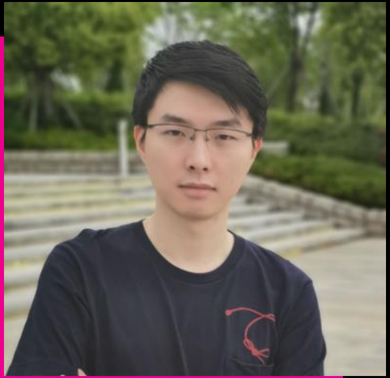


Hierarchical Label Distribution Learning for Disease Prediction

Reporter: Chengkai Wu

Postdoc.

Zhejiang Laboratory

CONTENTS

PART 01

Introduction

The prediction of disease can facilitate early intervention, comprehensive diagnosis and treatment, thereby benefiting healthcare and reducing medical costs.

Introduction

Disease Prediction

- Early intervention
- Comprehensive diagnosis and treatment
- Benefit healthcare
- Reduce medical costs

Electronic Health Records

- The widespread of EHR systems.
- Longitudinal experience of both patients and doctors.

Machine Learning

- Anticipates the information and develop predictive models.
- Single label or multi-label classification.

Introduction

While logical labels can indicate the risk of
diseases, they cannot determine which ones require more attention or which ones should
be prioritized for treatment.

Label Distribution

Since patients may suffer from multiple diseases and the pathological changes of
one organ or system may influence related ones.

Medical knowledge

Prediction becomes more challenging when considering more diseases, due to the class imbalance and the
data sparsity.

Hierarchical Label Distribution Learning (HLDL)

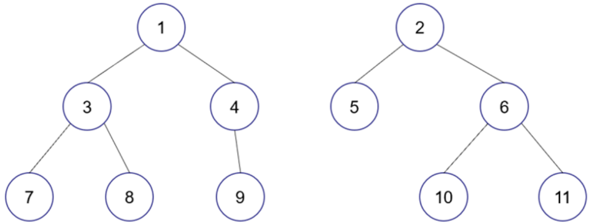
PART 02

Methods

Hierarchical Label Distribution Learning (HLDL) method includes a hierarchical neural network that combines global and local predictions.

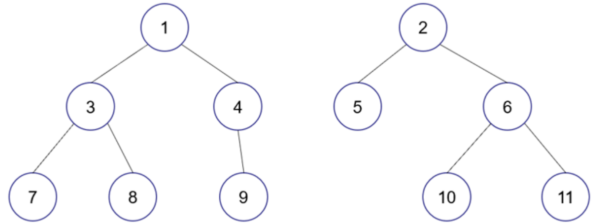
Hierarchical Label Enhancement

- Medical knowledge such as ICD-10 represents the hierarchy of medical concepts in the form of a parent-child relationship.
- $\mathcal{Y}^1 = \{y_1, y_2, \cdots, y_{m_1}\}$
- $\mathcal{Y}^2 = \{y_{m_1+1}, y_{m_1+2}, \cdots, y_{m_1+m_2}\}$
- $\mathcal{Y} = \{\mathcal{Y}^1, \mathcal{Y}^2, \cdots, \mathcal{Y}^K\}\{y_1, \cdots, y_{m_1}, y_{m_1+1}, \cdots, y_C\}$
- $C = \sum_{k=1}^{K} m_t$



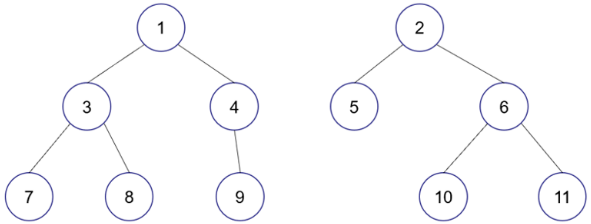
Hierarchical Label Enhancement

- Label distribution assigns a real number $d_x^{y_c} \in [0,1]$ to each label, representing the risk of disease y_c for patient x:
  - $\boldsymbol{d}_x^{\boldsymbol{y}} = \{d_x^{y_1}, d_x^{y_2}, \cdots, d_x^{y_C}\} \in \mathbb{R}^C$
- Quantify the diagnosis with the priority hierarchically:
  - $d_x^{y_i(k)} = \frac{r_i^{(k)}}{\sum_{j=1}^{m_k} r_j^{(k)}}$



