Survey on Rule Mining and Temporal Patterns Representation Technique

Shraddha Kalbhor, S. V. Kedar

Student, Department of Computer Engineering, Rajashri Shahu College of Engineering, Savitribai Phule Pune University, Pune, India

Asst. Prof., Department of Computer Engineering, Rajashri Shahu College of Engineering, Savitribai Phule Pune University, Pune, India

ABSTRACT: In recent years, the health monitoring systems are improved with wearable sensors. Such system has a huge amount of health care related data. Such data can be access with a popular data mining approach. Data mining is the most motivating approach increasingly used in health organizations. This can be helpful to various peoples associated with this concept like scientific community and industries. Data mining is used to predict the health status of the patient with their sensed data based on previous health records data of patients. Such data is getting from wearable sensor devices/technology. It requires the proper analysis and modeling of sensed data. This paper presents the several recent data mining approaches combined with wearable sensor technique for a wide area of health monitoring systems. Also, present some method and algorithms used to analyze the wearable sensor data in physiological monitoring healthcare systems.

KEYWORDS: Health informatics, Data-driven modeling, pattern abstraction, physiological sensor data, sensor data analysis, temporal rule mining.

I. INTRODUCTION

Nowadays, the research area of health monitoring frameworks has moved from simple reasoning of wearable sensor readings to the higher level of data processing with a specific end goal to give substantially more data that is imperative to the end clients. Wearable health monitoring frameworks have drawn a great deal of enthusiasm from the research community and the business amid the most last few of years. Wearable frameworks for wellbeing checking may contain different types of miniature sensors, wearable or even implantable. The fundamental target of data mining in healthcare monitoring frameworks is to recover data and there are several tasks considering wearable sensors that data mining strategies can do. In last few years, temporal association rule mining strategies have been applied to a clinical data stream to perceive complex relationships of the physiological sensor observations. Data-driven strategies capable of discovering hidden and potentially useful information through the physiological sensor data and to construct models based on the experimental data.

On the expansion of wearable sensor innovation in both clinical and at home settings, the aggregation of physiological sensor information requires a focused exertion on the investigation furthermore, demonstrating of this information. By means of sensor information examination furthermore, inevitable demonstrating, it is conceivable to accomplish a more profound comprehension of the relationships between's long haul estimations of physiological parameters and medicinal conditions.

Wearable technology, wearables, fashionable technology, wearable devices, tech togs, or fashion electronics are clothing and accessories incorporating computer and advanced electronic technologies. The designs often incorporate practical functions and features.

Commonly, this process infers utilizing assorted information mining methods on sensor information with a specific end goal to obtain persistent specific models. When all is said in done, such methodologies are either information or information driven. Utilizing information driven methodology prompts a managed model of data extraction, yet data is limited with master space information. Then again, information-driven systems empower us to find covered up and possibly helpful data through the physiological sensor information and to manufacture models based on the exploratory information. So as to influence from data-driven approaches, an answer whereby shrouded examples can be caught and made express in human consumable terms, i.e., semantics, is extremely valuable.

The existing system, for rule mining, association rule mining is used. In that Apriori algorithm is used to mining the data. But it not gives exact rule set at the output and required too much time to execute dataset.

II. RELATED WORK

In paper [1] influences from rule mining methods to extract and represent the temporal relation of prototypical patterns in clinical data streams. The methodology is completely data-driven, where the temporal rules are mined from physiological time series, for example, respiration rate, blood weight, and heart rate. To accept the rules, a novel similarity technique is presented, that compares the similarity between rule sets. An additional part of the proposed approach has been to use natural language generation techniques to represent the temporal relations between patterns.

In this paper [2] has displayed a methodology for programmed rule mining and representation from physiological sensor data considering the uniqueness of rules in clinical conditions. The fundamental part of rule mining as a data-driven technique is to display the conduct of temporal patterns in physiological data streams to create a subjective rule set in a clinical setting. The proposed approach has considered nine clinical conditions, for example, angina, sepsis, and respiratory disappointment, alongside three physiological sensor data (i.e., heart rate, circulatory strain, and breath rate). This study has tended to 1) data-driven rule mining for temporal patterns, 2) particular demonstrating of the rule sets of clinical conditions and 3) literary representation of the models in an intelligible yield. To assess the uniqueness of the gave rule sets, a rule set likeness, called event proportion, has been presented, which measure the event of rules in other rule sets.

This paper [3] addresses the issue of analyzing time series with temporal association rule mining systems. Since initially association rule mining was produced for the investigation of transactional information, as it happens for instance in business sector analysis, algorithms and time series must be adapted in order to apply these systems productively to the investigation of time series in general. Continuous time series of various beginnings can be discretized in order to mine several temporal association rules, what reveals interesting coherences in one and between sets of time series. Depending upon the area, the learning about these coherences can be utilized for several purposes, e.g. for the prediction of future estimations of time series.

