第 44 卷 第 9 期 计 算 机 工 程 2018 年 9 月
Vol. 44 No. 9 Computer Engineering September 2018

·人工智能及识别技术· 文章编号：1000-3428(2018)09-0159-05 文献标志码：A 中图分类号：TP393

基于隐空间代价敏感学习的微博水军识别方法

王 磊，任 航，王之怡

（西南财经大学 经济信息工程学院，成都 610074）

摘 要： 根据微博水军活动的特点，提出一种基于隐空间代价敏感学习的半监督水军识别方法。从内容、行为、社交关系 3 个视角选取微博账户的 22 个特征，结合矩阵隐空间分解、代价敏感学习和社交关系正则技术，构造代价敏感的半监督最大间隔分类模型，并利用随机梯度下降算法求解模型的线性复杂度。实验结果表明，该方法在准确率、召回率和 F1 指标上均优于 SMFSR 和 L2-SVMs 方法，并且具有接近线性的学习速度。

关键词： 水军识别；矩阵分解；代价敏感学习；社交关系正则；隐空间

中文引用格式： 王 磊，任 航，王之怡. 基于隐空间代价敏感学习的微博水军识别方法[J]. 计算机工程，2018，44(9)：159-163，170.

英文引用格式： WANG Lei，REN Hang，WANG Zhiyi. Microblog spammer identification method based on cost-sensitive learning in latent space[J]. Computer Engineering，2018，44(9)：159-163，170.

Microblog Spammer Identification Method Based on Cost-sensitive Learning in Latent Space

WANG Lei，REN Hang，WANG Zhiyi

（School of Economic Information Engineering，Southwestern University of Finance and Economics，Chengdu 610074，China）

【Abstract】 According to the characteristics of microblog spammers，this paper proposes a semi-supervised spammer identification method based on cost-sensitive learning in latent space. Firstly，it selects twenty-two features of microblog account from perspectives of contents，activities and social relations. Then，it obtains latent account vectors using matrix factorization method and constructs a novel cost-sensitive semi-supervised classification model with the maximum margin theory in latent space. In addition，a social relation regularization from following behaviors is formulated on the model. Finally，it develops a linear-complexity algorithm for solving the model with the stochastic gradient descent method. Experimental results show that the proposed method outperforms existing methods significantly，such as SMFSR and L2-SVMS，in terms of the evaluation measures of accuracy，recall and F1 score. It also obtains nearly linear training speeds.

【Key words】 spammer identification；matrix factorization；cost-sensitive learning；social relation regularization；latent space

DOI： 10.19678/j.issn.1000-3428.0048993

0 概述

网络水军是指为了商业和政治目的，有组织、有计划地在社交网络平台上集中炒作某个事件，以达成宣传、推销或诋毁效果的大量专业或非专业技术团体[1]。他们在社交网络平台上肆意传播虚假谣言、垃圾广告、恶意链接和木马等危险信息，严重影响了网络秩序、公民利益、社会及国家政治稳定。其中，微博平台因具有信息传播速度快、影响范围广、沟通自由随意等明显特点，成为网络水军猖獗活动的理想场所和重灾区[2-3]。因此，研究如何准确识别微博平台上的水军活动已成为近年来的研究热点之一。

现有的微博水军识别方法主要分为 2 类：无监督识别方法和监督识别方法[1]。无监督识别方法一般依据微博用户的行为和交互数据建立社交关系图及其邻接矩阵，然后采用图聚类[4]、谱分析[5-6]、信用传播[7-8]等技术识别出水军用户。监督识别方法通常利用数据集上的用户类别标签，基于各种机器学习算法训练分类模型，用于对未知用户进行分类和

基金项目： 中央高校基本科研业务费重大理论基础研究项目(JBK151127)；中央高校基本科研业务费创新团队项目(JBK130503，JBK150503)；教育部人文社会科学研究西部和边疆地区项目(16XJAZH002)。

作者简介： 王 磊(1978—)，男，副教授、博士，主研方向为机器学习、数据挖掘；任 航，硕士研究生；王之怡，副教授、博士。

收稿日期： 2017-10-17 **修回日期：** 2017-11-22 **E-mail：** wanglei_t@swufe.edu.cn

识别。这些算法包括决策树(Decision-making Tree, DT)[9]、朴素贝叶斯(Naive Bayes, NB)[10-11]、支持向量机(Support Vector Machine, SVM)[12]、人工神经网络(Artificial Neural Network, ANN)[2]等。然而,上述方法在构造水军识别模型时均一定程度存在缺陷,例如:无监督方法普遍存在模型计算复杂度高的问题[9-11];监督学习方法普遍存在用户标签数据获取困难、不适用于处理存在属性缺失的数据集等问题[2,4-8,12]。

