

ContrastNet: A Contrastive Learning Framework for Few-shot Text Classification

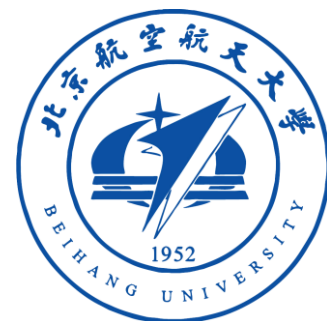
陈俊帆¹ 张日崇¹ 茆永轶² 徐洁³

¹北京航空航天大学

²渥太华大学

³利兹大学

2022 年 1 月 8 日



一、任务定义

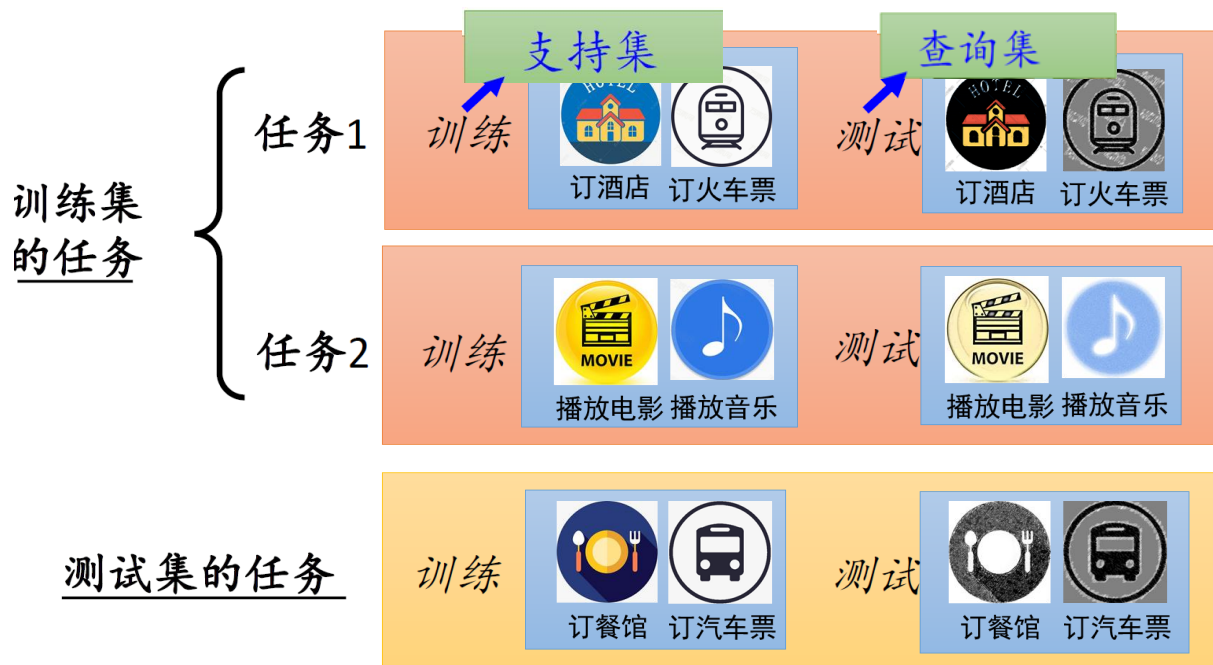
二、现存问题

三、解决方法

四、实验分析

● 少样本文本分类

- 元学习范式
- 训练集和测试集由小任务组成
- n-way k-shot
- 每个任务由支持集和查询集组成
- 训练和测试集文本类别集合不相交，分别称为源类别、目标类别



