Association for Information Systems
AIS Electronic Library (AISeL)

PACIS 2016 Proceedings

Pacific Asia Conference on Information Systems (PACIS)

Summer 6-27-2016

ANALYZING MEDICAL TRANSACTION DATA BY USING ASSOCIATION RULE MINING WITH MULTIPLE MINIMUM SUPPORTS

Shiang-Lin Lin
National Chengchi University, shiang0623@gmail.com

Chen-Shu Wang
National Taipei University of Technology, wangcs@ntut.edu.tw

Hui-Chu Chiu
Chung Hua University, zinniachiiou@yahoo.com.tw

Chun-Jung Juan
National Defense Medical Center Chief of Neuroradiology, peterjuancj@yahoo.com.tw

Follow this and additional works at: http://aisel.aisnet.org/pacis2016

ANALYZING MEDICAL TRANSACTION DATA BY USING ASSOCIATION RULE MINING WITH MULTIPLE MINIMUM SUPPORTS

Shiang-Lin Lin, Department of Management Information System, National Chengchi University, Taipei, Taiwan, shiang0623@gmail.com

Chen-Shu Wang, Department of Information and Finance Management, National Taipei University of Technology, Taipei, Taiwan, wangcs@ntut.edu.tw

Hui-Chu Chiu, Ph.D Program of Technology Management, Chung Hua University, Hsinchu, Taiwan, zinniachiiou@yahoo.com.tw

Chun-Jung, Juan, National Defense Medical Center Chief of Neuroradiology, Department of Radiology, Tri-Service General Hospital, Taipei, Taiwan, peterjuancj@yahoo.com.tw

Abstract

The quick development of IS has a huge impact on the healthcare industry. almost all the existing hospitals, clinics and other healthcare-related institutes have adopted a functionally powerful and highly integrated Hospital Information System (HIS) for management of clinic or medical-related affairs. The medical data stored in the HIS are collected from many different medical subsystems, However, problems of failed data sharing and inconsistent data content often occur among these subsystems, resulting in many hospitals collect a large amount of medical data, but not the ability to process and analyse these data properly, letting the valuable data in the HIS all go to waste. In this study, we made a practical visit to a certain hospital in Taiwan and collected radioimmunoassay (RIA) data from the Laboratory Information System (LIS) and the Departmental Registration System (DRS) of this hospital. Further, we proposed a method of the association rule mining in combination with the concept of multiple minimum supports to analyse and find valuable association rules from the RIA data. The analytical results found the method we proposed can indeed find association rules that would not be able to be found with the traditional association mining methods. It is very helpful in improving doctor-patient relationship and upgrading health care quality.

Keywords: Association rule mining, Multiple minimum support, Hospital information system, Radioimmunoassay,

1. INTRODUCTION

With the rapid development in network and scientific technologies, various types of information systems (IS) have been constantly developed in the past few years, which has made a huge impact on the healthcare industry (Reichertz 2006). This is because the healthcare industry involves much more complicated operational models and organizational relationship than other profit-seek industries. In view of this fact, almost all the existing hospitals, clinics and other healthcare-related institutes have adopted a functionally powerful and highly integrated Hospital Information System (HIS) for management of administrative, financial, clinic and many other medical-related affairs (Ismail et al. 2013).

The hospital information system (HIS) is a type of information system designed to manage the hospital's medical and administrative information (Ozkan et al. 2008). Collen (1991) defined the HIS as a system that integrates computer, medical examination, manually-operated, and communication apparatuses to give different departments of a hospital the ability to collect, store, process, retrieve and communicate patient care information and administration information. Through the HIS, automation of medical and administrative operating procedures in hospital can be realized to thereby improve patient safety as well as hospital efficiency and performance (Ammenwerth et al. 2003; Reichertz 2006).

