An ensemble framework with l_{21}-norm regularized hypergraph laplacian multi-label learning for clinical data prediction

Peng Cao
College of Computer Science and Engineering, Key Laboratory of Intelligent Computing in Medical Image, Ministry of Education,
Northeastern University
Shenyang, China
caopeng@cse.neu.edu.cn

Shanshan Tang
College of Information Science and Engineering, Northeastern University
Shenyang, China
1700892@stu.neu.edu.cn

Min Huang
College of Information Science and Engineering, Northeastern University
Shenyang, China
mhuang@ise.neu.edu.cn

Jinzhu Yang*
College of Computer Science and Engineering, Northeastern University
Shenyang, China
yangjinzhu@cse.neu.edu.cn

Dazhe Zhao
College of Computer Science and Engineering, Northeastern University
Shenyang, China
zhaodz@neusoft.com

Amine Trabelsi
Department of Computing Science,
University of Alberta
Edmonton, Canada
atrabels@ualberta.ca

Osmar Zaiane
Department of Computing Science
University of Alberta
Edmonton, Canada
zaiane@ualberta.ca

***Abstract*—Previous work has shown that machine learning algorithms lend themselves to clinical decision-making and are a valuable tool for physicians. For clinical data, it is often necessary to assign multiple labels to a patient record by choosing from a large number of potential labels. A key problem in learning from multi-labelled data is how to exploit the information contained in the correlations between labels. The hypergraph-based multi-label learning method learns from data by exploiting the spectral property of the hypergraph that encodes the correlation structure of labels. However, the problem with this method is the difficulty with which interpretations can be made. This is mainly due to its inability to recognize the importance of key features in the original feature space. Moreover, it is hard to comprehensively capture the complex structure of the correlations between labels. To overcome these difficulties and improve interpretability, we propose an l_{21}-norm regularized Graph Laplacian multi-label learning to perform feature selection and label embedding simultaneously. In-depth experimental studies, using the publicly available Medical Information Mart for Intensive Care (MIMIC-III) database, validate the effectiveness of our approach.**

Keywords—Multi-label learning, High dimensionality, Feature selection, Ensemble classification, MIMIC-III

I. INTRODUCTION

With recent advances and the success of machine learning models, many researchers have adopted these models for predictive tasks, which is a major problem in critical health research [1-2]. Clinical medical data consist of multivariate time series of observations involving laboratory tests, physiological values or electrocardiograms [3-4]. Availability of large health care databases, such as Medical Information Mart for Intensive Care (MIMIC-II and III) [5-6], has accelerated research in this area as it provides sufficient data and the ability to train and evaluate machine learning algorithms, detection of physiological decline, and phenotypic classification of a patient [7]. Patient phenotyping is a classification task to determine if a patient has a health problem [8] and is a popular machine learning application in recent years [9-11].

Diseases can often co-occur and many patients may suffer from other diseases related to the main disease. For this reason, we formulate patient phenotyping as a multi-label classification problem. In traditional label learning, each instance is associated to a single label. For multi-label learning task, instances may be associated to more than just one label. Various multi-label learning methods have been proposed to capture the dependency between labels. For clinical data, the difficult problem of classifying multi-tagged data is its high dimensionality. In addition, multi-label data often has irrelevant and redundant features that hinder the performance of multi-label learning.

The Hypergraph-based Multi-label Learning Method (HypergraphMLL) is an alternative solution for simultaneously modeling multi-label data and reducing the dimensionality of the data space by deriving a latent label space [14]. HypergraphMLL captures correlation information between multiple labels, using a small subspace shared by all labels. The purpose of the Laplacian hypergraph multi-label learning method is to capture higher label correlations. However, there remain two issues to be solved:

1) It is usually difficult to make good interpretations and conclusions from the results produced by HypergraphMLL models. The difficulty lies in the fact that many learning methods learn a projection, that is a linear combination (compression or summary) of all the original features. It is essentially a transformation of the input features into a low dimensional space. From a clinical point of view, a disease

diagnostic model should be able to accurately identify biologically significant biomarkers. Relevant biomarkers can help detect the early stages of the disease. Therefore, it is necessary to do manifold learning and feature selection at the same time in order to reduce the negative influence of noisy features.

