WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

SHG-Health Algorithm for Risk Prediction using Graph Based Approach

G. MURUGAN

Assistant Professor, Dept. of Computer Science and Business Systems, E. G. S. Pillay Engineering College (Autonomous), Nagapattinam

ABSTRACT-

Modern computer systems and fast evoluation of information technology, traditional healthcare is moving towards a more eletronic world. The fundamental challenge of learning a classification model for risk forecast lies within the unlabeled knowledge that establishes the bulk of the collected dataset.. There's no ground truth for discriminating their states of health. Significantly, the unlabeled knowledge describes the contributors in health investigations whose health conditions will vary greatly from healthy to very-ill.. 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. Extensive experiments based on both real health examination datasets and synthetic datasets are performed to show the effectiveness and efficiency of our method.

Index Terms- Health examination records, semi-supervised learning, heterogeneous graph extraction

1. INTRODUCTION

Data mining refers to extracting or mining knowledge from large amounts of data. The term is actually a misnomer. Remember that the mining of gold from rocks or sand is referred to as gold mining rather than rock or sand mining Thus, data mining should have been more appropriately named "knowledge mining from data," Which is unfortunately somewhat long," knowledge mining," a shorter term, may not reflect the emphasis on mining from large amounts of data.

Never the less, mining is a vivid term characterizing the process that finds a small set of precious nuggets from a great deal of raw material. Thus, such a misnomer that carries both "data" and "mining" became a popular choice. Many other terms carry a similar or slightly different meaning to data mining, such as a knowledge mining from data, knowledge extraction, data/pattern analysis, data archaeology, and data dredging. Many people treat data mining as a synonym from another popularly used term health mining

Large amounts of electronic health records(HER) collected over the years have provided a rich base for risk analysis and prediction. An HER contains digitally stored healthcare information about an individual, such as observations, laboratory test, diagnostic reports, medications, procedures, patient indentifying infromation and allergies. A special type of HER is the Health Examination Records(HER)from annual general health check-ups. 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 person's EHR. Identifying participants at risk based on their current and past HER'S is important for early waring and preventive intervention. By risk we mean unwanted outcomes such as mortality and morbidity.

There is no ground truth for differentiating their states of health 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.

Most existing classification methods on healthcare data do not consider the issue of unlabeled data. In this study we formulated the task of risk prediction as a multi-class classification problem using the Cause of Death (COD) information as labels, regarding the health-related death as the "highest risk".

A participant may have a sequence of irregularly time stamped longitudinal records, each of which is likely to be sparse in terms of abnormal results, and test items are naturally in categories, each conveying different semantics and possibly contributing differently in risk identification.

1. The existing classification methods on health data do not consider the issue of unlabeled data.
2. They either have expert-defined low risk or control classes or simply treat non-positive cases as negative.
3. Methods that consider unlabeled data are generally based on Semi-Supervised Learning (SSL) that learns from both labeled and unlabeled data.

Amongst these SSL methods, only handle large and genuinely unlabeled health data.

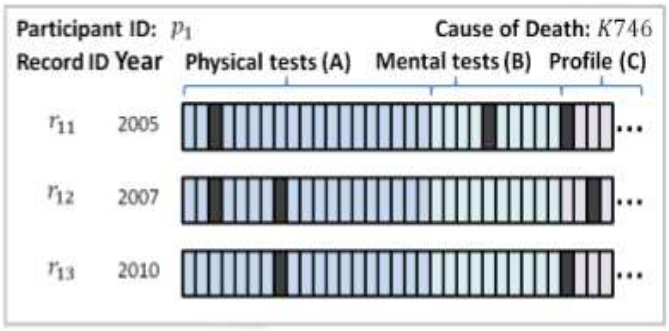

Fig.1: An example of health examination records of participant p1 who took examinations in three non-consecutive years, 2005, 2007, and 2010. Test items are in different categories (A,B,...) and the abnormal results are marked black. The main cause of death of p1 was cirrhosis of liver encoded as K746.