一些学者利用矩阵分解技术降低模型的复杂度和解决数据属性缺失问题[13-15]。文献[13]利用矩阵分解技术获得账户低维的隐因子向量,然后在隐空间内建立基于最大间隔分类方法的半监督识别模型 SMFSR,并结合好友社交关系对模型进行正则。该模型的识别精度和学习速度显著优于传统的 SVM 方法。此外,文献[14-15]分别基于矩阵分解和非负矩阵分解技术构建了类似的水军识别模型。但这些模型在设计时均未考虑微博水军识别应用中严重的数据不平衡问题[1-3,16],并且微博用户在社交平台中的影响力是显著不同的,他们被模型错误识别后所造成的后果和代价也显著不同。然而,现有识别模型多数都将所有微博账户同等对待。

针对大规模微博水军识别问题,本文在文献[13]研究的基础上,通过设计综合代价函数并采用基于用户关注关系的社交正则项,提出隐空间中基于代价敏感学习的半监督微博水军识别方法。

1 研究基础

1.1 微博水军识别问题定义

在微博社交网络中,识别水军的基本任务是:在给定的微博账户数据集上(包括少量带类别标签的监督数据和大量无标签的非监督数据),建立水军账户识别模型,以准确、快速地判别新样本是否属于水军。

令矩阵 $\boldsymbol{X}\in\mathbb{R}^{n\times d}$ 表示已知规模为 n 的微博账户数据集,d 是特征空间维度,并有 $\boldsymbol{X}=\boldsymbol{X}^{\mathrm{l}}\cup\boldsymbol{X}^{\mathrm{u}}$,其中,$\boldsymbol{X}^{\mathrm{l}}=\{\boldsymbol{x}_i\in\mathbb{R}^d\}_{i=1}^{l}$ 是带标签的监督数据集(满足 $l\ll n$),其类别标签是 $\boldsymbol{Y}^{l}=\{y_i\mid y_i=\pm1\}_{i=1}^{l}$。如果某个微博账户 $\boldsymbol{x}_i$ 的类别标签是 $y_i=-1$,则判断其为水军;否则,判断其为正常账户。$\boldsymbol{X}^{\mathrm{u}}=\{\boldsymbol{x}_i\in\mathbb{R}^d\}_{i=l+1}^{n}$ 是不带标签的非监督数据集。

同时,令 $\boldsymbol{X}^{\mathrm{test}}\in\mathbb{R}^{m\times d}$ 是未知类别的测试数据集,$f(\boldsymbol{x})$ 是学习的水军识别模型。因此,本文的研究目标是:对于 $\forall\boldsymbol{x}\in\boldsymbol{X}^{\mathrm{test}}$,建立最优的识别模型,以准确快速地判断 $\boldsymbol{x}$ 的类别,即 $y=f(\boldsymbol{x})$。

1.2 特征选取

微博水军善于伪装自己的行为和踪迹。因此,为了能够准确识别这些隐藏的水军账户,首先应该选取有效的特征对他们进行描述。一些学者从微博消息内容、用户行为、社交关系、时间/位置等视角提取特征,并通过实验验证了采用多视角的混合特征更容易捕获水军活动的踪迹。其中:文献[12]从内容、行为、关系 3 个视角选取 18 个特征,并通过统计方法分析了它们对于识别水军的重要性;文献[15]则采用相同的视角获取了类似的 20 个特征。

本文综合文献[12,15]的研究成果,从内容、行为、关系 3 个视角选取 22 个微博账户特征用于水军识别模型的研究,如表 1 所示。

表 1 本文采用的微博账户特征

类别	特征
消息内容特征	F1:消息平均长度
	F2:含 URL 平均数
	F3:含图片平均数
	F4:含 Hashtag 平均数
	F5:含有 URL 的消息比
	F6:含有图片的消息比
	F7:含有 Hashtag 的消息比
用户行为特征	F8:平均转发数
	F9:平均评论数
	F10:平均点赞数
	F11:平均收藏数
	F12:日均发消息数
	F13:日均接收消息数
	F14:日均@次数
	F15:日均被@次数
	F16:原创消息比
	F17:注册时间
社交关系特征	F18:粉丝数
	F19:关注数
	F20:好友数
	F21:关注/粉丝比
	F22:关注/好友比

