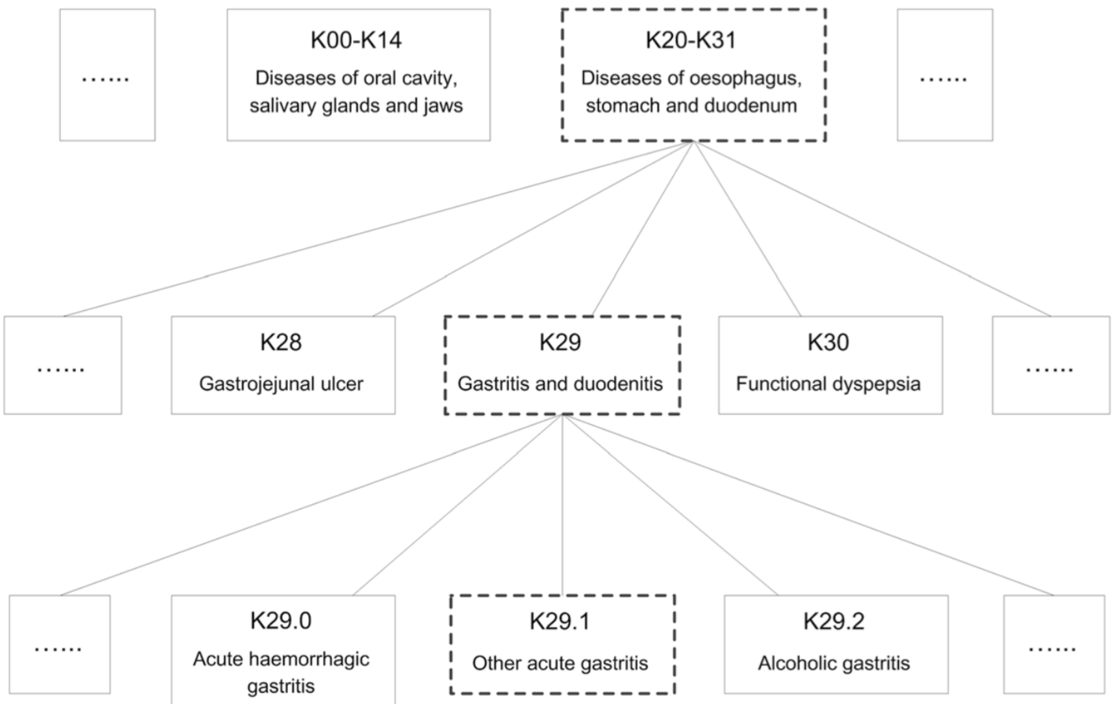
ICD Coding

Zhenghui Wang

Apex Data & Knowledge Management Lab
Shanghai Jiao Tong University

ICD

- ICD: The **I**nternational Statistical **C**lassification of **D**iseases and Related Health Problems
- Hierarchical architecture

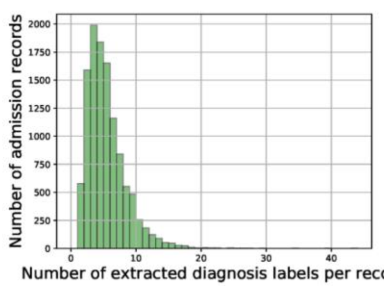


APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

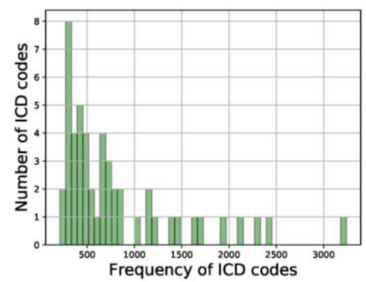
Data

- MIMIC III

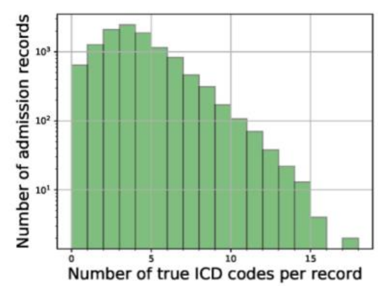
HADMID	189797
Original Texts of Discharge Summary	... DISCHARGE DIAGNOSIS: 1. Prematurity at 35 4/7 weeks gestation 2. Twin number two of twin gestation 3. Respiratory distress secondary to transient tachypnea of the newborn 4. Suspicion for sepsis ruled out ...
Extracted Diagnosis Descriptions	1. Prematurity at 35 4/7 weeks gestation 2. Twin number two of twin gestation 3. Respiratory distress secondary to transient tachypnea of the newborn 4. Suspicion for sepsis ruled out
Assigned ICD Diagnostic Codes	‘V3100’, ‘76518’, ‘7756’, ‘7706’, ‘V290’, ‘V053’



(a) Distribution of diagnosis description count



(b) Distribution of ICD code frequency

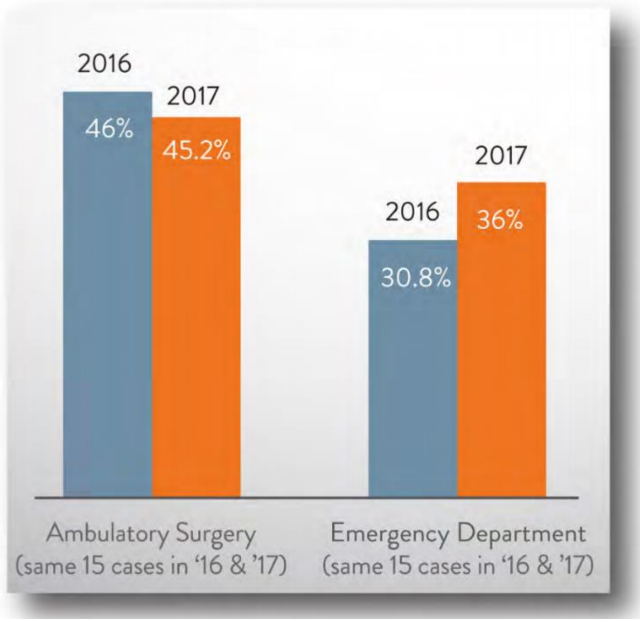


(c) Distribution of ICD code count

数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

ICD-10 Coding Contest

Average Accuracy Scores



"The code entry method was very easy and user-friendly."

"This is really good for students. And to further their education."

Actual Contestant Feedback

[https://www.centrallearning.com/codingcontest/]

Solutions

- String to string comparison

绒癌史 → Z85.406, 绒毛膜癌个人史

- Multi-label text classification

 →

Q21.100,房间隔缺损

Q21.203,部分性房室隔缺损

I37.000,肺动脉瓣狭窄

Str2Str comparison method

- Longest common subsequence
- Semantic similarity with HowNet

Longest common subsequence

- New LCS

$$C[i][j] = \begin{cases} 0 & (i = 0 \text{ or } j = 0) \\ c[i-1][j-1] + 1 & (i, j > 0, sim[i-1][j-1] > \varepsilon) \\ \max\{c[i-1][j], c[i][j-1]\} & (i, j > 0, a_i \neq b_j) \end{cases} \quad (2)$$

$sim[i-1][j-1] \leq \varepsilon$

- New similarity

$$\mathrm{sim}(A, B) = \frac{LCSL}{\max\{L(A), L(B)\}}$$

$$\mathrm{sim}(A, B) = \frac{2 * LCSL}{L(A) + L(B)} \qquad \mathrm{Sim}(A, B) = \frac{(LCSL + 1) * LCSL}{L(A) * LCSL + L(B)} \quad L(A) \leq L(B)$$

[Chen Y Z, Lu H J, Li L J. Automatic ICD-10 coding algorithm using an improved longest common subsequence based on semantic similarity[J]. PloS one, 2017, 12(3): e0173410.]

Semantic similarity with HowNet

- Sentence similarity:

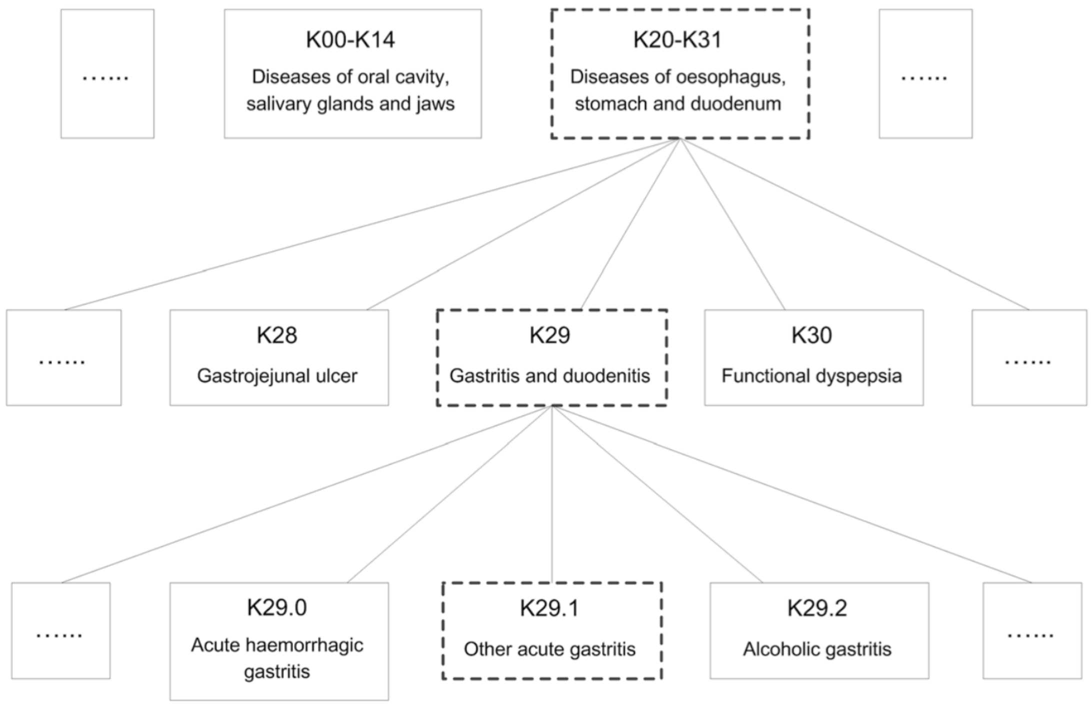
$$sim(T_1, T_2) = \frac{1}{2}\left(\frac{\sum_{w\in S(T_1,T_2,\theta)} (\max Sim(w, T_2)\cdot idf(w))}{\sum_{w\in\{T_1\}} idf(w)} + \frac{\sum_{w\in(T_2,T_1,\theta)} (\max Sim(w, T_1)\cdot idf(w))}{\sum_{w\in\{T_2\}} idf(w)}\right),$$

