Maintenance of Domain Knowledge for Nursing Care using Data in Hospital Information System

Haruko Iwata, Shoji Hirano and Shusaku Tsumoto

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

Abstract. Schedule management of hospitalization is important to maintain or improve the quality of medical care and application of a clinical pathway is one of the important solutions for the management. Although several kinds of deductive methods for construction for a clinical pathway have been proposed, the customization is one of the important problems. This research proposed an inductive approach to support the customization of existing clinical pathways by using data on nursing actions stored in a hospital information system. Since hospital data include temporal trends of clinical symptoms and medical services, we can discover not only knowledge about temporal evolution of disease, but also one about medical practice from hospital information system. This paper proposes temporal data mining process and applied the method to capture temporal knowledge about nursing practice. The results show that the reuse of stored data will give a powerful tool for management of nursing schedule and lead to improvement of hospital services.

Keywords: temporal data mining; clustering; multidimensional scaling; hospital information system; visualization

1 Introduction

Twenty years have passed since clinical data were stored electronically as a hospital information system (HIS)[1–3]. 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. All the clinical inputs are shared through the network service in which medical staff can retrieve their information from their terminals [4, 3].

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 management in which spatiotemporal data mining, social network analysis and...
other new data mining methods may play central roles[5, 2]. This paper proposes a temporal data mining method to maintain a clinical pathway used for schedule management of clinical care. Since the log data of clinical actions and plans are stored in hospital information system, these histories give temporal and procedural information about treatment for each patient. The method consists of the following four steps: first, histories of nursing orders are extracted from hospital information system. Second, orders are classified into several groups by using clustering and multidimensional scaling method. Third, by using the information on groups, feature selection is applied to the data and important features for classification are extracted. Finally, original temporal data are split into several groups and the first step will be repeated. The method was applied to a dataset whose patients had an operation of cataracta. The results show that the reuse of stored data will give a powerful tool for maintenance of clinical pathway, which can be viewed as data-oriented management of nursing schedule.

The paper is organized as follows. Section 2 briefly explains clinical pathway, which is schedule management of clinical process. Section 3 gives explanations on data preparation and mining process. Section 4 gives application of temporal data mining process to analysis of nursing orders, which will lead to revision of clinical pathway. Finally, Section 5 concludes this paper.

2 Background: Clinical Pathway

Since several clinical actions should be repeated appropriately in the treatment of a disease, schedule management is very important for efficient clinical process[8, 9]. Such a style of schedule management is called a clinical pathway. Such each pathway is deductively constructed by doctors or nurses, according to their experiences. For example, Table 1 illustrates a clinical pathway on cataracta in our university hospital. The whole process of admission will classified into three period: preoperation, operation and post-operation. The preoperation date is denoted by -1 day, and operation date is by 0 day. BT/PR denotes Body Temperature and Pulse Rate, BP denotes Blood Pressure.

3 Data Preparation and Analysis

3.1 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 heterogeneous 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

1 Application of ordinary statistical methods are shown in [6, 7].
Table 1. An Example of Clinical Pathway

<table>
<thead>
<tr>
<th>Preoperation</th>
<th>Operation</th>
<th>Postoperation</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1day</td>
<td>0day</td>
<td>1day</td>
</tr>
<tr>
<td>2day</td>
<td>3day</td>
<td>4day</td>
</tr>
<tr>
<td>BT/PR</td>
<td>BT/PR</td>
<td>BT/PR</td>
</tr>
<tr>
<td>BP</td>
<td>BP</td>
<td>BP</td>
</tr>
<tr>
<td>Nausea</td>
<td>Nausea</td>
<td>Nausea</td>
</tr>
<tr>
<td>Vomitting</td>
<td>Vomitting</td>
<td>Vomitting</td>
</tr>
</tbody>
</table>

Coaching Coaching Coaching Coaching Coaching

Pain Pain Pain Pain Pain Pain

Notations. BT/PR: Body Temperature/Pulse Rate BP: Blood Pressure

number of orders to capture temporal global characteristics of clinical activities, whose scheme is given as Fig 1. Here, data-warehousing has two stages: first, we compile the data from heterogeneous datasets with a given focus as the first DWH. Then, from this DWH, we split the primary DWH into two secondary DWHs: contents DWH and histories DWH. In this analysis, we focus on the latter DWH and we count the number of orders with a given temporal section, compiled into this DWH. Data mining process is applied to the generated data sets from this DWH.