In this paper [4], authors develop and study a probabilistic clustering model intended to moderate the impacts of temporal sparsity inherent in electronic health care records information. We assess the model subjectively by visualizing the learned cluster parameters and, quantitatively regarding its capacity to predict mortality results associated with patient episodes.

The paper [5] presents a system of a natural language generation framework that gives a text-based representation of the extracted numeric data from physiological sensor signals. All the more particularly, a new partial pattern identification algorithm is acquainted with catching the specific changes and events of health parameters. The extricated data is then represented to considering semantic characterization of numeric features.

In [6], Subsequence time arrangement clustering investigates the subsequence clusters of time arrangement data. Numerous studies have focused on related algorithms as a subroutine in rule revelation, indexing, order, and oddity location. Authors particularly survey this clustering from the viewpoint of fundamental techniques, similitude/separation measures, challenges, assessment measurements, applications, and datasets. To outline late

improvements around there of exploration, they audit 25 standard articles on subsequence time arrangement clustering. The development of this clustering is ordered into three gatherings, to be specific, between the pre and post proof periods.

In [7], Data decrease is worried about diminishing the volume of data while holding its crucial attributes. In that capacity, testing gives a general methodology which scales well also, offers more adaptability than just following tally insights. Also, the example can better be utilized for preparing reason or assist factual examination. For the full advantage of examining, notwithstanding, it is best to tailor a testing methodology to the current issue. Examining is one of the essential systems to build the proficiency of affiliation rule mining For a particular data mining assignment under particular data set, one testing methodology might work superior to anything others as far as exactness or effectiveness.

In [8], the algorithm authors propose in this paper is focused on finding subjective rules. In any case, the rules we find are not classification rules. They have no specified classes. Maybe, they have every one of the rules that depict the relationship between sets of things. An algorithm, called the KID3 algorithm, has been introduced that can be utilized to find the kind of affiliation rules we have considered. The KID3 algorithm is genuinely clear. Credits are not limited to be double in this algorithm

Table 1: Survey Table

Sr. No.	Title	Paper Details	Method Used	Advantages	Limitations
1.	Data-Driven Rule Mining and Representation of Temporal Patterns in Physiological Sensor Data	This paper leverages from rule mining techniques to extract and represent temporal relation of prototypical patterns in clinical data streams.	Temporal Rules in Physiological Sensor Data, Temporal Rule Set Similarity, Temporal Rule Representation	This approach was able to extract more distinctive rule set, which most likely is related to the expression of richer temporal relations.	The semantic model used to improve the quality of text is limited.
2.	PPSGen: Learning-Based Presentation Slides Generation for Academic Papers	Investigate a very challenging task of automatically generating presentation slides for academic papers.	PPSGen system	This method can generate much better slides than traditional methods.	Need to improve this system by using both text and graphical elements in the paper and make slides more comprehensible and vivid
3.	About the analysis of time series with temporal association rule mining	This paper addresses the issue of analyzing time series with temporal association rule mining techniques.	Standard and temporal association rule mining approaches	Successfully extract temporal association rules which describe the close relationship between environmental factors.	could not find major differences in the results of their clustering and SAX-based discretizations
4.	Unsupervised pattern discovery in electronic health care data using probabilistic clustering models	Develop and study a probabilistic clustering model designed to mitigate the effects of temporal sparsity inherent in electronic health care records data.	Probabilistic Clustering	This model can discover distinct, recognizable physiologic patterns with prognostic significance.	Discretizing time does not solve direct computation of standard similarity scores and distances difficult problem since it yields missing data.
5.	A framework for	Presents a framework of a	Natural Language	The influence of the	Need to work in

	automatic text generation of trends in physiological time series data	natural language generation system that provides a text-based representation of the extracted numeric information from physiological sensor signals.	Generation (NLG) system	detected trends and linguistic characterisation of numeric information is studied to assist the end user by generating readable output text.	physiological data analysis may also consider the dependency of multi health parameters.
6.	A Review of Subsequence Time Series Clustering	This paper reviews some definitions and backgrounds related to subsequence time series clustering.	Subsequence time series clustering	Elucidated and compared the strengths and weaknesses of the previous literature and presented theoretical and practical issues for future study.	High Memory Usage, Unsuccessful Outcomes with Large Parameters
7.	Sampling based Association Rules Mining- A Recent Overview	Provide an overview of existing sampling based association rule mining algorithms.	Sampling based association rule mining algorithms.	Sampling is one of the important techniques to increase the efficiency of association rule mining	It is necessary to study how different strategies are in a specific data mining task given specific data sets in order to provide users a set of guidelines for them to make decisions on which context it will be more suitable to use which sampling strategy.
8.	Fuzzy rule generation for diagnosis of coronary heart disease risk using substractive clustering method	Propose to generate the rules automatically from the data collection using subtractive clustering and fuzzy inference TagakiSugeno Kang orde-1 method.	Subtractive clustering method	It is used for the diagnosis of coronary heart disease risk.	It is applicable for only this particular application.

