

DOI:10.19651/j.cnki.emt.2106165

基于深度神经网络和模糊规则的文本分类方法^{*}

王建华 冉煜琨

(成都理工大学工程技术学院 乐山 614000)

摘要:传统的单个模糊分类器方法采用固定的去模糊化规则,在情感数据分类上容易引起文本歧义,针对该问题,提出一种基于深度神经网络和模糊规则的文本分类方法。该方法分为两个主要阶段。第1个阶段,利用模糊规则形成算法与不同的模糊范数,利用特征提取方法(词袋法、词嵌入向量),以及基于关联的特征子集选择方法来制备特征,从而训练多个模糊分类器;然后,进行分类器融合以识别歧义实例。第2个阶段,整理歧义实例,生成第2个训练集,使用KNN对新出现的歧义实例进行分类。与当前已有的先进方法相比,所提方法在大部分情况下具有更优的分类性能,Wilcoxon秩检验的统计具有显著性。

关键词:文本分类;模糊分类器;文本歧义;模糊规则;神经网络;Wilcoxon秩检验

中图分类号: TP391.1 **文献标识码:** A **国家标准学科分类代码:** 520.6020

Research of text classification method based on deep neural network and fuzzy rule

Wang Jianhua Ran Yukun

(The Engineering & Technology College of Chengdu University of Technology, Leshan 614000, China)

Abstract: The traditional single fuzzy classifier method, which uses fixed de-fuzzification rules, is easy to case the problem of text ambiguity in emotional data. To solve the problem, a text classification method based on deep neural networks and fuzzy rules is proposed. The method is divided into two main stages. In the first stage, we use fuzzy rule formation algorithm and different fuzzy norms, feature extraction method (word bag method, word embedding vector), and feature subset selection method based on association to prepare features, so as to train multiple fuzzy classifiers. Then, we fuse classifiers to identify ambiguous instances. In the second stage, disambiguation instances are sorted out, a second training set is generated, and KNN is used to classify new disambiguation instances. Compared with the current advanced methods, the proposed method has better classification performance in most cases, and the Wilcoxon rank test is statistically significant.

Keywords: text classification; fuzzy classifier; disambiguation instances; fuzzy rule; neural networks; Wilcoxon rank test

0 引言

文本分类^[1]技术已经发展了几十年,随着机器学习和模糊理论的快速发展与完善,其逐渐成为文本分类的强大工具,已被广泛用于各种文本分类应用领域中,例如电影评论^[2]、网络仇恨言论识别^[3],以及针对船舶新闻的文本分类^[4]等。但模糊性和不确定性的文本处理依然是个具有挑战性的任务,例如,对于含有正面和负面因素的文本,人类很容易判别其正面与否,但判别式算法较难处理。

按照目前的研究成果,文本分类方法可以分为3类:一般特征提取方法、机器学习方法、模糊分类方法。

一般特征提取方法一般包括词频-逆文档频率^[5](TF-IDF)、词袋^[6](bag of words, BoW)模型。这些方法通过词频或词向量的形式对文本进行分类,其层次较浅,局限性较大。

机器学习方法在特征提取的基础上,通过分类器的学习和归纳,具有较高的分类准确性。如文献[7]提出了基于多层次SVM的分类方法。该方法准确度非常高,但处理时间比其他标准学习方法更长。此外,研究人员还使用了其他学习算法,如朴素贝叶斯^[8](naive bayes, NB),逻辑回归^[9](logistic regression, LR)等。近些年,深度学习方法被广泛用于特征提取和分类器训练。如文献[10]使用卷积

收稿日期:2021-03-26

^{*} 基金项目:四川省重点实验室开放基金重点项目(scsxdz2019zd01)资助

神经网络(convolutional neural network, CNN)进行文本分类,使用不同类型的词向量作为特征,采用双向门控循环单元进行池化。文献[11]采用多个深度神经网络(deep neural network, DNN)架构学习分类器训练时所使用的词向量,结果表明,嵌入特征的分类性能优于 BoW。

模糊分类以模糊逻辑为基础^[12],旨在处理实例中的语言不确定性。这种情况下,每个实例并不明确,在不同分类中有着不同的隶属度。如文献[13]提出的模糊方法自动构建语料库以进行文本相似度比较。结果表明,与传统度量相比,模糊度量与手动标注的相关性更高。文献[14]提出了 3 层情感扩散模型,用于确定情感分类的模糊隶属度,结果表明该方法在 7 个情感预测数据集上的均方误差小于 SVM 和其他方法,同时取得与其他机器学习算法相似的性能。

