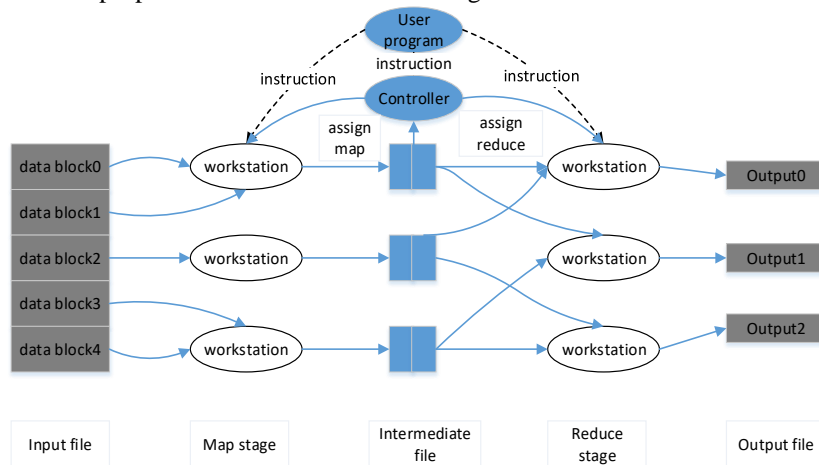The flowchart of the proposed resolution is shown in Fig.3.

Fig.3. DMPCCS program flow chart

3 Details of the proposed procurement cost control scheme

To facilitate the modeling process and experimental results of elaborating, this article only lists four materials' forecasting model with the experimental results.They are oxygen tube, insulin needle, anaesthetic mask, insulin pump infusion line. Next, we use oxygen tube data as an example to set up the medical supplies procurement forecasting model to illustrate the modeling process.

3.1 Stationarity test of time series

The stability of the sequence means that the fitting curve obtained through the sample time series can still "inertially" continue for a period of time in the future.If the data is non-stationary, the shape of the sample fitting curve does not have the characteristic of "inertia".That is, the curve fitting based on the sample time series to be obtained in the future will be different from the current sample fitting curve.From this we can see that the smoothness of the time series is the basic assumption of classical regression analysis.Only predictions based on stationary time series are valid. If the data is not stable, the "consistency" requirement, which is the basis for statistical inference under large samples, is destroyed and the prediction based on non-stationary time series fails.

The identification of sequence stationarity can be identified by sequence timing diagrams or Augmented Dickey-Fuller (ADF) test.

Fig.4 shows the time series of oxygen tube from January 2005 to July 2017.

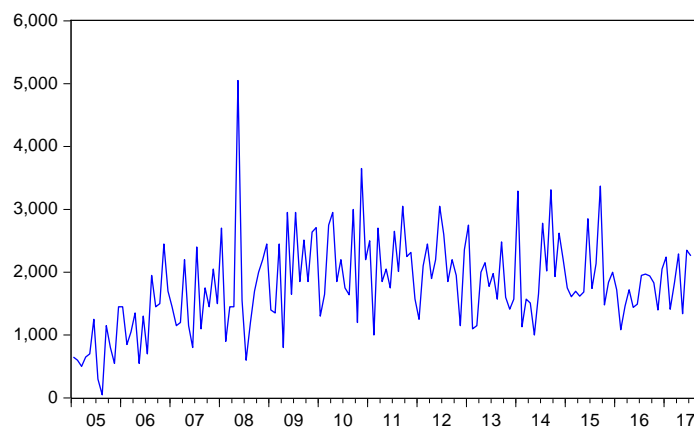**Fig. 4.** The time series of hospital oxygen tube consumption from January 2005 to July 2017

It can be seen from Fig.4 that the sequence is gradually increasing trend. So,the initial judgment is that the sequence is non-stationary. In order to further determine the stability of the sequence, it needs to be accurately judged by ADF test.

The results of ADF test for the sequence are shown in Table 1.