The HIS has many different subsystems, such as Outpatient and Emergency System, Laboratory and Examination System, Medical Support System, etc. Some of the HIS subsystems are particularly developed in response to the specific needs of different departments. Owing to hospitals are extremely complex institutions which consist of large departments to facilitate treatment for patients every day, the subsystems of all medical organizations have to record and process a huge amount of different types of records every day (Julia & David, 2010), such as records of medical treatment, appointments with doctors, medicine pick-up, etc., and then gather and send all the records to the HIS for data integration, so that cross-reference of data in different information systems is available for different departments of the hospital. It is no doubt a huge quantity of data will accumulate in the HIS over time. Besides, the medical data stored in the HIS are collected from many different medical subsystems, however, these medical subsystems might use different development platforms or different data formats to often cause the problems of failed data sharing and inconsistent data content among the subsystems. As a result, the HIS of many hospitals have the ability to collect a large amount of medical data, but not the ability to process and analyze these data properly, letting the valuable data in the HIS all go to waste. Therefore, the existing hospitals and other healthcare organizations all eagerly look forward to finding an appropriate way of processing, analyzing and utilizing the huge amount of medical data in their HIS, so as to achieve the purpose of providing patients with further improved healthcare quality.

Generally speaking, the analysis of a huge amount of complicated data necessitates the use of data mining methods. Currently, there are already many studies that applied data mining to analyze medical data. For example, Ceglowski et al. (2007) applied the data mining approach to analyze the complex

relationship between patient urgency, treatment and disposal, and the occurrence of queues for treatment; Obenshain (2004) applied the data mining approach for mining huge quantities of medical data; Tsumoto et al. (2011) applied temporal data mining and exploratory data analysis techniques to hospital management data, such as laboratory data and patient records; Tsumoto & Hirano (2010) applied the data mining technology in medical risk management to detect, analyze and evaluate risks potentially existing in clinical environments; Yoo et al. (2015) used process mining technology to analyze process changes based on changes in the hospital environment; and Chen et al. (2015), basing on the medical homepage records, analyzed the key indicators of the hospital medical quality by using decision tree C4.5 algorithm. There are also many studies that used association rule mining to analyze medical data. For instance, Sasirekha & Punitha (2015) applied an associative classification method in medical dataset mining; and Nguyen et al. (2015) applied the Apriori-based method of association analysis for discovering toxicity progression patterns in the form of temporal association rules.

Through the use of association mining to analyze large quantities of medical data, it is able to find out patients' patterns of doctor visits or laboratory tests, and such patterns can be used in future similar medical treatments to provide useful suggestions for relevant medical treatment behavior and accordingly improve the medical quality. In the past, only one minimum support is used for the whole database when conducting the association mining. However, there could be several decades of treatment items, some items appear very frequently in the medical data, while others rarely appear. If the minimum support is set too high, those rules that involve rare items will not be found. To find rules that involve both frequent and rare items, the minimum support has to be set very low, which may, however, cause combinatorial explosion because those frequent items will be associated with one another in all possible ways (Liu et al. 1999), this dilemma is called the rare item problem (Han et al., 2004; Han et al., 2000). As rare items also contain worthwhile information, it is an important issue of how to use data mining to find useful rules involving those rare items (Bhatt & Patel, 2015).

To solve the aforesaid rare item problem, Liu et al. (1999) proposed a mining method with multiple minimum supports (MMSs) that allows users to specify different support values for different items. In this method, the support of an itemset is defined as the minimum support of all items contained in the itemset. The specification of MMSs allows frequent itemsets to potentially contain rare items which are nevertheless deemed important (Mobasher et al., 2001). There are already many studies that proposed the application of the concept of MMSs to association rule. For instance, Lee et al. (2005) provided an algorithm combing the concept of maximum constraint based on the Apriori approach, in order to define the minimum supports of itemsets when items have different minimum supports; Hu et al. (2013) developed an efficient mining algorithm of sequential patterns with MMSs to discover the complete set of patterns; Siji et al. (2016) proposed a fuzzy association rule model integrating the concept of fuzzy theory and the Multiple Support Apriori (MSapriori) approach to enhance prediction performance; and Hu et al. (2015) proposed a mining method including the concept of MMSs to enable users to specify

multiple minimum item repetition support (MIR) according to the natures of items.

Since mining association rules with MMSs is a critical generalization of the association mining problem, it also has a criticized problem. That is, according to the definition of association mining with MMSs, the minimum support for a rule is the smallest value among the minimum item supports (MIS) of all items that appear in the rule. Due to the above-defined feature, it is difficult to calculate the minimum confidence value of the rule so found. In other words, without a minimum confidence for validating the accuracy of the found rule (Hu & Chen, 2006).

