基于半监督协同训练算法的微博水军识别

韩晴晴 张艳梅 牛 娃
(中央财经大学信息学院 北京 102206)

摘 要 在迅速发展的互联网时代,微博产生了大量的信息,但是在微博话题等地带存在着较多水军,水军在一定程度上影响了普通用户了解某人或者某事的真实情况。因此,为了高效、准确地识别水军,针对水军样本数量少、非水军样本数量庞大等问题,综合考虑使用半监督协同训练算法。该算法通过研究微博用户的多个特征并对其进行综合分析,重新定义了 6 个属性特征值,包括账户关注度、每日发表微博数、微博影响力等。依据算法的特点,将 6 个属性特征值分为两个属性集,每个属性集对应一个视图,每个视图利用 Scikit-Learn 机器学习库中的 7 种分类方法训练出分类器,以对微博用户进行水军识别,最后在爬取的微博用户数据集上进行实验。实验结果表明,两个视图在分别使用朴素贝叶斯算法、逻辑回归算法训练分类器时,分类结果的准确率、召回率、精度和 F1-measure 值都较高。因此,综合分析微博用户特征并且使用符合实际情况的半监督协同训练算法,能够准确、高效、快速地识别微博水军。

关键词 半监督,协同训练,水军识别,分类器
中图法分类号 TP393 **文献标识码** A **DOI** 10.11896/jsjkx.180901617

Microblogging Water Army Identification Based on Semi-supervised Collaborative Training Algorithm

HAN Qing-qing ZHANG Yan-mei NIU Wa
(Information School, Central University of Finance and Economics, Beijing 102206, China)

Abstract In the fast-developing Internet era, Weibo brings a large amount of information, but there exists water army in Weibo topic. To a certain extent, the water army affects ordinary users to understand the real situation. In order to efficiently and accurately identify the water army, the semi-supervised collaborative training algorithm is considered comprehensively in view of the small number of water military samples and the large number of non-water military samples. By studying and analyzing multiple characteristics of Weibo users, the proposed algorithm redefines six attribute feature values, such as account attention, daily microblog number, and microblog influence. According to the characteristics of the algorithm, the six attribute feature values are divided into two attribute sets, each attribute set corresponds to one view, and each view uses seven classification methods in the Scikit-Learn machine learning library to train the classifier to identify the water army. Finally, experiments are conducted on dataset. The results show that the accuracy, recall rate, accuracy and F1-measure value of the classification results are higher when the two views use the naive Bayes algorithm and the logistic regression algorithm to train the classifier. Therefore, comprehensive analysis of Weibo user characteristics and the use of semi-supervised collaborative training algorithms in line with the actual situation can accurately, efficiently and quickly identify Weibo water army.

Keywords Semi-supervised, Collaborative training, Water army identification, Classifier

1 引言

网络信息时代的到来和发展使得生活发生了巨大的改变,微博的出现也使得信息得到快速传播,并且具有较强的影响力。然而,信息时代的发展是一把双刃剑,在使用微博的同时不可避免地要应对它带来的问题,即水军问题。刘姝雯等[1]认为微博水军往往出于宣传某事物或引导言论的目的,在微博平台上发布大量的虚假意见或没有任何意义的信息,并对某个特定的信息进行推广。水军的存在会导致微博的信息质量严重下降,使微博用户的心情受到影响,甚至会导致用户注销账户、删除微博客户端等后果。因此,如何准确、高效地识别水军是亟待解决的问题。

为了对微博水军识别进行深入研究,本文分析了较多的机器学习算法。通过学习发现,在机器学习领域中,近年来半监督学习被研究得最多,并且出现了许多半监督学习算法。其代表性的算法有:协同训练(co-training)算法、半监督 EM

到稿日期:2018-09-01 返修日期:2018-12-04 本文受国家自然科学基金项目(61602536,61773415),北京市社会科学基金重点项目(16YJA001)资助。

韩晴晴(1993—),硕士生,主要研究方向为数据分析;**张艳梅**(1976—),女,博士,副教授,主要研究方向为智能数据分析和服务计算,E-mail:jlzym0309@sina.com(通信作者);**牛 娃**(1968—),硕士,副教授,主要研究方向为智能数据分析。

算法等。在诸多的研究领域中，通过应用先进的数据采集等技术可以获得大量无标记样本，而有标记样本的获得难度很大。例如，微博用户在收到的未关注人的消息中可能存在诈骗信息的链接、推销、黄色内容等，很多用户因为不了解对这种情况的处理方式而将其直接忽视，并没有举报发送人或将其拉入黑名单，所以该发送人没有被标记。由此可知，实际生活中有标记样本的数量很少，获取也较困难。无监督学习利用无标记样本，有监督学习利用有标记样本，但是半监督学习能够充分利用少量有标记样本和大量无标记样本进行学习。协同训练算法充分利用两个视图，每个视图利用数据集训练得到分类器，两个分类器之间的交互性可以提高分类器的精度。因此，在水军识别的过程中，协同训练算法能提高水军识别的准确性。