Table 1. Results of ADF test for oxygen tube comsumption

Variable	significance level	t-Statistic	Prob.*
ADF test statistic		-0.350309	0.5573
Test critical values:	1% level	-2.580897	
Test critical values:	5% level	-1.943027	
Test critical values:	10% level	-1.615260	

In table 1,the first column is the variable name, the second column is the significance level, the third column is t statistics, and the fifth is the bilateral probability value of the t test. As can be seen from Table 1, the companion probability of ADF test statistic is 0.5573, which is greater than 5%. Therefore, the null hypothesis of unit root is accepted and the sequence is non-stationary. Therefore, we need to take the first-order difference method to stabilize the sequence.

The results of ADF test for the first-order differential sequence are shown in Table 2.

Table 2. Results of the ADF test for the first-order difference sequence of oxygen tube consumption

Variable	significance level	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-9.764514	0.0000
Test critical values:	1% level	-2.580897	
Test critical values:	5% level	-1.943027	
Test critical values:	10% level	-1.615260	

As can be seen from Table 2, the companion probability of ADF test statistic is less than 5%, so the null hypothesis is not established and the unit root does not exist. That is, the first-order differential sequence is stationary.

3.2 Sequence randomness test

As we can see from the previous step, the sequence of the oxygen tube after the first-order difference is stationary. If a sequence is stable, then we will judge whether the data is white noise or not, and white noise has no significance. In this paper, the Ljung-Box test (also called Q test) is used to verify whether the first-order differential sequence is white noise or not, and if the p-value of the Q test is less than 5%, the sequence is a non-white noise sequence.

Fig.5 shows the autocorrelation and partial autocorrelation plots after the first-order differential of oxygen tube consumption time series from January 2005 to July 2017 in the hospital.
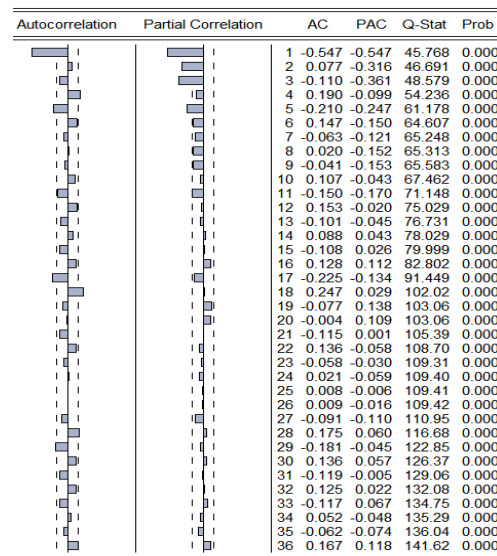
Fig. 5. First-order differential correlation diagram of the oxygen tube consumption sequence

In Fig.5,the left part is sequence autocorrelation and partial autocorrelation, and the right half includes five columns.The number of natural numbers K in the first column represents the delay order,AC is the autocorrelation coefficient,PAC is partial autocorrelation coefficient,Q-Stat represents the value of Q statistics for correlation test of sequences.Prob represents the P value of the concomitant probability of the Q statistics.,the P value is much less than 5%, so the sequences are not white noise sequences.

3.3 Parameter Selection for the Improved ARIMA Model

As can be seen from Fig.5 among the autocorrelation coefficients of the sequence, the lagged one has a significantly non-zero coefficient, while in the partial autocorrelation coefficient, the lagged 1,2,3 coefficients are obviously not zero. Therefore, the three parameters p, d, q of the ARIMA model can be chosen from (3,1,1), (2,1,1), (1,1,1).

ARIMA-based predictive models after the parameters have been estimated, the fitness of the model needs to be checked by a diagnostic test. There are two types of questions about the diagnostic test of the model, one is the adaptive test to examine whether the model is sufficient for the information extraction, and the other is the significance test of the model parameters to examine whether the model structure is the simplest.

An adaptive fitness test. The improved ARIMA model fitness test needs to be judged by white noise test on the model residual. If the residual is not a white noise series, it means that some features in the sequence are not extracted by the model. In this case, we need to further improve the model until the residual is a white noise sequence.

Fig.6, Fig.7 and Fig.8 show the autocorrelation and the partial correlation diagram of the residual when the model's parameters (p, d, q) are (3,1,1), (2,1,1), (1,1,1), respectively .