In view of these facts, this study conducted a case study and made a practical visit to a hospital in Taiwan to get an idea about the problems currently encountered by the hospital's HIS. Further, this study collected the hospital's radioimmunoassay (RIA) data from the Laboratory Information System (LIS) and the Departmental Registration System (DRS) of the hospital. Then, we applied the association mining analysis in combination with the concept of multiple supports to analyze and find valuable association rules among different RIA examination items, so that the hospital can use these association rules as a reference in its subsequent efforts to upgrade the medical quality and improve the HIS.

2. LITERATURE REVIEW

2.1 Association Rule Mining

In the domain of data mining, it is always one of the important core issues of how to mine frequent itemsets in an original massive transaction database and discover the association between transactional items. By "frequent itemset", it means an item or an itemset whose number of occurrence in an original dataset is relatively high. Association rule mining is the best-known method for frequent itemset mining, with which we can discover in the massive transaction database the association between merchandise items to thereby mine valuable and human-unknown association rules, which can be provided to decision makers for reference during decision-making.

The earliest association rule mining algorithm is Apriori algorithm, which was proposed by Agrawal et al. in 1990s. According to the concept of Apriori algorithm, candidate itemsets are first generated from a large quantity of itemsets and a minimum support is set as a threshold value. When a database is scanned using Apriori algorithm and the frequency of occurrence of a candidate itemset is larger than or equal to the threshold value, the candidate itemset is referred to as a frequent itemset. Thereafter, meaningful association rules are mined based on the frequent itemsets (Agrawal et al. 1993).

Nowadays, the association rule mining has been applied to mine big data in a variety of domains. For example, Maul et al. (2015) applied Apriori algorithm to analyze customer data of retail markets; Sasirekha & Punitha (2015) mined medical datasets using the Association Classification method; and Ismail et al. (2015) applied the association rule mining to analyze e-commerce transaction data.

2.2 Multiple Minimum Support

Traditionally, the association mining method uses the smallest acceptable frequency of occurrence of item in the transaction database as the minimum support threshold value. However, by doing this, it is possible to miss some potentially valuable information. For example, in a hypermarket, there are a variety of merchandise items, which are quite different in their selling prices and profits. In the case only one minimum support value is used as a criterion to determine whether an item is a frequent item, a high-profits merchandise item having low number of occurrences in transaction will not be considered a frequent item. However, it is possible these high-profits items are the real things that decision makers are more concerned about. In view of this fact, it seems necessary to lower the minimum support value to enable mining of the association rules between the items that create high profits but has relatively low frequency of occurrence in transactions. However, the use of a lower minimum support value would also lead to the generation of a large number of unnecessary or meaningless association rules from the analysis, which will inevitably confuse the decision makers (Han et al., 2004; Han et al., 2000).

There were already many relevant studies that proposed some multiple-support association mining algorithms to overcome the above-mentioned problems and improve the traditional association rule mining. For example, Liu et al. (1999) proposed MSApriori algorithm that was based on the traditional Apriori algorithm with the concept of MMSs being added thereto. The concept of MSApriori algorithm is to set different support values for items of different characteristics, so that association rules that comply with real existing conditions and are needed by and attract decision makers can be satisfactorily found when mining association rules between merchandise items.