此外,本文对于数据矩阵 $\boldsymbol{X}$ 每一维特征值均采用 z-score 方法进行归一化处理,以消除某一维特征上的极端值对模型精度的影响。

1.3 SMFSR 模型

文献[13]应用矩阵分解技术在隐空间中构造了一种新颖的半监督水军识别模型 SMFSR。该文献作者认为矩阵 $\boldsymbol{X}$ 可以分解为 2 个低秩的隐因子矩阵 $\boldsymbol{U}\in\mathbb{R}^{k\times n}$ 和 $\boldsymbol{V}\in\mathbb{R}^{k\times d}$,即 $\boldsymbol{X}\approx\boldsymbol{U}^{\mathrm{T}}\boldsymbol{V}$。其中,$k$ 是隐空间的维度且满足 $k<d$,它们的任意列向量 $\boldsymbol{u}_i$ 和 $\boldsymbol{v}_i$ 分别称为账户隐因子向量和特征隐因子向量。显然,$\boldsymbol{u}_i$ 是原始账户向量 $\boldsymbol{x}_i$ 在隐空间中的表示。

该文分别定义矩阵分解误差函数 $J_1(\boldsymbol{U},\boldsymbol{V})$、隐空间内最大间隔分类误差函数 $J_2(\boldsymbol{U},\boldsymbol{w})$、好友社交关系正则函数 $J_3(\boldsymbol{U})$,如下式所示:

$$J_1(\boldsymbol{U},\boldsymbol{V}) = \frac{1}{2}\sum_{x_{ij}\in\boldsymbol{X}} I_{ij}(x_{ij} - \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{v}_j)^2 + \frac{\lambda_1}{2}(\|\boldsymbol{U}\|_{\mathrm{F}}^2 + \|\boldsymbol{V}\|_{\mathrm{F}}^2) \tag{1}$$

$$J_2(\boldsymbol{U},\boldsymbol{w}) = C\sum_{i=1}^{l} h(y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{u}_i)) + \frac{\lambda_2}{2}\|\boldsymbol{w}\|_2^2 \tag{2}$$

$$J_3(\boldsymbol{U}) = \frac{\lambda_3}{2}\sum_{\boldsymbol{x}_i\in\boldsymbol{X}^{\mathrm{l}}}\sum_{\boldsymbol{x}_j\in N(\boldsymbol{x}_i)} y_i(\boldsymbol{u}_i - \boldsymbol{u}_j)^2 \tag{3}$$

其中,λ_1、λ_2、λ_3 是正则参数,$\|\cdot\|_{\mathrm{F}}$ 是 Frobenius 范数,$I_{ij}\in\{0,1\}$ 是指示变量,表明数据矩阵 $\boldsymbol{X}$ 相应位置上是否存在缺失值,C 是误分类时的惩罚参数,$\boldsymbol{w}$ 是最大间隔超平面的法向量,函数 $h(\cdot)$ 是惩罚分类错误的光滑 Hinge 损失函数,$N(\boldsymbol{x}_i)$ 表示账户 $\boldsymbol{x}_i$ 的好友集合。

由此,SMFSR 模型的目标函数可定义为:

$$\min J(\boldsymbol{U},\boldsymbol{V},\boldsymbol{w}) = J_1(\boldsymbol{U},\boldsymbol{V}) + J_2(\boldsymbol{U},\boldsymbol{w}) + J_3(\boldsymbol{U}) \tag{4}$$

通过简单分析可知,该模型实质上是在隐空间内,以矩阵分解误差和好友社交关系约束作为正则项,求解线性最大软间隔分类器的半监督模型。

2　隐空间中代价敏感的半监督水军识别模型

2.1　问题分析

通过对 SMFSR 分析可知,该模型存在以下不足:

1)微博数据存在严重的类别不平衡现象,水军账户只占很小比例[1-2,16]。而 SMFSR 模型采用的最大软间隔方法训练的超平面将更靠近多数类,导致了模型泛化性能的降低。

2)微博账户分为不同级别(如黄 V,蓝 V)[3],他们的社交圈规模和活跃程度有巨大差异。因此,不同账户被错误分类所带来的影响是不同的。然而在 SMFSR 模型中,所有账户被同等对待。

3)式(3)中的社交正则项只考虑了带标签账户的好友关系。事实上,水军的好友集通常很小,因此,该社交关系正则项通常难以充分发挥作用。

针对这些问题,本文采用代价敏感学习技术和基于关注关系的社交正则项给予解决,提出一种改进的基于隐空间代价敏感学习的半监督水军识别方法。

2.2　代价敏感学习

在分类模型中,当不同数据样本被模型错误分类造成不同的代价时,该分类问题被称为代价敏感学习问题。按照粒度的不同,代价分为基于类的代价和基于样本的代价[17]。

针对微博账户类别不平衡的问题,本文采用类代价函数 $c_1(\boldsymbol{x})$,依据每类样本的规模为水军和正常账户设置不同的误分类代价,具体如下:

$$c_1(\boldsymbol{x}) = \begin{cases} C, y = 1 \\ \beta C, y = -1 \end{cases} \tag{5}$$

