International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 07 | July 2020 www.irjet.net p-ISSN: 2395-0072

Chronic Disease Mining

Abhijith J

Student, Dept. of Dual Degree Computer Applications, Sree Narayana guru institute of science and technology paravur, Kerala, India

---

Abstract: Existing disease progression methods often ignore complex relations, such as the time-gap and pattern of disease occurrence. They also do not take into account the different medication stages of the same chronic disease, which is of great help when reducing healthcare costs. A heterogeneous network based chronic disease progression mining method is used to improve the current understanding on the progression of chronic diseases, including orphan diseases. The method also considers the different medication stages of the same chronic disease.

Key Words: hncdpm, chronic disease,mpm

1. INTRODUCTION

Detecting chronic disease at an earliest stage is an important and difficult task. With the rapid evolution of computer software and hardware technologies, various types of data from the pharmaceutical companies and information captured from wearable devices are becoming increasingly available. However, challenges such as sparsity, heterogeneity, noise, and bias, have been encountered when working directly with these records. Effective mining of these data can help us obtain actionable insights into chronic disease progression. If we can understand the process of chronic disease progression, we will be able to identify patients at risk of developing chronic disease based on their previous healthcare history. Preventive measures can then be taken, to increase the quality of healthcare and reduce costs.

Various data-mining methods have effectively been applied in the healthcare setting. These methods include a variety of supervised learning algorithms, such as decision tree and artificial neural network, to predict heart disease, cancer, and other diseases. Clustering and vector similarity-based collaborative filtering methods have also been proposed to predict individual disease risk. Although these data mining methods can capture the comorbidity of diseases, they often ignore complex relations, such as time-gap and pattern of disease occurrence.

2. EXISTING SYSTEM

This section reviews the existing work on disease progression mining. Charlson Comorbidity Index was proposed in1987 to predict the 10-year mortality of patients by ranking a range of demographic and comorbid conditions, such as heart disease, cancer, and AIDS. The Elixhauser index showed slightly better prediction performance than this index, especially when predicting mortality beyond 30 days. Similar models, such as APACHE-II and Mortality Probability Models (MPM), have also been used to assess the condition of ICU patients and determine the aggressiveness of treatment.

A more recent approach introduced in healthcare informatics is derived from social network analysis methods. Based on graph theory, these methods treat the healthcare data as complex relations between different entities, including physicians, diseases, and hospitals. The goal of this approach is to mainly understand the interactions between healthcare entities, improve collaboration efficiency among physicians, map the knowledge structure in healthcare research, and understand the progression of chronic comorbidity, among others. However, the actual context of healthcare settings, approaches, and entities considered in social network analysis varies widely. Although these methods can help us understand chronic disease progression to some extent, they are often only applicable to a specific kind of chronic disease and lack generality. Besides, they consistently ignore complex relations, such as time-gap and pattern of disease occurrence.

Frequent Sub graph Mining (FSM) is the essence of graph mining. The objective of FSM is to extract all frequent sub graphs in a given data set with occurrence counts above a specified threshold. The straight forward idea behind FSM is to grow candidate sub

graphs in either a breadth-first or depth-first manner (candidate generation) and then determine whether the identified candidate sub graphs occur frequently enough in the graph data set for them to be considered interesting (support counting). However, in existing FSM methods, such as gSpan, nodes that appear less often than the threshold are removed and do not appear in the obtained frequent sub graph. This issue may result in our inability to obtain information on orphan diseases.

Complex network communities can be categorized as overlapping or non-overlapping community structures. Of particular interest to this work is the overlapping community, which is a notable feature in many networked systems. In healthcare networks, a disease can belong to multiple groups. Hence, the detection of overlapping communities in complex networks can be expressed as a disease progression mining problem. Among the existing community detection methods currently available, Info map shows the most rapid calculations and the highest accuracy. This method considers a random walk over the network. The more extensively the nodes are connected one with each other, the more likely the walker will stay within them and, thus, form a community. Analysis of the flows over the network gives access to the underlying community structure. However, Info map considers only the topological structure of the network; it cannot obtain an optimal community detection result in healthcare networks. Thus, we improve Info map by combining the semantic information of the nodes and make the community detection results more meaningful.

