基于逻辑回归的微博用户可信度建模

徐建民[1]，粟武林[1+]，吴树芳[2]，武晓波[1]

（1. 河北大学 数学与计算机学院，河北 保定 071002；2. 河北大学 管理学院，河北 保定 071000）

摘 要：针对微博虚假用户问题，以新浪微博为研究平台，对微博用户的行为进行分析，从在线时长、发帖时间、互动程度等方面，提取用于区分用户类别的特征变量，运用逻辑回归算法，提出一个基于逻辑回归的微博用户可信度评价模型。实验结果表明，该模型能够对传统的虚假用户“僵尸粉”进行识别，对新型虚假用户有较高的识别率，可以根据置信值的大小对用户进行大致分类，实用性较强。

关键词：微博；逻辑回归；可信度；虚假用户；僵尸粉

中图法分类号：TP391 **文献标识号**：A **文章编号**：1000-7024（2015）03-0772-06

doi：10.16208/j.issn1000-7024.2015.03.042

Modeling user reliability based on logistic regression in micro-blog

XU Jian-min[1], SU Wu-lin[1+], WU Shu-fang[2], WU Xiao-bo[1]

(1. College of Mathematics and Computer, Hebei University, Baoding 071002, China;
2. College of Management, Hebei University, Baoding 071000, China)

Abstract: For the problem of fake users, behavioral characteristics of Sina micro-blog users were analyzed, and then feature variables for distinguishing the category of users were extracted according to online time, posting time, interaction degree and so on. Applying logistic regression algorithm, a user reliability evaluation model was put forward. The experimental result shows that the proposed model not only can recognise the traditional fake users zombie fans, but also has high recognition rate for new fake users. Sina micro-blog users can be roughly classified according to the confidence. The proposed model is stronger on practicability.

Key words: micro-blog; logistic regression; reliability; fake users; zombie fans

0 引 言

目前，学术界对于如何解决微博虚假用户问题[1]已有了一些研究成果。原福永等[2]从链接分析的角度提出了一种用户被关注度的概念，采用不平均分配权值的 PageRank 算法思想来解决虚假用户问题。方明等[3]采用文本分类的思想，对微博用户名进行 13 维的特征抽取，建立微博用户名特征库，并结合 SVM 和 ANN 等方法对微博用户进行智能分类。Chu 等[4]利用信息熵结合机器学习的方法对 Twitter 上的虚假用户进行识别，也取得了一定效果。但这些方法都局限于对“僵尸粉”进行识别，而对目前广泛流行的“活粉”的识别率并不高。

本文针对目前已有识别方法的不足，对微博用户的行为进行分析，提取了相应的特征变量，运用逻辑回归算法，提出了一个微博用户可信度评价模型，对微博用户的可信度进行度量，并根据置信值给出一个用户为真实用户的可能性，从而对虚假用户进行有效的识别。实验结果表明，该方法能较好识别微博中的虚假用户，是对虚假用户当中的“活粉”也有较高的识别率。

1 相关知识及准备工作

1.1 微博虚假用户

关于微博虚假用户一词，目前学术界使用的还较少，此前研究多使用“僵尸粉”一词。对于“僵尸粉”的概念，在百度百科上将其定义为：微博上的虚假粉丝，指花钱就可以买到“关注”，有名无实的微博粉丝，它们通常是由系

收稿日期：2014-05-22；**修订日期**：2014-07-25

基金项目：河北省自然科学基金项目（F2011201146）

作者简介：徐建民（1966-），男，河北邯郸人，教授，博士生导师，研究方向为信息检索、不确定信息处理；+通讯作者：粟武林（1987-），男（侗族），湖南怀化人，硕士研究生，研究方向为信息检索与信息系统；吴树芳（1980-），女，河北邯郸人，博士研究生，研究方向为信息检索与信息系统；武晓波（1986-），男，河北邯郸人，硕士研究生，研究方向为信息检索与信息系统。

E-mail: 380903647@qq.com

统自动产生的恶意注册用户[5]。但是，从“僵尸粉”的发展变化来看，传统的“僵尸粉”已经逐渐退出了微博的舞台，取而代之的是可以跟真实用户一样进行发微博，转发、评论他人微博的“活粉”了，其存在的目的已经不仅限于增加特定人的粉丝数量。显然，“僵尸粉”一词已经不再适用于微博的现状了。所以，本文结合当前这类用户的行为特征，对其进行重新定义。