To implement MSApriori Algorithm (Liu et al. 1999), first set a MIS value for each item. Then, all the items are listed in ascending order of item's MIS value. Assume that there are k items c. Then, these items c are represented as $<c_1, c_2, ..., c_k>$, where $MIS(c_1) \leq MIS(c_2) \leq ... \leq MIS(c_k)$. For example, there are five items in the database, namely, item 1, item 2, item 3, item 4 and item 5; and the five items 1-5 respectively have an MIS value of $MIS(1) = 10\%, MIS(2) = 20\%, MIS(3) = 5\%, MIS(4) = 6\%$ and $MIS(5) = 15\%$. In this case, the sort order of the itemset is $<c_3, c_4, c_1, c_5, c_2>$, where $MIS(c_3) \leq MIS(c_4) \leq MIS(c_1) \leq MIS(c_5) \leq MIS(c_2)$. After the itemset is listed in order of MIS value, the database is scanned a first time to find candidate itemset F and frequent itemset $L1$ that have a length of 1. Wherein, in the candidate itemset F, the smallest one of the MIS values for the items is defined as a minMIS; and in the frequent itemset $L1$, each item must have a frequency of occurrence larger than or equal to the MIS value that was set at the very beginning. Then, the frequent itemset $L1$ is scanned a second time to find candidate itemsets that have a length no less than 2. Like the previous example, when the counts of the five items obtained from the scanning of 100 entries of data are $C_3=6, C_4=3, C_1=9, C_5=20$ and $C_2=25$, the obtained $F=\{c3, c1, c5, c2\}$ because the minMIS is 5%, and $L1=\{c3, c5, c2\}$.

3. RESEARCH METHOD

In this study, the analyzed case concerns the RIA test data of a certain hospital. The left half of Figure 2 shows the workflow in the hospital.
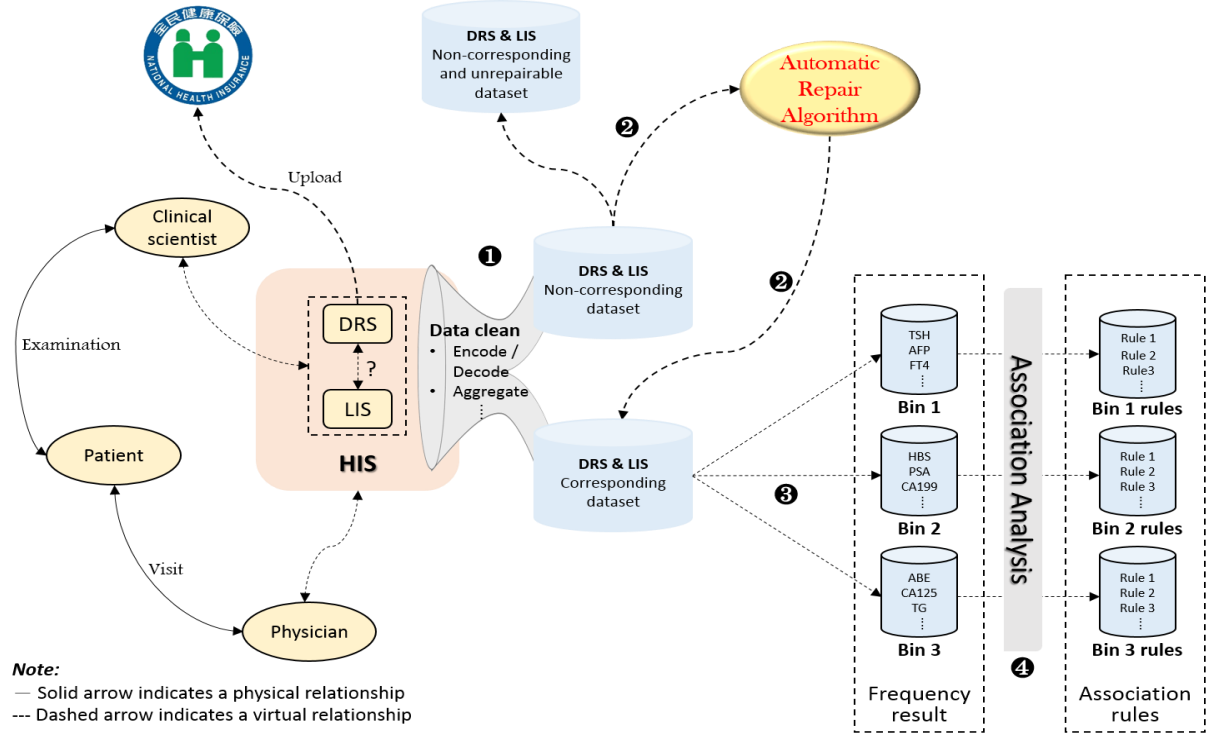