4. METHODOLOGY

A Heterogeneous Network-based Chronic Disease Progression Mining (HNCDPM) method help to understand the progression of chronic disease, including orphan diseases, detect chronic disease fraud, and reduce healthcare costs. HNCDPM considers the different medication stages of the same disease and obtains two types of rules: the pattern between different periods of different chronic diseases, which indicates the relationship between different types of chronic disease, and the pattern between different stages of the same chronic disease, which shows the clinical path of the chronic disease. These two types of rules can be used to detect chronic disease fraud. Extensive experiments show that our method can outperform the existing methods by 20% in terms of precision.

We denote P= (p_1,p_2..p_n) as a set of patients who have chronic healthcare insurance records during period (T_s,T_e), where T_s indicates the start time of the time range and T_e is the end time of the time range. The purpose of our algorithm is to mine the progression of chronic disease based on these records. We consider the different stages of the same chronic disease and mine two types of progression rules of the disease. One type of rules describes the pattern between different stages of different chronic diseases, which indicates the relationship between different kinds of chronic diseases. The second type of rules describes the patterns between different stages of the same chronic disease, which shows the clinical path of the disease.

4.1 ALGORITHM FOR HETEROGENEOUS NETWORK BASED CHRONIC DISEASE PROGRESSION MINING

HNCDPM method is used to mine the chronic disease progression. The proposed method can be divided into five steps.

Step 1: Construct a health seeking temporal graph for each patient.

Step 2: Mine frequent disease process sub graphs from the graph set in Step 1 using Constrained Frequent Sub graph Mining (CFSM), and recode the health seeking temporal graph set with the mined frequent disease process sub graphs.

Step 3: Construct the base disease progression network by statistical aggregation of the recoded graph set in Step 2.

Step 4: Conduct community detection on the base disease progression network and transform them into chronic disease progression rules.

Step 5: Conduct chronic disease based healthcare insurance fraud detection according to the rules obtained in Step 4.

4.2 DEFINITIONS FOR HEALTH SEEKING TEMPORAL GRAPH CONSTRUCTION

Suppose the time range of healthcare insurance records is (T_s, T_e), where T_s indicates the start time of the time range and T_e is the end time of the time range.

Definition 1: Health-seeking Behavior

Each health seeking behavior bi can be denoted as b_i = (p,d,t) where p is the patient, d denotes the diagnose, and t is the health-seeking time of the health-seeking behavior. A health seeking behavior may contain multiple kinds of drugs or treatments.

Definition 2: Health-seeking Temporal Graph

Health-seeking temporal graph G is a heterogeneous information network with three types of nodes: patient, health seeking-behavior, and process. Three types of edges are observed in G.

The edge between patient node pi and health seeking behavior node b_j shows that patient p_i conducts the health-seeking behavior b_j. The edge between health-seeking behavior node b_u and health-seeking behavior node b_v indicates that b_v occurs after b_u, and the weight of edge e_{uv} is defined as:

$$W_{euv} = \frac{1}{|t_u - t_v| + 1}$$

Where, t_u and t_v refer to the time of b_u and b_v. The shorter the time interval between b_u and b_j, the less is the weight of edge e_{uv}. The edge between health seeking behavior node b_j and process m_r indicates that process m_r is used in b_j, and the weight of edge e_{jr} shows the dose of m_r in b_j. Each patient has a health seeking temporal graph G, which can be donated as Figure 4.1 shows an example of the health seeking temporal graph of patient p_1.

