

## 基于特征扰动的半监督专家发现方法

陈卓,张樊星<sup>†</sup>,杜军威,袁玺明  
(青岛科技大学信息科学技术学院,山东青岛 266061)

**摘要:**专家可为社区问答提供权威的答复,高效精准的专家发现有助于提升问答社区的服务质量.现有社区用户数据中存在噪声标签数据,且由于专家数量较少造成分类数据不平衡,从而降低了监督学习模型的专家发现精度.针对上述问题,本文提出一种基于特征扰动的半监督专家发现方法.该方法构建了一种无标签数据特征扰动策略,利用 Sharpening 算法实现无标签数据的伪标签化;基于 ADASYN 算法,通过构建专家用户邻近样本的方式扩充专家样本数据量,缓解分类数据的不平衡;构建联合损失函数,利用有标签和伪标签数据共同训练分类器,增强模型的泛化性能.实验结果表明,该方法在多个评价指标上优于已有模型和方法.

**关键词:**专家发现;社区问答;半监督学习;特征扰动

**中图分类号:**TP18 **文献标志码:**A

## Semi-supervised Expert Discovery Method Based on Feature Perturbation

CHEN Zhuo, ZHANG Fanxing<sup>†</sup>, DU Junwei, YUAN Ximing

(School of Information Science and Technology, Qingdao University  
of Science and Technology, Qingdao 266061, China)

**Abstract:** Experts can provide authoritative answers for community question answer (CQA). Efficient and accurate expert discovery can help improve the service quality of CQA. The expert discovery accuracy of the supervised learning model is reduced by the noise label data existing in the community user data as well as the unbalanced classification data due to the small number of experts. A semi-supervised expert discovery method based on feature perturbation is proposed to solve the mentioned problems. In this method, a feature perturbation strategy for unlabeled data is constructed, using the Sharpening algorithm to label the pseudo-label of unlabeled data. Based on the ADASYN algorithm, expert sample data is expanded by constructing neighbor samples of expert users to alleviate the imbalance of classification data. A joint loss function is constructed, which trains the classifier by both the labeled and pseudo-labeled data to enhance the generalization performance of the method. The experimental results show that this method is superior to the existing models and methods in several evaluation indexes.

**Key words:** expert discovery; community question answer; semi-supervised learning; feature perturbation

\* 收稿日期:2021-09-27

基金项目:国家自然科学基金资助项目(F030810, 6217072142), National Natural Science Foundation of China (F030810, 6217072142); 山东省重点研发计划资助项目(2018GGX101052), Shandong Provincial Key Research and Development Plan Project Funding (2018GGX101052); 山东省自然科学基金资助项目(ZR2021MF092), Natural Science Foundation of Shandong Province(ZR2021MF092)

作者简介:陈卓(1978—),女,山东青岛人,青岛科技大学副教授,博士

<sup>†</sup> 通信联系人, E-mail:1915515342@qq.com

问答社区模拟现实中的社区,使人们能够突破时间与空间的限制,通过网络聚集到一起,进行即刻交流和知识获取.专家在问答社区中发挥着重要作用,因专家可为社区问答提供权威的答复,专家认可的问题和回答也更可信、更具推广价值.

目前,问答社区普遍采用自动和人工两种方式进行权威专家的评定.在自动评定方式中,用户的权威性可根据用户所在用户组进行判定.用户组等级越高,其在问答社区中的权威性也越高,但用户需要依靠累计数据的评价制度才能成为权威专家.在人工评定方式中,用户可以向问答社区版主或负责人发送自我介绍以及参与的高质量问答信息,申请成为权威专家.该方式虽然不需要用户长期积累问答社区互动数据,但是人工审核不仅需要耗费大量的人力和时间成本,审核效率低下,而且该方式需要用户发起申请,社区缺乏主动发现潜在权威专家的能力.

因此,如何高效精准地主动发现问答社区用户中的权威专家,成为提升问答社区服务质量、提高用户参与社区互动积极性、保障社区持续发展的关键技术问题.目前,专家发现问题普遍采用的监督学习模型难以有效解决问答社区数据中存在的以下两个突出问题:社区用户数据集中存在噪声标签数据、社区用户数据集类别不平衡问题.