其中,系数 β 是监督数据集 $\boldsymbol{X}^{\mathrm{l}}$ 中正常账户和水军之间数量的比值。显然,式(5)中水军具有更大的误分类代价,从而防止分类超平面靠近多数类。

针对微博账户在社交网络中的级别和影响力的不同,本文采用样本代价函数 $c_2(\boldsymbol{x})$,并通过账户的粉丝数 $nr(\boldsymbol{x})$ 和关注数 $ne(\boldsymbol{x})$ 之和来度量账户在社交平台的影响力。其中,粉丝数反映了账户当前的影响力,关注数体现了账户的潜在影响力。这样,样本代价函数定义为:

$$c_2(\boldsymbol{x}) = \ln(nr(\boldsymbol{x}) + ne(\boldsymbol{x}) + 2) \tag{6}$$

通过分析可知,粉丝数和关注数越高的水军,其社交影响力越大,潜在的危害也越大。因此,式(6)将为他们设置高的误分类代价,以使得分类器更加重视他们被正确识别。同理,对于影响力高的正常账户,也设置高的误分类代价。

利用上述 2 种代价函数,本文针对微博账户 $\forall\boldsymbol{x}\in\boldsymbol{X}^{\mathrm{l}}$ 设计新的综合代价函数为:

$$cost(\boldsymbol{x}) = c_1(\boldsymbol{x})\cdot c_2(\boldsymbol{x}) \tag{7}$$

在此基础上,定义隐空间中如下的代价敏感的最大间隔分类误差函数如下:

$$J_2'(\boldsymbol{U},\boldsymbol{w}) = \sum_{i=1}^{l} cost(\boldsymbol{x}_i)h(y_i(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{u}_i)) + \frac{\lambda_2}{2}\|\boldsymbol{w}\|_2^2 \tag{8}$$

Hinge 损失函数 $h(t)$ 的定义如下:

$$h(t) = \begin{cases} 0.5 - t, t\leqslant 0 \\ 0.5(1-t)^2, 0 < t < 1 \\ 0, t\geqslant 1 \end{cases} \tag{9}$$

通过分析可知,根据式(8)设计的最大软间隔分类器将在隐空间内对少数类样本和影响力大的样本设置更大的误分类惩罚,使分类器能有效处理不平衡数据集和提升对影响力高的样本的分类准确度,这与微博账户数据集的特点一致。

2.3　基于关注关系的社交正则项

在微博社交过程中,关注是一种主动性社交行为,体现了其他用户的价值认同。同时,水军通常通过大量关注合法用户,试图虚构其社交关系网络以躲避检测。因此,可以做出合理假设:在以关注关系建立的社交关系网络中,水军与其关注账户之间存在明显差异,而正常账户与其关注账户之间有较高相似性。令 $\boldsymbol{S}\in\mathbb{R}^{n\times n}$ 表示基于关注行为建立的社交关系矩阵,其中,若微博账户 $\boldsymbol{x}_i$ 单向关注账户 $\boldsymbol{x}_j$,则 $S_{ij}=1$,反之为 0。设 $E(\boldsymbol{x}_i)$ 表示 $\boldsymbol{x}_i$ 已关注的其他账户集合,$|E(\boldsymbol{x}_i)|$ 是集合的规模。本文在微博数据集上定义如下的社交关系正则项:

$$J_3'(\boldsymbol{U}) = \frac{\lambda_3}{2}\sum_{\boldsymbol{x}_i\in\boldsymbol{X}^{\mathrm{l}}}\sum_{\boldsymbol{x}_j\in E(\boldsymbol{x}_i)}\frac{y_i(\boldsymbol{u}_i - \boldsymbol{u}_j)^2}{|E(\boldsymbol{x}_i)|} \tag{10}$$

显然,最小化式(10)相当于在隐空间中最大化水军和其关注账户之间的差异,并最小化正常账户

和其关注账户之间的差异。由于关注账户集合的规模远大于好友集合,因此该正则项相比于文献[13]采用的方法能更准确地反映微博账户的社交关系。

2.4 本文方法

基于上述分析,本文在文献[13]的基础上提出一种基于隐空间代价敏感学习的半监督水军识别方法,其优化目标函数为:

$$\min J(\boldsymbol{U},\boldsymbol{V},\boldsymbol{w}) = J_1(\boldsymbol{U},\boldsymbol{V}) + J_2'(\boldsymbol{U},\boldsymbol{w}) + J_3'(\boldsymbol{U}) \quad (11)$$