为了更好地处理歧义或者不确定性文本,提出了一种基于模糊规则和 DNN 实例学习的文本分类方法,不同于传统模糊方法单独使用每个分类器进行分类,所提方法对不同参数设置生成的多个模糊分类器进行融合(合并每个分类的隶属度)。且使用半固定的去模糊化规则,根据该实例与训练集其他歧义实例的余弦相似度进行分类。实验结果验证了所提方法对网络言论分类的适用性,性能较优。

1 模糊逻辑与规则

模糊逻辑是确定性逻辑的扩展,采用从 0~1 的连续真实值,而非二进制真实值(0 或 1)。模糊集中,每个元素 e_1, e_2, \dots, e_n 均对集合 A 有一定的隶属度,隶属度的数值取决于针对模糊集 A 而定义的隶属函数 μ_A ,其中 $\mu_A(e_i)$ 表示元素 e_i 在模糊集 A 中的隶属度, $\mu_A(e_i) \in [0, 1], 1 \leq i \leq n$ 。

在基于模糊规则的系统中,训练阶段的主要操作是对连续(数值)属性进行模糊化。可通过将每个数字属性转换为若干定性属性来完成上述操作。例如被定义为数字属性的“年龄”,可转换为 3 个定性属性:“青年”,“中年”和“老年”。每个定性属性均被视为 1 个模糊集,利用模糊隶属函数定义,因此每个定性属性的域必须为 $[0, 1]$ 。

隶属函数可为不同形状,包括梯形、三角形和高斯形状等。定义隶属函数的目的是估计其参数。梯形隶属函数包含 4 个参数(a, b, c, d),这 4 个参数决定着定性属性 A 中的数字的隶属度:

$$f_A(x) = \begin{cases} 0, & x \leq a \text{ || } x \geq d \\ (x-a)/(b-a), & a < x < b \\ 1, & b \leq x \leq c \\ (d-x)/(d-c), & c < x < d \end{cases} \quad (1)$$

梯形隶属函数可被视为三角和矩形隶属函数的推广。若 $b=c$,则隶属函数的形状为三角形。同理,若 $a=b$ 且 $c=d$,则隶属函数的形状为矩形。文献[15]将区间(a, b)或(c, d)称为支持区,代表一个元素对某个集合具有隶属度的软边界;区间 $[b, c]$ 称为核心区,代表一个元素完全属于某个集合的硬边界。

现实应用中,可通过专家系统,或通过对数据的统计学习,估计出隶属函数的参数。后者适用于高维数据。一般来说,基于模糊规则的系统可表现为如下形式。

- 1) 规则 1(r_1):若 x_1 为 A_{11}, x_2 为 A_{12} ,依次类推, x_n 为 A_{1n} ,则分类为 C_1 。
- 2) 规则 2(r_2):若 x_1 为 A_{21}, x_2 为 A_{22} ,依次类推, x_n 为 A_{2n} ,则分类为 C_2 。
- 3)...
- 4)...
- 5) 规则 m (r_m):若 x_1 为 A_{m1}, x_2 为 A_{m2} ,依次类推, x_n 为 A_{mn} ,则分类为 C_q 。

为使用上述模糊规则(r_1, r_2, \dots, r_m)对新实例 v_i 分类,必须确认在每个维度 x_j 中 v_i 对模糊集 A_{ij} 的隶属度。其后,通过合并 v_i 对所有与规则 r_i 的前因相关的模糊集($A_{i1}, A_{i2}, \dots, A_{in}$)隶属度,计算出每个规则 r_i 的触发强度。该计算可使用三角模 $T(\cdot) \in [0, 1]$ 计算,例如:

$$T(\mu_{A_{i1}}(x_1), \mu_{A_{i2}}(x_2), \dots, \mu_{A_{in}}(x_n)) = \min_{j=1}^n \{\mu_{A_{ij}}(x_j)\} \quad (2)$$

此外,通过合并 C_k 的所有规则的触发强度,计算出对每个分类 C_k 的整体隶属度。该计算可使用三角余模,最后,通过分配具有最大隶属度的分类,对新实例进行分类。