It present the SHG-Health algorithm to handle a challenging multi-class classification problem with substantial unlabeled cases which may or may not belongs to the known classes. This work pioneers in risk prediction based on health examination records in the presence of large unlabeled data. A novel graph extraction mechanism is introduced for handling heterogeneity found in longitudinal SHG Health that combines the advantages health examination records. The proposed graph-based semi-supervised learning algorithm.

Health examination records are represented as a graph that associates all relevant cases together.

1. Multi-typed relationships of data items can be captured and naturally mapped into a heterogeneous graph.
2. Features can be weighted in their own type through a label propagation process on a heterogeneous graph.

2. SHG-HEALTH:

To solve the problem of health risk prediction based on health examination records with heterogeneity and large unlabeled data issues, we present a semi-supervised heterogeneous graph-based algorithm called SHG-Health. The semi-supervised learning problem is formulated as follows: Problem Definition 1. Given a set of health examination records of n participants S = {s1,...,sl,sl+1,...,sn}, where si = {ri1,...,rini} is the set of ni records of participant i and rij is a tuple (xij,tij) such that xij∈ Rd is a d-dimensional vector for the observations at time tij, and a set of labels C = {1,...,c},thefirst l participants si (i ≤ l) are labeled as yi∈ C and the remaining u = n−l participants sl+1,...,sl+u are unlabeled (l u). The goal is to predict for unlabeled si(l <i ≤ n) a label yi∈ ˜ C = {1,...,c,c+1}where c+1 gives a mechanism to handle an additional class for unknown cases. An overview of our proposed solution to the problem is included .Our SHG-Health algorithm takes health examination data (GHE) and the linked cause of death (COD) labels described in as inputs. Its key components are a process of Heterogeneous Health Examination Record (HeteroHER) graph construction and a semi-supervised learning mechanism with label propagation for model training. Given the records of a participant pi as a query, SHG-Health predicts whether pi falls into any of the high-risk disease categories or “unknown” class whose instances do not share the key traits of the known instances belonging to a high-risk disease class.

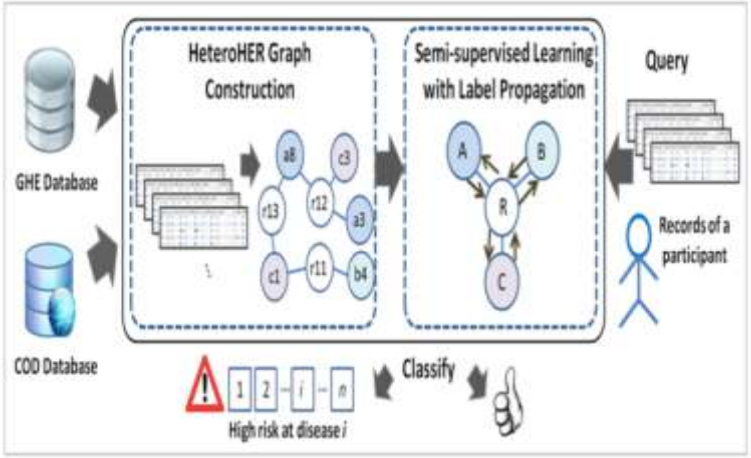

Fig.2 Proposed view of SHG Algorithm

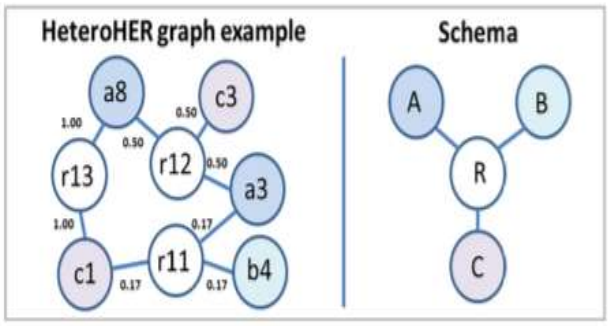

Fig.3 Proposed System of SHG Health Algorithm

2.1 HeteroHER

Graph A graph representation allows us to model data that is sparse. 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. Graph Construction The process of HeteroHER graph construction includes the following steps:

Step 1. Binarization: As a preparatory step, all the record values are first discretized and converted into a 0/1 binary representation, which serves as a vector of indicators for the absence/presence of a discretized value. Specifically, real values, such as age, are first binned into fixed intervals (e.g., 5 years). Then, all the ordinal and categorical values are converted into binary representations.

