A Framework for Personalized Health Trend Analysis

Guo-Cheng Lan\textsuperscript{a}, Chao-Hui Lee\textsuperscript{a}, Jin-Shang Wu\textsuperscript{d}, Huan-Chung Li\textsuperscript{d}, Shing-Hua Ho\textsuperscript{a}, Vincent S. Tseng\textsuperscript{a,b}

\textsuperscript{a}Department of Computer Science and Information Engineering
National Cheng-Kung University
Tainan, 701, Taiwan, ROC.
rrfoheiay@idb.csie.ncku.edu.tw; lobby@idb.csie.ncku.edu.tw

\textsuperscript{b}Institute of Computer Science and Information Engineering and Institute of Medical Informatics
National Cheng-Kung University
Tainan, 701, Taiwan, ROC.
tsengsm@mail.ncku.edu.tw

\textsuperscript{c}Department of Family Medicine
National Cheng-Kung University
Tainan, 701, Taiwan, ROC.
jins@mail.ncku.edu.tw

\textsuperscript{d}Innovative Digitech-Enabled Applications and Services Institute
Institute for Information Industry
Taipei, 105, Taiwan, ROC.
wiselyli@iii.org.tw

\textsuperscript{e}Innovative Digitech-Enabled Applications and Services Institute
Institute for Information Industry
Kaohsiung, 806, Taiwan, ROC.
hosh@iii.org.tw

Abstract—In recent years, with the development of the economy and the advancement of the national income, personal healthcare has been becoming an important issue. Since the health examination can help people understand their own health conditions clearly, many people make regularly health examinations to avoid missing the best treatment time. So the health examination is an important role for people’s health statuses. However, people only get aware of the result of health examination report but they are not clear whether their health trends are high-risk or not. In this paper, we proposed a novel framework for discovering health risk patterns and the relationships between the health risk patterns and some target diseases by mining historic health examination data. Moreover, a prediction model can be constructed effectively by the analyzed information for the target disease. At last, the physicians can early provide health alerts and medical treatments for people to effectively increase the quality of disease prevention by the analyzed information.

Keywords - data mining; health examination; health risk pattern; disease analysis; prediction model.

I. INTRODUCTION

In recent years, with development of the economy and the advancement of the national income, the health has been becoming an important issue. This is because people more care their own health conditions. According to the published statistics information of Department of Health in Taiwan [11], most of the top-10 death factors are the chronic diseases, such as high blood pressure, diabetes, heart disease. However, those chronic diseases can be easily found and prevented. In addition, since the type of diseases is recently changed, and the age of illness is decreased gradually, the role of health examination is more important [5][7].

Since people want to understand their own health conditions whether they suffer from some chronic diseases or not, they have to do a regular health examination. Through the evaluation, people can avoid missing the best treatment if they do suffer from some diseases. However, under the development prospect of the health examination, it is more important and critical to develop the health-related products and techniques.

After each health examination, persons only know the result from the examination report. If we could observe the historic health examination data, we can obtain the health trends, and those health trends can be analyzed and more useful information can be provided for people. For example, one person’s examined values of cholesterol, triglyceride and blood sugar are not higher than the range of health alert, but the values of ones from historic health examination are increasing gradually. Hence, it is necessary to observe the health trends from persons’ historic health examination data.

With rapid development of data mining technology, traditional medical systems have evolving the intelligent medical systems [3], such as the intelligent auto-diagnosing physiology signal system and the fundamentals of health care system. Therefore, we proposed the framework of
health examination data analysis system which is based on the data mining technology to discover the health trend patterns from people’s historic health examination data.

Moreover, since the relationship between health trend patterns and chronic diseases is less discussed in the past, we further discuss the relationship from the health examination data. Fig. 1 shows the framework of proposed personal health management system.

Figure 1. The framework of proposed personal health analysis system.