目前，使用半监督协同训练算法进行水军识别的研究相对较少，但是这种方式能够利用少量已标记样本来判断大量的未标记样本，且可以提高水军识别的准确性。另一方面，很多研究使用的属性特征值并不是综合考量水军可能存在的特征，陈亮等[2]研究的属性特征值为关注数、评论数、转发数等，Gayo-Avello 等[3]关注账户注册微博的时间、账户粉丝值、账户朋友值等。基于此，本文抓取了用户的多个特征，包括微博的发表时间、用户名称、VIP 等级、关注数、粉丝数、性别、微博认证、微博等级、阳关信用、注册时间等，并按照实际情况对多个特征进行分析后将其重新组合为 6 个特征。

本文利用半监督协同训练算法对微博水军进行分类识别，该方法较好且符合实际情况。本文通过使用数据清洗等手段来对在微博平台爬取到的微博用户数据集进行人工标记，利用标记好的数据集提取已经定义的属性，在半监督协同训练算法中使用多种分类算法对样本进行分类，最后选择最优的分类算法。本文的主要贡献如下：1）根据有标记样本少的实际情况，使用了半监督协同训练算法，减少了对大量有标记样本的需求，降低了数据处理的难度，提高了水军分类识别的准确性能；2）依据抓取的微博用户的多个特征做出了较为准确的人工标记，并利用多个特征重新组合得到的 6 个属性特征值得到较高的准确率、召回率等。

本文第 2 节论述了相关工作；第 3 节论述了半监督协同训练算法；第 4 节进行了实验；最后总结全文。

2　相关工作

对微博水军进行识别的方法有 3 类：半监督学习方式、监督学习方式、无监督学习方式。对微博水军的属性特征值的选取方法也呈多样化。本节分别对这两方面的工作进行阐述。

2.1　水军识别方法的分析

1）半监督学习方式

Chen 等[4]提出了一种半监督线索融合模型，该模型对微博用户的内容、行为、关系、交互 4 个方面存在的线索进行识别，使用线性加权函数并以半监督的方式利用标记样本对未标记样本进行识别，以实现对不同话题和时间段的活跃水军进行很好地识别。Zhang 等[5]在半监督协同训练算法的基础上提出了一种 CoTrade 方法，该方法分两步传递标记信息，旨在提升不同视图之间标记信息传递的可靠性。经实验证明，该方法可以有效地利用未标记数据来实现更好的泛化性能。Blum 等[6]在利用协同训练算法的基础上使用 PAC 样式分析，从而使用网页小部分标记数据预测未标记的数据。

2）无监督学习

Miller 等[7]对两种流聚类算法 StreamKM＋＋和 DenStream 进行了改进，并将两种算法相结合对 Twitter 的水军进行识别，结果达到了 100％的召回率和 2.2％的错误率，效果良好。韩忠明等[8]在对电子商务平台上的水军团体进行分析后，利用加权用户关系图模型定位出很多隐藏较深的电子商务水军团体，其性能优于非加权用户关系图。

3）监督学习

Kim 等[9]利用增强的特征选择方法，使用朴素贝叶斯分类器对社会编书系统的水军进行识别。张艳梅等[10]结合贝叶斯模型和遗传算法，使用优化后的阈值矩阵训练分类器，能准确地对微博用户进行分类。Zheng 等[11]在分析微博用户的消息内容和社交行为后提取了一组特征，并使用支持向量机算法对微博的水军进行了识别。

目前，传统的无监督学习方式利用的是大量的无标记样本，有监督学习只利用少量的有标记样本。然而，半监督学习方式的突出优点是能够同时利用数量少的有标记样本和大量的无标记样本进行学习。因此，在标记样本少、无标记样本多的微博水军识别领域，利用半监督学习方式可以降低数据收集、处理等操作的难度，简化样本处理的过程等。同时，在多个特征的基础上重新定义的 6 个属性体现了用户的账户特征和行为特征，这与协同训练算法从两个角度分析数据的特点相符合。

2.2　微博用户特征值的选取分析

1）微博用户账户特征

袁旭萍等[12]在利用综合指数法和熵值法识别水军的过程中，选取的微博账户特征包括博主的粉丝数与关注数之比、微博博主发布的微博数、博主微博等级。程晓涛等[13]在基于关系图特征识别水军的研究中，从用户双向关注比（即用户的双方关注数与该用户关注其他用户的数量的比值）、账户注册时间这两个角度来分析微博用户是否是水军；基于 MapReduce 的随机森林算法选择账号关注度、互粉比，此外在内容特征中选择 URLs 比例、转发微博占比、发布平台类型等进行深入学习。

