

北京大学学报(自然科学版)  
Acta Scientiarum Naturalium Universitatis Pekinensis  
doi: 10.13209/j.0479-8023.2018.063

# 结合表示学习和迁移学习的跨领域情感分类

廖祥文<sup>1,2,3,†</sup> 吴晓静<sup>1,2</sup> 桂林<sup>1</sup> 黄锦辉<sup>4</sup> 陈国龙<sup>1,2</sup>

1. 福州大学数学与计算机科学学院, 福州 350116; 2. 福建省网络计算与智能信息处理重点实验室(福州大学), 福州 350116;  
3. 福建省信息处理与智能控制重点实验室(闽江学院), 福州 350116; 4. 香港中文大学系统工程与工程管理学系, 香港;

† 通信作者, E-mail: liaoxw@fzu.edu.cn

**摘要** 针对现有跨领域情感分类方法中文本表示特征忽略了重要单词与句子的情感信息, 且在迁移过程中存在负面迁移的问题, 提出一种将文本表示学习与迁移学习算法相结合的跨领域情感分类方法。该方法首先利用低维稠密的词向量对文本进行初始化, 通过分层注意力网络对文本中重要单词与句子的情感信息进行建模, 从而学习源领域与目标领域的文档级分布式表示。随后采用类噪声估计方法, 对源领域中的迁移数据进行检测, 剔除负面迁移样例, 挑选高质量样例以扩充目标领域的训练集。最后训练支持向量机对目标领域文本进行情感分类。在大规模公开数据集上进行的两个实验结果表明, 与基准方法相比, 该方法在均方根误差上分别降低了1.5%和1.0%, 说明该方法可以有效提高跨领域情感分类性能。

**关键词** 文本表示学习; 迁移学习; 类噪声估计; 跨领域; 情感分类

## Cross-Domain Sentiment Classification Based on Representation Learning and Transfer Learning

LIAO Xiangwen<sup>1,2,3,†</sup>, WU Xiaojing<sup>1,2</sup>, GUI Lin<sup>1</sup>, HUANG Jinhui<sup>4</sup>, CHEN Guolong<sup>1,2</sup>

1. School of Mathematics and Computer Science, Fuzhou University, Fuzhou 350116; 2. Fujian Provincial Key Laboratory of Networking Computing and Intelligent Information Processing (Fuzhou University), Fuzhou, 350116; 3. Fujian Provincial Key Laboratory of Information Processing and Intelligent Control (Minjiang University), Fuzhou, 350116; 4. Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong; † Corresponding author, E-mail: liaoxw@fzu.edu.cn

**Abstract** Most of existing cross-domain sentiment classification methods are not expressive enough to capture rich representation of text, and class noise accumulated during transfer process can lead to negative transfer which can adversely affect performance. To address these issue, the authors propose a method combining textual representation learning and transfer learning algorithm for cross-domain sentiment classification. This method first builds a hierarchical attention network to generate document representations with local semantic information. Afterwards, the authors utilize the class noise estimation algorithm to detect the negative transfer samples in transferred samples and remove them. Finally, the sentiment classifier is trained on the expanded dataset from samples in target domain and transferred ones in source domain. Compared with the baselines, two experiments on large-scale product review datasets show that the proposed method is able to effectively reduce RMSE of cross-domain sentiment classification by 1.5% and 1.0% respectively.

国家自然科学基金(61772135, U1605251)、中国科学院网络数据科学与技术重点实验室开放基金课题(CASNDST201708, CASNDST201606)、可信分布式计算与服务教育部重点实验室主任基金(2017KF01)和赛尔网络下一代互联网技术创新项目(NGII20160501)资助

收稿日期: 2018-04-15; 修回日期: 2018-08-18; 网络出版时间: 2018-08-22 15:46:57

**Key words** textual representation learning; transfer learning; class-noise estimation; cross-domain; sentiment classification

随着 Amazon、Yelp 和 IMDB 等在线评论网站的迅速发展,大量用户对不同领域的产品进行评论、表达观点。文本情感分析(sentiment analysis)可以利用这些大规模用户生成文本进行情感分析,发现用户对事件或产品的态度与倾向性<sup>[1]</sup>。跨领域情感分类(cross-domain sentiment classification)作为情感分析和自然语言处理(natural language processing, NLP)的重要部分<sup>[2]</sup>,旨在利用源领域与目标领域的有用知识,对目标领域的文本进行情感分析,挖掘用户的观点与态度,这对消费者意见反馈<sup>[3]</sup>和社交媒体研究<sup>[4]</sup>等有重要意义。

在跨领域情感分类研究中,们通常利用迁移学习的方法,从有大量标注数据的源领域中获取有用的知识,迁移到标注数据缺乏的相关目标领域中,以提高算法在目标领域上的性能。这些方法主要分为两大类<sup>[5]</sup>。

1)基于特征表达的跨领域情感分类。根据不同领域间的通用潜在特征来寻找新的特征空间,主要采用结构对应学习<sup>[6]</sup>(structural correspondence learning, SCL)、谱特征对齐算法<sup>[7]</sup>(spectral features alignment, SFA)、迁移成分分析<sup>[8]</sup>(transfer component analysis, TCA)或高斯迁移过程<sup>[9]</sup>等方法,来减少领域间的特征分布差异。这些方法大多数使用人工设计的规则或 N 元文法提取句子特征,往往忽略了上下文关系及重要单词与句子的情感信息,无法很好地对跨领域大规模文档进行表示。