求解出 $\boldsymbol{U}$、$\boldsymbol{V}$、$\boldsymbol{w}$ 后,对于任意的待识别账户 $\boldsymbol{x} \in \mathbb{R}^d$,水军识别模型的判别函数如下:

$$f(\boldsymbol{x}) = \mathrm{sgn}(\boldsymbol{w}^{\mathrm{T}}\boldsymbol{u}) = \mathrm{sgn}(\boldsymbol{w}^{\mathrm{T}}(\boldsymbol{V}^{\dagger})^{\mathrm{T}}\boldsymbol{x}) \quad (12)$$

其中,$\boldsymbol{V}^{\dagger}$ 是隐因子矩阵 $\boldsymbol{V}$ 的 Moore-Penrose 广义逆,用以近似它的逆矩阵,$\mathrm{sgn}(t)$ 是指示函数,当满足 $t \geqslant 0$ 时,$\mathrm{sgn}(t)=1$,否则 $\mathrm{sgn}(t)=-1$。

2.5 随机梯度下降求解算法

容易验证 $J(\boldsymbol{U},\boldsymbol{V},\boldsymbol{w})$ 是一阶连续可微的,则其对 $\boldsymbol{u}_i$、$\boldsymbol{v}_i$ 和 $\boldsymbol{w}$ 的梯度分别如下:

$$\frac{\partial J}{\partial \boldsymbol{u}_i} = \sum_{x_{ij}\in X} I_{ij}(x_{ij} - \boldsymbol{u}_i^{\mathrm{T}}\boldsymbol{v}_j)(-\boldsymbol{v}_j) + \lambda_1 \boldsymbol{u}_i + cost(\boldsymbol{x}_i) \cdot h'(y_i \boldsymbol{w}^{\mathrm{T}}\boldsymbol{u}_i) \cdot y_i \boldsymbol{w} + \lambda_3 \sum_{x_j \in E(x_i)} \frac{y_i(\boldsymbol{u}_i - \boldsymbol{u}_j)}{|E(\boldsymbol{x}_i)|} \quad (13)$$

$$\frac{\partial J}{\partial \boldsymbol{v}_i} = \sum_{x_{ij}\in X} I_{ij}(x_{ij} - \boldsymbol{u}_j^{\mathrm{T}}\boldsymbol{v}_i)(-\boldsymbol{u}_j) + \lambda_1 \boldsymbol{v}_i \quad (14)$$

$$\frac{\partial J}{\partial \boldsymbol{w}} = \sum_{i=1}^{l} cost(\boldsymbol{x}_i) \cdot h'(y_i \boldsymbol{w}^{\mathrm{T}}\boldsymbol{u}_i) y_i \boldsymbol{u}_i + \lambda_2 \boldsymbol{w} \quad (15)$$

其中,$h'(t)$ 是损失函数,是 $h(t)$ 的梯度函数。

在此基础上,本文给出一种求解上述模型的随机梯度下降算法 LCSIM,如算法 1 所示。

算法 1 LCSIM

输入 数据矩阵 $\boldsymbol{X} = \boldsymbol{X}^{\mathrm{l}} \cup \boldsymbol{X}^{\mathrm{u}}$,监督数据的标签向量 $\boldsymbol{Y}^{l}$,社交关系矩阵 $\boldsymbol{S}$,隐空间的维度 k,惩罚参数 C,学习速率 η

输出 隐因子矩阵 $\boldsymbol{U}$、$\boldsymbol{V}$ 和法向量 $\boldsymbol{w}$

步骤 1 在[0,1]范围内随机初始化 $\boldsymbol{U}$、$\boldsymbol{V}$ 和向量 $\boldsymbol{w}$。

步骤 2 对任意用户 x_i,$i=1,2,\cdots,n$,按下式更新其隐向量:

$$\boldsymbol{u}_i \leftarrow \boldsymbol{u}_i + \eta \frac{\partial J}{\partial \boldsymbol{u}_i}, \boldsymbol{v}_i \leftarrow \boldsymbol{v}_i + \eta \frac{\partial J}{\partial \boldsymbol{v}_i}$$

步骤 3 更新分类超平面的法向量 $\boldsymbol{w}$:

$$\boldsymbol{w} \leftarrow \boldsymbol{w} + \eta \frac{\partial J}{\partial \boldsymbol{w}}$$

步骤 4 若所有梯度 $\frac{\partial J}{\partial u_i}$、$\frac{\partial J}{\partial v_i}$、$\frac{\partial J}{\partial w}$ 的变化量均小于预设阈值 ε,则算法终止;否则返回步骤 2。

