Capturing Behavior of Medical Staff: A Similarity-Oriented Temporal Data Mining Approach*

Shusaku Tsumoto[1], Shoji Hirano[1], Haruko Iwata[1], and Yuko Tsumoto[1]

[1] Department of Medical Informatics, School of Medicine, Faculty of Medicine
Shimane University
89-1 Enya-cho Izumo 693-8501 Japan
{tsumoto,hirano,haruko23}@med.shimane-u.ac.jp

[2] Department of Fundamental Nursing, School of Nursing, Faculty of Medicine
Shimane University
89-1 Enya-cho Izumo 693-8501 Japan
tsumotoy@med.shimane-u.ac.jp

Abstract. This paper presents data mining results in which temporal behavior of global hospital activities are visualized. The results show that the reuse of stored data will give a powerful tool for hospital management and lead to improvement of hospital services.

Keywords: temporal data mining, visualization, clustering, hospital information system.

1 Introduction

Twenty years have passed since clinical data were stored electronically as a hospital information system. Stored data give all the histories of clinical activities in a hospital, including accounting information, laboratory data and electronic patient records. Due to the traceability of all the information, a hospital cannot function without the information system.

However, reuse of the stored data has not yet been discussed in details, except for laboratory data and accounting information to which OLAP methodologies are applied. Data mining approach just started ten years ago [3,4].

In this paper, we first propose a scheme for innovation of hospital services based on data mining. Then, based on this scheme, we applied data mining techniques to data extracted from hospital information systems. The results show several interesting results, which suggests that the reuse of stored data will give a powerful tool to improve the quality of hospital services.

The paper is organized as follows. Section 2 proposes a general framework on innovation of hospital services based on data mining. Section 3 briefly explains how hospital information system works, which is a background on this study.

* This research is supported by Grant-in-Aid for Scientific Research (B) 21300052 from Japan Society for the Promotion of Science (JSPS).

T.-h. Kim et al. (Eds.): FGIT 2011, LNCS 7105, pp. 1–16, 2011.

Section 4 gives explanations on data preparation and mining process. Section 5 shows the results of visualization of hospital activities by using HIS data. Section 6 shows clustering-based analysis of similarities between divisions. Section 7 applies trajectories mining technique to temporal analysis the number of orders. Finally, Section 9 concludes this paper.

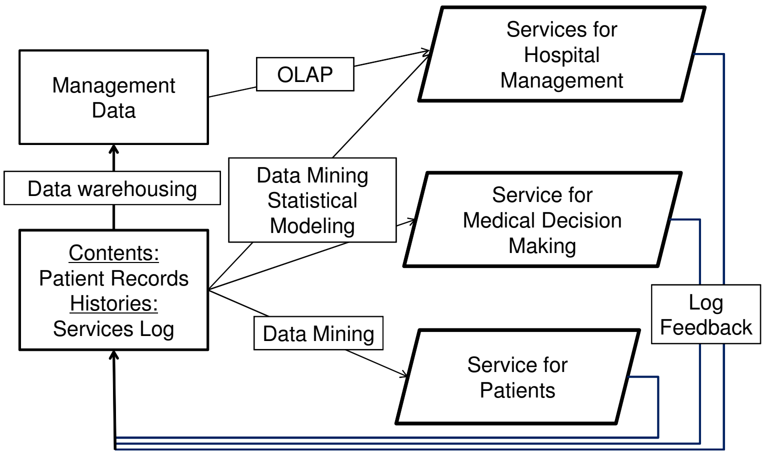

Fig. 1. Service-Oriented Hospital Management

2 Our Goal: Data-Mining Based Hospital Services

Figure 1 shows our goal for hospital services, which consists of the following three layers of hospital management: services for hospital management, devices for medical staff and services for patients. Data mining in hospital information system plays a central role in achieving these layers.

The first layer is called services for patients. It supports the improvement of healthcare service delivery for patients. This is a fundamental level of healthcare services in which medical staff directly gives medical services to the patients. Patient records and other results of clinical examinations support the quality of this service. The second layer is called services for medical staff. It supports decision making of medical practitioner. Patient histories and clinical data are applied to data mining techniques which gives useful patterns for medical practice. Especially, detection of risk of patients, such as drug adverse effects or temporal status of chronic diseases will improve the qualities of medical services. The top layer is called services for hospital management. This level is achieved by capturing global behavior of a hospital: the bridging between microscopic behavior of medical staff and macroscopic behavior of hospital is very important to deploy medical staff in an optimal way for improving performance of the hospital.

