

# 基于多视角多标签学习的读者情绪分类

温 雯<sup>1</sup> 陈 颖<sup>1</sup> 蔡瑞初<sup>1</sup> 郝志峰<sup>1,2</sup> 王丽娟<sup>1</sup>

(广东工业大学计算机学院 广州 510000)<sup>1</sup>

(佛山科学技术学院数学与大数据学院 广东 佛山 528000)<sup>2</sup>

**摘 要** 传统的读者情绪分类主要从情感分析的角度出发,着重考量读者评论中体现出来的情感极性。然而现实中,读者评论的缺失有可能影响情绪分类的有效性和及时性。如何融合包括新闻文本和评论在内的多视角信息,对读者情绪进行更加准确的研判,成为了一个具有挑战性的问题。针对这一问题,构建了一种融合多视角信息的多标签隐语义映射模型(Multi-view Multi-label Latent Indexing, MV-MLSI),将不同视角下的文本特征映射到低维语义空间,同时建立特征和标签之间的映射函数,通过最小化重构误差对模型进行求解,并设计了相关算法,从而实现了对读者情绪的有效预测。相比于传统模型,该模型不仅可以充分利用多视角的信息,而且考虑了标签之间的相关性。在新闻文本数据集上的实验表明,该方法可以获得更高的准确率和稳定性。

**关键词** 情绪分类,情感分析,多标签学习,多视角学习,LSI

中图法分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2018.08.034

## Emotion Classification for Readers Based on Multi-view Multi-label Learning

WEN Wen<sup>1</sup> CHEN Ying<sup>1</sup> CAI Rui-chu<sup>1</sup> HAO Zhi-feng<sup>1,2</sup> WANG Li-juan<sup>1</sup>

(Department of Computers, Guangdong University of Technology, Guangzhou 510000, China)<sup>1</sup>

(Department of Mathematics and Big Data, Foshan University, Foshan, Guangdong 528000, China)<sup>2</sup>

**Abstract** The traditional emotion classification for readers mainly focuses on the emotional polarity embodied in the reader's comments, which is from the perspective of sentiment analysis. However, the readers' comments are occasionally not collected due to some reasons, which tends to reduce the effectiveness and timeliness of emotional classification. How to integrate the multi-perspective information, including news texts and readers' comments, and to make a more accurate judgment of reader's emotions has become a challenging problem. In this paper, a multi-view multi-label latent indexing (MV-MLSI) model was proposed, which maps the multi-view text features from different perspectives to the low-dimensional semantic space. Meanwhile, the mapping function among the features and labels was established, and the model could be solved by minimizing the reconstruction error. The optimization algorithm was also presented in this paper so as to make the effective prediction of reader's emotion. Compared with the traditional model, the proposed model can not only take full advantage of multi-view information, but also take into account the correlation among labels. Experiments on the multi-view news text dataset demonstrate that the method can achieve higher accuracy and stability.

**Keywords** Emotion classification, Sentiment analysis, Multi-label learning, Multi-view learning, Latent semantic indexing

## 1 引言

随着互联网的快速发展和 Web 2.0 时代的到来,特别是近年来新闻、微博、论坛等用户参与性极强的各类应用蓬勃兴起,用户可以直接表达自己的观点和情感,使得网络中产生了

大量饱含读者情绪的文本信息。通过分析读者阅读时产生的情绪,可以帮助商家进行个性化推荐,还可以辅助政府、企业等机构了解社会舆论和民众的诉求,以便及时发现当下社会热点进而进行舆情监控。此外,随着数据采集技术的快速发展,情绪语料的获取变得轻而易举,对用户进行情绪分析已经

到稿日期:2017-07-26 返修日期:2017-10-06 本文受国家自然科学基金(61202269,61472089)资助。

温 雯(1981—),女,博士,副教授,CCF 会员,主要研究方向为机器学习、模式识别、信息检索,E-mail:wwen\_gdut@qq.com(通信作者);陈 颖(1993—),女,硕士生,主要研究方向为数据挖掘、模式识别,E-mail:511093965@qq.com;蔡瑞初(1983—),男,博士,教授,CCF 高级会员,主要研究方向为数据挖掘、机器学习、信息检索;郝志峰(1968—),男,博士,教授,CCF 会员,主要研究方向为机器学习、人工智能;王丽娟(1978—),女,博士,副教授,主要研究方向为机器学习、高维数据聚类分析。

成为自然语言处理等领域的研究热点<sup>[1-2]</sup>。

情绪表达形式多样,例如,新闻读者的情绪不仅可以通过新闻内容来判断,而且可以通过新闻评论来分析,而新闻内容和新闻评论就构成了对读者情绪分析的两个视角。然而,已有的大多数工作只利用单个视角信息对读者情绪进行分类<sup>[3]</sup>,属于单视角学习。在真实数据中,单视角语料往往存在缺失。例如,对于一篇新闻报道,读者读完后若感觉无聊,则往往不会提交相应的评论,导致该情绪标签“无聊”没有出现在评论信息中,只用评论信息作为分类特征会使得分类不够准确。因此,结合不同的视角信息使其相互补充和融合,显得尤为重要。

另一方面,人的情绪也是复杂多样的,情绪是人们对社会上各种现象拥有的不同观点和立场的表达。考虑到每个人的背景和经历存在差异,对事物也会持有不同的观点,甚至同一个人对同一件事物也会产生多种复杂的情绪,Ekman<sup>[4]</sup>将人的情绪划分为6种基本状态:喜、哀、怒、惧、惊、恶。由于一篇文档可能使读者产生多种情绪,因此,读者情绪分析还是一个多标签分类问题<sup>[5]</sup>。

综上所述,读者情绪分类事实上是一个多视角多标签分类问题。对于这类问题,现有的工作主要面临以下挑战:

1) 文本数据的高维度问题。由于文本数据有成千上万个词汇,不同视角的文本特征维度较高,高维度数据的处理相对复杂。

2) 多标签间的相关性问题。多标签之间往往隐含着标签之间的特殊相关性,同时也导致了不同标签之间数据的不平衡性,如某些标签语料充足,某些标签语料则十分少。多标签特性对传统分类器的依赖性较大。