通过分析可知,算法中步骤 2 和步骤 3 中一次迭代的时间复杂度为 $O(ndk + n\bar{E}k)$,其中,$\bar{E}$ 是微博账户的平均关注数量。在实际中,可以把 k、d、$\bar{E}$ 以及迭代次数 k 视为常数。因此,算法的时间复杂度仅与微博账户的数量 n 呈线性关系,该算法是一种快速识别算法。

此外,由于算法利用矩阵分解技术在隐空间中进行迭代,能有效处理原始数据矩阵存在大量缺失值的问题,因此具有良好的适应性。

3 数值实验与结果分析

为了客观评价提出的 LCSIM 方法的有效性,本文在大规模的微博数据集上对模型算法的性能进行实验,并与 SMFSR、半监督线性支持向量机(L2-SVMs)[18]的实验结果进行对比分析。所有算法均采用 C++编码实现,运行环境为一台 3.9 GHz 四核 CPU、16 GB 内存的计算机。

实验设计爬虫程序从新浪微博开放平台获取微博账户数据,经预处理后得到 255 917 个账户及其 22 个特征(如表 1 所示)。经过领域专家的人工判别,从中识别出 2 200 个水军和 12 500 个正常账户。实验从中随机选择约 40% 样本作为监督训练数据集(900 个水军和 5 000 个正常账户),剩余部分作为测试集(1 300 个水军和 7 500 个正常账户),其他 241 217 个样本作为非监督训练数据集。

本文选用准确率(*Precision*)、召回率(*Recall*)、F1 值(F)作为模型性能的评价指标。若令 TP、FN 分别为数据集中的水军样本被预测为水军和正常账户的数量,TN、FP 分别为正常账户被预测为正常账户和水军的数量,则上述指标定义为:

$$Precision = \frac{TP}{TP + FP} \quad (16)$$

$$Recall = \frac{TP}{TP + FN} \quad (17)$$

$$F = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (18)$$

在实验中,LCSIM 的训练参数设置为:$k=15$,$C=4$,正则参数 $\lambda_1=\lambda_2=\lambda_3=1$,学习速率 $\eta=0.1$,阈值 $\varepsilon=10^{-4}$。公平起见,SMFSR 的参数采用相同设置,L2-SVMs 的正则参数按原文设置为 $\lambda=0.001$。

在监督和非监督训练集上,将上述 3 种方法分别运行 10 次,其识别结果比较如表 2 所示。可以看出:LCSIM 和 SMFSR 由于采用了社交关系正则项和矩阵隐空间分解技术,它们取得的识别结果明显高于 L2-SVMs 方法;而 LCSIM 由于采用了代价敏感学习技术,避免了分类超平面过度靠近多数类,其准确度、召回率和 F1 值显著优于 SMFSR 方法,3 个指标分别提高了 7.0%、1.3% 和 4.4%。此外,通过分析被误判的水军样本和正常样本,笔者发现它们粉丝数和关注数之和的均值分别为 29.1 和 12.8,远低于所有样本的均值,这与上文的讨论是一致的,即本文方法更倾向影响力大的账户被正确分类。

表 2　3 种方法的识别结果比较

方法	准确率	召回率	F1 值
LCSIM	0.922 ±0.019	0.989 ±0.010	0.954 ±0.013
SMFSR	0.852 ±0.046	0.976 ±0.031	0.910 ±0.037
L2-SVMs	0.785 ±0.092	0.862 ±0.058	0.822 ±0.055

此外,本文通过实验检验隐空间维度 k 和惩罚参数 C 对性能的影响,实验结果如图 1 和图 2 所示。由图 1 可知,LCSIM 的 F1 值随着维度 k 而变化,当其在区间[14,20]范围内取值时,获得最佳 F1 值。考虑到维度 k 和 LCSIM 的计算复杂度有关,本文取 $k=15$。由图 2 可知,惩罚参数 C 的最佳取值范围在[1,32]之间。若其取值过小,则对分类器错误分类的惩罚很小,且代价损失函数无法发挥作用,导致其的精度较低;若其取值过大,将对分类错误进行过度惩罚,导致出现过度拟合现象而降低精度。因此,本文取最优值惩罚参数 $C=4$。

