Mining Health Examination Records by using Graphical Approach

Jayashri A. Sonawane[1], Dr. Swati A. Bhavsar [2]
Department of Computer Engineering
[1,2] MCERC,Eklahare,Nashik

Abstract- *EHR(Electronic Health Records) collects data on yearly basis and it is used in many countries for healthcare.HER(Health Examination Records) collects the data on regular basis and identifies the participants at risk that is important for early warning and prevention.the fundamental challenge is for learning classification model for risk prediction with unlabelled data and live data string that established the majority of the collected dataset.the unlabelled data string describes the participants in health examintions whose health conditions can be vary from healthy to highly risky or very ill.in this paper, we propose a graph based,semisupervised learning algorithm called*
SHG health (semi-supervised heterogenous graph on Health) for risk prediction and assessment to classify a progressively developing condition with the majority of the data unlabelled. An efficient iterative algorithm is designed and developed to proof the convergence is given.extensive experiments based on both real health examination dataset and live datasets to show effectiveness of our method.

Keywords- live data string,heteroHER,classifier.

I. INTRODUCTION

Huge amounts of Electronic Health Records (EHRs) collected over the years have provided a rich base for risk analysis and prediction. An EHR contains digitally stored healthcare information about an individual, such as observations, laboratory tests, diagnostic reports, medications, procedures, patient identifying information, and allergies. A special type of HER is the Health Examination Records (HER) from annual general health check-ups. For example, governments such as Australia, U.K., and Taiwan , offer periodic geriatric health examinations as an integral part of their aged care programs. Since clinical care often has a specific problem in mind, at a point in time, only a limited and often small set of measures considered necessary are collected and stored in a persons EHR. By contrast,HERs are collected for regular surveillance and preventive purposes, covering a comprehensive set of general health measures , all collected at a point in time in a systematic way paper proposes a semi-supervised heterogeneous graph-based algorithm called SHG-Health (Semisupervised Heterogeneous Graph on Health) as an evidence-based risk prediction approach to mining longitudinal health examination records. To handle heterogeneity, it explores a Heterogeneous graph based on Health Examination Records called HeteroHER graph, where examination items in different categories are modelled as different types of nodes and their temporal relationships may be time-consuming, finding ways of alleviating the labelling costs is critical for our ability to automatically learn such models.Sheng, W Ruan, X Li, S Wang, Z Yang[8]proposes The health risks are calculated using the information from the cause of death (COD) dataset that is linked to the GME dataset. a data mining-based method for prediction of personal health index based on annual geriatric medical examination records. Eichelberg M., Aden T., Riesmeier J., Dogac A., Laleci G.[10]propose introduction of electronic health records, boosting the efficiency of medical services at a lower cost, at the same time offering still a vast range of research challenges.In this,The analysis of the documents that were gathered through these terms yielded additional keywords and references to additional document sources.The following keywords or combinations were used: software, quality,certification,Electronic/Personal,Medical/Health Record, HER Standards, EHR certification

II. REVIEW OF LITERATURE

MF Ghalwash, V Radosavljevic, Z Obradovic[1]proposed an approach , a temporal data mining method is proposed for extracting interpretable patterns from multivariate time series data, which can be used to assist in providing interpretable early diagnosis. The problem is formulated as an optimization based binary classification task addressed in three steps. in this classification is often employed as a data exploration step, where summa- rization of the data in a target class using interpretable distinct features becomes the central task. To the best of our knowledge, the problem of extracting interpretable features for early classification on time series.Tran, T., Phung,D., Luo, W., Venkatesh, S[2]This constructs a novel ordinal regression framework for predicting medical risk stratification from EMR. First, a conceptual view of EMR as a temporal image is constructed to extract a diverse set of features.Second, ordinal modeling is applied for predicting cumulative or progressive risk. Mas S Mohktar, Stephen J Redmond, Nick C Antoniades[3] proposed The use of telehealth technologies to

remotely monitor patients suffering chronic diseases may enable preemptive treatment.As a means of detecting exacerbation earlier, and at the resolution of a single day, it has been proposed that patients with COPD might use a home telehealth service daily to evaluate their health status . Existing home telehealth services offer a range of vital sign monitoring modalities, for measurements including lungs.Jin-Mao Wei, Shu-Qin Wang, Xiao-Jie Yuan[4] proposes Cancer classification is the critical basis for patient-tailore therapy. Conventional histological analysis tends to be unreliable because different tumors may have similar appearance. Various machine learning methods can be employed to classify cancer tissue samples based on microarray data.J. Simon, Pedro J. Caraballo, Terry M. Therneau, Steven S[7] In this paper to maintain a EMR (Electronic Medical Record) and apply association rule mining to discover sets of risk factors and their. Association Rules, Survival Analysis, Association Rule Summarization. Yanbing Xue and Milos Hauskrecht[6] Learning of classification models in medicine often relies on data labeled by a human expert. Since labeling of clinical may be time-consuming, finding ways of alleviating the labeling costs is critical for our ability to automatically learn such models.Sheng, W Ruan, X Li, S Wang, Z Yang[8]proposes The health risks are calculated using the information from the cause of death (COD) dataset that is linked to the GME dataset. a data mining-based method for prediction of personal health index based on annual geriatric medical examination records.Eichelberg M., Aden T., Riesmeier J., Dogac A., Laleci G.[10]propose introduction of electronic health records, boosting the efficiency of medical services at a lower cost, at the same time offering still a vast range of research challenges.In this,The analysis of the documents that were gathered through these terms yielded additional keywords and references to additional document sources.The following keywords or combinations m were used:software, quality, certification,Electronic/Personal Medical/Health Record, HER Standards, EHR certification

III. SYSTEM ARCHITECTURE / SYSTEM OVERVIEW

Health risk prediction is necessary for prevention and proper diagnosis before disease completely developed. The proposed system is used efficient and robust classification algorithm based on live data string.the electronic health records is not good for live or current data because it collects the records on yearly basis.so, the proposed system is used to predict the future risk of the participants on live data string for prevention and early diagnosis before the disease completely developed.