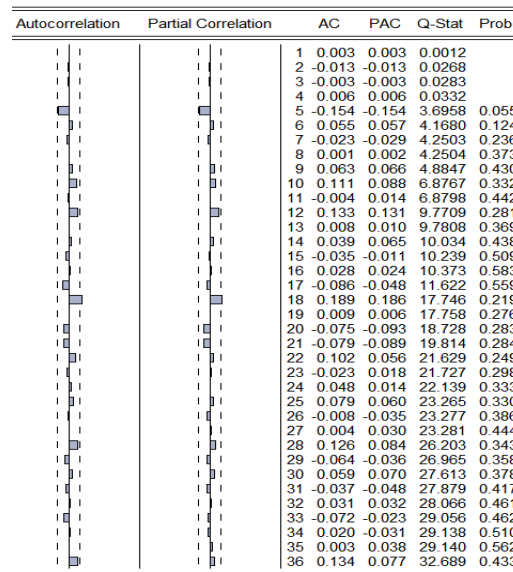| Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.003	0.003	0.0012	
		2	-0.013	-0.013	0.0268	
		3	-0.003	-0.003	0.0283	
		4	0.006	0.006	0.0332	
		5	-0.154	-0.154	3.6958	0.055
		6	0.055	0.057	4.1680	0.124
		7	-0.023	-0.029	4.2503	0.236
		8	0.001	0.002	4.2504	0.373
		9	0.063	0.066	4.8847	0.430
		10	0.111	0.088	6.8767	0.332
		11	-0.004	0.014	6.8798	0.442
		12	0.133	0.131	9.7709	0.281
		13	0.008	0.010	9.7808	0.369
		14	0.039	0.065	10.034	0.438
		15	-0.035	-0.011	10.239	0.509
		16	0.028	0.024	10.373	0.583
		17	-0.086	-0.048	11.622	0.559
		18	0.189	0.186	17.746	0.219
		19	0.009	0.006	17.758	0.276
		20	-0.075	-0.093	18.728	0.283
		21	-0.079	-0.089	19.814	0.284
		22	0.102	0.056	21.629	0.249
		23	-0.023	0.018	21.727	0.298
		24	0.048	0.014	22.139	0.333
		25	0.079	0.060	23.265	0.330
		26	-0.008	-0.035	23.277	0.386
		27	0.004	0.030	23.281	0.444
		28	0.126	0.084	26.203	0.343
		29	-0.064	-0.036	26.965	0.358
		30	0.059	0.070	27.613	0.378
		31	-0.037	-0.048	27.879	0.417
		32	0.031	0.032	28.066	0.461
		33	-0.072	-0.023	29.056	0.462
		34	0.020	-0.031	29.138	0.510
		35	0.003	0.038	29.140	0.562
		36	0.134	0.077	32.689	0.433

Fig. 6. The autocorrelation and partial autocorrelation diagrams of the residuals with the model parameters (3,1,1)