In the past, most of the related studies [4][8] only provide the query function and the statistics function of health data, and they detected and diagnosed patients through the clinical method. However, they did not consider advanced information, such as the health trend patterns and the relationship between single examination item or multiple examination ones and diseases. The reason is that the above information can provide more useful information for us, and the obtained information can be used to build the intelligent prediction models to increase the quality of health prevention.

In view of the issues as described above, in this paper, we proposed a novel framework for discovering the health risk patterns and the relationship between health risk patterns and diseases by mining the historic health examination data. The whole processes can be divided into two phases. In the first phase, the data of all health examination items are converted into the specified format. After that, the related data mining techniques are applied to discover a set of described patterns. In the second phase, the set of discovered patterns are used to analyze the relationship of single health examination item or multiple health examination items and diseases. Finally, our proposed health examination data analysis system can be built by the discovered information through the two phases. The main contributions of this paper are listed below.

1. A novel framework for health examination data analysis system is proposed to discover and analyze the health risk patterns of each health examination item.

2. The proposed system can also discover the relationship between health risk patterns and target diseases from the historic health examination data.

II. RELATED WORKS

In this section, we further illustrate the problem definitions and the importance of health examination. Besides, association rules mining, sequential patterns mining and the classification on data mining are briefly reviewed.

A. Problem Definition

In general, a health examination is composed of multiple examination items. By the examined value of each item in the health examination report, people can clearly understand their currently health conditions through the health examination in the hospitals. TABLE I shows that one person’s multiple health examination item results.

<table>
<thead>
<tr>
<th>Item</th>
<th>Item A1</th>
<th>Item A2</th>
<th>Item (N-1)</th>
<th>Item N</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st evaluation</td>
<td>3.2</td>
<td>45</td>
<td>...</td>
<td>97</td>
</tr>
<tr>
<td>2nd evaluation</td>
<td>3.8</td>
<td>30</td>
<td>...</td>
<td>119</td>
</tr>
<tr>
<td>3rd evaluation</td>
<td>5.1</td>
<td>28</td>
<td>...</td>
<td>139</td>
</tr>
<tr>
<td>4th evaluation</td>
<td>4.8</td>
<td>30</td>
<td>...</td>
<td>133</td>
</tr>
<tr>
<td>5th evaluation</td>
<td>5</td>
<td>35</td>
<td>...</td>
<td>157</td>
</tr>
</tbody>
</table>

People only understand their own health conditions after making the health examination. However, they do not know their health trends from their own historic health examination data. The health trends may provide the health change information for people from their historic examinations.

Hence, the health trend of each examination item could be seen as a time-series record to be processed in this study. Fig. 2 shows that one person’s health trend of one examination item in his/her historic health examination data.

Figure 2. Health trends of one health examination item.

The purpose of this study is to discover and to analyze useful information from historic health examination data, and those analyzed information can not only be used as auxiliary information but also be used for early prevention of chronic diseases. Therefore, we focus the discussion in this study on construction of the health trend prediction model and the chronic disease prediction model. The two models are introduced below.

1) The Health Trend Prediction Model

From one person’s historic health examination data, a health trend of one examination item can be composed of results of all time points of this examination item. After that, we can then obtain each person’s health trends of each examination item from their own health examination historic data. And finally, the analysis techniques on data mining are applied to discover the relationship between health trends of each examination item and the results of health examination.
2) The Chronic Disease Prediction Model

To effectively increase the quality of health prevention, the relationships between the health trend patterns and the target diseases are discovered and analyzed. Moreover, the value distributions of examination items are analyzed with the target diseases. Hence, the chronic disease prediction model can be then constructed according to above analyzed information.

B. Importance of Health Examination

In the past, the earliest disease prevention concept in the world is proposed by Dr. Horace Dobell, in 1861. That is, the disease examination is composed of multiple diseases examination items, and the examination combination is then called health examination later. Subsequently, the physicians also think that the health combination examination is the basis of prevention medical science [5][7].

