Zero-shot Text Classification

Dan Yuhao

2021.5.19

Overview

- Application Scenarios
- Problem Definition
- Learning Settings
- Vector Spaces
- Methods
- Papers in Chronological Order

Application Scenarios？

- **Too many classes**

  比如，对于目标识别任务，人类能识别的类别有30000，而我们不可能为这每个类别都收集样本。另一个例子是人类动作识别。人类能做的动作非常多，而我们能定义的动作是有限的，很多动作在现有的数据集中都没有样例。
- **Some classes are rare**

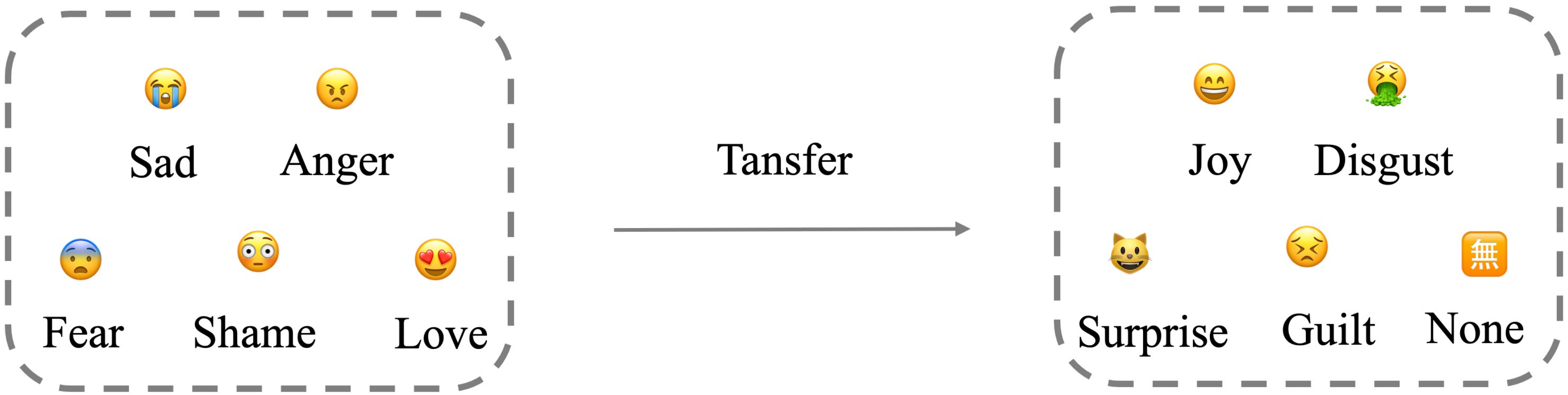
  例如我们不可能为一种花的每种细分品种都收集足够的图片样例。
- **Samples are changing over time**

  例如识别特定品牌或者风格的产品。而随着时间的推进，会有新的品牌以及新的产品出现。
- **High annotation cost**

  例如对于语义分割问题，需要对每个像素点的所属类别进行标记，带标签样本的数量是有限的。此外，现有数据集所覆盖的对象的类别的数量也是有限的。许多类别都没有带有标签的样例 。

A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Problem Definition



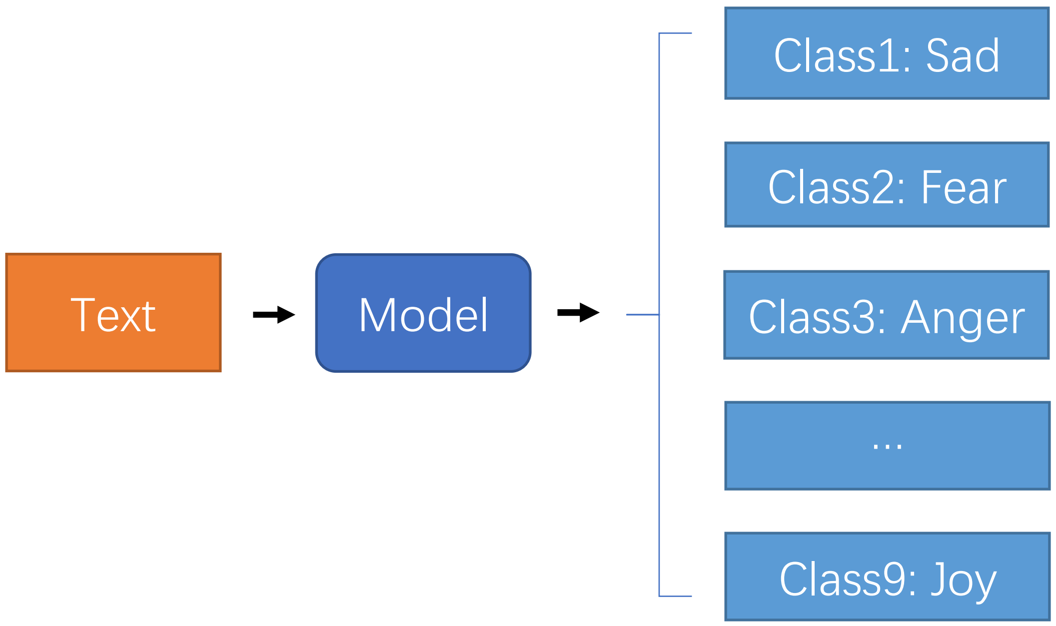
Seen Classes **Unseen Classes**

Training set Testing set

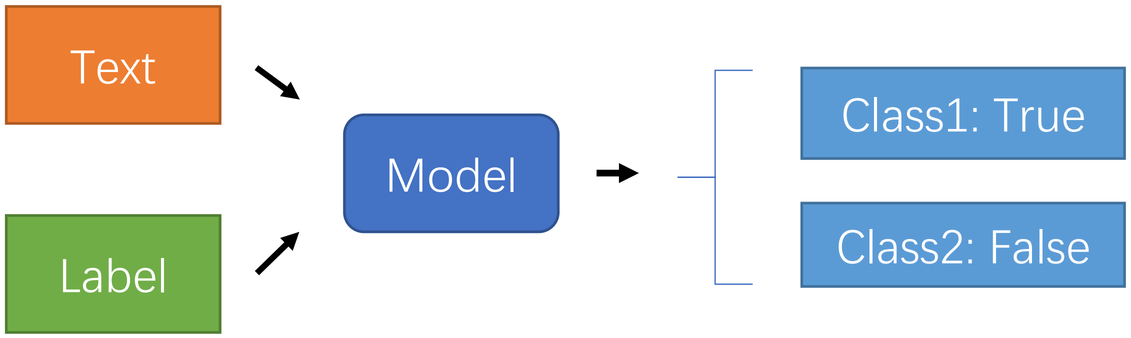
$$\mathcal{S} = \{c_i^s | i = 1, \ldots, N_s\} \qquad \mathcal{S} \cap \mathcal{U} \qquad \mathcal{U} = \{c_i^u | i = 1, \ldots, N_u\}$$

c_i^s is a seen class

c_i^u is an unseen class.