- Word similarity:
  - Both in HowNet: $sim(s_1, s_2) = \frac{\alpha}{distance(s_1, s_2) + \alpha}$
  - Otherwise： $sim(w_1,, w_2) = \frac{len(LCS(w_1,, w_2))}{len(w_1) + len(w_2) - len(LCS(w_1,, w_2))}$

[Ning W, Yu M, Zhang R. A hierarchical method to automatically encode Chinese diagnoses through semantic similarity estimation[J]. BMC medical informatics and decision making, 2016, 16(1): 30.]

Semantic similarity with HowNet

- Predict in a hierarchical way



[Ning W, Yu M, Zhang R. A hierarchical method to automatically encode Chinese diagnoses through semantic similarity estimation[J]. BMC medical informatics and decision making, 2016, 16(1): 30.]

Multi-label text classification

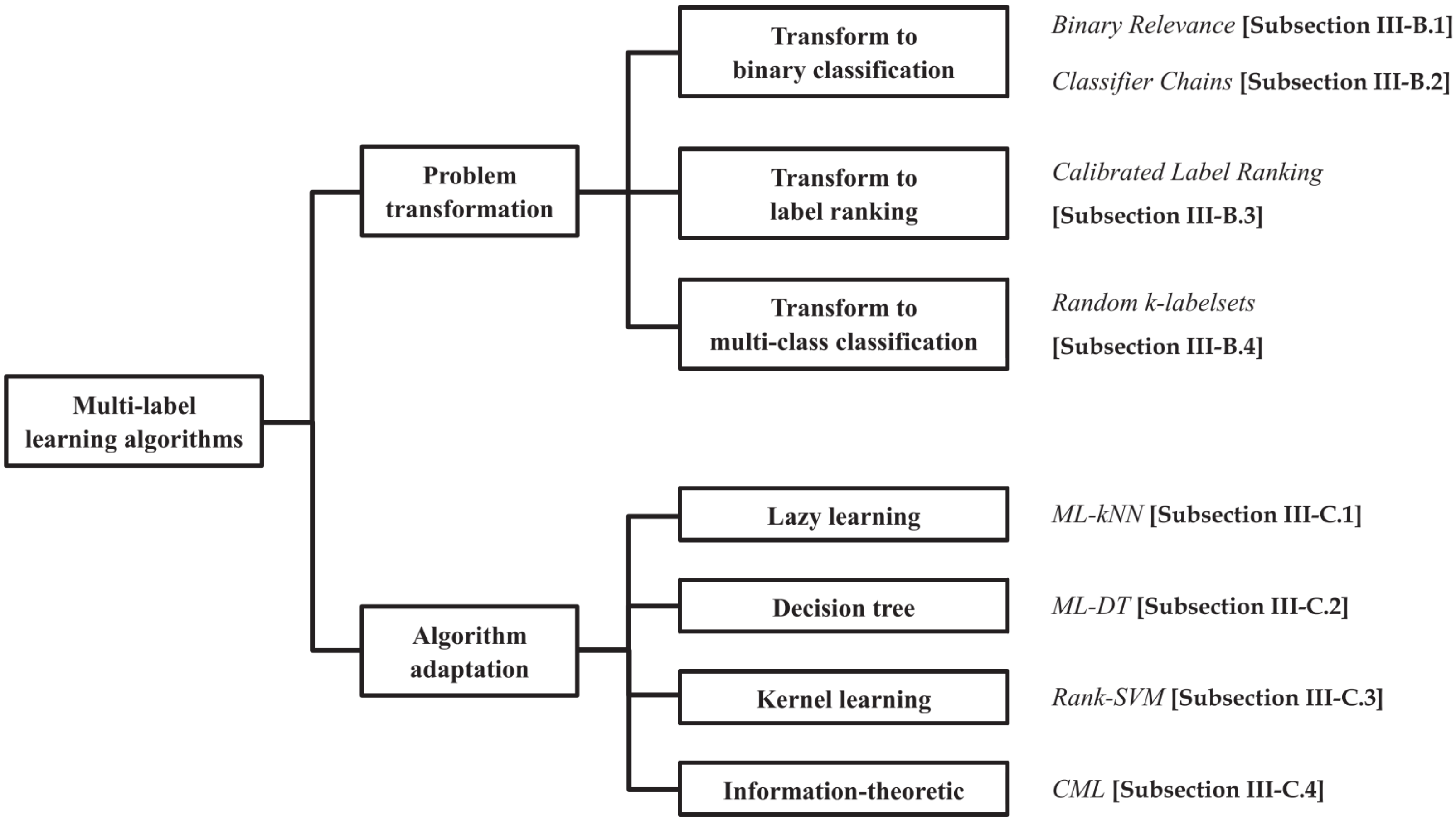
- Multi-label classification
  - Algorithms
  - Evaluation metrics
- Multi-label text classification
  - Binary Relevance
  - Label correlation
  - Label specific text representation
  - Label embedding
  - Others

Multi-label classification

Multi-Label Problem: **Output vector:**

Instance	Classes	A	B	C
1	A, B	1	1	0
2	A	1	0	0
3	A, B	1	1	0
4	C	0	0	1

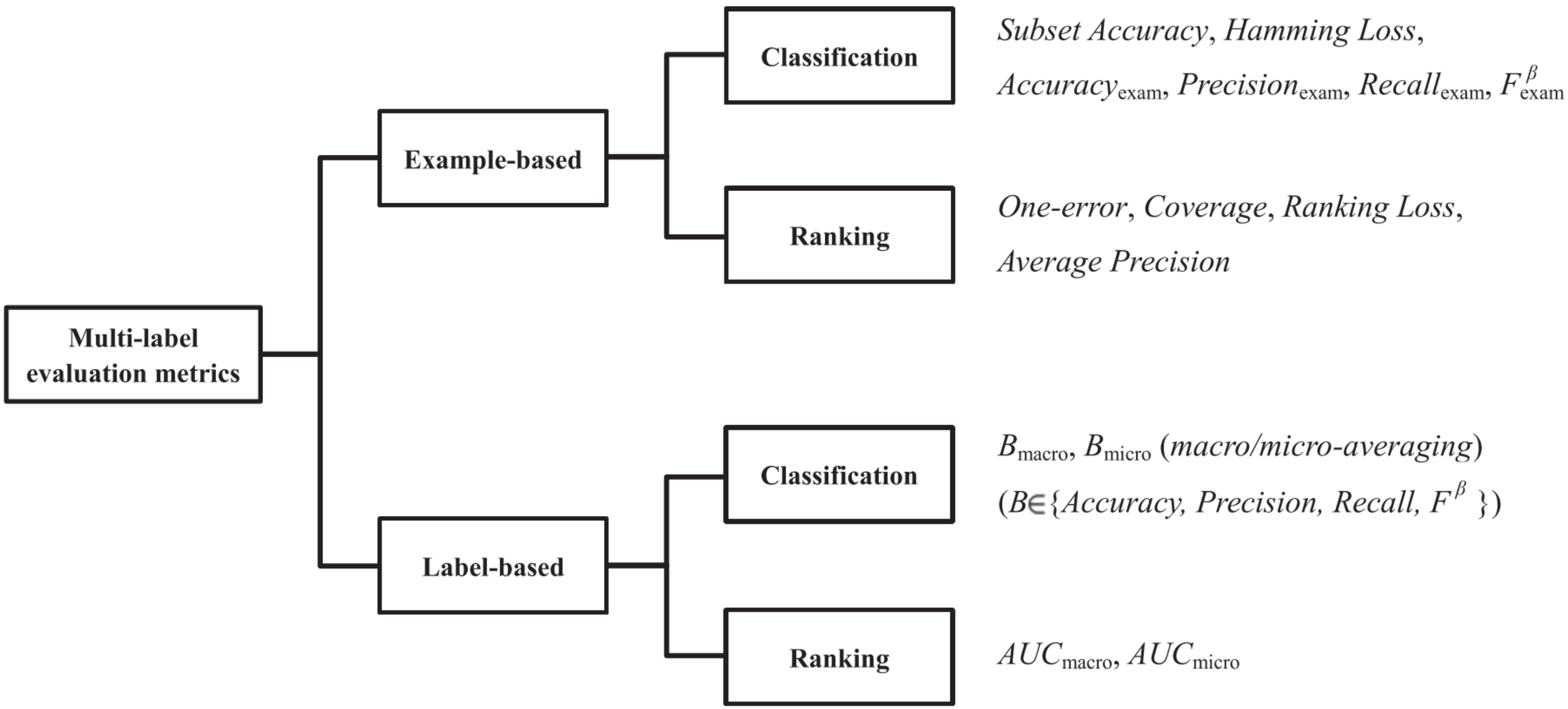
Multi-label classification algorithms



[Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE transactions on knowledge and data engineering, 2014, 26(8): 1819-1837.]

