3S-1

On Mining Causal Relation From Nursing Record Data

Ahmad Zufri† Kazuto Hachiya† Antonia Scheidel†, ‡ Kazuo Hashimoto†
†Graduate School Information Science, Tohoku University
‡Computational Linguistics Department. Saarland University

1 Abstract

In the geriatric facilities, health condition of elderly people, are recorded regularly by nurses. There is a need to analyze nursing records together with incident reports, called HIYARI-HATTO reports, to detect possible risks in the daily life of the elderly. However, it is difficult for service providers to analyze such records because the amount of nursing record data is increasing significantly in addition to the daily burden of elderly care. In the hope of facilitating such analysis, the authors are analyzing nursing records data using topic detection technology, to enable efficient discovery of health status changes. This paper proposes an analysis of nursing record data with existing two causal mining approaches; Sequential Pattern Mining Approach and Dynamic Bayesian Network Approach.

2 Introduction

The nursing record data used in this research is from Sakuranomori geriatric facility. This data contains the examination results from each patient and also accident data in that facilities. Because patients are elderly, there were many accidents and health problems happened at that facility. As a follow up from the accident, nurses in that facility has to make an accident report for each accident. This becomes difficult thing because several reasons. First is lots of accidents happened at that facilities. The second is the amount of data before the accident that need to be analyzed is considerably a lot. Because for 1 day, there might be 2-3 records. And also there is another problem about time span of data that need to be analyzed. In what time span the data needs to be analyzed to find the causal of the event or accident. Another problem is the data analysis problem. Nursing record data characteristic, which are written in natural language and unstructured form make it become difficult to do automatic analysis. Because every nurse has different written language style and also the word choice is different for each other. These make the automatic analysis become difficult. To overcome these problems, Firstly we extract the main component from nursing record data which are vital sign data, health status data, and mental state and then use the combination of those three elements to find the causal from accident or particular health condition by using sequential pattern approach and DBN approach.

3 Nursing Record Analysis

The main components from nursing record data can be divided into 3 categories data.

Patient Health Topic at time t is the result of topic detection technology which is applied to record at certain time t in the nursing record data.
Let $P = \{p_1, p_2, ..., p_n\}$ be a list patient in the geriatric facilities. Let $L = \{l_1, l_2, ..., l_i\}$ be the set of health condition topics from topic detection technology result. Let $r(t, j)$ be the record of patient j at time t and P_{tjk} is the probability of $r(t, j)$ belong to topic l_k, $l_k \in L$. Then the Patient Health topic of a patient j at time t H_{tj} can be defined as $< P_{tj1}, ..., P_{tji} >$.

Vital Signs at time t which are blood pressure, heart beat rate, and body temperature.

Mental Condition at time t is the patient's emotional condition(positive,neutral,negative), which is extracted from nursing record data at time t [4].

3.1 Input Data

Before the causal relation mining process, the input data need to be adjusted in order to fit our approaches. The adjustments that are made to the input data are:

- Patient Health Topic: For certain time t, we can not directly classify each of health topic data into one topic with the highest probability. Because the probability differences between each topic is very small. In order to simplify this problem, patient health topic data here is adjusted by using clustering method. Assume there is a N number of health status topic. We can regard the data as N dimensional data. We use the K-Means clustering algorithm with euclidan distance with initial cluster N. Then for each H_{tj} we give the label according to our clustering method result.

- Vital Signs: We categorized the vital sign data into some categories. For blood pressure, we divided it into 6 categories (High I, High II, High III, High IV, Normal, Low). For heart beat and body temperature, we divided it into 3 categories (High,Normal,Low). And then we define some classes which are the combination of the three components category.
For example Class A(Blood Pressure=Normal,heart beat=low, temperature=normal).

- Mental Condition: This data needs no adjustment because it is already categorical data.

After doing above steps, input data of our approaches can be defined as follow:

- For each patient P_i, let $H_i(t), M_i(t), V_i(t)$ be health topic label, mental state, and vital sign class of patient P_i at time t. The input sequence for each patient S_i can be define as the a time series of tuple

$< H_i(t), M_i(t), V_i(t) >$. Then, the input of our approaches is $< S_1, ..., S_n >$ where n is the number of patients in geriatric facilities.

3.2 Causal relation mining process

After data preparation step, the next step is the causal relation mining process. In this process, we use 2 approaches.