Classification Model



Zero-shot Classification Model

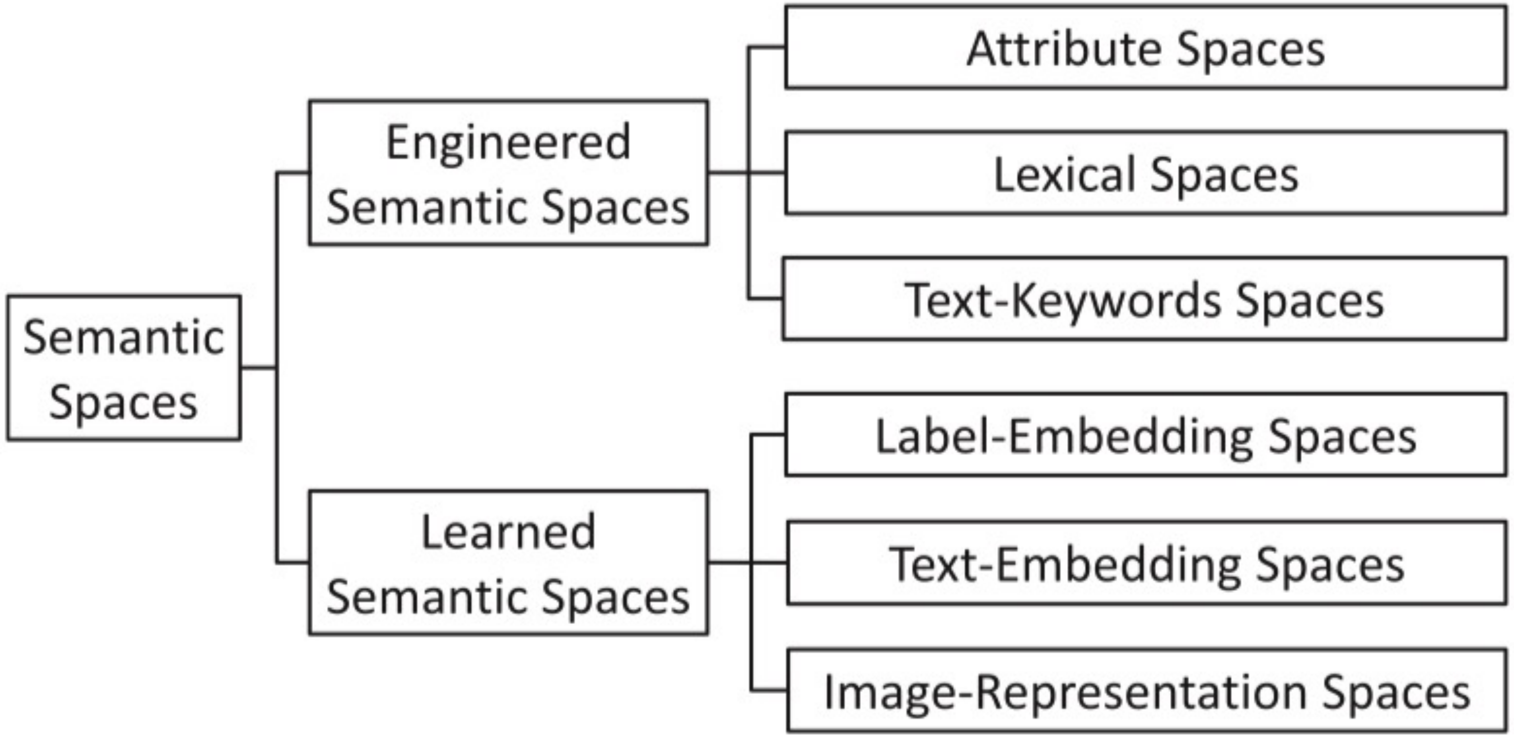
Learning settings

- 类别和样例都靠推测（CIII）
- 类别迁移，样例推测（CTII）
- 类别迁移，样例迁移（CTIT）

	带标签的训练样例	Seen class prototypes	Unseen class prototypes	不带标签的测试样例
CIII	√	√		
CTII	√	√	√	
CTIT	√	√	√	√

A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Semantic Spaces（如何构建prototype？）



A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Engineered Semantic Spaces

- **Attribute Space**

向量的每一个维度都是人工设定的某种性质，例如是否有尾巴。

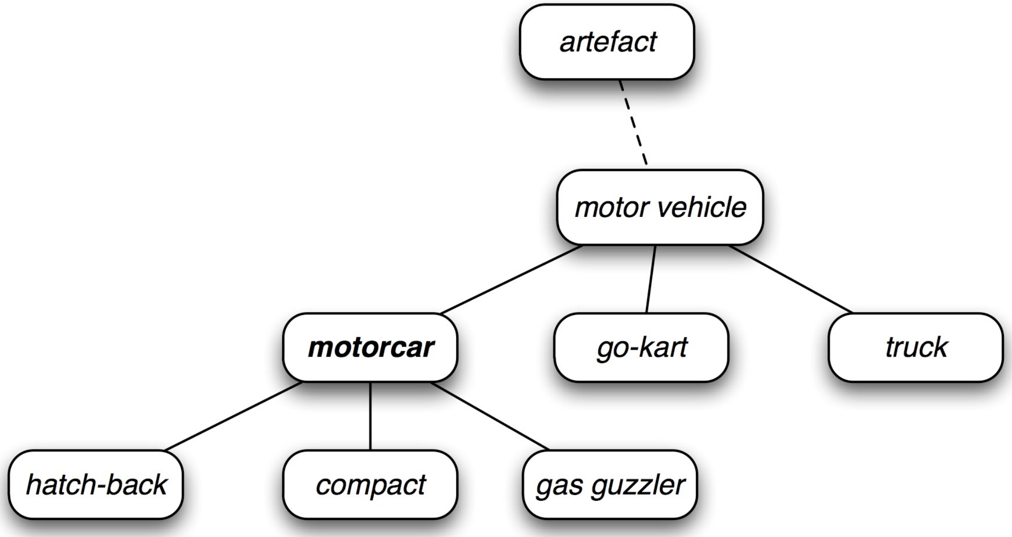
有条纹 →	0	1
有尾巴 →	1	1
吃草 →	1	1
可爱 →	1	0

A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Engineered Semantic Spaces

- **Lexical Space**

每一个维度都是一个和类别相关的词语，例如使用WordNet构建原型。



A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Engineered Semantic Spaces

- **Text-keyword Space**

从公开知识库，例如wiki中搜索类别的描述，再从描述中抽取词语，例如选择名词，来构成原型的每个维度。

Educational institution is a place where people of different ages gain an education

A Survey of Zero-Shot Learning: Settings, Methods, and Applications [ACM Transactions on Intelligent Systems and Technology 2019]