定义 1 微博虚假用户：在微博网络中，以某种利益为目的，由特定的算法产生，进而去关注指定的用户，并且能以特定的算法或人工操控的方式去模拟真实用户进行发帖，转发、评论微博，以此来维持自身一定活跃度的用户。这类用户不具备感情、逻辑和互动性。

1.2 逻辑回归模型

对微博用户的身份类别进行预测，是一种典型的二分类问题。逻辑回归（logistic regression）是一种可以用来分类的常用统计分析方法，并且可以得到概率型的预测结果，属于一种概率型非线性回归[6-8]。

考虑具有 n 个变量的向量 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$，设条件概率 $\boldsymbol{P}(Y=1\mid \boldsymbol{x})=p$ 为根据观测量相对于某事件发生的概率，则逻辑回归模型可表示为

$$\mathrm{P}(Y=1\mid \boldsymbol{x})=\pi(\boldsymbol{x})=\frac{1}{1+e^{-g(x)}} \tag{1}$$

其中，$g(\boldsymbol{x})=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n$，$\beta_0$ 为截距项，$\boldsymbol{\beta}=(\beta_1,\beta_2,\cdots,\beta_n)$ 为自变量的回归系数。显然，$\pi(\boldsymbol{x})$ 的值域为 [0，1]，因此我们可以根据其取值来估计因变量 $Y=1$ 时发生的概率。对逻辑回归模型的参数进行估计，通常采用极大似然函数法。设 y 是 0－1 型变量，m 个观测值为 $\{y_1,y_2,\cdots,y_m\}$，于是 m 个观测值的似然函数为

$$L(\boldsymbol{\beta})=\prod_{i=1}^{m}p(y_i)=\prod_{i=1}^{m}\mathrm{p}(\boldsymbol{x})^{y_i}[1-\mathrm{p}(\boldsymbol{x})]^{1-y_i} \tag{2}$$

对式（2）两边求自然对数，可得对数似然函数

$$\ln L=\sum_{i=1}^{m}[y_i(\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_nx_{in})-\ln(1+e^{\beta_0+\beta_1x_{i1}+\beta_2x_{i2}+\cdots+\beta_nx_{im}})] \tag{3}$$

最大似然估计就是选取 $\beta_0,\beta_1,\beta_2,\cdots,\beta_n$ 的估计值，使得 $\ln L$ 的值最大化。对式（3）求导，应用牛顿－拉斐森（Newton-Raphson）方法进行迭代求解，即可获取模型的截距项和回归系数。将求得的参数代入式（1）即可建立逻辑回归的预测模型。

1.3 样本数据的形成

本文以国内目前应用最为广泛的新浪微博作为研究平台，通过新浪微博官方提供的 API 数据接口和自行设计的微博爬虫程序相结合的方式，有针对性的采集了不同种类的 10 000 多个微博用户的相关信息，去噪、过滤后筛选出有效用户信息 10 279 个。经人工辨别分类后分别获得了 5755 个真实用户信息和 4524 个虚假用户信息，将这两类用户信息各提取 50％进行混合形成训练样本集，共计 5140 个用户信息，剩余部分作为测试样本集，共计 5139 个用户信息。

2 微博用户可信度评价模型

2.1 模型的构建

定义 y_u 为样本数据中用户的类别，当 $y_u=1$ 时代表用户为真实用户，$y_u=0$ 时代表用户为虚假用户，置信值 CM（confidence measure）定义为值域在［0，1］之间的一个实数，用来表征微博用户 y_u 的可信度大小，其值越大表示用户 y_u 为真实用户的可能性越高[9-11]。设 $\boldsymbol{x}=(x_1,x_2,\cdots,x_n)$ 为对用户 y_u 行为特征进行提取的 n 个变量，则由式（1）可得微博用户置信值 CM 的计算公式为

$$CM=\mathrm{P}(y_u=1\mid \boldsymbol{x})=\frac{1}{1+e^{-(\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_nx_n)}} \tag{4}$$

通过样本数据对模型进行训练，采用极大似然法即可求出模型各个参数的估计值，从而得到最终的微博用户可信度评价模型 WUREM（Weibo user reliability evaluation model），具体流程如下：

