Discovering Association between Metabolic Syndrome and its Related Chronic Diseases Represented by ICD-10 Code

Supak Iamongkot, Sotarat Thammaboosadee, and Supaporn Kiattisin

Technology of Information System Management Division, Faculty of Engineering, Mahidol University, Nakhon Pathom, Thailand

Email: supuk_note@hotmail.com, sotarat.tha@mahidol.ac.th, supaporn.kit@mahidol.ac.th

Abstract—This paper applies the association rules method to discover the relationship between metabolic syndrome and its chronic diseases. The sample data used in this research is medical records specified to metabolic syndrome patients in a large government hospital. The Apriori and FP-Growth algorithms are chosen to be compared in the performance and applicable results of extracting the relationship of the metabolic syndrome patient records represented by ICD-10 code. The results show that the Apriori can extract 6 rules and 724 rules from FP-Growth. The comparative results between Apriori and FP-Growth found that 6 rules are common. The overall results show that the metabolic syndrome patients mostly have strong relationships with hypertension, obesity and diabetes. Interestingly, these diseases often occur with the patients was diagnosed that was metabolic syndrome. Additionally, the results would bring to the suggestion in metabolic syndrome patients to know about the relationship of these chronic diseases. Moreover, the physicians could use this guide for the treatment strategy in the future.

Index Terms—association rules, Apriori, FP-Growth, metabolic syndrome, chronic diseases, ICD-10

I. INTRODUCTION

Metabolic syndrome [1] is obesity condition to especially the lumbar. It makes adversely affect various body systems the multiple system. This condition refers to a group disease caused by metabolic the disorders which affect pose the hyperlipidemia, hypertension, has insulin resistance and would cause heart disease and high artery.

The main cause of metabolic syndrome has two main respects: obesity and insulin resistance. The obesity is imbalance of food intake and energy consumption of the body which makes the fat accumulation in part the abdominal surface and the waistline. Additionally, the insulin resistance was caused by genetic causes and external causes such as the obesity, increasing age and uses the certain drugs. The term "Chronic diseases" [2] means a disease that may without symptoms but the symptoms will gradually intensified as when it is not treated. It often needs to be treated continued and caused the symptoms acute redundancy is always such as diabetes, hypertension, heart disease, hypercholesterolemia and cancer.

In the currently there is a classification of various diseases and symptoms of diseases by using classification codes of various diseases such as myocardial disease, renal disease, aborticide and metabolic syndrome. This classification code is called the ICD-10 [3] which stands for International Classification of Diseases and Related Health Problem 10th Revision. It is the code of the disease and symptoms prepared by the World Health Organization (WHO). It has been developed for recording and gathering statistical data in health planning internationally such as E88.9 Metabolic disorder unspecified, E66.9 Obesity unspecified, E51.9 Heart disease unspecified, etc.

From problems and diseases that occurred with patients that are metabolic syndrome motivates authors to study metabolic syndrome by using data represented by the ICD-10 code for discover the relationship of metabolic syndrome with its chronic diseases caused followed from it by the association rules method is to find relationship of two data sets or more to set up within a larger data group. The expected results of this research is that the usage of the discovered relationship from data of metabolic syndrome patients to raise awareness in caring, medical protection, medical suggestion, attitudes, and also policy, according to metabolic syndrome.

II. RELATED THEORIES

This section presents theory about diseases, data mining and its approaches.

A. Metabolic Syndrome

Metabolic syndrome [4] is obesity condition to especially the lumbar. It makes adversely affect various body systems the multiple system. This condition refers to a group disease caused by metabolic the disorders which affect pose the hyperlipidemia, hypertension, has insulin resistance and would cause heart disease and high

Manuscript received April 27, 2015; revised July 20, 2015.

This work was supported by the Faculty of Graduate Studies, Mahidol University grant to support graduate students in academic presentations in Singapore.

artery. Olden would call this disease groups that syndrome X, insulin resistance syndrome.

The main cause of metabolic syndrome has two main respects: obesity and insulin resistance. The obesity is imbalance of food intake and energy consumption of the body which makes the fat accumulation. Especially precinct part the abdominal surface and the waistline. Additionally, the insulin resistance was caused by genetic causes and external causes such as the obesity, increasing age and uses the certain drugs. Usually, the metabolic syndrome patients will have insulin resistance rather than obesity patient hip precinct and pose the Hypertension, sugar in the low blood and also cause heart disease and artery subsequent. Besides relaxation is not enough, sleep deprivation, stress and work has the style of part-time job affects being metabolic syndrome as well [5].