2 提出的模糊分类方法

传统的基于模糊规则的系统^[15]通常基于预定义隶属度或关联规则对每个数字属性进行去模糊化,即针对类别区分为所有数字属性预定义分区。此类模糊方法在处理例如文本等高维数据时表现不佳。同时,仇恨言论数据通常是不平衡的,其中少数类的重要性很高。针对此问题,本文提出的用于网络言论检测的模糊方法在训练阶段包含两个阶段,如图 1 所示。第 1 个阶段,利用混合模糊规则形成算法与不同的模糊范数,训练多个模糊分类器,其后进行分类器融合以识别歧义实例。第 2 个阶段,整理歧义实例,生成第 2 个训练集,从而使用 DNN 对新出现的歧义实例进行分类。图 1 中, i, c 和 n 分别表示实例、分类和分类器的索引。

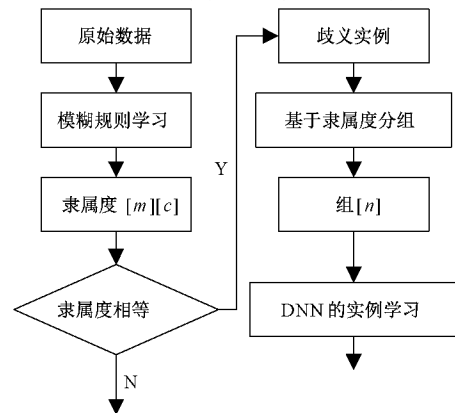


图 1 提出的文本分类学习框架

混合模糊规则形成算法的程序包括以逐个实例的方式顺序生成新规则并修改已有规则,即检查每个实例,并向规则集添加新规则或修改已有规则。整个程序中,每个规则 r_i 涉及到 n 个隶属函数(规则前因部分中的 n 个维度)和两个额外参数 w 和 λ ,其中 w 表示规则 r_i 覆盖的实例数量, λ 为记忆触发该规则 r_i 生成的原始实例的“锚”。这个过程有点类似模糊C均值^[16]的样本分类模式,通过目标函数获得类中心隶属度,从而决定样本类属。

在检查每个实例 $x_i \in C_k$ 后,存在3种可能情况,即覆盖、提交和缩小。

1)覆盖:若实例 x_i 处于已有规则 r_i 所覆盖的支持区域内,即实例 x_i 在每个维度 j 中的数值 v_{ij} 对相应的隶属函数 $\mu_{A_{ij}}$ 存在非零隶属度,则判定实例 x_i 被规则 r_i 所覆盖。若实例 x_i 不在规则 r_i 的核心区域中,则需要对规则 r_i 在每个维度 j 中的核心区域 $[b_{ij}, c_{ij}]$ 进行调整,以使 x_i 位于核心区域中。经过上述调整后,实例 x_i 将完全隶属于规则 r_i 。

2)提交:若实例 $x_i \in C_k$ 未被规则集中分类 C_k 的任何已有规则所覆盖,则判定满足提交情况。这种情况下,需要生成新的规则并添加到规则集,实例 x_i 将被记录为该规则的锚 λ 。根据实例 x_i 在每个维度 j 中的值 v_{ij} ,对新规则在每个维度 j 中的核心区域进行初始化,例如,若 $v_{i1} = 2$,则将核心区域初始化为 $[2, 2]$ 。

3)缩小:检查新实例 $x_i \in C_k$ 后,一旦添加新规则(提交案例)或修改已有规则(覆盖案例),有必要触发第3个案例(缩小)以避免分类冲突,即检查是否存在类别 $C_l \neq C_k$ 的任何已有规则覆盖实例 x_i 。若出现冲突,则需要调整类别 $C_l \neq C_k$ 的规则,以使得实例 x_i 对该规则无隶属度。为修改类别 $C_l \neq C_k$ 的规则 r_l 以避免分类冲突,需要确认实例 x_i 是否仅位于支持区域中(在核心区域之外)。若如此,则可对规则 r_l 进行调整,且不会丢失类别 C_l 的覆盖实例。在调整隶属函数时,仅需要考虑实例 x_i 落入规则 r_l 的支持区域的维度。在这些需要考虑的维度中,仅选择会造成最小容量损失的维度进行调整:

$$j_{\min} = \operatorname{argmin}_{j=1}^n \{V_j\} \quad (3)$$

容量损失可计算为:

$$V_j = d_j^*(x_i, r_i) \cdot \prod_{h=1, h \neq j}^n d_h^\times(x_i, r_i) \quad (4)$$