2) It is difficult to learn complex label correlations directly from data samples when the number of labels increases [15]. It is well known that exploiting label correlations is important for multi-label learning. Existing approaches typically exploit label correlations globally. However, as the number of labels increases, the correlation structure of labels becomes difficult to evaluate directly from data samples.

In order to solve the issues raised above, we reformulate the learning problem and use l_{21}-norm [16] on a projection matrix to achieve sparsity in rows. This leads to relevant feature selection and dimensionality reduction simultaneously. In this respect a hypergraph is designed to account for multi-labelled data correlations. The proposed formulation is obtained by solving a generalized eigenvalue problem. Moreover, we propose to combine RAndom k-labELsets (RAKEL) ensemble with l_{21}-norm regularized Graph Laplacian multi-label learning, to exploit potential higher-order correlations between multiple instances sharing the same label only in the label subset with smaller label size. Combined with local label subset-based RAKEL ensemble [17], the l_{21}-norm regularized HypergraphMLL is able to capture the local instance-label dependencies more effectively.

In summary, the main contributions of this paper are:

- Combining joint feature selection with sparsity and Hypergraph Laplacian multi-label learning into a single framework to select the most informative features when learning a low-dimensional embedding for multi-labeled data;
- Designing an efficient algorithm for the optimization of the proposed non smooth objective function associated to the formulation of the l_{21}HypergraphMLL model;
- Combining local label subset-based RAKEL ensemble and l_{21}-norm regularized HypergraphMLL to capture the local instance-label correlations more efficiently. Each component model builds a hypergraph and locally trains an l_{21}HypergraphMLL classifier based on a small subset of labels, while removing the effect of noisy label correlations. No previous work simultaneously takes into account feature selection and locally complex correlation modeling for multi-label learning;
- Improving traditional Hypergraph Laplacian based multi-label learning and outperforming the state-of-the-art multi-label learning methods on the basis of the publicly available MIMIC III ICU data sets. The results show that our methods are effective in tackling the complex clinical multi-label data with curse of dimensionality.

The rest of this paper is organized as follows. In Section 2, we present the formulation of Hypergraph Laplacian based multi-label learning. We introduce the formulation and optimization procedure of proposed l_{21}HypergraphMLL and RAKEL-l_{21}HypergraphMLL in Section 3. In Section 4, we discuss experiment results and Section 5 concludes the paper.

II. Multi-label learning HyperGraph

The aim of our work is to classify diagnoses of each patient visit (or episode) given multivariate Intensive Care Unit (ICU) time series. We formally define the multi-label classification problem as follows: let $X = \{x_1, \dots, x_n\}$ denotes the space of instances and $Y = \{y_1, \dots, y_l\}$ the class labels, and $T = \{(x_1, Y_1), \dots, (x_n, Y_n)\}$ the multi-label training dataset. Note that $|Y| = l$ and $|T| = n$.

In [14], the hypergraph is used to capture the correlation information among different labels while higher-order correlations are exploited by the HypergraphMLL algorithm. The purpose of hypergraph embedding is to find the optimal low-dimensional vector representation that maintains the original relationship between the data points. The procedure of hypergraph Laplacian multi-label learning involves: (1) hypergraph construction, (2) Laplacian matrix estimation, and (3) low dimensional embedding learning for the transformation matrix.

A. Hypergraph construction

Hypergraph is a generalization of the traditional graph in which an edge can connect arbitrary non-empty subsets of the vertex set. In a hypergraph $G = (V, E)$, V is the vertex set and E is the edge set, where each $e \in E$ is a subset of V. Given a multi-label dataset, the samples with their labels are represented as a single hypergraph $G = (V, E)$. Some concepts are introduced as follow:

$d(v)$ is the degree of a vertex as defined as:

$$d(v) = \sum_{v \in e, e \in E} w(e) \tag{1}$$

where $\delta(e) = |e|$ and $w(e)$ is the weight associated with the hyperedge e.

The vertex-edge incidence matrix $J \in \mathbb{R}^{|V| \times |E|}$ is defined as

$$J(v, e) = \begin{cases} 1 & \text{if } v \in e \\ 0 & otherwise. \end{cases} \tag{2}$$

B. Hypergraph Laplacian estimation

The Laplacian matrix from a traditional graph is widely used for learning from graphs. It is a commonly used technique for capturing the relationship among nodes in the hypergraph and has been used in spectral clustering. The normalized hypergraph Laplacian can be obtained as follow:

$$L_z = I - S_z \tag{3}$$

$$S_z = I - L_z = D_v^{-\frac{1}{2}} J W_H D_e^{-1} D_v^{-\frac{1}{2}} \tag{4}$$

where D_e, D_v and W_H are the diagonal matrix forms for $\delta(e)$, $d(v)$ and $w(e)$, respectively.