B. Chronic Diseases

The term "Chronic diseases" means a disease that when the onset the symptoms may without symptoms but the symptoms will be gradually intensified as when it is not treated. It often needs to be treated continued may lifelong and often is not completely cured and caused the symptoms acute redundancy is always such as diabetes, hypertension, heart disease, hypercholesterolemia and cancer. Causes the chronic diseases be several cause are that is [6] disorders of the immune system such as Allergies, Systemic Lupus Erythematosus, etc., disorders of genetic such as psoriasis and diabetes type 1, caused by use everyday life invalid prolonged until led to mechanism the health disorders and caused by the deterioration of mechanisms field the various health as age increases.

C. International Classification of Diseases and Related Health Problem 10th Revision (ICD-10)

International Classification of Diseases and Related Health Problem 10th Revision or ICD-10 [7] is the code of the disease and symptoms. It has been developed for recording and gathering statistical data in health planning internationally.

ICD started use in 1893. For use in specified code the death was an international code and in 1948 the WHO has come responsible in improvement the ICD to be the 6th Edition. WHO has developed ICD-10 codes continuously and added information to be modern. The latest revised edition is ICD 10: 2010 which is improved on the A.D. 2010 or B.E. 2553.

ICD-10 codes are alphanumeric codes. By each code to begin with the English characters A-Z then followed by Arabic numbers 0-9, two to four digits which provide diseases codes for the individual patients. In general, providing the ICD-10 codes will provide the code to disease that the patient is just one disease. Providers the code wants to name the disease a clear and has good looks.

D. Data Mining

Data mining [8] liken evolution in the storage and interpreting the data easily one from the original there the data storage easily until enter store in the form of database that can be extract information come used until data mining that can discover knowledge hidden in the data. Generally, data mining consists of three tasks which are classification, clustering and association. In this paper, the method of association rules extraction is considered.

Association rules [8] is to find the relationship rules of data by finding relation of at least two data sets, setting up within data groups at large, and displaying the relationship of events or objects that occur simultaneously. An example of the applied uses the association rules such as if it is found that people who usually bought a video tape often bought adhesive tape together by consider products that buyers often bought simultaneously. Stores maybe set stores for both products closer together to increase sales. There are several approaches in the finding of association rules as follows:

1) Apriori algorithm: This is the most basic which Apriori was use to find all frequent itemsets in the regulative database. The main concept of Apriori algorithm is to make multiple pass over the database. It justly depend the Apriori characteristic which states that "All non-empty itemsets of a frequent itemset must be the frequent" [9]. Also explain the anti-monotonic characteristic which if the system could not pass the minimum support test, all its supersets will fail to passes the test.

2) FP-Growth algorithm: FP-Growth [9] is other efficiently frequent pattern mining method. It creates frequent itemsets without candidacy generation which use tree based structure. The problem of Apriori algorithm is manage by introduction a novel, compact data structure, called frequent pattern tree or FP-tree then dependent this structure an FP-tree-based pattern segment growth method has been the develops. Which construct conditional frequent pattern tree and conditional pattern base from database which satisfaction the minimum support and FP-growth traces the set of concomitant item.

III. METHODOLOGY

This paper uses the association rules techniques to find the relationship between metabolic syndrome and its chronic diseases. In the preparation of the data used in find association rules there are various process as shown in Fig. 1.



Figure 1. Research process.

A. Research Interests Study

Starts from searching research that is interested before that the authors has interested about research any aspect and then find research about field to the authors interested. Then, the research summarizes those that to make study that those research is research about anything. Use any method technique in make above research and results that get from research is what.

B. Data Gathering

This paper gathered data requested the data of metabolic syndrome patients from a government hospital in Thailand, in the period of the year 2009 - 2014. The data that used in this study is a secondary data which has been already collected. By the data structure has 13 attribute, 3677 instances.

C. Data Selection

This step will select data used in research by bringing information obtained from collected data of patient which has been requested data of the patients was diagnosed that was metabolic syndrome from a government hospital. Consequently, authors will select data that preferred to use in find the relationship between metabolic syndrome and its related chronic diseases consists of two attribute that is enc_id and dx_code that would use in this paper.