2）微博用户行为特征

张良[14]在微博水军识别研究中使用用户提及率（即用户所发微博包含的提及量与数据集中该用户的微博条数的比值）、文本 URL 率（即用户所发微博包含的 URL 总数与数据集中该用户微博总数的比值）、文本话题标签率（即用户微博内容中包含的话题标签的总数与数据集中该用户所发微博数的总数比）。吕晨[15]通过研究天涯社区用户的用户头像信息（全局）、账号存活期（全局）、平均每日积分（全局）、粉丝数关注数之比（全局）等，基于用户行为的特征来识别该社区中可能存在的水军用户。

在众多水军识别的研究中，基于微博用户的账户特征和行为特征这两个角度的分析不全面。很多文献[16-21]往往忽

略了对微博用户的微博等级、是否开通会员、阳光信用类别进行综合分析，而这种分析方式能使对水军的识别更加准确和具有参考性。同时，微博用户发表的微博会被其他用户转发、评论及点赞。水军发表的微博内容往往不会得到很多用户的评论，从而它的微博转发量、评论量及点赞量会非常少，因此水军的这种属性特征值得深一步的研究。

3 半监督协同训练算法

本节主要对属性特征值的选择、算法框架、算法的具体步骤进行阐述。

3.1 属性特征值的选择

微博数据集的属性特征主要包含：微博的发表时间、所用设备、微博具体内容、点赞数、评论数、转发数、用户 ID、用户名称、VIP 等级、关注数、粉丝数、性别、微博认证、微博等级、阳光信用、注册时间。将上述特征重新组合定义为 6 个属性特征值，并且将其分为两个视图，分别表示微博用户的账户特征和行为特征，如表 1 所列。

表 1 微博用户属性特征

Table 1 Weibo user attribute characteristics

账户特征	行为特征
账户关注度	@比例
微博等级	每日发表微博数
综合素质评估	微博影响力

1)账户关注度(AA)。微博用户之间的相互关注是微博中很重要的关系。将与某用户有关系的账户分为“关注”和“粉丝”两类，其中“关注”指该微博用户主动关注其他账户，“粉丝”指因对该微博账户发布的微博内容感兴趣等原因而设置关注的用户。在新浪微博中，该用户可以在关注页面收到关注用户发布的微博信息，同时该用户的粉丝也可以在自己的关注页面收到该用户发布的微博信息。

通过观察很多微博用户发现，具有影响力的用户的粉丝数往往多于该用户的关注数，例如某些明星的微博粉丝数是几千万而关注数在一百左右。但是，水军账户的关注数往往比粉丝数多。根据以上分析，定义账户关注度指标如下：

$$\text{账户关注度}=\frac{\text{粉丝数}}{\text{粉丝数}+\text{关注数}} \tag{1}$$

2)微博等级(WL)。微博等级是微博用户在微博的活跃度和所获荣誉的综合体现。微博用户在微博活跃的天数越多、使用的功能越多，该用户的微博等级就会越高。本文判断微博用户的等级是否大于 15，若 WL 值大于 15，则令 WL 值为 1，否则为 0。等级越高的用户的活跃度越高，而等级低的用户的活跃天数少，不排除其为了炒话题从而存在活跃时间集中而短暂的特点。

3)综合素质评估(QE)。该评估利用 3 个属性：用户等级是否大于 15(G)，是否开通会员(M)，阳光信用是否为极好、较好或一般(C)。设置经过微博认证的用户的阳光信用为极好。微博等级是展现微博用户在微博的活跃程度和成长的一个属性。微博用户在想要拥有更多特权和福利并且完善自己对账户的经营度时会考虑开通会员，而水军用户更多的是使用未开通会员的账户，不太注重账户本身享有的特权和福利。微博阳光信用是对每一个微博用户的发布的微博、对其他微博点赞、评论、转发、是否实名制、是否与其他用户建立起良好的社交关系、是否存在违规记录等信息进行综合评估的指标，用户的阳光信用较高说明该用户发表的微博数量较多、社交关系良好、可信任度很高等。

$$QE=0.3G+0.2M+0.5C \tag{2}$$

4)@关系(R)。用户在发布微博时，为了提醒与该微博相关的用户及时查看该微博内容或者应某些博主要求进行抽奖等活动时，往往需要@其他用户。被@的用户登录微博时能够看到该微博的提醒信息，提高了沟通的效率。为了衡量微博中@的情况，定义@比例指标如下：

$$@\text{比例}=\frac{\text{该用户含@的微博数}}{\text{选取的该用户发表的微博数}} \tag{3}$$

5)每日发表微博数(AW)。其为该用户在一段时间内发布的微博数(包括原创、转发等)与该用户发表微博的天数差的比值。

$$\text{每日发表微博数}=\frac{\text{该用户发表的微博数}}{\text{该用户发表微博的天数差}} \tag{4}$$

6)微博影响力(WI)。微博中用户发表的微博拥有一定的阅读量，较多用户在浏览结束后会在其微博中留下评论，以与该用户交流想法、表达自己的观点，有些用户会为其点赞或者转发该条微博来表达对此微博内容的支持。该账户在一定程度上在微博上扩散了自己的影响力，获得了较多的关注。为了测评该账户的影响力，定义账户影响力为该账户发表的微博的转发数(TN)、评论数(C)、点赞数(L)之和。其计算公式如下：