This is because the result values of some examination items are abnormal, the physicians can provide corresponding diagnostic suggestions for those persons. After adapting these diagnostic suggestions, those persons can avoid missing the best treatments themselves.

The regular health examination can achieve early detection of early treatment for diseases. Hence, people can clearly understand their own health conditions through the regular health examination. In addition, the chronic diseases, such as high blood pressure, diabetes, heart disease can be prevented through the regular examination.

However, people only know the result of once health examination report. If the health trends of many times health examinations can instead of the result of once health examination, the physicians can get more useful information about persons’ health conditions to provide more precise diagnoses and suggestions for people. The useful information may reduce their health risks.

For example, although the health alert is not occurred in the result of one person’s health examination report, the health trend is similar as the health trend in Fig. 2. Then, the physicians can early provide not only the health alert information but also prevention suggestions for the person. Hence, the development of practical application of health examination data analysis techniques is an imperious demand to achieve the practical effect on disease prevention and disease treatment.

C. Association Rules Mining

The association rule mining, which is proposed by Agrawal et al. [1], in 1993, is to discover the relationship among items from the databases. They assume that transactions include products bought by customers in the transactional databases. The representation format of association rules is denoted as \( \{X\} \rightarrow \{Y\} \), where both \( X \) and \( Y \) are a set of items. For example, the best-known association rule is that \( \{\text{diapers}\} \rightarrow \{\text{beers}\} \). That is, when most of customers bought diapers, they usually bought beers together. In our study, the techniques of association rules mining are applied to discover the relationship among attribute values of different examination items, and the discovered rules could be used as the auxiliary information of our proposed models.

D. Sequential Patterns Mining

Sequential patterns mining is proposed by Agrawal et al. [2], in 1995. Its goal is to discover purchasing behaviors with time sequence from transactional databases. The transactional database is composed of identification numbers of customers, the time of transaction occurred, and items bought in the transactions.

To understand further purchasing behaviors of customers, Agrawal et al. use the sequential patterns mining technology [2] to discover useful information. For example, assume there is a sequential pattern like \( \{\text{Computer}\} \rightarrow \{\text{Printer}\} \). Its meaning is the most customers will buy the printers after buying computers. The example is shown in Fig. 3. But the purchasing behaviors are not the successive behaviors. The purchasing behaviors are only listed by time. Such purchasing behaviors thus called sequential patterns mining.

![Figure 3. The example for sequential patterns mining.](image)

E. Classification Mechanism

The main goal of classification mechanisms is to build the classification model. In the previous studies [6][9], many classification mechanisms are proposed to build the classification model, but their building methods on classification mechanisms are not the same. Fig. 4 shows the classifying process of classification mechanism on data mining.

![Figure 4. The classifying process of classification mechanism.](image)

In Fig. 3, the classification model is constructed and trained by the past known data, and the unknown data can be classified by the trained classification model. That is, the classification model of the classification mechanism is similar as the role of doctors with diagnosis experiences. Therefore, the better classification model could provide the precise diagnosis suggestions for patients according to their described symptoms.

F. Association Rules Classification

CBA (Classification Based on Associations) algorithm is proposed to apply the classification domain of data mining by Liu et al. [6], in 1998. Its whole process can be divided into the two phases. In the first phase, the association rules,
which satisfy minimum support threshold and minimum confidence threshold, are discovered from the training data. An example rule $X \rightarrow Y$, where $X$ is an frequent itemset and $Y$ is the class. In the second phase, the association rules discovered are used to construct the prediction model, and the prediction model can be tested the accuracy from the testing data.

In [10], those researchers applied CBA (Classification Based on Associations) algorithm [6] to the asthma disease of medical applications. For example, assume there is an association rule like “{temperature difference: large, suspended substances: many, cold: yes} $\rightarrow$ {asthma}”. Its meaning is that the person may occur the asthma in the next few days when all the three conditions occurred in the past few days.