D. Data Preprocessing and Transformation

After the selected data is ready, the data would be grouped by dx_code which is the attribute about diseases that patient came for treatment, by merging diseases that have character the related or similar symptoms into same group such as E119 (Non-insulin-dependent diabetes mellitus without complications) and E113 (Non-insulindependent diabetes mellitus with ophthalmic complication) are grouped as EX04 (diabetes), E669 (Obesity, unspecified) and E668 (Other obesity) are grouped as EX03 (Obesity), and so on. To reduce the amount of the attributes that would be fed into the modeling, the authors will determines code to each group and must be unique to the ICD-10 codes, which would has 85 attribute and then, brings the enc_id to transform from the numeric as nominal by transform them to ID01, ID02 to ID1024.

E. Modeling and Evaluation

The preprocessed data is used for modeling by using association rules techniques, which are selected as the Apriori and FP-Growth method. The results from both methods will be shown and compared to find model that are reliability and acceptable the most. To measure the reliable and acceptable results, the association rules discovery methods usually use two common indicators which are support and confidence [8]. Support value is a value that represents the accuracy of the association rules that use or the probability of an event occurring simultaneously. Confidence value is value that says rules acquired has an actuality how much.

F. Conclusions and Comparing the Extracted Association Rules

The association rules should be interpreted and analyzed its results and conclude research. For find the relationship between metabolic syndrome and chronic diseases caused by metabolic syndrome and compare association rules that same between Apriori and FP-Growth method. Thus, there presentation about suggestions, attitudes and preventing of metabolic syndrome for used in further treatment.

IV. EXPERIMENTAL RESULTS

In this paper, chosen techniques are Apriori and FP-Growth method applied to the metabolic syndrome patient data represented by ICD-10 code. Modeling association rules for find relationship between metabolic syndrome and its related chronic diseases in metabolic syndrome patients which have 1,024 patients. In Modeling has define the minimum support to 0.1 and minimum confidence to 0.1.

A. Result of Apriori

Result of modeling the association rules between metabolic syndrome and its related chronic diseases by Apriori method has six rules, as shown in Table I.

 TABLE I.
 Association Rules Result of Apriori

No.	Association Rules	Support (%)	Confidence (%)
1	EX04 ==> EX02	10	67
2	EX03 ==> EX02	10	52
3	EX02 ==> EX04	10	51
4	EX04 ==> EX03	10	45
5	EX02 ==> EX03	10	43
6	EX03 ==> EX04	10	42

From the results shown in Table I, it can be interpreted the discovered rules, for example:

Rule 1: If patients were Metabolic syndrome (EX01) and Diabetes (EX04) they would were Hypertension (EX02).

Rule 2: If patients were Metabolic syndrome (EX01) and Obesity (EX03) they would were Hypertension (EX02).

B. Result of FP-Growth

Results of modeling the association rules between metabolic syndrome and its related chronic diseases by FP-Growth method have 724 rules. Total 10 rules were discovered as shown in Table II.

From the results shown in Table II, it can be interpreted the discovered rules, for examples:

Rule 1: If patients were Metabolic syndrome (EX01) and Disease involving the urinary system (EX13) they would were Personal history of congenital malformations, deformations and chromosomal (Z877).

Rule 2: If patients were Metabolic syndrome (EX01) and Diseases of the respiratory (EX20) they would were Personal history of congenital malformations, deformations and chromosomal (Z877).

No.	Association Rules	Support (%)	Confidence (%)
1	EX13 ==> Z877	98.1	100
2	EX20 ==> Z877	88.1	100
3	EX03 ==> Z877	69.4	100
4	EX02 ==> Z877	62.8	100
5	EX04 ==> Z877	28.3	100
6	EX05 ==> Z877	16.4	100
7	R730 ==> Z877	9.8	100
8	EX07 ==> Z877	6.2	100
9	EX05 ==> EX13	16.4	100
10	R730 ==> EX13	9.8	100

TABLE II. ASSOCIATION RULES RESULT OF FP-GROWTH

C. Comparison of Apriori and FP-Growth Method

According to the results of Apriori and FP-Growth, their comparative results showed that there are six common association rules are extracted by both methods, as shown in Table III. There are also association rules not common to several the rules from the comparison of Apriori and FP-Growth which these rules are extracted by FP-Growth only.