一、任务定义

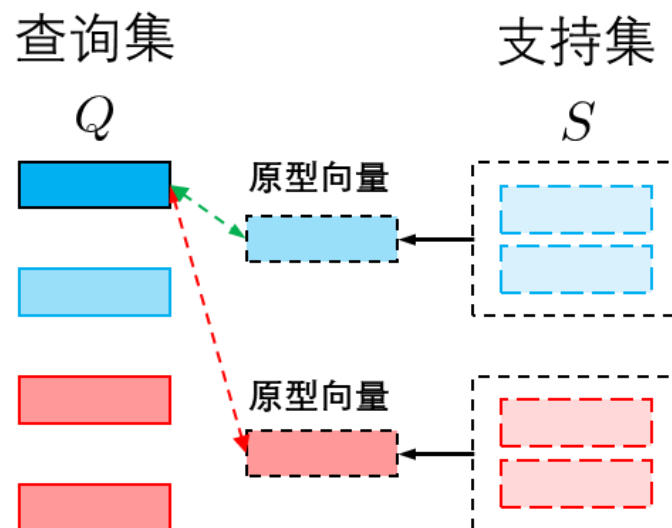
二、现存问题

三、解决方法

四、实验分析

● 表征区分度不够

- 原型网络没有直接对查询文本的关系建模，导致语义相似但类别不同的查询文本表征区分度不够，易出现分类错误



who covered the song one more cup of coffee

表征区分度不够

0.53 music-play
0.47 music-query



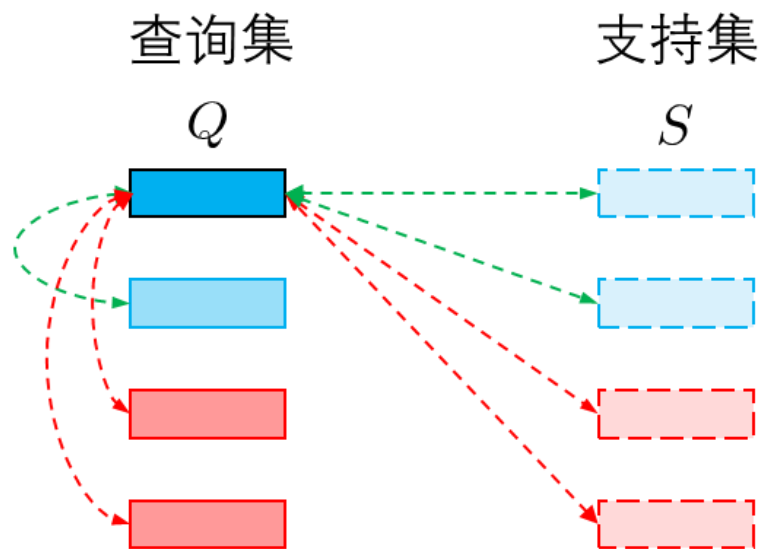
play the song one more cup of coffee

0.52 music-play
0.48 music-query



● 表征区分度不够

- 原型网络没有直接对查询文本的关系建模，导致语义相似但类别不同的查询文本表征区分度不够，易出现分类错误
- 解决方法：利用监督对比表征学习直接对查询文本的关系建模

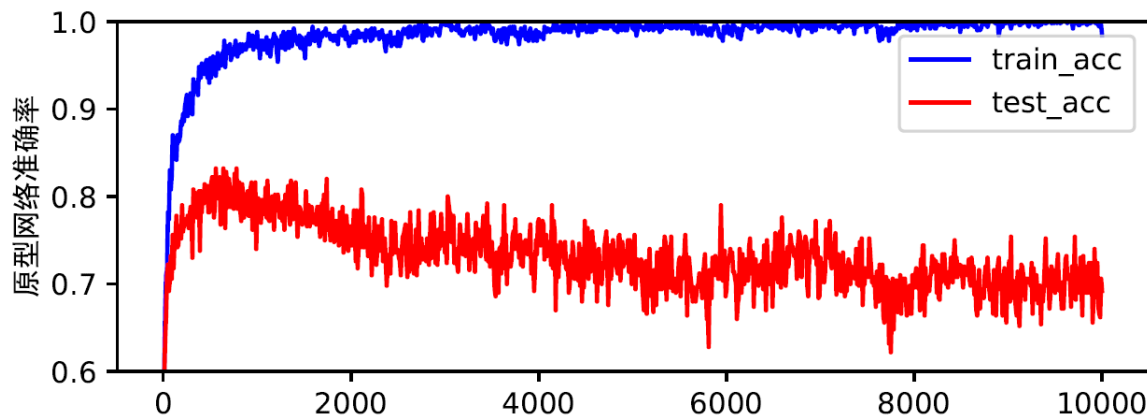


对比网络(ContrastNet)

提高表征区分度	who covered the song one more cup of coffee	0.43	music-play	✓
	play the song one more cup of coffee	0.57	music-query	✓
		0.58	music-play	✓
		0.42	music-query	

