International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 4 (2017), pp. 615-620

http://www.ripublication.com

Survey on the Multiple Time Series Data with Data Mining Techniques

Dr. P. Radha[1] and R.Divya[2*]

[1]Assistant professor, PG and Research Department of Information Technology, Govt. Arts, Coimbatore, Tamil Nadu, India.

[2]Research scholar, Department of Computer Science, Govt. Arts, Coimbatore, Tamil Nadu, India.

Abstract

In modern years in healthcare areas, Data mining helps to predict the diseases by advanced techniques. Data mining is deal with searching the information from the enormous datasets. It is incredibly difficult task to the researchers to predict the disease from the huge medical datasets. To conquer this problem the researcher used data mining techniques such as classification, clustering, association rules. In Data mining techniques, the classification algorithms are used to prediction process. In this paper the study is about on the Multiple-Time series and the classification techniques. In this paper, the survey is also done on the various data mining diseases which are used for disease prediction. This survey helps to know current techniques which are required in prediction of diseases.

Keywords: Classification, Multiple-Time series, temporal classification, SVM, Data mining Techniques.

1. INTRODUCTION

Data Mining is a non-trival process of classifying the valid, novel, potentially useful and ultimately understandable patterns in data. Data Mining is used to discover knowledge from data. In other words Knowledge mining from data, knowledge extraction, data/patterns analysis, data archeology and data dredging. Multiple time series data are very perceptive to analysis and predict the disease. In multiple time series data contains multiple measurement data are collected from different time interval .The selection of the data individuality are the one of the most important drawback of data processing. There are two types of data: time series data and cross-

section data. Time-series data are a progression of examine of a particular feature, which are ordered in time, and cross-section data are accumulated by examining the many features at the same time. In the time series, the features can vary over time and these modifications contain important information. Changing the data from a low-level quantitative form to a high level qualitative description is called the temporal abstraction. The method of temporal abstraction takes any raw or preprocessed data as input. Temporal classification of time-related data has assured properties that discriminate it from other classification methods and by using the characteristics of temporal data in theory, there is an improvement in the temporal classification. There are various techniques are reviewed: Support vector clustering Support vector data description, Hidden Markov Model, Pattern-Based Decision Tree, Pattern Based class-Association Rule, Classify-by-Sequence, Minimal Predictive temporal patterns.

1.2 Prediction of the Diseases with Data mining techniques

Zhong Yin et al [1] presented a novel mental workload (MWL) detection framework based on a combination of unsupervised and supervised learning strategies to improve the prediction accuracy of mental workload. The use of EEG recording for the data acquisition causes the problems of high dimensionality of the candidate EEG features vectors and also reduces the ability to determine the MWL variations and the target class labels. In the presented approach, the locally linear embed-ding (LLE), support vector clustering (SVC) and support vector data description (SVDD) techniques are combined to overcome the problems of using EEG recording. The LLE technique is employed to find the low-dimensional MWL features in the high-dimensional EEG feature space in order to extract the representative EEG markers from different cortical regions. Then the SVC-SVDD hybrid framework is employed in which the SVC technique is used to find the data clusters in EEG data space and SVDD technique is used to distinguish the blurred and overlapped cluster into two classes. Thus the presented approach improves the prediction accuracy in the three class MWL temporal data classification. Still the time delays effects, non-automatic execution of SVDD and the real-time MWL assessment problems reduce the efficiency of the approach.

Ranganatha Sitaram et al [2] presented an approach for determining the feasibility of using a multi-channel Near-infrared spectroscopy (NIRS) in the development of brain–computer interface (BCI) system. The presented approach employs temporal classification of the multi-channel NIRS signals of the motor images to improve the prediction accuracy. Initially the signal acquisition is performed and the signals are analyzed to test the presence of significant patterns in the hemodynamic response to motor imagery. Then the classification of the NIRS signals is performed offline using two pattern recognition techniques, Support Vector Machines (SVM) and Hidden Markov Model (HMM). Thus the classification problem can be resolved and the development of NIRS-BCI system can be visualized. The major drawback is the slow process of the long time constants of the hemodynamic response making NIRS–BCI system. The inability to cope with the fast NIRS signal is also a major concern.