本文提出了一种基于特征扰动的半监督专家发现算法(Semi-supervised expert discovery method based on feature perturbation, SSEd),其贡献可以总结为以下两点:

1)针对社区用户数据集中存在噪声标签数据的问题,本文将疑似噪声数据归为无标签数据,并构建一种无标签数据的特征扰动策略,利用Sharpening算法实现无标签数据的伪标签化.

2)针对用户类别不平衡问题,本文基于ADASYN(adaptive synthetic sampling)算法,通过构建专家用户邻近样本的方式扩充专家样本数据量,缓解分类数据的不平衡;构建联合损失函数,利用有标签和伪标签数据共同训练分类器,增强模型的泛化性能.

## 1 相关工作

### 1.1 专家发现算法

专家发现的目的是找到社区内拥有强大专业知识与解决问题能力的权威专家,学者们根据不同的

数据特征,构建了相应的专家推荐模型.

从用户个人特征角度出发,张高明<sup>[1]</sup>综合考虑项目与专家的知识匹配度、项目间的相似性、用户选择偏好以及专家的历史表现等因素,融合了基于内容特征、潜在主题特征、协同过滤的方法对用户进行专家发现.Campos等<sup>[2]</sup>从政客的个人资料和演讲稿中获取政客的专业知识,然后将政治会议的主题与政客的专业知识进行相似度计算,为会议寻找特定领域的专家.史玉珍<sup>[3]</sup>将h指数应用到个人学术水平评价中,并结合文献特征使用层次聚类算法实现学术领域的专家发现.

从用户网络特征角度出发,龚凯乐<sup>[4]</sup>构建问题-用户的权威值传播网络,利用答案质量改进加权的HITS实现专家发现.薛凌云<sup>[5]</sup>融合用户的线下行为信息和线上的关系信息,通过位置特征将线上虚拟空间和线下现实社会进行连接,提出了基于评论的本地专家发现算法.

通过进一步分析,不难发现利用上述专家发现方法进行专家发现时,都对训练数据标注的准确性和数据规模有很高的要求,没有充分考虑由于专家人数少导致的分类数据不平衡以及训练集中存在噪声标签数据的问题.

### 1.2 半监督学习

半监督学习是机器学习的一个分支,用以解决在传统监督学习任务中,训练样本不足导致的模型性能退化问题,常被应用于物品分类问题和图片分类问题.半监督学习算法多数是对传统的机器学习算法进行改进<sup>[6]</sup>,将现有的监督算法扩展到半监督学习,即加入无标记样本.

自训练最早由Yarowsky<sup>[7]</sup>提出,应用于文本文档消除词义歧义,并由Triguero等<sup>[8]</sup>将其归类于最基本的伪标记方法.Kingma等<sup>[9]</sup>采用叠加的生成模型来学习标记样本和无标记样本,使用SVM对伪标记样本进行分类.Rasmus等<sup>[10]</sup>通过在每层解码器和编码器之间添加跳跃连接,在最高层添加分类器,将ladderNet变成一个半监督模型.Weston等<sup>[11]</sup>在网络输出层的目标函数和中间隐含层的目标函数中分别加入图的拉普拉斯正则化项,实现半监督的分类和特征学习.Tanha等<sup>[12]</sup>采用局部度量距离确定实例之间的置信度等级,提出一种决策树分类器的半监督自训练方法.毛铭泽等<sup>[13]</sup>从学习器的不同角度充分扩展模型的多样性,提出了一种基于权值多样性的半监督分类学习算法.

## 2 SSED 模型

问答社区中专家人数比例极少,存在较严重的分类数据不平衡性问题,而且社区中现有的用户级别信息无法准确评价用户权威性,例如部分近一两年内新注册的用户虽然级别不高,但积极参与社区

问答且问题回答质量高.有鉴于半监督学习在处理小样本标注信息<sup>[14]</sup>上的独特优势,本文提出SSED算法进行问答社区专家发现.