2)基于实例加权的跨领域情感分类。其原理是调整源领域的训练数据权重,使其尽可能符合目标领域分布,例如采用迁移学习适应提升算法<sup>[10-11]</sup>(boosting for transfer learning, TrAdaBoost)、核均值匹配<sup>[12]</sup>(kernel mean matching, KMM)和直接重要性估计<sup>[13]</sup>等方法,学习实例权重参数。另外,有些研究通过主动学习<sup>[14-15]</sup>对迁移数据进行有选择地标注,从而进行跨领域情感分类。但是,在迁移过程中不可避免出现噪声数据,即存在负面迁移的问题,反而会降低跨领域情感分类的性能。并且,对于大规模标注文本,在深度学习方法<sup>[16-17]</sup>上已经取得高性能,是否有必要进行迁移学习值得探讨。

本文提出一种将文本表示方法与迁移学习算法相结合的跨领域文本情感分析方法。该方法首

先利用训练好的词向量和分层注意力网络<sup>[17]</sup>(hierarchical attention network),对源领域与目标领域的文档进行特征提取,从而获得文档级的分布式表示。接着,通过类噪声估计方法从源领域的文本样本中筛选出与目标领域分布相符的高质量样例,迁移这些样例以扩充目标领域的数据集,进一步利用支持向量机<sup>[18]</sup>(support vector machine, SVM)训练情感分类器。在目前公开的大规模产品评价数据集上进行的实验表明本文提出的方法性能优于目前主流的情感分类方法,表明本文提出的思路可以有效提高迁移样本的质量,从而提高跨领域情感分类的性能。

## 1 结合文本表示与迁移学习的跨领域情感分类

为了方便研究,本文的跨领域情感分类问题可描述为:给定目标领域的训练数据集  $D^t$  和源领域的训练数据集  $D^s$ ,  $Y = \{c_i\}$  为情感类别 ( $i = 1, 2, \dots, n_c$ ),  $n_c$  为情感类别数,且源领域与目标领域的数据分布不一致。利用  $T = \{D \times Y\}$  找到一个最优映射  $F$  使得  $F: S \rightarrow Y$ , 其中  $D = D^t \cup D^s$ ,  $S$  为目标领域的测试数据。

本文的跨领域情感分类方法框架可分为三个阶段,如图 1 所示。

1)基于分层注意力网络的跨领域情感文本表示学习。首先,对目标领域训练数据和源领域训练数据进行预处理和词向量化;随后,构建分层注意力网络模型,将目标领域的训练数据视为无噪声数据,在此基础上训练模型,接着分别对源领域和目标领域的数据进行特征提取,获取文本数据的分布式表示。

2)基于类噪声估计的样例迁移学习。对于源领域数据中的每一个样本,利用基于 Rademacher 和分布的类噪声估计方法计算其与目标领域分布的误差,并根据类噪声估计结果对源领域数据进行筛选,剔除类噪声概率大于阈值的样本,得到高质量的迁移数据。

3)基于支持向量机的情感分类。将目标领域的标注数据与第二阶段得到的源领域迁移数据进行融合,得到最终的训练数据。在此基础上利用支持向量机情感分类器进行训练,并在目标领域的测试集上得到情感类别值。

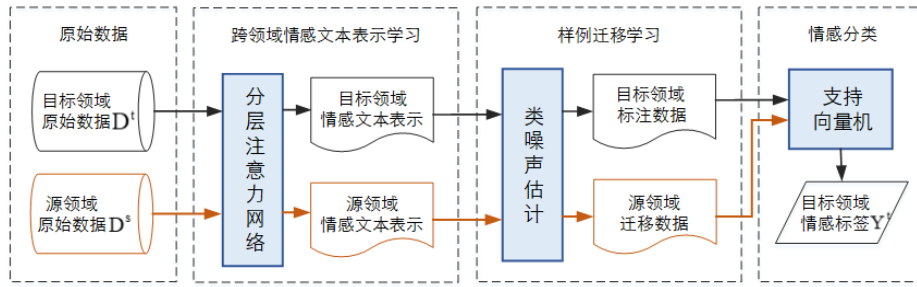


图 1 结合表示学习和迁移学习的跨领域情感分类框架

Fig. 1 Architecture of cross-domain sentiment analysis

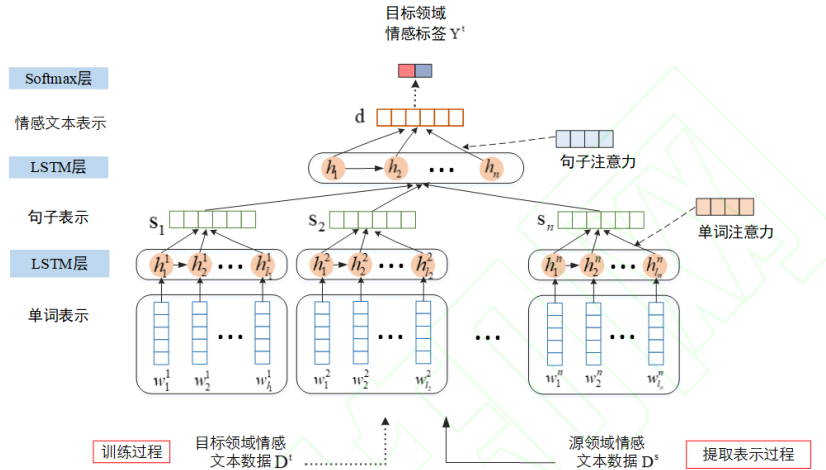


图 2 基于分层注意力网络的跨领域情感文本表示学习模型

Fig. 2 Model architecture of cross-domain representation learning based on hierarchical attention network