3) 多视角的信息融合问题。多视角信息在一定程度上可以相互补充,有效的多视角信息融合将会比单视角取得更好的性能,无效的融合反而会降低学习性能。

针对以上挑战,本文提出了一种基于多视角多标签学习的情绪分类模型(Multi-view Multi-label Latent Semantic Indexing, MV-MLSI)。该模型基于LSI进行改进,结合标签信息将不同视角的特征投影在低维隐语义子空间,不仅解决了特征维度过高的问题,而且考虑了特征与标签之间的映射以及不同情绪标签之间的相关性。在此基础上训练得到的多标签分类器,能有效地将不同视角的特征和标签进行有机融合。结果表明,与已有的相关方法相比,本文提出的方法能够提高新闻文本的读者情绪分类的准确率。

## 2 相关工作

与本文方法最为相关的工作有3类:用户情绪分析、多标签学习和多视角学习。

### 2.1 用户情绪分析

情绪分析的主要目的是自动识别用户可能产生的情绪类别。已有的研究成果大致可以分为两大类:基于情绪词典的方法和基于机器学习的方法。基于情绪词典的方法主要依赖一些情感词典或者领域词典以及文本中一些带有感情色彩的组合评价单元来计算情感得分,进而判断情感的极性。例如,文献[6-9]通过分析句子/篇章的评价词语或者组合评价短语

的极性进行加权求和。牛耕等人<sup>[10]</sup>提出了基于情感词典和规则的方法对微博文本的情绪进行分析,该文选取了HowNet和C-LIWC两个被广泛应用的情感词典,根据情感词典查找微博包含的情绪词,情绪词数量最多的则为该微博的主要情绪。Wang等人<sup>[11]</sup>提出了基于机器学习和词典相结合的方法来构建统一的情绪分类器。基于词典和规则的方法存在情绪词语义模糊、难以处理客观和隐晦的文本等问题,效果并不理想。为了解决这类问题,研究者提出使用机器学习的方法对用户情绪进行分析。例如,Pang等人<sup>[12]</sup>首次将机器学习的方法应用在篇章级情感分类中。他们使用n-gram作为特征,用传统的机器学习模型进行分类与比较,效果明显优于人工选取情感特征的方法。此外,考虑到情绪和主题的相关性,叶璐等人<sup>[13]</sup>利用主题模型实现了读者情绪的预测,并且在LDA的基础上采用加权方法进行改进,利用W LDA主题模型实现降维,进而对读者情绪进行预测。还有一些学者使用word2vec对各个情绪类别构建有效的语义词簇,采用词簇与词簇之间的余弦距离对其进行筛选,进而将传统的词向量转换成语义词簇上的表达,并在此基础上结合使用多标签分类器MLKNN进行新闻读者情绪的分类<sup>[14]</sup>。在情绪标签中,标签之间常常存在一定的相关性。例如,情绪标签“火大”和“难过”经常同时出现,标签之间并不是相互独立的。在以往的情绪分析研究中,很少有研究者考虑标签之间的相关性。

考虑到标签之间的相关性,学者们相继提出了结合标签相关性的多标签学习方法。例如,XU等人<sup>[15]</sup>将读者情绪预测看作多标签分类问题,利用PLDA将情绪类型词与主题相结合,将PLDA的结果作为特征进行情绪的预测。Luyckx等人<sup>[16]</sup>设计了一种阈值的方法,为单个实例分配多个标签,利用SVM分类器返回概率估计,通过对返回的概率设置阈值来分配情感标签。刘欢欢等人<sup>[17]</sup>则从多视角的角度提出结合作者情绪和读者情绪来对新闻的读者情绪做粗粒度的预测,即直接针对新闻文本和相应的评论文本分别构建两个分类器,最后采用多种基于后验概率的方法进行融合,从而得出最终的分类结果。但该方法没有考虑到不同视角间以及不同标签之间的相关性。

### 2.2 多标签学习

在近十年的研究中,多标签学习得到了全面的发展。这些研究从算法层面上大致可以分为两类:问题转换法和算法适应法。问题转换法的主要思想是将多标签问题转换成单标签问题,并利用已有的单标签算法(如SVM、贝叶斯分类器、KNN等)来解决。例如,殷会等人<sup>[18]</sup>提出了基于LS-SVM的多标签分类算法,他们将 $K$ 个标签的数据集分解为 $K(K-1)/2$ 个两类单标签和两类双标签子集,然后用LS-SVM分类器对每个子集进行分类。申超波等人<sup>[19]</sup>提出了基于标签聚类的多标签分类算法,该算法通过改进K-Means聚类算法来挖掘训练集中重要的标签集合,并将其用于形成新的训练集,从而进行多标签分类。该类方法常常独立区分各个标签,无法有效地结合标签之间的相关性。算法适应法则是通过改进现有的算法来解决多标签的问题,Zhang等<sup>[20]</sup>提出了基于 $k$ 近邻的多标签(MLKNN)学习算法来处理文本的标签信息。该方法通过计算 $k$ 近邻所属的类别来确定自己所

属的类别,对不同的分类任务都取得了不错的效果。现有的研究大多致力于利用标签之间的相关性来提高分类器的性能<sup>[21-22]</sup>,Yu 等<sup>[23]</sup>考虑到标签之间的相关性,提出了有监督的 LSI 特征映射方法来进行多标签分类。该方法将文本特征联合标签信息映射到低维隐语义子空间,得到了与标签关联的文本特征,再采用 SVM 分类器分别对每个类别进行分类。近年来,启发式技术也被用来探索标签的特性,以改进多标签分类<sup>[24]</sup>。此外,还有部分学者把多标签学习看成是对训练特征矩阵、测试特征矩阵和标签矩阵所组成的矩阵进行完善的过程<sup>[21,25]</sup>。尽管很多复杂的方法被相继提出以解决多标签学习问题,但大多数解决方案主要还是从单一视角出发。