（1）训练流程

步骤 1 样本数据的处理，对样本数据进行去噪、筛选和分类。

步骤 2 模型自变量的提取，对微博用户的行为进行分析，提取合适的特征变量。

步骤 3 自变量的筛选，选择显著性强的变量作为模型的最终变量。

步骤 4 回归系数的求解，采用最大似然函数对回归系数进行迭代求解。

步骤 5 将自变量和回归系数代入式（4），形成最终的微博用户可信度评价模型。

（2）分类流程

步骤 1 对测试用户的行为特征进行量化，形成模型的自变量值。

步骤 2 运用 WUREM 模型计算用户为真实用户的概率，即用户的置信值 CM。

步骤 3 根据设定的分类标准值（阈值）进行微博用户类别的判定。

2.2 模型自变量的提取

对于特征的提取有很多方法，常见的有主成分分析法（principal component analysis，PCA）、信息增益（information gain，IG）、互信息（mutual information，MI）等。然而对于微博用户身份的鉴别，单从用户的原始信息特征很难较好的进行区分，传统的特征提取方法也都不适用。因此，本节在基于对微博用户行为分析的基础上，通过合理的逻辑归纳，对微博用户的原始特征进行组合和变换，提取了以下特征作为模型的自变量，并对变量的取值给出了具体的计算公式。

（1）亲近率（ufr，ufa）

亲近率表征了用户对关注对象或者粉丝对用户的重视程度。用户与关注对象的亲近率的大小由用户转发关注对象的微博数来衡量，从转发的深度和广度两个方面考虑，即用户转发过的微博用户中，关注对象越多，转发关注对象的微博数越多，其亲近率也就越高。用户与关注对象之间的亲近率 ufr 计算如下

$$ufr = \frac{Rt_{FromFr}}{Rt_{Wb}} \times \frac{Fr_{Rt}}{Fr} \times \eta \tag{5}$$

式中：Rt_{FromFr} ——用户转发来自关注对象的微博数，Rt_{Wb} ——用户所有转发的微博数，Fr_{Rt} ——用户转发过的关注对象数，Fr ——用户的关注对象数，η ——修正值。研究发现，虚假用户转发的微博大多呈现出两种极端：一是转发微博的对象基本不重复；二是转发的微博集中来自某几个对象。而对于真实用户来说，如果将其按每个转发对象被转发的微博数进行排序，则呈现出明显的长尾分布，如图 1 所示。

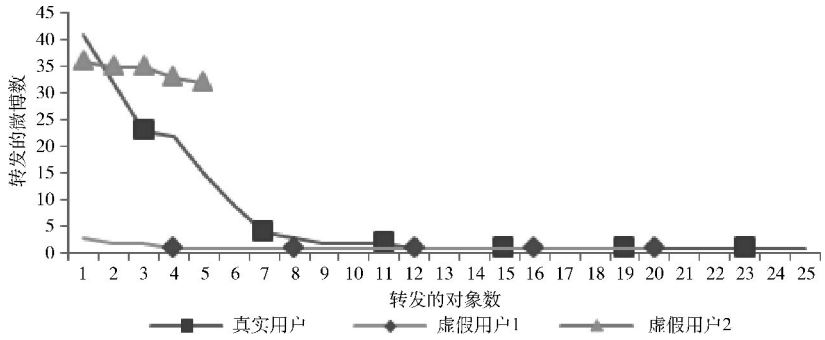

图 1 用户转发关注对象的微博数

从图 1 中可以看出，真实用户转发微博的对象集中在少数关注对象中，大部分关注对象的微博只被转发一次或没有被转发。对于用户转发关注对象的微博数是否服从长尾分布的判断，可以近似的用 $\eta = \frac{1}{1+e^{-(\sigma-2)}}$ 来代替，其值域约为 (0.1，1)，σ 为转发对象之间被转发微博数的离散度，即标准差 $\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$，其中 X_i 表示第 i 个转发对象的被转发微博数，n 为用户转发过的关注对象数。因此，修正值 η 的意义可以定义为：当用户转发关注对象的微博数的离散度越大时，其行为越接近真实用户，η 的值就越接近 1，反之 η 的值越小。

同理，我们可以得出粉丝与用户之间的亲近率 ufa 为

$$ufa = \frac{Rtd_{ByFa}}{Rtd_{Wb}} \times \frac{Fa_{Rtd}}{Fa} \times \rho \tag{6}$$

式中：Rtd_{ByFa} ——用户被粉丝转发的微博数，Rtd_{Wb} ——用户所有被转发的微博数，Fa_{Rtd} ——转发过用户微博的粉丝数量，Fa ——用户的粉丝数，ρ ——亲近率的修正值。