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Evaluation Metrics

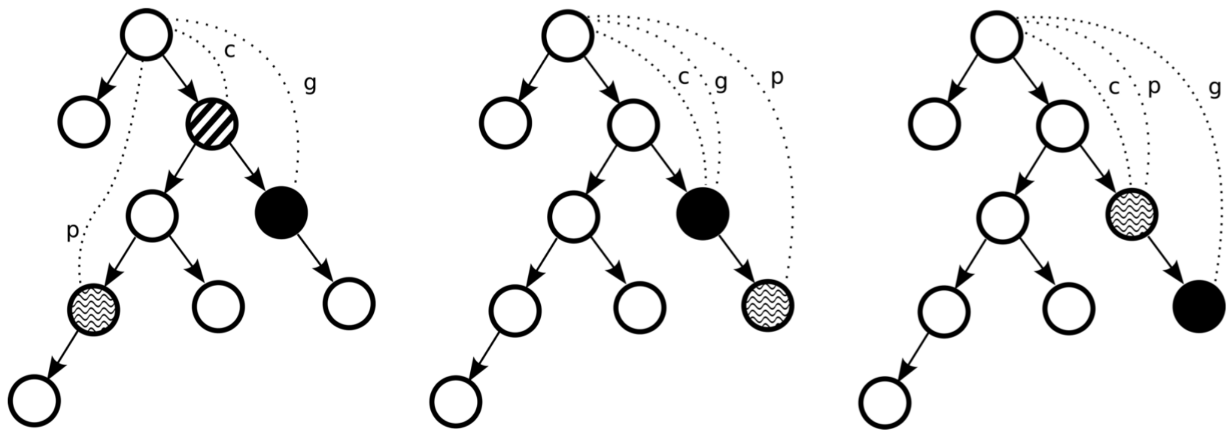


[Zhang M L, Zhou Z H. A review on multi-label learning algorithms[J]. IEEE transactions on knowledge and data engineering, 2014, 26(8): 1819-1837.]

Evaluation Metrics for ICD Coding

- Over coding & under coding problems

Figure 1 Quantities used in novel evaluation metrics for evaluation of automated ICD9 coding for different cases (left: prediction path diverges from the gold-standard path; middle: prediction is on the correct path but is too granular; and right: prediction is on the correct path, but is not granular enough).



Normalized divergent path to gold standard ((g−c)/g)

Normalized divergent path to predicted ((p−c)/p)

[Perotte A, Pivovarov R, Natarajan K, et al. Diagnosis code assignment: models and evaluation metrics[J]. Journal of the American Medical Informatics Association, 2013, 21(2): 231-237.]

Multi-label text classification

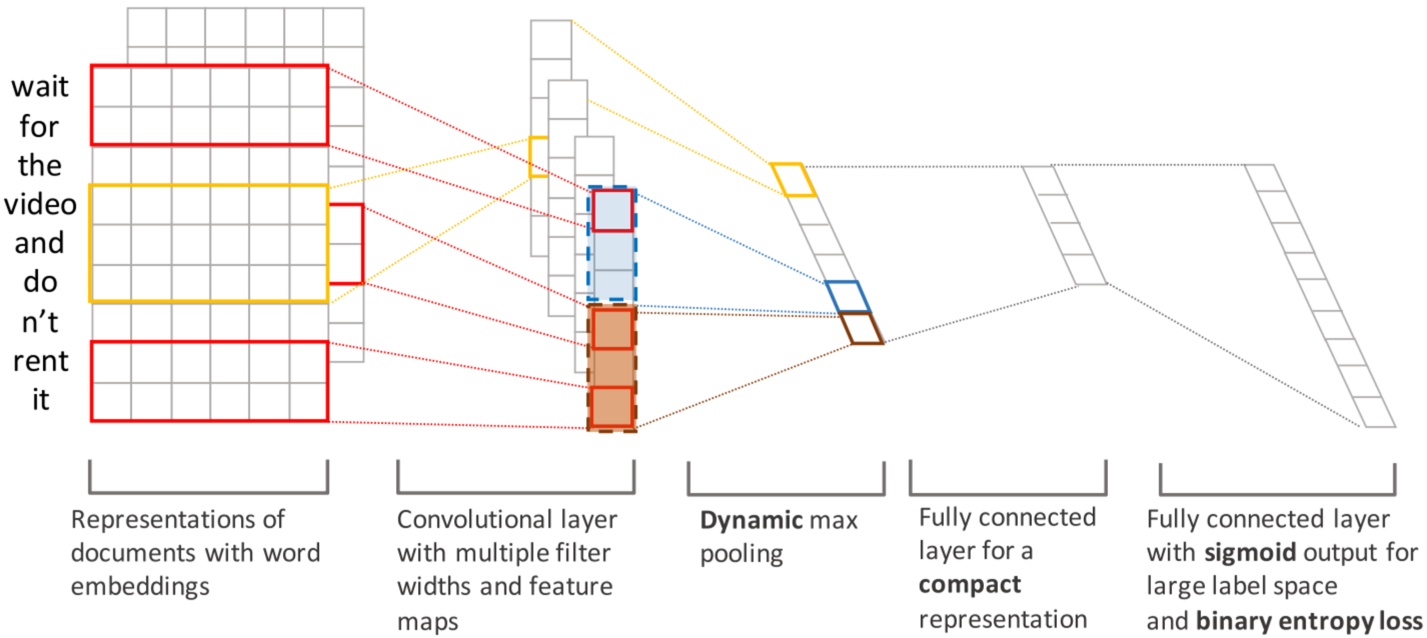
- Binary Relevance
- Label correlation
- Label specific text representation
- Label embedding
- Others

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Multi-label text classification

- Binary Relevance
- Label correlation
- Label specific text representation
- Label embedding
- Others

SIGIR 2017



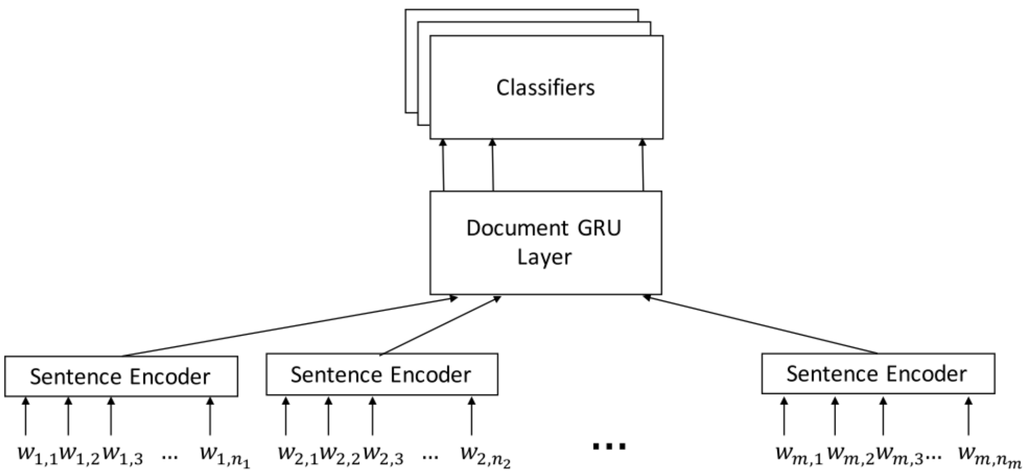
- Binary Cross-entropy objective

$$\min_{\Theta} \quad -\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}[y_{il}\log(\sigma(f_{il})) + (1-y_{il})\log(1-\sigma(f_{il}))]$$

[Liu J, Chang W C, Wu Y, et al. Deep Learning for Extreme Multi-label Text Classification[C]//SIGIR, 2017]
[Dembczynski K, Kotlowski W, Hüllermeier E. Consistent multilabel ranking through univariate losses[C] // ICML, 2012.]