$$\text{微博影响力}=\text{转发数}+\text{评论数}+\text{点赞数} \tag{5}$$

3.2 算法框架

半监督协同训练算法的框架如图 1 所示。该算法利用微博用户数据具有多个属性的特点，将属性分为两个属性集，每个属性集构成了一个视图。在每个视图中，利用有标记数据集训练出一个分类器。选取一部分未标记数据集放入缓冲池，两个分类器各自从缓冲池中挑选 p 个分类置信度高的正例和 n 个分类置信度高的反例，为其赋予伪标记，并对这些数据进行分类和标记，在处理完成后将这些分类好的数据返回到有标记数据集，再在无标记数据集中取出一部分数据放入缓冲池，至此完成一次迭代。在下一轮迭代中，新的有标记数据集将重新训练两个分类器，训练出来的分类器将利用缓冲池中的未标记数据进行新的训练。在每轮迭代中如果新训练的两个分类器都没有发生变化则停止迭代，或者如果达到预先设定好的迭代次数则停止训练。

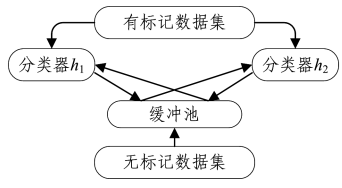

图 1 半监督协同训练算法框架

Fig. 1 Semi-supervised collaborative training algorithm framework

3.3 算法的具体步骤

半监督协同训练算法的具体步骤如下。利用有标记的数

据集 L 和未标记数据集 U 从未标记数据集中随机选取大小为 $2(p+n)$的数据放入缓冲池 U_1，P 和 n 分别为正例数和反例数，设置迭代次数为 3。在每次迭代中，先利用有标记数据集 L 中的两个子数据集 x_1 和 x_2 分别训练得到新的分类器 h_1 和 h_2，h_1 和 h_2 分别挑选分类置信度高的大小为$(p+n)$的正反例给对方用于训练更新。如果两个分类器没有发生变化则迭代停止，否则将标记好的数据放入有标记数据集 L 中，再次从未标记数据集 U 中随机挑选大小为 $2(p+n)$的样本放入缓冲池中。如果达到预先设定的迭代次数或者未标记数据集 U 为空时，则算法完成。具体步骤如算法 1 所示。

算法 1　半监督协同训练算法

输入：有标记样本集 L，未标记样本集 U，大小为(2p+2n)的缓冲池 U_1，每轮挑选的正例数 p，每轮挑选的反例数 n

输出：迭代次数 k，分类器 h_1 和 h_2 的两个 F1-measure 的平均值，两个准确率的平均值，两个召回率的平均值，两个精度的平均值

1. For k=1,2,3 do
2. $x_1 \leftarrow L, x_2 \leftarrow L$
3. x_1 训练得到 h_1，x_2 训练得到 h_2。
4. h_1 在 U_1 中挑选 p 个置信度高的正例和 n 个置信度高的反例，并为其赋予伪标记，给 h_2。
   h_2 在 U_1 中挑选 p 个置信度高的正例和 n 个置信度高的反例，并为其赋予伪标记，给 h_1。
5. 如果 h_1 和 h_2 没有发生改变，break；
   否则，将 x_1，x_2 的标记好的样本加入 L。
6. 从 U 中取出(2p+2n)个未标记样本放入缓冲池 U_1。
7. End for

4　实验

本节将详细介绍微博用户的数据集来源、对微博用户数据的属性特征值的选择以及判断微博用户水军识别分类器分类性能的 4 个指标。设置迭代次数 k 为 3，并采用 Scikit-Learn 机器学习库中的 7 种分类算法对微博用户进行水军识别。

本文进行了 3 个实验。实验 1 设置两个视图，选择相同的分类算法训练分类器，在 3 次迭代中将对 7 种分类算法的准确率、召回率、精度、F1-measure 值进行对比分析。实验 2 利用实验 1 中分类评价指标较高的 3 种分类算法进行实验。实验 2 设置两个视图，采用不同的分类算法训练分类器，经过实验得到分类效果最好的分类算法组合，从而达到准确识别水军的目的。实验 3 选取实验 2 中得到的最好的分类算法组合，将该算法组合与张艳梅等[10]研究的贝叶斯算法均在同一数据集上进行水军识别，并采用相同的 4 个评价指标对实验结果进行分析。