TABLE III. COMPARATIVE COMMON ASSOCIATION RULES BETWEEN APRIORI AND FP-GROWTH METHOD

		Apriori		FP-Growth	
No.	Association Rules	Support	Confidenc	Support	Confidenc
		(%)	e (%)	(%)	e (%)
1	EX04 => EX02	10	67	9.5	33.4
2	EX03 => EX02	10	52	48	69.2
3	EX02 => EX04	10	51	9.5	15.1
4	EX04 => EX03	10	45	15.6	55.2
5	EX02 => EX03	10	43	48	76.5
6	EX03 => EX04	10	42	15.6	22.5

V. CONCLUSIONS

This paper is to apply the association rules, for discover the relationship between metabolic syndrome and its chronic diseases represented by ICD-10 which authors chosen using Apriori and FP-Growth method in extracting the relationship of the metabolic syndrome patient data. The results of Apriori have 6 rules and the results of FP-Growth have 724 rules. Which the result from comparison between Apriori and FP-Growth method found has 6 rules that same in finding this relationship. The results show that the metabolic syndrome patients mostly have relationship with Hypertension, Obesity and Diabetes. The metabolic syndrome patients often would were hypertension or obesity or diabetes along or were more than one disease, Moreover, these diseases often occur with metabolic syndrome patients. Additionally, the results would bring to the suggestion in metabolic syndrome patients to know about the relationship of these chronic diseases and physicians can guide for the treatment metabolic syndrome patients in the future.

REFERENCES

S. M. Grundy, H. B. Brewer, J. I. Cleeman, S. C. Smith, and C. [1] Lenfant, "Definition of metabolic syndrome report of the National Heart, Lung, and Blood Institute/American Heart Association conference on scientific issues related to definition," Circulation, vol. 109, no. 3, pp. 433-438, 2004.

- [2] D. Yach, C. Hawkes, C. L. Gould, and K. J. Hofman, "The global burden of chronic diseases: Overcoming impediments to prevention and control," Jama, vol. 291, no. 21, pp. 2616-2622, 2004
- Contexo Media. 2014 ICD-10-CM Official Coding Guidelines [3] [Online]. Released. Available: https://aafp.codeitrightonline.com/ciri/2014-icd-10-cm-officialcoding-guidelines-released.html.
- [4] Metabolic Syndrome. [Online]. Available: http://emedicine.medscape.com/article/165124-overview#a0101.
- S. Linlawan, Incidence and Risk Factors of Metabolic Syndrome [5] in Personal of Phrachomklao Hospital, 2011.
- F. G. Schellevis, J. van der Velden, E. van de Lisdonk, J. T. M. [6] Van Eijk, and C. V. van Weel, "Comorbidity of chronic diseases in general practice," Journal of Clinical Epidemiology, vol. 46, no. 5, pp. 469-473, 1993.
- [7] ICD-10-PCS Official Guidelines for Coding and Reporting. [Online]. Available: https://www.cms.gov/Medicare/Coding/ ICD10/Downloads/PCS-2014-guidelines.pdf
- [8] R. Sakdulyatham, "Factors for successful analysis of trigger finger treatment of faculty of medicine vajira hospital," University of Bangkok Using Data Mining Techniques, 2010.
- [9] A. Saxena and S. Gadhiya, "A Survey on frequent pattern mining methods Apriori, Eclat," FP Growth, vol. 2, no. 1, 2012.



Supak Iamongkot was born in Suphan Buri, Thailand in February 1990. She received the B.B.A. (first-class honors) degree in Computer Information System - Business Computer from Rajamangala University of Technology Thanyaburi, Thailand, in 2012 and M.Sc. degree in Information Technology Management from Mahidol University, Thailand, in 2015. Her research interests include information systems, information technology management, data mining, knowledge discovery and Future

computer technologies.



Sotarat Thammaboosadee was born in 1982 and received his B.Eng. and M.Sc. degrees in Computer Engineering and Technology of Information System Management from Mahidol University, Thailand, in 2003 and 2005 respectively. He also received a Ph.D. in Information Technology from King Mongkut's University Technology Thonburi, Thailand, in 2013. He is now a lecturer at Technology of Information System

Management Division at Faculty of Engineering, Mahidol University, Thailand. His research interests include data mining in several domains such as the legal domain, healthcare domain and management domain. He also interests the research filed of technology valuation and business process improvement in CRM section.



Supaporn Kiattisin received her B.Eng. and M.Eng. degrees in Computer Engineering and Electrical Engineering from Chiang Mai University, Thailand, in 1996 and King Mongkut's University Technology Thonburi, Thailand, in 2000 respectively. She also received a Ph.D. in Electrical and Computer Engineering from King Mongkut's University Technology Thonburi, Thailand, in 2007. She is now a vice-director at

Technology of Information System Management Division and program chair of Information Technology Management program at Faculty of Engineering, Mahidol University, Thailand. Her research interests include enterprise architecture, risk management, business process improvement, customer relationship management, and human resource management.