Step 2. Node Insertion: Every element in the binary representation obtained in Step 1 with a value "1" is modeled as a node in our HeteroHER graph, except that only the abnormal results are modeled for examination items (both physical and mental). This setting is primarily based on the observation that physicians make clinical judgements generally based on the reported symptoms and observed signs, and secondarily Sor the reduction of graph density.

Step 3. Node Typing: Every node is typed according to the examination category that its original value belongs to, for example, the Physical tests (A), Mental tests (B), and Profile (C) in Fig. 1. In addition, a new type of nodes is introduced to represent individual records such as r11, r12, and r13 in the same figure. All the other non-Record type nodes that are linked to the Record type nodes can be seen as the attribute nodes of these Record type nodes. In other words, categories A, B, and C in Fig. 1 can be regarded as the attributes of the Record type at a schema level. This leads to a graph schema with a star shape as shown on the right which is known as a star schema. Note that type scan often be hierarchically structured and thus choosing the granularity of node type may require domain knowledge or be done experimentally.

Step 4. Link Insertion: Every attribute (non-Record) type node is linked to a Record type node representing the record that the observation was originally from. The weight of the links is calculated based on the assumption that the newer a record the more important it is in terms the graph on the left shows a HeteroHER graph extracted from the example in Fig. 1. For instance, there is a link between r11 (the first record of p1) and a3 (the third item of category A) if the result of a3 is abnormal in r11. The link is weighted using Eq. 1. The star-shaped schema on the right is a type-level schema of such a graph. of risk prediction. A simple function g(·) can be defined as:

$$g(t) = (t - s + 1)/l \qquad (1)$$

Wheretisthetimeofcurrentrecord, list he time window of interest, and s is the starting time of the time window. Other functions such as truncated Gaussian distribution and Chi Squared distribution can also be used. The window length is the time period of records considered by the model. Note that the window length only sets the scope. It is the link weighting function that controls the contribution of time t records to the model. The two should be considered together according to domain knowledge and/or experimentally.

We include as an example based on the records of participant p1 in Fig. 1 to illustrate the process. In this simplified example, we assume all the values of examination items are binary. Different types of examination items are treated as different types of nodes on the graph. An abnormal result of the item of type Z in the jth record of the kth participant is represented as a link between nodes rkj and zi. For instance, there is a link between r11 and a3 in the left sub-figure , and the weight of the link is (2005 − 2005 + 1)/6 ∼ = 0.17 using Eq. (1) with a window width equal to 6 years. The output of the graph construction process is a heterogeneous graph represented as a set W of sparse matrices Wij for any two node types i,j that are linked to each other in the schema.

2.1.1 Normalized Weights To strengthen the weights in the low density region and weaken the weights in the high density region, the weights Wij for i,j = 1,...,m are further normalized by the row sum and column sum as in :

$$\tilde{W}_{ij} = D_{ij}^{-1/2} W_{ij} D_{ji}^{-1/2} \qquad (2)$$

Where dij,pp = $\sum_q$ Wij,pq is the sum of row p in Wij and Dij is an ni-by-ni diagonal matrix with the (p,p) element as dij,pp.

2.2 Semi-supervised Learning on HeteroHER Graph The second component of our method is a semi-supervised learning algorithm for the constructed HeteroHER graph

The algorithm combines the advantages for class discovery and for handling heterogeneity to solve a specific problem induced by evidence-based risk prediction from health examination records. In this section, we first define an objective function for the learning problem and show its convexity, followed by an optimization procedure to solve the problem. Then we derive an efficient iterative algorithm and show its convergence. Finally, time complexity is discussed.