4.1　数据来源

为了使分类模型的分类效果更加准确和贴近实际情况，考虑到微博热门话题具有可以利用水军进行炒作的性质，从新浪微博自身的 API 接口捕获微博用户样本集，在爬取时选取 3 个不同舆论方向的新浪微博的热门话题："偶像练习生""CBA 总决赛""美国空袭叙利亚"。在每个话题主页随机选择部分参与话题的用户，其中用户的个人信息包括用户名、用户 ID、VIP 等级、关注数、粉丝数、微博认证、性别、等级、阳光信用、注册时间。同时，选取了 216 位用户在 2017 年 10 月 15 日至 2018 年 4 月 15 日所发的 32 万条微博，包含的信息有微博发表时间、所用设备、微博内容、该条微博的点赞数、评论数、转发数。为了使人工标记的水军用户和非水军用户的结果更加准确和更有可信度，邀请了微博水军分类识别研究领域的 15 位资深学者对所有微博用户样本数据进行人工识别，最后逐条对每一位用户的标识结果进行最终确认，从而得到了 171 个水军用户样本数据集合与 45 个非水军用户样本数据集合。然后再对人工标记完成后的样本数据进行预处理，将用户的特征属性信息整理为 6 个属性特征值，分别为账户关注度、每日发表微博数、综合素质评估、微博等级、@关系和微博影响力。

实验所用语言为 Python，实验环境为 Windows10 操作系统，Inter(R)Core(TM)i5-4200U cpu @ 1.60GHz 2.30GHz。

4.2　分类性能评价指标

本文参考相关研究[10]，利用分类结果混淆矩阵评估文中分类算法的性能，其中混淆矩阵如表 2 所列。TP(True Positive)表示实际为水军的微博用户被正确分类为水军用户。FP(False Positive)表示该用户实际是非水军用户，但是被错误地判断为水军用户。FN(False Negative)表示该微博用户实际是水军用户，但是被判断为非水军用户。TN 表示(True Negative)表示该微博用户实际为非水军用户，同时也被正确地判断为非水军用户。

表 2　分类结果的混淆矩阵

Table 2　Confusion matrix of classification results

真实情况	分类结果	
	水军	非水军
水军	TP	FN
非水军	FP	TN

利用该微博用户分类结果的混淆矩阵可以计算得到分类性能评价指标：准确率($Accuracy=\frac{TP+TN}{TP+FP+FN+TN}$)、召回率($Recall=\frac{TP}{TP+FP}$)、精度($Precision=\frac{TP}{TP+FN}$)、$F1\text{-}measure$($F1=2\times\frac{precision\times recall}{precision+recall}$)。

4.3　实验设置

实验中，两个视图将使用 Scikit-Learn 机器学习库中的 7 种分类算法训练分类器：随机森林算法(RF)、朴素贝叶斯算法(NB)、k 近邻分类算法(KNN)、逻辑回归算法(LR)、决策树算法(DT)、支持向量机算法(SVM)、梯度提升决策树算法(GBDT)；4 个分类性能度量指标为：准确率、召回率、精度、F1-measure；设置的迭代次数为 3。另外，实验 1 和实验 2 分别记录了程序处理数据的时间 T_1、3 次迭代过程中训练好的分类器识别水军用户所用的时间 T_2 和半监督协同训练算法每一次迭代的时间 T_3，$T_{总}$为每次程序运行的总时间。

4.3.1　实验 1

本实验设置半监督协同训练算法的两个视图均使用相同

的分类算法训练分类器，对 4 个评价指标进行综合分析对比，最后选择在 4 个评价指标中值较高的 3 种分类算法，利用这 3 种算法进行实验 2。实验 1 的结果如图 2－图 5 所示，Inter1，Inter2，Inter3 表示 3 次迭代过程。图中的横轴表示 7 种机器学习分类算法，纵轴表示分类评价指标。

