ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol.2, Issue 12, December 2013*

# A Survey of Data Mining Techniques on Medical Data for Finding Temporally Frequent Diseases

**Mohammed Abdul Khaleel[1], Sateesh Kumar Pradhan[2], G.N.Dash[3], F. A. Mazarbhuiya[4]**

Research Scholar, Sambalpur University, India [1]

Post Graduate Department of Computer Science, Utkal University, India[2]

Post Graduate Department of Physics, Sambalpur University, India[3]

Albaha University,Albaha, KSA[4]

**ABSTRACT:** Health care domain is flooded with huge amount of data that holds sensitive information pertaining to patients and their medical conditions. Medical data mining can help obtain latent patterns or actionable knowledge. Data mining techniques can discover such latent patterns or hidden relationships among the objects in the medical data sources. This will give know how to ascertain the progression of diseases over a period of time. As medical data sources contain set of observations that are made from time to time with clinical parameters, considering temporal dimension of the data as fundamental parameter can give valuable insights related to temporal nature of diseases. The classical sequence pattern mining is not sufficient to know the temporal nature of diseases that prevail in a region or country. This is because the sequential patterns do not consider the elapsing time between events. Time-annotated sequences can bestow a novel paradigm in data mining. As temporal data mining has potential advantages, this paper focuses on finding data mining techniques that can be used to extract temporally frequent diseases. We analyze the techniques using for temporal data mining on medical data sets.

**Index Terms** –Data mining, medical data mining, data mining techniques, temporally frequent diseases

## 1. INTRODUCTION

Data mining techniques have potential to discover hidden relationships in the data of medical databases. This will help in understanding the prevailing situations in healthcare domain with respect to patients, their medical conditions and treatments. Medical databases are very bulky that need computerized programs to find latent trends that will help in medical diagnosis and treatment. In the wake of data mining techniques, especially medical data mining techniques, the health care domain has made significant progress in using the technologies in prevention and diagnosis of disease. With respect to data mining techniques, the traditional frequent pattern discovering techniques [1], [2], [3], [4], [5], [6], [7], [8] are not sufficient to know the temporal nature of diseases. These techniques do not consider the elapsed time between two events and thus cannot produce valuable insights into temporally frequent diseases as they do not take the time dimension as variable in their framework.

Spenceley and Warren [9] explored temporal data mining with respect to taking intelligent inputs to an online medical application.Catley, Stratti, and McGregor [10] emerging techniques related to temporal data mining on medical time series data sets.Catley et al. [11] have extended their work later with respect to multi-dimensional medical data.Meamarzadeh, Khayyambashi and Saraee [12] applied temporal data mining techniques that helped in discovering hidden relationships in medical data sets.Shuxia and Zheng [13] proposed fuzziness approach for mining interminacy temporal data.Tsumoto, Hirano and Iwata [14] applied temporal data mining for characterization of medical practice.Adaptive fuzzy cognitive maps are used by Froelich and Wakulicz-Deja [15] for mining temporal data in medical data sets.Berlingerio, Bonchi, Giannotti and Turini [16] believed that clinical databases contain temporal data that can be exploited to discover intelligence that supports in making decisions pertaining to patients' health and diagnosis.Abe, Yokoi, Ohsaki and Yamaguchi [17] proposed an integrated environment for medical data mining.Tsumoto and Hirano [18] explored the mining possible trajectories from medical data sets. More details about all these researches can be found in section II.

Our contributions in this paper include the analysis of state-of-the-art of the existing data mining techniques that are used for temporal data mining on medical data besides summarizing and providing future directions of the research. This remainder of this paper is structured as follows. Section II reviews related literature. Section III summarizes the findings pertaining to extracting temporally frequent diseases while section IV concludes the paper.

## II.RELATED WORKS

Temporal data mining has promising impact on the medical data mining. This is evident in literature as number of researchers worked out techniques for extracting temporally

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol.2, Issue 12, December 2013*

frequent diseases from medical data sets. The time dimension of the data sets is exploited for various purposes. Spenceley and Warren [9] explored temporal data mining with respect to taking intelligent inputs to an online medical application. These authors considered two approaches that are with case and without case approaches. Regarding with case approach, they considered individuals medical history in order to predict input requirements in the next visit. The case independent approach considers the temporal relationship among events irrespective of patients. Catley, Stratti, and McGregor [10] emerging techniques related to temporal data mining on medical time series data sets. Catley et al. mainly focused on multi-dimensional data. The dataset considered is "Neonatal Intensive Care". Six trends were identified in four categories namely knowledge base, integration, results and data. Their discoveries with multi-dimensional medical data help to address challenges associated with "course of dimensionality" in medical data besides driving towards next generation knowledge discovery. Catley et al. [11] have extended their work later with respect to multi-dimensional medical data. They applied temporal data mining techniques to emerging data streams using the same "Neonatal Intensive Care" dataset. Their approach helped in clinical investigations on multi-dimensional time-series data.