Laplacian matrix plays an important role in learning. In this paper, we use Zhou's normalized Laplacian for calculating the hypergraph Laplacian.

C. Low-dimensional embedding learning

Based on the hypergraph and Laplacian matrix, the goal of the HypergraphMLL algorithm is to learn a low-dimensional feature transformation W, which is also called the projection matrix.

The formulation of learning a low-dimensional embedding through a linear transformation W is:

$$\min_{W} \quad trace(W^T X^T LXW)$$
$$\text{subject to} \quad W^T X^T LXW = I_k, \tag{5}$$

The aim of the formulation is to encourage the instances sharing many common labels to be close to each other in the transformed low dimensional space.

To improve the efficiency of the formulation, an approximate hypergraph low-dimensional embedding learning formulation is designed as follow:

$$W = \text{argmin} \quad L(W) = \|XW - QU\|_F^2 \tag{6}$$

where $U = \text{svd}(R), Q, R = \text{qr}(H), S_z = HH^T$

The optimization procedure of the approximate algorithm is shown in Algorithm 1.

Algorithm 1: The optimization of approximate HypergraphMLL algorithm

Input: Training data $\{ X \in \mathbb{R}^{n\times k}, Y \in \mathbb{R}^{n\times k}\}$,
regularization parameter λ;
Output: mapping matrix W .
1: Construct D_v, D_e, W_H, J based on Y;
2: Similarity matrix $S_z \leftarrow D_v^{-\frac{1}{2}} JW_H D_e^{-1} D_v^{-\frac{1}{2}}$;
3: $H \leftarrow D_v^{-\frac{1}{2}} JW_H^{\frac{1}{2}} D_e^{-\frac{1}{2}}$;
4: $Q, R \leftarrow \text{qr}(H)$;
5: $U \leftarrow \text{svd}(R)$;
6: $W \leftarrow \min L_2(W,\lambda) = \|XW - QU\|_F^2 + \lambda\|W\|_F^2$;

III. An Ensemble framework with L21-norm regulaized Hypergraph Laplacian multi-label learning

A. *l_{21}-norm regularized Graph Laplacian multi-label learning, l_{21}HypergraphMLL*

In this section, we introduce the proposed model, l_{21}HypergraphMLL, which is extended to joint sparse-based feature selection and lower dimensional embedding learning for modeling the correlation of multiple labels.

For patients, not all features of the original input space are useful in phenotyping. Some are unrelated to the tasks at hand. It is generally not known which is the best descriptor of discriminant features. Although the multi-label classification has attracted a lot of attention in recent years, very little research effort has been devoted to multi-label feature selection. Sparsity-based feature selection approaches provide a solution to this problem by assessing the strength of potential correlations between different features. Among these approaches, the l_{21}-norm regularization has shown to be effective for sparse feature selection. The objective function of the l_{21}HypergraphMLL is specified in the following optimization formulation:

$$\min L_{21}(W_i,\lambda) = \|XW - QU\|_F^2 + \lambda\|W_i\|_{21} \tag{7}$$

where $\lambda > 0$ is the regulation parameter.

The l_{21}-norm regularization term is imposed on W to ensure that W is sparse in rows. Each row of W measures the importance of i-th feature in the original space. The l_{21}-norm regularization automatically selects the most relevant features. The optimization Eq.(7) is presented in Algorithm 2.