Learned Semantic Spaces

- **Label-embedding spaces**

利用Word2vec或者Glove这样的词嵌入来构建类别的原型。

- **Text-embedding spaces**

将类别的文本描述整个放到一个编码器中，将输出的嵌入作为原型。

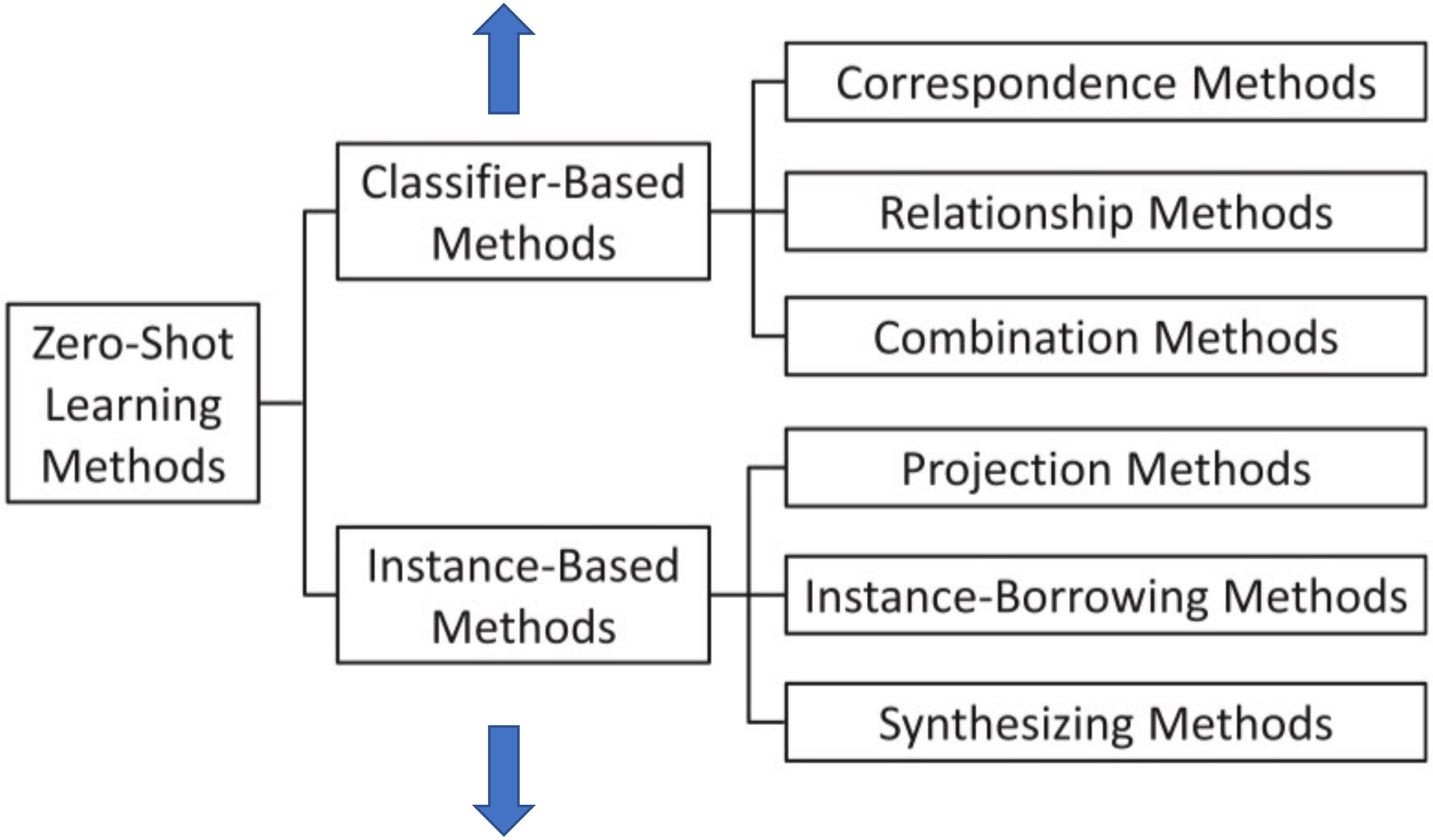
- **Image-representation spaces**

将属于某个类别的图片输入预训练模型，将得到的嵌入结合起来作为原型。

A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Methods

如何为unseen classes建立分类器？



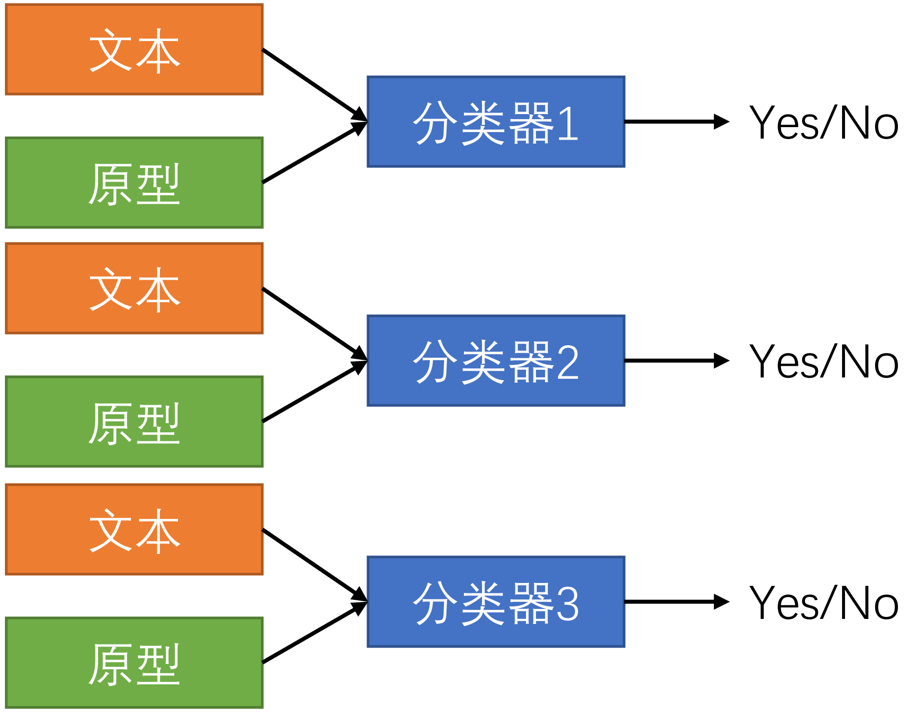
如何为无标签数据获得伪标签，并用这些数据训练模型？

A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Classifier-Based Methods

- **Correspondence methods**

为每个类别建立一个one vs rest的分类器，并与对应的原型关联起来。

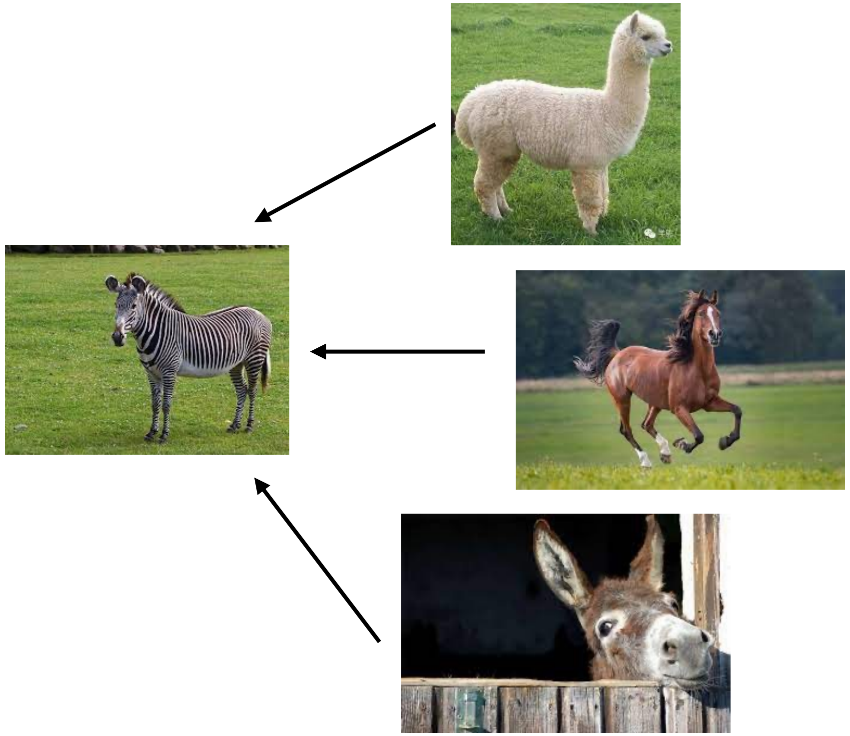


A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Classifier-Based Methods

- **Relationship methods**

根据类别间的关系来为没有见过的类别建立分类器 。



$$f_i^u(\cdot) = \sum_{j=1}^{K} \delta(c_i^u, c_j^s) \cdot f_j^s(\cdot),$$

Unseen类别的原型，
例如斑马的原型

Seen类别的原型，
例如马的原型

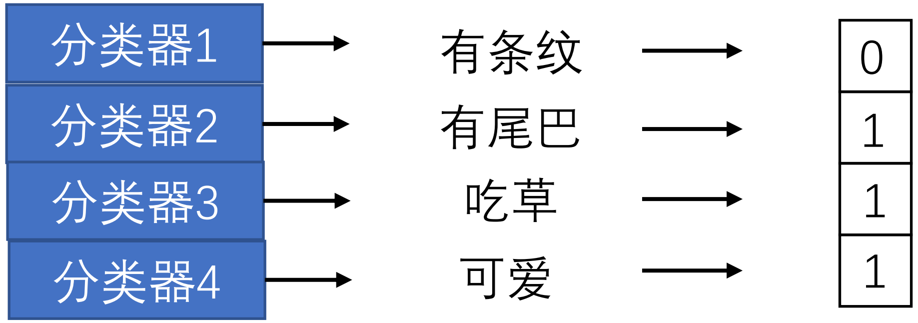
Seen类别的分类器，
例如马的分类器

A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

Classifier-Based Methods