### 1.1 基于分层注意力网络的跨领域情感文本表示方法

分层注意力网络将文档中的局部语义作为注意力，利用长短时记忆网络<sup>[19]</sup>(long short-term memory, LSTM)学习文本的单词级和句子级表示，进而得到文档级的分布式表示，很好地捕获文本的局部语义特征。本文采用基于注意力机制的分层 LSTM 网络<sup>[17]</sup>对跨领域文档进行表示学习，利用 LSTM 作为基础模型，采用 word2vec<sup>[20]</sup>预训练好的词向量作为分层注意力网络的输入，通过两层 LSTM 提取单词级和句子级表示，并分别引入单词和句子的局部语义注意力，最终得到情感文本的分布式表示，模型结构如图 2 所示。

假设  $d^s$  为源领域的一个情感文本，且  $d^s = \{s_1, s_2, \dots, s_n\}$ ， $s_i$  表示文本中的第  $i$  个句子，且  $s_i = \{w_1^i, w_2^i, \dots, w_{l_i}^i\}$ ，其中  $w_j^i \in \mathbb{R}^n$  表示文本中第  $i$  个句子中第  $j$  个单词的  $n$  维词向量。设  $l_i$  表示句子  $s_i$  包含的单词总数，该文本的情感标签为  $y$ 。

分层注意力网络可分为单词级和句子级表示学习。在单词级别的表示学习中，句子中的每个单词映射到一个低维语义空间，即用一个  $n$  维词向量表示。在 LSTM 的每一步中，给定一个输入单词  $w_j^i$ ，当前单元状态  $c_j^i$  和隐藏状态  $h_j^i$  可以根据前一个单元状态  $c_{j-1}^i$  和隐藏状态  $h_{j-1}^i$  进行更新：

$$\begin{bmatrix} i_j^i \\ f_j^i \\ o_j^i \end{bmatrix} = \begin{bmatrix} \sigma \\ \sigma \\ \sigma \end{bmatrix} (W \cdot [h_{j-1}^i, w_j^i] + b), \quad (1)$$

$$\hat{c}_j^i = \tanh(W \cdot [h_{j-1}^i, w_j^i] + b), \quad (2)$$

$$c_j^i = f_j^i \odot c_{j-1}^i + i_j^i \odot \hat{c}_j^i, \quad (3)$$

$$h_j^i = o_j^i \odot \tanh(c_j^i), \quad (4)$$

其中  $i, f, o$  是门限函数， $\odot$  表示矩阵乘法， $\sigma$  表示 sigmoid 函数， $W, b$  是模型中需要学习的参数。但并不是所有单词对句子表示的贡献程度都相同，因此，利用注意力机制来获取每个单词对于句子表示的重要程度。具体来说，先通过一层全连接层得到隐藏状态  $h_j^i$  的表示  $u_j^i$ ：

$$\mathbf{u}_j^i = \tanh(\mathbf{W}_w \cdot \mathbf{h}_j^i + \mathbf{b}_w), \quad (5)$$

其中,  $\mathbf{W}_w, \mathbf{b}_w$  为网络参数。然后计算单词的重要性, 即单词级的注意力权重  $\alpha_j^i$ :

$$\alpha_j^i = \frac{\exp(\mathbf{u}_j^{i\top} \cdot \mathbf{u}_w)}{\sum_j \exp(\mathbf{u}_j^{i\top} \cdot \mathbf{u}_w)}, \quad (6)$$

其中, 单词上下文向量  $\mathbf{u}_w$  经过随机初始化后, 在训练期间与模型参数共同学习更新。最后, 整合单词级的注意力权重  $\alpha_j^i$ , 得到句子表示  $\mathbf{s}_i$ :

$$\mathbf{s}_i = \sum_j \alpha_j^i \mathbf{h}_j^i. \quad (7)$$

在句子级别的表示学习中, 将句子向量  $[\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n]$  送入 LSTM 层, 用类似的方法, 通过整合句子级的注意力权重  $\alpha_i$ , 得到文档表示  $\mathbf{d}$ :

$$\mathbf{u}_i = \tanh(\mathbf{W}_s \cdot \mathbf{h}_i + \mathbf{b}_s), \quad (8)$$

$$\alpha_i = \frac{\exp(\mathbf{u}_i^\top \cdot \mathbf{u}_s)}{\sum_i \exp(\mathbf{u}_i^\top \cdot \mathbf{u}_s)}, \quad (9)$$

$$\mathbf{d} = \sum_i \alpha_i \mathbf{h}_i, \quad (10)$$

其中,  $\mathbf{W}_s, \mathbf{b}_s$  为网络参数,  $\mathbf{u}_s$  为句子上下文向量。

在基于分层注意力网络的跨领域情感文本表示学习的训练过程中, 将目标领域训练集  $\mathbf{D}^t \times \mathbf{Y}$  作为模型的训练数据, 通过上述的特征提取阶段, 利用一个 softmax 函数, 对文本表示进行预测, 得到情感类别预测值  $\hat{y}$ :

$$\hat{y} = \text{softmax}(\mathbf{W}_c^\top \mathbf{d}), \quad (11)$$

其中  $\mathbf{W}_c$  是 softmax 层的权值矩阵。通过最小化交叉熵损失函数, 逐层更新网络参数, 可获得一个在目标领域上的最优情感分类器。在文本表示学习的特征提取过程中, 目标领域的训练数据  $\mathbf{D}^t$  与源领域的训练数据  $\mathbf{D}^s$  分别通过训练好的模型, 提取其中最具代表性的文本特征作为文本的表示。