Algorithm 2: The optimization of l_{21}HypergraphMLL

Input: Training data $\{ X \in \mathbb{R}^{n\times k}, Y \in \mathbb{R}^{n\times k}\}$,
regularization parameter λ;
Output: mapping matrix W .
1: Construct D_v, D_e, W_H, J based on Y_i;
2: Similarity matrix $S_z \leftarrow D_v^{-\frac{1}{2}} JW_H D_e^{-1} D_v^{-\frac{1}{2}}$;
3: $H \leftarrow D_v^{-\frac{1}{2}} JW_H^{\frac{1}{2}} D_e^{-\frac{1}{2}}$;
4: $Q, R \leftarrow \text{qr}(H)$;
5: $U \leftarrow \text{svd}(R)$;
6: $W \leftarrow \min L_{21}(W,\lambda) = \|XW - QU\|_F^2 + \lambda\|W\|_{21}$;

B. *Ensemble learning classification combined with l_{21}HypergraphMLL*

Laplacian multi-label learning captures the correlation among different labels globally. However, label correlations are naturally local [15]. RAndom k-labELsets (RAKEL) algorithm, an effective ensemble method for solving multi-label classification, is proposed in [17]. RAKEL randomly breaks the initial set of labels into a number of small-sized label subsets from the original set of labels. These subsets are referred to as k-labelsets. In each label subset, the proposed l_{21}HypergraphMLL is used to train a corresponding multi-label learning model. Only the correlation of labels with hyperedge for each label in the label subset can be captured. Finally, the final prediction of RAKEL is made by voting of the l_{21}HypergraphMLL models in the ensemble. The pseudocode of the ensemble learning classification combined with l_{21}HypergraphMLL is illustrated in Algorithm 3.

Algorithm 3 RAKEL combined with l_{21}HypergraphMLL (RAKEL- l_{21}HypergraphMLL)

Input: Training data $\{X\in \mathbb{R}^{n\times d}, Y \in \mathbb{R}^{n\times l}\}$,
size of label subset k, number of label subsets m,
MHSL regularization parameter λ, BR base classifier parameter C, new instance feature $\vec{X}$;
Output: classification result of new instance $\vec{Y}$.
1: $\{sub_1, \ldots, sub_m\} \leftarrow$ *random_k_label* (*l,k,m*);
2: **for** i =1 **to** m **do**
3: Construct a hypergraph
4: Calculate a Laplacian matrix according to Eq. (3)
5: W_i←compute mapping matrix with X_i,Y_i,λ according to **Algorithm 2**;
6: $X_i \leftarrow X_i W_i$ transform training data with W_i;
7: $\vec{X_i} \leftarrow \vec{X}\, W_i$ transform new instance with W_i;
8: $H_i \leftarrow$ Train a base classifier H_i based on X_i, Y_i, C;
9: $P_i^{1\times k} \leftarrow H_i(\vec{X_i})$ label vector based on sub_i;
10: $\vec{Y} \leftarrow$ multi-label voting($\{P\}$, $\{sub\}$);

In RAKEL, the diversity of classifiers is achieved by randomly selecting label subsets. The classification of a new instance is achieved by thresholding the average of the binary decisions of each model for each label. The pseudocode of the ensemble learning classification combined with l_{21}HypergraphMLL is illustrated in Algorithm 3.

IV. Experiments

A. Data

We evaluate our model on the publicly available MIMIC-III ICU database. MIMIC- III includes all patients admitted to an ICU at the Beth Israel Deaconess Medical Center from 2001 to 2012. Table 1 shows some useful statistics of MIMIC dataset. In total, we obtain 17×7×6 = 714 features per time series.

Table 1. Statistics for six benchmark datasets used in our experiments.

Instances			Statistics			
Training	**Val**	**Test**	***Feat.***	***Labels***	***Card.***	***Dens.***
27180	6371	6281	714	25	4.34	0.174

