International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES)
Impact Factor: 5.22 (SJIF-2017), e-ISSN: 2455-2585
Volume 4, Issue 8, August-2018

A Mining Health Examination Records Using Graph-based Approach

V.Manimalar[#1], Mr. V.P.Muthukumar[#2]

M.Phil (full-time) Research Scholar, Assisstant Professor,
PG and Research Department of Computer Science,
Vivekananda College of Arts and Sciences for Women,
Elayampalayam, Tiruchengode, Namakkal, Tamilnadu, India.

ABSTRACT

In lots of real-international applications, classified instances are usually limited and expensively Amassed, whilst the most times are unlabeled and the amount is regularly sufficient. Consequently, semi supervised mastering (ssl) has attracted much attention, on account that it's far an powerful device to find out the unlabeled instances. Generally, fitness examination is an critical approach which can be utilized in more than one nations to perceive the health statistics. To identify the chance factors which can be warning and prevention in lots of illnesses is crucial. This is the predominant challenge to classify this risk elements utilized in unlabeled information which includes the dataset. Health kingdom situation can adjustments unexpectedly from healthful to very-unwell.

There's no specific base for differentiating the nation of fitness method. To advocate a graph-based, semi-supervised gaining knowledge of algorithm referred to as shg health (semi-supervised heterogeneous graph on health) is used for risk predictions. So many efficient healths gaining knowledge of approach is to be had to recognize any unlabeled dataset.

Key words: Health Examination Records, Heterogeneous Graph Extraction, Semi-Supervised Learning.

I. INTRODUCTION

Recently, the interest of Machine Learning (ML) has increased in artificial intelligence (AI), including the theoretical and practical aspects. The performance of most ML algorithms heavily depends on the scale of labeled instances, such as deep learning. huge amount of data collected to provide a rich base for electronic health records (EHRs) for risk analysis and prediction. An EHR stored healthcare information digitally about an individual. This digital information contains healthcare information such as analysis, various laboratory tests, records diagnostic reports, used medications, procedure used, patient recognize data, various allergies and diseases Health Examination Records (HERs). This is the major challenge to classify risk factors used in unlabeled data which contains the datasets. Health condition can change rapidly from healthy to very-ill. So, unlabeled data contains records of such health examination.

There is no special base facility for differentiating state of health process. Most semi-supervised learning approaches design specialized learning algorithms to effectively utilize both labeled and unlabeled data. However, it is often the case that a user already has a favorite (well-suited) supervised learning algorithm for his application, and would like to improve its performance by utilizing the available unlabeled data. In this light, a more practical approach is to design a technique to utilize the unlabeled samples, regardless of the underlying learning algorithm.

II. LITERATURE REVIEW

Leveraging temporal observations to predict a patient's 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 [1]. Information for this kind of early diagnosis could be extracted by use of temporal data mining methods for handling complex multivariate time series [2]. 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 [3].

The algorithms used in this system are association rule mining and k-anonymity which are comprised under one single algorithm, which is named as Semi Supervised Heterogeneous graph on Health (SHG-Health) algorithm [4]. The association rule mining is a data mining technique used in situations involving a huge amount of tabular represented data, which on a certain condition needs to be classified together into a single table, this makes that classification easier to get implemented[5][6]. It basically associates the data taken from different tuples together as per the condition mentioned as once it is met, the associatively is made and the records are automatically organized into the new table[7][8].

III. Semi Supervised Heterogeneous Algorithm (Sshl)

1) SVM-Algorithm

Step 1. Start
Step 2. . First select the query Q (Z)
Step 3. Remove unwanted word from query
Step 4.Vector model process, Divide neural word from− one site and non-neural word from one site
Step 5. Matching Relevance word-–>Semi- Supervised Learning
Step 6. Result

2) SHG-Algorithm

Step 1. Start
Step 2. Node Insertion
Step 3. Node Typing
Step 4. Link Insertion
Step 5. Result

3.1 MATHEMATICAL MODEL

S is the system such that

S = I, F, O

Where,

I is the input to the system

F is system functions

O is Systems output

Function F can be defined as:

F=Q, Sv, U, P, D

Where,

Q(Z)= Query

S(Z)= solved Disease Problem

U(Z)= Un-Solved Disease Problem

P(Z)= Patient

D(Z)= Doctor

Such that,

Q(Z) = q1,q2,q3…….