● 过拟合问题

- 少量训练样本形成有偏差的分布使模型容易过拟合到源类别
- 包括实例级过拟合和任务级过拟合(类别的组合)
- 解决方法: 引入任务级和实例级无监督对比损失作为正则项



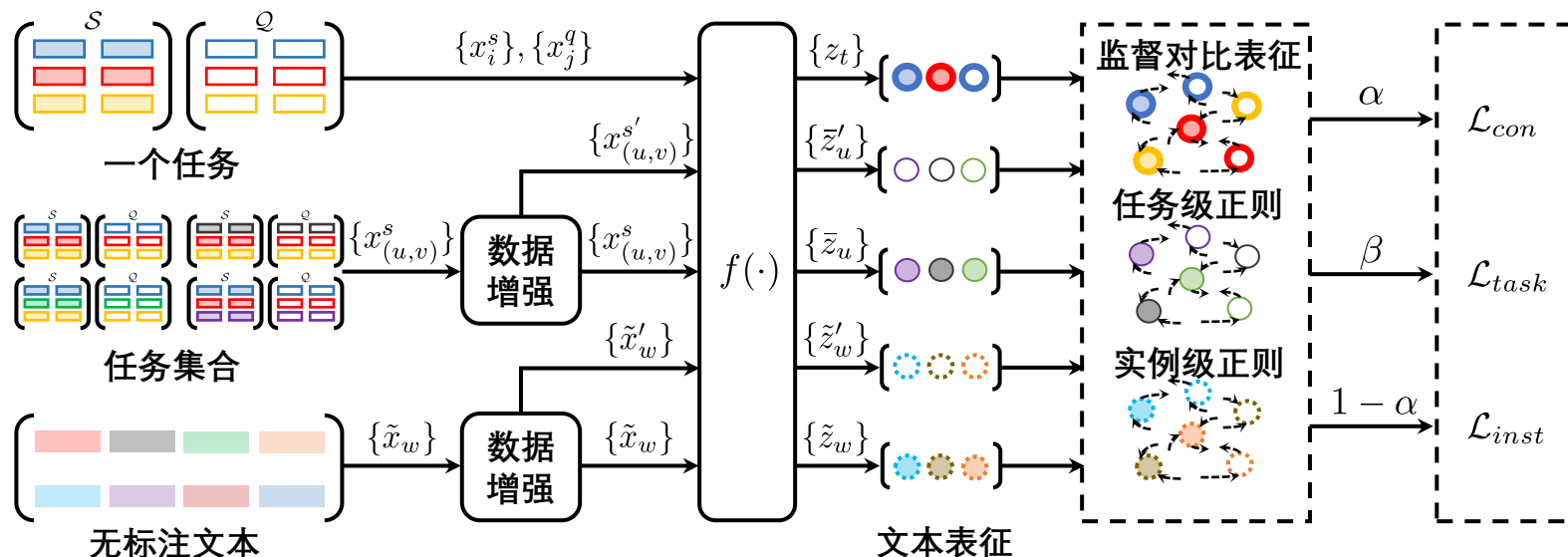
一、任务定义

二、现存问题

三、解决方法

四、实验分析

监督对比学习



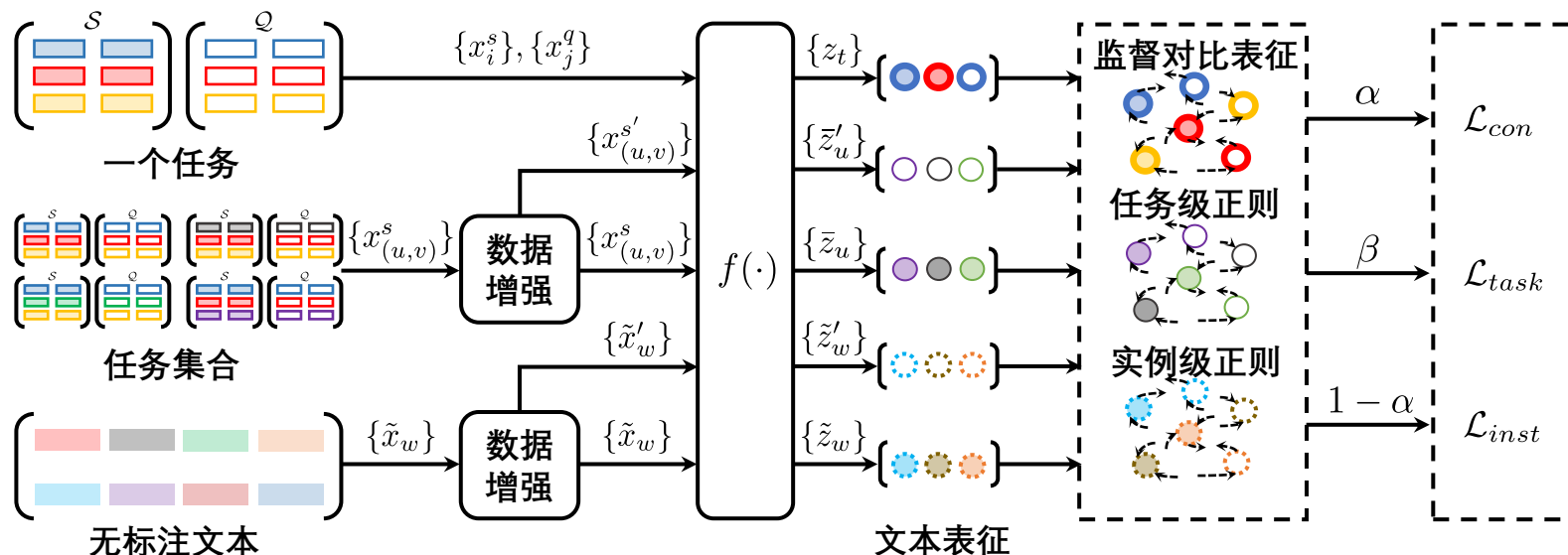
$$\mathcal{L} = \alpha \mathcal{L}_{con} + (1 - \alpha) \mathcal{L}_{inst} + \beta \mathcal{L}_{task}$$

$$\mathcal{L}_{con} = - \sum_{x_t \in \mathcal{B}} \frac{1}{c} \log \frac{\sum_{y_r=y_t} \exp(z_t \cdot z_r / \tau)}{\sum_{y_r=y_t} \exp(z_t \cdot z_r / \tau) + \sum_{y_{r'} \neq y_t} \exp(z_t \cdot z_{r'} / \tau)}$$

一个任务 $\frac{1}{c}$ $k+m-1$ 相同类别 正例 不同类别 负例

监督对比学习拉近同类别的语句表征，推远不同类别的语句表征，缓解相似语义不同意图类别的语句易预测错的问题

实例级对比正则



$$\mathcal{L} = \alpha \mathcal{L}_{con} + (1 - \alpha) \mathcal{L}_{inst} + \beta \mathcal{L}_{task}$$