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

AAAI 2018



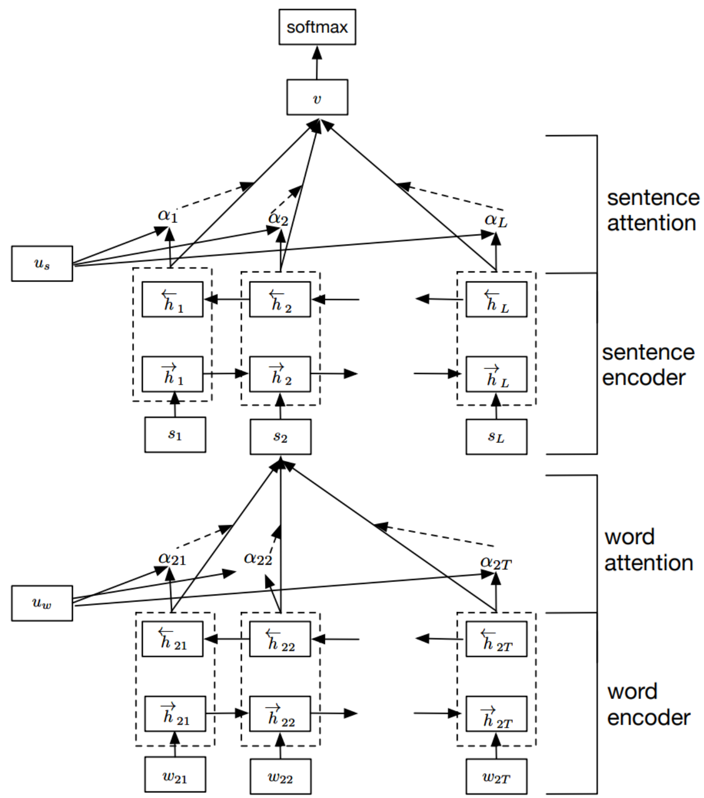
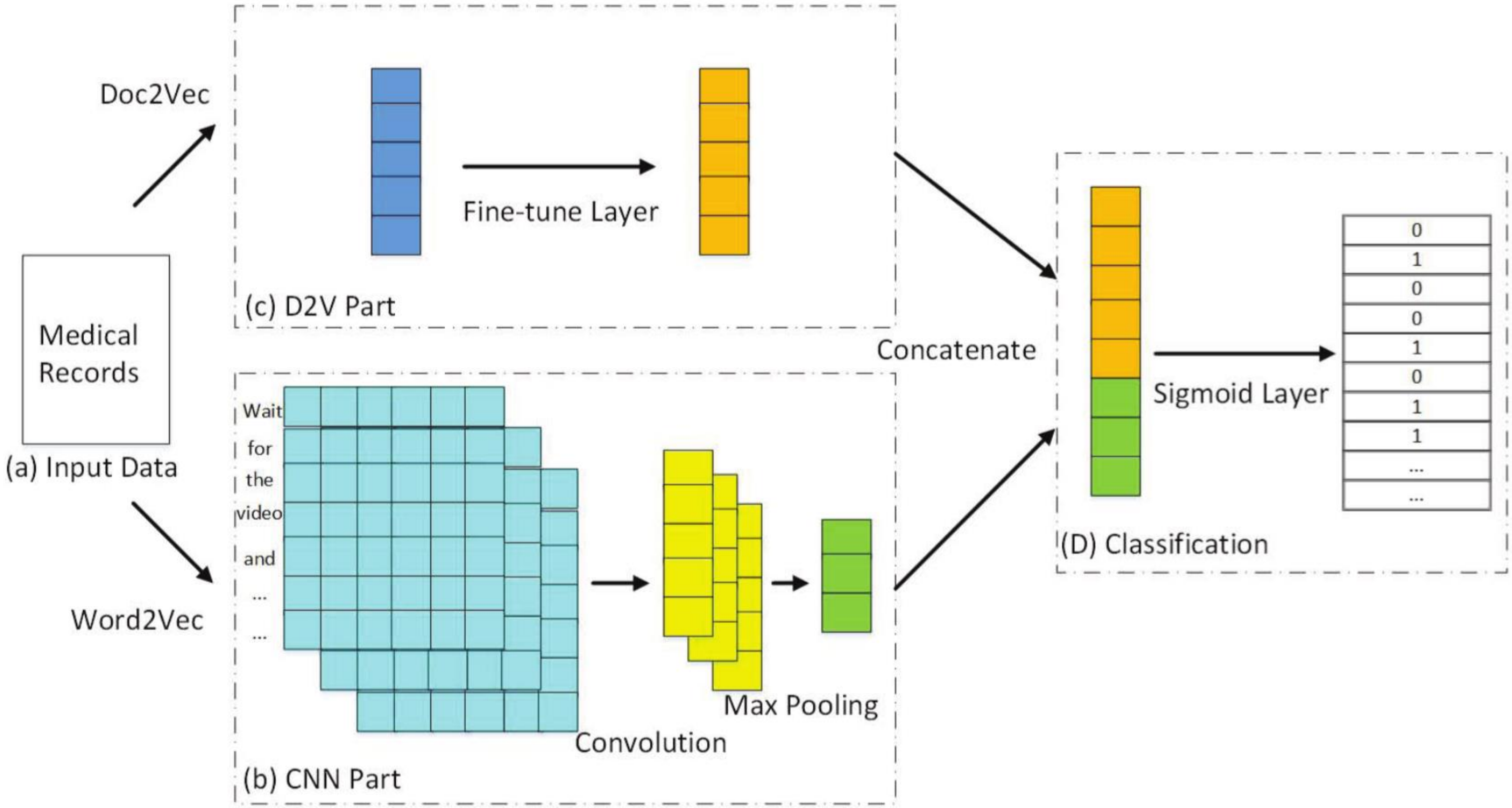


Figure 2: Hierarchical Attention Network.

[Baumel T, Nassour-Kassis J, Elhadad M, et al. Multi-Label Classification of Patient Notes a Case Study on ICD Code Assignment[J]. arXiv preprint arXiv:1709.09587, 2017.]

CNN+D2V



[Li M, Fei Z, Zeng M, et al. Automated ICD-9 Coding via A Deep Learning Approach[J]. IEEE/ACM Transactions on Computational Biology and Bioinformatics, 2018.]

Multi-label text classification

- Binary Relevance
- Label correlation
- Label specific text representation
- Label embedding
- Others

Modeling Label co-occurrence

$$P(C_i|C_j) = \frac{\exp(w_0 + \sum_{k=1}^{K} w_k \cdot F_k(C_i, C_j))}{1 + \exp(w_0 + \sum_{k=1}^{K} w_k \cdot F_k(C_i, C_j))}$$

where $\exp(\cdot)$ is the natural exponent, $F_k(C_i, C_j)$ are a set of K feature functions tracking various aspects of the codes C_i and C_j, as explained below, and w_k are the model weights estimated during the training phase.

This model is intended to capture solely the trends of code co-occurrence, leaving prediction of individual codes from the document to the primary auto-coder. Therefore, it does not use features that

[Subotin M, Davis A R. A method for modeling co-occurrence propensity of clinical codes with application to ICD-10-PCS auto-coding[J]. Journal of the American Medical Informatics Association, 2016, 23(5): 866-871.]

Modeling Label co-occurrence

Input:

$D_1 \dots D_M$: a set of M documents with manually assigned codes
$MAN(D_1) \dots MAN(D_M)$: sets of manually assigned codes
$GEN(D_1) \dots GEN(D_M)$: top-scoring outputs of primary auto-coder

For each D_i **in** $D_1 \dots D_M$:

  For each C_j^{man} **in** $MAN(D_i)$:

    For each C_k^{pred} **in** $GEN(D_i) \cup MAN(D_i)$:

      Extract features for estimate $P(C_k^{pred} | C_j^{man})$

      If $C_k^{pred} \in MAN(D_i)$:
        Generate positive training instance
      else:
        Generate negative training instance

[Subotin M, Davis A R. A method for modeling co-occurrence propensity of clinical codes with application to ICD-10-PCS auto-coding[J]. Journal of the American Medical Informatics Association, 2016, 23(5): 866-871.]

Modeling Label co-occurrence

Input:
$GEN(D_1) \dots GEN(D_M)$: top-scoring outputs of primary auto-coder
Data structures:
CURRENT: map of codes to current scores
FINAL: map of codes to final scores
QUEUE: priority queue of scored codes
For each D_i **in** $D_1 \dots D_M$:
Initialize *CURRENT* with $GEN(D_i)$ using primary auto-coder scores
Initialize *QUEUE* with $GEN(D_i)$ using primary auto-coder scores
Initialize *FINAL* to be empty
For i **from** 1 **to** depth of exploration d:
Pop C^{top} **from** *QUEUE*
$FINAL(C^{top}) \leftarrow CURRENT(C^{top})$
For each C_k **in** *QUEUE*:
$CURRENT(C_k) \leftarrow CURRENT(C_k) \times P(C_k | C^{top})$
Update *QUEUE* **with** *CURRENT*
For each C_k **in** *QUEUE*:
$FINAL(C_k) \leftarrow CURRENT(C_k)$
Output *FINAL*

[Subotin M, Davis A R. A method for modeling co-occurrence propensity of clinical codes with application to ICD-10-PCS auto-coding[J]. Journal of the American Medical Informatics Association, 2016, 23(5): 866-871.]