Table 2. The information of data and phenotype in MIMIC dataset

phenotype	Prevalence			No.	%
	Train	Val	Test		
Essential hypertension	0.453	0.410	0.423	17573	44.1
Coronary atherosclerosis and related	0.347	0.317	0.331	13540	34.0
Cardiac dysrhythmias	0.346	0.316	0.323	13458	33.8
Disorders of lipid metabolism	0.314	0.286	0.289	12162	30.5
Fluid and electrolyte disorders	0.288	0.276	0.265	11254	28.3
Congestive heart failure; non hypertensive	0.289	0.264	0.268	11220	28.2
Acute and unspecified renal failure	0.232	0.207	0.212	8964	22.5
Complications of surgical/medical care	0.223	0.201	0.213	8695	21.8
Diabetes mellitus without complication	0.209	0.186	0.192	8074	20.3
Respiratory failure; insufficiency; arrest	0.194	0.184	0.177	7566	19.0
Septicemia (except in labor)	0.154	0.146	0.139	5975	15.0
Pneumonia	0.151	0.135	0.135	5815	14.6
Chronic kidney disease	0.145	0.132	0.132	5607	14.1
Hypertension with complications	0.143	0.131	0.130	5547	13.9
Chronic obstructive pulmonary disease	0.142	0.128	0.126	5455	13.7
Acute myocardial infarction	0.110	0.103	0.108	4337	10.9
Diabetes mellitus with complications	0.103	0.095	0.094	3988	10.0
Other liver diseases	0.095	0.091	0.089	3723	9.3
Pleurisy; pneumothorax; pulmonary collapse	0.092	0.090	0.091	3658	9.2
Shock	0.085	0.075	0.082	3291	8.3
Acute cerebrovascular disease	0.080	0.075	0.066	3079	7.7
Gastrointestinal hemorrhage	0.077	0.075	0.079	3067	7.7
Conduction disorders	0.078	0.070	0.071	3011	7.6
Other lower respiratory disease	0.055	0.049	0.057	2168	5.4
Other upper respiratory disease	0.044	0.037	0.043	1702	4.3

Table 1 shows some useful statistics of MIMIC datasets, such as the number of instances in the training and test sets, the number of features (Feat.), the number of labels, label cardinality (Card.) and label density (Dens.). The 25 care conditions (labels) are described in Table 2. The feature description is shown in Table 3. In total, we obtain 17×7×6 = 714 features per time series.

B. Setting

We evaluate the performance of the competing approaches on the basis of five commonly used multi-label assessing criteria: Hamming loss, F1-micro, F1-macro, F1-weighted and Jaccard score (see Table 4).

We use a validation dataset to tune the hyperparameters. The regularization parameter of λ (0.0000001, 0.000001, …, 0.0001), the labels subset size parameter (3,5,7,10) and the ensemble size (13,26,38,51) are optimized. The reported results were the best results of each method with the optimal parameters shown in Table 5.

Table 3. The feature set used in our experiment.

Feature set
Capillary refill rate
Diastolic blood pressure
Fraction inspired oxygen
Glascow coma scale eye opening
Glascow coma scale motor response
Glascow coma scale total
Glascow coma scale verbal response
Glucose
Heart Rate
Height
Mean blood pressure
Oxygen saturation
Respiratory rate
Systolic blood pressure
Temperature
Weight
pH

Table 5. The tuned value of hyperparameters

Hyperparameters	Optimal value
λ	0.000001
k	3
m	38

C. Experiment I

In Experiment I, we assess the impact of the l_{21}-norm based feature selection and the performance of RAKEL ensemble. A comparison is carried out between our proposed methods (ensemble version RAKEL- l_{21}HypergraphMLL and single version l_{21}HypergraphMLL), the intermediate method RAKEL-HypergraphMLL, and the basic methods, such as HypergraphMLL and basic binary relevance (BR) method. The binary relevance (BR) [18] splits the multi-label learning problem into several binary classification problems using the one-against-all strategy. From the results in Table 5, we may make the following observations:

1) Except for the Hamming loss measure, RAKEL-l_{21}HypergraphMLL achieved high predictive

performance compared to both baseline methods: BR and single HypergraphMLL methods.

2) Compared to the l_{21}HypergraphMLL simple model, the RAKEL-l_{21}HypergraphMLL offers a better performance for all measurements, except for the Hamming loss. This result illustrates the contribution of the ensemble component to improving the performance of the single multi-label learning by modelling the local structure of label correlations.
3) It is surprising that l_{21}HypergraphMLL does not improve the performance of HypergraphMLL. However, when they are both associated with RAKEL (i.e. RAKEL-l_{21}HypergraphMLL) improvement is obtained in terms of F1-micro score and Jaccard. This demonstrates that exploiting label correlations using a subset of random tags improves prediction performance in terms of F1-micro score and Jaccard.

Besides focusing on the classification performance, we are interested in assessing the advantages of feature selection procedure with l_{21}-norm. Table 6 shows the number of features selected by l_{21}HypergraphMLL and RAKEL-l_{21} HypergraphMLL models. Both models are able to remove some irrelevant and redundant features, while RAKEL-l_{21}HypergraphMLL selects fewer features than single l_{21}HypergraphMLL. This can be attributed to the fact that RAKEL-l_{21}HypergraphMLL can identify specific features that are important for learning label correlation only on a smaller label subset. It appears that the combined RAKEL-l_{21}HypergraphMLL models, with local learning of label correlations, contribute to improve the performance of feature selection and classification. The fewer selected features improve the efficiency when predicting new instances.