可以使用一些缩小启发式来测量该容量损失,例如基于规则的缩小(式(5)和(6)),基于锚的缩小(式(5)和(7)),以及基于边界的缩小(式(5)和(8))。

$$d_j^*(x_i, r_i) = \begin{cases} v_{ij} - a_{ij}, v_{ij} \leq \lambda_{ij} \\ d_{ij} - v_{ij}, v_{ij} > \lambda_{ij} \end{cases} \quad (5)$$

$$d_h^\times(x_i, r_i) = d_{ij} - a_{ij} \quad (6)$$

$$d_j^\times(x_i, r_i) = \begin{cases} \lambda_{ij} - a_{ij}, v_{ij} \leq \lambda_{ij} \\ d_{ij} - \lambda_{ij}, v_{ij} > \lambda_{ij} \end{cases} \quad (7)$$

$$d_j^\times(x_i, r_i) = \begin{cases} b_{ij} - a_{ij}, v_{ij} \leq \lambda_{ij} \\ d_{ij} - c_{ij}, v_{ij} > \lambda_{ij} \end{cases} \quad (8)$$

式中: j_{\min} 为指向最小容量损失的维度的索引; V_j 为实际容量损失量; $d_j^*(\cdot)$ 为容量损失函数; $d_j^\times(\cdot)$ 为容量损失加权函数; v_{ij} 为 x_i 在维度 j 中的值; λ_{ij} 为规则 r_i 的锚在维度 j 中的值; a_{ij} 和 d_{ij} 分别为规则 r_i 在维度 j 中的支持区域的左边界和右边界; b_{ij} 和 c_{ij} 分别为规则 r_i 在维度 j 中的核心区域的左边界和右边界。

当实例 $x_i \in C_k$ 处于类别 $C_l \neq C_k$ 的规则 r_l 的核心区域中,则需要对该规则的支持区域和核心区域均进行调整,以使得实例 x_i 不隶属于规则 r_l ,因此无法在不损失 C_l 覆盖实例的情况下修改规则。这种情况下,容量损失的测量函数 $d_j^*(\cdot)$ 和加权函数 $d_j^\times(\cdot)$,以及缩小启发式均需要进行修改:

$$d_j^*(x_i, r_i) = \begin{cases} v_{ij} - b_{ij}, v_{ij} \leq \lambda_{ij} \\ c_{ij} - v_{ij}, v_{ij} > \lambda_{ij} \end{cases} \quad (9)$$

$$d_j^\times(x_i, r_i) = c_{ij} - b_{ij} \quad (10)$$

$$d_j^\times(x_i, r_i) = \begin{cases} \lambda_{ij} - b_{ij}, v_{ij} \leq \lambda_{ij} \\ c_{ij} - \lambda_{ij}, v_{ij} > \lambda_{ij} \end{cases} \quad (11)$$

如式(10)所示,基于规则的缩小和基于边界的缩小是一样的。

按照混合模糊规则形成算法的程序,训练一组模糊规则。由于文本歧义性,使用固定规则(选择隶属度最大的类别)对新实例进行分类会造成误差增加。为此,本文利用不同的模糊范数训练多个模糊分类器,以鼓励模糊分类器之间的多样化,使用混合模糊规则形成算法时,不同的模糊范数会对模糊分类器的训练产生不同影响,使这些模糊分类器导向不同的歧义实例集合^[17]。由此,这些模糊分类器的融合有助于减少训练阶段中歧义实例的数量,并在测试阶段可能会消除对两个类别有着相等隶属度的新出现实例的歧义。

但模糊分类器的融合并不能确保消除所有歧义实例,因此,本文采用基于实例的学习方法,以余弦相似度和K最近邻(KNN)为基础。在多维向量空间中收集所有歧义实例,并在相同特征空间中划分新的歧义实例。向新出现的实例分配“最近邻”类标签(即与新出现的实例匹配度最高的特征)。如图2所示的样例,5个已标注实例被选为新出现样例的最近邻,由此共享最相似的特征,而大多数(4个已标注实例)隶属于正面类,1个为负面类,因此新出现样例被分配到正面类别。这种情况下,需要基于数据集,确定用于新出现实例分类的最近邻数量(通常为奇数,以避免平局),以及是否需要处理类别不平衡问题。