Mohamed F Ghalwash et al [3] proposed a new early detection method called Multivariate Shapelets Detection (MSD) for early and patient-specific classification of multivariate time series clinical data. The presented approach extends the concept of univariate shapelets to multivariate shapelets to improve the prediction accuracy. The approach utilized the information gain-based distance threshold and the weighted information-gain based utility score of a shapelet to incorporate the earliness and assigns high utility score to the shapelet to improve the early detection of disease pattern change. Thus the approach can improve the early classification of the multivariate time series data. The drawback in the presented approach is that all the multivariate time series shapelets have the same starting positions which cannot be possible at all situations due to the increasing number of shapelets.

Iyad Batal et al [4] suggested an approach called the minimal predictive temporal patterns (MPTP) framework by integrating classification and pattern mining techniques for classifying Multivariate temporal data to predict the patients with disease developing risks. The presented framework resolves the problem of excessive irrelevant pattern generation caused in temporal pattern mining techniques. The electronic health records consisting of multivariate time series data are collected from the patients and the temporal domain is incorporated to determine the temporal patterns by the temporal abstractions and temporal logic to construct the classification features. The MPTP framework, which effectively filters the non-predictive and spurious temporal patterns, is utilized to automatically mine the predicted temporal patterns by integrating the pattern selection and frequent pattern mining.

Rainer Schmidt et al [5] presented a prognostic model for temporal courses for early detection of the disease risks using the multiple time series data. The prognostic model was presented by combining the temporal abstractions with case-based reasoning (CBR) for efficiently classifying the time series data. The Temporal courses are characterized by domain-dependent trend descriptions to detect the early risks of kidney courses and the influenza diseases. The prognostic model maintains the functioning details of the normal kidney courses and compares it with the current functioning of the kidney to find out the dissimilarities for detecting the abnormal functioning. Thus the multiple time series data can be utilized for effective and early detection of diseases using the prognostic model.

Riccardo Bellazzi et al [6] presented a temporal mining approach for the assessment of the clinical performance of hemodialysis (HD) services based on the automatically collected time series data. The presented approach uses two new methods for association rule discovery and temporal rule discovery are applied to the time series for executing the pre-processing techniques such as data reduction, multi-scale filtering and temporal abstractions. Initially, the time series data are represented using the temporal abstractions. Then multi-scale filtering methods are utilized to pre-process the median time series data. Then the association between the temporal abstractions and the dialysis outcomes are searched using a search algorithm based on the association rules and finally the temporal rules are utilized to classify the patterns. Thus the dialysis services can be assessed effectively. But the slow processing of the time series data is a major drawback.

Chao-Hui Lee et al [7] presented a novel data mining mechanism for predicting the occurrence of chronic diseases by utilizing both bio-signals of patients and environmental factors which are time series data. The proposed mechanism uses two data mining methods, Pattern Based Decision Tree (PBDT) and Pattern Based Class-Association Rule (PBCAR). The PBDT method integrates the concepts of sequential pattern mining technique to extract the chronic disease features and constructs classifiers using the decision tree technique. The PBCAR also uses sequential pattern mining technique but utilizes the association rule approach to construct the classifier. Thus the proposed data mining mechanism can provide easy and early prediction of the chronic disease attacks with high accuracy.

Vincent S. Tseng et al [8] proposed a novel pattern-based data mining method called classify-by-sequence (CBS) for classifying the large temporal datasets with time series data. The proposed CBS approach is based on the integration of sequential pattern mining and probabilistic induction. The CBS is performed in two stages. The representative sequential patterns are extracted as the main features for each class in the first stage while a classifier is built by scoring the features extracted from the first stage based on probabilistic induction in the second stage. Thus the CBS can provide effective classification of the time series temporal data. The CBS technique has many advantages such as easy implementation, stable accurate classification, and better execution time. But still the pre-processing techniques used in CBS reduce the overall performance to some extent.

Themis P. Exarchos et al [9] presented an optimized sequential pattern matching methodology for efficient sequence classification. The presented technique employs a two stage process to generate sequence classification model automatically. A sequential pattern mining algorithm is utilized in the first stage to extract the sequential patterns and the score of each pattern is calculated. Then the score of each class is estimated by summing the pattern scores and multiplied by a weight and the output of the first stage is taken as the classification confusion matrix of the sequences. In the second stage, an optimization technique is utilized to finding a set of weights which minimize an objective function using the classification confusion matrix. Thus the classification of the time series can be improved along with selecting an optimal set of weights for minimizing the processing time. The proposed technique has the advantages of automatic weight assignment to classes using optimization techniques and knowledge discovery in the sequential domain of application. The disadvantages of the proposed method is that it produces more patterns including the irrelevant patterns that increases the max gap values and also increases the processing time. The optimization technique used in the proposed method increases the computational effort and the overall training time which is also a major drawback.