Table 2 The RIA test and research process

First, when an outpatient physician logs in the HIS menu and signs a laboratory test request form for a patient via the Outpatient and Emergency System (OES), the Laboratory Information System (LIS) will receive the physician's test request form and then retrieves from the HIS all relevant data needed for the requested test, such as the patient's data, test items and so on. Thereafter, the Departmental Registration System (DRS) will also access data related to the test requested by the physician. When the patient's data and the test request-related data have been transmitted to the LIS and the DRS, respectively, a clinical scientist enters the LIS to do relevant recording, tests and analyses, and sends test reports back to the LIS, from where the test reports are further sent to the HIS and archived for future use and check by other related persons. When the test reports are output, the clinical scientist also has to log in the DRS to record, verify and make an entry of account for the test. Meanwhile, the account record is also sent by the DRS to the HIS for archiving and using as a basis for future request of laboratory fee from the National Health and Insurance Administration.

The right half of Figure 2 shows the steps of data processing and analysis, which includes six data processing steps as follows:

- Step 1. Data clean: Original medical data existing in the LIS and the DRS of the hospital's HIS are aggregated and then subjected to a number of data preprocesses, and finally, data having a

correspondence found between them in the two systems and data having no correspondence found between them in the two systems are stored in two separate databases. The following is a brief description of the data processes involved in data clean:

- o Encode/decode: Delete unnecessary fields from the original data. Further, since different abbreviations/acronyms are possibly used in the LIS and the DRS to define test/examination items, it is necessary for the two systems to use consistent test/examination item names to avoid errors during data matching between systems.

- o Aggregate: It is possible that, in either the LIS or the DRS, two or more entries of the same one patient appeared in the same day. Therefore, it is necessary to delete the repeated data or combine the data into one complete data.

- o Consistency validate: Match datasets between the LIS and the DRS to find the corresponding dataset between the two systems (i.e. data that exist in both of the two systems) and the non-corresponding dataset between the two systems (i.e. data that exist in only one of the two systems), and store the two types of datasets in a "database of corresponding dataset" and a "database of non-corresponding dataset", respectively.

- Step 2. Data repair: Divide the non-corresponding dataset into unrepairable data and repairable data, which are independently processed as follows:

- o Unrepairable records: For data that are inconsistent with one another between the two systems and could not be further processed (e.g., inconsistent laboratory test items), they are stored in a "database of unrepairable dataset" for further checking later.

- o Repairable records: For data that are inconsistent with one another between the two systems but are automatically repairable, such as data showing inconsistent patient names, medical record numbers and examination dates, they are automatically repaired using an automatic repair algorithm to become consistent with one another. The consistent data are then transferred to the database of corresponding dataset.

- Step 3. Data grouping: According to interviews with professional physicians and suggestions given by them, all laboratory test items in the corresponding dataset are classified into multiple bins based on the times of appearance of individual test items. These bins of test items are separately stored in different databases to facilitate subsequent multiple-support association rule analysis.

- Step 4. Conducting association analysis:

- o Set the MS value for each bin: According to the interviews with physicians and suggestions given by them, an appropriate minimum support (MS) value is set for the physician orders (PO) in each bin.

- o Conduct association mining: Association analysis are conducted for each of the bins based on the preset MS value, and meaningful rules or rules valuable for reference found for each bin from the association mining are stored in a Database of Association Rules. These association rules can be recommended to physicians for their future reference.

4. CASE STUDY

This study took a certain hospital's radioimmunoassay (RIA) data in Taiwan as a sample case to analyze. RIA is a sensitive method for measuring tiny substance in the blood. Generally, RIA-related data processing can include two major parts, namely, laboratory test data processing and registration data processing. The laboratory test data processing further includes laboratory test request, specimen collection and processing, test result report output, etc. The purpose of registration data processing is to recognize the performed test item, which is accounted for only after the recognition and qualified for use as a basis for future request of laboratory fee from the National Health and Insurance Administration.

The hospital's RIA received data from the Laboratory Information System (LIS) and the Departmental Registration System (DRS) of the hospital. Generally speaking, laboratory test data and registration data should be consistent with one another. However, from preliminary interviews with hospital staffs, it was found the hospital's systems lack full compatibility with one another. Some of the laboratory test data and registration data were processed through manual matching, which not only consumed more time and effort, but also tended to cause errors in matching.