Table 6. The number of feature selected by both l_{21}HypergraphMLL and RAKEL-l_{21}HypergraphMLL (*k=3, m=13, λ=1e-6*).

RAKEL-l_{21}HypergraphMLL		l_{21}HypergraphMLL
label subset	no. of feature selected	no. of feature selected
1	504	
2	557	
3	544	
4	500	
5	494	
6	532	
7	498	663
8	496	
9	502	
10	514	
11	496	
12	527	
13	512	

D. Experiment II

We investigate five modern models for the task of multi-label classification on the MIMIC datasets. The comparative results are shown in Table 7. The results confirm the advantages of our approach for multi-label data learning. More specifically, the experimental results show that the proposed RAKEL-l_{21}HypergraphMLL outperforms the state-of-the-art methods in most cases. These results reveal several interesting points:

1) Both label powerset (LP) [19] and l_{21}HypergraphMLL have the capacity of capturing high-order correlations among labels. This can help exploiting the relationships of multiple labels more effectively and intrinsically. Figures in Table 7 show that adding RAKEL can improve l_{21}HypergraphMLL. Moreover RAKEL-l_{21}HypergraphMLL outperforms RAKEL-LP in most cases. The results justify our claim that modeling label correlation with hypergraph leads to improved performance.
2) The classifier chain (CC) method has been shown to improve the classification accuracy of the BR method on a number of regular datasets [20]. However, it does not outperform BR without considering the label correlation. The CC effectiveness dramatically drops when the complexity of the dataset increases. The MIMIC dataset is complex with respect to the number of labels, cardinality and label dependency. This is the reason of the poor performance obtained by CC.
3) The results in [9] show that the traditional multi-label learning including CC, LP, and ML-kNN are less performant than BR on the MIMIC dataset. The results we obtain for CC, ML-kNN [21] and RAKEL-LP in Table 6 are in accordance with the observations in [9]. Although we find similar results, we don't agree with the reasons provided in [9]. For the MIMIC dataset with high dimensional features and complex label dependency, it is critical to perform feature selection and label modeling locally during the multi-label learning.

E. Experiment III

To investigate the effect of feature selection in our l_{21}-norm regularized l_{21}HypergraphMLL and RAKEL-l_{21}HypergraphMLL, we vary the value of λ to control the effect of l_{21}-norm. Figure 1 and Figure 2 show the number of selected features and the classification performance according to the value of λ.

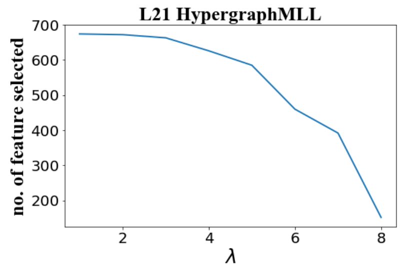

Fig 1. The number of feature selected by l_{21}HypergraphMLL according to λ values. The x-axis denotes the values of λ: [1*e*-7, 5*e*-7, 1*e*-6, 5*e*-6, 1*e*-5, 5*e*-5, 0.0001, 0.0003].

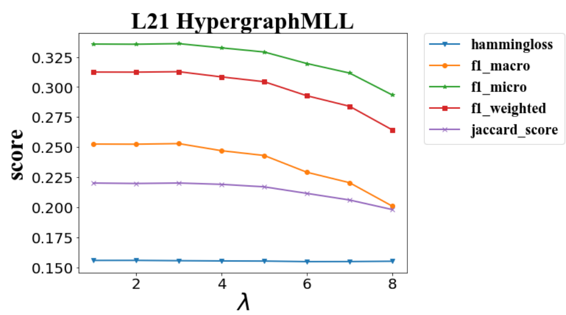

Fig 2. Metrics on l_{21}HypergraphMLL according to λ values. The x-axis denotes the values of λ: [1*e*-7, 5*e*-7, 1*e*-6, 5*e*-6, 1*e*-5, 5*e*-5, 0.0001, 0.0003].

Table 4. The description of the metrics used in our experiment.