- **Combination methods**

采用Attribute space作为原型，构建分类器预测原型上的每一个维度（属性）。



A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

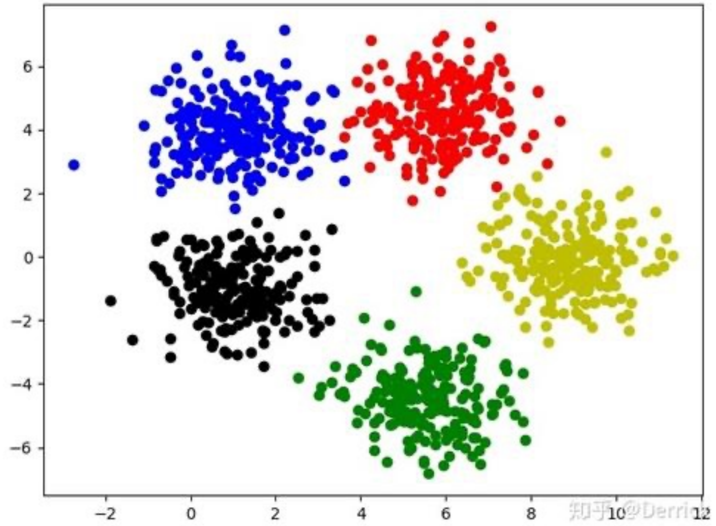
Instance-Based Methods

- **Projection methods**

将feature space(样例)和semantic space(原型)投影到一个共同的向量空间projection space。

$$\mathcal{X} \rightarrow \mathcal{P} : \mathbf{z}_i = \theta(\mathbf{x}_i),$$

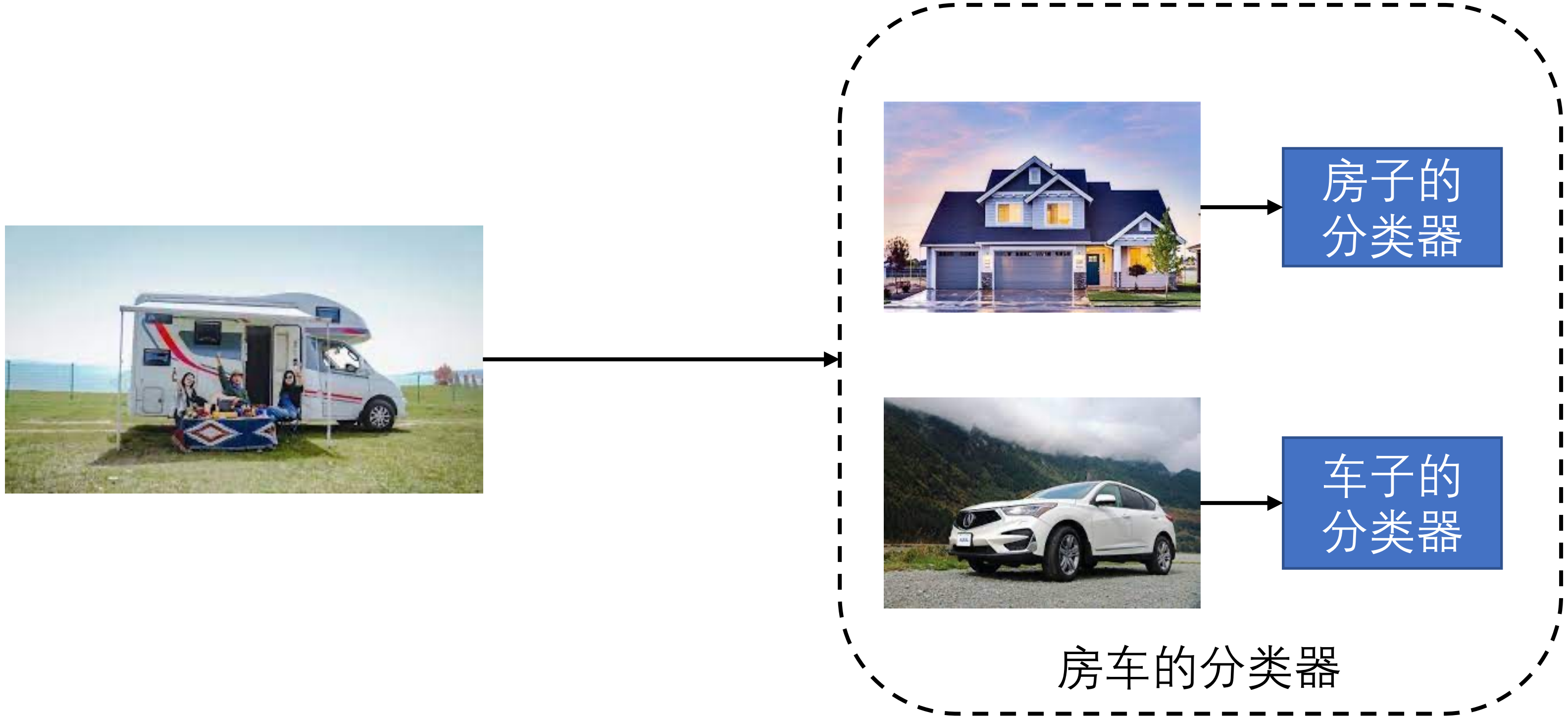
$$\mathcal{T} \rightarrow \mathcal{P} : \mathbf{b}_j = \xi(\mathbf{t}_j).$$



A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

- **Instance-borrowing methods**

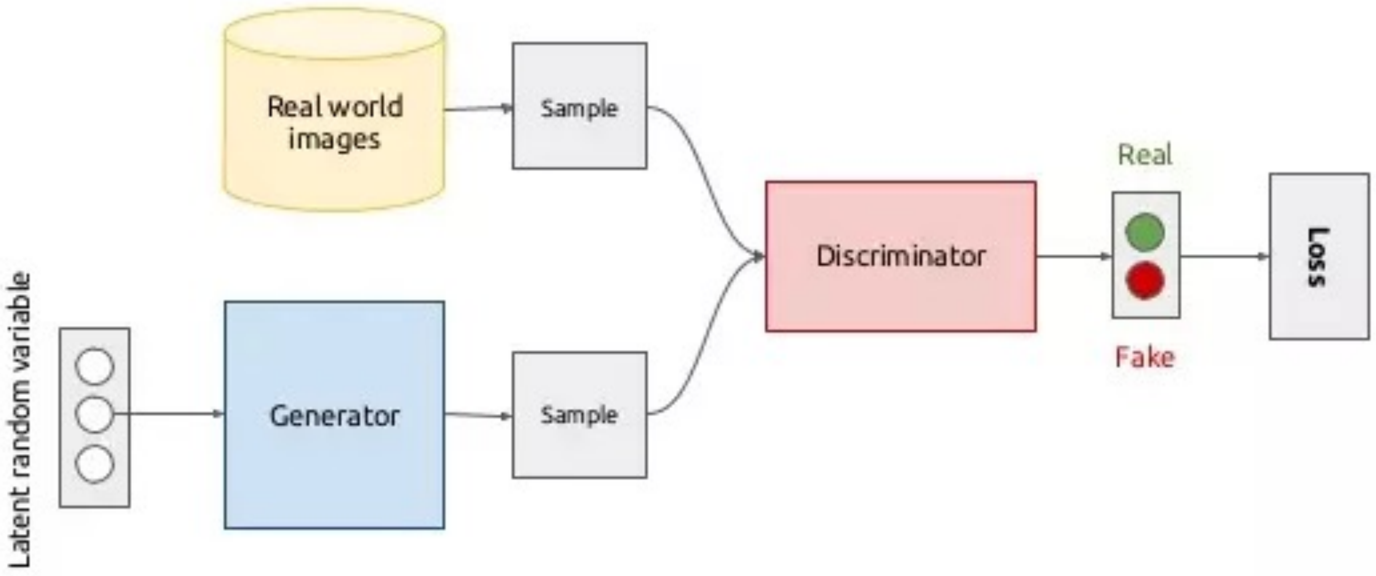
对于某个unseen class，我们从seen classes里头借用和它相似的样本，并打上这个unseen class的标签，对于此类方法来说，unseen classes需要被提前确定。



A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

- **Synthesizing methods**

首先为unseen class生成样例（例如使用GAN），然后将此样例打上标签，用于训练。此类方法的unseen classes也需要被提前确定。



A Survey of Zero-Shot Learning: Settings, Methods, and Applications
[ACM Transactions on Intelligent Systems and Technology 2019]