3.2.1 Sequential Pattern Approach

We use this approach to find the long term causal relation. Because this approach relies on the sequence without concerning about the time difference between one event and another event. For this Approach, we used the prefixspan algorithm [1]. To apply this approach, we define a digitization method to handle our data. Because our data is a set of tuple while prefixspan input data is sequence of symbol. The output of the algorithm is a set of sequential pattern. After that, we store this pattern as the knowledge database. Given the current states of patient, the causal of current states is analyzed by using the result of prefixspan algorithm.

3.2.2 Dynamic Bayesian Network Approach

Dynamic Bayesian Network is graphical model, which represent the dynamic condition in real world that changes over time, by using the direct acrylic graph (DAG). Causal relation between variables in DBN, is represented by direct edge between node [2] [3]. To apply this approach to nursing record data, we do following processes:

- Our data is modeled by the Figure 1. From that picture, we can see that health topic status H depends on mental state M and also vital sign V. This is because the mental state and the mental state usually contain in the examination record, which is a free natural language text at time t. We can say that, the Health status at time t only depend on Health status at time t-1. Vital sign and mental state also only depend directly on the previous state.

- After we define a structure of DBN, we build the complete DBN with the belief. To do that, we have to define 3 parameters to complete the DBN. The first is $Pr(H_t \mid H_{t+1})$, $Pr(V_t \mid V_{t+1})$ and $Pr(M_t \mid M_{t+1})$ which denotes the probability of horizontal dependency between one variable state in time t and another variable in time t+1. The second is $Pr(V_t \mid H_{t+1})$ and $Pr(M_t \mid H_{t+1})$ which denotes the probability of vertical dependency between one variable state in time t with another variable in the same time slice. The last is the initial probability of $Pr(H_0)$, $Pr(V_0)$, and $Pr(M_0)$. We do this process by selecting 5 random patients' data, and then extract the value for H, M and V. An then, the learning process is applied to obtain the parameter of our current DBN. The learning process is done by using The Expectation Maximum (EM) algorithm to maximize the likelihood from joint probability distribution.

- After we get the complete DBN with the belief degree, given the current state $< H_{(t)}, M_{(t)}, V_{(t)} >$, we obtain the causal relation from current state by doing inference in DBN with back propagation.

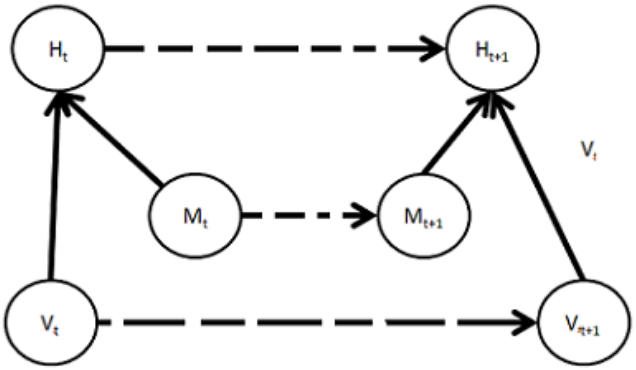

Figure 1: Dynamic Bayesian Network Model

4 Acknowledgements

This Work was supported by Intermediate Nursing Home, Sakuranomori for the nursing record and hiyari-hatto report. This work also supported by Grant-in-Aid for Scientific Research (B) (22360619) from Japan Society for Promotion of Science(JSPS).

5 Conclusion and Future Work

In this paper, we have proposed an analysis of causal mining on nursing record data with current existing causal mining technology, Dynamic Bayesian Network (DBN) approach and the sequential pattern mining approach. The future work of this research is to evaluate how effective these methods to handle the problem in geriatric facilities, especially in assisting the nurse to find the causal relation from a certain accident or health trouble.

References

[1] J.Pei and J.Han and Behzad M.A and J.wang and H.Pinto and Q.Chen and U.Dayal and M.C Shu, "Mining Sequential Patterns by Pattern-Growth:The PrefixSpan Approach", *IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING*, Volume 16, (2004).

[2] Stuart Russel, Peter Norvig "Artificial Intelligence A Modern Approach Second Edition", Prentice Hall, 2002.

[3] V. Mihajlovic, M.Petkovic "Dynamic Bayesian Network, The state of the Art", Website, 2001. http://doc.utwente.nl/36632/1/0000006a.pdf

[4] Antonia Scheidel, Kazuto Hachiya, Ahmad Zufri, Kazuo Hashimoto "On Mental State Detection and Tagging in Nursing Records", *In Proceeding 73th Information Processing Society of Japan Conference*, In Press.