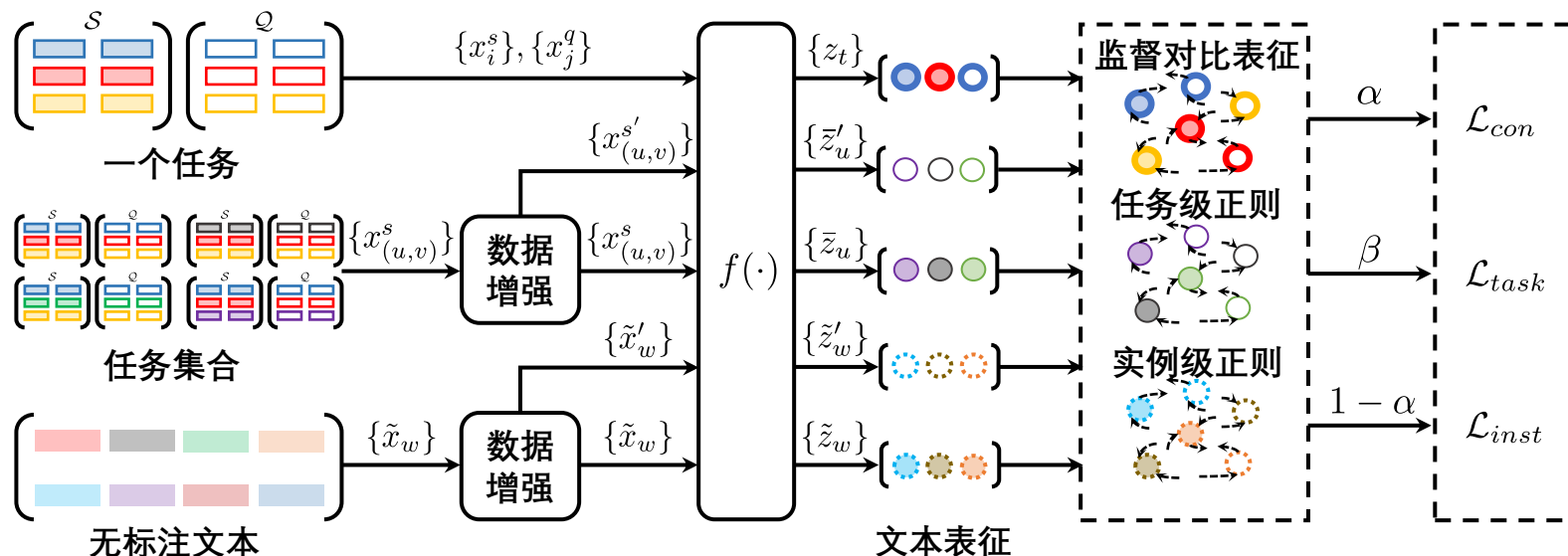
$$\mathcal{L}_{inst} = - \sum_{w=1}^{2N_{inst}} \log \frac{\exp(\tilde{z}_w \cdot \tilde{z}'_w / \tau)}{\exp(\tilde{z}_w \cdot \tilde{z}'_w / \tau) + \sum_{\tilde{z}_{w'} \neq \tilde{z}'_w} \exp(\tilde{z}_w \cdot \tilde{z}_{w'} / \tau)}$$

正例：数据增强样本

负例

实例级对比正则增加语句表征的区分度，鼓励目标类别语句表征与源类别语句表征彼此远离，减轻实例级过拟合问题

任务级对比正则



$$\mathcal{L} = \alpha \mathcal{L}_{con} + (1 - \alpha) \mathcal{L}_{inst} + \beta \mathcal{L}_{task}$$



$$\mathcal{L}_{task} = - \sum_{u=1}^{2N_{task}} \log \frac{\exp(\bar{z}_u \cdot \bar{z}'_u / \tau)}{\exp(\bar{z}_u \cdot \bar{z}'_u / \tau) + \sum_{\bar{z}_{u'} \neq \bar{z}'_u} \exp(\bar{z}_u \cdot \bar{z}_{u'} / \tau)}$$

正例：任务数据增强

负例

任务级对比正则增加任务表征的区分度，鼓励训练集任务表征和测试集任务表征彼此远离，减轻任务级过拟合问题

一、任务定义

二、现存问题

三、解决方法

四、实验分析

实验结果

模型	Banking77		HWU64		Liu		Clinic150		平均结果	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Prototypical Networks	86.28	93.94	77.09	89.02	82.76	91.37	96.05	98.61	85.55±2.20	93.24±1.22
PROTAUGMENT	86.94	94.50	82.35	91.68	84.42	92.62	94.85	98.41	87.14±1.36	94.30±0.60
PROTAUGMENT (bigram)	88.14	94.70	84.05	92.14	85.29	93.23	95.77	98.50	88.31±1.43	94.64±0.59
PROTAUGMENT (unigram)	89.56	94.71	84.34	92.55	86.11	93.70	96.49	98.74	89.13±1.13	94.92±0.57
ContrastNet ($\mathcal{L}_{task} \& \mathcal{L}_{inst} / o$)	88.53	95.22	84.62	91.93	80.53	93.47	94.29	98.09	86.99±1.57	94.68±0.74
ContrastNet (\mathcal{L}_{inst} / o)	89.75	95.36	85.14	91.69	86.79	93.28	96.32	98.25	89.50±1.30	94.65±0.64
ContrastNet	91.18	96.40	86.56	92.57	85.89	93.72	96.59	98.46	90.06±1.02	95.29±0.53