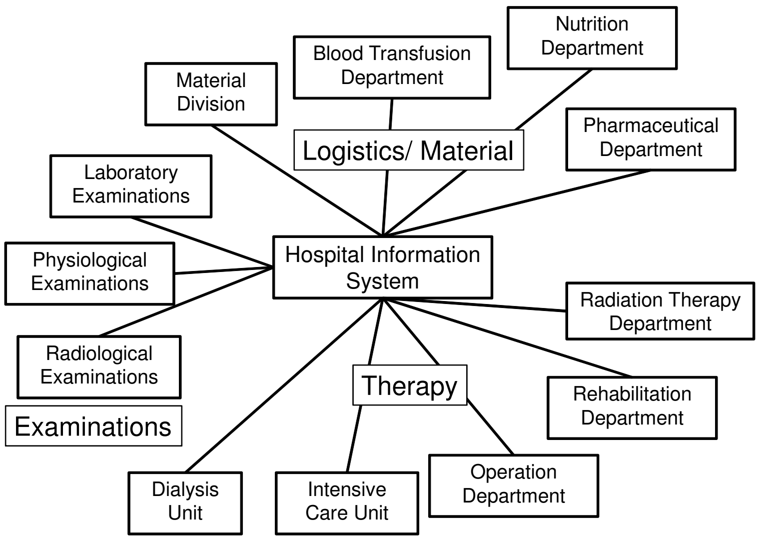

Fig. 2. Hospital Information System in Shimane University

3 Background

3.1 Hospital Information System: Cyberspace in Hospital

On the other hand, clinical information have been stored electronically as a hospital information system (HIS). The database stores all the data related with medical actions, including accounting information, laboratory examinations, and patient records described by medical staffs. Incident or accident reports are not exception: they are also stored in HIS as clinical databases. For example, Figure 2 shows the structure of the HIS in Shimane University Hospital. As shown in the figure, all the clinical inputs are shared through the network service in which medical staff can retrieve their information from their terminals [1,5].

Since all the clinical data are distributed stored and connected as a large-scale network, HIS can be viewed as a cyberspace in a hospital: all the results of clinical actions are stored as "histories". It is expected that similar techniques in data mining, web mining or network analysis can be applied to the data. Dealing with cyberspace in a hospital will give a new challenging problem in hospital

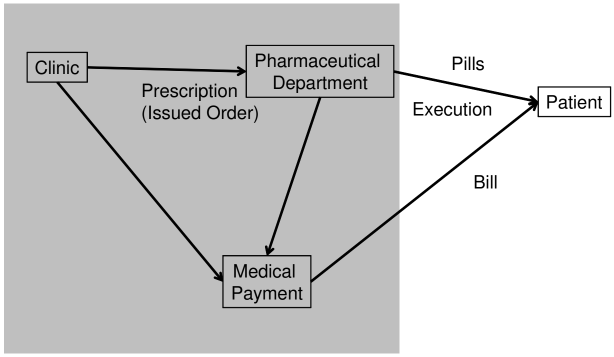

Fig. 3. Workflow of Prescription Order

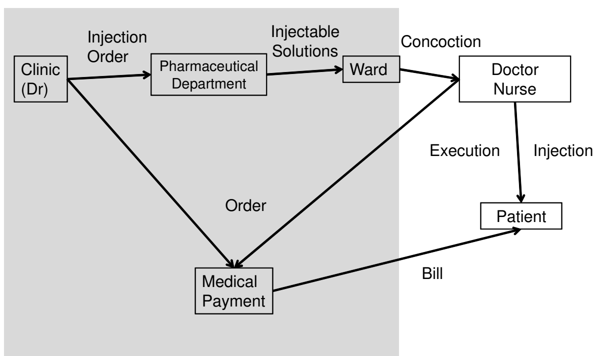

Fig. 4. Workflow of Injection Order

management in which spatiotemporal data mining, social network analysis and other new data mining methods may play central roles [8].

3.2 Basic Unit in HIS: Order

The basic unit in HIS is an "order", which is a kind of document or message which conveys an order from a medical practitioner to others. For example, prescription can be viewed as an order from a doctor to a pharmacist and an prescription order is executed as follows.

1. Outpatient Clinic
2. A prescription given from a doctor to a patient
3. The patient bring it to medical payment department
4. The patient bring it to pharmaceutical department

5. Execution of order in pharmacist office
6. Delivery of prescribed medication
7. Payment

The second to fourth steps can be viewed as information propagation: thus, if we transmit the prescription through the network, all the departments involved in this order can easily share the ordered information and execute the order immediately. This also means that all the results of the prescription process are stored in HIS.

Figure 3 depicts the workflow of prescription between doctors, pharmacologist, patients and pay desk. For comparison, Figure 4 shows that of injection.

These sharing and storing process, including histories of orders and their results, are automatically collected as a database: HIS can also be viewed as a cyberspace of medical orders.