### 2.3 多视角学习

多视角学习是机器学习领域中的另一个研究热点<sup>[26-27]</sup>。在传统的机器学习中,大多数研究基于一个视角,但是由于数据的采集方式和特征提取的多样化,在现实问题中常常存在多个视角。例如,在网页分类问题中,不仅有网页本身,还有网页间的链接关系等。单一的视角很难全面地描述事物,需要综合多种视角来共同完成对事物的描述。多视角学习已有许多研究成果<sup>[28-29]</sup>,现有的方法主要从特征融合和分类器融合角度出发,典型的方法有子空间学习法,该方法将多个视角映射到相同的子空间,将该子空间特征用于分类、聚类等任务。该子空间的维度低于输入数据的任意视角,因此可以同时解决维度灾难问题。Xia 等人<sup>[29]</sup>提出了一种多视角谱嵌套 (Multi-view spectral embedding) 算法,其利用不同视角间的互补信息,在不同视角上学习得到一个低维平滑嵌套。还有一些方法通过对矩阵的分解和重构实现对多个视角的信息融合<sup>[26,30]</sup>,这些方法大多需要预先给定相似矩阵,并且没有考虑到标签信息。近年来,也有其他一些解决该类问题的相关技术被相继提出<sup>[31-32]</sup>。

## 3 模型

### 3.1 问题定义

综合上述分析,本文把读者情绪分析建模成多视角多标签分类问题。读者情绪多标签分类不仅存在特征维度过高、单一视角特征不足的问题,而且还需要考虑标签之间的相关性。鉴于此,本文提出以下机器学习模型来解决上述问题。

对于给定的具有多视角特征的  $N$  篇文档集  $D = \{d_1, d_2, \dots, d_N\}$ , 其相应的情绪标签集为:  $Y = \{\vec{y}_1, \vec{y}_2, \dots, \vec{y}_N\}$ ,  $\vec{y}_i = \{y_{i1}, y_{i2}, \dots, y_{iL}\}$ ,  $Y \in R^{N \times L}$ 。其中,  $y_{ij} = 1$  表示第  $i$  篇文档中有第  $j$  个标签,  $y_{ij} = -1$  则表示第  $i$  篇文档中没有第  $j$  个标签。给定特征矩阵  $X_1, X_2$ , 其中  $X_1 \in R^{N \times M_1}$  为视角一的特征,  $X_2 \in R^{N \times M_2}$  为视角二的特征。

为了便于理解,本文在提出多视角多标签分类模型 (MV-MLSD) 之前,首先介绍 LSI 模型,然后在此基础上引入有监督的、适用于多视角数据的多标签分类模型。

### 3.2 LSI 模型

LSI 模型<sup>[20]</sup>旨在找到一个线性映射  $x \mapsto v$ , 即输入特征  $x$  在低维隐语义空间上的映射  $v$ , 使得  $v$  保留尽可能多的原始信息。因此,通过最小化重构误差构建以下模型:

$$\begin{aligned} \min_{A, V} & \|X - VA\|^2 \\ \text{s. t.} & V^T V = I \end{aligned} \quad (1)$$

其中,  $V \in R^{N \times K}$ ,  $A \in R^{K \times M}$ ,  $K \leq M$ ,  $V$  为  $X$  在低维空间上的映射, 矩阵  $V$  中的列即为矩阵  $X$  中行相对应的隐变量,  $A$  为  $X$  到  $V$  的映射因子,  $I$  表示单位矩阵。优化目标函数(1), 使  $VA$  近似等于矩阵  $X$ , 即在投影的空间中保留尽可能多的原始信息。此处, 约束条件  $V^T V = I$  保证了隐变量之间的正交性。通过优化模型(1), 即可求得  $V$  的最优解。

### 3.3 MV-MLSI

在模型 LSI 中构建了从  $X$  到  $V$  的射, 该模型是无监督、单视角学习模型。LSI 模型对于读者情绪分类这类具有多标签以及多视角特性的数据并不适用。LSI 模型并没有考虑输入特征  $X$  和输出标签  $Y$  之间的相关性, 以及标签内部之间的相关性, 并且只适用于单视角数据的学习。针对以上不足, 本文提出了有监督的多视角多标签模型, 将两个视角特征  $X_1$  和  $X_2$  分别映射到不同的隐语义空间, 同时将标签  $Y$  与  $X_1, X_2$  的隐语义投影在同一个空间, 进而建立多个视角语义与输出标签的相关性。最后, 通过最小化重构误差求解, 定义模型如下:

$$\begin{aligned} \min_{A, B, C, D, V_1, V_2} & a_1 \|X_1 - V_1 A\|^2 + a_2 \|X_2 - V_2 B\|^2 + \\ & \|Y - V_1 C - V_2 D\|^2 \\ \text{s. t.} & \begin{cases} V_1^T = I \\ V_2^T V_2 = I \end{cases} \end{aligned} \quad (2)$$

其中,  $V_1 \in R^{N \times K_1}$  为视角一的特征  $X_1$  在隐语义空间  $\Omega_1$  的基,  $V_2 \in R^{N \times K_2}$  为视角二的特征  $X_2$  在隐语义空间  $\Omega_2$  上的基。其中, 参数  $A$  和参数  $B$  分别为  $X_1$  到  $V_1, X_2$  到  $V_2$  的线性权重,  $C$  和  $D$  为标签  $Y$  投影到  $V_1, V_2$  空间的线性权重。参数  $a_1$  和  $a_2$  是平衡标签预测误差和特征重构误差之间的参数。以上模型旨在通过优化  $A, B, C, D, V_1, V_2$ , 使得目标函数最小。

本文首先给出一个关键定理, 该定理是求解模型(2)的基础, 其详细推导过程如下。

$$\begin{aligned} \text{对于任意矩阵 } X, \text{ 有 } \|X\|^2 &= \text{Tr}[XX^T], \text{ 因此可以得到:} \\ J(A, B, C, D, V_1, V_2) &= a_1 \|X_1 - V_1 A\|^2 + a_2 \|X_2 - V_2 B\|^2 + \|Y - V_1 C - \\ & V_2 D\|^2 \\ &= a_1 \text{Tr}[(X_1 - V_1 A)(X_1 - V_1 A)^T] + a_2 \text{Tr}[(X_2 - V_2 B) \\ & (X_2 - V_2 B)^T] + J_p(C, D, V_1, V_2) \end{aligned}$$