![Fig. 1. Datawarehousing](image)

### 3.2 Mining Process

Except for the basic process, we will propose temporal data mining process, which consists of the following three steps, shown in Figure 2. We count tem-
poral 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 [10], multidimensional scaling MDS, and other methods based on similarities are applied. In this paper, all the analysis is conducted by R2-15-1.

3.3 Clinical Pathway Maintenance Process

Figure 3 shows the process for maintenance of a clinical pathway. The right cycle shows the repetition of procedure of temporal data mining process proposed in Figure process2. The procedure will be terminated when grouping becomes stable and a pathway based on the classification will be constructed. If a clinical pathway is apriori given, it will be compared with the induced pathway.

4 Construction of Clinical Pathway

Here we apply the above temporal mining process to revision and construction of clinical pathway.

4.1 Data Preparation

We focused on histories of patients who were admitted to the hospital for operations of cataracta. The number of patients is 121 in 2010. For this disease, the clinical pathway mentioned in the above section is used to optimize the schedule of treatment. We counted the nursing orders during the stay of each patient and regard chronological change of each order as a temporal sequence.

2 The process follows the method proposed by Tsumoto [11]
4.2 First Cycle: Grouping

In the first cycle, clustering, MDS and correspondence analysis were applied as the second step in the temporal data mining process. Figure 4 to 6 show the results of clustering, MDS and correspondence analysis of nursing orders with respect to #orders.

Clustering results give two major groups: one includes the orders indispensable to this disease and the other includes those which are rather specific to the status of each patient, except for preoperation instruction. MDS gives further classification of the first group into three subgroups and the second one into two subgroups. The former three subgroups consist of vital signs (BP, BT and PR), body care (Coaching and Wash), and watchlist (Eye Symptoms and Nausea). The latter two groups consist of preoperation instruction and other symptoms which may specific to the status to the patients.

On the other hand, correspondence analysis did not give good information about correspondence between length of stay and nursing orders. This may be because the length of stay was too short (within a week) and all the nursing orders were well scheduled due to the application of clinical pathway.

The results also give another interesting observation: comparison of the above results with the existing clinical pathway shows that the pathway lacks two orders, Wash and Coach, although their temporal patterns are very similar to other orders indispensable to the treatment. So, they should be included into clinical pathway. In this way, these method can be used to evaluate the existing pathway for a disease and revise it.

4.3 First Cycle: Feature Selection

With the labels obtained, decision tree induction[12] was applied to the dataset. Figure 7 shows the result where only selected attribute is “d1”, the first day of
postoperation period. In the induced tree, \( d1 < 23 \) to the right means that if the number of a given order is less than 23, the order belongs to the right class (which is specific to the status of patients.) In other words,

Table 2 shows the list of information gain for the labels. These values show that for \( d1 \), \( d2 \) and \( d3 \) give complete classification for the labels, and \( preoperation \), \( postoperation \), \( d4 \) and \( d5 \) do not have enough information for complete classification. Thus, the data should be split into three groups: the first interval is
Fig. 6. Results of Correspondence Analysis of Nursing Orders (Cataracta)

preoperation and operation, the second one is $d_1$ to $d_3$, and the final interval is $d_4$ and $d_5$. In the next cycle, the same procedures will be applied to each subset of data.