4 Data Preparation and Analysis

4.1 DWH

Since data in hospital information systems are stored as histories of clinical actions, the raw data should be compiled to those accessible to data mining methods. Although this is usually called "data warehousing", medical data warehousing is different from conventional ones in the following three points. First, since hospital information system consists of distributed and heterogenous data sources. Second, temporal management is important for medical services, so summarization of data should include temporal information. Third, compilation with several levels of granularity is required. In this paper, we focus on the number of orders to capture temporal global characteristics of clinical activities, whose scheme is given as Figure 5. Here, data warehousing has two stages: first, we compile the data from heterogenous data sets with a given focus as the first DWH. Then, we split the primary DWH into two secondary DWHs: contents and histories. In this analysis, we focus on the latter DWH and we count the number of orders within a given temporal section. Data mining process is applied to the data sets generated from such obtained DWH.

4.2 Mining Process

We propose temporal data mining process, which consists of the following three steps, shown in Figure 6. We count temporal change of #orders per hour or per days in the second DWH. Then, since each order can be viewed as a temporal sequence, we compare these sequences by calculating similarities. Using similarities, clustering, multidimensional scaling (MDS) , and other similarity-based method are applied.

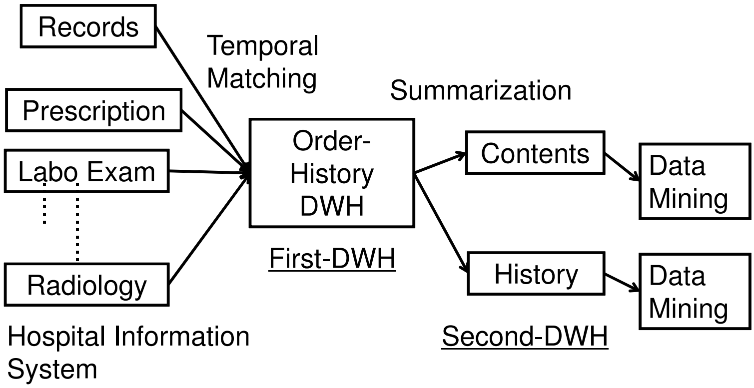

Fig. 5. Data warehousing

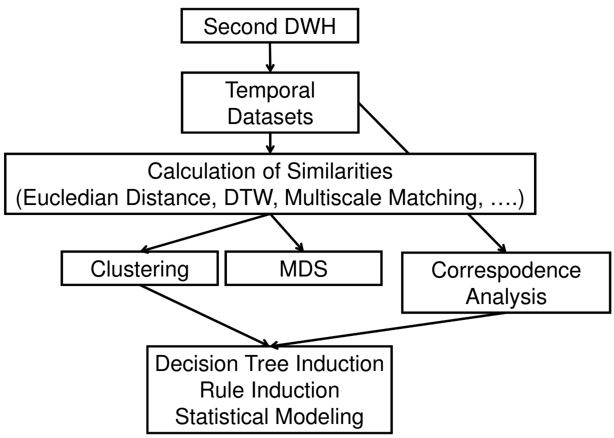

Fig. 6. Mining Process

4.3 Mining Methods

In this paper, the whole analysis is conducted by R2-13-1, except for trajectories mining. As for MDS and correspondence analysis, default methods were applied and for clustering, the ward method was applied. The library mvpart was used for decision tree mining. Multiscale matching was used for preprocessing for trajectories mining [6].

5 Visualizing Hospital Actions from Data

Let us show the primitive mining results of HIS. Table 1 shows the averaged number of each order during the same period. Although these values do not

Table 1. Averaged Number of Orders per day (Oct, 2006 to Feb, 2011)

	Outpatient			Ward		
	Average	Per Patient	Percentage	Average	Per Patient	Percentage
Prescription	403.010	0.457	11%	259.855	0.626	7%
Labo Exam	287.953	0.326	8%	221.373	0.533	6%
Phys Exam	78.595	0.089	2%	45.620	0.109	1%
Radiology	218.781	0.248	6%	56.352	0.135	2%
Operation			0%	12.853	0.031	0%
Transfusion	1.168	0.001	0%	11.089	0.027	0%
Meal			0%	186.293	0.448	5%
Pathology	21.909	0.025	1%	16.737	0.040	0%
Injection	84.414	0.096	2%	469.394	1.131	13%
Reservations	663.192	0.752	18%	77.465	0.186	2%
Documents	454.485	0.516	13%	156.966	0.378	5%
Nursery	11.087	0.0126	0%	846.334	2.0395	24%
Process	422.718	0.480	12%	99.0420	0.239	3%
Records	951.955	1.080	26%	963.017	2.321	28%
Rehabilitation	3.048	0.0035	0%	3.065	0.0074	0%
In/Out			0%	55.403	0.134	2%
Total	3602.315	4.088	100%	3480.856	8.388	100%
#Patients	881.209			414.972		

remove the effects of holidays, all the characteristics reflect those shown in Figure 2. Patient records and Nursing cares are the major part of orders (39%). Prescription, reservation of clinics, injection are top three orders in the hospital.