定义  $J_p(C, D, V_1, V_2)$  如下:

$$\begin{aligned} J_p(C, D, V_1, V_2) &= \text{Tr}[(Y - V_1 C - V_2 D)(Y - V_1 C - V_2 D)^T] \\ &= \text{Tr}[YY^T + V_1 CC^T V_1^T + V_2 DD^T V_2^T + 2V_2 CD^T V_2^T - \\ & 2V_1 CY^T - 2V_2 DY^T] \end{aligned}$$

令  $J_p$  对  $A, B, C, D$  求偏导并等于 0:

$$\frac{\partial J}{\partial A} = 0 \Rightarrow A = V_1^T X_1$$

$$\frac{\partial J}{\partial B} = 0 \Rightarrow B = V_2^T X_2$$

$$\frac{\partial J}{\partial C} = 0 \Rightarrow 2V_1^T V_1 C - 2V_1^T Y + 2V_1^T V_2 D = 0$$

$$\frac{\partial J}{\partial D} = 0 \Rightarrow 2V_2^T V_2 D - 2V_2^T Y + 2V_2^T V_1 C = 0$$

注:上述公式根据以下矩阵求导公式得到:

$$\frac{\partial}{\partial x} \text{Tr}(AXBX^T C) = A^T C^T X B + C A X B$$

$$\frac{\partial}{\partial x} \text{Tr}(A X B) = A^T B^T$$

因此,推算得到以下公式:

$$C + V_1^T V_2 D = V_1^T Y$$

$$V_2^T V_1 C + D = V_2^T Y$$

即:  $C = V_1^T Y - V_2^T V_1 D$

将  $C$  代入到公式中可得:

$$V_2^T V_1 V_1^T Y - V_2^T Y = V_2^T V_1 V_1^T V_2 D - D$$

(Notes:  $V_2^T V_2 = I$ )

$$\Rightarrow V_2^T (V_1 V_1^T - I) Y = V_2^T (V_1 V_1^T - I) V_2 D$$

$$\Rightarrow Y = V_2 D$$

同理可以得到  $Y = V_1 C$ 。即  $C = V_1^T Y, D = V_2^T Y$ ,用得到的  $A, B, C, D$  最优解替换式(2)中的参数,可以得到:

$$\begin{aligned} J(A, B, C, D, V_1, V_2) &= a_1 \text{Tr}(X_1 X_1^T) + a_1 \text{Tr}(X_2 X_2^T) + \text{Tr}(Y Y^T + V_1 V_1^T Y Y^T \\ &\quad V_2 V_2^T + V_2 V_2^T Y Y^T V_1 V_1^T) - \text{Tr}(V_1^T O_1 V_1) - \\ &\quad \text{Tr}(V_2^T O_2 V_2) \end{aligned}$$

将  $Y = V_1 C, Y = V_2 D$  代入  $V_1 V_1^T Y Y^T V_2 V_2^T$  和  $V_2 V_2^T Y Y^T V_1 V_1^T$  中,由于模型有约束条件  $V_1^T V_1 = I, V_2^T V_2 = I$ ,因此  $V_1 V_1^T Y Y^T V_2 V_2^T = Y Y^T, V_2 V_2^T Y Y^T V_1 V_1^T = Y Y^T$ ,推出原式等价于以下形式:

$$\begin{aligned} J_{\text{opt}} &= \text{Tr}[a_1 X_1 X_1^T + a_2 X_2 X_2^T] + 3 \text{Tr}[Y Y^T] - \\ &\quad \text{Tr}[V_1^T O_1 V_1] - \text{Tr}[V_2^T O_2 V_2] \end{aligned}$$

**定理 1** 给定  $O_1 = a_1 X_1 X_1^T + Y Y^T, O_2 = a_2 X_2 X_2^T + Y Y^T$ ,令  $\lambda_1^1 \geq \dots \geq \lambda_{k_1}^1$  为  $O_1$  的特征向量  $v_1^1, \dots, v_{k_1}^1$  对应的特征值,  $\lambda_1^2 \geq \dots \geq \lambda_{k_2}^2$  为  $O_2$  的特征向量  $v_1^2, \dots, v_{k_2}^2$  对应的特征值。当  $A, B, C, D, V_1, V_2$  为式(2)的最优解时,有:

- 1)  $A = V_1 X_1, B = V_1 X_2, C = V_1^T Y, D = V_2^T Y$ ;
- 2)  $V_1 = [v_1^1, \dots, v_{k_1}^1], V_2 = [v_1^2, \dots, v_{k_2}^2]$ ;
- 3) 在最优解下,式(2)等价于以下形式:  

$$\text{Tr}[a_1 X_1 X_1^T + a_2 X_2 X_2^T] + \text{Tr}[Y Y^T] - \text{Tr}[V_1^T O_1 V_1] - \text{Tr}[V_2^T O_2 V_1]$$

其中,  $x_1, x_2, Y$  为输入数据,因此  $\text{Tr}[a_1 X_1 X_1^T + a_2 X_2 X_2^T] + \text{Tr}[Y Y^T]$  等于常数。因此式(2)等价于:

$$\begin{aligned} \max_{V_1, V_2} &\text{Tr}[V_1^T O_1 V_1] + \text{Tr}[V_2^T O_2 V_2] \quad (3) \\ \text{s. t.} &\begin{cases} V_1^T V_1 = I \\ V_2^T V_2 = I \end{cases} \end{aligned}$$

根据定理 1 可知,原始问题  $V_1, V_2$  的求解可以转换成  $O_1$  和  $O_2$  的特征向量求解问题。因此,式(2)的求解可以通过求解  $O_1$  和  $O_2$  特征向量得到,即转化为求解  $O_1 v_i^1 = \lambda_1 v_i^1, O_2 v_j^2 = \lambda_2 v_j^2$  对应的前  $K_1, K_2$  个最大的特征值对应的特征向量问题。其等价于求解以下模型:

$$\begin{aligned} \max_{v_1, v_2 \in R^M} &v_1^T O_1 v_1 + v_2^T O_2 v_2 \quad (4) \\ \text{s. t.} &\begin{cases} V_1^T V_1 = I \\ V_2^T V_2 = I \end{cases} \end{aligned}$$

在模型(2)的基础上,仍需考虑以下两个问题:1)对于缺少标签信息的文档,无法对其进行情绪预测,因此模型的预测不能依赖标签信息;2)模型的稳健性,即真实数据常常存在过拟合问题。