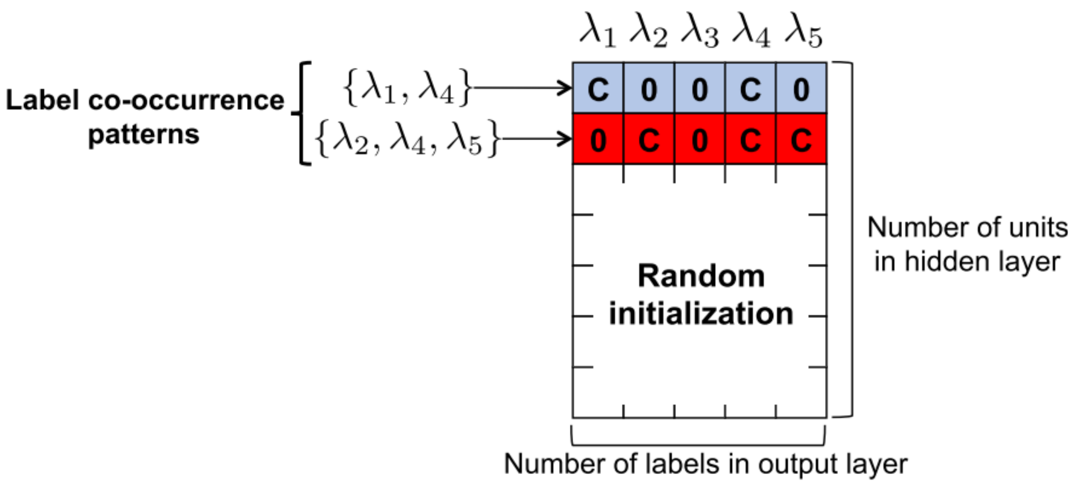
APEX 数据和知识管理实验室
Data & Knowledge Management Lab

Exploiting Associations between Class Labels

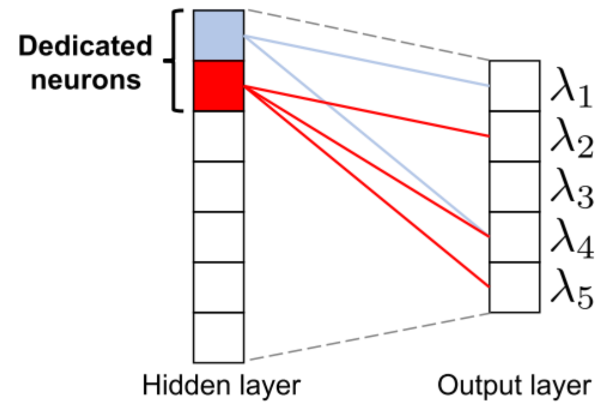
- Association rule extraction
  - E.g., (laptop → wireless mouse) [support: 20%, confidence: 80%]
- Types
  - Positive relationship: $y_1\ y_2 \rightarrow y_5$
  - Negative relationship: $\sim y_6 \rightarrow \sim y_3$
  - Hybrid relationship: $y_3 \rightarrow \sim y_6$ or $y_3 \sim y_2 \rightarrow y_1$
- Algorithm
  - Filter the extracted rules and keep the high quality rules
  - Apply the final rules in the prediction phase in order to correct the errors (where possible) to improve the classification results.

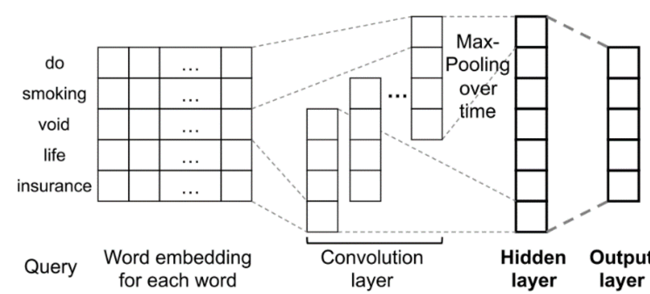
[Mirzamomen Z, Ghafooripour K. Exploiting Associations between Class Labels in Multi-label Classification[J]. Journal of AI and Data Mining, 2018.]

Better Weight Initialization



Weight matrix between hidden and output layers initialized by proposed method





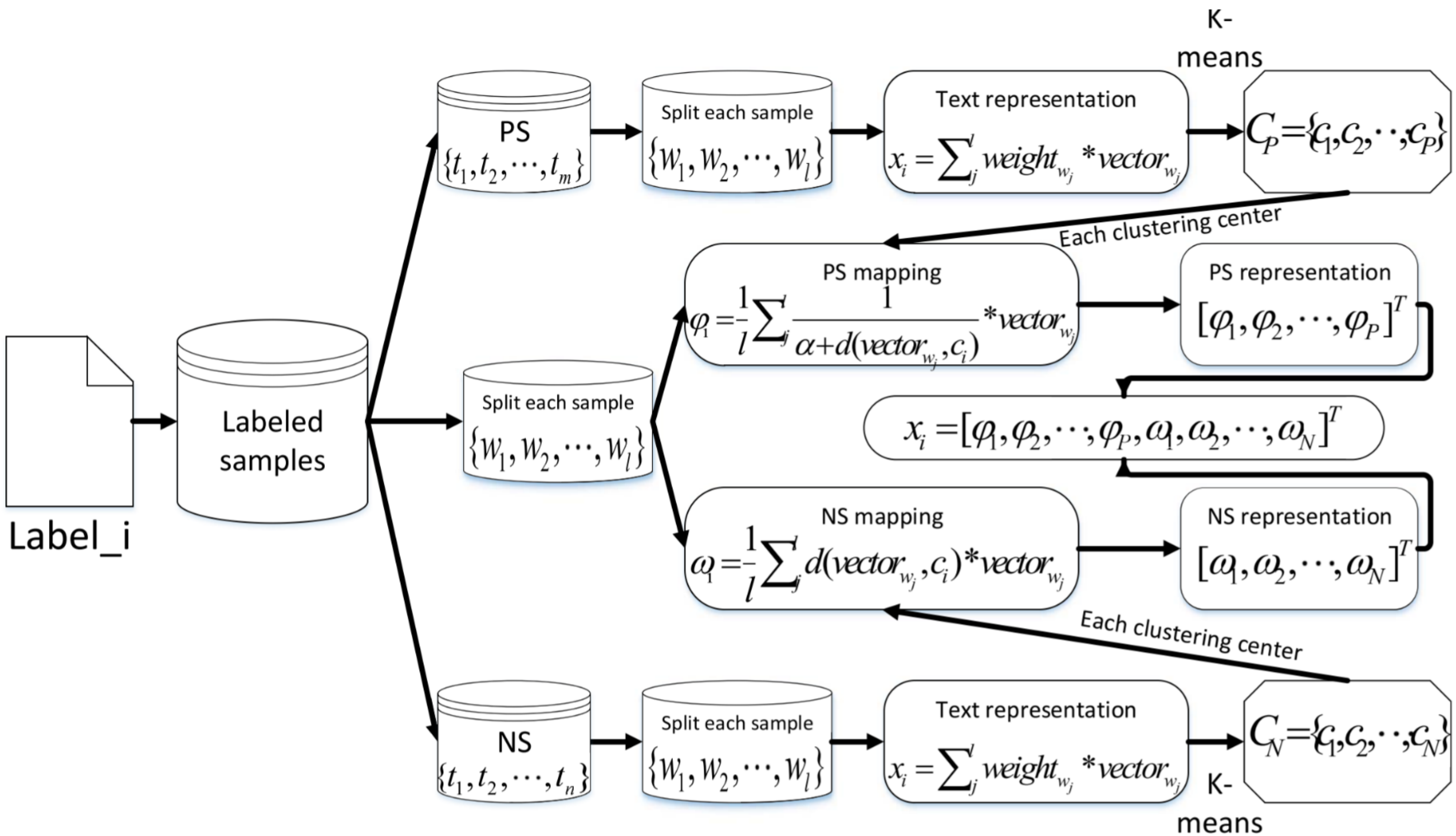
[Kurata G, Xiang B, Zhou B. Improved neural network-based multi-label classification with better initialization leveraging label co-occurrence[C]//Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 521-526.]

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Multi-label text classification

- Binary Relevance
- Label correlation
- Label specific text representation
- Label embedding
- Others

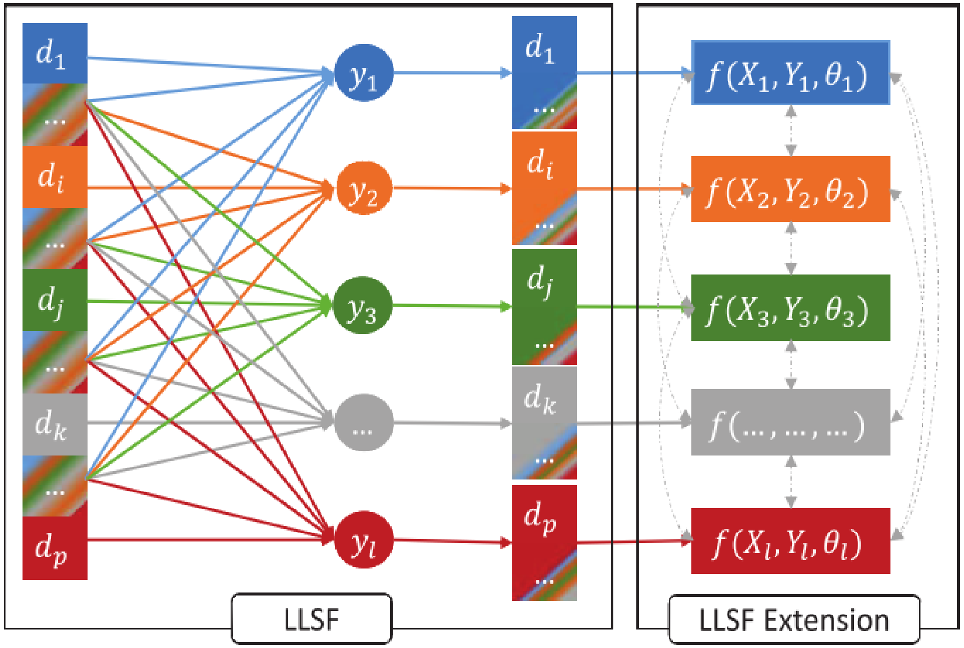
Labels Information Based Feature Mapping



[Wang T, Luo T, Li J, et al. Reasearch on feature mapping based on labels information in multi-label text classification[C]//Electronics Information and Emergency Communication (ICEIEC), 2017 7th IEEE International Conference on. IEEE, 2017: 452-456.]