III. PROPOSED SYSTEM

Pattern Abstraction is a method where the system generates a time series of each parameters design. Time series generate a series of parameters at an equal interval of time. Series generation means that to discrete the series in a segmentation (St = s1, s2,....,Sn). After that, with different ideas time series discretization, a sliding window method is the most famous algorithm. In a sliding window concepts window of size w with a given overlap on two regular windows. Clustering technique is used to form the cluster of subsequence series. Using a clustering algorithm the advantage is that the prototypical patterns are provided in a data-driven way without involving any domain knowledge to customize the typical patterns. Applying a K-mean clustering algorithm to form clusters on each segment. Natural language text is to put the relation and the definition of the itemsets as the antecedent and consequent in a textual format. The proposed system use Eclat algorithm for mining. In association rule, Apriori algorithm is used, in existing system but it not provides a sati-factional result and also takes more time to execute. To overcome this thing, in proposed used the Eclat algorithm. Eclat is time efficient algorithm and also save memory.
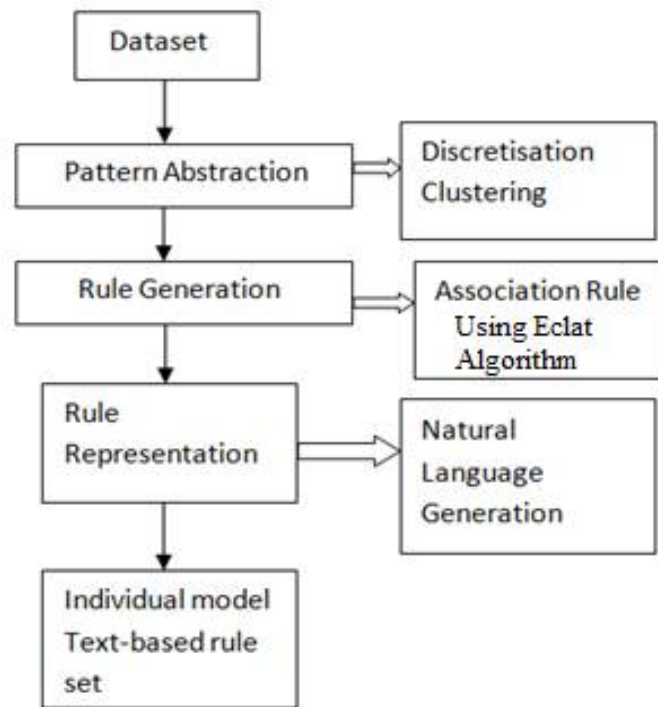
Fig 1. Propose System

IV. CONCLUSION

From this survey, we conclude that the data mining is a very important and interesting aspect in the wide area of medicine. In this paper, we make the comparative analysis of some recent data mining systems combined with wearable sensors to improve the accessibility of health careorganizations data. Such techniques are comparing on the basis of key concepts and their respective advantages and disadvantages. Knowledge mined by using such data mining techniques is useful in healthcare organizations to make an accurate decision of patient's health. This data is obtained through various wearable sensor devices.

REFERENCES

[1] HadiBanaee and Amy Loutfi, "Data-Driven Rule Mining and Representation of Temporal Patterns in Physiological Sensor Data", IEEE Journal of Biomedical and Health Informatics, Vol. 19, No. 5, September 2015.
[2] Yue Hu and Xiaojun Wan, ``PPSGen: Learning-Based Presentation Slides Generation for Academic Papers", in IEEE transactions on knowledge and data engineering, vol. 27, no. 4, April 2015.
[3] T. Schluter and S. Conrad, "About the analysis of time series with temporal association rule mining," in Proc. IEEE Symp. Comput. Intell. Data Mining, 2011, pp. 325–332.
[4] B. M. Marlin, D. C. Kale, R. G. Khemani, and R. C. Wetzel, "Unsupervised pattern discovery in electronic health care data using probabilistic clustering models," in Proc. 2nd ACM SIGHIT, 2012, pp. 389–398.
[5] H. Banaee, M. U. Ahmed, and A. Loutfi, "A framework for automatic text generation of trends in physiological time series data," in Proc. IEEE Int. Conf. Syst., Man, Cybern., 2013, pp. 3876–3881.
[6] SeyedjamalZolhavarieh, SaeedAghabozorgi, and Ying Wah The, "A Review of Subsequence Time Series Clustering", Scientific World Journal Volume 2014, Article ID 312-521.
[7] V.Umarani, Dr. M. Punithavalli, "Sampling based Association Rules Mining- A Recent Overview", V. Umarani et al. / (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 02, 2010, 314-318.
[8] L. Muflikhah and Y. Wahyuningsih, "Fuzzy rule generation for diagnosis of coronary heart disease risk using substractive clustering method," Softw. Eng. Appl., vol. 6, pp. 372–378, 2013.