Meamarzadeh, Khayyambashi and Saraee [12] applied temporal data mining techniques that helped in discovering hidden relationships in medical data sets. These authors believed that medical data has much temporal information that needs to be exploited in order to get intelligence on temporally frequent medical events. Mining temporal relational rules was their main focus. They presented the rules that have been mined in the form of a graph for further processing. The temporal intervals are used to know the temporally frequent events that exist in medical data. Their work helped in finding the temporal frequency of early detection of high risk patients, births, deaths, and pre-mature newborns. Shuxia and Zheng [13] proposed fuzziness approach for mining interminacy temporal data. Using fuzziness concept, the authors explored the degree of interminancy in patient records of medial data set. The prototype built them has utility to demonstrate the proof of concept.

Tsumoto, Hirano and Iwata [14] applied temporal data mining for characterization of medical practice. The idea is taken from the fact that medical data has details of patients, physicians, and temporal evaluation of events. By discovering trends in the medical data, the authors proposed a tool that can help in reusing data to exploit best medical practices. Adaptive fuzzy cognitive maps are used by Froelich and Wakulicz-Deja [15] for mining temporal data in medical data sets. The technique is used to discover medical concepts from temporal data. The medical concepts thus extracted include changes in patients' conditions over a period of time, the drugs prescribed and the health effects expressed by patients. For effective knowledge representation fuzzy cognitive maps are used.

Berlingerio, Bonchi, Giannotti and Turini [16] believed that clinical databases contain temporal data that can be exploited to discover intelligence that supports in making decisions pertaining to patients' health and diagnosis. They proposed a novel data mining approach known as "Time-Annotated Sequences". Berlingerio et al. could extract interesting TAS patterns besides making a general methodology that can be used further to explore the possible discovery of temporal dimensions in medical data for better prediction of diseases and prevention of the same.

Abe, Yokoi, Ohsaki and Yamaguchi [17] proposed an integrated environment for medical data mining. The technique they used is related to time-series data mining. These authors could extract useful medical information from medical data set. They studied on the rules evolving from the data sets besides comparing their approach with other related works. Abe et al. mined time related rules from medical data set. They used visual human-system interfaces that are user-friendly to demonstrate their proof of concept.Tsumoto and Hirano [18] explored the mining possible trajectories from medical data sets. The data set considered was related to chronic hepatitis that could reflect details such as choline esterase, albumin and temporal covariance of platelets. Cjos [19] identified need for building new methods that can produce patterns considering time parameter. The disease data has been exploited in order to find trends in the data.

Tsumoto [20] opined that temporal data is one of the challenges for medical data mining. IUI dataset has been explored by Kooptiwoot [21] for medical data mining considering temporal nature of data and achieved results that satisfied domain experts. Mao et al. [22] exploited temporal data mining to build a system that warns early deterioration of patients' medical condition. Olukunle and Ehikioya [23] opined that medical data sets are characterized by temporal nature besides exhibiting missing values and long patterns. Pradhan and Prabhakaran [24] exploited hidden frequent patterns in medical data considering temporal relationships. Saraee, Ehghaghi, and Meamarzadeh [25] applied temporal data mining techniques to discover mortality related to accidents caused to children. The ensuing section summarizes the techniques used for temporal data mining.