针对以上问题,本文提出以下假设:  $X_1 \mapsto V_1, X_2 \mapsto V_2$  之间存在线性映射关系,表示如下:

$$V_1 = X_1 W_1, V_2 = X_2 W_2$$

即  $v_i^1 = X_1 w_i^1, v_j^2 = X_2 w_j^2$ ,其中  $i = 1, 2, \dots, K_1, j = 1, 2, \dots, K_2$ 。将线性约束条件  $V_1 = X_1 W_1, V_2 = X_2 W_2$  代入式(4)中,将对  $V_1, V_2$  的优化问题转换成对特征向量  $w^1, w^2$  的优化:

$$\begin{aligned} \max_{w^1, w^2 \in R^M} &(X_1 w^1)^T O_1 (X_1 w^1) + (X_2 w^2)^T O_2 (X_2 w^2) \quad (5) \\ \text{s. t.} &\begin{cases} (X_1 w^1)^T O_1 (X_1 w^1) = I \\ (X_2 w^2)^T O_2 (X_2 w^2) = I \end{cases} \end{aligned}$$

进一步地,为了增加模型的稳健性,在模型(5)的基础上添加正则项  $\|w^1\|^2, \|w^2\|^2$ ;同时将式(4)中的最大化问题转换成最小化  $v_1^T O_1^{-1} v_1 + v_2^T O_2^{-1} v_2$  的问题,以便优化求解,从而得到以下模型:

$$\begin{aligned} \max_{w^1, w^2} &(X_1 w^1)^T O_1 (X_1 w^1) + (X_2 w^2)^T O_2 (X_2 w^2) + \\ &\gamma_1 \|w^1\|^2 + \gamma_2 \|w^2\|^2 \quad (6) \end{aligned}$$

$$\text{s. t.} \begin{cases} (X_1 w^1)^T O_1 (X_1 w^1) = I \\ (X_2 w^2)^T O_2 (X_2 w^2) = I \end{cases}$$

其中,  $\|w_i^1\|^2 = (w_i^1)^T (w_i^1), \|w_j^2\|^2 = (w_j^2)^T (w_j^2)$  为正则项,  $\gamma_1, \gamma_2$  为调节参数。以下推导可以证明正则项  $\|w_i^1\|^2, \|w_j^2\|^2$  可以减少映射函数的模糊性,提高模型的稳健性。

**定理 2** 假设  $w^1, w^2$  为式(6)的特征向量,则  $w^1, w^2$  必为  $x_i^1, x_j^2$  的线性加和:

$$w_1 = X_1^T \beta_1 = \sum_{i=1}^N (\beta_1)_{i,x_1} x_i, w_2 = X_2^T \beta_2 = \sum_{i=1}^N (\beta_2)_{i,x_2} x_i, i = 1, \dots, N$$

采用拉格朗日乘法对式(6)求  $w^1, w^2$  的偏导,并令其等于 0,可得到以下等式:

$$[X_1^T O_1^{-1} X_1 + \gamma_1 I] w^1 = \tilde{\lambda}_1^1 X_1^T X_1 w^1$$

$$[X_2^T O_2^{-1} X_2 + \gamma_2 I] w^2 = \tilde{\lambda}_2^2 X_2^T X_2 w^2$$

通过上式可求解  $w^1, w^2$ ,取最小的  $K_1, K_2$  个特征值对应的特征向量,即为模型的最优解。 $x$  和  $w$  的映射函数如下:

$$\Phi_i(x_1) = w_i^1 x_1, \Phi_j(x_2) = w_j^2 x_2,$$

$$i = 1, 2, \dots, K_1; j = 1, 2, \dots, K_2$$

考虑到求最小特征值时的不稳健性,令  $\lambda_1 = \frac{1}{\tilde{\lambda}_1^1}, \tilde{\lambda}^2 = \frac{1}{\tilde{\lambda}_2^2}$ ,

将求前  $K_1, K_2$  个最小特征值的问题转换成求最大特征值的问题,即:

$$X_1^T X_1 w^1 = \lambda_1 [X_1^T O_1^{-1} X_1 + \gamma_1 I] w^1$$

$$X_2^T X_2 w^2 = \lambda_2 [X_2^T O_2^{-1} X_2 + \gamma_2 I] w^2$$

整个算法的过程如算法 1 所示。

**算法 1** MV-MLSI 算法

输入:  $X_1 \in R^{N \times M_1}, X_2 \in R^{N \times M_2}, \gamma_1, \gamma_2 \geq 0; K_1, K_2 > 0; a_1, a_2 \in R$

输出: 映射函数  $\Phi_i(x_1) = w_i^1 x_1, \Phi_j(x_2) = w_j^2 x_2, i = 1, 2, \dots, k_1, j = 1, 2, \dots, k_2$

步骤 1  $O_1 = a_1 X_1 X_1^T + Y Y^T$

$$O_2 = a_2 X_2 X_2^T + Y Y^T$$

求解特征值:

$$X_1^T X_1 w_1 = \lambda_1 [X_1^T O_1^{-1} X_1 + \gamma_1 I] w_1$$

$$X_2^T X_2 w_2 = \lambda_2 [X_2^T O_2^{-1} X_2 + \gamma_2 I] w_2$$

步骤 2 求解  $w_1^2, w_2^2, i=1, \dots, K_1, j=1, \dots, K_2, \lambda_1^1 \geq \dots \geq \lambda_{K_1}^1, \lambda_1^2 \geq \dots \geq \lambda_{K_2}^2$

由式(2)可知,标签的最优解为  $Y = V_1 C + V_2 D$ ,因此采用  $Y = V_1 C + V_2 D$  的值作为标签的预测结果,其中  $Y \in R^{N \times L}$ 。其中  $y_{ij} > 0$  时表示第  $i$  篇文档有第  $j$  个标签,  $y_{ij} < 0$  时表示第  $i$  篇文档没有第  $j$  个标签。