Table 2. Time Differences between Ordered and Performed Date (Oct, 2010)

	Average (Days)	Median (Days)
Prescription	0	0
Laboratory Exam.	28.55	14
Physiological Exam.	29.97	7
Radiology	29.22	9
Transfusion	8.897	6
Pathology	-0.003	0
Injection	9.799	2
Reservation	41.85	28
Nursery	1.397	0
Process	-0.0003 0	
Rehabilitation	0	0

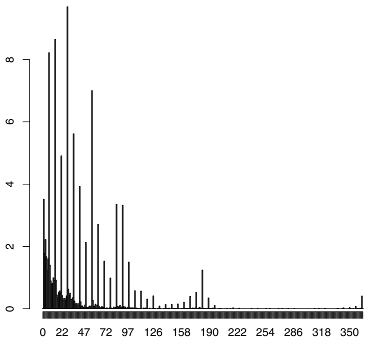

Fig. 7. Distribution of Differences between Executed and Issued Dates for Reservation

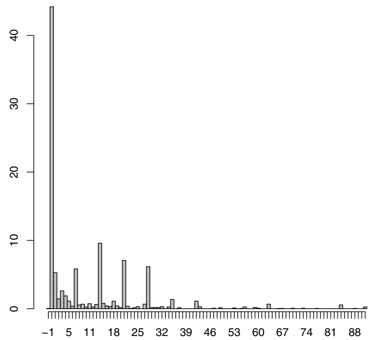

Fig. 8. Distribution of Differences between Executed and Issued Dates for Laboratory Examinations

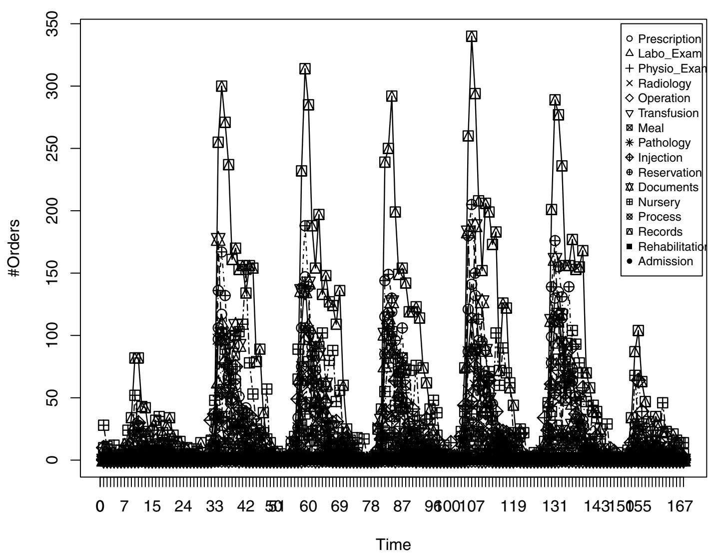

Fig. 9. Trends of Number of Orders (June 1 to 6, 2008)

Table 2 gives the statistics of time differences between ordered and performed date. For example, laboratory examinations are performed 28.55 days (averaged) and 14 days (median) after they are ordered.

Figure 7 shows the histogram (distribution) of time-difference on reservations. The peaks are given as 14, 21 , 28, 56, 90 and 180, which reflects the follow up period frequently used by clinicians.

Figure 8 shows the histogram (distribution) of time-difference on injections. The peaks are given as 7, 14, 21 and 28, which reflects the follow up period frequently used by clinicians.

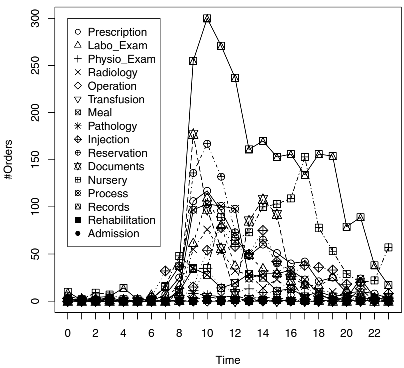

Fig. 10. Trends of Number of Orders (June 2, 2008)

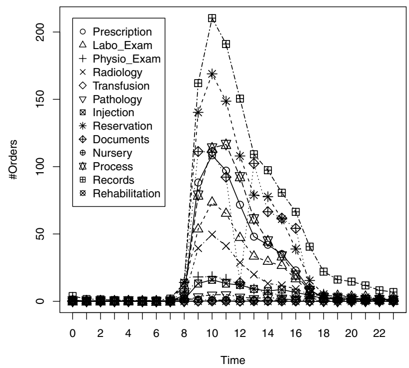

Fig. 11. Trends of Number of Orders of Outpatient Clinic (2010)

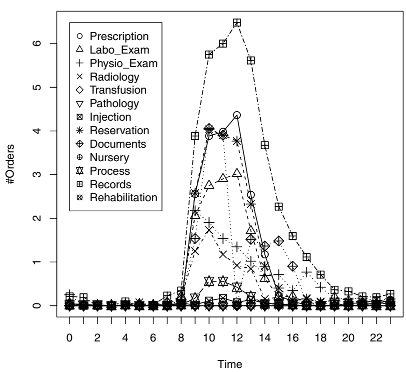

Fig. 12. Trends of Number of Orders of Cardiology (2010)