SSED的模型图如图1所示,图中包括了有标签数据的不平衡处理、无标签数据的轻微扰动操作、无标签数据的伪标签化以及损失计算四部分内容.

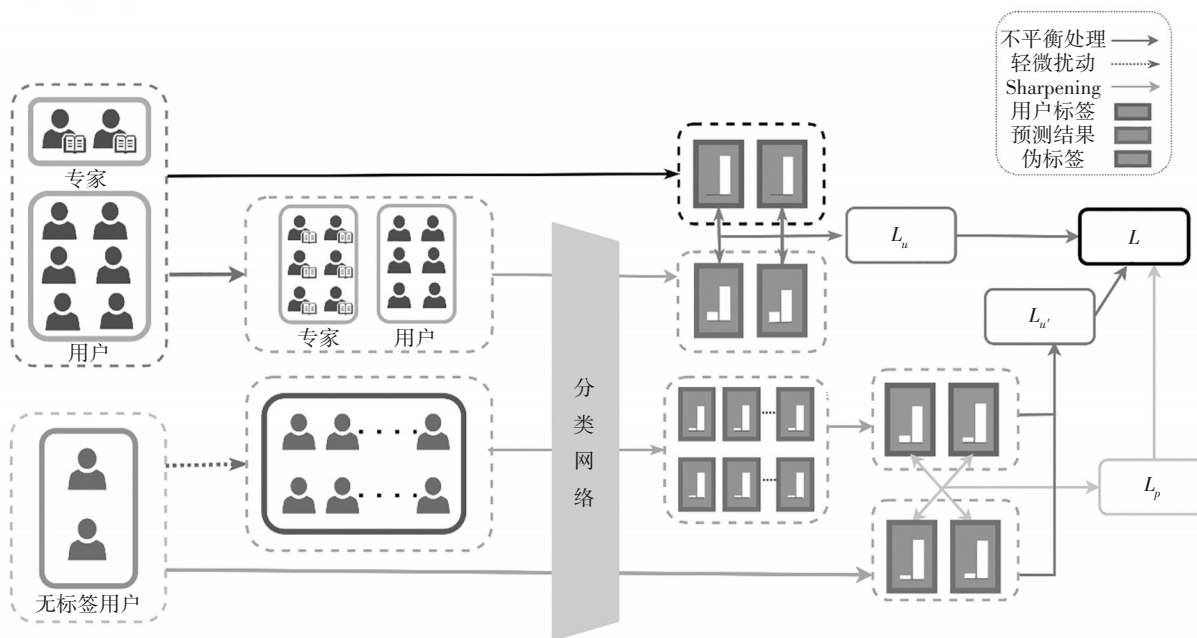


图 1 SSED 模型图

Fig.1 SSED model diagram

### 2.1 有标签数据的不平衡处理

本文实验数据中有标签的专家与普通用户比例为 1:26,不平衡的数据会严重干扰损失函数的优化,进一步影响到分类器的性能,因此在输入分类器之前,需要先降低数据的不平衡率.

ADASYN 算法由 He 等人<sup>[15]</sup>提出,其原理是与负样本的距离近的样本仍是负样本,该方法解决了随即向上抽样中简单的复制所带来的随机性和重复问题.本文首先遍历专家用户,寻找与专家用户特征相近的  $k$  个用户,在专家和邻近用户之间生成新样本,生成公式见式(1).

$$u_{new} = u_{expert} + \text{rand}(0, 1) \times (u_i - u_{expert}) \quad (1)$$

式中:  $u_{expert}$  表示专家用户,  $u_i (i=1, 2, \dots, k)$  表示与  $u_{expert}$  邻近的用户,  $k$  为超参数,  $\text{rand}(0, 1)$  表示  $(0, 1)$  之间的一个随机数.

将产生的人造专家不断加入原始有标签数据集,计算专家与普通用户的比例,直至专家与普通用户比例为 1:1.