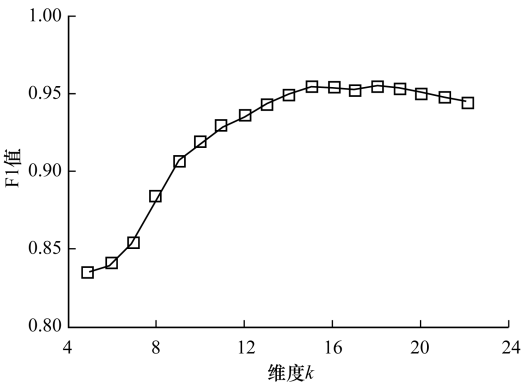

图 1　F1 值随隐空间维度 k 的变化情况

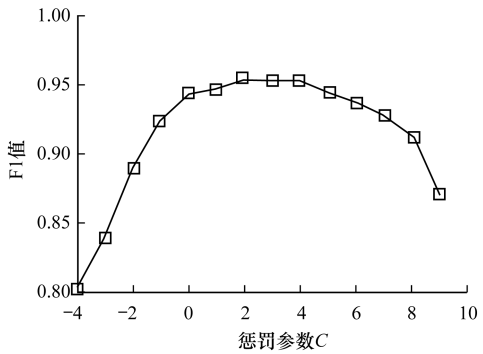

图 2　F1 值随隐空间惩罚参数 C 的变化情况

另外,本文分别选用了一定比例的训练集进行实验,考察 LCSIM 方法所需的训练时间,结果如图 3 所示。可以看出,其所需的训练时间基本上和数据集规模呈线性关系,仅仅略低于 SMFSR 方法。在全部 247 117 个样本上所需的训练时间仅为 332.5 s,速度明显快于线性核函数的 L2-SVMs 方法。

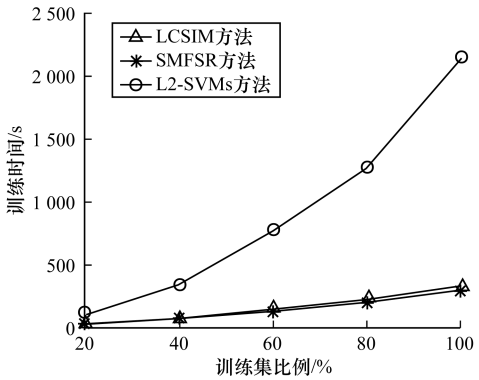

图 3　不同方法所需训练时间比较

4　结束语

本文针对微博水军识别问题,提出一种基于隐空间代价敏感学习的半监督识别方法。首先从 3 个视角筛选出 22 个微博账户特征,然后结合矩阵隐空间分解技术、综合代价函数、基于关注关系的社交正则项构建半监督水军识别模型,并采用随机梯度下降算法进行求解。实验结果表明,在大规模微博数据集上,该方法在各项性能指标上均明显优于 SMFSR 和 L2-SVMs 方法。下一步将结合流形学习技术提高其识别精度,并研究正则参数 λ_1、λ_2、λ_3 的最优取值。

参考文献

[1] ADEWOLE K S, ANUAR N B, KAMSIN A, et al. Malicious accounts: dark of the social networks[J]. Journal of Network and Computer Applications, 2017, 79(1): 41-67.

[2] 莫　倩,杨　珂. 网络水军识别研究[J]. 软件学报, 2014,25(7):1505-1526.

[3] 丁兆云,贾　焰,周　斌. 微博数据挖掘研究综述[J]. 计算机研究与发展,2014,51(4):691-706.

[4] MILLER Z, DICKINSON B, DEITRICK W, et al. Twitter spammer detection using data stream clustering[J]. Information Sciences, 2014, 260(1): 64-73.

[5] 韩忠明,杨　珂,谭旭升. 利用加权用户关系图的谱分析探测大规模电子商务水军团体[J]. 计算机学报, 2017,40(4):939-954.

[6] WU Leting, WU Xintao, LU Aidong, et al. A spectral approach to detecting subtle anomalies in graphs[J]. Journal of Intelligent Information Systems, 2013, 41(2): 313-337.

[7] GHOSH S, VISWANATH B, KOOTI F, et al. Understanding and combating link farming in the Twitter social network[C]//Proceedings of International Conference on World Wide Web. New York, USA: ACM Press, 2012: 61-70.

[8] XUE Jilong, YANG Zhi, YANG Xiaoyong, et al. VoteTrust: leveraging friend invitation graph to defend against social network Sybils[C]//Proceedings of IEEE INFOCOM'13. Washington D. C., USA: IEEE Press, 2013: 2400-2408.

(下转第 170 页)

[7] ZELENKO D, AONE C, RICHARDELLA A. Kernel methods for relation extraction[J]. Journal of Machine Learning Research, 2003, 3(3): 1083-1106.

[8] CULOTTA A, SORENSEN J. Dependency tree kernels for relation extraction[C]//Proceedings of Meeting of the Association for Computational Linguistics. [S. l.]: Association for Computational Linguistics, 2004: 423-429.

[9] BUNESCU R C, MOONEY R J. A shortest path dependency kernel for relation extraction[C]// Proceedings of Conference on Human Language Technology and Empirical Methods in Natural Language Processing. [S. l.]: Association for Computational Linguistics, 2005: 724-731.