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

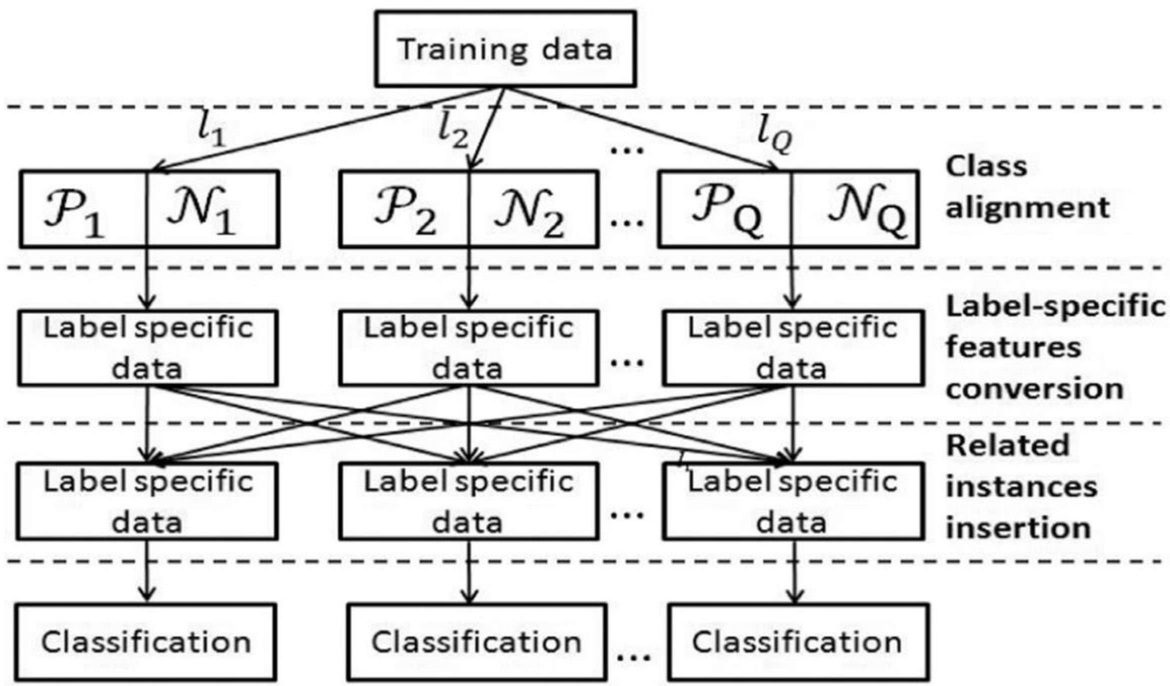
label specific features



[Huang J, Li G, Huang Q, et al. Learning label specific features for multi-label classification[C]//Data mining (ICDM), 2015 IEEE international conference on. IEEE, 2015: 181-190.]

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

label-specific features and local pairwise label correlation



[Weng W, Lin Y, Wu S, et al. Multi-label learning based on label-specific features and local pairwise label correlation[J]. Neurocomputing, 2018, 273: 385-394.]

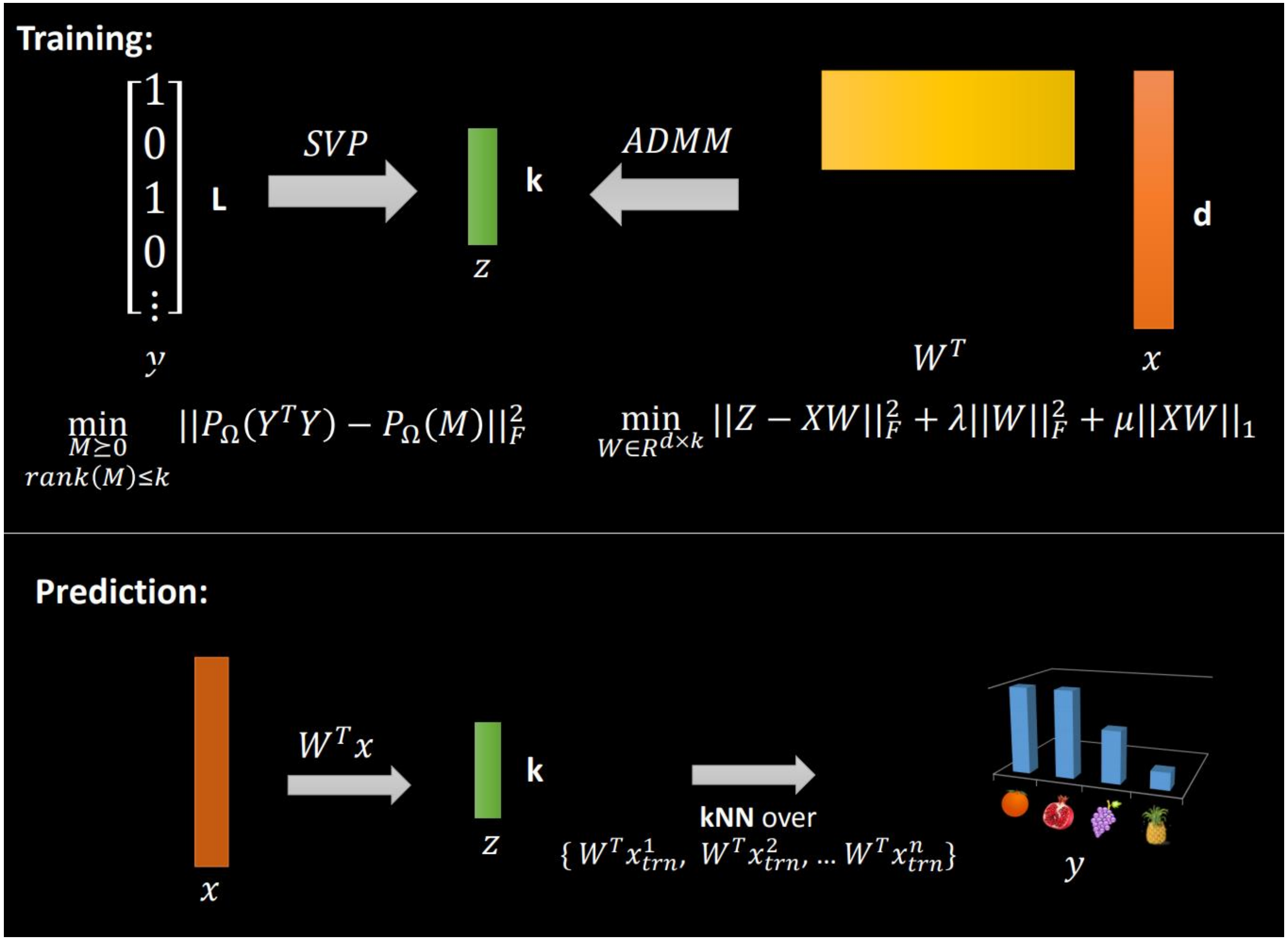
APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Multi-label text classification

- Binary Relevance
- Label correlation
- Label specific text representation
- Label embedding
- Others

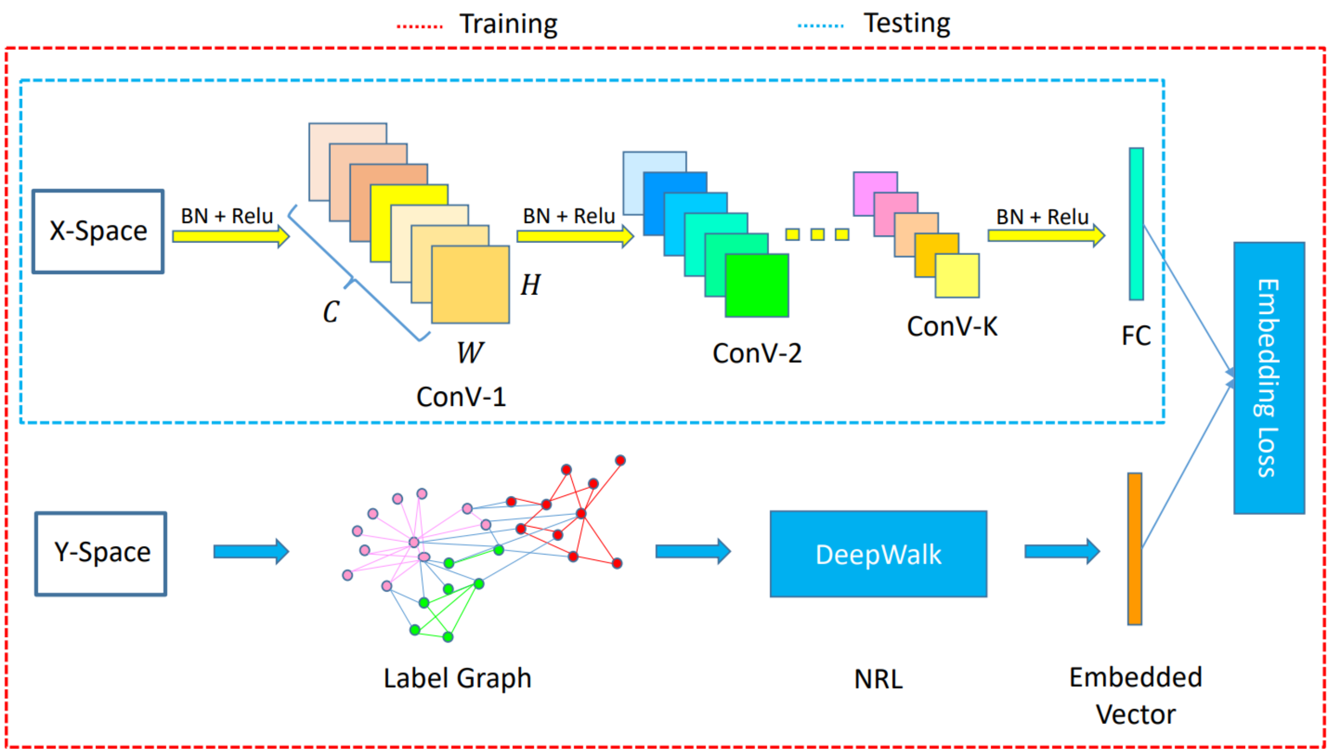
APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