### 2.2 无标签数据的轻微扰动

#### 2.2.1 一致性正则化

半监督学习在训练模型时,通过充分利用无标签数据来增强模型的泛化能力<sup>[16]</sup>,思路是对未标记的数据进行样本特征的扰动操作,将这些增强的数据重新输入分类器,同一样本的不同扰动数据的预测结果应该保持一致.模型损失函数的传统定义形式见式(2).

$$\text{Loss} = \| p_{\text{model}}(y_n | \text{Augment}_1(x_n); \theta) - p_{\text{model}}(y_n | \text{Augment}_2(x_n); \theta) \|_2^2 \quad (2)$$

式中:  $x_n$  表示无标签数据,  $\text{Augment}$  表示对  $x_n$  做随机增强产生的新数据,  $\theta$  表示模型参数,  $y_n$  表示模型的预测结果.

#### 2.2.2 轻微扰动

借鉴一致性正则化思想,本文对无标签用户进行轻微扰动,将轻微扰动后的数据输入分类器,其分类结果也应与未经扰动的无标签数据预测结果一致.因为 2.1 节使用的 ADASYN 算法的设计思想为:



在样本的附近人工生成的样本标签与原样本标签保持一致.因此,本文使用ADASYN算法对无标签用户数据进行轻微扰动,生成 $k$ 个无标签数据的轻微扰动样本并输入分类器,对于目标用户的 $k$ 个轻微扰动的输出进行平均化,计算公式见式(3).

$$p_{\text{avg}} = \frac{\sum_{i=1}^k c(u_i)}{k} \quad (3)$$

式中: $p_{\text{avg}}$ 表示平均化后的预测分布, $c$ 表示分类器, $k$ 为扰动生成的样本个数, $u_i$ 表示无标签样本轻微扰动数据.

### 2.3 无标签数据的伪标签化

在2.2节中,本文首先对无标签数据进行轻微扰动,将轻微扰动后的数据输入模型,再将得到的结果平均化.对于大部分输出结果应当是高置信度的,即专家的高密度区域和普通用户的高密度区域应当有明显边界.

半监督学习普遍基于一个假设,即分类器的分类边界不应该穿过任何一类的高密度区域.因此,本文对无标签数据轻微扰动后输出的结果进一步使用Sharpening算法<sup>[17]</sup>实现最小化分类.Sharpening算法对数据预测进行最小化熵的操作,使得输出结果的交叉熵更低,即在预测专家时,要么大概率是专家,要么大概率是普通用户,具体公式见式(4).

$$\text{Sharpen}(p, T)_i = \frac{p_i^{1/T}}{\sum_{j=1}^L p_j^{1/T}} \quad (4)$$

式中: $p$ 表示用户属于专家或普通用户的概率, $i$ 表示类别, $L$ 表示类别数, $T$ 为超参数. $T$ 用于调节分类熵,通过降低 $T$ ,促使模型产生低熵的预测, $T$ 越接近0,算法输出的结果越接近独热编码的分布,即社区用户要么是专家,要么是普通用户.

本文通过对分类器输出的结果进行锐化,将预测的置信度提高.若该无标签样本不满足锐化条件,则在本轮训练中不计算该样本.锐化判别公式见式(5).

$$\max(p_{\text{as}}) > T_c \quad (5)$$

式中: $p_{\text{as}}$ 表示对同一个用户轻微扰动 $k$ 次的样本预测平均值进行Sharpening后的分布. $\max(p_{\text{as}})$ 所属的用户类别作为该无标签用户的伪标签, $T_c$ 为手动设置的置信度阈值.

### 2.4 联合损失计算

在基于特征扰动的半监督专家发现方法中,模型的损失函数由三部分组成:监督损失 $L_u$ 、非监督损

失 $L_{u'}$ 和相似对损失 $L_p$ ,公式见式(6).

$$L = L_u + \alpha L_{u'} + \beta L_p \quad (6)$$

式中: $\alpha$ 和 $\beta$ 是 $L_{u'}$ 和 $L_p$ 的权重参数,接下来本文将分别介绍这三部分损失.