## 4 实验

### 4.1 数据集

实验数据集来自 Yahoo Qimo 的新闻语料,共有 49000 篇包含各种主题的新闻,每篇包括标题、新闻内容、新闻的评论、投票数及情绪标签所占的比例等数据。由于网络采集数据的局限性,个人行为数据等用户相关信息的获取相对困难,采集到的实验数据仅有新闻内容和评论文本两个视角的信息,因此本文主要基于用户评论内容出发,即从新闻的用户评论中挖掘、理解用户的情绪,从而对新闻自身的文本信息进行补充,再结合投票比例对新闻读者情绪进行分类。

本实验数据中存在少量新闻文本没有投票,部分新闻投票数量较少,考虑到标签的准确性,选用投票数大于 15 的数据进行分析。筛选后,剩余 13233 篇新闻和相应的评论文本。文本特征采用 TF-IDF,并经过卡方特征选择进行初步筛选,最后保留 5595 个新闻文本特征和 6100 个评论文本特征,将文档-词矩阵作为输入特征。各个情绪标签的分布如表 1 所列。

表 1 单个标签下各个标签的文档数目

Table 1 Number of documents for each label under single label

标签	标签数量	标签	标签数量
实用	2325	无聊	1602
感人	394	害怕	221
开心	5495	难过	984
超扯	2197	火大	5988

如表 1 所列,在单个标签情况下,数据的分布很不均匀,存在不平衡现象。例如,标签“火大”和“开心”的语料较多,而标签“感人”和“害怕”的语料不足。根据实验所得,将投票比例阈值设置为 23%,即当标签投票比例为 23%及以上时认为该文档属于这个标签,并将其标记为 1,否则标记为 -1。标签的分布情况如表 2 所列。

表 2 标签数目及比例

Table 2 Label number and ratio

标签数目	文档数	百分比
1	7487	0.57
2	5245	0.40
3	501	0.037

从表 2 可以看出,对于该数据集,情绪标签有两个及两个以上的占整个数据集的 43%,由此可见,将对用户的情绪分析看成多标签问题是合理的。

### 4.2 实验结果

本文实验结果采用多标签分类中常用的评价指标 Macro-F1 和 Micro-F1<sup>[33]</sup>进行衡量。F1 兼顾了 Precision 和 Recall,其中 Macro-F1 为 F1 的平均,而 Micro-F1 则为加权平均。本文同时给出了具体到每个标签的 F1 值并进行评价。实验随机选取数据集的 70%作为训练样本,剩余的 30%作为测试样本。为了保证实验的公平性和准确性,重复 5 次实验再取平均值。

表 3 7 种方法在最优参数下的实验结果

Table 3 Experimental results of seven algorithms under optimal parameters

模型	Macro-F1	Micro-F1
MV-MLSI	<b>0.5498</b>	<b>0.6379</b>
V1-MLSI	0.4505	0.5745
V2-MLSI	0.3485	0.4822
V1V2-MLSI	0.4629	0.5874
LSI-MLKNN	0.0937	0.3080
Original-MLKNN	0.2644	0.4320
LDA-MLKNN	0.0980	0.2701
$\tau\%(\uparrow)/\%$	<b>8.69</b>	<b>5.05</b>

注:粗体表示所有方法中的最好结果;斜体表示基线方法中的最好结果; $\tau\%(\uparrow)$ 表示 MV-MLSI 相对于最好的基线算法的绝对提升比例

实验采用了以下几种方法进行对比。

1)Original-MLKNN:直接将视角一的新闻内容  $X_1$  和视角二的新闻评论特征  $X_2$  拼接为  $X = [X_1, X_2]$ 后作为输入特征,再采用 MLKNN 分类器进行分类。该实验结果为 Baseline。

2)LSI-MLKNN:将两个视角的特征拼接为  $X = [X_1, X_2]$ 后,采用 LSI 对  $X$  降维,再使用 MLKNN 分类器分类。

3)V1-MLSI:MLSI 是文献[23]提出的一种多标签分类方法,单独采用新闻内容特征  $X_1$  并使用 MLSI 分类。

4)V2-MLSI:只采用新闻评论特征  $X_2$ ,用 MLSI 分类器分类。

5)V1V2-MLSI:将两个视角信息拼接融合为  $X = [X_1, X_2]$ ,再用 MLSI 分类器分类。

6)MV-MLSI:本文提出的多视角多标签的分类方法。

7)LDA-MLKNN:将两个视角的特征表示为  $X = [X_1, X_2]$ 后,采用主题模型 LDA<sup>[34]</sup>提取特征,对特征使用 MLKNN 分类器分类。

表 3 给出了 7 种分类方法在最优参数下的 Macro-F1 值和 Micro-F1 值。下面对结果进行详细分析:1)表 3 的结果表明,MV-MLSI 方法在 Macro-F1 指标上比 Original-MLKNN 方法提升了 28.54%,Micro-F1 也提升了 20.59%。而 LSI-MLKNN 和 LDA-MLKNN 方法的分类效果并不理想。由此可见,对于新闻读者情绪分类这一问题,倘若脱离分类标签的指引,即使采用较好的文本特征提取方法并结合较强的多标签分类器,也难以获得预期的结果。2)对比 MV-MLSI 与单视角下 MLSI(见参考文献[23])的分类结果发现,相比于只用新闻视角的 V1-MLSI, MV-MLSI 的 Macro-F1 值增加了 10%,其 Micro-F1 提高了 7%;而相比于只用新闻评论视角信息的 V2-MLSI, MV-MLSI 的 Macro-F1 值提升了 20%,其 Micro-F1 提升了 15.57%。由此可见,采用单视角的信息时,

对新闻读者情绪分类的预测准确度十分有限。3)进一步对比在标签指引下获得视角特征并进行特征拼接的方法 V1V2-MLSI 发现, MV-MLSI 比 V1V2-MLSI 在 Macro-F1 指标上提高了 8.69%, 在 Micro-F1 上则提高了 5.05%, 提

升效果依然相当可观。以上说明, 本文提出的 MV-MLSI 既能有效融合视角特征, 又能有效利用多标签的指引信息, 是一种有效的适用于新闻读者情绪分类的多视角多标签学习算法。