SLEEC



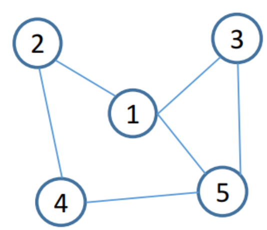
[Bhatia K, Jain H, Kar P, et al. Sparse local embeddings for extreme multi-label classification[C]//Advances in Neural Information Processing Systems. 2015: 730-738.]

Label Embedding with Graph



sample1	1, 3, 5
sample2	2, 4
sample3	1, 2
sample4	4, 5

Label space



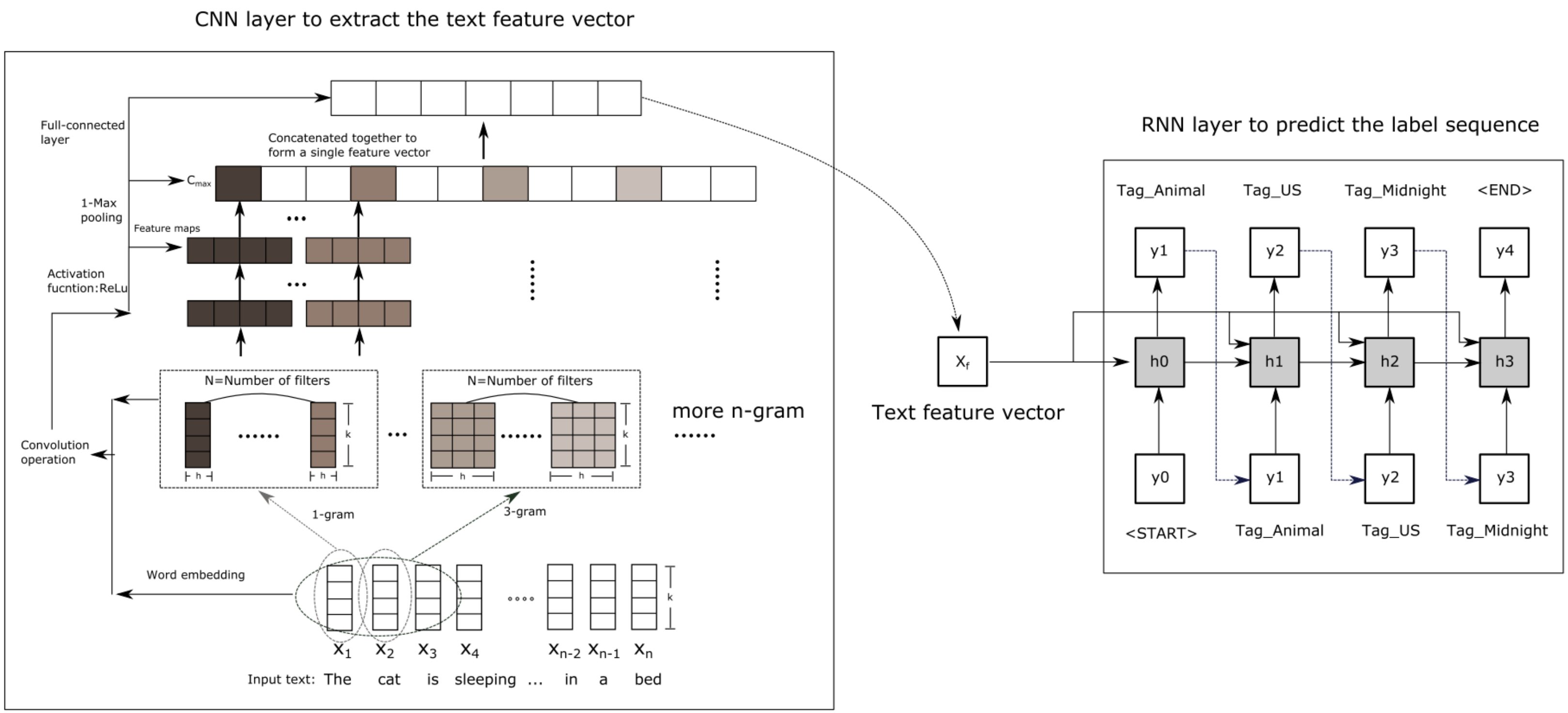
Label graph

[Zhang W, Wang L, Yan J, et al. Deep Extreme Multi-label Learning[J]. arXiv preprint arXiv:1704.03718, 2017.]

Multi-label text classification

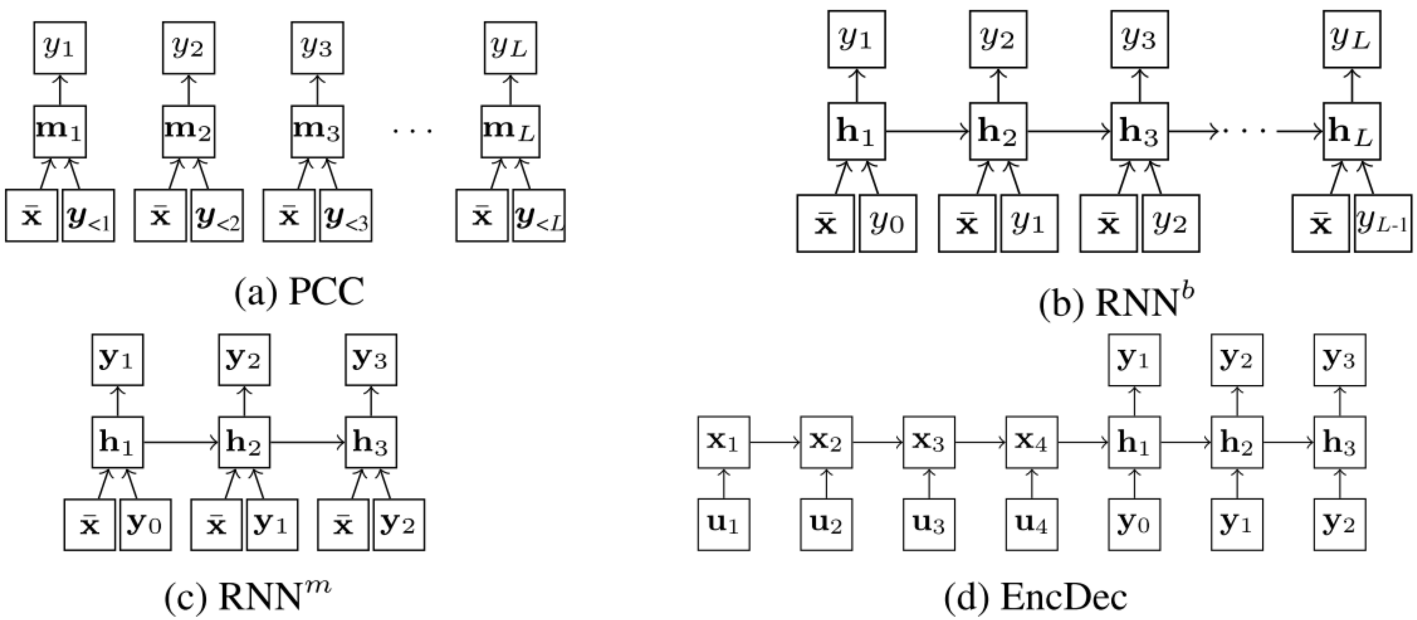
- Binary Relevance
- Label correlation
- Label specific text representation
- Label embedding
- Others
    - Classifier Chain
    - Code embedding

CNN-RNN Model



[Chen G, Ye D, Xing Z, et al. Ensemble application of convolutional and recurrent neural networks for multi-label text categorization[C]//Neural Networks (IJCNN), 2017 International Joint Conference on. IEEE, 2017: 2377-2383.]

RNN Model



[Nam J, Mencía E L, Kim H J, et al. Maximizing Subset Accuracy with Recurrent Neural Networks in Multi-label Classification[C]//Advances in Neural Information Processing Systems. 2017: 5419-5429.]

label Decomposition

- Fix ‘There are some combination class labels which are associated with records less frequently than others in training datasets.’