2.1.2 Notations

Let us start with definitions and notations for the following discussions. Assume there are c classes and there is one additional"unknown"classforthecasesthatarenotknown to belong to any of the c disease classes. In this work we attach label information of a participant to the Record type nodes representing their examination records. However, the modelis generalenoughtoinclude labelsfordifferenttypes ofnodes.DefineY = [Y1,...,Ym]T ∈{0,1}Pm ini×(c+1) such that Yi = [yi1,...,yini]T ∈{0,1}ni×(c+1) encodes the labels of type i nodes. Let y(k) ip be the kth element of vector yip. If xip, i.e., node p of type i, is labeled, y(k) ip = 1 if xip belongs to class k; otherwise y(k) ip = 0. If xip is unlabeled, y(c+1) ip = 1. By doing so, we set the initial labels of the unlabeled data to be the unknown class. However, we will show later that these initial labels for the unlabeled data have little influence on learning their labels. In addition, we designed the computed labels to be soft labels. Soft labels are especially desirable for medical applications because knowing to what degree of certainty a person is classified into is sometimes as important as knowing the class itself. Let F = [F1,...,Fm]T ∈RPmini×(c+1) be the computed soft labels of m node types such that Fip∈ Rc+1 is a vector indicating the degree of certainty that xip belongs to any of the c+1 classes. The class label of xip iscomputedas argmaxk≤(c+1)F(k) ip . Fi canbeinitialized uniformly amongst type i nodes for i = 1,...,m.

2.1.3 Objective Function

We considered a regularized framework on a heterogeneous graph for our problem. Denote tr(·) as trace and denote k·kF as the Frobenius norm of matrix, i.e., kMk2 F = tr(MTM). The classification problem can be viewed as an optimization problem that minimizes an objective function J(F):

$$J(F) = \sum_{ij}^{m} \gamma_{ij} \sum_{p}^{n_i} \sum_{q}^{n_j} \tilde{W}_{ij,pq} \|F_{ip} - F_{jq}\|_F^2 + \sum_{i}^{m} \sum_{p}^{n_i} \mu_{ip} \tilde{d}_{ip} \|F_{ip} - Y_{ip}\|_F^2 \quad (3)$$

where ˜ Wij is the normalized weights on the links between type i and j nodes as defined in Eq. (2), and F and Y are the same as defined in .The first term is the smoothness constraint based on the assumption that the computed labels between the connected nodes in the graph should be close. Let z = [z1,...,zm]T such that 0 ≤ zi≤ 1 be the weights for m node types. Then, γij is defined as between-type weight of type i and type j nodes as follows:

$$\gamma_{ij} = \begin{cases} \frac{1}{2} z_j & \text{if } i = j \\ z_j & \text{otherwise} \end{cases} \quad (4)$$

where the weight is reduced to half for links between sametype nodes to avoid double counting in the summation. The second term is the fit ness constraint that penalizes when the computed F is different from labels Y. Let ˜ dij,pp = $\sum_q$ Wij,pq and ˜ Dij be an ni-by-ni diagonal matrix with the (p,p) element as ˜ dij,pp. Then ˜ dip in Eq. (3) is the degree of node p of type i, weighted by its connected node type, i.e.,

$$\tilde{d}_{ip} = \sum_{j}^{m} \sum_{p}^{n_i} z_j \tilde{D}_{ij,pp}.$$

On the other hand,µip> 0controls the trade off between the smoothness and fitness co nstraints for node p of type i.

2.1.4 Convexity The strict convexity of Eq. (3) minimization is derived in this section. For clarity, we will discuss two terms in Eq. (3) separately. The first term of the objective function denoted as J1(F) can be derived as follows:

$$\begin{aligned} J_1(F) &= \sum_{i}^{m} \Big(\sum_{j}^{m} \gamma_{ij} \sum_{p}^{n_i} \sum_{q}^{n_j} \tilde{W}_{ij,pq} \|F_{ip} - F_{jq}\|_F^2 \Big) \\ &= \sum_{i}^{m} \sum_{j}^{m} \gamma_{ij} tr(F_i^T \tilde{D}_{ij} F_i - 2F_i^T \tilde{W}_{ij} F_j + F_j^T \tilde{D}_{ji} F_i) \\ &= tr\Big(\sum_{i}^{m} \sum_{i \neq j}^{m} z_j (F_i^T \tilde{D}_{ij} F_i - 2F_i^T \tilde{W}_{ij} F_j + F_j^T \tilde{D}_{ji} F_i) \\ &\quad + z_i F_i^T (\tilde{D}_{ii} - \tilde{W}_{ii}) F_i \Big) \end{aligned} \quad (5)$$