表 4 7 种方法下每个标签的 F1 值的对比

Table 4 F1-score comparison for each label under seven algorithms

标签	实用	感动	开心	超扯	无聊	害怕	难过	火大
MV-MLSI	<b>0.4499</b>	<b>0.5758</b>	<b>0.6920</b>	<b>0.4598</b>	<b>0.4420</b>	<b>0.4413</b>	<b>0.5749</b>	<b>0.7620</b>
V1-MLSI	0.3940	0.4021	0.6537	0.3321	0.3710	0.2620	0.4491	0.7191
V2-MLSI	0.3112	0.2505	0.5223	0.3179	0.3046	0.1528	0.3201	0.6083
V1V2-MLSI	0.3992	0.4040	0.6526	0.3773	0.3947	0.2743	0.4682	0.7327
LSI-MLKNN	0.0567	0.0000	0.5274	0.0075	0.0007	0.0000	0.0000	0.1573
Original-MLKNN	0.1795	0.1182	0.5092	0.1088	0.2519	0.2094	0.1525	0.5854
LDA-MLKNN	0.0282	0.0000	0.4682	0.0313	0.0566	0.0000	0.0232	0.1766
$\tau\%(\uparrow)/\%$	5.07	17.18	3.83	8.25	4.73	16.7	10.67	2.93

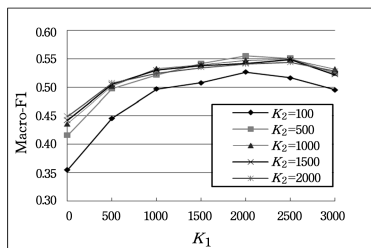
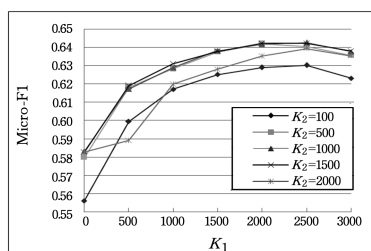
注: 粗体表示所有方法中的最好结果;  $\tau\%(\uparrow)$  表示 MV-MLSI 相对于最好的基线算法的绝对提升比例

表 4 列出了具体到每个标签在最优参数时的 F1 值。其中, 各个标签的 F1 值存在较大差异, 这是由数据不平衡所致。从表 4 可以看出, MV-MLSI 方法在 8 个情绪标签上都提升, 特别是在语料相对缺乏的“感人”“害怕”等标签。其效果更优。相比于其他方法, 尤其相对于 Original-MLKNN, LSI-MLKNN 和 LDA-MLKNN 方法, 本文方法有较大提升, 部分标签提升近 46%。通过实验发现, MLKNN 分类器对正样本过少的标签的分类效果较差; 而 MV-MLSI 在正样本语料缺乏的情况下也能取得不错的效果。

### 4.3 参数分析

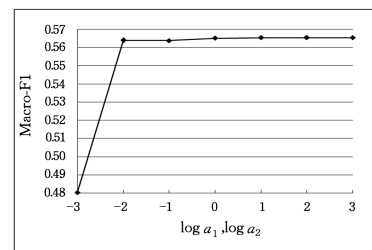
为了考查模型中参数对实验的影响, 分别给出了  $K_1$ ,  $K_2$ ,  $a_1$ ,  $a_2$  以及  $\gamma$  在不同取值下的实验结果。

1) 参数  $K_1, K_2$ 。 $K_1, K_2$  的取值即视角一和视角二的特征在低维空间隐语义映射的维度。图 1、图 2 分别为 Macro-F1, Micro-F1 在  $K_1, K_2$  取不同值时的变化曲线。如图 1 所示, 当  $K_1=2000, K_2=1500$  时, 模型的效果最好。

图 1 不同  $K_1, K_2$  下的 Macro-F1 值Fig. 1 Macro-F1 value under different  $K_1$  and  $K_2$ 图 2 不同  $K_1, K_2$  下的 Micro-F1 值Fig. 2 Micro-F1 value under different  $K_1$  and  $K_2$ 

2) 参数  $a_1, a_2$ 。 $a_1, a_2$  是平衡标签预测误差和特征重构误差之间的参数。图 3 为本实验数据集  $a_1, a_2$  取多组值时的 Macro-F1 的变化曲线, 其中  $a_1, a_2$  的变化范围为  $0.001 \sim$

1000。可以看出, 当  $a_1, a_2$  的取值大于 0.01 时, 模型性能趋于稳定。对于不同的数据集,  $a_1, a_2$  的取值差异较大。

图 3 不同  $a_1, a_2$  下的 Macro-F1 值Fig. 3 Micro-F1 value under different  $a_1$  and  $a_2$ 

3)  $\gamma$  为控制内核组合权重和正规化的强度。本实验中  $\gamma$  取值 0.01。

**结束语** 本文深入研究读者情绪分类问题, 针对真实的情绪文本复杂多样、具有多视角多标签的特性, 在 LSI 模型的基础上提出了一种改进的模型——MV-MLSI。将不同视角下的文本特征映射到低维语义空间, 构建了有监督的多视角融合的多标签分类模型; 同时建立了特征与标签之间的映射函数, 在新闻语料的读者情绪分类中取得了较好的结果。本文所提出的模型可以较容易地推广到多于两个视角的情况, 在将来的工作中, 还可以在特征  $X$  和  $V$  的映射假设中做进一步改进, 即  $X$  与  $V$  的关系将不局限于线性映射。

### 参考文献

- [1] TANG D, QIN B, LIU T. Document Modeling with Gated Recurrent Neural Network for Sentiment Classification[C]// Conference on Empirical Methods in Natural Language Processing. 2015:1422-1432.
- [2] LIU H, LI S, ZHOU G, et al. Joint Modeling of News Reader's and Comment Writer's Emotions[C]// Meeting of the Association for Computational Linguistics. 2013:511-515.
- [3] CHEN J, LIU H, LI S, et al. Active Learning on News Reader Emotion Classification[J]. Journal of Shanxi University (Natural Science Edition), 2017, 40(1): 21-26. (in Chinese) 陈敬, 刘欢欢, 李寿山, 等. 基于主动学习的新闻读者情绪分类方法[J]. 山西大学学报(自然科学版), 2017, 40(1): 21-26.
- [4] EKMAN P. Facial expression and emotion[J]. American psychologist, 1993, 48(4): 384-392.
- [5] BHOWMICK P K. Reader perspective emotion analysis in text