4.1 Data Collection

In this study, we collected, from the RIA Division of the hospital's Department of Nuclear Medicine, data statements produced by the LIS and the DRS during the period from December 2009 to July 2010. The distribution of relevant data is shown in Table 1. Since the data is sufficiently large in amount, it can be suitably used in this study for a stress test on our analytical model. Further, with the high dispersion of distribution of items (i.e. physician orders), it is also appropriate to apply the multiple-support association rule mining in the data analysis.

Year	2009	2010							Total
Month	Dec.	Jan.	Feb.	Mar.	Apr.	May.	Jun.	Jul.	
Patients	5,748	2,671	4,446	7,778	5,972	6,427	6,515	5,010	44,567
Number of Test Items	12,688	5,801	9,646	1,7156	1,2839	1,4623	15,063	11,684	99,500
Note: The Source of data is original medical data from the hospital's LIS.									

Table 1. Number of Patients Received RIA Test and Number of RIA Test Items Performed

As can be found from Table 1, during the eight months from December 2009 to July 2010, the monthly average number of patients received RIA test was 5570 and the monthly average number of RIA test items performed was 12,437 (requested via total 25 different types of physician orders). In addition, the number of patients received RIA test and RIA test items performed are very unevenly distributed in the eight months. This situation is just the same as the transaction records of merchandise items in the real world.

4.2 Data Matching

Further, when conducting data matching between the LIS and the DRS system, we found the number of laboratory test records in the LIS is 44,567, while only 34,826 test records were shown in the DRS. That is, a significant difference in the number of records of data was existed between the two systems. Therefore, before conducting the subsequent association analysis, data matching must be conducted for all laboratory test data stored in the two systems one by one to locate all records of data that appeared in both of the two systems. However, among the records of data located through the above data matching, only 32,514 records of data are completely consistent in content. That is, there was still a small part of matched data records that actually showed data heterogeneity between them. In other words, these data records are not fully consistent in their content. Table 2 shows the number of records of different types of data that were found inconsistent in content.

Inconsistent Data Type	Patient Name	Medical Record Number	Date of Doctor Visit	Physician Order
Number of Records	60	878	61	459

Table 2. Number of Records of Different Types of Data Found Inconsistent in Content

In this study, the content-inconsistent records of data shown in Table 2 were divided into repairable records, including inconsistent records of patient name, medical record number and date of doctor visit, and unrepairable records, including inconsistent records of physician order. Wherein, the repairable records were automatically repaired through an automatic repair algorithm, while the unrepairable records were not automatically repairable and accordingly, stored in an unrepairable dataset database and excluded from subsequent analysis. Finally, there were total 33,825 records of data that were usable in subsequent analysis and separately stored in several mappers.

4.3 Data Grouping

These 33,825 records of data came from 25 types of physician orders, and the number of laboratory tests requested by and performed according to different physician orders varied significantly. In this circumstance, if the concept of MMSs were not used in the association analysis and only one minimum support value, such as minimum *supp=0.1*, was set for more than thirty thousand records, it was very possible that association rules could be found only the physician orders which appearing frequently. The results from such an analysis would obviously have very limited help to decision makers. Therefore, in this study, we applied association mining with MMSs in data analysis. However, we had first conducted interviews and discussions with professional physicians before the analysis was started, and divided the physician orders into 3 bins based on the professional physicians' suggestions and the number of occurrences of each physician order type. The types of physician orders and the counts thereof in different bins are aggregated and shown in Table 3.

Bin 1		Bin 2		Bin 3			
PO	Count	PO	Count	PO	Count	PO	Count
TSH	10117	HBS	4783	CA125	1506	HAV	82
		PSA	4196	TG	1477	FBHCG	53
AFP	9243	CA199	3464	THY	1152	ACHR	42
		TRI	3429	FPSA	890	ACIGM	29
FT4	8321	HBE	3019	ABC	889	T3UP	7
		CA153	2918	SCC	600		
CEA	7858	ABS	2599	HAIGM	132		
		ABE	2107	B2M	87		

Table 3. RIA Data Grouping

4.4 Association Rule Mining Conducting

According to the grouping results shown in Table 3, there are 25058 physician orders in bin 1, 19218 physician orders in bin 2, and 6509 physician orders in bin 3. In this study, we also set different MIS values for different bins based on suggestions given by professional physicians, and then started association rule mining to find the association rule between items in each bin that is valuable for reference. Part of the analytical results are shown in Table 4.