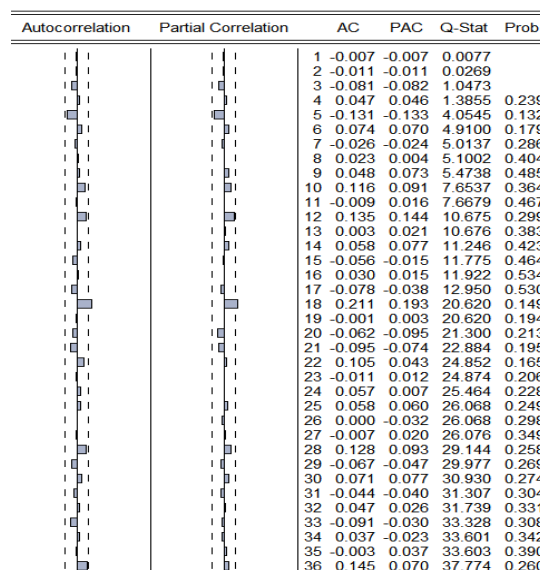| Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.0077	
		2	-0.011	-0.011	0.0269	
		3	-0.081	-0.082	1.0473	
		4	0.047	0.046	1.3855	0.239
		5	-0.131	-0.133	4.0545	0.132
		6	0.074	0.070	4.9100	0.179
		7	-0.026	-0.024	5.0137	0.286
		8	0.023	0.004	5.1002	0.404
		9	0.048	0.073	5.4738	0.485
		10	0.116	0.091	7.6537	0.364
		11	-0.009	0.016	7.6679	0.467
		12	0.135	0.144	10.675	0.299
		13	0.003	0.021	10.676	0.383
		14	0.058	0.077	11.246	0.423
		15	-0.056	-0.015	11.775	0.464
		16	0.030	0.015	11.922	0.534
		17	-0.078	-0.038	12.950	0.530
		18	0.211	0.193	20.620	0.149
		19	-0.001	0.003	20.620	0.194
		20	-0.062	-0.095	21.300	0.213
		21	-0.095	-0.074	22.884	0.195
		22	0.105	0.043	24.852	0.165
		23	-0.011	0.012	24.874	0.206
		24	0.057	0.007	25.464	0.228
		25	0.058	0.060	26.068	0.249
		26	0.000	-0.032	26.068	0.298
		27	-0.007	0.020	26.076	0.349
		28	0.128	0.093	29.144	0.258
		29	-0.067	-0.047	29.977	0.269
		30	0.071	0.077	30.930	0.274
		31	-0.044	-0.040	31.307	0.304
		32	0.047	0.026	31.739	0.331
		33	-0.091	-0.030	33.328	0.308
		34	0.037	-0.023	33.601	0.342
		35	-0.003	0.037	33.603	0.390
		36	0.145	0.070	37.774	0.260

Fig. 7. The autocorrelation and partial autocorrelation diagrams of the residuals with model parameters (2,1,1),

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.007	-0.007	0.0068	
		2	-0.055	-0.055	0.4620	
		3	-0.077	-0.078	1.3693	0.242
		4	0.074	0.070	2.2201	0.330
		5	-0.125	-0.134	4.6636	0.198
		6	0.067	0.071	5.3771	0.251
		7	-0.003	-0.008	5.3786	0.371
		8	0.024	0.007	5.4716	0.485
		9	0.039	0.071	5.7193	0.573
		10	0.112	0.089	7.7648	0.457
		11	-0.009	0.019	7.7774	0.557
		12	0.141	0.163	11.038	0.355
		13	0.010	0.023	11.054	0.439
		14	0.046	0.071	11.414	0.494
		15	-0.062	-0.017	12.061	0.523
		16	0.018	-0.005	12.116	0.597
		17	-0.102	-0.073	13.903	0.533
		18	0.196	0.175	20.514	0.198
		19	0.000	-0.017	20.514	0.249
		20	-0.090	-0.118	21.917	0.236
		21	-0.108	-0.091	23.954	0.198
		22	0.101	0.006	25.744	0.174
		23	0.031	0.044	25.914	0.210
		24	0.058	0.022	26.516	0.230
		25	0.053	0.054	27.017	0.255
		26	0.017	0.000	27.069	0.301
		27	-0.012	0.033	27.096	0.351
		28	0.116	0.110	29.594	0.285
		29	-0.079	-0.043	30.764	0.281
		30	0.051	0.071	31.252	0.306
		31	-0.046	-0.035	31.653	0.335
		32	0.052	0.033	32.165	0.360
		33	-0.073	-0.012	33.185	0.361
		34	0.025	-0.042	33.307	0.403
		35	0.052	0.082	33.848	0.426
		36	0.159	0.081	38.868	0.260

Fig. 8. The autocorrelation and partial autocorrelation diagrams of the residuals with the model parameters (1,1,1)

From Fig.6,Fig.7,Fig.8,we know the residual probabilities of the Q-statistics for all the models are all greater than 0.05,therefore, we accept the null hypothesis that the residuals are independent of each other at the level of 5%.That is, the residuals are white noise sequences.Therefore, under the 5% confidence level, when the value of (p, d, q) is (3,1,1) or (2,1,1) or (1,1,1),the ARIMA model for the extraction of features in the sequence is sufficient, the model is significant.