[10] 黄瑞红, 孙 乐, 冯元勇, 等. 基于核方法的中文实体关系抽取研究[J]. 中文信息学报, 2008, 22(5): 102-108.

[11] ZHANG M, ZHANG J, SU J, et al. A composite kernel to extract relations between entities with both flat and structured features[C]//Proceedings of International Conference on Computational Linguistics and Meeting of the Association for Computational Linguistics. [S. l.]: Association for Computational Linguistics, 2006: 17-21.

[12] QIAN L, ZHOU G, KONG F, et al. Exploiting constituent dependencies for tree kernel-based semantic relation extraction[C]//Proceedings of International Conference on Computational Linguistics. [S. l.]: DBLP, 2008: 697-704.

[13] 庄成龙, 钱龙华, 周国栋. 基于树核函数的实体语义关系抽取方法研究[J]. 中文信息学报, 2009, 23(1): 3-8.

[14] SOCHER R, PENNINGTON J, HUANG E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. [S. l.]: DBLP, 2011: 151-161.

[15] LAI S, XU L, LIU K, et al. Recurrent convolutional neural networks for text classification[C]//Proceedings of the 29th AAAI Conference on Artificial Intelligence. Austin, USA: [s. n.], 2015: 2267-2273.

[16] ZENG D, LIU K, CHEN Y, et al. Distant supervision for relation erxtraction via piecewise convolutional neural networks[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: [s. n.], 2015: 1753-1762.

[17] SANTOS C N D, XIANG B, ZHOU B. Classifying relations by ranking with convolutional neural networks[J]. Computer Science, 2015, 86: 132-137.

[18] XU K, FENG Y, HUANG S, et al. Semantic relation classification via convolutional neural networks with simple negative sampling[J]. Computer Science, 2015, 71: 941-949.

[19] YAN X, MOU L, LI G, et al. Classifying relations via long short term memory networks along shortest dependency path[J]. Computer Science, 2015, 42: 56-61.

[20] LI J, LUONG T, JURAFSKY D, et al. When are tree structures necessary for deep learning of representations[C]//Proceedings of Conference on Empirical Methods in Natural Language Processing. Lisbon, Portugal: [s. n.], 2015: 2304-2314.

[21] MIWA M, BANSAL M. End-to-end relation extraction using LSTMs on sequences and tree structures[C]// Proceedings of Meeting of the Association for Computational Linguistics. [S. l.]: Association for Compuational Linguistics, 2016: 1105-1116.

[22] KIM Y. Convolutional neural networks for sentence classification[J]. Empirical Methods in Natural Language Processing, 2014, 14: 1746-1751.

编辑 金胡考

(上接第 163 页)

[9] IGAWA R A, JR S B, KIDO G S, et al. Account classification in online social networks with LBCA and wavelets[J]. Information Sciences, 2016, 332(3): 72-83.

[10] 张艳梅, 黄莹莹, 甘世杰, 等. 基于贝叶斯模型的微博网络水军识别算法研究[J]. 通信学报, 2017, 38(1): 44-53.

[11] CHEN C M, GUAN D J, SU Q K. Feature set identification for detecting suspicious URLs using Bayesian classification in social networks[J]. Information Sciences, 2014, 289(6): 133-147.

[12] ZHENG Xianghan, ZENG Zhipeng, CHEN Zheyi, et al. Detecting spammers on social networks[J]. Neurocomputing, 2015, 159(C): 27-34.

[13] ZHU Yin, WANG Xiao, ZHONG Erheng. Discovering spammers in social networks[C]//Proceedings of AAAI Conference on Artificial Intelligence. New York, USA: AAAI Press, 2012: 171-177.

[14] HU Xia, TANG Jiliang, ZHANG Yanchao, et al. Social spammer detection in microblogging[C]//Proceedings of International Joint Conference on Artificial Intelligence. New York, USA: AAAI Press, 2013: 2633-2639.

[15] YU Dingguo, CHEN Nan, JIANG F, et al. Constrained NMF-based semi-supervised learning for social media spammer detection[J]. Knowledge-Based Systems, 2017, 125: 64-73.

[16] ZHANG Yi, LU Jianguo. Discover millions of fake followers in Weibo[J]. Social Network Analysis and Mining, 2016, 6(1): 1-15.

[17] 董建设, 袁占亭, 张秋余. 代价敏感支持向量机在垃圾邮件过滤中的应用[J]. 计算机工程, 2008, 34(10): 131-132.

[18] KEERTHI S S, DECOSTE D. A modified finite Newton method for fast solution of large scale linear SVM[J]. Journal of Machine Learning Research, 2005, 6: 341-361.

编辑 金胡考