## 1.2 基于类噪声估计的样例迁移学习方法

为解决迁移迭代过程中可能产生负面迁移, 造成跨领域情感分类器的性能下降问题, 本文改进了基于 Rademacher 和分布的类噪声估计方法<sup>[21]</sup>, 并将其引入样例迁移学习算法中, 利用类噪声检测方法筛选出迁移样本中的错误标注样本并剔除(即源领域中带有类噪声的样本), 从而避免负面迁移的影响。

假设无噪声的训练数据集记为  $\mathbf{D}$ , 且  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  是  $n$  个来自数据集  $\mathbf{D}$

的数据, 其中  $x_i$  为文本数据,  $y_i$  为该文本数据的真实标签,  $y_i = \pm 1 (i = 1, 2, \dots, n)$ 。在类噪声问题中, 真实标签往往难以获得, 因此将实际数据集记作  $\tilde{\mathbf{D}}$  (即有噪声数据集  $\tilde{\mathbf{D}}$ ), 其中的训练数据为  $(x_1, \tilde{y}_1), (x_2, \tilde{y}_2), \dots, (x_n, \tilde{y}_n)$ , 这里  $\tilde{y}_i$  是实际标签, 可能与真实标签  $y_i$  不同。因此, 合理假设一个样本的真实标签与其最相似样本的标签相同。采用基于边权统计(cut edge weight statistic)的错误样本估计方法, 对每一个训练样本计算其与其他样本之间的相似度并排序, 选取相似度最高的  $k$  个样本, 与之连接, 得到一个  $K$  近邻图。对于图中任意两个相连的节点  $(x_i, y_i)$  和  $(x_j, y_j)$ , 定义其边的权重为

$$w_{ij} = \text{sim}(x_i, x_j) \quad (12)$$

这里,  $\text{sim}(\cdot)$  是相似度计算函数, 用于度量两个样本之间的相似度, 可以是余弦相似度(Cosine similarity)或基于欧氏距离(Euclidean Distance)的相似度, 且归一化后  $w_{ij} \in [0, 1]$ 。同时, 定义一个符号函数:

$$\mathbf{I}_{ij} = \begin{cases} 1, & \text{if } y_j \neq y_i \\ -1, & \text{if } y_j = y_i \end{cases}. \quad (13)$$

符号函数  $\mathbf{I}_{ij}$  表示了一个样本和与之最相似的样本之间的差异。显然, 当两个输入样本的标签一致时, 符号函数的输出为  $-1$ , 否则为  $1$ , 且  $P(\mathbf{I}_{ij} = 1) = P(\mathbf{I}_{ij} = -1) = \frac{1}{2}$ 。

依据文献[21-22]可得, 对于任意样本  $(x_i, \tilde{y}_i) \in \tilde{\mathbf{D}}$ , 其类噪声率定义为

$$P_c(x_i) = 1 - e^{-r(x_i)/2}, \quad (14)$$

$$r(x_i) = \frac{\left(\sum_{j=1}^k w_{ij} \mathbf{I}_{ij}\right)^4}{(\|\mathbf{w}_i\|_1 \|\mathbf{w}_i\|_2)^2}, \quad (15)$$

其中,  $\|\cdot\|_1$  和  $\|\cdot\|_2$  分别为  $L_1$  和  $L_2$  范数,  $\mathbf{I}_{ij}$  为两个样本  $(x_i, y_i)$  和  $(x_j, y_j)$  之间符号函数的计算结果,  $k$  表示选取最相似样本的个数,  $\mathbf{w}_i = (w_{i1}, w_{i2}, \dots, w_{ik}) \in L_2$ 。类噪声率表示迁移样本在目标领域中被错误标注的概率, 类噪声率大的样本有更大的概率对迁移过程产生负面影响。

根据文献[23], 在有类噪声训练数据集上训练的分类器与真实最优假设之间的误差(即分类器的性能), 对于一个确定的机器学习问题, 其影响因素仅有类噪声率。在估计类噪声的基础上, 可以进一步估计迁移样本集合  $TS$  的误差界:



$$\varepsilon = \sqrt{\frac{2 \ln(\frac{2N}{\delta}) |TS|}{[|TS| - \sum_i P_c(x_i)]^2}}, \quad (16)$$