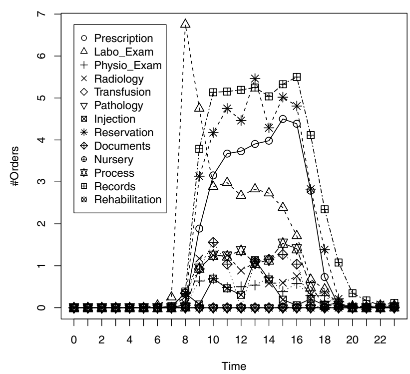

Fig. 13. Trends of Number of Orders of Rheumatology (2010)

5.1 Temporal Trend of #Orders

Although the above tables overview the total behavior of the hospital, we can also check the temporal trend of each order as shown in Figure 9 and 10. The former figure depicts the chronological overview of the number of each order from June 1 to 7, 2008, and the latter shows that of June 2, 2008. Vertical axes denote the averaged number of each order, classified by the type of orders. Horizonal axis give each time zone. The plots show the characteristics of each order. For example, the number of records of doctors has its peak in 11am, which corresponds to the peak of outpatient clinic, whose trend is very similar to reservation of outpatient clinic. The difference between these two orders is shown in 1pm to 5pm, which corresponds to the activities of wards.

The trends can capture the differences between division. Figures 12 and 13 show those in orders of outpatient clinics of cardiology and rheumatology on Tuesday. Compared with the total trends in outpatient clinic shown in Figure 11, those of cardiology are much closer than those of rheumatology. These results show that we can measure and visualize the dynamics of clinical activities in the university hospital by exploratory methods. If we can detect some abnormalities different from the usual behavior in these measurements, this may give some knowledge about risks in the clinical activities. Thus, it is highly expected that data mining methods, especially spatiotemporal data mining techniques play crucial roles in analyzing data in hospital information system and understanding the dynamics of hospital [8,9].

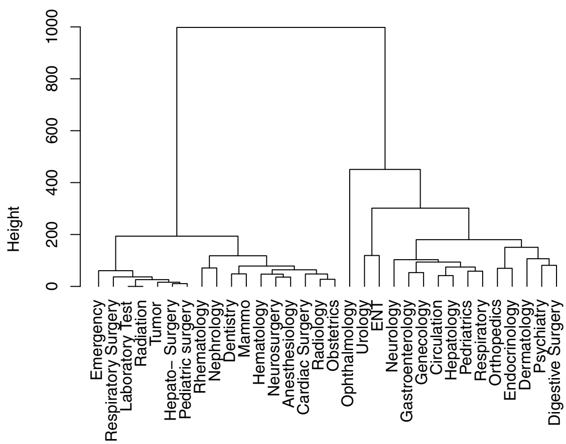

Fig. 14. Clustering of Chronological Patterns of Divisions

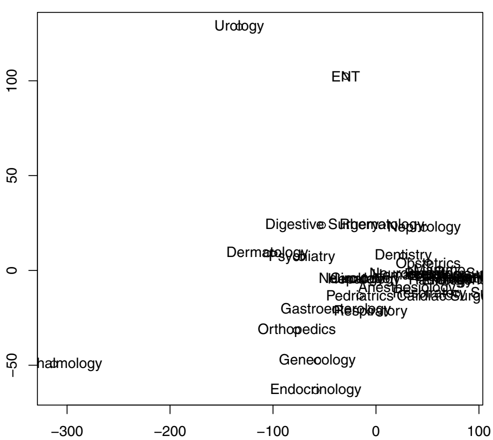

Fig. 15. MDS Results of Chronological Patterns of Divisions

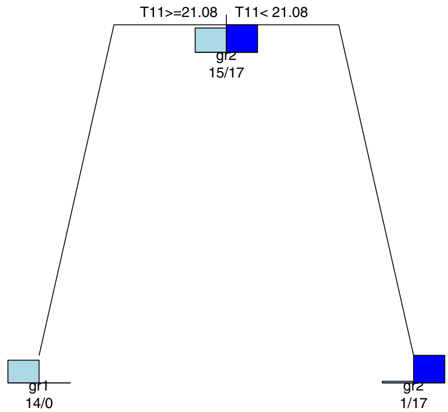

Fig. 16. Decision Tree induced by Chronological Patterns of # Total Orders

6 Clustering of Temporal Trends of Divisions

If we regard temporal trends as sequences for divisions in a hospital, we can classify divisions by using clustering methods.[1]

Figure 14 shows the grouping of divisions of hospitals by using Ward's method, in which the metrics are calculated from the chronological trends of the number of total orders in outpatient clinic. Figure 15 shows the results of multidimensional scaling, which gives a two-dimensional complementary view of similarities among divisions.