3 实验结果与分析

本节将对网络仇恨言论分类进行实验分析,涉及4类仇恨言论(宗教、种族、残疾和性取向)。对于每个数据集,分别利用BOW和Doc2vec制备4个不同特征集,并利用

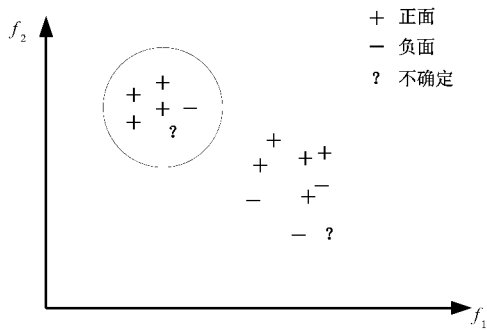


图 2 基于实例的推理样例

基于相关性的特征子集选择方法^[16]进行特征选择。同时,使用提出的模糊方法对分类器进行训练,并从精度、召回率和 F 得分方面,将提出的方法与 SVM^[7]、NB^[8]、DNN^[11]以及其他模糊方法相比较。

3.1 数据集

使用基于命名实体(事件焦点)的搜索术语进行数据采集。这些术语包括很多对事件的引用和围绕事件的主要标签。每个事件生成的数据集包含 300 000~1 200 000 个文本,本文从中随机采样 2 000 个文本以供人工编码。按照文献[19],数据集通过人工注释得到 4 个标准数据集:宗教,1 901 条微博,其中 222 个实例包含攻击性或敌意内容(占已注释样本 11.68%);种族,1 876 条微博,70 个实例包含攻击性或敌意内容(占已注释样本 3.73%);残疾,1 914 条微博,其中 51 个实例包含攻击性或敌意内容(占已注释样本 2.66%);以及性取向,1 803 条微博,183 个实例包含攻击性或敌意内容(占已注释样本 10.15%)。后将包含攻击性或敌意内容的实例称为网络仇恨言论,这部分文本在整体样本中占比相对较小。这些文本在针对每个事件的完整数据集中随机采样获得,具有代表性。

3.2 预处理与参数设置

文本处理方面,移除停用词、标点符号和数字,提取剩余词的词干。然后,在 BOW 特征提取前,计算每个词的频率(TF 和 IDF)。仅选择满足预定义的最小绝对 TF 的特征用于分类器训练,并使用 TF-IDF 作为每个特征的值。一般将最小 TF 设为 2,但针对宗教相关数据,将该频率设为 5,以避免维数过高。

对于 Doc2vec 特征提取,文本预处理操作程序与 BOW 特征提取相同,但排除词干化步骤,以避免用于文本识别的语义信息丢失。使用 DBOW,通过预测给定目标词的上下文词,学习词和文档(微博文本)的表征,其中学习率设为 0.025,上下文窗口大小为 2,以更有效地处理小型语料库中的短文本。所有词(待转换为向量)均需要满足数值为 2 的最小绝对 TF。在经过 50 代的嵌入向量训练(每批词数 10 000)后,每个数据集中所有 n 条微博最终转换为 n 个文档向量,维数(特征)为 100。

对于使用 BOW 和 Doc2vec 从每个数据集提取出的两

个特征集,分别单独使用基于相关性的特征子集选择方法,以减少维数,得到两个较小的特征集。由此针对每个数据集准备共计 4 个特征集,下文给出了这些特征集的维数。其中,BOW(全集)表示将 BOW 提取出的全部特征集用于分类器训练,BOW(子集)表示仅使用 BOW 特征的选定子集。对通过 Doc2vec 提取出的嵌入向量也做同样处理,以区分使用全部特征集和选定特征子集的情况。

机器学习阶段,使用 C4.5 算法以及减少误差修剪法训练分类器。使用线性核训练 SVM 分类器。对于 DNN 训练,为适应较小数据集,选择包含两个全连接层,每层 100 个单元的网络架构。使用 MSE 损失函数和线性修正线性单元(ReLU)激活函数(学习率 0.01)训练分类器,使用随机梯度下降,对 DNN 参数进行 20 代的优化(批大小为 500)。

在通过 KNN 进行基于实例的推理时,K 值通常设为 5,除非不需要基于实例的推理阶段,即模糊分类器学习和融合后没有出现歧义实例的情况,此时将 K 值设为 0,即整个程序中不使用 KNN。

3.3 结果分析

为表示方便,将 M1 和 M2 分别表示使用模糊范数 min/max 和 Yager 对模糊分类器进行训练。M3 表示上述 2 个模糊范数训练的 2 个模糊分类器的融合案例。M4 表示提出的模糊方法,其中利用 M3 进行分类器融合,并通过 KNN,使用余弦相似度进行基于实例的推理。