#### 2.4.1 有标签数据损失的定义

将有标签数据和扰动生成的无标签数据输入模型,计算预测结果与实际标签的交叉熵,利用该交叉熵表示有标签数据损失,计算公式见式(7).

$$L_u = \frac{1}{|u_1 + u_{\text{ADASYN}}|} \sum_{x, y \in u_1 + u_{\text{ADASYN}}} H[y, p_{\text{model}}(\tilde{y}|x; \theta)] \quad (7)$$

式中: $u_1$ 表示原始的有标签用户数据, $u_{\text{ADASYN}}$ 表示ADASYN算法生成的无标签用户数据, $\tilde{y}$ 表示分类器对于样本的预测结果, $H$ 表示交叉熵函数, $\theta$ 表示模型参数集合.

#### 2.4.2 无标签数据损失的定义

计算分类网络预测的轻微扰动样本与原始无标签样本之间的距离,利用该距离表示无标签数据损失,计算公式见式(8).

$$L_{u'} = \frac{\sum_{f, p \in u'} \mathbf{1}_{[\max(p_{\text{as}}) > T_c]} [p_{\text{as}} - p_{\text{model}}(\tilde{y}|x; \theta)]_2^2}{L|u'|} \quad (8)$$

式中: $u'$ 表示无标签用户数据; $p_{\text{as}}$ 表示对无标签数据进行伪标签化后的分布,若分布大于阈值 $T_c$ ,则将其看作样本的伪标签,进一步计算分类器对于原始无标签数据的分类结果 $\tilde{y}$ ;  $L_{u'}$ 则表示所有数据的伪标签与原始数据预测标签的L2损失的均值.

#### 2.4.3 相似对损失的定义

相似对损失的计算过程如图2所示,对于无标签数据,其原始数据输入分类器的预测结果不仅应该和该用户本身的伪标签一致,与相似伪标签用户的伪标签也应该保持一致.因此本文引入了一个新的损失项——相似对损失,它允许信息在不同的无标签用户之间隐式传播.在相似对损失中,本文使用无标签数据的一个高置信度的预测作为伪标签 $p$ .

对于用户 $u_1$ 和用户 $u_2$ ,若其伪标签 $p_1$ 和 $p_2$ 的相似度超过阈值 $T_{\text{sim}}$ ,那么分类器对于用户 $u_2$ 预测的概率向量分布与伪标签 $p_1$ 之间也应该保持一个较短的距离.相似对损失的公式见式(9).

$$L_p = \sum_{i, j \in |u'|, i \neq j}^H f_{\text{dist}}[p_i, p_{\text{model}}(\tilde{y}|x_j; \theta)] \quad (9)$$

式中: $i$ 和 $j$ 表示伪标签样本, $p_{\text{model}}(\tilde{y}|x_j; \theta)$ 表示模型对 $x_j$ 预测的概率向量, $H$ 表示所有伪标签的概率向量之间相似对的数量, $u'$ 表示无标签用户.

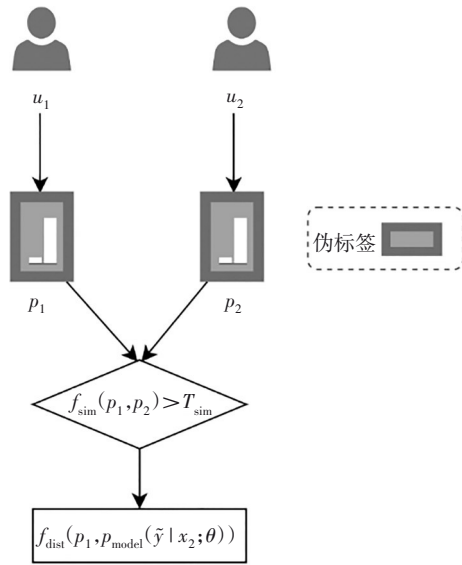


图2 计算相似对损失的流程

Fig.2 Process of calculating the loss of similarity pairs

本文采用巴氏距离计算两个伪标签的相似度,计算公式见式(10).