其中,  $\delta$  是平滑系数,  $N$  是与特征空间相关的常

数。迁移样本集合的误差界用于对当前迭代周期所训练的情感分类器质量进行评估。

基于类噪声估计的样例迁移学习算法的具体过程如算法 1 所示。

**算法 1 基于 Rademacher 和分布类噪声估计的样例迁移学习算法**

输入: 目标领域上的有标注数据  $L$ ; 源领域上的无标注数据  $U$ ; 分类器  $C$ ; 最大迭代次数  $T$ ; 检测周期  $T_D$ ; 当前周期  $n$ ; 当前周期的候选迁移样本集合  $TS_n = \phi$ ; 已迁移样本集合  $TS = \phi$ ; 每轮迭代添加的正面样本数  $p$  和负面样本数  $q$ ; 每个周期估计的测试误差边界  $\varepsilon = 1$  ( $i = 1, 2, \dots, n$ )

输出: 迁移样本集合  $TS$

```

1   $n = 1$ 
2  for  $iter \leftarrow 1$  to  $T \times T_D$  do
3      在  $L$  上训练分类器  $C$ 
4      使用  $C$  对  $U$  进行分类, 将置信度最高的  $p$  个正面样本和  $q$  个负面样本加入  $TS_n$ 
5      if  $iter \% T_D == 0$  then
6          for  $x_i \in TS \cup TS_n$  do
7              对样本  $x_i$  建立  $K$  近邻图, 并根据式(14)计算类噪声率  $P_c(x_i)$ 
8          end for
9          根据式(16)进一步估计  $TS \cup TS_n$  的误差界  $\varepsilon_n$ 
10         使用类似方法估计  $TS$  的误差界  $\varepsilon_{n-1}$ 
11         if  $\varepsilon_n > \varepsilon_{n-1}$  then
12             break
13         else
14             对于  $TS_n$  中的每个样本  $x_i$ , if  $P_c(x_i) > \theta$  then  $TS_n \leftarrow TS_n - \{x_i\}$ 
15              $TS = TS \cup TS_n$ ,  $n++$ ,  $TS_n = \phi$ 
16         end if
17     end if
18 end for
    
```

### 1.3 基于支持向量机的情感分类方法

支持向量机<sup>[24]</sup>是一种监督学习方法, 在情感分类任务<sup>[18]</sup>上表现出了优异的性能, 且具有泛化错误率低, 计算开销小等优势。支持向量机建立一个超平面作为决策面, 使类别间隔最大化, 即:

$$\begin{cases} \max \frac{2}{\|w\|} \\ \text{s.t. } y_i(w^T x_i + b) \geq 1, i = 1, 2, \dots, N \end{cases} \quad (17)$$

其中,  $x_i$  与  $y_i$  分别表示训练样本及其对应标签,  $N$  为训练集大小,  $w$  和  $b$  为待求解的参数。

第一阶段, 目标领域训练数据  $D^t$  与源领域训练数据  $D^s$  通过基于分层注意力网络, 分别得到对应的情感文本表示  $X^t$  与  $X^s$ ; 第二阶段, 利用基于类噪声估计的迁移样例筛选算法, 从源领域

样本中挑选出高质量的迁移样本数据  $X^s \subset X^s$ ; 合并目标领域与源领域的训练数据, 得到  $X = X^t \cup X^s$ ; 第三阶段, 利用合并数据集  $X \times Y$ , 训练支持向量机情感分类器, 并在目标领域的测试集  $S$  上得到情感类别的预测值。

## 2 实验结果及分析

### 2.1 数据集

为了与基础实验进行对比, 本文选用的数据集是 IMDB 电影评论数据集以及 Yelp 网站上 2013 年和 2014 年的餐馆评论数据集<sup>[17]</sup>进行实验, 具体的统计信息如表 1 所示。这 3 个公开数据集均包含大规模的英文评论文本以及对应的情感类别, 且数据分布不一致。

表 1 实验数据集的统计信息  
Table 1 Statistics of the datasets

数据集	情感类别	产品总数	文档总数	词语总数	平均字数/条	训练文档数	测试文档数
IMDB	1~10	1635	84919	91808	395.21	67426	9112
Yelp2013	1~5	1633	78966	96817	188.76	62522	8671
Yelp2014	1~5	4194	231163	183541	196.06	183019	25399

由于 Yelp2013 和 Yelp2014 的情感类别(1~5)与 IMDB(1~10)有所差异, 因此要对数据进行处理。本文采取归并方案, 利用  $\lfloor (label + 1) / 2 \rfloor$  将 IMDB 的情感类别  $label$  映射到 5 分类标注体系中。由于对于 5 分类体系较难向更多分类映射, 所以本文实验仅仅针对 Yelp 不同年份数据之间的迁移以及 IMDB 数据集, 向 Yelp 数据集的迁移。

为了评估算法的性能, 我们选用分类算法中常用的两个评价指标<sup>[25]</sup>: 准确率(Accuracy)和均方根误差(root mean square Error, RMSE), 计算公式如下所示。

$$Accuracy = \frac{\sum_{i=1}^N \hat{y}_i = y_i}{N} \cdot 1, \quad (17)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}}, \quad (18)$$