	Bin 1			Bin 2			Bin 3		
Qualify instances/ instances	7517/25058			1922/19218			651/6509		
MIS value	30%			10%			10%		
Association rules	Left	Right	Conf.	Left	Right	Conf.	Left	Right	Conf.
	FT4	TSH	0.99	TRI	TSH	0.99	FPSA	PSA	1.00
	TSH	FT4	0.82	CA153	CEA	0.97	TG	TSH	0.97
				ABE	HBE	0.94	TG	FT4	0.92
				ABS	HBS	0.89	ABC	HBS	0.89
				TRI	FT4	0.80	THY	TRI	0.79
				CA199	CEA	0.78	ABC	ABS	0.75
				HBE	AFP	0.66			

Table 4. Association Rule between Physician Orders in Each Bin

In Table 4, we listed some meaningful and valuable association rules found between the physician orders in each bin. For example, in this study, we found that among 2107 records in bin 2 that showed the physician order ABE, 1971 records also showed the physician order HBE (*confidence = 0.94*). This means ABE and HBE are two laboratory test items often simultaneously requested by physicians for their patients. After a discussion with the chief physician of the Lab Division of the hospital, we found ABE and HBE all are laboratory test items related to hepatitis B. When an experienced physician requests the laboratory test item of ABE for a patient, he/she will also request the laboratory test item of HBE for the patient. The chief physician of the Lab Division of the hospital also stated that other physician orders appeared in other mined rules also have some association between them. This proves the method proposed and implemented in this study can indeed find many meaningful and valuable association rules.

5. CONCLUSION

Hospitals are extremely complex institutions which consist of large departments to facilitate treatment for patients, and the medical subsystems of all medical organizations have to record and process a huge amount of different types of records every day. However, problems of failed data sharing and inconsistent data content often occur among these subsystems, resulting in many hospitals have not the ability to process and analyze these data properly.

In this study, the concept of MMSs was applied to mine and analyze the association rules between the RIA-related medical data of a certain hospital in Taiwan. Unlike the traditional MSApriori, in this study, we first classified the data into multiple bins according to the number of occurrences of items (i.e. the physician orders) and then set a single support threshold for each bin before the association mining, so as to find valuable association rules for each bin. Since the bins were different in their instances and MIS values, what we conducted in this study is somewhat similar to the concept of MMSs.

From the analytical results of this study, it is found the method we proposed can indeed find association rules that would not be able to be found with the traditional association mining methods. With the inspiring study result, we have also passed the mined association rules to the physician of the hospital's Laboratory Division for his verification. According to the physician of Lab Division, the analytical results from this study are indeed valuable for reference and can be provided to new physicians for their reference in medical diagnosis and treatment. For example, our study result can help physicians to know which laboratory test items being requested are in association with one another. Thus, our study result is very helpful in improving doctor-patient relationship and upgrading health care quality.

References

Agrawal, R., Lmielinski, T.& Swami, A. (1993). Mining association rules between sets of items in large databases. in Proc. of the ACM SIGMOD Int. Conf. on Management of Data, 22(2).

Ammenwerth, E., Gräber, S., Herrmann, G., Bürkle, T., & König, J. (2003). Evaluation of health information systems—problems and challenges. International journal of medical informatics, 71(2), 125-135.

Bhatt, U., & Patel, P. (2015). A Novel Approach for Finding Rare Items Based on Multiple Minimum Support Framework. Procedia Computer Science, 57, 1088-1095.

Ceglowski, R., Churilov, L., & Wasserthiel, J. (2007). Combining data mining and discrete event simulation for a value-added view of a hospital emergency department. Journal of the Operational Research Society, 58(2), 246-254.

Chen, L., Bi, L., Si, H., Zhang, J., & Ren, Y. (2015, September). Research on prediction method for pivotal indicator of hospital medical quality using decision tree. In Software Engineering and Service Science (ICSESS), 2015 6th IEEE International Conference on (pp. 247-250). IEEE.