4.4 Second Cycle: Grouping

Figure 8 and 9 show the results for preoperation and operation (-1 to 0 days), which gives much clearer results. Here we can see three clusters, compared with
Table 2. Information Gain for Each Attribute

<table>
<thead>
<tr>
<th>Node</th>
<th>Information Gain</th>
</tr>
</thead>
<tbody>
<tr>
<td>preoperation &lt; 8 to the right</td>
<td>2.103810</td>
</tr>
<tr>
<td>operation &lt; 23 to the right</td>
<td>3.534545</td>
</tr>
<tr>
<td>d1 &lt; 23 to the right</td>
<td>10.08</td>
</tr>
<tr>
<td>d2 &lt; 21 to the right</td>
<td>10.08</td>
</tr>
<tr>
<td>d3 &lt; 39 to the right</td>
<td>10.08</td>
</tr>
<tr>
<td>d4 &lt; 27.5 to the right</td>
<td>6.48</td>
</tr>
<tr>
<td>d5 &lt; 18.5 to the right</td>
<td>6.48</td>
</tr>
</tbody>
</table>

Fig. 8. Clustering Results of Nursing Orders (Cataracta: -1 to 0 days)

Next, Figure 10 and 11 show the results for 1 to 3 days. Here the results are very similar to those obtained from the whole data. However, Coaching is similar to BP, and Wash is close to BT/PR and Nausea/Vomitting. Pain and Eye symptoms. Finally, Figure 12 and 13 show the results for 4 and 5 days. Here the grouping is slightly different from the second one and Coaching and BP are not similar to BT/PR, Wash, Nausea/Vomitting, Pain and Eye symptoms.
In either case, the labels were provided by the grouping methods for each period, and decision tree induction was applied. Each attribute gave complete classification, so further data partition was not needed, and the repetition of grouping and feature selection was terminated.
4.5 Construction of Pathway

From those results, coaching and wash, whose chronological characteristics are similar to the orders indispensable to the treatment of cataracta, were not included in the existing pathway, but should be added to the pathway. Furthermore, coaching and wash should be treated as postoperation follow-up and routine process, respectively. Thus, the pathway was revised as shown in Table 3.
In summary, temporal data mining of nursing may be useful for to construct clinical pathway for a disease where a pathway is not introduced and to revise the existing pathway.

5 Conclusions

In this paper, we propose a general framework on innovation of hospital services based on temporal data mining process. This process can be called similarity-based visualization approach in which similarity-based methods, such as clustering and multidimensional scaling (MDS) and correspondence analysis. We applied the process to datasets of #nursing orders for cases for operation of cataracta where clinical pathway has been introduced. By using Clustering and MDS, we obtained two major groups in the nursing orders: ones were indispensable to the treatment, and the others were specific to the status of patients. Then, in the step for feature selection, the first day of postoperation could be be viewed as a threshold in the original datasets. Thus, periods before and after operation should be dealt as independent datasets. Repeating these steps, we could characterize the temporal aspects of nursing orders, and then found missing information in the existing pathway. 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.

Acknowledgments This research is supported by Grant-in-Aid for Scientific Research (B) 24300058 from Japan Society for the Promotion of Science(JSPS).
Table 3. Revised Clinical Pathway for Cataracta

<table>
<thead>
<tr>
<th>Preoperation Operation</th>
<th>Postoperation</th>
</tr>
</thead>
<tbody>
<tr>
<td>-1day</td>
<td>0day</td>
</tr>
<tr>
<td>BT/PR</td>
<td>BT/PR</td>
</tr>
<tr>
<td>BP</td>
<td>BP</td>
</tr>
<tr>
<td>Wash</td>
<td>Wash</td>
</tr>
<tr>
<td>Nausea</td>
<td>Nausea</td>
</tr>
<tr>
<td>Vomitting</td>
<td>Vomitting</td>
</tr>
<tr>
<td>Coaching</td>
<td>Coaching</td>
</tr>
</tbody>
</table>

Pain Pain Pain Pain Pain Pain

Preoperation Instruction

Notations. BT/PR: Body Temperature/Pulse Rate. BP: Blood Pressure.
Eye Symp: Eye Symptoms.

References