其中,  $\hat{y}_i$  表示第  $i$  个样本的预测标签,  $y_i$  则表示其正确标签,  $N$  为测试集的大小。

## 2.2 实验参数设置

对于 word2vec 与分层注意力网络中的参数,

为便于对比, 用文献[17]中的参数设置, 将每个情感文本映射为一个 200 维的情感文本表示向量。

在基于类噪声估计的样例筛选过程中, 选取统计学中常用的置信度标准 0.05 作为阈值<sup>[26]</sup>, 即: 若一个样本的类噪声率  $P_c(x_i) > 0.95$ , 则这一样本  $x_i$  被认为是负面迁移样本。另外, 由于在文献[21]的类噪声检测算法中仅考虑二分类问题, 但本文所解决的是多分类问题, 故对原始的类噪声检测算法进行适当改进以统计相似样本中的标签差异: 针对五分类的数据集, 对于情感类别相差 1 的相似样本, 不认为这两个样本的情感有明显差异; 对于情感类别相差为 2 及以上的相似样本, 则认为有明显差异。因此, 本文将 5 分类标签分为正中负 3 档: 1 和 2 为负面, 3 为中性, 4 和 5 为正面。总之, 重新定义类噪声估计中的符号函数, 当且仅当两个样本的标签有正负面差异时才输出为 1, 否则输出-1。

本文实验参数的具体设置如表 2 所示。

表 2 参数设置  
Table 2 Parameters setup

参数类型	可调参数	参数值
word2vec	词向量维度	200
	词向量训练模型	Skip-Gram
分层注意力网络	LSTM 单元状态维度	200
	LSTM 隐含状态维度	200
	注意力权重维度	200
基于类噪声估计的 样例迁移方法	优化方法	Adadelta
	近邻数 $k$	8
支持向量机	类噪声阈值 $\theta$	0.95
	核函数	径向基函数( $\gamma=0.005$ )
	惩罚系数	1.0

### 2.3 基于文本表示学习的跨领域情感分类实验

在 Yelp 两个不同年份数据集上, 分别采用卷积神经网络(CNN)、分层 LSTM 网络(NSC)以及基于分层注意力网络(NSC+LA)的文本表示学习方法, 进行情感分类实验, 结果如表 3 所示。

由表 3 可以看出, 分层注意力网络能较好地提取文本中的重要局部特征, 从而达到较好的情感分类性能。

作为参照, 先简单地向两个数据集添加来自其他训练集的样本, 不进行迁移数据的筛选, 利用支持向量机训练情感分类器, 实验结果如表 4 所示。

从表 4 可以看出, 将两个训练集直接合并后, 分类效果有明显的下降, 主要原因在于两个数据集来自于不同的领域, 领域之间的数据分布具有较大差异, 直接迁移会产生负面影响。从 Yelp2013 与 IMDB 之间进行的迁移实验可以看出, 简单合并之后出现的负面效果尤为明显, 这是由于 Yelp2013 与 IMDB 的数据规模相差不大, 负面迁

移影响更为显著。在 Yelp2014 上进行的实验中, 加入其他两个数据集之后, 性能略微提升, 并且由于 Yelp2013 与 Yelp2014 的数据分布较为类似, 性能提升较明显。这仍说明简单合并数据集的方法并不能有效地迁移不同领域的的数据。

### 2.4 结合类噪声检测的跨领域情感分类实验

利用本文提出的算法对迁移数据进行基于类噪声估计的筛选, 并使用支持向量机训练分类器, 实验结果如表 5 所示。

对比表 4 和 5 的实验结果可以看出, 与简单合并训练集相比, 对迁移数据进行筛选后, 各个数据集上的性能都有较大的提高, 说明利用基于类噪声估计的样例迁移算法可以有效提升跨领域情感分类性能, 表明对迁移数据进行基于类噪声估计的筛选可以避免负面迁移, 从而有效提高迁移样本的质量。特别是在 Yelp2013 数据集上的实验结果, 比表 4 中简单合并的准确率(0.618 和 0.628)有大幅的提升, 说明了进行高质量数据迁移的必要性。

表 3 基于文本表示学习的基准分类实验

Table 3 Results of textual representation learning

文本表示学习模型		CNN		NSC		NSC+LA	
训练数据	测试数据	Accuracy	RMSE	Accuracy	RMSE	Accuracy	RMSE
Yelp2013 训练集	Yelp2013 测试集	0.577	0.812	0.627	0.701	<b>0.631</b>	<b>0.706</b>
Yelp2014 训练集	Yelp2014 测试集	0.585	0.808	<b>0.637</b>	<b>0.686</b>	0.630	0.715

表 4 合并训练集的实验结果

Table 4 Results of merging training set

训练数据	测试数据	Accuracy	RMSE
Yelp2013 训练集 + IMDB 训练集	Yelp2013 测试数据	0.618	0.742
Yelp2013 训练集 + Yelp2014 训练集	Yelp2013 测试数据	0.628	0.724
Yelp2014 训练集 + IMDB 训练集	Yelp2014 测试数据	0.632	0.711
Yelp2014 训练集 + Yelp2013 训练集	Yelp2014 测试数据	0.639	0.709

表 5 基于类噪声估计的样例迁移学习实验结果

Table 5 Results of transfer learning based on class noise estimation

训练数据	测试数据	Accuracy	RMSE
Yelp2013 标注数据 + IMDB 迁移数据	Yelp2013 测试数据	0.634	0.686
Yelp2013 标注数据 + Yelp2014 迁移数据	Yelp2013 测试数据	<b>0.642</b>	<b>0.672</b>
Yelp2014 标注数据 + IMDB 迁移数据	Yelp2014 测试数据	0.645	0.676
Yelp2014 标注数据 + Yelp2013 迁移数据	Yelp2014 测试数据	<b>0.650</b>	<b>0.662</b>