Measure	Formulation	Description
Hamming loss	$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{l}\sum_{j=1}^{l}[\![h_{ij} \neq y_{ij}]\!]$	The fraction of misclassified labels
F1-micro	$\frac{2\sum_{j=1}^{l}\sum_{i=1}^{n} y_{ij}h_{ij}}{\sum_{j=1}^{l}\sum_{i=1}^{n} y_{ij} + \sum_{j=1}^{l}\sum_{i=1}^{n} h_{ij}}$	F-measure averaging on the prediction matrix
F1-macro	$\frac{1}{l}\sum_{j=1}^{l}\frac{2\sum_{i=1}^{n} y_{ij}h_{ij}}{\sum_{i=1}^{n} y_{ij} + \sum_{i=1}^{n} h_{ij}}$	F-measure averaging on each label
F1-weighted	$\frac{1}{l}\sum_{j=1}^{l} weight_j \frac{2\sum_{i=1}^{n} y_{ij}h_{ij}}{\sum_{i=1}^{n} y_{ij} + \sum_{i=1}^{n} h_{ij}}$	F-measure averaging on each label by their weighted average
Jaccard score	$\frac{\lvert y_{pred} \cap y_{true}\rvert}{\lvert y_{pred} \cup y_{true}\rvert}$	the size of the intersection divided by the size of the union of two label sets

Table 6. The performance of the four HypergraphMLL methods (The rank is also shown)

Methods	Hamming loss	F1-macro	F1-micro	Weighted F1	Jaccard score
BR	0.2319(3)	0.3512(3)	0.4193(3)	0.4171(3)	0.2512(3)
HypergraphMLL	0.1631(2)	0.2839(4)	0.4015(4)	0.3638(4)	0.2501(4)
RAKEL-HypergraphMLL	0.2751(5)	**0.3722**(1)	0.4483(2)	**0.4330**(1)	0.2821(2)
l_{21}HypergraphMLL	**0.1560**(1)	0.2528(5)	0.3358(5)	0.3126(5)	0.2203(5)
RAKEL-l_{21}HypergraphMLL	0.2404(4)	0.3659(2)	**0.4551**(1)	0.4284(2)	**0.2871**(1)

Table 7. The performance of the our RAKEL-l_{21}HypergraphMLL compared with four state-of-the-art MLL methods (The rank is also shown)

Methods	Hamming loss	F1-macro	F1-micro	Weighted F1	Jacc
BR	0.2319(5)	0.3512(2)	0.4193(2)	0.4171(2)	0.2512(4)
CC	0.2082(4)	0.3270(3)	0.4188(3)	0.4020(3)	0.2518(3)
LP	**0.1763**(1)	0.1571(6)	0.2668(6)	0.2114(6)	0.1772(6)
RAKEL-LP	0.1854(3)	0.2965(4)	0.4118(4)	0.3771(4)	0.2574(2)
ML-kNN	**0.1763**(1)	0.2390(5)	0.3359(5)	0.3083(5)	0.2109(5)
RAKEL - l_{21}HypergraphMLL	0.2404(6)	**0.3659**(1)	**0.4551**(1)	**0.4284**(1)	**0.2871**(1)

V. Conclusion

We combine l_{21}-norm regularized hypergraph Laplacian multi-label learning and RAKEL ensemble algorithms to perform multi-label classification on medical records of ill patients. The unified framework can handle high dimensionality and the local complex correlation structure of labels, simultaneously. Experimental results indicate that the classification performance of RAKEL-l_{21}HypergraphMLL compares favorably with that of other state-of-the-art approaches over multiple evaluation measures. These promising results support our contention that modeling label correlations with hypergraph leads to improved performance. In a future work we will extend our approach to the dynamic modelling of patient's health state from its longitudinal electronic medical record. Extensions include nonlinear models with kernel mapping or deep learning.

References

[1] Choi, E., Bahadori, M.T., Schuetz, A., Stewart, W.F. and Sun, J., 2016, December. Doctor ai: Predicting clinical events via recurrent neural networks. In Machine Learning for Healthcare Conference (pp. 301-318).

[2] Luo, Y., Xin, Y., Joshi, R., Celi, L. and Szolovits, P., 2016, February. Predicting ICU mortality risk by grouping temporal trends from a multivariate panel of physiologic measurements. In Thirtieth AAAI Conference on Artificial Intelligence.