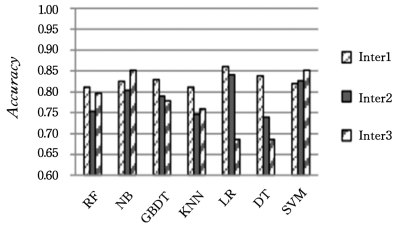

图 2 分类算法的准确率

Fig. 2 Accuracy rate of classification algorithms

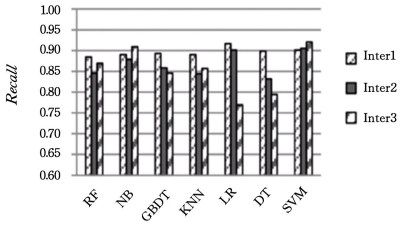

图 3 分类算法的召回率

Fig. 3 Recall rate of classification algorithms

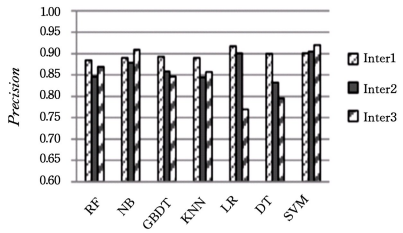

图 4 分类算法的精度

Fig. 4 Accuracy of classification algorithms

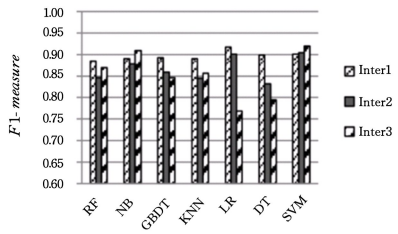

图 5 分类算法的 F1 值

Fig. 5 F1-measure of classification algorithms

由图 2 可知，在第一次迭代过程中，准确率最高的是 LR；在第二次迭代过程中，准确率最高的是 LR，SVM 的准确率略低；在第三次迭代过程中，准确率最高的是 NB 和 SVM，两者的准确率相同，均为最高。由图 3 可知，在第一次迭代过程中，召回率最高的是 LR；在第二次迭代过程中，召回率最高的是 SVM，LR 的准确率略低于 SVM；在第三次迭代过程中，召回率最高的是 SVM，NB 的召回率略低于 SVM。由图 4 可知，在第一次迭代过程中，精度最高的是 LR；在第二次迭代过程中，精度最高的是 SVM，LR 略低于 SVM；在第三次迭代过程中，精度较高的两个分类算法是 NB 和 SVM。由图 5 可知，在第一次迭代过程中，$F1$ 值最高的是 LR，SVM 的值略低于 LR；在第二次迭代过程中，$F1$ 值较高的两个分类算法是 LR 和 SVM；在第三次迭代过程中，$F1$ 值较高的两个分类算法是 NB 和 SVM。

综上，在半监督协同训练算法的两个视图均采用同一种机器学习分类算法 LR，N，SVM 训练分类器时，在 3 次迭代过程中，准确率、召回率、精度和 F1-measure 的值均较高，都在 0.85 左右，因此选择逻辑回归算法(LR)、朴素贝叶斯算法(NB)、支持向量机算法(SVM)进行实验 2。

4.3.2 实验 2

本实验设置半监督协同训练算法的两个视图采用不同的分类算法训练分类器，利用实验 1 中保留下来的 4 个指标较高的 3 种分类算法 LR，NB 和 SVM 进行实验。将 3 种分类算法两两组合，共分为了 6 个组，如表 3 所列。图 6－图 9 中，Inter1，Inter2，Inter3 表示 3 次迭代过程。图中的横轴表示 3 种机器学习分类算法的不同组合，纵轴表示分类评价指标。

表 3 分类算法的组合

Table 3 Combination of classification algorithms

	组 1	组 2	组 3	组 4	组 5	组 6
h_1	NB	LR	LR	SVM	NB	SVM
h_2	LR	NB	SVM	LR	SVM	NB

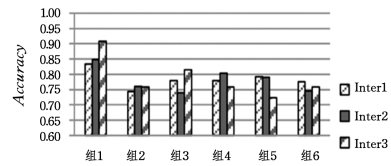

图 6 分类算法的准确率

Fig. 6 Accuracy rate of classification algorithms

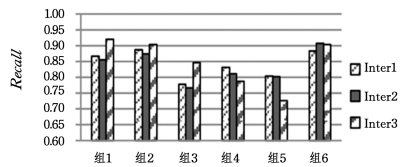

图 7 分类算法的召回率

Fig. 7 Recall rate of classification algorithms

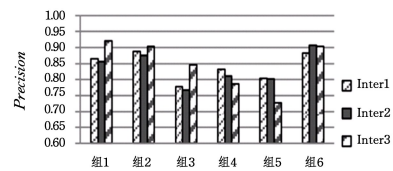

图 8 分类算法的精度

Fig. 8 Accuracy of classification algorithms

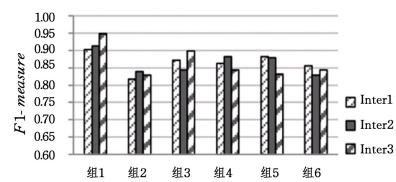

图 9 分类算法的 $F1$ 值

Fig. 9 $F1$-measure of classification algorithms

由图 6 可知，6 个不同组合的准确率在 3 次迭代过程中的变化很大。在 3 次迭代过程中，准确率最高的是组合 1，即 h_1＝NB 和 h_2＝LR，该组的准确率最高达到 90%，最低也在 83%以上。而组合 2 在 3 次迭代过程中的准确率都较低，其他组合的准确率也都比组合 1 低很多。由图 7 可知，在第三次迭代过程中 6 个不同组合的召回率变化很大，组合 5 在第三次迭代过程中的召回率最低。在第一、第二次迭代过程中，组合 6（h_1＝SVM，h_2＝NB）的召回率维持在 90%左右，组合 1（h_1＝NB，h_2＝LR）在第一、第二次迭代过程中的召回率大于 85%，在第三次迭代过程中的召回率大于 90%。经综合分析，组合 1 和组合 6 的召回率在 6 组中的表现较好。由图 8 可知，在第一、第二次迭代过程中组合 6（h_1＝SVM，h_2＝NB）的精度在 6 个组合中最高，在第三次迭代过程中组合 5 的精度最低，组合 1（h_1＝NB，h_2＝LR）的精度最高大于 90%，也是在 6 个组合的 3 次迭代过程中精度最高的。由图 9 可知，在 3 次迭代过程中组合 1（h_1＝NB，h_2＝LR）的 $F1$ 值最高且均大于 90%，组合 2（h_1＝LR，h_2＝NB）的 $F1$ 值最低，其他组合的 $F1$ 值在 3 次迭代过程中相差不多。