模型	HuffPost		Amazon		Reuters		20News		平均结果	
	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
MAML	35.9	49.3	39.6	47.1	54.6	62.9	33.8	43.7	40.9	50.8
Prototypical Networks	35.7	41.3	37.6	52.1	59.6	66.9	37.8	45.3	42.7	51.4
Induction Networks	38.7	49.1	34.9	41.3	59.4	67.9	28.7	33.3	40.4	47.9
HATT	41.1	56.3	49.1	66.0	43.2	56.2	44.2	55.0	44.4	58.4
DS-FSL	43.0	63.5	62.6	81.1	81.8	96.0	52.1	68.3	59.9	77.2
MLADA	45.0	64.9	68.4	86.0	82.3	96.7	59.6	77.8	63.9	81.4
ContrastNet ($\mathcal{L}_{task} \& \mathcal{L}_{inst} / o$)	52.74	63.59	74.70	84.47	83.74	93.28	70.61	80.04	70.45±3.28	80.35±3.32
ContrastNet (\mathcal{L}_{inst} / o)	52.85	64.88	75.33	84.21	85.10	93.65	70.35	80.19	70.91±3.00	80.73±2.79
ContrastNet	53.06	65.32	76.13	85.17	86.42	95.33	71.74	81.57	71.84±2.81	81.85±2.03