$$f_{sim}(p_1, p_2) = \sqrt{p_1}^T \sqrt{p_2} \quad (10)$$

两个概率向量的差异计算公式见式(11).

$$f_{dist}(p_1, p_2) = 1 - f_{sim}(p_1, p_2) \quad (11)$$

### 3 实验与分析

#### 3.1 实验数据集

为了验证系统的有效性和实用性,本文使用海川化工论坛的用户及问答信息构建实验数据集.在该论坛数据中选取的6 017名用户中,仅有148名专家用户,有3 884人作为普通用户,其余1 985人作为无标签数据,选取结果见表1.

表1 问答社区数据集

Tab.1 Dataset of the CQA

用户标签	训练集
专家	148
无标签用户	1 985
普通用户	3 884

本文将用户历史回答能力的向量、用户认可度表示和用户自编码后的个人特征拼接后作为用户特征,选用DeepFM作为分类器进行模型训练,并对其实验结果进行统计与分析.

#### 3.2 评价指标

为了能够评估模型的准确性和泛化性能,在每次评估模型效果时,本文对所有用户进行重新专家评定,采用  $Answer_{avg}$ 、 $Money_{avg}$  和  $Rank_{avg}$  作为模型性能的评价指标.

平均回答问题数计算公式见式(12).

$$Answer_{avg} = \frac{T}{P} \quad (12)$$

式中: $T$ 表示该组用户回答的总条数, $P$ 表示该组用户的总人数.

平均回答单条问题所获财富值数的计算公式见式(13).

$$Money_{avg} = \frac{\sum_{i=1}^T m_i}{T} \quad (13)$$

式中: $m_i$ 表示用户的回答*i*所获得的财富值.

用户平均回答单条问题所获财富值排名的计算公式见式(14).

$$Rank_{avg} = \frac{\sum_{i=1}^T r_i}{T} \quad (14)$$

式中: $r_i$ 表示用户的回答*i*所获得的财富值在所属问题中的排名, $ct_i$ 表示*i*所属问题下回答的总数.

#### 3.3 参数设定

SSED模型的超参数有轻微扰动的参数*k*、伪标签化的参数 $T_c$ 、损失函数中的 $\alpha$ 和 $\beta$ ,共四个.本文通过实验对比了问答社区数据集中,不同参数*k*、 $T_c$ 以及 $\alpha$ 、 $\beta$ 取值时,SSED模型的专家发现效果,实验结果见表2至表4.

表2 不同轻微扰动参数*k*的专家发现效果对比

Tab.2 Comparison of expert discovery of different slight perturbation parameters *k*

<i>k</i>	专家数	$Answer_{avg}$	$Money_{avg}$	$Rank_{avg}$
2	425	31.6	0.81	0.50
3	410	33.9	0.84	0.49
5	423	33.5	0.82	0.50

表3 不同伪标签化参数 $T_c$ 的专家发现效果对比

Tab.3 Comparison of expert discovery of different pseudo-labeled parameters  $T_c$

$T_c$	专家数	$Answer_{avg}$	$Money_{avg}$	$Rank_{avg}$
0.75	427	33.4	0.83	0.49
0.80	410	33.9	0.84	0.49
0.85	405	32.7	0.81	0.50

表4 不同损失函数参数 $\alpha$ 和 $\beta$ 的专家发现效果对比Tab. 4 Comparison of expert discovery of different loss function parameters  $\alpha$  and  $\beta$ 

损失参数	专家数	Answer <sub>avg</sub>	Money <sub>avg</sub>	Rank <sub>avg</sub>
$\alpha=75, \beta=100$	411	29.9	0.80	0.50
$\alpha=100, \beta=120$	410	33.9	<b>0.84</b>	<b>0.49</b>
$\alpha=100, \beta=150$	401	30.7	0.82	0.50

由表2至表4可知,当轻微扰动的参数 $k$ 取3,伪标签化的参数 $T_c$ 取0.80,损失函数中 $\alpha$ 取100、 $\beta$ 取120时,专家发现效果最好.因此,在下文的消融和对比实验中,选取了本节中各最优参数值作为各超参数的取值.