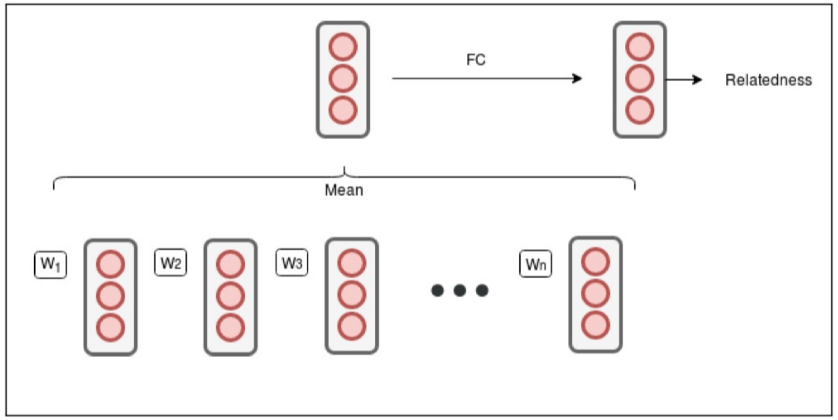


Figure 2: Architecture 1

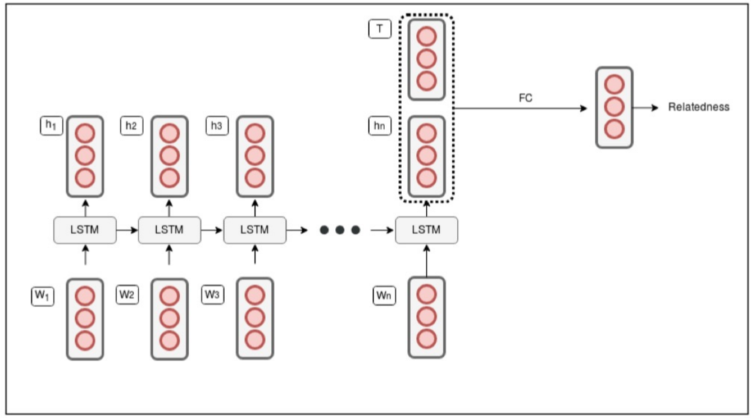


Figure 3: Architecture 2

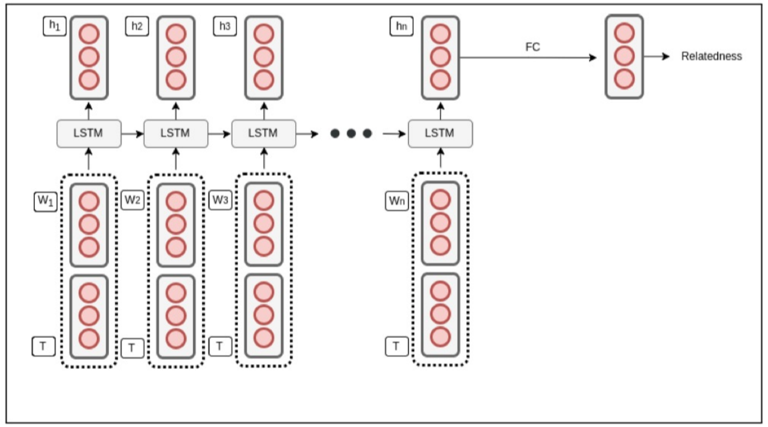
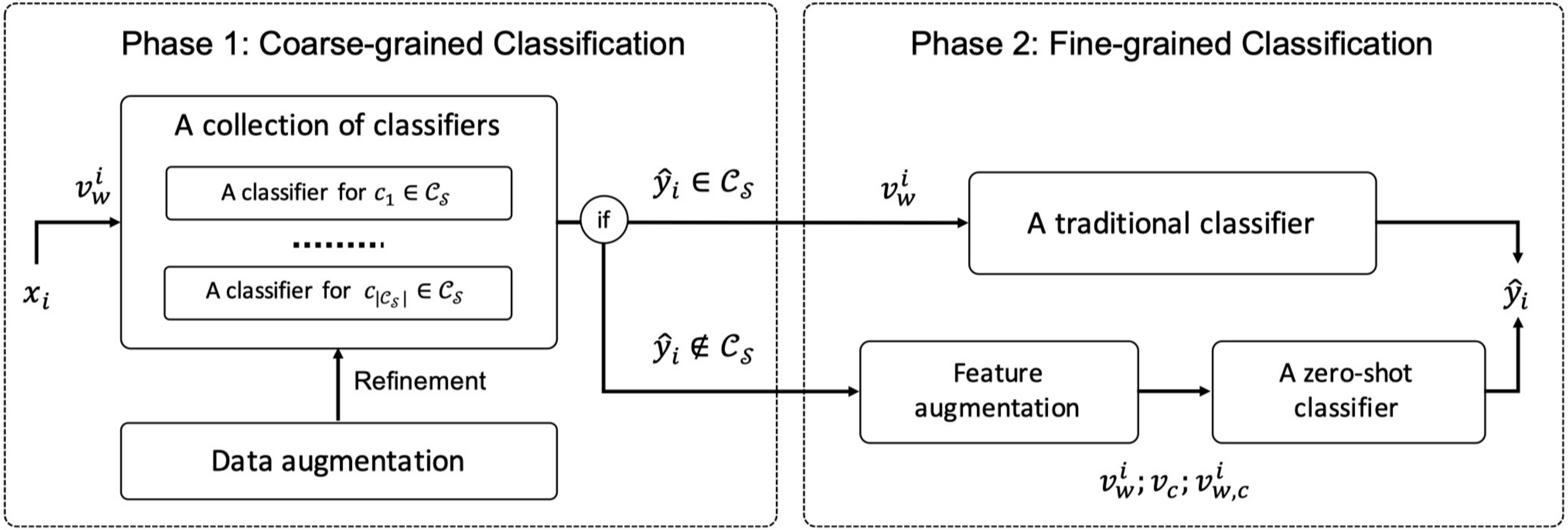


Figure 4: Architecture 3

Train once, test anywhere zero-shot learning for text classification [ICLR 2018]



Integrating semantic knowledge to tackle zero-shot text classification [NAACL 2019]

Data augmentation

$$c{:}w :: c'{:}?$$

Animal (Original)	Mitra perdulca is a species of sea snail a marine gastropod mollusk in the family Mitridae the miters or miter snails.
Animal → Plant	Arecaceae perdulca is a flowering of port aster a naval mollusk gastropod in the fabaceae Clusiaceae the tiliaceae or rockery amaryllis.
Animal → Athlete	Mira perdulca is a swimmer of sailing sprinter an Olympian limpets gastropod in the basketball Middy the miters or miter skater.

Integrating semantic knowledge to tackle zero-shot text classification [NAACL 2019]

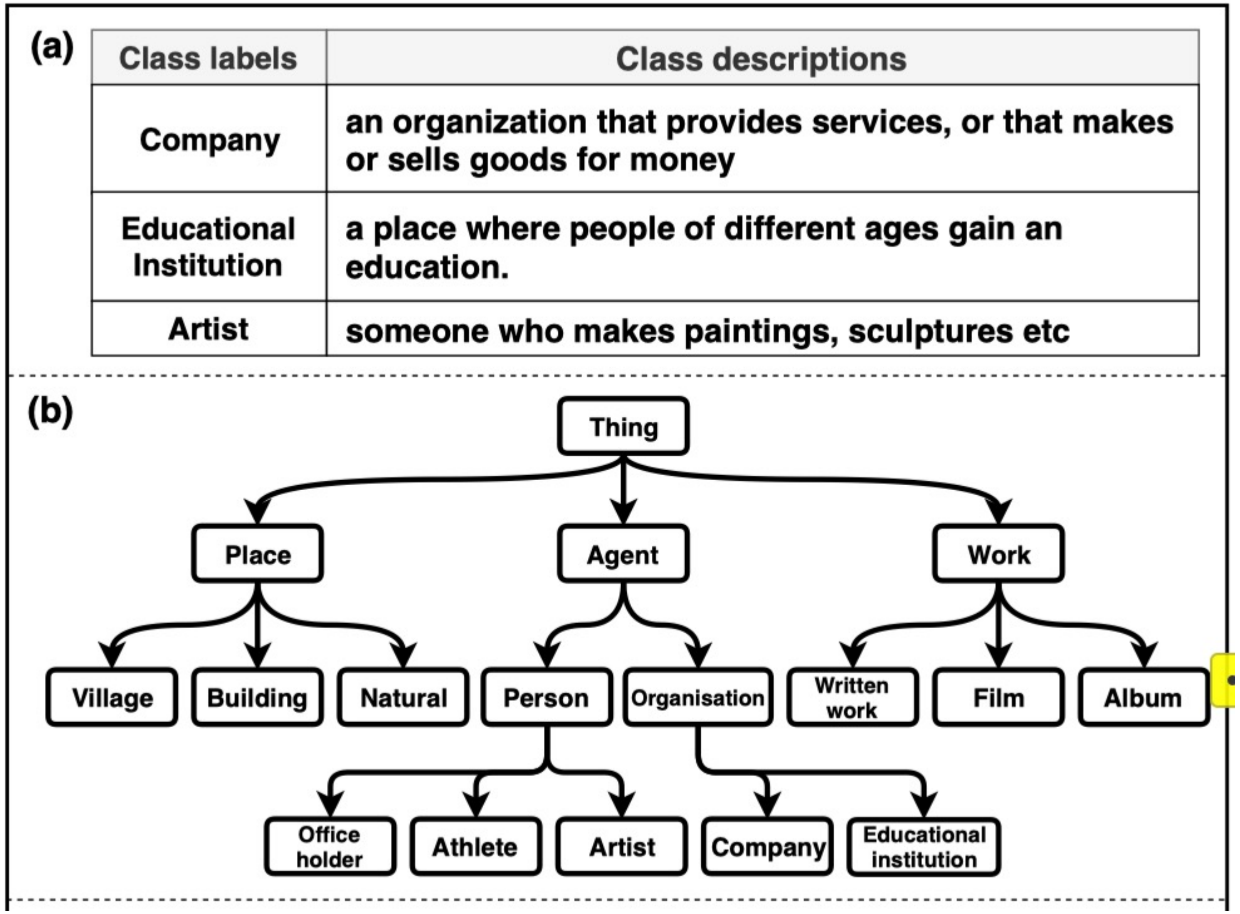
Feature augmentation

为每个Label创建三类词语集合：

- ✓类别本身包含的词语
- ✓ConceptNet中的上位词
- ✓描述中的名词

Educational Institution

- ✓educational institution,
  educational, institution
- ✓organization, agent
- ✓place, people, ages, education.

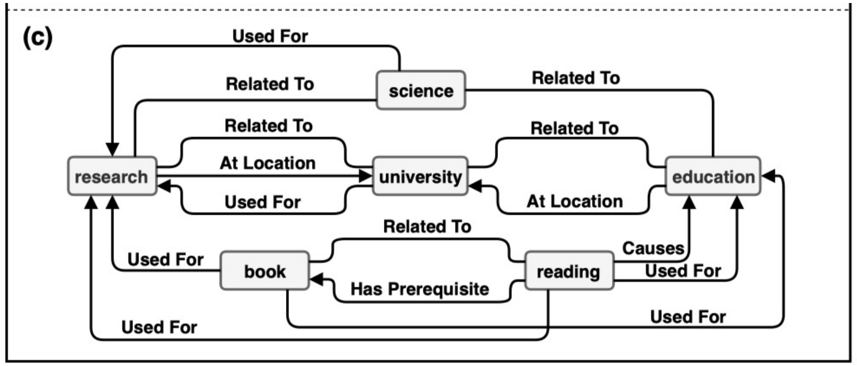


(a)

Class labels	Class descriptions
Company	an organization that provides services, or that makes or sells goods for money
Educational Institution	a place where people of different ages gain an education.
Artist	someone who makes paintings, sculptures etc

Integrating semantic knowledge to tackle zero-shot text classification [NAACL 2019]