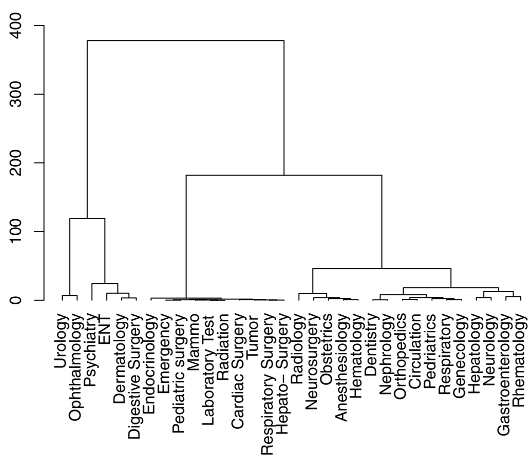

Fig. 17. Clustering of Division with T12

[1] Attributes may not be dependent, thus the usage of all the attributes will give us overfitted classification. Thus, feature selection should be considered: an approach to this problem is discussed in [7].

6.1 Decision Tree Induction

With the labels obtained, decision tree induction was applied to the data set. Figure 16 shows the result where only selected attribute is "T11", Tuesday 11pm. Thus, this time zone is the best attribute for classification of two major groups.

With this selected attribute, clustering method was refined as shown in Figure 17. Intuitively, this group is much better than the former one. Then, decision tree can be applied to data with newly generated groups. These results and repetitive process for temporal data mining is proposed in [7].

7 Analysis of Trajectory of #Orders

If we take two variables of each orders shown in Figure 10, then we can depict the trend of two attributes as a trajectory, as shown in Figures 18 and 19.

Tsumoto and Hirano proposes a clustering method of trajectories, which calculates dissimilarity measures via multiscale matching and apply clustering methods to trajectories by using the dissimilarities between trajectories [6]. By applying this method to the data shown in Section 5, the dendrogram shown in Figure 20 was obtained. An example of clusters are shown in Figure 21, which gives a pattern where orders are given both in wards and outpatient clinics. The other one gives a pattern where orders are provided mainly in the wards. A typical example in the first cluster is shown in Figure 18, while one in the second cluster is in Figure 19.

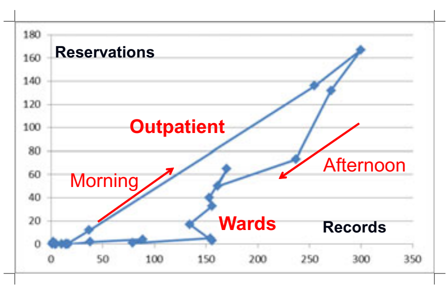

Fig. 18. Trajectory between #Reservations and #Records (June 2, 2008)

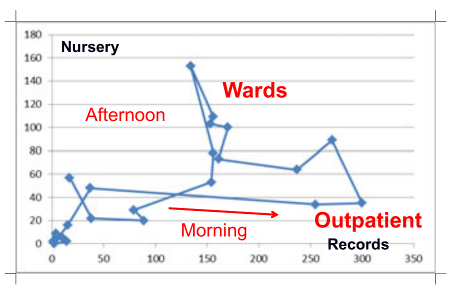

Fig. 19. Trajectory between #Nursery Orders and #Records (June 2, 2008)

The next step is to introduce three-dimensional trajectories mining proposed in [2], although selection of three variables plays an important role in making efficient classifications. It will be our future work to apply this method to the data.

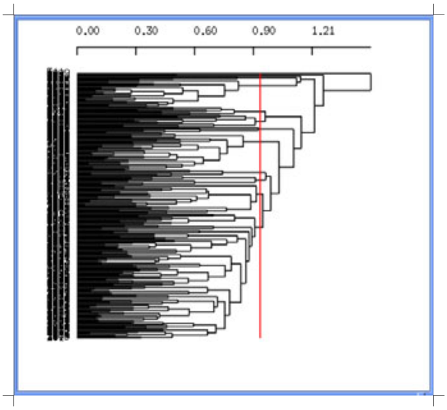

Fig. 20. Dendrogram of Trajectories (June 1 to 6, 2008)

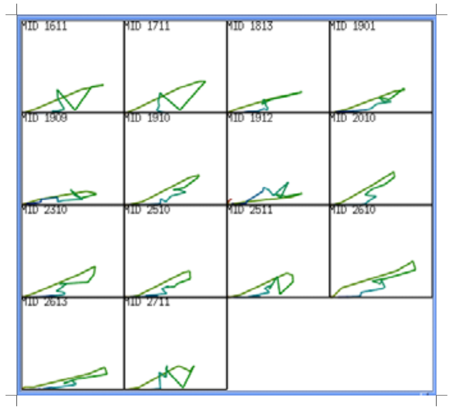

Fig. 21. Cluster No.1 (June 1 to 6, 2008)