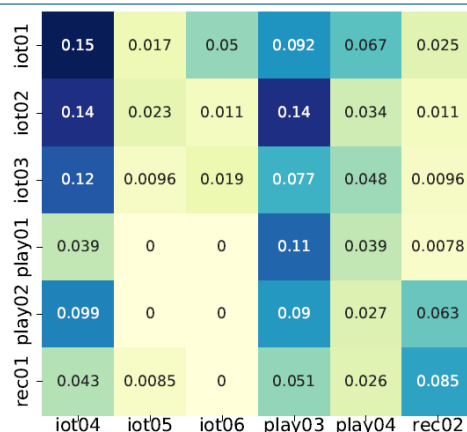
不使用无监督正则 模型 仅使用任务级正则

在8个少样本数据集共20项评估中，提出的方法在其中17项上超过SOTA结果，在1-shot下优势明显。消融实验结果证明了无监督正则的有效性。

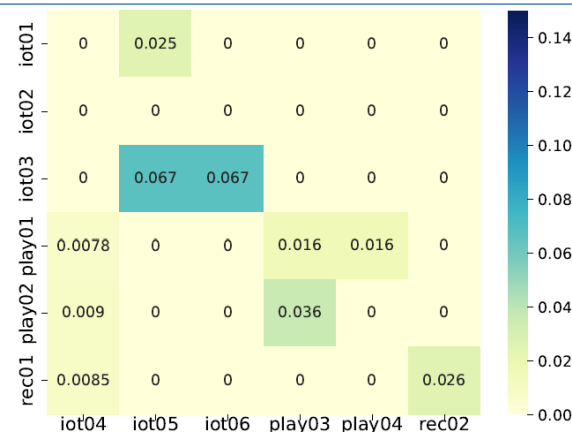
相似类别上的结果分析

● 错分率分析

- 一个类别的文本被错分为另一相似类别的概率
- 原型网络中，某些相似类别错分率高达15%
- 同比之下，对比网络相似类别错分率降至约0%



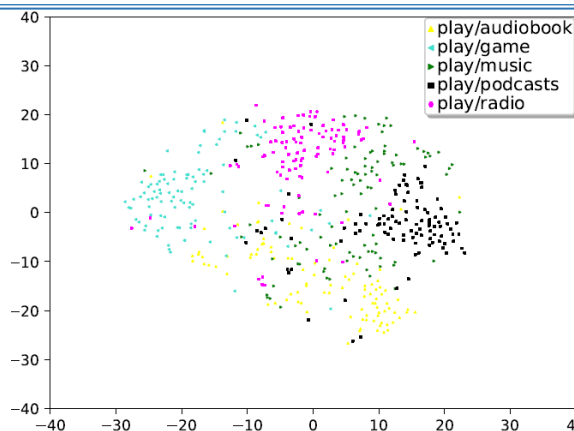
原型网络



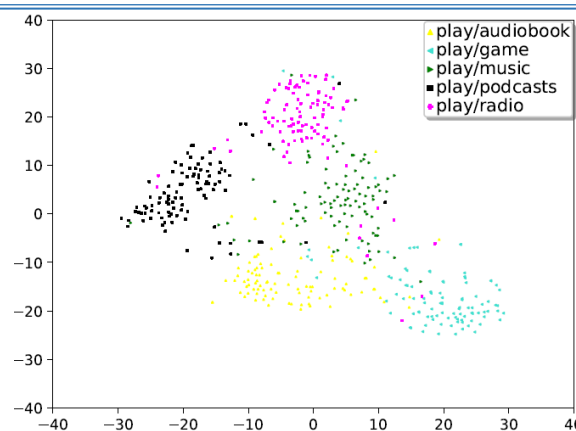