表 1 所示为模糊分类上的结果,表 1 结果表明,使用不同模糊范数训练的模糊分类器的融合,一般能够有效提升分类性能,且采用 KNN 进行基于实例的推理,可在大部分情况下进一步改善性能。

如表 2 所示,对于从“种族”数据集提取的选定特征子集,M4 性能优于 M3,因为在模糊分类器融合后,没有剩余的歧义实例。换言之,M3 将歧义实例数量成功降为 0,则 M4 已没有进一步改善的空间。表 3 结果表明,模糊分类器融合一般能够减少歧义实例数量。最差情况下,歧义实例数量也不会高于使用单个模糊分类器后的最大歧义实例数量。

使用 BOW 进行特征提取时,歧义实例数量显著增加,向 M4 中基于实例的推理提供了更大的改善空间。歧义实例数量的增加可能源自于维数约简,增加了每个规则的触发强度。例如当维数被大幅减少时,使用三角模的 min 函数,规则前因部分中一个维度所得到的最小隶属度可能会变大。采用基于相关性的特征子集选择,目的是减少特征到特征的相关性,并增加特征到类别的相关性,因此选定特征被考虑为与类别高度相关,移除的特征则考虑为冗余或不相关特征。这种情况下,若 1 个实例在 1 个特征维度中得到高隶属度,则其很可能在其他相关的特征维度中也得到高隶属度。通过使用三角模,1 个规则的触发强度不高于规则前因部分中的特征维度中的最小隶属度。同时,

表1 模糊分类性能

方法		宗教			种族			残疾			性取向		
特征	学习	P	R	F	P	R	F	P	R	F	P	R	F
BOW 全集	M1	0.779	0.491	0.602	0.644	0.543	0.589	0.536	0.588	0.561	0.655	0.311	0.422
	M2	0.767	0.518	0.618	0.694	0.486	0.571	0.517	0.588	0.550	0.642	0.333	0.439
	M3	0.771	0.532	0.629	0.649	0.529	0.583	0.527	0.569	0.547	0.597	0.404	0.482
	M4	0.679	0.667	0.673	0.627	0.671	0.648	0.517	0.608	0.559	0.519	0.525	0.522
BOW 子集	M1	0.986	0.324	0.488	0.857	0.343	0.491	1.000	0.569	0.725	0.890	0.355	0.508
	M2	0.987	0.342	0.508	0.929	0.371	0.531	1.000	0.569	0.725	0.926	0.344	0.502
	M3	0.988	0.383	0.552	0.889	0.457	0.604	1.000	0.569	0.725	0.878	0.393	0.543
	M4	0.878	0.617	0.725	0.877	0.714	0.787	0.912	0.608	0.729	0.875	0.536	0.664
Doc2vec 全集	M1	0.694	0.387	0.497	0.880	0.314	0.463	0.906	0.569	0.699	0.461	0.404	0.427
	M2	0.931	0.243	0.386	0.917	0.314	0.468	0.933	0.549	0.691	0.456	0.284	0.350
	M3	0.778	0.378	0.509	0.926	0.357	0.515	0.967	0.569	0.716	0.744	0.333	0.46
	M4	0.748	0.401	0.522	0.933	0.410	0.560	0.969	0.608	0.747	0.688	0.350	0.464
Doc2vec 子集	M1	0.577	0.356	0.440	0.756	0.443	0.559	0.935	0.569	0.707	0.449	0.454	0.451
	M2	0.804	0.333	0.471	0.737	0.400	0.519	0.879	0.569	0.901	0.468	0.404	0.434
	M3	0.839	0.351	0.495	0.861	0.443	0.585	0.935	0.569	0.707	0.848	0.366	0.511
	M4	0.835	0.365	0.508	0.861	0.443	0.585	0.939	0.608	0.738	0.789	0.388	0.520

表2 歧义实例数量

方法		宗教		种族		残疾		性取向	
特征	学习	总数	仇恨数量	总数	仇恨数量	总数	仇恨数量	总数	仇恨数量
BOW 全集	M1	212	53	414	10	73	4	393	69
	M2	194	46	403	11	68	4	349	60
	M3	169	40	311	10	56	3	249	40
BOW 子集	M1	1 263	150	1 661	46	1 805	20	1 031	113
	M2	1 259	146	1 661	44	1 805	20	1 041	117
	M3	1 250	137	1 653	38	1 804	20	1 023	107
Doc2vec 全集	M1	25	11	5	5	2	2	83	15
	M2	5	2	4	3	3	3	72	11
	M3	19	8	3	3	2	2	83	15
Doc2vec 子集	M1	16	5	5	3	3	3	84	15
	M2	8	3	3	3	2	2	77	12
	M3	11	0	0	0	2	2	67	14