IV. SYSTEM ANALYSIS

The general architecture of the project is described below

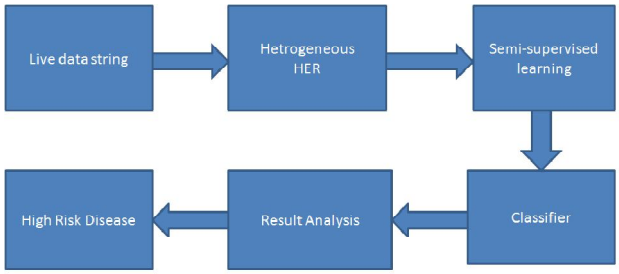

Fig. 1. fig1.Overview of the system Architecture

Live Data String :

in this, we give live data to the system which consist of known and unknown symptoms.on the basis of this data the future risks of the participants in predicted.

HeterogenousHER :

A graph represents model data that is meager. To capture the heterogeneity naturally found in health examination items, we constructed a graph called HeteroHER consisting of multi-type nodes based on health examination records.health risk prediction based on health examination records with heterogeneity in line and large unlabeled data problems, we present a semi-supervised heterogeneous graph-based algorithm called SHG-Health.

Semi-Supervised Learning :

The third component of our method is a semi-supervised learning algorithm for the constructionof HeteroHER graph The algorithm combines the advantages of for class discovery and for handling heterogeneity to isolate a specific problem caused by evidence-based risk prediction from health examination records.

Classifier :

in the system solves the problem of unsupervised learning by applying semi-supervised approach.this can be done by maintaining graph of known and unknown symptoms.these graphs are given to the classifier basically,it consist of two types of data that is,training data and testing data.in training data we have to learn classier that which symptoms are found and what to say that disease then classier gives the specified prediction of risk.

Result Analysis : in this section the high risk disease are analyzed on the basis of records obtained from classifier.

V.ADVANTAGES

1) The SHG-Health algorithm to handle a challenging multi-class classification problem with substantial unlabeled cases which may or may not belong to the known classes. This work pioneers in risk prediction based on health examination records in the presence of large unlabeled data.
2) A novel graph extraction mechanism is introduced for handling heterogeneity found in longitudinal health examination records.
3) The proposed graph-based semi-supervised learning algorithm SHG-Health that combines the advantages from heterogeneous graph learning and class discovery shows significant performance gain on a large and comprehensive real health examination dataset of participants as well as synthetic datasets

VI. CONCLUSION

The proposed system, shows data fusion for the health examination records to be integrated with other types datasets such as hospital based electronic health records and participants living conditions.a SHG algo- rithm makes use of heteroHER and semi-supervised learning for finding various known and unknown symptoms in live data which is given to the system and predict the future risk.

ACKNOWLEDGMENT

Inspiration and guidance are invaluable in every aspect of life, especially in the field of education, which I have received from my respected guide Prof. S.A.Bhavsar who has guided me and gave earnest co-operation whenever required. I would like to express my sincere gratitude towards her. She always provided me with access to the latest technology and facilities and encouragement at every point and took active participation in the achievement of my objective.

REFERENCES

[1] M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic, Extraction of interpretable multivariate patterns for early diagnostics, 2013.
[2] T. Tran, D. Phung, W. Luo, and S. Venkatesh, Stabilized sparse ordinal regression for medical risk stratication, Knowledge and In- formation Systems, Mar. 2014.
[3] M. S. Mohktar, S. J. Redmond, N. C. Antoniades, P. D. Rochford, J.J. Pretto, J. Basilakis, N. H. Lovell, and C. F. McDonald, Pre- dicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data, Articial Intelligence in Medicine, vol. 63, 2015.
[4] Q. Nguyen, H. Valizadegan, and M. Hauskrecht, Learning classication models with soft-label information. vol. 21, 2014.
[5] G. J. Simon, P. J. Caraballo, T. M. Therneau, S. S. Cha, M. R. Castro,and P. W. Li, Extending Association Rule Summarization Techniques to Assess Risk of Diabetes Mellitus, vol. 27, 2015.
[6] L. Chen, X. Li, S. Wang, H.-Y. Hu, N. Huang, Q. Z. Sheng, and M.Sharaf, Mining Personal Health Index from Annual Geriatric Medical Examinations, 2014.
[7] S. Pan, J. Wu, and X. Zhu, CogBoost: Boosting for Fast Costsensitive Graph Classication, vol. 6, 2015.
[8] M. Eichelberg, T. Aden, J. Riesmeier, A. Dogac, and G. B. Laleci,A survey and analysis of Electronic Healthcare Record standards,vol. 37,2005.
[9] C. Y. Wu, Y. C. Chou, N. Huang, Y. J. Chou, H. Y. Hu, and C. P. Li,Cognitive impairment assessed at annual geriatric health examinations predicts mortality among the elderly, vol. 67, 2014. 21
[10] L. Krogsbll, K. Jrgensen, C. Grnhj Larsen, and P. Gtzsche, General health checks in adults for reducing morbidity and mortality from disease (Review),vol no. 10, 2012.
[11] J. Kim and H. Shin,Breast cancer survivability prediction using labeled,unlabeled, and pseudo-labeled patient data, vol. 20, 2013.