where ˜ is defined in the same way as in Eq. (2) and ˜ Dij and γij the same previous section. Suppose the total number of nodes n =Pm ini. Let ˜ L be a n-by-n block matrix. Let its (i,j) block ˜ Lij = ˜ Dij − ˜ Wij be a Laplacian matrix with normalized weights. Eq. (5) can be transformed to the following matrix expression:

$$J_1(F) = tr(F^T I_z \tilde{L} F) = tr(F^T H F) \qquad (6)$$

Where Iz is a block diagonal matrix with the elements of the diagonal of (i,i) block equal to zi, and H = Iz ˜ L. The second term of the objective function can be derived as follows:

$$\begin{aligned} J_2(F) &= \sum_{i}^{m} \sum_{p}^{n_i} \mu_{ip} \tilde{d}_{ip} \|F_{ip} - Y_{ip}\|_F^2 \\ &= tr(\sum_{i}^{m} (F_i - Y_i)^T U_i \tilde{D}_i (F_i - Y_i)) \end{aligned} \qquad (7)$$

where the diagonal matrix ˜ Di = Pm j zj ˜ Dij and its (p,p) entry is ˜dip asdefinedearlierinUi isadiagonal matrix such that Ui,pp = µip. Combining Eq. (6) and Eq. (7), the objective function can be transformed into:

$$J(F) = tr(F^T H F) + tr((F - Y)^T U \tilde{D} (F - Y)) \qquad (8)$$

where ˜ D is a block diagonal matrix with the diagonal of the (i,i) block equal to ˜ Di.

It is easy to verify that ˜ L is positive semi-definite and likewise, H, U, ˜ D, U ˜ D and their traces. Therefore the objective function is strictly convex.

4.2.4 Optimization Procedure The closed-form solution for minimizing Eq. (8) can be obtained by setting the partial derivative of J(F) with respect to F to zero:

$$\left.\frac{\partial J(F)}{\partial F}\right|_{F=F^*} = 2(HF^* + \tilde{D}U(F^* - Y)) = 0 \qquad (9)$$

where we use the fact that H = Iz ˜ L is symmetrical and that ˜ D and U are diagonal. By multiplying ˜ D−1 on both sides of Eq. 9 and rearranging the equation, we have:

$$F^* = (\tilde{D}^{-1} I_z \tilde{L} + U)^{-1} U Y \qquad (10)$$

By using the fact that U = I−Iα Iα = Iβ Iα , we have:

$$\begin{aligned} F^* &= (\tilde{D}^{-1} I_z \tilde{L} I_\alpha + I_\beta)^{-1} I_\beta Y \\ &= (I_\alpha (I - I_z P) + I_\beta)^{-1} I_\beta Y \\ &= (I - I_\alpha I_z P)^{-1} I_\beta Y \\ &= (I - \hat{P})^{-1} I_\beta Y \end{aligned} \qquad (11)$$

where ˜ L = ˜ D− ˜ W as before, P = ˜D−1 ˜ W, and ˆ P = IαIzP. Note that the ∞−norm of IαIzP is lower than 1 given 0 ≤ ziαi< 1(i = 1,...,n). Hence the spectral radius of ˆ P is not greater than the∞−norm. So (I − ˆ P) is invertible.