The significance test. The significance test of the ARIMA model was judged by the significance of the parameter t-test.When the parameters (p, d, q) are (3,1,1), (2,1,1), (1,1,1),the results of the t-test for the oxygen tube consumption estimation model are shown in Table 3,Table 4,Table 5.

Table 3. The parameter estimates and test magnitudes of the model when the parameter is (3,1,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.214753	5.968581	1.041245	0.2995
AR(1)	-0.187873	0.099101	-1.895769	0.0600
AR(2)	-0.101899	0.099947	-1.019524	0.3097
AR(3)	-0.106475	0.095162	-1.118877	0.2651
MA(1)	-0.857408	0.055665	-15.40296	0.0000

Table 4. The parameter estimates and test magnitudes of the model when the parameter is (2,1,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.187155	5.855725	1.056599	0.2925
AR(1)	-0.161897	0.095606	-1.693383	0.0925
AR(2)	-0.067061	0.094105	-0.712617	0.4772
MA(1)	-0.877370	0.046113	-19.02645	0.0000

Table 5. The parameter estimates and test magnitudes of the model when the parameter is (1,1,1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	6.305768	5.851216	1.077685	0.2829
AR(1)	-0.144528	0.091167	-1.585313	0.1150
MA(1)	-0.885544	0.040611	-21.80546	0.0000

In table 3,table 4 and table 5,the first column is the explanatory variable name (including the constant term), the second column is the corresponding parameter estimation, the third column is the standard error of the parameter, the fourth column is t statistics, and the fifth is the bilateral probability value of the t test.Table 4 shows that the absolute value of the second-order self-regression coefficient of the model is0.067061,which is less than 0.1.So the parameter is significantly zero. From Table 3 and Table 5, when

the ARIMA model parameters (p, d, q) are (3,1,1) or (1,1,1), the absolute value of the coefficients in the model is greater than 0.1, so the parameters can be considered as significant.

Optimization of the Proposed Material Consumption Estimation Model. At this point, the model can fully extract the features of the sample data when the parameter (p, d, q) of the ARIMA model is (3,1,1) or (1,1,1).And the parameters of the model are all significant . At this time, AIC (Akaike Information Criterion) and SC (Schwartz's Criterion) are needed to select the best of the two models. The information criteria corresponding to the two models are shown in Table 6.

Table 6. These are the information criteria for the model when the parameters are (3,1,1) and (1,1,1)

Model	AIC	SC
ARIMA(3,1,1)	15.91063	16.01189
ARIMA(1,1,1)	15.88620	15.94641

As shown in Table 6, when the parameter is (1,1,1), the AIC or SC values of the model is smaller than the values of the model when the parameter is (3,1,1). Therefore, when the parameter is (1,1,1), ARIMA model is the best model among the three models.

According to the same steps, we can find the best model of the other three materials. The initial prediction model of the four materials procurement is shown in Table 7.

Table 7. The forecasting model of the four types of material

Material Name	Forecasting Model
oxygen tube	(1,1,1)
insulin needle	(5,1,1)
anaesthetic mask	(4,0,3)
insulin infusion line	(1,1,1)

Correct model to avoid inventory deficiencies. According to the prediction model obtained from the previous prediction, we get the initial prediction value, the initial prediction value minus the actual usage quantity and get the redundant amount. The redundancy of the four materials is shown below.
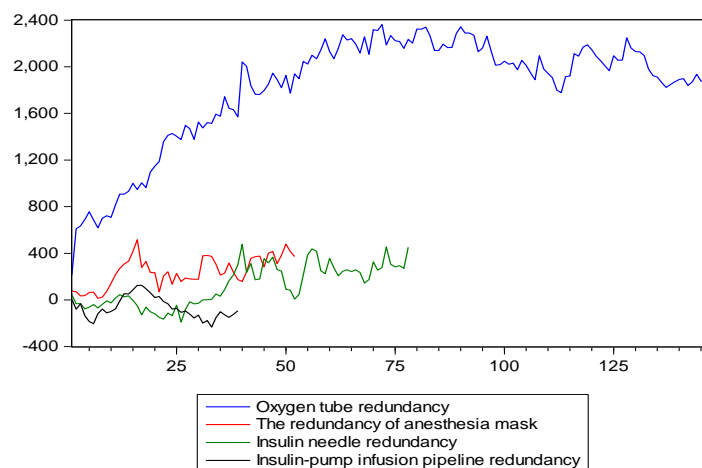