综上，在实验二中，组合（h_1＝NB，h_2＝LR）在 3 次迭代过程中的 4 个度量指标均较高，因此两个视图分别使用朴素贝叶斯算法、逻辑回归算法时能够较为准确地对微博水军进行识别。

对两个实验的时间进行对比，由表 4 可知，实验 1 中应用逻辑回归算法的运行总时间最长，应用朴素贝叶斯算法（LR）的运行时间最短；由表 5 可知，实验 2 中分类算法组合 1（h_1＝NB，h_2＝LR）的程序运行总时间最长，与其他组合运行程序所用时间相差较大。

表 4　实验 1 的时间对比

Table 4　Time comparison of experiment 1

（单位：s）

时间		KNN	LR	RF	DT	GBDT	NB	SVM
T_1		0.37	0.36	0.39	0.36	0.36	0.36	0.37
Inter1	T_{21}	0.01	0.10	0.02	0.01	0.02	0.01	0.02
	T_{31}	0.01	0.10	0.03	0.01	0.02	0.01	0.02
Inter2	T_{22}	0.01	0.01	0.02	0.01	0.02	0.01	0.01
	T_{32}	0.02	0.01	0.02	0.01	0.02	0.01	0.01
Inter3	T_{23}	0.01	0.01	0.03	0.01	0.02	0.01	0.01
	T_{33}	0.01	0.01	0.03	0.01	0.02	0.01	0.01
$T_{总}$		0.41	0.48	0.47	0.40	0.42	0.39	0.41

表 5　实验 2 的时间对比

Table 5　Time comparison of experiment 2

（单位：s）

时间		组合					
		组 1	组 2	组 3	组 4	组 5	组 6
T_1		0.44	0.38	0.37	0.36	0.39	0.39
Inter1	T_{21}	0.01	0.01	0.01	0.01	0.01	0.01
	T_{22}	0.01	0.01	0.01	0.01	0.01	0.01
Inter2	T_{31}	0.01	0.01	0.01	0.01	0.01	0.01
	T_{32}	0.01	0.01	0.01	0.01	0.01	0.01
Inter3	T_{23}	0.01	0.01	0.01	0.01	0.01	0.01
	T_{33}	0.01	0.01	0.01	0.01	0.01	0.01
$T_{总}$		0.47	0.41	0.40	0.39	0.42	0.42

4.3.3　实验 3

为了实现半监督协同训练算法与其他算法在水军识别领域的对比，本实验设置半监督协同训练算法与张艳梅等[10]研究的贝叶斯算法进行对比。选取半监督协同训练算法分类效果最好的算法组合朴素贝叶斯算法（NB）和逻辑回归算法（LR），并根据实际情况选取了第三次迭代的结果。将贝叶斯算法选取水军的实验结果与半监督协同训练算法的结果进行对比分析。评价指标为准确率、召回率、精度和 F1-measure，结果如图 10 所示。

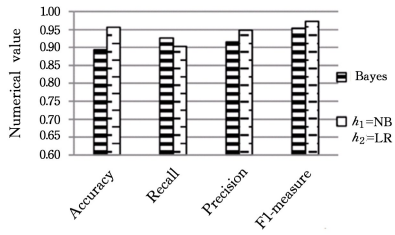

图 10　不同算法的对比

Fig. 10　Comparison of different algorithms

由图 10 可知，在两种算法均采用相同数据集（即 216 位用户的 32 万条微博）的情况下，贝叶斯算法的召回率高于半监督协同训练算法，而半监督协同训练算法的准确率、精度和 $F1$ 值均大于贝叶斯算法。由此可知，在水军识别过程中，半监督协同训练算法模型的水军识别效果在一定程度上比贝叶斯算法模型好，半监督协同训练算法能更好地区分水军和非水军。

结束语　通过对微博水军识别相关研究的学习，本文对微博水军存在的多个属性特征进行分析，利用半监督协同训练算法模型，并使用 Scikit-Learn 机器学习库中 的 7 种分类方法对水军进行识别，最终发现 7 种分类算法中的逻辑回归算法（LR）、朴素贝叶斯算法（NB）、支持向量机算法（SVM）在水军识别的 4 个分类性能评价指标中的值相对较高，并且当两个视图分别使用朴素贝叶斯算法（NB）、逻辑回归算法（LR）训练分类器时能更好地对微博用户进行水军识别。同时，通过研究发现半监督协同训练算法的水军识别效果在一定程度上比贝叶斯算法模型好。然而，研究中使用的数据集较小，另外，本文重点研究离线状态下的微博账户在某时刻下的静态特征，并没有对其在一段时间内或实时状态下的研究价值更高的数据进行分析，没有实现在线动态检测水军账户。因此，下一步的工作重点一方面将利用更大的数据集对水军识别领域进行深入研究，以提升结果的准确性；另一方面将会关注微博水军账户，收集其动态特征数据，并利用更适用的算法实现在线动态检测水军用户，这将推动微博水军识别研究的进步与发展，具有十分重要的意义。

参 考 文 献

[1]　LIU S W，XU Y，WANG B L，et al. Water Army Detection of Weibo Using User Representation Learning[J]. Journal of Intelligence，2018，37(7)：95-100.（in Chinese）
刘姝雯，徐扬，王冰璐，等. 基于用户表示学习的微博水军识别研究[J]. 情报杂志，2018，37(7)：95-100.