Collen, M. F. (1991). A brief historical overview of hospital information system (HIS) evolution in the United States. International journal of bio-medical computing, 29(3), 169-189.

Han, J., Pei, J., & Yin, Y. (2000). Mining frequent patterns without candidate generation. In ACM Sigmod Record, 29(2), 1-12).

Han, J., Pei, J., Yin, Y., & Mao, R. (2004). Mining frequent patterns without candidate generation: A frequent-pattern tree approach. Data mining and knowledge discovery, 8(1), 53-87.

Hu, Y. H., & Chen, Y. L. (2006). Mining association rules with multiple minimum supports: a new mining algorithm and a support tuning mechanism. Decision Support Systems, 42(1), 1-24.

Hu, Y. H., Tsai, C. F., Tai, C. T., & Chiang, I. C. (2015). A novel approach for mining cyclically repeated patterns with multiple minimum supports. Applied Soft Computing, 28, 90-99.

Hu, Y. H., Wu, F., & Liao, Y. J. (2013). An efficient tree-based algorithm for mining sequential patterns with multiple minimum supports. Journal of Systems and Software, 86(5), 1224-1238.

Ismail, M., Ibrahim, M. M., Sanusi, Z. M., & Nat, M. (2015). Data Mining in Electronic Commerce: Benefits and Challenges. International Journal of Communications, Network and System Sciences, 8(12), 501.

Ismail, N. I., Abdullah, N. H., Shamsudin, A., & Ariffin, N. A. N. (2013). Implementation differences of Hospital Information System (HIS) in Malaysian public hospitals. International Journal of Social Science and Humanity, 3(2), 115.

Julia A.M., & David W.B. (2010). Paperless healthcare: Progress and challenges of an IT-enabled healthcare system. Business Horizons, 53(2), 119-130.

Lee, Y. C., Hong, T. P., & Lin, W. Y. (2005). Mining association rules with multiple minimum supports using maximum constraints. International Journal of Approximate Reasoning, 40(1), 44-54.

Liu, B., Hsu, W., & Ma, Y. (1999, August). Mining association rules with multiple minimum supports. In Proceedings of the fifth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 337-341). ACM.

Maul, N., Jagtap, G., Kale, P., Waydande, P., & Lokhande, M. Parallel Implementation of Apriori Algorithm on Multicore System for Retail Market.

Mobasher, B., Dai, H., Luo, T., & Nakagawa, M. (2001). Effective personalization based on association rule discovery from web usage data. In Proceedings of the 3rd international workshop on Web information and data management (pp. 9-15). ACM.

Nguyen, D., Luo, W., Phung, D., & Venkatesh, S. (2015). Understanding Toxicities and Complications of Cancer Treatment: A Data Mining Approach. In AI 2015: Advances in Artificial Intelligence (pp. 431-443). Springer International Publishing.

Obenshain, M. K. (2004). Application of data mining techniques to healthcare data. Infection Control, 25(08), 690-695.

Ozkan S., Byakal N., & Sincan M. (2008). Assessing Hospital Information Systems Processes: A Validation of PRISE Information Systems Success Model in Healthcare. AMCIS 2008 Proceedings. Paper 250.

Reichertz, P. L. (2006). Hospital information systems—Past, present, future. International Journal of Medical Informatics, 75(3), 282-299.

Sasirekha, D., & Punitha, A. (2015). A Comprehensive Analysis on Associative Classification in Medical Datasets. Indian Journal of Science and Technology, 8(33).

Siji, P. D., VALARMATHI, D. M. L., & MOHANA, S. (2016). A new soft computing technique for efficient rule mining. Journal of Theoretical and Applied Information Technology, 83(3).

Tsumoto, S., & Hirano, S. (2010). Risk mining in medicine: Application of data mining to medical risk management. Fundamenta Informaticae, 98(1), 107-121.

Tsumoto, S., Hirano, S., & Tsumoto, Y. (2011). Towards data-oriented hospital services: Data mining in hospital management. In SRII Global Conference (SRII), 349-356..

Yoo, S., Cho, M., Kim, E., Kim, S., Sim, Y., Yoo, D., ... & Song, M. (2015). Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital. International Journal of Medical Informatics.