[3] Zhang, W., Liu, F., Luo, L. and Zhang, J., 2015. Predicting drug side effects by multi-label learning and ensemble learning. BMC bioinformatics, 16(1), p.365.

[4] Weng, H., Liu, Z., Maxwell, A., Li, X., Zhang, C., Peng, E., Li, G. and Ou, A., 2018, December. Multi-Label Symptom Analysis and Modeling of TCM Diagnosis of Hypertension. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1922-1929). IEEE.

[5] Purushotham, S., Meng, C., Che, Z. and Liu, Y., 2017. Benchmark of deep learning models on large healthcare MIMIC datasets. arXiv preprint arXiv:1710.08531.

[6] Johnson, A.E., Pollard, T.J., Shen, L., Li-wei, H.L., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L.A. and Mark, R.G., 2016. MIMIC-III, a freely accessible critical care database. Scientific data, 3, p.160035.

[7] Harutyunyan, H., Khachatrian, H., Kale, D.C., Steeg, G.V. and Galstyan, A., 2017. Multitask learning and benchmarking with clinical time series data. arXiv preprint arXiv:1703.07771.

[8] Oellrich, A., Collier, N., Groza, T., Rebholz-Schuhmann, D., Shah, N., Bodenreider, O., Boland, M.R., Georgiev, I., Liu, H., Livingston, K. and Luna, A., 2015. The digital revolution in phenotyping. Briefings in bioinformatics, 17(5), pp.819-830.

[9] Zufferey, D., Hofer, T., Hennebert, J., Schumacher, M., Ingold, R. and Bromuri, S., 2015. Performance comparison of multi-label learning algorithms on clinical data for chronic diseases. Computers in biology and medicine, 65, pp.34-43.

[10] Bromuri, S., Zufferey, D., Hennebert, J. and Schumacher, M., 2014. Multi-label classification of chronically ill patients with bag of words and supervised dimensionality reduction algorithms. Journal of biomedical informatics, 51, pp.165-175.

[11] Peng, Y., Tang, C., Chen, G., Xie, J. and Wang, C., 2017, November. Multi-label learning by exploiting label correlations for TCM diagnosing Parkinson's disease. In 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 590-594). IEEE.

[12] Zhang, M.L. and Zhou, Z.H., 2013. A review on multi-label learning algorithms. IEEE transactions on knowledge and data engineering, 26(8), pp.1819-1837.

[13] Silva, I., Moody, G., Scott, D.J., Celi, L.A. and Mark, R.G., 2012, September. Predicting in-hospital mortality of icu patients: The physionet/computing in cardiology challenge 2012. In 2012 Computing in Cardiology (pp. 245-248). IEEE.

[14] Sun, L., Ji, S. and Ye, J., 2008, August. Hypergraph spectral learning for multi-label classification. In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 668-676). ACM.

[15] Huang, S.J. and Zhou, Z.H., 2012, July. Multi-label learning by exploiting label correlations locally. In Twenty-sixth AAAI conference on artificial intelligence.

[16] Nie, F., Huang, H., Cai, X. and Ding, C.H., 2010. Efficient and robust feature selection via joint ℓ2, 1-norms minimization. In Advances in neural information processing systems (pp. 1813-1821).

[17] Tsoumakas, G. and Vlahavas, I., 2007, September. Random k-labelsets: An ensemble method for multilabel classification. In European conference on machine learning (pp. 406-417). Springer, Berlin, Heidelberg.

[18] O. Luaces, J. Dez, J. Barranquero, J. del Coz, A. Bahamonde, Binary relevance efficacy for multilabel classification, Prog. Artif. Intell. 1 (4) (2012) 303–313.

[19] Boutell, M.R., Luo, J., Shen, X., Brown, C.M.: Learning multi-label scene classification. Pattern Recognition 37(9), 1757- 1771 (2004)

[20] Zhang, M.L. and Zhou, Z.H., 2007. ML-KNN: A lazy learning approach to multi-label learning. Pattern recognition, 40(7), pp.2038-2048.

[21] Read, J., Pfahringer, B., Holmes, G. and Frank, E., 2009, September. Classifier chains for multi-label classification. In Joint European Conference on Machine Learning and Knowledge Discovery in Databases (pp. 254-269). Springer, Berlin, Heidelberg.