以 Yelp13 标注数据+IMDB 迁移数据的实验为例, 在近邻数  $k=8$  时, 选取 IMDB 训练集中两个类噪声率不同的样本进行对比, 如图 3 所示。图 3(a)表示一个文本样本的情感标签为 5, 通过类噪声估计方法计算得出类噪声率较小, 因为其其与相近邻的 8 个文本样本的标签的标签差异不大, 可以认为该文本样本标注正确, 较符合目标领域的分布。图 3(b)中, 文本样本的标签为 4, 与其周围的近邻样本标签相差较大, 类噪声率较大, 说明该样本在目标领域属于错误标注样本, 与目标领域的分布相差较大, 有极大可能性会造成分类器性能下降, 产生负面迁移。由图 3 可以看出, 基于 Rademacher 和分布的类噪声估计方法可以有效过滤错误标注的样本, 从而进一步提升跨领域情感分类的性能。

图 4 表示近邻数  $k$  不同时, 在不同数据集上基于类噪声估计的样例迁移学习算法的准确率变化。可以看出, 当  $k$  为 7 或 8 时, 本文方法能达到较好的效果。

为了进一步说明本文提出的情感分类方法的

有效性, 我们将该方法与目前主流的情感分类方法进行对比: 1)Paragraph Vector, 利用 Le 等<sup>[27]</sup>提出的文档向量表示训练支持向量机进行分类; 2)RNTN+Recurrent: 利用 Socher 等<sup>[28]</sup>提出的递归神经网络进行情感分类; 3)CNN, 利用 Kim<sup>[16]</sup>提出的卷积神经网络进行情感分类; 4)NSC 和 NSC+LA, 利用 Chen 等<sup>[17]</sup>提出的两种文档级情感分类方法进行情感分类; 5)ATL+DGP, 基于深度高斯过程的非对称迁移学习<sup>[9]</sup>进行跨领域情感分类, 使用 IMDB 训练集作为辅助数据, 利用分层注意力网络提取文本特征后利用深度高斯过程进行迁移学习。

为了便于对比, 前 5 个对比实验的参数沿用文献[17]中实验的参数设置, ATL+DGP 方法中高斯过程的参数与文献[9]中一致, 实验结果如表 6 所示。值得说明的是, ATL+DGP 方法在迁移学习中未能对负面迁移样本进行筛选, 因而分类效果不具优势。由表 6 可看出, 本文提出的方法能有效过滤造成负面迁移的样本, 在负面迁移出现之前停止迭代, 从而在文本表示学习的基础上提高大规模文本跨领域情感分类的性能。

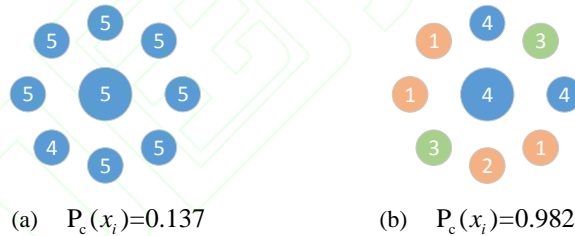


图 3 不同类噪声率样本的比较

Fig. 3 Comparison of samples over different class noise rate

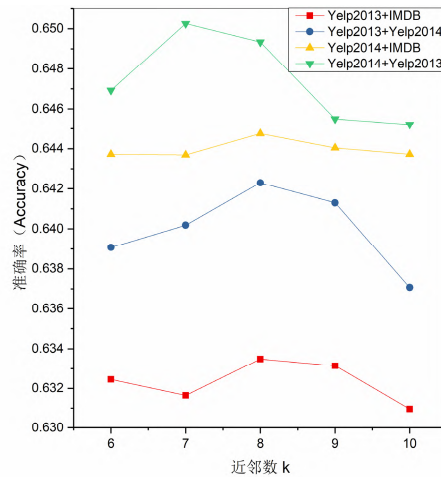


图 4 近邻数  $k$  的取值对准确率的影响

Fig. 4 Accuracy over parameter  $k$



表 6 跨领域情感分类实验结果

Table 6 Results of cross-domain sentiment classification

模型方法	Yelp2013		Yelp2014	
	Accuracy	RMSE	Accuracy	RMSE
Paragraph Vector	0.554	0.832	0.564	0.802
RNTN+Recurrent	0.574	0.804	0.582	0.821
CNN	0.577	0.812	0.585	0.808
NSC	0.627	0.701	0.637	0.686
NSC+LA	0.631	0.706	0.630	0.715
ATL+DGP	0.627	0.716	0.639	0.686
本文方法	<b>0.634</b>	<b>0.686</b>	<b>0.645</b>	<b>0.676</b>

### 3 结论

针对跨领域情感分类方法中存在的文本表示忽略局部语义特征和负面迁移问题, 本文提出一种基于分层注意力网络的文本表示方法与基于样例的迁移学习算法相结合的跨领域情感分类方法, 引入基于类噪声估计的迁移数据筛选方法, 从源领域中选取高质量样本迁移到目标领域中, 以提高文本跨领域情感分类系统的性能。在公开的大规模产品评价数据集上的实验表明, 表示学习与迁移学习相结合的方法可以提高文本跨领域情感分类系统的性能, 同时验证了该方法可以有效提升迁移样本质量, 从而说明了在大规模文本中引入迁移学习的必要性。但方法达到较好性能需要大量数据集, 未来将进一步研究半监督任务中的大规模跨领域情感分类问题, 以更好地应用于实际场景, 提高情感分类性能。