### 3.4 消融实验

本文选取了2.4节中的三个损失,由于问答社区即使不使用无标签数据,也依然可以完成专家推荐.因此,当只输入损失 $L_u$ 时,分类器模型为监督学习.本文通过输入不同损失计算模型的预测效果,实验结果见表5.

表5 不同损失的专家发现效果对比

Tab. 5 Comparison of expert discovery between loss

损失项	专家数	Answer <sub>avg</sub>	Money <sub>avg</sub>	Rank <sub>avg</sub>
$L_u$	430	29.5	0.77	0.51
$L_u + \alpha L_{u'}$	414	30.4	0.80	0.50
$L_u + \alpha L_{u'} + \beta L_p$	410	<b>33.9</b>	<b>0.84</b>	<b>0.49</b>

通过分析表5可知,监督学习模型由于数据较少,发现的专家回答问题的质量不如半监督学习的效果出色.这证明了对于问答社区数据使用半监督算法的有效性,同时证明了加入本文提出的相似对损失能够进一步提升分类器的性能.

### 3.5 对比实验

为了评估SSED算法的性能,本文选择以下三种半监督模型作为基线方法进行对比实验.

S4VM<sup>[18]</sup>模型:该模型试图找到能够正确划分有标记样本且穿过特征空间中密度最低的区域的一个超平面.

SSWL<sup>[19]</sup>模型:通过考虑实例相似度和标签相似度来弥补缺失的标签.

UDEED<sup>[20]</sup>模型:通过对未标记的数据增加多样性来促进集成学习对于有监督数据的训练结果.

SSED与基线方法在数据集上,按照9:1划分训练集和测试集.关于模型对比实验结果如表6所示.

表6 不同模型之间的专家发现效果对比

Tab.6 Comparison of expert findings between different models

模型	专家数	Answer <sub>avg</sub>	Money <sub>avg</sub>	Rank <sub>avg</sub>
UDEED	399	31.9	0.81	<b>0.49</b>
S4VM	425	31.1	0.79	0.50
SSWL	414	29.5	0.79	0.52
SSED	410	<b>33.9</b>	<b>0.84</b>	<b>0.49</b>

表6中显示了在三个评价指标下,SSED算法在社区论坛数据上的表现优于其他三种半监督基线模型,验证了本文方法的有效性和可行性.

## 4 结论

专家能为问答社区提供高质量回复,提高问题解决效率,因此准确高效的专家发现是问答社区持续发展所面临的一个重要问题.已有专家发现方法通常采用监督学习模型,难以有效处理社区用户数据集中存在的噪声标签数据以及分类数据的不平衡问题.

本文提出SSED模型,利用过采样技术平衡有标签数据样本,采用无标签数据特征扰动策略进行数据增强,通过有标签和伪标签数据共同训练分类器,增强模型的泛化性能.实验结果表明,SSED模型在多个评价指标上优于已有模型和方法.

由于ADASYN算法在生成人造样本和扰动数据时存在一定误差,在计算无标签数据轻微扰动时容易发生误差累积的问题.后续可以基于用户短期的问答行为进一步改进专家样本生成方法,使其在保持专家样本主要特征的基础上,具备更加丰富的个性化特征,进一步避免专家样本的单一化,减少累积误差.

## 参考文献

- [1] 张高明,张善从.一种改进的组合策略评审专家推荐算法[J].科技管理研究,2021,41(1):175-180.  
ZHANG G M, ZHANG S C. An improved expert recommendation algorithm for combined strategy evaluation[J]. Science and Technology Management Research, 2021, 41(1): 175-180. (In Chinese)
- [2] CAMPOS L M, FERNÁNDEZ-LUNA J M, HUETE J F, et al. LDA-based term profiles for expert finding in a political setting [J]. Journal of Intelligent Information Systems, 2021, 56(3): 529-559.