S(Z) = s1,s2,s3,.....

U(Z) =u1,u2,u3.....

P(Z) = p1,p2,p3.....

D(Z) = d1,d2,d3....

Patient check own query in this technique through Semi-Supervised Learning Algorithm. Easily Known which type of disease occurs which disease is very easily recover. Also, easily Crack which doctor treatment is best.

3.2Semi-supervised vs. transductive learning

Labeled data (Xl , Yl) = {(x1:l , y1:l)}

Unlabeled data Xu = {xl+1:$_n$}, available during training

Test data X_{test} = {x_n+1:}, not available during training

Supervised learning (classification, regression) {($x_{1:n}$, $y_{1:n}$)}

Semi-supervised classification/regression {($x_{1:l}$, $y_{1:l}$), $x_{l+1:n}$, x_{test}}

Transductive classification/regression {($x_{1:l}$, y1:l), $xl_{+1:n}$}

Semi-supervised clustering {$x_{1:n}$, must-, cannot-links}

Unsupervised learning (clustering) {$x_{1:n}$}

3.3Graph-based semi-supervised learning

The graph Nodes: Xl ∪ Xu

Edges: similarity weights computed from features, e.g.,

k-nearest-neighbor graph, unweighted (0, 1 weights)

fully connected graph, weight decays with distance

$w = \exp(-\|x_i - x_j\|^2/\alpha^2)$

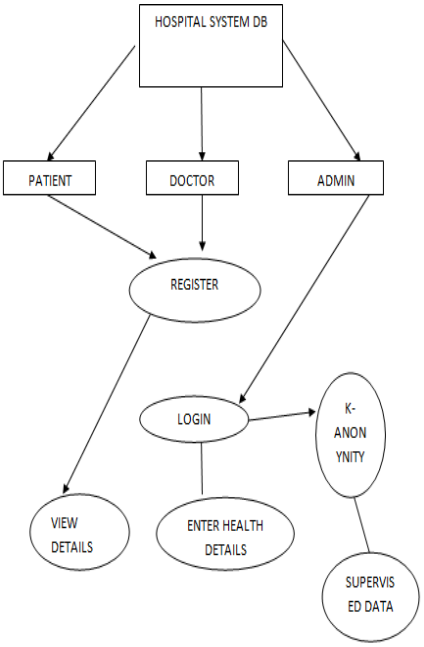

Fig: Architecture diagram

IV. ANALYSIS AND RESULT

Association rule mining to identify sets of risk factors and the corresponding patient subpopulations that are at significantly increased risk of progressing to diabetes. The system found that the most important differentiator between the algorithms is whether they use a selection criterion to include a rule in the summary based on the expression of the rule or based on the patient subpopulation that the rule covers. Also the regular diet information are given according to the patients risk stage which is maintained as an updating diet specification of record examined based on Semi Supervised as well as Naive classification.

V. CONCLUSION

In this work , In enhancement technique, as per patient queries we can recognize which precaution needs to be taken. Confirmation about precaution can be given by mail or by own account, also can be recommended to relevant patients friend. Finally we conclude that how we can manage health care dataset and also help people live happily. In future work, by using SHG algorithm techniques with efficient classification techniques we can provide better service to the patients. The most important results obtained from the above techniques are reorganization of data and information for easier access, the more effective Efficiency and better data classification in order to obtain best results . Data mining can identify important customers in data warehouse. Our approach has follow advantages: firstly, semi-supervised machine learning techniques are used to automatically construct customer behavior modeling to improve accuracy. Secondary, a neural network can be used for data visualization purposes. Data visualization is an important step in data analysis and can provide valuable insights into all steps of classification.

VI. REFERENCES

[1] J. M. Toi, S. Q. Wang, and X. J. Yuan,Ensemble rough hypercuboid approach for classifying cancers, IEEE Transactions on Knowledge and Data Engineering, vol. 22, no. 3, pp. 381391, 2010.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] C. Blake, E. Keogh, and C. J. Merz, "UCI repository of machine learning databases,"[http://www.ics.uci.edu/~mlearn/MLRepository.html], Department of Information and Computer Science, University of California, Irvine, CA, 1998.

[7] A. Blum and S. Chawla, "Learning from labeled and unlabeled data using graph mincuts," in Proceedings of the 18th International Conference on Machine Learning, Williamston, MA, 2001, pp. 19–26.