对比网络

● 查询集文本可视化

- 5个类别，100个小任务的查询集文本表征
- 相比于原型网络，对比网络不同类别的表征区分度更高



原型网络



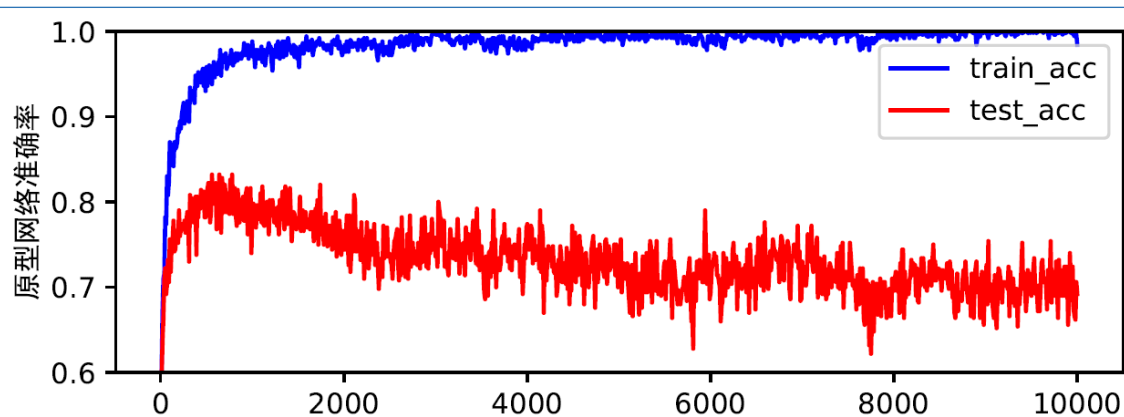
对比网络

原型网络中相似类别间的错分率较高，对比网络明显地降低了相似类别间的错分率。可视化结果表明对比学习提高了不同类别表征的区分度

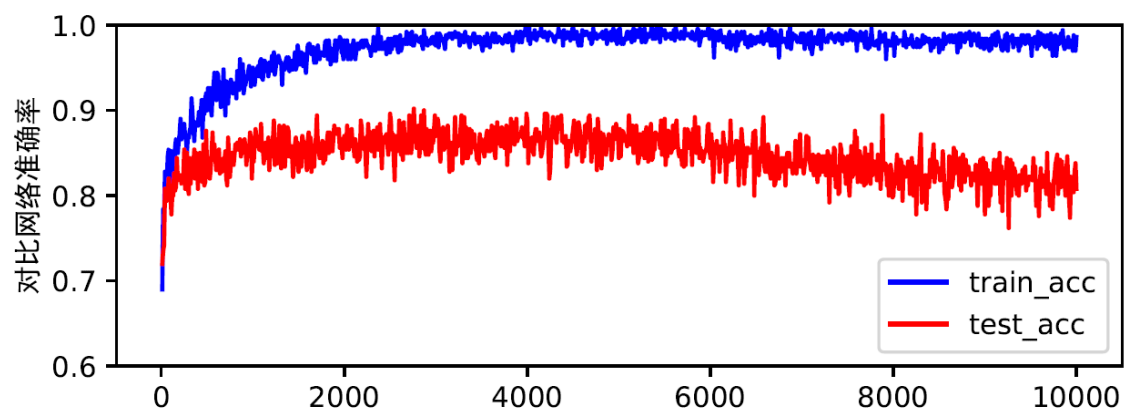
过拟合分析

●过拟合分析

- 原型网络训练和测试之间的**误差更大**，测试性能约600轮以后开始波动下降，**过拟合较严重**
- 对比网络训练和测试之间的**误差更小**，测试性能约4000轮以后开始波动下降，**过拟合程度相对原型网络更轻**