## III.SUMMARY OF TECHNIQUES FOR TEMPORAL DATA MINING

In medical data mining temporal dimension has significant impact on discovering actionable knowledge related to prediction and detection besides forecasting of temporally frequent diseases and take necessary steps in order to ensure that such diseases are prevented in time with proper information in hand and planning priori. Various data mining techniques that have considered temporal dimension of the medical data sets are presented in table 1.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol.2, Issue 12, December 2013*

Table 1 – Summary of researches on temporally frequent data mining on medical data sets

| Research Reference (s) | Data Mining Technique (s) | Advantages | Disadvantages | Remarks |
|---|---|---|---|---|
| Spenceley and Warren [9] | Tool for online electronic medical records | Incremental data mining of existing records, ease of future data entry with respect to patients | The tool does not have provision for finding temporally frequent diseases | Two approaches such as case dependent and case independent are used for experiments |
| Catley, Stratti, and McGregor [10] | Emerging trends and patterns extraction | Discovers six patterns related to knowledge base, integration, results and data from "Neonatal Intensive Care" data set | No comprehensive framework has been proposed. | Knowledge discovery through medical data mining and inputs from domain expert are combined |
| Catley et al. [11] | Extension to Cross Industry Standard Process for Data Mining (CRISP-DM) | Helps in clinical investigations on medical data sets | Evaluation of the proposed approach has not been made | Temporal data mining techniques are applied to emerging data streams |
| Meamarzadeh, Khayyambashi and Saraee [12] | Temporal rule mining | Mining rules with temporal information can help in characterization of medical practices | Temporal data mining can be applied to other areas of medical data. | Authors presented the rules that have been mined in the form of a graph for further processing |
| Shuxia and Zheng [13] | Fuzziness approach | Mining interminacy of temporal data | Query optimization has not been made on temporal information | Authors built a tool to demonstrate the proof of concept |
| Tsumoto, Hirano and Iwata [14] | Discovers not only discovering temporally frequent diseases but also the practices made by physicians | Characterization of medical practice | More analysis can be made further | History of clinical actions is considered for research |
| Froelich and Wakulicz-Deja [15] | Adaptive fuzzy cognitive maps are used to represent knowledge | Discovers concepts from temporal data | There is no tool support for this research | Focused on application of discovered fuzzy cognitive maps |
| Berlingerio, Bonchi, Giannotti and Turini [16] | Time – Annotated Series (TAS) approach | Reusable methodology, extracting interested TAS patterns | Only few variables are considered for research | Authors considered a case study to elaborate their findings |
| Abe, Yokoi, Ohsaki and Yamaguchi [17] | Time-Series data mining | Integrated environment is created for medical data mining, rules related to time are extracted | No tool support is given | Pseudo code is provided to demonstrate the proof of concept |
| Tsumoto and Hirano [18] | Multiscale matching and clustering | Mining results reflect details such as choline esterase, albumin and temporal covariance of platelets. | Only chronic diseases are explored. | Trajectory data mining techniques are applied on chronic hepatitis dataset. |

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

*International Journal of Advanced Research in Computer and Communication Engineering*
*Vol.2, Issue 12, December 2013*

As seen in table 1, it is evident that researchers exploited for extracting trends of patterns using time as one of the fundamental variables in medical data mining. Especially characterization of diseases, effects of them on the patients, medical practices and so on are the essence of the temporal medical data mining.

## IV.CONCLUSION

In this paper we analyzed temporally frequent mining of diseases. Time dimension in medical data is considered as a fundamental variable for the analysis of frequency of diseases that prevail with respect to time. Classical frequent pattern mining cannot utilize the time interval between events and therefore it is not suitable for exploring the temporally frequent diseases. The latent trends related to this can be discovered using data mining techniques that consider the time dimension. In this paper we analyzed such data mining techniques and found that they are very useful in extracting patterns that can help in obtaining valuable insights into the temporally frequent diseases that will help medical authorities of a region or country or the world to know the frequency of diseases over a period of time. This know how helps them to take well informed decisions pertaining to public health in order to eliminate or reduce the occurrence of diseases. In fact, such information has high significance pertaining to medical research. The essence of findings include that with temporal data mining it is possible to characterize medical practices besides the frequency of diseases and their effects. This will definitely help when incorporated into an expert decision making system. Our future work is to build a tool that discovers temporally frequent diseases which will be a part of intelligent decision making system used.