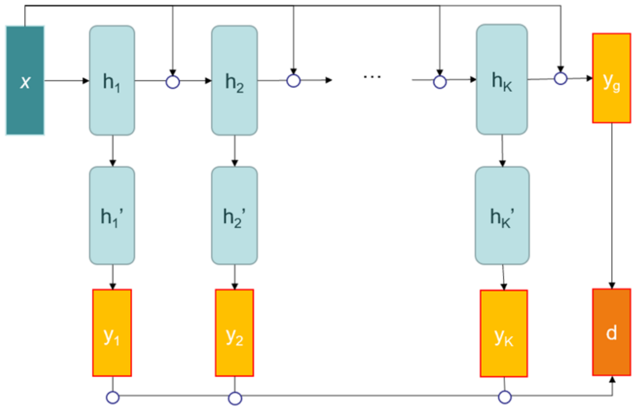
Hierarchical Label Enhancement

- Diagnosis: $\{y_7, y_8, y_{10}\}$
- $r_7 = 3$, $r_8 = 2$, and $r_{10} = 1$
- $d_x^{y_7} = 0.5$, $d_x^{y_8} = 0.3$, and $d_x^{y_{10}} = 0.2$.
- Second layer: $\{y_3, y_6\}$
- $d_x^{y_3} = 0.67$ and $d_x^{y_6} = 0.33$
- First layer: $\{y_1, y_2\}$
- $d_x^{y_1} = 0.67$ and $d_x^{y_2} = 0.33$
- Label distribution: $\boldsymbol{d}_x^{\boldsymbol{y}} = \{d_x^{y_1}, d_x^{y_2}, \cdots, d_x^{y_{11}}\} = \{0.67,0.33,0.67,0,0,0.33,0.5,0.3,0,0.2,0\}$



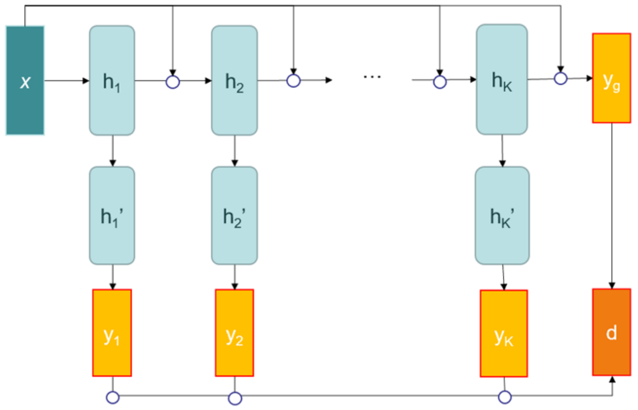
Hierarchical Label Distribution Learning

- Feature space: $\mathcal{X} = \mathbb{R}^q$
- Label distribution space: $\mathcal{D} = \mathbb{R}^C$
- Training set: $\mathcal{S} = \{(\boldsymbol{x}_i, \boldsymbol{d}_i) | i = 1,2,\cdots N\}$
- *Hierarchical Label Distribution Learning* (HLDL):
  - $f: \mathcal{X} \rightarrow \mathcal{D}$



Hierarchical Label Distribution Learning

- Global prediction network:
  - $\boldsymbol{h}_1 = \sigma(\boldsymbol{w}_1\boldsymbol{x} + \boldsymbol{b}_1)$
  - $\boldsymbol{h}_2 = \sigma(\boldsymbol{w}_2[\boldsymbol{h}_1 \oplus \boldsymbol{x}] + \boldsymbol{b}_2)$
  - $\boldsymbol{h}_K = \sigma(\boldsymbol{w}_K[\boldsymbol{h}_{K-1} \oplus \boldsymbol{x}] + \boldsymbol{b}_K)$
  - $\boldsymbol{y}_g = sigmoid(\boldsymbol{w}_g\boldsymbol{h}_k + \boldsymbol{b}_g)$
- Local prediction network:
  - $\boldsymbol{h}'_k = \sigma(\boldsymbol{w}'_k\boldsymbol{h}_k + \boldsymbol{b}'_k)$
  - $\boldsymbol{y}_k = softmax(\boldsymbol{w}_{lk}\boldsymbol{h}'_k + \boldsymbol{b}_{lk})$