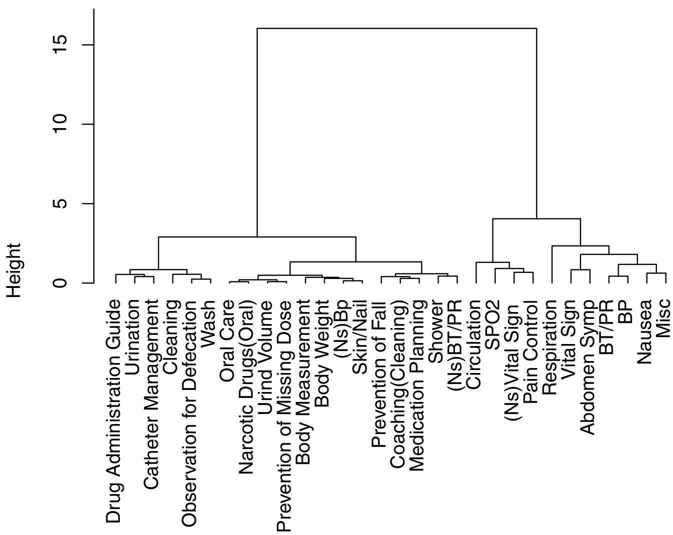

Fig. 22. Clustering Results of Nursing Orders (Lung Cancer)

8 Characterization of Nursing Orders

Finally, we focus on one disease, say lung cancer and count the nursing orders during the stay of each patient and regard chronological change of each order as a temporal sequence. Figures 22 to 24 show the results of clustering, MDS and correspondence analysis of nursing orders with respect to #orders.

Clustering results gave two major groups: one included the orders indispensable to this disease and the other included those which are rather specific to the status of each patient (Figure 28). MDS gave further classification of the

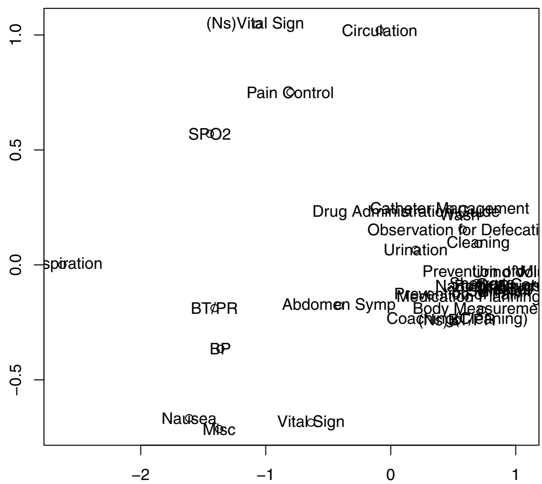

Fig. 23. MDS Results of Nursing Orders (Lung Cancer)

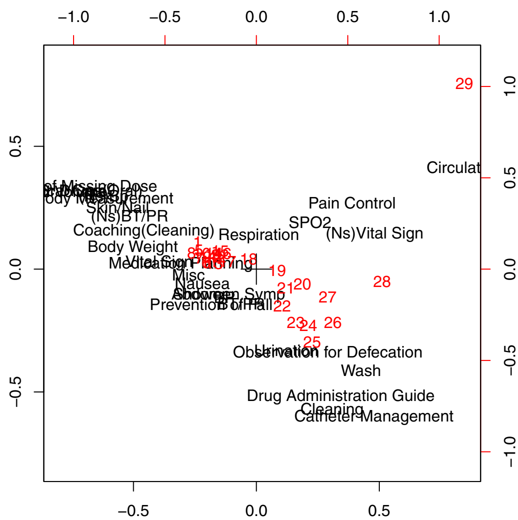

Fig. 24. Results of Correspondence Analysis of Nursing Orders (Lung Cancer)

first group into the following three subgroups: (1) core orders (Figure 25), (2) core and indispensable orders (Figure 27), both of which are not influenced by patients' status, and (3) core, indispensable orders which are influenced by their status (Figure 27). These results show that nursing orders can be classified by the proposed process, the deviation of this classification may give important information for a clinical action.