Fig. 9. Four kinds of consumables redundancy

(the horizontal coordinate is the sequence position rather than the time axis)

From Fig.9, we can see that the redundancy of oxygen tube and anesthetic mask is not negative. It doesn't need to consider the problem of inventory shortage. We only need to consider whether the predicted value is smaller than that of artificial prediction.However, the redundancy of the insulin syringe and the insulin pump infusion pipeline is negative. It is necessary to establish a model for redundant volume and correct the initial prediction value.

4   Performance Evaluation

The following Table 8 is a modified model of the two consumables.

Table 8. Correction model of insulin needle and insulin pump infusion pipeline.

Material Name	Forecasting Model	Modified Model
insulin needle	ARIMA(5,1,1)	ARIMA(0,1,4)
insulin infusion line	ARIMA(1,1,1)	ARIMA (0,1,0)

After fitting the redundant model with the revised model, it is concluded that the correction parameter of the infusion pipeline is 235 (the absolute value of the minimum negative value) and the correction parameter of the insulin needle is 204. The Fig.10 is the revised redundancy.
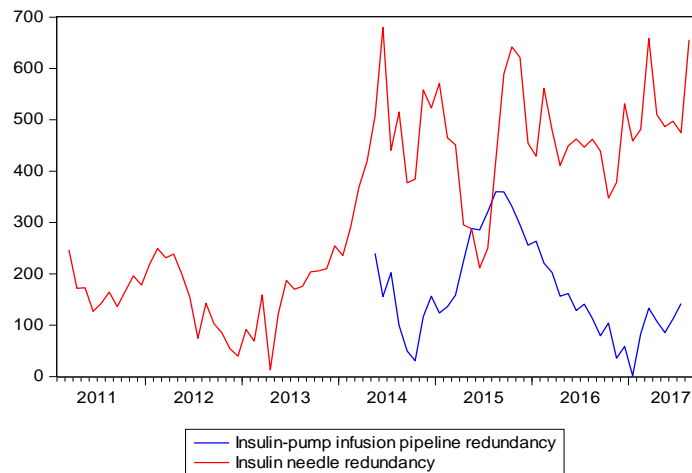
Fig. 10. The redundant after the model of insulin injection and insulin pump infusion pipeline are corrected

The FIg.10 shows that all the redundancy is bigger than zero.From this we can see that by modifying the model, the initial prediction values are corrected, redundancise are greater than zero, so there is no shortage of stock. Therefore, the correction effect of the modified models meets the requirements.

The following Table 9 is the comparison of the average redundancy and the average redundancy amount for the prediction and artificial prediction of the four models of consumables.

Table 9. The comparison between the redundancy in manual prediction and the redundancy in model prediction

Material Name	Artificial redundancy Number/per month	Artificial redundancy Amount(yuan)/per month	model redundancy number/per month	model redundancy amount(yuan)/per month	Save amount(yuan)/ per month
oxygen tube	3707	5931.2	1821	2913.6	3017.6
insulin needle	722	11624.2	328	5280.8	6343.4
anaesthetic mask	316	12008	244	9272	2736
insulin pump infusion line	307	35305	168	19320	15985
Total	5052	64868.4	2561	36786.4	28082

From Table 9, we can see that the prediction results of four consumables, including insulin syringe, oxygen tube, anesthetic mask and insulin pump pipeline, can save 28082 yuan per month than artificial prediction. This proves that the medical supplies procurement forecasting model based on the ARIMA model can draw the prediction results more accurate than manual estimation, while avoiding the occurrence of inventory shortage, effectively reducing the redundancy of material procurement, saving the cost of procurement and warehouse management cost. Under the background of the new medical reform, it provides a feasible scheme for the sustainable development of the hospital and the implementation of the cost control.