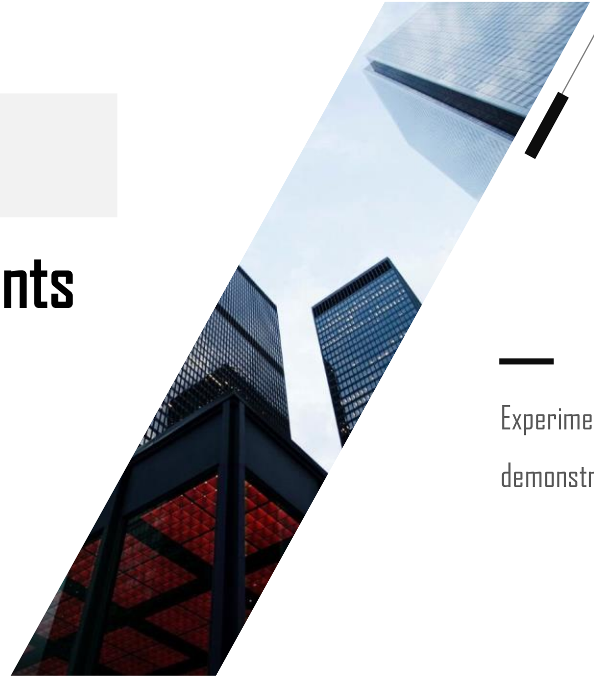


Hierarchical Label Distribution Learning

- Final prediction:
  - $\widehat{\boldsymbol{d}} = \rho[\boldsymbol{y}_1; \boldsymbol{y}_2; \cdots; \boldsymbol{y}_K] + (1-\rho)\boldsymbol{y_g}$
- Loss function : $L = L_g + \tau L_p$
- Global loss function:
  - $L_g = L_{kl} + L_{ls} = \sum_{i=1}^{n} \boldsymbol{d}_i ln\frac{\boldsymbol{d}_i}{\boldsymbol{g}_i} + \sum_{i=1}^{n} \|\boldsymbol{g}_i - \boldsymbol{u}_i\|^2$
  - $\boldsymbol{u}_i$: initial logical labels
- Local loss function:
  - $L_p = L_{kl} + L_{hv} = \sum_{i=1}^{n} \boldsymbol{d}_i ln\frac{\boldsymbol{d}_i}{\boldsymbol{p}_i} + \sum_{i=1}^{n} \boldsymbol{A}(\boldsymbol{1} - \boldsymbol{p}_i^{\top})\boldsymbol{p}_i$
  - $\boldsymbol{A} \in \{0,1\}^{c \times c}$: the relationship matrix

PART 03

Experiments



Experiments conducted on two real-world datasets demonstrate the better performance of the proposed method.

Datasets and Experimental Settings

Dataset	patients	# of visits	# for pretrain	# for train and test
MIMIC-III	46517	58929	39018	7499
MIMIC-IV	190173	453905	83660	106513

ICD	First layer	Second layer	Third layer
ICD-9	19	158	1233
ICD-10	22	261	2269

Embedding	Hidden layer	Activation	Optimization	ρ	τ
64	[256, 516, 1024, 2048]	ReLU	Adam	0.5	2

Experimental Settings

Baselines	Introductions
Doctor AI	It is a generic predictive model for disease prediction.
RETAIN	It is a two-level neural attention model for interpretable.
HMCN	It is a classical hierarchical multi-label classification method.
SA-BFGS	It is a widely used label distribution learning method.
'+LDL'	Trained with label distribution.

$$top - k\ recall = \frac{\#\ of\ true\ positives\ in\ the\ top\ k\ predictions}{\#\ of\ true\ positives\ in\ the\ top\ k\ true\ labels}$$

Experimental Results

Table 1. The performance of compared methods on disease prediction task.

Methods	MIMIC-III			MIMIC-IV		
	top@10	top@30	top@50	top@10	top@20	top@30
Doctor AI	60.16	53.88	53.35	47.95	44.53	46.00
Doctor AI+LDL	63.21	56.56	55.72	49.18	46.06	47.26
RETAIN	61.25	53.62	53.02	48.89	44.54	45.97
RETAIN+LDL	63.45	57.35	56.44	48.64	45.23	46.67
HMCN	65.44	57.67	56.58	55.37	50.40	51.36
HMCN+LDL	68.77	61.54	60.31	58.02	53.83	54.86
SA-BFGS	58.89	52.74	52.72	40.90	36.81	36.16
HLDL	69.25	60.34	59.19	59.76	54.79	55.56
HLDL+LDL	**70.73**	**62.55**	**61.35**	**61.17**	**55.84**	**56.19**

PART 04

Conclusion

The proposed HLDL method, which utilizes a hierarchical neural network ensemble to perform both global and local predictions, as well as incorporating medical ontologies and disease relations to improve model accuracy.

Conclusion

- The proposed HLDL method, which utilizes a hierarchical neural network ensemble to perform both global and local predictions, as well as incorporating medical ontologies and disease relations to improve model accuracy.
- The use of label distribution allows for more details of the risk of diseases for a given patient, leading to more accurate predictions.
- The experimental results on two real-world datasets demonstrate the superiority of HLDL over other methods and highlight the importance of label distribution in diagnosis prediction.
- HLDL has the potential to be a valuable tool for healthcare professionals in making more informed decisions regarding patient care.