Damian Bargiel et al [10] presented a multi-temporal classification of agriculture lands based on high resolution spotlight TerraSAR-X images. The presented classification approach uses the Maximum Likelihood classification that is based on a high amount of ground truth samples. The classification of the agricultural lands using the proposed approach based on the dual-polarized radar images improves the classification accuracy than the use of normal satellite images. The multi-temporal

DeGrandi filter is used to pre-process the images while the Maximum-Likelihood Classifier (ML) is utilized to obtain accurate classification. But the approach suffers from low accuracy in classification due to the use of Maximum Likelihood classification.

Joseph O. Sexton et al [11] suggested multi-temporal classification for the classification of land cover based on the Landsat-5 images. In this approach, a signature extension approach for dense time-series of Landsat Thematic Mapper images that is based on a single supervised classification trained over a spatially and temporally distributed reference sample. Then the image processing, sampling, and signature-modeling methods are used to maximize the classification robustness and other sources of imageto-image variation. Thus the land images can be classified by the resulting temporal patterns. The advantage of this approach is the use of high dimensional images increases the classification accuracy. But there are many drawbacks in the proposed approach. The bias correction in the image processing reduces the efficiency of classification. The changes in the atmospheric conditions also cause variant results in the classification.

CONCLUSION

In this paper, the study is done on the various methodologies in data mining used for disease prediction. Classification Accuracy has been increased by these methodologies. For clinical dataset, classification techniques are used for disease prediction with high accuracy. The time consumption of prediction of diseases was reduced by using different classification techniques.

REFERENCES

[1] Z. Yin and J. Zhang, "Identification of temporal variations in mental workload using locally-linear-embedding-based EEG feature reduction and support-vector-machine-based clustering and classification techniques," Comput. Methods Programs Biomed., vol. 115, no. 3, pp. 119–134, 2014.

[2] R. Sitaram, H. Zhang, C. Guan, M. Thulasidas, Y. Hoshi, A. Ishikawa, K. Shimizu, and N. Birbaumer, "Temporal classification of multichannel near-infrared spectroscopy signals of motor imagery for developing abrain–computer interface," NeuroImage, vol. 34, no. 4, pp. 1416–1427, 2007.

[3] M. F. Ghalwash and Z. Obradovic, "Early classification of multivariate temporal observations by extraction of interpretable shapelets," BMC Bioinformat., vol. 13, art. no. 195, 2012.

[4] I. Batal, H. Valizadegan, G. F. Cooper, and M. Hauskrecht, "A pattern mining approach for classifying multivariate temporal data," in Proc. IEEE Int. Conf. Bioinformat. Biomed., vol. 2011, Nov. 12, 2011, pp. 358–365.

[5] R. Schmidt and L. Gierl, "A prognostic model for temporal courses that combines temporal abstraction and case-based reasoning," Int. J. Med. Informat., vol. 74, nos. 2–4, pp. 307–315, 2005.

[6] R. Bellazzi, C. Larizza, P. Magni, and R. Bellazzi, "Temporal data mining for the quality assessment of hemodialysis services," Artif. Intell. Med., vol. 34, no. 1, pp. 25–39, 2005.

[7] C. H. Lee, J. C. Chen, and V. S. Tseng, "A novel data mining mechanism considering bio-signal and environmental data with applications on asthma monitoring," Comput.Methods Programs Biomed., vol. 101, no. 1, pp. 44–61, Jan. 2011.

[8] V. S. Tseng and C.-H. Lee, "Effective temporal data classification by integrating sequential pattern mining and probabilistic induction," Expert Syst. Appl., vol. 36, no. 5, pp. 9524–9532, 2009.

[9] T. Exarchos,M. Tsipouras, C. Papaloukas, and D. Fotiadis, "An optimized sequential pattern matching methodology for sequence classification," Knowl. Inf. Syst., vol. 19, no. 2, pp. 249–264, 2009.

[10] D. Bargiel and S. Herrmann, "Multi-temporal land-cover classification of agricultural areas in two European regions with high resolution spotlight terraSAR-X data," Remote Sens., vol. 3, no. 5, pp. 859–877, 2011.

[11] J. O. Sexton, D. L. Urban, M. J. Donohue, and C. Song, "Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record," Remote Sens. Environ., vol. 128, pp. 246–258, 2013.