5   Conclusions

The experimental results show that the proposed scheme solves the problem that large amount of data in medical group is difficult to analyze and process in case of decentralized storage through distributed data mining technology. The proposed scheme can also solve the over-procurement problem caused by the over-estimation and avoid the situation of insufficient stock through the improved ARIMA model. So the program effectively saves the procurement costs and warehouse management costs. Under the background of new medical reform, hospital integration platform, hospital information interconnection and regional health information interconnection, the program provides a feasible solution for the hospital to achieve sustainable development and implement cost control.

References

[1]   Wu Jing, Sun Chaowei, Liu Hairong. Research Report on medical consumables and expenditure management [J]. Financial Circles: Academic Edition, 2015 (5): 72-72.

[2]   Cai Zeling, Lu Xizhi, Yin Jiming, et al. Strengthening management of medical consumables management to reduce hospital operating cost [J]. Management Observation, 2013 (17): 141-142.

[3]   Parmar R, Roy S, Bhattacharaya D, et al. Large Scale Encryption in Hadoop Environment: Challenges and Solutions[J]. IEEE Access, 2017, PP(99):1-1.

[4]   Chang R S, Liao C S, Fan K Z, et al. Dynamic Deduplication Decision in a Hadoop Distributed File System[J]. International Journal of Distributed Sensor Networks, 2014, 2014(5):1-14.

[5]   Chai X Q, Dong Y L, Li J F. Profit-oriented task scheduling algorithm in Hadoop cluster[J]. Eurasip Journal on Embedded Systems, 2016, 2016(1):6.

[6]   Ma N, Hu R, Zhang R. Parallel Process of Virtual Screening Result File Based on Hadoop[J]. Lecture Notes in Computer Science, 2013, 7929:511-517.

[7]   Ibrahim S, Moise D, Chihoub H E, et al. Towards Efficient Power Management in MapReduce: Investigation of CPU-Frequencies Scaling on Power Efficiency in Hadoop[J]. Lecture Notes in Computer Science, 2014, 8907:147-164.

[8]   Fu X, Gao Y, Luo B, et al. Security Threats to Hadoop: Data Leakage Attacks and Investigation[J]. IEEE Network, 2017, 31(2):67-71.

[9]   Box, G. E., & Jenkins, G. M. (1976). Time series analysis: forecasting and control rev. ed. Oakland, California, Holden-Day, 1976, 31(4), 303-303.

[10]   Contreras J, Espinola R, Nogales F J, et al. ARIMA models to predict next-day electricity prices[J]. IEEE Transactions on Power Systems, 2003, 18(3):1014-1020.

[11]   Zhang G P. Time series forecasting using a hybrid ARIMA and neural network model[J]. Neurocomputing, 2003, 50(1):159-175.

[12]   Conejo A J, Plazas M A, Espinola R, et al. Day-ahead electricity price forecasting using the wavelet transform and ARIMA models[J]. IEEE Transactions on Power Systems, 2005, 20(2):1035-1042.

[13]   Williams B M, Hoel L A. Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results[J]. Journal of Transportation Engineering, 2003, 129(6):664-672.

[14]   Kavasseri R G, Seetharaman K. Day-ahead wind speed forecasting using -ARIMA models[J]. Renewable Energy, 2009, 34(5):1388-1393.

[15]   Chen P, Pedersen T, Bak-Jensen B, et al. ARIMA-Based Time Series Model of Stochastic Wind Power Generation[J]. IEEE Transactions on Power Systems, 2010, 25(2):667-676.

[16]   Liu Yong, Wang Xuhui.Application of ARIMA Model in Energy Consumption Prediction in China [J]. Economic Survey, 2007 (5): 11-13.

[17]   Liu H, Tian H Q, Li Y F. Comparison of two new ARIMA-ANN and ARIMA-Kalman hybrid methods for wind speed prediction[J]. Applied Energy, 2012, 98(1):415-424.

[18]   Montanari A, Rosso R, Taqqu M S. A seasonal fractional ARIMA Model applied to the Nile River monthly flows at Aswan[J]. Water Resources Research, 2000, 36(5):1249-1259.