Feature augmentation



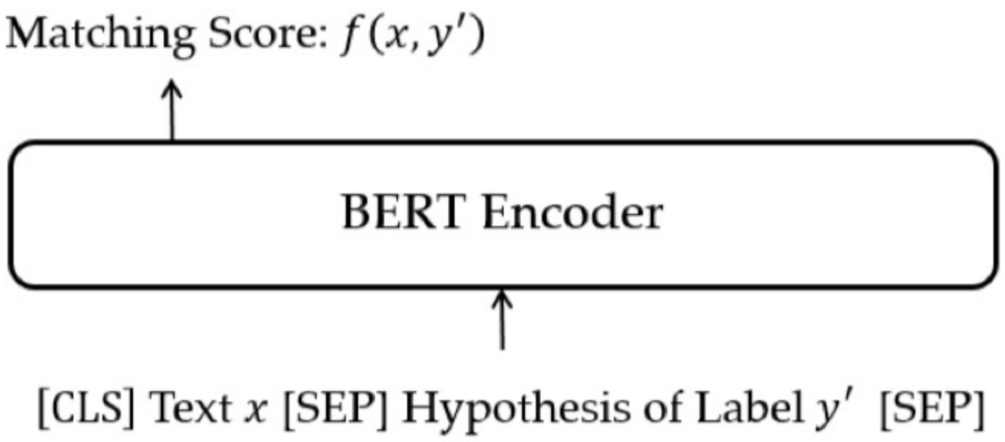
- $v[0] = 1$ if w_j is a node in that set; otherwise, $v[0] = 0$. → 词语在set中？
- for $k = 0, \ldots, K-1$:
  - $v[3k+1] = 1$ if there is a node in the set whose shortest path to w_j is $k+1$. Otherwise, $v[3k+1] = 0$. → 如果set中有词语到当前词语的距离是k跳，则为1
  - $v[3k+2]$ is the number of nodes in the set whose shortest path to w_j is $k+1$. → set中的词语到当前word的距离为k跳的数量。
  - $v[3k+3]$ is $v[3k+2]$ divided by the total number of nodes in the set. → set中的词语在word的第k跳中的数量和set中词语总数的比值

Integrating semantic knowledge to tackle zero-shot text classification [NAACL 2019]

Examples

Premise	Label	Hypothesis
Fiction		
The Old One always comforted Ca'daan, except today.	*neutral*	Ca'daan knew the Old One very well.
Letters		
Your gift is appreciated by each and every student who will benefit from your generosity.	*neutral*	Hundreds of students will benefit from your generosity.
Telephone Speech		
yes now you know if if everybody like in August when everybody's on vacation or something we can dress a little more casual or	*contradiction*	August is a black out month for vacations in the company.
9/11 Report		
At the other end of Pennsylvania Avenue, people began to line up for a White House tour.	*entailment*	People formed a line at the end of Pennsylvania Avenue.

Benchmarking zero-shot text classification: datasets, evaluation and entailment approach [EMNLP 2019]



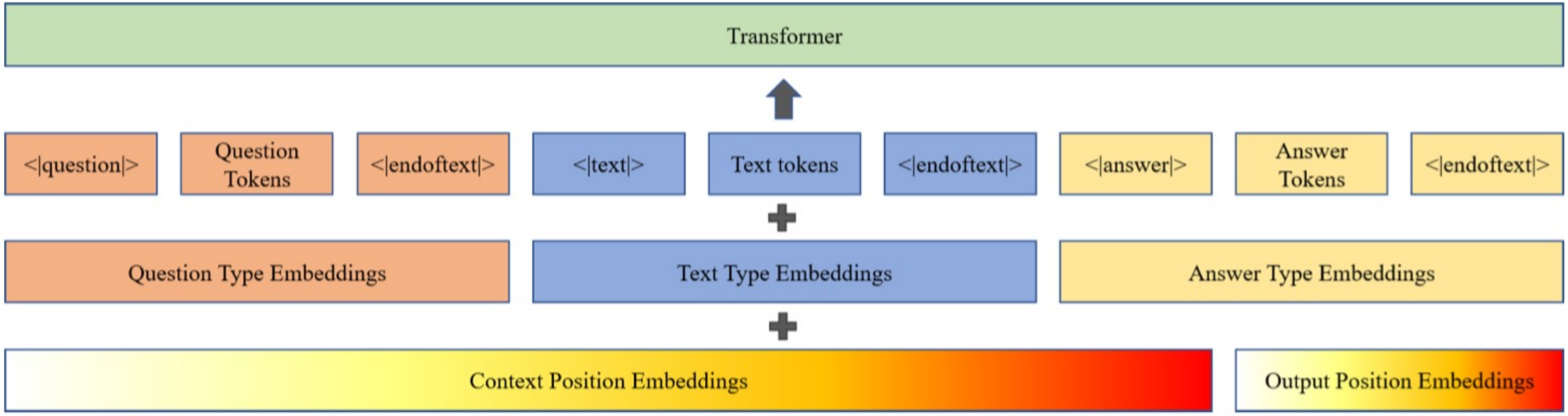
情感分类(joy)：
输入：[CLS] Text [SEP] the person feels joyful [SEP]
输出：二分类

情感分类(fear)：
输入：[CLS] Text [SEP] the person is afraid of something [SEP]
输出：二分类

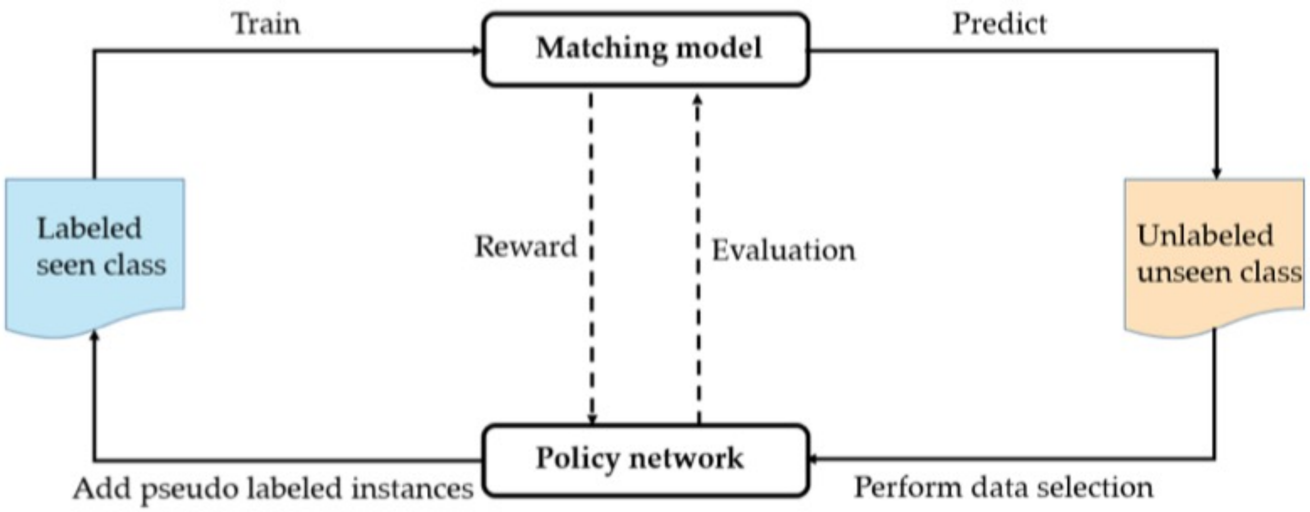
Benchmarking zero-shot text classification: datasets, evaluation and entailment approach [EMNLP 2019]

Dataset	Question	Text	Answer
Title Prediction Pretraining	Which of these choices best describes the following document? : " A pool For All Bodies " , " Lawmakers say they'd take pay cut, but they can't " , " Raiders' Gareon Conley faces civil suit " , " Prolific cybercriminal suspected of spreading ransomware arrested by Polish Police [Europol] "	Story highlights Members of Congress also preparing for potential sharp cuts in federal spending\n\nBut lawmakers will not see any change to their annual salary of $174,000...	Lawmakers say they'd take pay cut, but they can't
AGNews Zero-shot Classification	How is the text best described? : " Science & Technology " , " Business " , " Sports " , or " World News "	An Entertaining Holiday Pick \n Hastings, a multimedia retailer, trims losses and raises full-year guidance.	Business

Zero-shot Text Classification With Generative Language Models [NeurIPS 2019 Workshop]



Zero-shot Text Classification With Generative Language Models [NeurIPS 2019 Workshop]



State:

[CLS] of input: p_{x,y^*}

Prediction Confidence: c_{x,y^*}

Action:

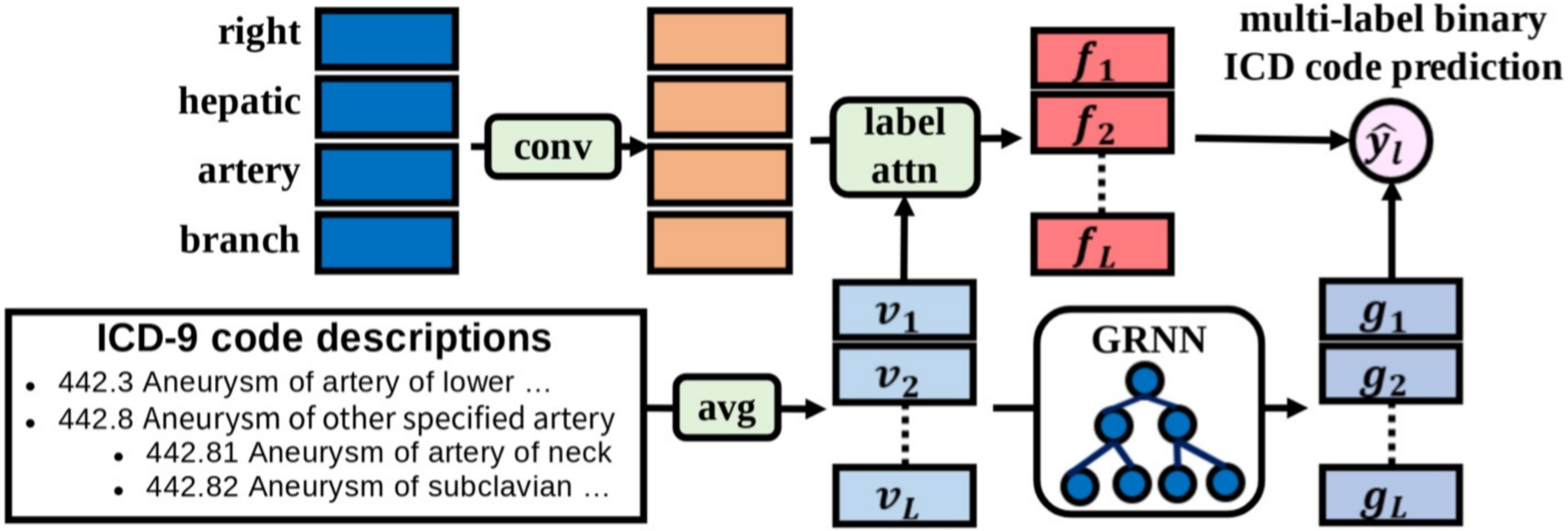
$$z_t = ReLU(W_1^T c_{x,y^*} + W_2^T p_{x,y^*} + b_1),$$

$$P(a|s_t) = softmax(W_3^T z_t + b_2)\ .$$

Reward:

$$r_k = \frac{(F_k^s - \mu^s)}{\sigma^s} + \lambda \cdot \frac{(F_k^u - \mu^u)}{\sigma^u}$$

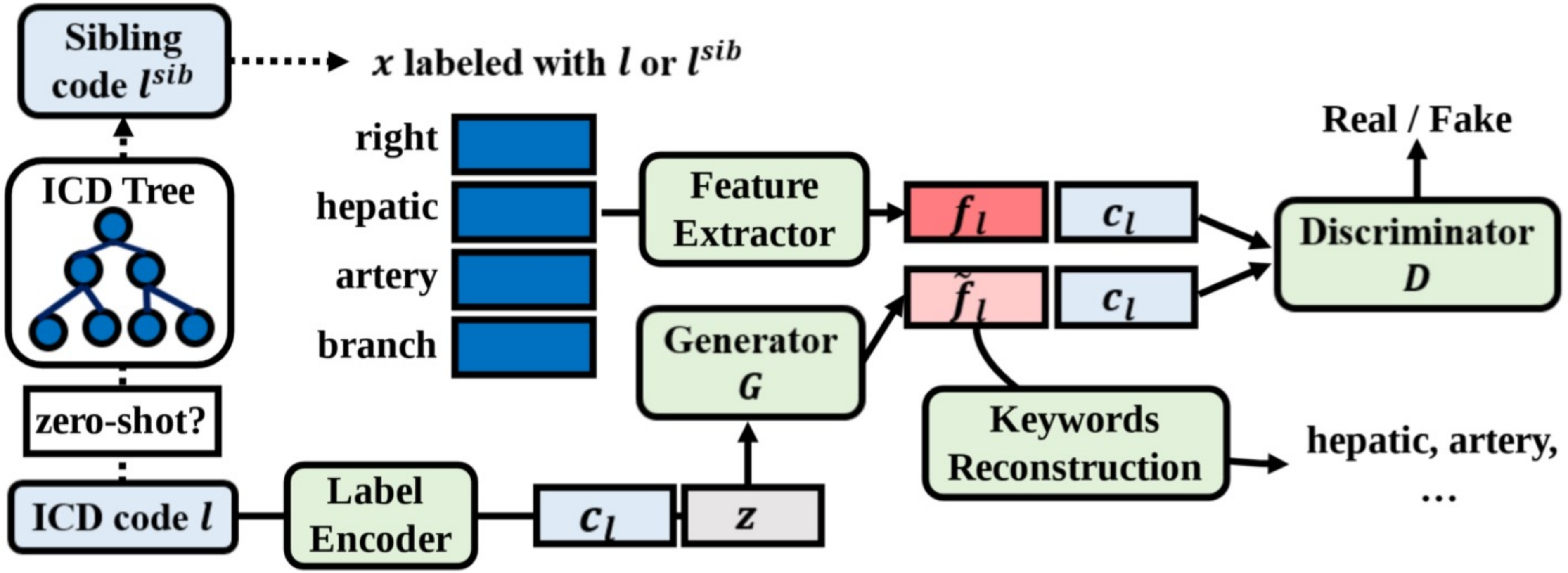
Zero-shot text classification via reinforced self-training [ACL 2020]



$$s_l = \text{softmax}(\tanh(H \cdot W_a^\top + b_a) \cdot v_l),\ a_l = s_l^\top \cdot H \qquad h_l^t = \frac{1}{|\mathcal{V}(l)|}\Sigma_{j \in \mathcal{V}(l)} g_j^{t-1},\ g_l^t = \text{GRU}(h_l^t, g_l^{t-1})$$

$$f_l = \text{rectifier}(W_o \cdot a_l + b_o), \quad \hat{y}_l = \sigma(g_l^\top \cdot f_l) \qquad \mathcal{L}_{\text{BCE}}(y, \hat{y}) = -\sum_{l=1}^{L}[y_l \log(\hat{y}_l) + (1 - y_l)\log(1 - \hat{y}_l)]$$

Generalized zero-shot text classification for ICD coding [IJCAI 2020]



Gaussian noise vector

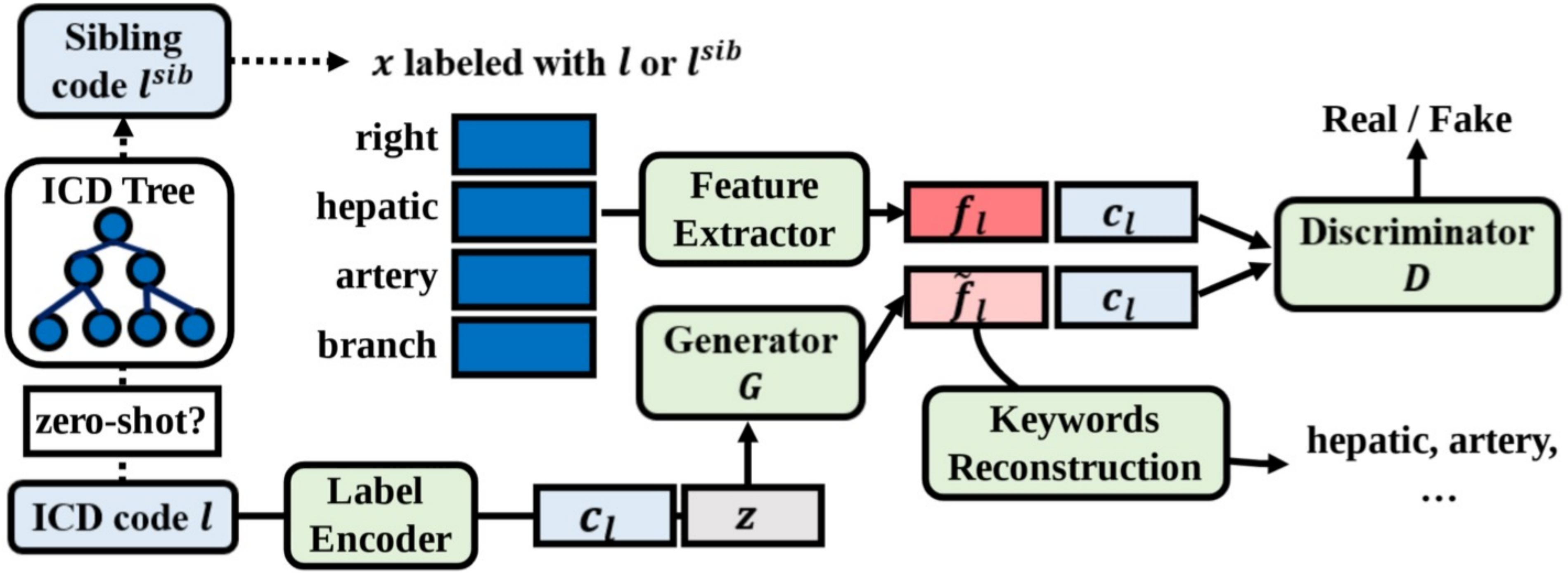
Generator $\longrightarrow$ $\tilde{f}_l$