When association rules discovered are obtained, we can further construct the classification model. To construct the effectively classification model, those association rules discovered are sorted by their support and confidence values in the second phase. The main goal of sorting method is to select the rules predicted from a large number of rules discovered. The advantage of association rules classification is that it did not require to load the dataset to the memory, and it can select the rules predicted through the sorting method from the association rules discovered. It can thus have good results of the memory usage and the execution performance. In this study, the concept of association rules can be used to find the relationship among health examination items, and the relationship can discover the information of impact factors from examination results for all the health examination items.

III. PROPOSED FRAMEWORK

In this section, we explain our proposed framework in details.

In this subsection, since the shown symptoms of each disease are different, their corresponding health examination items are also not the same. For the different target diseases, hence, we should construct corresponding analysis and prediction models. Fig. 5 shows the framework of health trend pattern prediction with target diseases. The required processes are introduced below.

![Figure 5. The process of finding health risk patterns.](image)

1) Part 1: The Health Examination Data Labeled by Target Disease

In the 1st part, we first have to label the health examination data whether each one is normal condition by its health examination result or not. However, we need physicians’ help to label the health examination data by their diagnosis experiences because they have medical domain knowledge. After labeling process, these persons who suffer from the diseases are found from the health examination data. Note that the case history or the last health examination report can be used to determine one person whether he/she has the target disease or not. Next, health examination data which those persons suffer from the disease are marked “Abnormal”. Otherwise, other data is marked “Normal”.

2) Phase 2: Emerging Sequence Pattern Discovery

In the 2nd phase, the successive values of each health examination item from the earliest time to the latest time can be seen as a time-series record. After that, we can then obtain the time-series data of two clusters, such as a cluster with disease risk and the other without disease risk. According to difference in attribute value between the time-series data of two clusters, the attributes with large difference are found.

The above processes are introduced below. Firstly, the successive values of each health examination item from each person’s historic health examination data can be seen as a time-series curve. The reason is that the time-series curve can show the health change. After processing all persons’ historic data, hence, each person’s time-series curves at each health examination item can be then obtained. Next, those curves with the large difference between the time-series data of normal cluster and abnormal cluster are found by emerging pattern technique.

And finally, those curves with large difference can be seen as the features. This is because the support values of curves of these health examination items exist the obvious differences between the two clusters. Hence, those features can be used for not only constructing the classification mechanism but also regarding as the difference between the two clusters.

3) Part 3: Finding of Frequent Attribute Set

In fact, since the number of each person’s health examination is not many, the static features are also considered in this study. In the 3rd part, the result of each person’s the latest health examination is labeled “Abnormal” or “Normal” by physicians. Hence, TABLE III shows the content of processed data, where the high-risk attribute represents the person whether he/she suffers from the diseases or not.

<table>
<thead>
<tr>
<th>Item A1</th>
<th>Item A2</th>
<th>Item (N-1)</th>
<th>Item B</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test T1</td>
<td>32</td>
<td>27</td>
<td>97</td>
<td>1.3</td>
</tr>
<tr>
<td>Test T2</td>
<td>38</td>
<td>30</td>
<td>110</td>
<td>0.95</td>
</tr>
<tr>
<td>Test T3</td>
<td>51</td>
<td>28</td>
<td>130</td>
<td>1</td>
</tr>
<tr>
<td>Test T4</td>
<td>48</td>
<td>30</td>
<td>113</td>
<td>0.95</td>
</tr>
<tr>
<td>Test T5</td>
<td>5</td>
<td>35</td>
<td>127</td>
<td>1.3</td>
</tr>
</tbody>
</table>

Next, the information gain method is applied to select those attributes having higher information gain with class values. Then, the cutting point of the domain value of each selected attribute is found. Such way can let these selected
attributes having the highest degree of information gain with class values.