## REFERENCES

[1]HAI-BING MA, JIN ZHANG, YlNG-JIE FAN, YUN-FA W. (2004). MINING FREQUENT PATTERNS BASED ON IS+-TREE. *IEEE*. 0 (0), P1208-1213.
[2] Cong-Rui Ji and Zhi-Hong Deng. (n.d). Mining Frequent Ordered Patterns without Candidate Generation. *IEEE*. 0 (0), P1-5.
[3] Hai-Tao He and Shi-Ling Zhang. (2007). A New method for Incremental Updating Frequent patterns mining. *IEEE*. 0 (0), p1-4.
[4] Carson Kai-Sang Leung∗ Christopher L. Carmichael and Boyu Hao. (2007). Efficient Mining of Frequent Patterns from Uncertain Data. *IEEE*. 0 (0), p489-494.
[5] Shariq Bashir, Zahid Halim, A. Rauf Baig. (2008). Mining Fault Tolerant Frequent Patterns using Pattern Growth Approach. *IEEE*. 0 (0), p172-179.
[6] Sunil Joshi and Dr. R. C. Jain. (2010). A Dynamic Approach for Frequent Pattern Mining Using Transposition of Database. *IEEE*. 0 (0), p498-501.
[7] Thanh-Trung Nguyen. (2010). An Improved Algorithm for Frequent Patterns Mining Problem. *IEEE*. 0 (0), p503-507.
[8] Xiaoyong Lin and Qunxiong Zhu. (2010). Share-Inherit: A novel approach for mining frequent patterns. *IEEE*. 0 (0), p2712-2717.
[9] Susan E. Spenceley and James R. Warren. (1998). The Intelligent Interface for On-Line Electronic Medical
Records using Temporal Data Mining. IEEE, p1-9.
[10] Christina Catley, Heidi Stratti and Carolyn McGregor. (2008).Multi-Dimensional Temporal Abstraction and Data Mining of Medical Time Series Data: Trends and Challenges. P4322-4325.
[11] Christina Catley, Kathy Smith, Carolyn McGregor, and Mark Tracy. (2009). Extending CRISP-DM to Incorporate Temporal Data Mining of Multidimensional Medical Data Streams: A Neonatal Intensive Care Unit Case Study. IEEE. P1-5.
[12] Hoda Meamarzadeh, Mohammad Reza Khayyambashi and Mohammad Hussein Saraee. (2009). Extracting Temporal Rules from Medical data. IEEE, p327-331.
[13] Ren Shuxia and Zhao Zheng, (2010). Mining of Indeterminacy Temporal Data Based on Fuzziness. IEEE, p391-393.
[14] Shusaku Tsumoto, Shoji Hirano and Haruko Iwata, (2012). Temporal Data Mining of Order Entry Histories for Characterization of Medical Practice. IEEE, p1-4.
[15] Wojciech Froelich, Alicja Wakulicz-Deja (2009). Mining Temporal Medical Data Using Adaptive Fuzzy Cognitive Maps. *IEEE*. P1-8.
[16] Michele Berlingerio (n.d). Mining Clinical Data with a Temporal Dimension: a Case Study. *IEEE*. p1-8.
[17] Hidenao Abe AND Hideto Yokoi (n.d). Developing an Integrated Time-Series Data Mining Environment for Medical Data Mining. *IEEE*. P1-6.
[18] Shusaku Tsumoto and Shoji Hirano, (2008). Mining Trajectories of Laboratory Data using Multiscale Matching andClustering. IEEE ISCBMS, p626-631.
[19] Krzyrztof J. Cior, (2000). Medical Data Mining and Knowledge Discovery. IEEE ENGINEERING IN MEDICINE AND BIOLOGY, p15-16.
[20] Shusaku Tsumoto (n.d). Problems with Mining Medical Data. *IEEE*. p1-2.
[21]Suwimon Kooptiwoot, (2010). IUI Mining: Template Base Variation Approach, IEEE, p282-284.
[22] Yi Mao, Yixin Chen, Gregory Hackmann, Minmin Chen, Chenyang Lu, Marin Kollef and Thomas C.Bailey, (2011). Medical Data Mining for Early DeteriorationWarning in General HospitalWards. IEEE, p1042-1049.
[23] Adepele Olukunle and Sylvanus Ehikioya, (n.d). A Fast Algorithm for Mining Association Rules in Medical Image Data. *IEEE*. p1-7.
[24] Gaurav N. Pradhan AND B. Prabhakaran (n.d). Association Rule Mining In Multiple, Multidimensional Time Series Medical Data. *IEEE*. p1-4.
[25] Mohammad Hossein Saraee, Zahra Ehghaghi and Hoda Meamarzadeh, (2008). Applying Data Mining In Medical Data, IEEE, p160-164.