原型网络



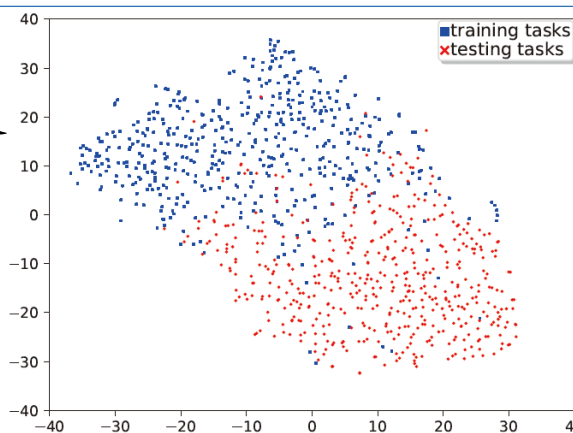
对比网络

训练过程中训练集和测试集的性能分析结果表明对比网络**减轻了**少样本文本分类中的**过拟合问题**

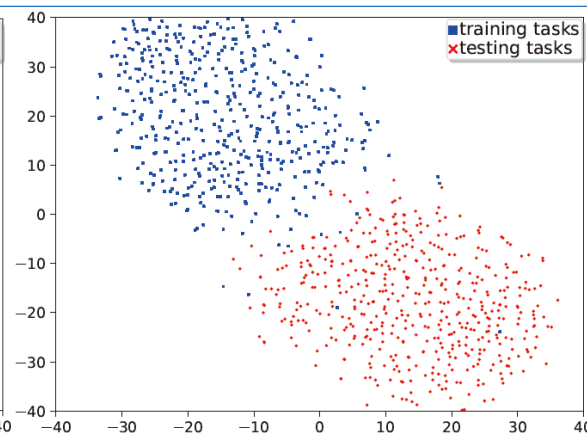
无监督对比正则可视化分析

● 任务表征可视化

- 训练、测试各200个任务
- 原型网络训练集和测试集任务表征部分重合
- 对比网络训练集和测试集任务表征明显分离
- 减轻了任务级过拟合



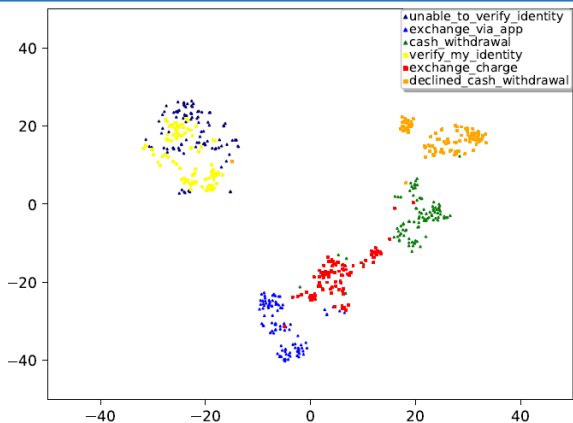
原型网络



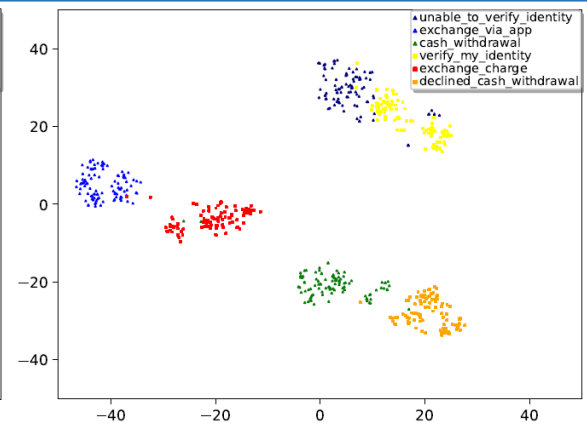
对比网络

● 实例表征可视化

- 训练和测试集分别选出3个类别可视化
- 对比网络比原型网络训练测试文本区分更开
- 减轻了实例级过拟合



原型网络



对比网络

可视化结果展示了对比网络下训练集和测试集在任务和实例表征上比原型网络下分离得更开，缓解了任务级和实例级过拟合问题

谢谢