- through ensemble based multi-label classification framework [J]. *Computer and Information Science*, 2009, 2(4): 64.
- [6] YU H, HATZIVASSILOGLOU V. Towards answering opinion questions; Separating facts from opinions and identifying the polarity of opinion sentences[C]// *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*. Association for Computational Linguistics, 2003: 129-136.
- [7] KIM S M, HOVY E. Automatic detection of opinion bearing words and sentences[C]// *Companion Volume to the Proceedings of the International Joint Conference on Natural Language Processing (IJCNLP)*. 2005: 8.
- [8] HU M, LIU B. Mining and summarizing customer reviews[C]// *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2004: 168-177.
- [9] TURNEY P D. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews[C]// *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 2002: 417-424.
- [10] NIU Y, PAN M H, WEI O, et al. Emotion Analysis of Chinese Microblogs Using Lexicon-based Approach[J]. *Computer Science*, 2014, 41(9): 253-258. (in Chinese)  
牛耘, 潘明慧, 魏欧, 等. 基于词典的中文微博情绪识别[J]. *计算机科学*, 2014, 41(9): 253-258.
- [11] WANG M, LIU M, FENG S, et al. A novel calibrated label ranking based method for multiple emotions detection in Chinese microblogs[M]// *Natural Language Processing and Chinese Computing*. Springer Berlin Heidelberg, 2014: 238-250.
- [12] PANG B, LEE L, VAITHYANATHAN S. Thumbs up?: sentiment classification using machine learning techniques[C]// *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing-Volume 10*. Association for Computational Linguistics, 2002: 79-86.
- [13] YE L. Research on emotion prediction of news article from reader's perspective[D]. Harbin: Harbin Institute of Technology, 2013. (in Chinese)  
叶璐. 新闻文本的读者情绪自动预测方法研究[D]. 哈尔滨: 哈尔滨工业大学, 2013.
- [14] WEN W, WU B, CAI R C, et al. Emotion classification for news readers based on multi-category semantic word clusters[J]. *Journal of Computer Applications*, 2016, 36(8): 2076-2081. (in Chinese)  
温雯, 吴彪, 蔡瑞初, 等. 基于多类别语义词簇的新闻读者情绪分类[J]. *计算机应用*, 2016, 36(8): 2076-2081.
- [15] XU R, ZOU C, XU J, et al. Reader's Emotion Prediction Based on Partitioned Latent Dirichlet Allocation Model[C]// *Proceedings of International Conference on Internet Computing and Big Data*. 2013.
- [16] LUYCKX K, VAASSEN F, PEERSMAN C, et al. Fine-grained emotion detection in suicide notes: A thresholding approach to multi-label classification [J]. *Biomedical Informatics Insights*, 2012, 5(Suppl 1): 61-69.
- [17] LIU H H. Automatic Classification on Reader's Emotion towards News[D]. Suzhou: Soochow University, 2015. (in Chinese)  
刘欢欢. 面向新闻的读者情绪自动分析方法研究[D]. 苏州: 苏州大学, 2015.
- [18] YIN H, XU J H, XU H. A Multi-Label Classification Algorithm Bases on LS-SVM[J]. *Journal of Nanjing Normal University (Engineering and Technology)*, 2010, 10(2): 68-73. (in Chinese)  
殷会, 许建华, 许花. 基于 LS-SVM 的多标签分类算法[J]. *南京师范大学学报(工程技术版)*, 2010, 10(2): 68-73.
- [19] SHEN C B, WANG Z H, SUN Y G. A Multi-Label Classification Algorithm Based on Label Clustering[J]. *SoftWare*, 2014, 35(8): 16-21. (in Chinese)  
申超波, 王志海, 孙艳歌. 基于标签聚类的多标签分类算法[J]. *软件*, 2014, 35(8): 16-21.
- [20] ZHANG M L, ZHOU Z H. ML-KNN: A lazy learning approach to multi-label learning[J]. *Pattern Recognition*, 2007, 40(7): 2038-2048.
- [21] CABRAL R S, TORRE F, COSTEIRA J P, et al. Matrix completion for multi-label image classification[C]// *Advances in Neural Information Processing Systems*. 2011: 190-198.
- [22] WANG H, DING C, HUANG H. Multi-label linear discriminant analysis[C]// *European Conference on Computer Vision-ECCV 2010*. 2010: 126-139.
- [23] YU K, YU S, TRESP V. Multi-label informed latent semantic indexing[C]// *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 2005: 258-265.
- [24] ZHANG M L, WU L. LIFT: Multi-label learning with label-specific features[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 37(1): 107-120.
- [25] GOLDBERG A, RECHT B, XU J, et al. Transduction with matrix completion; Three birds with one stone[C]// *Advances in Neural Information Processing Systems*. 2010: 757-765.
- [26] LIU M, LUO Y, TAO D, et al. Low-Rank Multi-View Learning in Matrix Completion for Multi-Label Image Classification[C]// *29th AAAI Conference on Artificial Intelligence*. 2015: 2778-2784.
- [27] XIA R, PAN Y, DU L, et al. Robust Multi-View Spectral Clustering via Low-Rank and Sparse Decomposition[C]// *28th AAAI Conference on Artificial Intelligence*. 2014: 2149-2155.
- [28] CHAUDHURI K, KAKADE S M, LIVESCU K, et al. Multi-view clustering via canonical correlation analysis[C]// *Proceedings of the 26th Annual International Conference on Machine Learning*. ACM, 2009: 129-136.
- [29] XIA T, TAO D, MEI T, et al. Multiview spectral embedding[J]. *IEEE Transactions on Systems Man and Cybernetics Part B*, 2010, 40(6): 1438-1446.
- [30] GREENE D, CUNNINGHAM P. A matrix factorization Approach for integrating multiple data views[C]// *Proceedings of the European Conference on Machine Learning and Knowledge Discovery in Databases*. 2009: 423-438.
- [31] XU C, TAO D, XU C. A survey on multi-view learning [J]. *arXiv preprint arXiv:1304.5634*, 2013.
- [32] SUN S. A survey of multi-view machine learning[J]. *Neural Computing and Applications*, 2013, 23(7/8): 2031-2038.
- [33] 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.
- [34] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation [J]. *Journal of Machine Learning Research*, 2003, 3(1): 993-1022.