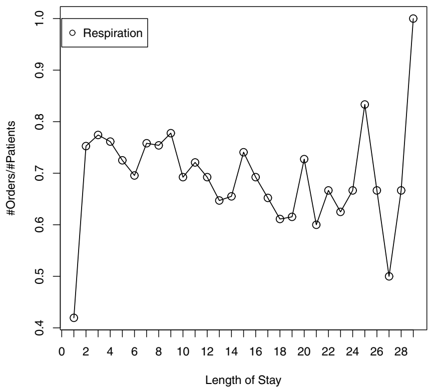

Fig. 25. The Main Nursing Orders

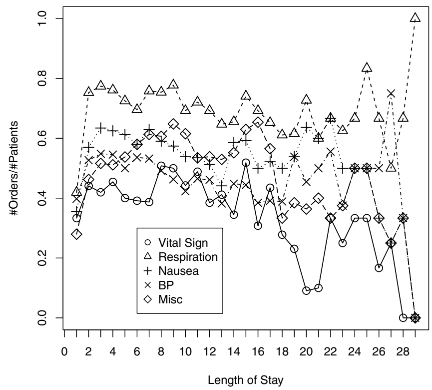

Fig. 26. Core orders which are not influenced by patients' status

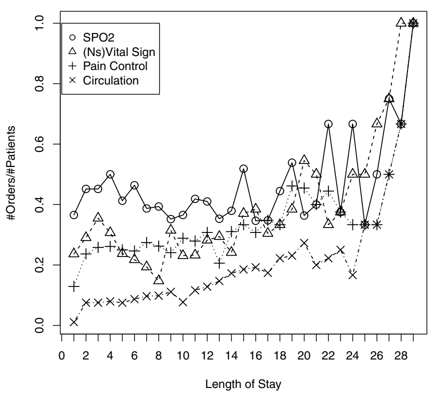

Fig. 27. Core Orders which may be influenced by patients' status

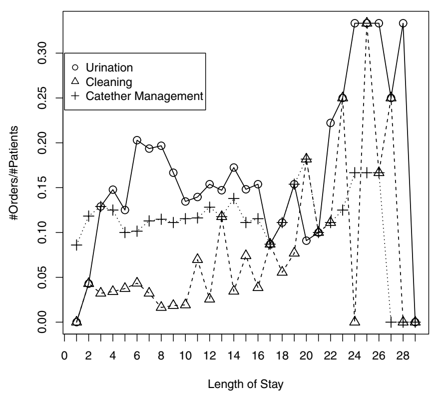

Fig. 28. Patient-specific Nursing Orders

9 Conclusions

In this paper, we propose a general framework on innovation of hospital services based on data mining. Then, we applied several data mining techniques to data extracted from HIS in order to capture the characteristics of clinical activities in hospital as follows. First, we shows the chronological overview of hospital activities: periodical behavior of the number of orders can be viewed as a "life-cycle" of a hospital. Secondly, we extracted a pattern of long-term follow up patients with respect to the number of orders by using HIS data. Section 6 shows the results of clustering analysis of divisions with respect to chronological change of total orders. Thirdly, we applied trajectories mining technique to temporal analysis the number of orders. The results of clustering analysis gave two groups of clinical actions. The one was a pattern where orders are given both in wards and outpatient clinics. The other one was a pattern where orders are provided mainly in the wards. Finally, we applied similarity-based analysis to temporal trends of the numbers of nursing orders for lung cancer. The results showed that nursing orders are automatically classified into two major categories, "disease-specific" and "patient-specific" ones. Furthermore, the former one is classified into three subcategories, according to the temporal characteristics.

This paper is a preliminary approach to data mining hospital management towards a innovative process for hospital services. More detailed analysis will be reported in the near future.

References

1. Hanada, E., Tsumoto, S., Kobayashi, S.: A "Ubiquitous Environment" through Wireless Voice/Data Communication and a Fully Computerized Hospital Information System in a University Hospital. In: Takeda, H. (ed.) E-Health 2010. IFIP AICT, vol. 335, pp. 160–168. Springer, Heidelberg (2010)
2. Hirano, S., Tsumoto, S.: Multiscale comparison and clustering of three-dimensional trajectories based on curvature maxima. International Journal of Information Technology and Decision Making 9(6), 889–904 (2010)
3. Tsumoto, S.: Knowledge discovery in clinical databases and evaluation of discovered knowledge in outpatient clinic. Information Sciences (124), 125–137 (2000)
4. Tsumoto, S.: G5: Data mining in medicine. In: Kloesgen, W., Zytkow, J. (eds.) Handbook of Data Mining and Knowledge Discovery, pp. 798–807. Oxford University Press, Oxford (2001)
5. Tsumoto, S., Hirano, S.: Risk mining in medicine: Application of data mining to medical risk management. Fundam. Inform. 98(1), 107–121 (2010)
6. Tsumoto, S., Hirano, S.: Detection of risk factors using trajectory mining. J. Intell. Inf. Syst. 36(3), 403–425 (2011)
7. Tsumoto, S., Hirano, S., Tsumoto, Y.: Clustering-based analysis in hospital information systems. In: Proceedings of GrC 2011. IEEE Computer Society (2010)
8. Tsumoto, S., Hirano, S., Tsumoto, Y.: Towards data-oriented hospital services: Data mining-based hospital management. In: Fan, W., Hsu, W., Webb, G.I., Liu, B., Zhang, C., Gunopulos, D., Wu, X. (eds.) ICDM Workshops, pp. 1076–1083. IEEE Computer Society (2010)
9. Tsumoto, S., Hirano, S., Tsumoto, Y.: Information reuse in hospital information systems: A data mining approach. In: IRI, pp. 172–176. IEEE Systems, Man, and Cybernetics Society (2011)