Code description

$$\mathcal{L}_{\mathrm{WGAN}} = \mathbb{E}_{(f,c)\sim P_{\mathbb{S}}^{f,c}}[D(f,c))] - \mathbb{E}_{(\tilde{f},c)\sim P_{\mathbb{S}}^{\tilde{f},c}}[D(\tilde{f},c))] + \lambda \cdot \mathbb{E}_{(\hat{f},c)\sim P_{\mathbb{S}}^{\hat{f},c}}[(||\nabla D(\hat{f},c))||_2 - 1)^2] \quad (6$$

Generator **Discriminator**

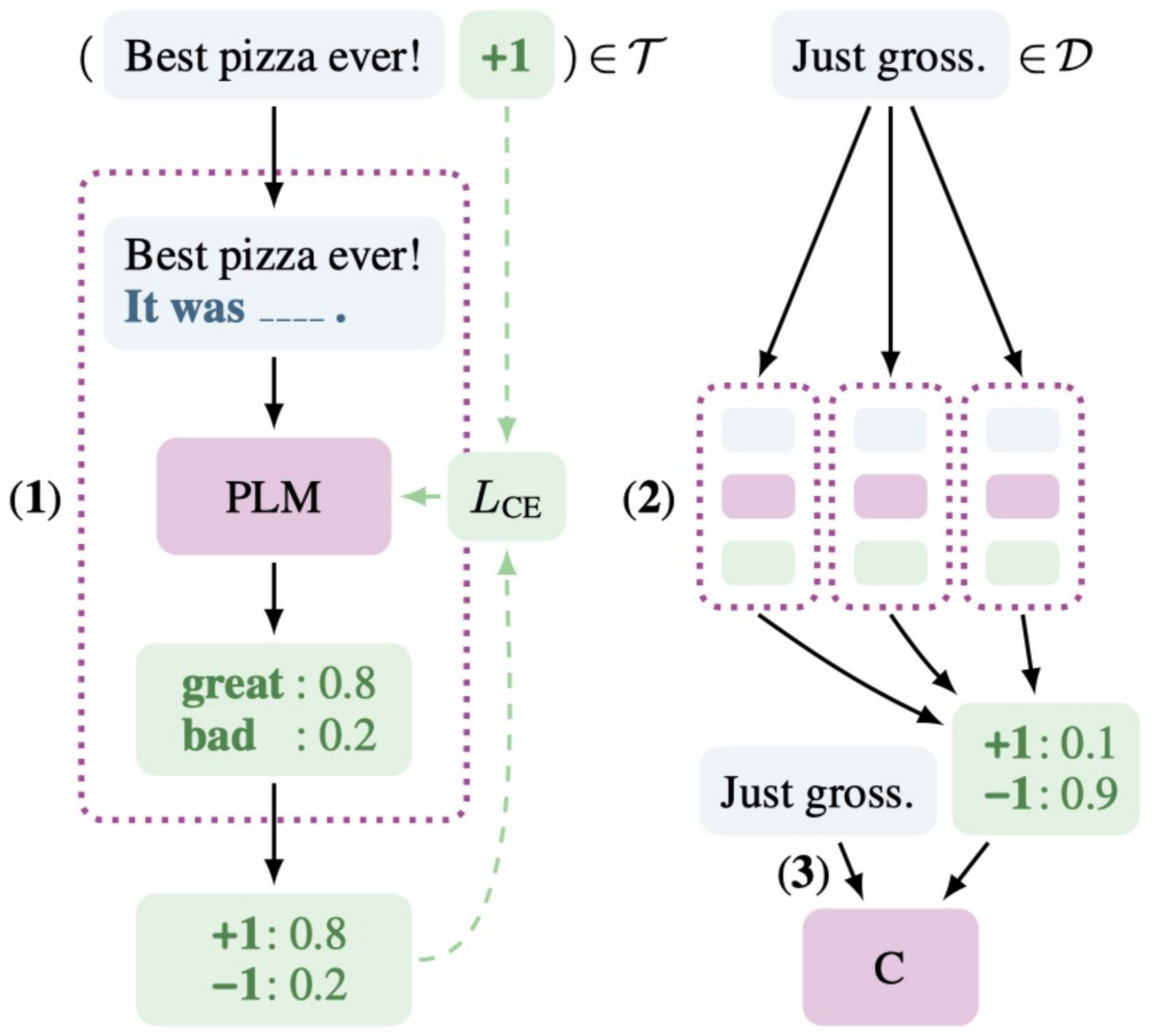
Generalized zero-shot text classification for ICD coding [IJCAI 2020]



$$\mathcal{L}_{\mathtt{KEY}} = -\log P(K_l|\tilde{f}_l) \approx -\sum_{w_k \in K_l} \pi(w_k, v_l) \cdot \log P(w_k|\tilde{f}_l)$$

Keyword Reconstruction

$$= -\sum_{w_k \in K_l} \pi(w_k, v_l) \cdot \log \frac{\exp(w_k^\top \cdot Q\tilde{f}_l)}{\sum_{w \in \mathcal{K}} \exp(w^\top \cdot Q\tilde{f}_l)} \qquad (7)$$



Exploiting cloze questions for few shot text classification and natural language inference [EACL 2021]

情感分类：

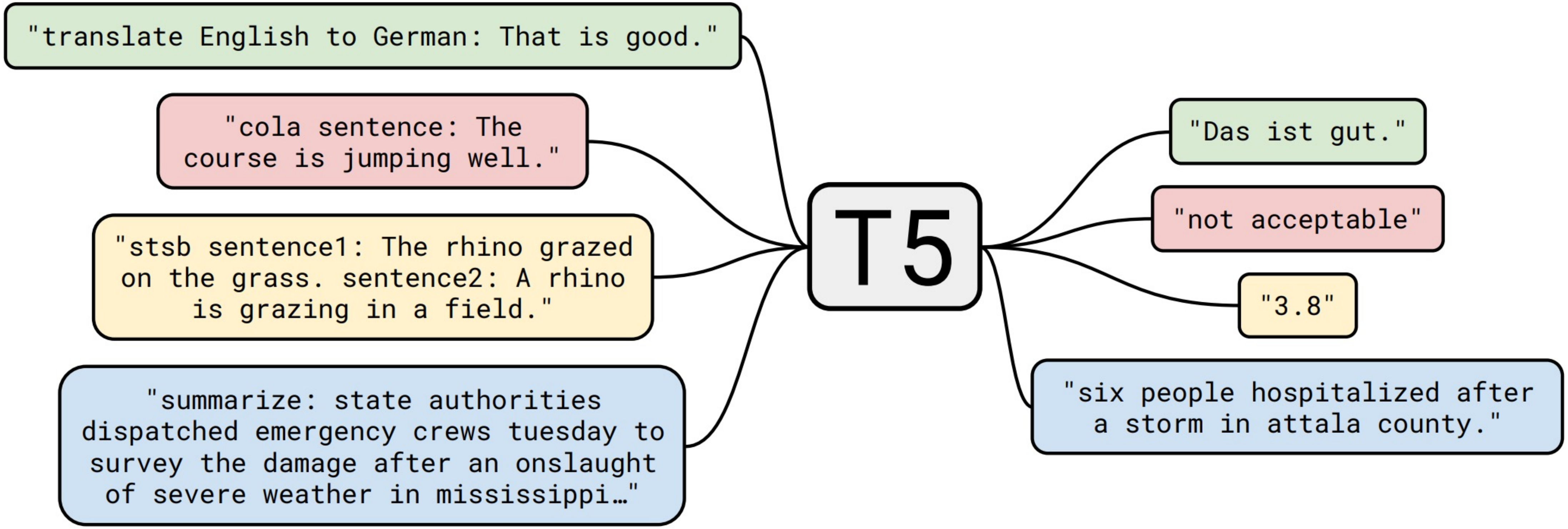
________满意。这趟北京之旅我感觉很不错。

{很, 挺, 太} → 正面　　{不, 难} → 负面

新闻主题分类：

下面报导一则________新闻。八个月了，终于又能在赛场上看到女排姑娘们了。

Exploiting cloze questions for few shot text classification and natural language inference [EACL 2021]



情感分类：
输入：识别该句子的情感倾向：这趟北京之旅我感觉很不错。
输出：正面

主题分类：
输入：下面是一则什么新闻？八个月了，终于又能在赛场上看到女排姑娘们了。
输出：体育

Exploiting cloze questions for few shot text classification and natural language inference [EACL 2021]

😉

Thank you ~