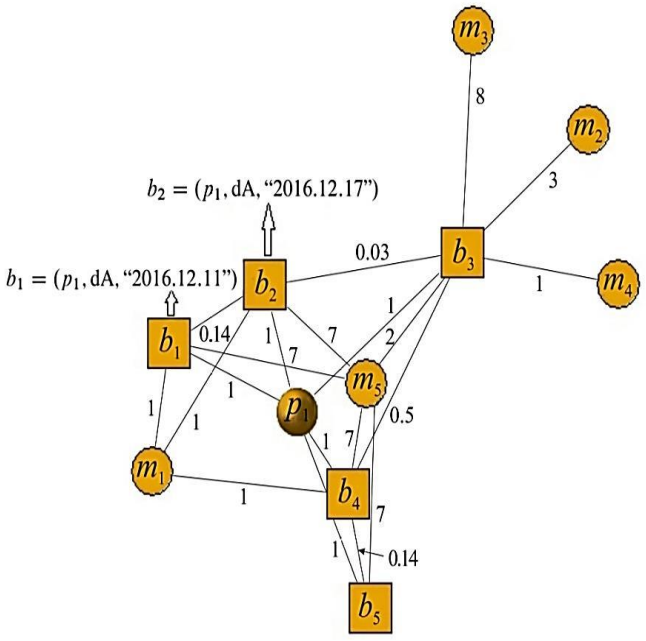

Figure 4.1 Health seeking temporal graph of patient p_1. The spherical node indicates a patient with chronic disease, each square node represents a health-seeking behavior, and each circular node is a process.

From Figure 4.1, the different health seeking behaviors of the same patient. Although the diagnoses are the same, they may indicate different periods of the same disease. Thus, we can identify the period of disease according to the process information. Conventional disease progression mining methods consider the same diagnoses as the same disease. To identify whether the two health seeking behaviors with the same diagnoses refer to the same disease period, we must mine the frequent disease-process patterns.

5. FUTURE WORK

In future we can introduce the healthcare informatics is derived from social network analysis methods. Based on graph theory, these methods treat the healthcare data as complex relations between different entities, including physicians, diseases, and hospitals. The goal of this approach is to mainly understand the interactions between healthcare entities, improve collaboration efficiency among physicians, map the knowledge structure in healthcare research, and understand the progression of chronic comorbidity,

among others. However, the actual context of healthcare settings, approaches, and entities considered in social network analysis varies widely. Although these methods can help us understand chronic disease progression to some extent, they are often only applicable to a specific kind of chronic disease and lack generality. Besides, they consistently ignore complex relations, such as time-gap and pattern of disease occurrence.

6. CONCLUSION

The developed method helps to understand the progression of chronic disease, including orphan diseases, and is helpful in detecting chronic diseases which helps in reducing healthcare costs. HNCDPM considers different medication periods of the same disease and produces two types of rules: the pattern between different stages of different chronic diseases, which indicates the relationship between different types of chronic disease, and the pattern between different stages of the same chronic disease, which shows the clinical path of the disease. Extensive experiments show that this method can outperform the existing methods by over 20% in terms of F-measure.

ACKNOWLEDGEMENT

In the name of almighty, I would like to extend my heartfelt thanks to our HoD Mrs.Kavitha C.R, Department of a Dual Degree Master of Computer Applications for the helps extended to me throughout my course of my study. I am deeply grateful to my guide Mrs. Anagha Pradeep.Assistant Professor, Department of a Dual Degree Master of Computer Applications for the valuable guidance

REFERENCES

[1] J.S.Ko,H.Chalfin,B.J.Trock,Z.Y.Feng,E.Humphreys, S. W. Park, H. B. Carter, K. D. Frick, and M. Han, Variability in Medicare utilization and payment among urologists, Urology, vol. 85, no. 5, pp. 1045–1051, 2015.

[2] M. E. Charlson, P. Pompei, K. L. Ales, and C. R. MacKenzie, A new method of classifying prognostic comorbidity in longitudinal studies: Development and validation, J. Chron. Dis., vol. 40, no. 5, pp. 373–383, 1987.

[3] A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey, Comorbidity measures for use with administrative data, Med. Care, vol. 36, no. 1, pp. 8–27, 1998.