2.2 Iterative Solution

An iterative algorithm is often more efficient than a closedform solution with matrix inverse. Here we describe an iterative solution for F and prove its convergence. The optimal Fi for type i nodes can be computed with the following update rule for i = 1,...,m:

$$F_i(t+1) = I_{\alpha_i} \sum_{j}^{m} z_j P_{ij} F_j(t) + I_{\beta_i} Y_i \qquad (12)$$

Where Pij = ˜ Di ˜ Wij, Iβi = I −Iαi, and zj is a type weight scalar as defined earlier in. Eq. (12) bears a label propagation interpretation. Each node iteratively spreads label information to its neighbors until a global stable state is achieved. Particularly, zjPij can be seen as the normalized links from type j nodes to type i nodes, scaled by the source type weight zj. Soft labels of type i nodes at (t + 1) are determined by two factor, , 1) the computed label scores of neighboring nodes at time t propagated via links, and 2) the initial labels for type i nodes.The diagonal matrix Iαi controls the trade-off between these two influences. It is important to note that Iαi provides a mechanism to learn an extra outlier class via an instance-level control over the trade-off. Specifically, αl and αu are introduced as parameters in the range of [0,1] that control the influence from the labeled data and unlabeled data respectively. The parameters αip (p = 1,...,ni) in Iαi are defined as αip = αl if xip is labeled, and αip = αu if otherwise. The larger αl and αu are, the less influence initial labels from Yhas.

Algorithm 1 SHG-Health

Input: a set of health examination records of n participants S, the corresponding encoded labels Y
Output: optimized F as the computed soft labels

1: $W \leftarrow$ graph construction from S (Section 4.1.1).
2: Calculate the normalized weights for $i, j = 1, ..., m$ by: $\tilde{W}_{ij} = D_{ij}^{-1/2} W_{ij} D_{ji}^{-1/2}$ (2)
3: Initialize F_i uniformly amongst type i nodes for $i = 1, ..., m$.
4: $t = 1$
5: **repeat**
6: Update F_i for $i = 1, ..., m$ by: $F_i(t+1) = I_{\alpha_i} \sum_j^m z_j P_{ij} F_j(t) + I_{\beta_i} Y_i$ (12)
7: $t = t + 1$
8: **until** convergence
9: **return** F

Particularly, when αu is set to a value extremely close to 1, it means that the initial labels of the unlabeled data play almost no role in the learning so that the computed label for annul labeled case is basically determined by its connectivity in the graph. This mechanism allows the algorithm to learn an additional (c + 1) class for nodes that are less connected to the labeled nodes from high risk disease classes. The complete algorithm of SHG-Health, combining the graph construction and iterative solution, is summarized in Algorithm 1.

2.2.1. Convergence

The proof of the convergence of Eq. (12) is as follows: Let Iα, Iβ, zj and P = ˆ D−1 ˆ W be the same as defined earlier . The update rule Eq. (12) for type i = 1,...,m can be reorganized as:

$$\begin{aligned} F_i(t+1) &= I_{\alpha_i} \sum_j^m z_j P_{ij} F_j(t) + I_{\beta_i} Y_i \\ &= I_{\alpha_i} P_i I_{z_i} F(t) + I_{\beta_i} Y_i \\ &= \hat{P}_i F(t) + I_{\beta_i} Y_i \end{aligned} \tag{13}$$

where ˆ Pi = IziIαiPi. It is equivalent to the following expression:

$$F(t+1) = \hat{P} F(t) + I_\beta Y \tag{14}$$

It has been proved in that F∗ = lim t→∞ F(t) = (I − ˆ P)−1Iβ Y , which is equivalent to the close-form solution expressed in Eq. (11) and hence our proof is completed.

Time Complexity Analysis Now we analyze the computational time complexity of the iterative solution (Step 3-8 of Algorithm 1). Step 3 takes O(k|V|) time for initialization, where k is the number of classes (i.e., k = c + 1) and |V| is the number of nodes in the graph. At each iteration of Step 6, every link needs to be processed twice, once for the node at each end of the link. This is done for every class, and consequently takes O(k|E|) time, where |E|is the number of links. Also, another O(k|V|) time is needed for incorporating IβiYi. Therefore,thetotaltimefor each iteration is (k(|E|+|V|)),

3. CONCLUSION

Mining health examination data is challenging especially due to its heterogeneity,intrinsic noise, and particularly the large volume of unlabeled data. In this paper, we introduced an effective and efficient graph-based semi-supervised algorithm namely SHG-Health to meet these challenges.

