Survey on Mining Health Examination Records : A Graph-based Approach

Rituja A.Bibave[1], Dr. B. L. Gunjal[2]

[1] *Student, Department of the Computer Engineering, AVCOE, Maharashtra, India*
[2] *Head,Department of the Information Technology, AVCOE, Maharashtra, India*

ABSTRACT

General health examination is an integral part of healthcare in many countries. Identifying the participants at risk is important for early warning and preventive intervention. The fundamental challenge of learning a classification model for risk prediction lies in the unlabeled data that constitutes the majority of the collected dataset. Particularly, the unlabeled data describes the participants in health examinations whose health conditions can vary greatly from healthy to very-ill. There is no ground truth for differentiating their states of health. In this paper, we propose a graph-based, semi-supervised learning algorithm called SHG-Health (Semi-supervised Heterogeneous Graph on Health) for risk predictions to classify a progressively developing situation with the majority of the data unlabeled. An efficient iterative algorithm is designed and the proof of convergence is given. Extensive experiments based on both real health examination datasets and synthetic datasets are performed to show the effectiveness and efficiency of our method.

Keyword : - *Health examination records, heterogeneous graph extraction , semi-supervised learning*

1.INTRODUCTION

Huge collected over the years have provided a rich base for UGE amounts of Electronic Health Records (EHRs) risk analysis and prediction . An EHR contains digitally stored healthcare information about an individual, such as observations, laboratorytests, diagnostic reports, medications, procedures, patient identifying information, and allergies . A special type of EHR is the Health Examination Records (HER) from annualgeneral 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 . Identifying participants at risk based on their current and past HERs is important for early warning and preventive intervention. By risk, we mean unwanted outcomes such as mortality and morbidity. Identifying the participants at risk is important for early warning and preventive intervention. The fundamental chal- lenge of learning a classification model for risk prediction lies in the unlabeled data that constitutes the majority of the collected dataset. Particularly, the unlabeled data describes the participants in health examinations whose health conditions can vary greatly from healthy to very-ill. There is no ground truth for differentiating their states of health.

2. REVIEW OF LITARATURE

Extraction of interpretable multivariate patterns for early diagnostics :-

Leveraging temporal observations to predict a patients health state at a future period is a very challenging task. Providing such a prediction early and accurately allows for designing a more successful treatment that starts before a disease completely develops. Information for this kind of early diagnosis could be extracted by use of temporal data mining methods for handling complex multivariate time series. However, physicians usually prefer to use interpretable models that can be easily explained, rather than relying on more complex black-box approaches. First,the time series data is transformed into a binary matrix representation suitable for application of classification methods. Second, anovel convex-concave optimization problem is defined to extract multivariate patterns from the constructed binary matrix. Then,a mixed integer discrete optimization formulation is provided to reduce the dimensionality and extract interpretable multivariate patterns. Finally, those interpretable multivariate patterns are used for early classification in challenging clinical applications. In the conducted experiments on two human viral infection datasets and a larger myocardial infarction dataset, the proposed method was more accurate and provided classifications earlier than three alternative state-of-the-art methods.

Stabilized sparse ordinal regression for medical risk stratification:-

The recent wide adoption of Electronic Medical Records (EMR) presents great opportunities and challenges for data mining. The EMR data is largely temporal, often noisy, irregular and high dimensional. This paper 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. The challenges are building a transparent predictive model that works with a large number of weakly predictive features, and at the same time, is stable against re-sampling variations. Our solution employs sparsity methods that are stabilized through domain-specific feature interaction networks. We introduces two indices that measure the model stability against data re-sampling. Feature networks are used to generate two multivariate Gaussian priors with sparse precision matrices (the Laplacian and Random Walk).

Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home tele-health measurement data :-

Chronic obstructive pulmonary disease (COPD) is responsible for significant morbidity and mortality worldwide.Recent clinical research has indicated a strong association between physiological homeostasis and the onset of COPD exacerbation. Thus the analysis of these variables may yield a means of predicting a COPD exacerbation in the near future.However, the accuracy of existing prediction methods based on statistical analysis of periodic snapshots of physiological variables is still far from satisfactory, due to lack of integration of long-term and interactive effects of the physiological variables. Therefore, developing a relatively accurate method for predicting COPD exacerbation is an outstanding challenge. In this paper, a regression-based machine learning technique was developed, using trend pattern variables extracted from COPD patients longitudinal physiological records, to classify subjects into low-risk and high-risk categories, indicating their risk of suffering a COPD exacerbation event. Experimental results from cross validation assessment of the classifier model show an average accuracy of 79.27 percent using this method.

3. EXISTING SYSTEM

Identifying the participants at risk is important for early warning and preventive intervention. The fundamental challenge of learning a classification model for risk prediction lies in the unlabeled data that constitutes the majority of the collected dataset. Particularly, the unlabeled data describes the participants in health examinations whose health conditions can vary greatly from healthy to very-ill. There is no ground truth for differentiating their states of health.

3.1.DISADVANTAGES OF EXISTING SYSTEM

- Record does not maintain proper classification.
- To analyzed Record classification is very difficult.

3.2. PROPOSED SYSTEM

To Propose a graph-based, semi-supervised learning al- gorithm called SHG-Health (Semi-supervised Heterogeneous Graph on Health) for risk predictions to classify a progres- sively developing situation with the majority of the data unlabeled. An efficient iterative algorithm is designed and the proof of convergence is given. Extensive experiments based on both real health examination datasets and synthetic datasets are performed to show the effectiveness and efficiency of our method.

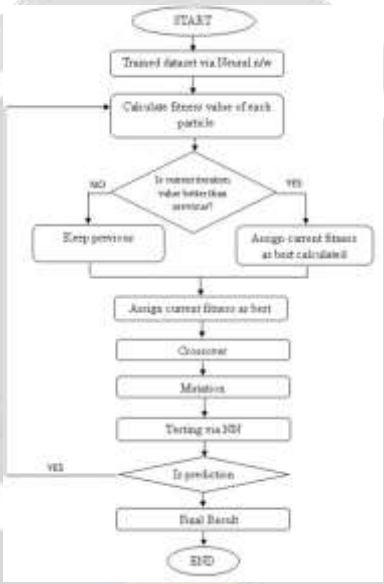

Fig.1:Proposed System

3.3. ADVANTAGES OF PROPOSED SYSTEM

- Record Map graph based to access very fastly.
- Record Examined based on Semi Supervised as well as Navie classification.
- DCG factor is very high other system.

4. CONCLUSIONS

All of the existing System having how to store record as well as some important thing related health care System.

6. REFERENCES

[1]. M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic, Extraction of interpretable multivariate patterns for early diagnostics, IEEE International Conference on Data Mining,pp.201210,2013.

[2] E. Kontio, A. Airola, T. Pahikkala, H. Lundgren-Laine, K. Junttila, H. Korvenranta, T. Salakoski, and S. Salantera, Predicting patient acuity from electronic patient records. Journal of Biomedical Informatics, vol. 51, pp. 813, 2014.
[3] T.Tran, D. Phung, W. Luo, and S. Venkatesh, Stabilized sparse ordinal regression for medical risk stratification, Knowledge and Information Systems, pp. 128, Mar. 2014.
[4] Q. Nguyen, H. Valizadegan, and M. Hauskrecht, Learning classification models with soft-label information. Journal of the American Medical Informatics Association : JAMIA, vol. 21, no. 3, pp. 5018, 2014.
[5] 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, in 2014 IEEE International Conference on Data Mining, 2014, pp. 761766 .