- [3] 史玉珍,彭智勇. 基于修正 h 指数的学科领域专家发现的研究[J]. 计算机工程与应用,2011,47(29):1-3.  
SHI Y Z, PENG Z Y. Research on experts community discovery based on modified h index[J]. Computer Engineering and Applications, 2011, 47(29):1-3. (In Chinese)
- [4] 龚凯乐,成颖. 基于“问题-用户”的网络问答社区专家发现方法研究[J]. 图书情报工作,2016,60(24):115-121.  
GONG K L, CHENG Y. Finding experts in community question answering based on “question-user”[J]. Library and Information Service, 2016, 60(24):115-121. (In Chinese)
- [5] 薛凌云. LBSN 中基于评论的本地专家发现[D]. 南京:东南大学,2017.  
XUE L Y. Review based local experts discovery in LBSN [D]. Nanjing: Southeast University, 2017. (In Chinese)
- [6] CASTELLI V, COVER T M. The relative value of labeled and unlabeled samples in pattern recognition with an unknown mixing parameter[J]. IEEE Transactions on Information Theory, 1996, 42(6):2102-2117.
- [7] YAROWSKY D. Unsupervised word sense disambiguation rivaling supervised methods [C]//33th Annual Meeting on Association for Computational Linguistics. 1995:189-196.
- [8] TRIGUERO I, GARCÍA S, HERRERA F. Self-labeled techniques for semi-supervised learning: taxonomy, software and empirical study [J]. Knowledge and Information Systems, 2015, 42(2):245-284.
- [9] KINGMA D P, REZENDE D J, MOHAMED S, *et al.* Semi-supervised learning with deep generative models [C]//Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2.2014:3581-3589.
- [10] RASMUS A, VALPOLA H, HONKALA M, *et al.* Semi-supervised networks [C]//Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2. Berlin: Springer, 2015: 3546-3554.
- [11] WESTON J, RATLE F, MOBAHI H, *et al.* Deep learning via semi-supervised Embedding [C]//Neural Networks: Tricks of the Trade. 2012:639-655.
- [12] TANHA J, SOMEREN M, AFSARMANESH H. Semi-supervised self-training for decision tree classifiers [J]. International Journal of Machine Learning and Cybernetics, 2017, 8(1):355-370.
- [13] 毛铭泽,曹芮浩,闫春钢. 基于权值多样性的半监督分类算法[J]. 计算机应用,2021,41(9):2473-2480.  
MAO M Z, CAO R H, YAN C G. Semi-supervised classification algorithm based on weight diversity [J]. Journal of Computer Applications, 2021, 41(9):2473-2480. (In Chinese)
- [14] BERTHELOT D, CARLINI N, CUBUK E D, *et al.* ReMixMatch: semi-supervised learning with distribution alignment and augmentation anchoring [EB/OL]. 2019: arXiv: 1911.09785 [cs. LG]. <https://arxiv.org/abs/1911.09785>
- [15] HE H B, BAI Y, GARCIA E A, *et al.* ADASYN: Adaptive synthetic sampling approach for imbalanced learning [C]//2008 IEEE International Joint Conference on Neural Networks. Hong Kong, China: IEEE, 2008:1322-1328.
- [16] LAINE S, AILA T. Temporal ensembling for semi-supervised learning [C]//International Conference on Learning Representations. New Orleans: ICLR, 2017: 1-12.
- [17] GOODFELLOW I, BENGIO Y, COURVILLE A, *et al.* Deep learning [M]. Cambridge: MIT press, 2016: 58-59.
- [18] LI Y F, ZHOU Z H. Towards making unlabeled data never hurt [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(1):175-188.
- [19] DONG H C, LI Y F, ZHOU Z H. Learning from semi-supervised weak-label data [C]//Proceedings of the AAAI Conference on Artificial Intelligence. Menlo Park: AAAI, 2018, 32(1):2926-2933.
- [20] ZHANG M L, ZHOU Z H. Exploiting unlabeled data to enhance ensemble diversity [J]. Data Mining and Knowledge Discovery, 2013, 26(1):98-129.