纳入规则前因部分一些维度的不相关特征,会造成该规则触发强度进一步下降。但不同类别的文本实例可能会存在一些作为冗余特征的常用词,这导致当实例包含此类常用词时,该实例会出现歧义。

使用 Doc2vec 进行特征提取时,特征选择后得到的歧义实例数量与使用整个特征集相似。这种情况下,使用整个特征集或选定特征子集时,除性取向数据集之外,M4 性能通常较为相似。

表3所示为模糊分类结果。表中“N/A”表示相应的分类器未输出仇恨类别,即完全偏倚到非仇恨类别。表3结果

表明,M4 在大部分情况下优于其他方法。具体来说,当使用完整 BOW 特征集时,本文 M4 的性能通常劣于其他方法,但采用特征选择能够显著提高 M1 的性能,使其除宗教数据集之外均优于其他方法。使用 Doc2vec 进行特征提取时,M4 在完整特征集和选定特征子集上的性能始终优于其他方法。这充分说明了提出的模糊方法能够有效处理文本歧义问题,克服了单个模糊分类器的局限性,通过使用不同模糊范数和基于实例的推理,对多个模糊分类器进行训练和融合,显著提升分类性能。同时,提出的方法利用多种方式处理文本中的模糊性,克服了概率性方法的局限。

表3 与其他方法的分类性能比较

方法		宗教			种族			残疾			性取向		
特征	学习	P	R	F	P	R	F	P	R	F	P	R	F
	SVM ^[7]	0.444	0.626	0.520	0.336	0.529	0.411	0.596	0.549	0.571	0.448	0.475	0.462
BOW	NB ^[8]	0.808	0.532	0.641	0.849	0.643	0.732	0.682	0.294	0.411	0.635	0.219	0.325
全集	DNN ^[11]	N/A	0	N/A	N/A	0	N/A	N/A	0	N/A	N/A	0	N/A
	本文 M4	0.679	0.667	0.744	0.627	0.671	0.648	0.517	0.608	0.559	0.519	0.525	0.565
	SVM ^[7]	0.850	0.662	0.744	0.592	0.871	0.705	0.654	0.667	0.660	0.739	0.557	0.636
BOW	NB ^[8]	0.854	0.608	0.711	0.857	0.600	0.706	0.882	0.588	0.706	0.868	0.503	0.637
子集	DNN ^[11]	0.847	0.500	0.629	0.833	0.571	0.678	0.903	0.549	0.683	0.811	0.164	0.273
	本文 M4	0.878	0.617	0.725	0.877	0.714	0.787	0.912	0.608	0.729	0.875	0.536	0.664
	SVM ^[7]	0.516	0.374	0.433	0.109	0.814	0.193	0.192	0.647	0.296	0.507	0.186	0.272
Doc2vec	NB ^[8]	0.705	0.633	0.473	0.706	0.171	0.276	1	0.451	0.622	0.704	0.311	0.432
全集	DNN ^[11]	N/A	0	N/A	N/A	0	N/A	N/A	0	N/A	N/A	0	N/A
	本文 M4	0.748	0.401	0.522	0.933	0.400	0.560	0.969	0.608	0.747	0.688	0.350	0.464
	SVM ^[7]	0.424	0.563	0.484	0.103	0.714	0.180	0.157	0.627	0.251	0.417	0.437	0.427
Doc2vec	NB ^[8]	N/A	0	N/A	N/A	0	N/A	1.000	0.451	0.622	0.636	0.077	0.137
子集	DNN ^[11]	0.667	0.045	0.084	N/A	0	N/A	N/A	0	N/A	0.611	0.120	0.201
	本文 M4	0.835	0.365	0.508	0.861	0.443	0.585	0.939	0.608	0.738	0.789	0.388	0.520

此外,本文利用 Wilcoxon 秩检验执行了统计分析,以确认本文模糊方法与其他方法的性能差异是否具有统计显著性。基于针对每个数据集准备的每个特征集来进行比较,即总计 16(4×4)个数据集。如表 4 所示,与其他方法相比,使用 M4 能够显著提升分类性能。