[4] M. T. A. Sharabiani, P. Aylin, and A. Bottle, Systematic review of comorbidity indices for administrative data, Med. Care, vol. 50, no. 12, pp. 1109–1118, 2012.

[5] D. T. Wong and W. A. Knaus, Predicting outcome in criticalcare: The current status of the APACHE prognostic scoring system, Can. J. Anaesth., vol. 38, no. 3, pp. 374– 383, 1991.

[6] M. J. Breslow and O. Badawi, Severity scoring in the critically ill: Part 1—Interpretation and accuracy of outcome prediction scoring systems, Chest, vol. 141, no. 1, pp. 245–252, 2012.

[7] M.Baglioni,S.Pieroni,F.Geraci,F.Mariani,S.Molinaro, M. Pellegrini, and E. Lastres, A new framework for distilling higher quality information from health data via social network analysis, in Proc. 13th Int. Conf. Data Mining Workshops, Dallas, TX, USA, 2013, pp. 48–55.

[8] J. G. Anderson, Evaluation in health informatics: Social network analysis, Comput. Biol. Med., vol. 32, no. 3, pp. 179–193, 2002.

[9] S. Uddin, A. Khan, and M. Piraveenan, Administrative claim data to learn about effective health care collaboration and coordination through social network, in Proc. 48th Hawaii Int.Conf.System Sciences, Kauai, HI,USA,2015, pp. 3105–3114.

[10] S. Uddin, A. Khan, and L. A. Baur, A framework to explore the knowledge structure of multidisciplinary research fields, PLoS One, vol. 10, no. 4, p. e0123537, 2015.

[11] H. Luijks, T. Schermer, H. Bor, C. Van Weel, T. Lagro Janssen M. Biermans and W. De Grauw, Prevalence and incidence density rates of chronic comorbidity in type 2 diabetes patients: An exploratory cohort study, BMC Med., vol. 10, p. 128, 2012.

[12] D. Chambers, P. Wilson, C. Thompson, and M. Harden, Social network analysis in healthcare settings: A systematic scoping review, PLoS One, vol. 7, no. 8, p. e41911, 2012.

[13] X.F.Yanand J.W.Han, gSpan: Graph-based sub structure pattern mining, in Proc. 2002 IEEE Int. Conf. Data Mining, Maebashi, Japan, 2002, pp. 721–724.

[14] M.Rosvall and C.T.Bergstrom, Maps of random walk son complex networks reveal community

structure, Proc. Natl. Acad. Sci. USA, vol. 105, no. 4, pp. 1118–1123, 2008.

[15] X. Y. Li, H. H. Cao, E. H. Chen, H. Xiong, and J. L. Tian, BP-growth: Searching strategies for efficient behavior pattern mining, in Proc. 13th Int. Conf. Mobile Data Management, Bengaluru, India, 2012, pp. 238–247.

[16] J. A. K. Suykens, Support vector machines: A nonlinear modelling and control perspective, Eur. J. Control, vol. 7, nos. 2&3, pp. 311–327, 2001.

[17] D. Chambers, P. Wilson, C. Thompson, and M. Harden, Social network analysis in healthcare settings: A systematic scoping review, PLoS One, vol. 7, no. 8, p. e41911, 2012.

[18] X.F. Y anand J. W.Han, gSpan: Graph-based sub structure pattern mining, in Proc. 2002 IEEE Int. Conf. Data Mining, Maebashi, Japan, 2002, pp. 721–724.

[19] M. Rosvalland C.T.Bergstrom, Maps of random walkson complex networks reveal community structure, Proc. Natl. Acad. Sci. USA, vol. 105, no. 4, pp. 1118–1123, 2008.

[20] X. Y. Li, H. H. Cao, E. H. Chen, H. Xiong, and J. L. Tian, BP-growth: Searching strategies for efficient behavior pattern mining, in Proc. 13th Int. Conf. Mobile Data Management, Bengaluru, India, 2012, pp. 238–247.