#### 参考文献

[1] Subramaniaswamy V, Logesh R, Abejith M, et al. Sentiment analysis of tweets for estimating criticality and security of events. *Journal of Organizational & End User Computing*, 2017, 29(4): 51–71

[2] Al-Moslmi T, Omar N, Abdullah S, et al. Approaches to cross-domain sentiment analysis: a systematic literature review. *IEEE Access*, 2017, 5(99): 16173–16192

[3] Wei Xiacong, Lin Hongfei, Yu Yuhai, et al. Low-resource cross-domain product review sentiment classification based on a CNN with an auxiliary large-scale corpus. *Algorithms*, 2017, 10(3): 81

[4] Zhao Chuanjun, Wang Suge, Li Deyu. Deep transfer learning for social media cross-domain sentiment classification // Chinese National Conference on Social Media Processing. Singapore, 2017: 232–243

[5] 李良豪. 跨领域文本分类算法研究. 北京: 清华大学, 2012

[6] Ziser Y, Reichart R. Neural structural correspondence learning for domain adaptation. *Conference on Computational Natural Language Learning*. Vancouver, 2017: 400–410

[7] Pan S J, Ni Xiaochuan, Sun J T, et al. Cross-domain sentiment classification via spectral feature alignment // *Proceedings of the 19th International Conference on World Wide Web*. Raleigh, 2010: 751–760

[8] Pan S J, Tsang I W, Kwok J T, et al. Domain Adaptation via Transfer Component Analysis. *IEEE Transactions on Neural Networks*, 2011, 22(2): 199–210

[9] 吴冬茵, 桂林, 陈钊, 等. 基于深度表示学习和高斯过程迁移学习的情感分析方法. *中文信息学报*, 2017, 31(1): 169–176

[10] Dai Wenyuan, Yang Qiang, Xue Guirong, et al. Boosting for transfer learning // *Proceedings of the 24th International Conference on Machine Learning*. Corvallis, 2007:193–200

[11] Huang Xingchang, Rao Yanghui, Xie Haoran, et al. Cross-domain sentiment classification via topic-related TrAdaBoost // *Proceedings of the 31st Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*. San Francisco, 2017: 4939–4940

[12] Seah C W, Tsang I W, Ong Y S. Transfer ordinal label learning. *IEEE Transactions on Neural Networks & Learning Systems*, 2013, 24(11): 1863–1876

[13] Sugiyama M, Nakajima S, Kashima H, et al. Direct

- importance estimation with model selection and its application to covariate shift adaptation // Proceedings of the 21st Annual Conference on Neural Information Processing Systems. Vancouver, 2007: 1433–1440
- [14] Hajmohammadi M S, Ibrahim R, Selamat A, et al. Combination of active learning and self-training for cross-lingual sentiment classification with density analysis of unlabelled samples. *Information Sciences*, 2015, 317: 67–77
- [15] Wu Fangzhao, Huang Yongfeng, Yan Jun. Active sentiment domain adaptation // Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, 2017: 1701–1711
- [16] Kim Y. Convolutional neural networks for sentence classification // Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing. Doha, 2014: 1746–1751
- [17] Chen Huimin, Sun Maosong, Tu Cunchao, et al. Neural sentiment classification with user and product attention // Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing. Austin, 2016: 1650–1659
- [18] Luo Fang, Li Cheng, Cao Zehui, et al. Affective-feature-based sentiment analysis using SVM classifier // Proceedings of the 20th International Conference on Computer Supported Cooperative Work in Design. Nanchang, 2016: 276–281
- [19] Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Computation*, 1997, 9(8): 1735–1780
- [20] Mikolov T, Sutskever I, Chen Kai, et al. Distributed representations of words and phrases and their compositionality // Proceedings of the 27th Annual Conference on Neural Information Processing Systems. Lake Tahoe, 2013: 3111–3119
- [21] Gui Lin, Xu Ruifeng, Lu Qin, et al. Negative transfer detection in transductive transfer learning. *International Journal of Machine Learning & Cybernetics*, 2017, 9(2): 1–13
- [22] Gui Lin, Lu Qin, Xu Ruifeng, et al. A novel class noise estimation method and application in classification // Proceedings of the 24th ACM International on Conference on Information and Knowledge Management. Melbourne, 2015: 1081–1090
- [23] Angluin D, Laird P D. Learning from noisy examples. *Machine Learning*, 1988, 2(4): 343–370
- [24] Ukil A. Support vector machine. *Computer Science*, 2002, 1(4): 1–28
- [25] Fatourechí M, Ward R K, Mason S G, et al. Comparison of evaluation metrics in classification applications with imbalanced datasets // Proceedings of the 7th International Conference on Machine Learning and Applications, San Diego, 2008: 777–782
- [26] 李贤平. 概率论基础. 2版. 北京: 高等教育出版社, 1997
- [27] Le Q V, Mikolov T. Distributed representations of sentences and documents // Proceedings of the 31th International Conference on Machine Learning. Beijing, 2014: 1188–1196
- [28] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment Treebank // Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing. Seattle, 2013: 1631–1642