Firstly, health examination records are represented as a graph that associates all relevant cases together. This is especially useful for modeling abnormal results that are often sparse.

Secondly, multi-typed relationships of data items can be captured and naturally mapped into a heterogeneous graph. Particularly, the health examination items are represented as different types of nodes on a graph, which enables our method to exploit the underlying heterogeneous sub graph structures of individual classes to achieve higher performance.

Thirdly, features can be weighted in their own type through a label propagation process on a heterogeneous graph. These in-class weighted features then contribute to the effective classification in an iterative convergence process.

4. FUTURE ENHANCEMENT

Our work shows a new way of predicting risks for participants based on their annual health examinations. Our future work will focus on the data fusion for the health examination records to be integrated with other types of datasets such as the hospital-based electronic health records and the participants living conditions (e.g.Diets and general exercises). By integrating data from multiple available information sources, more effective prediction may be achieved.

Reference:

[1]. M. F. Ghalwash, V. Radosavljevic, and Z. Obradovic, "Extraction of interpretable multivariate patterns for early diagnostics," IEEE International Conference on Data Mining, pp. 201–210, 2013.

[2]. T. Tran, D. Phung, W. Luo, and S. Venkatesh, "Stabilized sparse ordinal regression for medical risk stratification," Knowledge and Information Systems, pp. 1–28, 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, "Predicting the risk of exacerbation in patients with chronic obstructive pulmonary disease using home telehealth measurement data,"Artificial Intelligence in Medicine, vol. 63, no. 1, pp. 51–59, 2015.

[4]. J. M. Wei, 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. 381–391, 2010.

[5]. E. Kontio, A. Airola, T. Pahikkala, H. Lundgren-Laine, K. Junttila, H. Kovenranta, T.Salakoski, and S.Salanter"a,'Predicting patient acuity from electronic paitentrecords".Journal of Biomedical Informativs,vols.51, pp.8

[6]. Q. Nguyen, Hauskrecht, "Learning classification models with soft-label information." Journal of the AmericanMedicalInformaticsAssociation:JAMIA,vol.21,no.3,pp.501–8, 2014.

[7]. 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," IEEE Transactions Knowledge and Data Engineering, vol. 27, no. 1, pp. 130–141, 2015.

[8]. 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. 761–766.

[9]. S. Pan, J. Wu, and X. Zhu, "CogBoost: Boosting for Fast Costsensitive Graph Classification,"IEEETransactionson Knowledge and Data Engineering, vol. 6, no. 1, pp. 1–1, 2015.

[10]. M.Eichelberg,T.Aden,J.Riesmeier,A.Dogac,andG.B.Laleci,"A survey and analysis of Electronic Healthcare Record standards," ACM Computing Surveys, vol. 37, no. 4, pp. 277–315, 2005.

[11]. 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," Preventive Medicine, vol. 67, pp. 28–34, 2014.

[12]. "Health assessment for people aged 75years andolder," http://www.health.gov.au/internet/main/publishing.nsf/ Content/mbs primary care mbs item 75 and older, accessed: 201505-03.

[13]. "Health checks for the over-65s," http://www.nhs.uk/Livewell/ Screening/Pages/Checksover65s.aspx, accessed: 2015-05-03.

[14]. L. Krogsbøll, K. Jørgensen, C. Grønhøj Larsen, and P. Gøtzsche, "General health checks in adults for reducing morbidity and mortality from disease (Review)," Cochrane Database of Systematic Reviews, no. 10, 2012.

[15]. B. Qian, X. Wang, N. Cao, H. Li, and Y.-G. Jiang, "A relative similarity based method for interactive patient risk prediction," Data Mining and Knowledge Discovery, vol. 29, no. 4, pp. 1070–1093, 2015.

[16]. J. Kim and H. Shin, "Breast cancer survivability prediction using labeled,unlabeled,andpseudo-labeledpatientdata,"Journalofthe American Medical Informatics Association : JAMIA, vol. 20, no. 4, pp. 613–618, 2013.