After completing above processes, the values of those health examination items are converted into the specified representation format by the cutting point of corresponding attribute. For example, assume that the domain value of one health examination item B is from 0 to 20, and its cutting point is 10. If the value of one person’s health examination item B is 5, then the value could be converted into the representation format “<10”. Hence, the values of these attributes of each person’s latest health examination data are converted into the corresponding representation format by cutting information. TABLE IV shows the converted health examination data.

<table>
<thead>
<tr>
<th>TABLE III. THE CONTENT OF THE FEATURE ATTRIBUTE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
</tr>
<tr>
<td>Testec P1</td>
</tr>
<tr>
<td>F4(A1&lt;10.5), F3(A3&lt;20), F6(A8&lt;1)</td>
</tr>
<tr>
<td>Testec P2</td>
</tr>
<tr>
<td>F6(A8&lt;1), F7(A7&lt;2.45)</td>
</tr>
<tr>
<td>Testec P3</td>
</tr>
<tr>
<td>F1(A1&lt;10.5), F4(A5&lt;3), F6(A8&lt;1), F10(A11&lt;5.6)</td>
</tr>
<tr>
<td>Testec P4</td>
</tr>
<tr>
<td>F1(A1&lt;10.5), F3(A3&lt;20), F6(A8&lt;1)</td>
</tr>
<tr>
<td>Testec P5</td>
</tr>
<tr>
<td>F4(A5&lt;3), F3(A3&lt;20), F10(A11&lt;5.6)</td>
</tr>
</tbody>
</table>

Next, association rules mining technique on data mining is applied to discover the frequent itemsets which they have high frequency and high confidence relationships with class value. Then, the static features can be obtained via the processing method. For example, [the health examination item A1<10.5, the health examination item A3>20} → {the specific disease}.

4) Part 4: Construction of The Prediction Model

In the 4th part, we can use the obtained feature data to construct the classification model by using the CBA (Classification Based on Associations) algorithm [6]. The processes of building classification rules are introduced below. Firstly, all data are converted the boolean representation in the feature attribute table. The reason is that it can be easily observed the features of each person’s health examination items as shown in TABLE V. Secondly, the feature attribute table can be used to build the classification model.

<table>
<thead>
<tr>
<th>TABLE IV. THE FEATURE ATTRIBUTE TABLE.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
</tr>
<tr>
<td>Testec P1</td>
</tr>
<tr>
<td>F1(A1&lt;10.5), F2(A2&lt;1.1), F3(A3&lt;20)</td>
</tr>
<tr>
<td>Testec P2</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>Testec P3</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>Testec P4</td>
</tr>
<tr>
<td>0</td>
</tr>
<tr>
<td>Testec P5</td>
</tr>
<tr>
<td>0</td>
</tr>
</tbody>
</table>

After completing the disease prediction model, the model can be used to predict the prediction of specific diseases. Similarly, each person’s historic health examination data are converted into the representation format of feature attribute table. That is, the classification features of health examination data are found by the feature attribute table, and then are performed the classifying process by the built decision tree. And finally, we can obtain the classification result after classifying process.

IV. CONCLUSIONS

In this study, we have proposed a novel framework of the personal health data analysis by mining the health examination data. The adaptive techniques on data mining are used to discover the health risk patterns of each health examination item from history data of each examination one. Moreover, we also propose a method for analyzing the relationship between the health trend patterns and the target disease, such that the prediction model is further constructed. Information analyzed by our proposed system can be provided for physicians as the auxiliary information on treatments or alerts. By the analyzed information, the physicians can early provide not only the health alerts but also the adaptive medical treatments for people to increase the effect of early treatment.

For the future work, we would implement the proposed framework and apply it on real health examination datasets. Various experiments will be conducted to evaluate the performance of the proposed framework system.

ACKNOWLEDGMENT

This study is conducted under the “Applied Information Services Development & Integration, Phase II” of the Institute for Information Industry which is subsidized by the Ministry of Economy Affairs of the Republic of China.

REFERENCES