表4 学习方法性能的秩检验

比较方法	p 值	虚假设
NB ^[8] vs M4	0	拒绝
SVM ^[7] vs M4	0.001	拒绝
DNN ^[11] vs M4	0	拒绝

4 结 论

本文提出了用于网络仇恨言论分类的模糊方法,以处理文本的模糊性和不确定性,通过反映不同的情感强度提供更精确的输出,有效检测任何歧义实例,并通过多个模糊分类器的融合和基于实例的推理,对歧义文本进行深度分析。使用两种特征提取方法和一种特征选择方法制备不同的特征集,以评估特征制备的不同方式对模糊分类器训练的影响。实验结果表明,提出的模糊方法在大部分情况下性能优于其他方法,能够显著提升分类性能。

参考文献

- [1] 严佩敏,唐婉琪. 基于可分离卷积神经网络的文本分类[J]. 电子测量技术, 2020, 43(13): 7-12.
- [2] 余本功,张连彬. 基于 CP-CNN 的中文短文本分类研究[J]. 计算机应用研究, 2018, 35(4): 1001-1004.

- [3] BURNAP P, WILLIAMS M L. Cyber hate speech on Twitter: An application of machine classification and statistical modeling for policy and decision making[J]. Policy Internet, 2015, 7(2): 223-242.
- [4] 朱芳鹏,王晓峰. 面向船舶工业新闻的文本分类[J]. 电子测量与仪器学报, 2020, 34(1):154-160.
- [5] 叶雪梅,毛雪岷,夏锦春,等. 文本分类 TF-IDF 算法的改进研究[J]. 计算机工程与应用, 2019, 55(2): 104-109,161.
- [6] 郑啸,王义真,袁志祥,等. 基于卷积记忆神经网络的微博短文本情感分析[J]. 电子测量与仪器学报, 2018, 32(3): 195-200.
- [7] PASADAS D J, BASKARAN P, RAMOS H G, et al. Detection and classification of defects using ECT and multi-level SVM model[J]. IEEE Sensors Journal, 2019, 51(9): 85-96.
- [8] 贺鸣,孙建军,成颖. 基于朴素贝叶斯的文本分类研究综述[J]. 情报科学, 2016, 34(7):147-154.
- [9] WASEEM Z, HOVY D. Hateful symbols or hateful people predictive features for hate speech detection on Twitter[C]. In Proceeding NAACL-HLT, San Diego, CA, USA: IEEE Press, 2016: 88-93.
- [10] 周枫,李荣雨. 基于 BGRU 池的卷积神经网络文本分类模型[J]. 计算机科学, 2018, 45(6): 241-246.
- [11] BADJATIYA P, GUPTA S, GUPTA M, et al. Deep learning for hate speech detection in tweets[C]. 26th International Conference World Wide Web Companion,

- Perth, WA, Australia; IEEE Press, 2017; 759-760.
- [12] 樊蕊. 模糊信息流的逻辑研究[D]. 天津:南开大学, 2014.
- [13] CHANDRAN D, CROCKETT K A, DMCLEAN D, et al. An automatic corpus based method for a building multiple fuzzy word dataset [C]. IEEE International Conference Fuzzy System. Istanbul, Turkey; IEEE Press, 2015;1-8.
- [14] ZHAO CH J, WANG S G, LI D Y. Determining fuzzy membership for sentiment classification: A three-layer sentiment propagation model [J]. PLoS ONE, 2016, 11(11): 60-72.
- [15] 张创基. 基于BP学习的P2P网络信任度评价模型优化[J]. 国外电子测量技术, 2019, 38(6): 44-48.
- [16] 周世波, 徐维祥, 徐良坤. 融合密度峰值和空间邻域信息的FCM聚类算法[J]. 仪器仪表学报, 2019, 40(4): 140-147.
- [17] 崔建, 李强, 刘勇. 基于模糊分类关联规则的支持向量机分类器生成方法[J]. 计算机应用, 2011, 31(5): 196-198, 214.
- [18] IEEE A P K, MEMBER I S, SONAVANE S P. Improved genetic particle swarm optimization and feature subset selection for extreme learning machine [J]. International Journal of Computer Ences and Enginccring, 2018, 6(1): 48-54.
- [19] THELWALL M, BUCKLEY K, PALTOGLOU G, et al. Sentiment strength detection in short informal text[J]. Journal of Association Information Science Technology, 2010, 61(12): 2544-2558.

作者简介

王建华, 硕士, 讲师, 主要研究方向为智能算法、神经网络、数据挖掘。

E-mail: cc999687vip@sina.com

冉煜琨, 研究生, 主要研究方向为智能算法。