Physical Records	Disease Labels (Combination)
R	{A, B, C}

Physical Records	Disease Decompose Labels
R	{A, B}
R	{A, C}

[Li R, Zhao H, Lin Y, et al. Multi-label classification for intelligent health risk prediction[C]//Bioinformatics and Biomedicine (BIBM), 2016 IEEE International Conference on. IEEE, 2016: 986-993.]

Hierarchical Embedding

$$\min_{U,V} J(U,V) = \sum_{l=1}^{\mathcal{L}} \sum_{i \in S} h(y_{il}(x_i U V_l^T)) + \frac{\lambda}{2}(\|U\|_F^2 + \|V\|_F^2)$$

Algorithm 1: MLC-HMF ($\mathcal{X}$, $\mathcal{Y}$, k, $\mathcal{T}$, h).

input : Data Matrix: $\mathcal{X}$, Label Matrix: $\mathcal{Y}$, Size of Reduced Dimension Space: k, Threshold: $\mathcal{T}$, Depth of the Hierarchy: h
output: Tree with Mapping U and Label Feature Matrix V at Each Node

Divide $\mathcal{X}$ into $\mathcal{X}^1$ and $\mathcal{X}^2$ using *kmeans* clustering
for $i \in \{1,2\}$ **do**
  if $|\mathcal{X}^i|$ *is small or depth is exceed* h **then**
    Let its corresponding node as leaf node
    return
  end
  Learn the mapping U and label feature matrix V for $\mathcal{X}^i$ using Eq. (4).
  Let $\bar{\mathcal{X}} \subseteq \mathcal{X}^i$ is the set of instances whose hamming loss is less than the threshold $\mathcal{T}$ and $\bar{\mathcal{Y}}$ is their corresponding label matrix
  Maintain U, V and $\bar{\mathcal{X}}$ at the current node
  MLC-HMF ($\mathcal{X}^i \setminus \bar{\mathcal{X}}$, $\mathcal{Y}^i \setminus \bar{\mathcal{Y}}$, k, $\mathcal{T}$, h)
end

[Kumar V, Pujari A K, Padmanabhan V, et al. Multi-label classification using hierarchical embedding[J]. Expert Systems with Applications, 2018, 91: 263-269.]

Attention

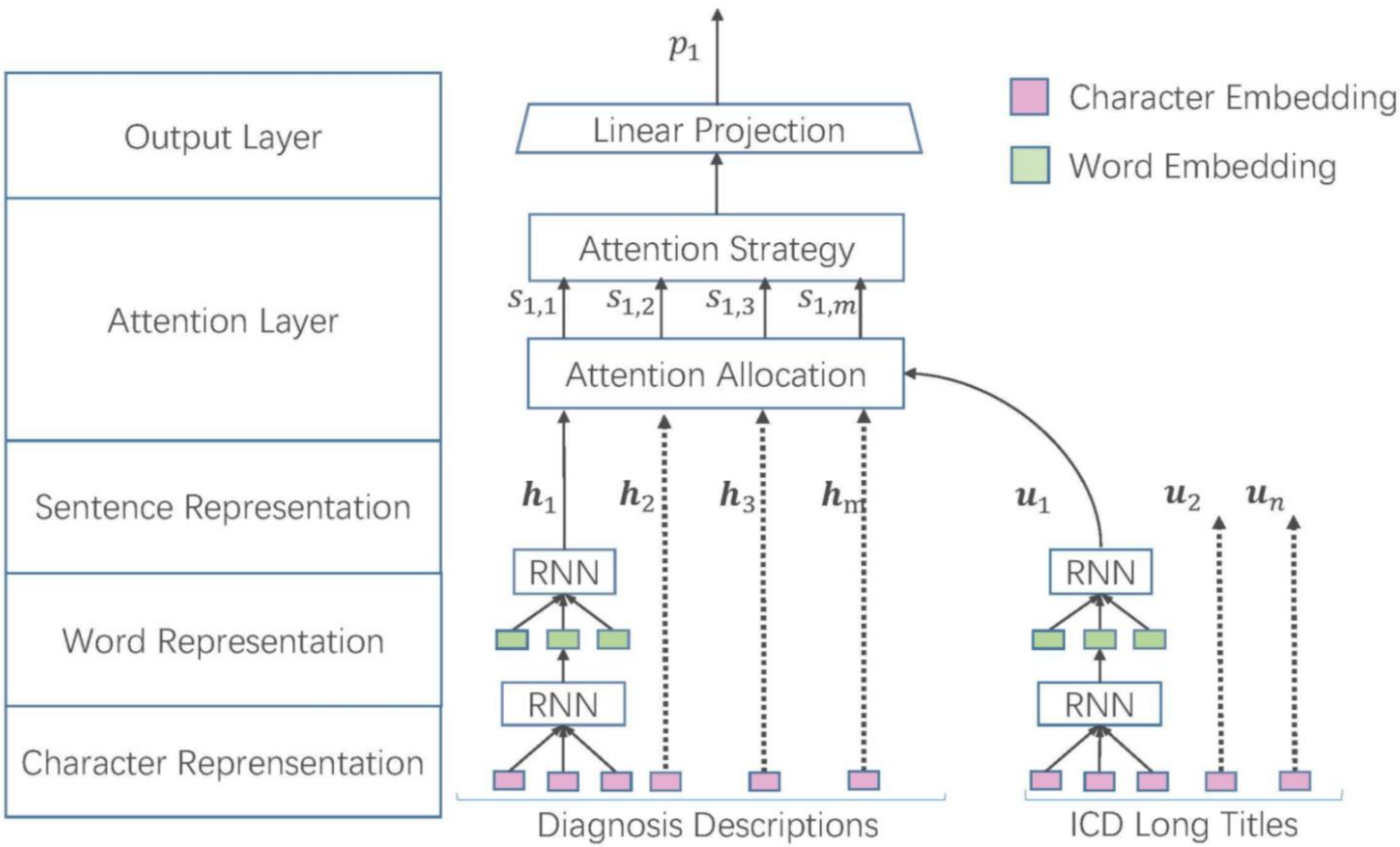
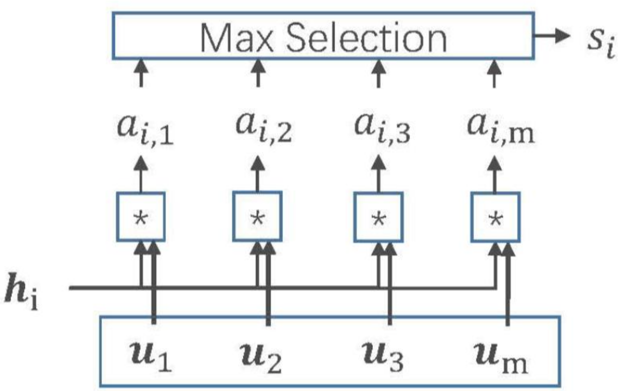


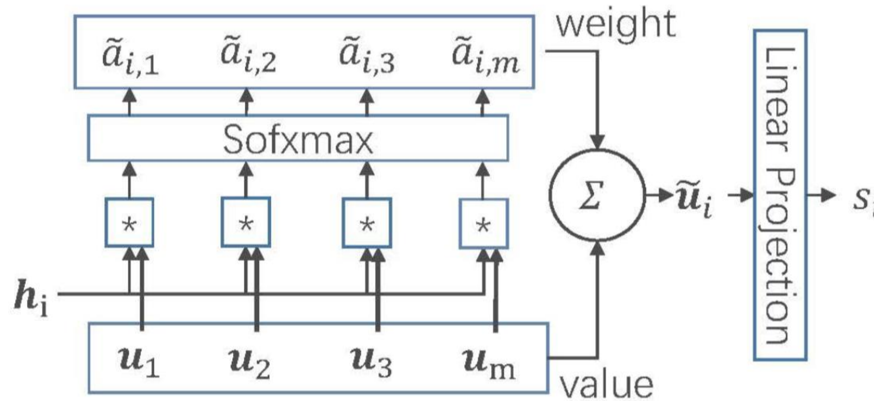
Figure 2. Model Architecture.

[Shi H, Xie P, Hu Z, et al. Towards Automated ICD Coding Using Deep Learning[J]. arXiv preprint arXiv:1711.04075, 2017.]

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB

Attention





$$\tilde{a}_{i,j} = \frac{exp(a_{i,j})}{\sum_{j=1}^{m}(exp(a_{i,j}))}$$

$$a_{i,j} = \sum_{k=1}^{d} u_{i,k} h_{j,k}$$

$$\tilde{\boldsymbol{u}}_i = \sum_{j=1}^{m} \tilde{a}_{i,j} * \boldsymbol{h}_j$$

$$p_i = \mathrm{sigmoid}(\max_{j=1,2,\dots m}(a_{i,j}))$$

$$s_i = \sum_{k=1}^{d} w_{i,k} \tilde{u}_{i,k}$$

$$p_i = \mathrm{sigmoid}(s_i)$$

[Shi H, Xie P, Hu Z, et al. Towards Automated ICD Coding Using Deep Learning[J]. arXiv preprint arXiv:1711.04075, 2017.]

Problems

- Performance
- Ontology (MeSH, SNOMED)
- Global and Local Label Correlation
- Cross & Multi Specialty

Thanks!

APEX 数据和知识管理实验室
DATA & KNOWLEDGE MANAGEMENT LAB