[2] CHEN K,CHEN L,ZHU P D,et al. Interaction based on method for spam detection in online social networks[J]. Journal on Communications,2015,36(7):120-128. (in Chinese)
陈侃,陈亮,朱培栋,等. 基于交互行为的在线社会网络水军检测方法[J]. 通信学报,2015,36(7):120-128.

[3] GAYO-AVELLO D,BRENES D J. Overcoming Spammers in Twitter-A Tale of Five Algorithms[C]// CERI. Madrid:Spain,2010:41-52.

[4] CHEN H ,LIU J ,LV Y ,et al. Semi-supervised Clue Fusion for Spammer Detection in Sina Weibo[J]. Information Fusion,2017:S1566253517300714.

[5] ZHANG M L,ZHOU Z H. CoTrade:Confident Co-Training With Data Editing[J]. IEEE Transactions on Systems Man & Cybernetics Part B Cybernetics A Publication of the IEEE Systems Man & Cybernetics Society,2011,41(6):1612-1626.

[6] BLUM A. Combining labeled and unlabeled data with co-training[C]// Conference on Computational Learning Theory. 1998:92-100.

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

[8] HAN Z M,YANG K,TAN X S. Analyzing Spectrum Features of Weight User Relation Graph to Identify Large Spammer Groups in Online Shopping Websites[J]. Chinese Journal of Computers,2017,40(4):939-954. (in Chinese)
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.

[9] KIM C,HWANG K. Naive Bayes Classifier Learning with Feature Selection for Spam Detection in Social Bookmarking[M]. Pennsylvania,USA:Penn State,2008.

[10] ZHANG Y M,HUANG Y Y,GAN S J,et al. Weibo spammers' identification algorithm based on Bayesian model[J]. Journal on Communications,2017,38(1):44-53. (in Chinese)
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] ZHENG X,ZENG Z,CHEN Z,et al. Detecting spammers on social networks[J]. Neurocomputing,2015,159(C):27-34.

[12] YUAN X P,WANG R W,ZHAI B Y. Automatic Recognition of Micro-blog Water Army Based on Multi-index Comprehensive Index Method and Entropy Method[J]. Journal of Intelligence,2014(7):176-179. (in Chinese)
袁旭萍,王仁武,翟伯荫. 基于综合指数和熵值法的微博水军自动识别[J]. 情报杂志,2014(7):176-179.

[13] CHENG X T,LIU C X,LIU S X. Graph-based Features for Identifying Spammers in Microblog Networks[J]. Acta Automatica Sinica,2015,41(9):1533-1541. (in Chinese)
程晓涛,刘彩霞,刘树新. 基于关系图特征的微博水军发现方法[J]. 自动化学报,2015,41(9):1533-1541.

[14] ZHANG L. The Research and Implementation on the Technology of Spammer Detection for Sina Mircoblog[D]. Changsha:National University of Defense Technology,2015. (in Chinese)
张良. 面向新浪微博的水军识别技术的研究与实现[D]. 长沙:国防科学技术大学,2015.

[15] LV C. Research and Implementation of Internet Forum Water Army Detection Based on User Behaiors[D]. Chengdu:Southwest Jiaotong University,2017. (in Chinese)
吕晨. 基于用户行为的网络论坛水军检测研究与实现[D]. 成都:西南交通大学,2017.

[16] BLUM A. Combining Labeled and unlabeled Data with Co-training [C]// Proc. of the Conference on Computational Learning Theory. 1998.

[17] ZHI-HUA Z. Disagreement-based Semi-supervised Learning [J]. Acta Automatica Sinica,2013,39(11):1871-1878.

[18] NIGAM K,GHANI R. Analyzing the effectiveness and applicability of co-training[C]// International Conference on Information and Knowledge Management. ACM,2000:86-93.

[19] PENG Y,ZHANG D Q. Semi-Supervised Canonical Correlation Analysis Algorithm[J]. Journal of Software,2008,19(11):2822-2832. (in Chinese)
彭岩,张道强. 半监督典型相关分析算法[J]. 软件学报,2008,19(11):2822-2832.

[20] LI F,HUANG M,YANG Y,et al. Learning to identify review spam[C]// International Joint Conference on Artificial Intelligence. AAAI Press,2011:2488-2493.

[21] ZHU J. Semi-supervised learning literature survey[J]. Computer Sciences Department,2005,37(1):63-77.