(2) 互动率（Ir）

互动率表征了用户和关注对象、粉丝之间进行逻辑交流的程度，这在很大程度上表明了用户身份的真实性，虚假用户是不具备逻辑交流能力的。由于新浪微博对数据获取的限制，实验中无法获取用户发出去的评论信息，因此作如下约定：在用户 A 的微博中有形如：用户 A：回复@用户 B、用户 B：回复@用户 A 的评论就认为用户和粉丝之间完成了一次互动行为，用户的互动率 Ir 可以计算如下

$$Ir = \frac{1}{w_c}\sum_{i=1}^{n}\left(\frac{c_{urf}^i}{c_{fcu}^i} + \frac{c_{fru}^i}{c_{urf}^i}\right) \tag{7}$$

式中：w_c ——用户有评论的微博数，n ——有互动行为的微博数，c_{urf}^i、c_{fru}^i ——用户的第 i 条微博中用户回复粉丝评论数和粉丝回复用户评论数，$c_{fcu}^i = c^i - c_{urf}^i - c_{fru}^i$ 为用户的第 i 条微博中粉丝的有效评论数，c^i 为第 i 条微博的评论数。

(3) 平均时间开销（$\overline{T_{wb}}$）

用户每发一条微博所需的时间越长，说明微博发布者对微博内容的重视性越高，为真实用户的可能性也越高，其值可以表示为

$$\overline{T_{wb}} = \overline{T}/w_f \tag{8}$$

式中：$\overline{T}$，w_f ——用户平均在线时长和微博的发布频率，特别的，$w_f = w/d_w$，w 为用户的微博数，d_w 为用户有微博发布的登录天数。由于微博平台不提供用户在线时长的相关数据，实验中采用用户登录的平均活跃天数来代替在线时长，即 $\overline{T} = (T_a - T_j)/d$，$T_a$ 为用户的活跃天数，T_j 为用户奖励得到的活跃天数，d 为当前时间至用户注册时间的间隔天数。

(4) 行为与反馈的相关度（r_u）

用户行为指的是发布微博，反馈指的是用户发布的微博被赞、转发或评论。当用户的微博被反馈的越多时，用户发微博的积极性越高，反之用户使用微博的积极性会越来越低，而虚假用户不具备这一特征。将用户发布的微博按一定的时间段 t 天（实验中设 $t = 10$）进行分割，可以得到个 n 组相同时间段内发布的微博数（$x_1, x_2, \cdots, x_n$）和对应的被反馈的微博数（$y_1, y_2, \cdots, y_n$）。用户行为与反馈的

相关度就可以表示为变量 x 和变量 y 之间的相关系数与用户的平均反馈率的乘积，其计算公式为

$$r_u = \bar{g} \times \sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y}) / \sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2 \sum_{i=1}^{n}(y_i-\bar{y})^2} \tag{9}$$

其中：平均反馈率 $\bar{g} = \frac{1}{n}\sum_{i}^{n}(wf_i/w_i)$ 表示用户在 n 组时间段内被反馈的微博数 wf_i 和所有发布的微博数 w_i 比值的平均值。从上式可以看出，用户的微博与被反馈的微博之间相关系数越高，被反馈的微博所占的比重越大，其行为与反馈的相关度就越高。

（5）发帖行为规律性（Lt）

用户的发帖行为越有规律，说明其发帖行为受算法操控的可能性越大，如果一个用户经常 24 小时不间断发帖，则其为虚假用户的可能性极大。为了量化用户发帖行为的规律性，我们作如下定义：

定义 2 命中次数：将 24 小时从 02：00 时开始平均分为 6 个时间段，用户发布微博的时间只要在任意一个时间段出现一次，则当天的命中次数加 1，重复出现在同一个时间段内，命中次数不变，用 ε_k 表示命中 k 次以上的天数。

定义 3 发帖时间的平均离散度：设 Δt（分钟）表示一天中任意两条连续发布微博的时间间隔，用户当天的发帖时间离散度就可以表示为 n 个 Δt 的标准偏差，即：$\sigma_t = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\Delta t_i - \overline{\Delta t})^2}$，从而可以得到用户发帖时间的平均离散度 $\overline{\sigma_t} = \frac{1}{d_w}\sum_{i=1}^{d_w}\sigma_{ti}$，$d_w$ 为用户有微博的天数。

根据以上定义，我们将用户发帖行为规律性定义为一个分类变量 Lt，用不同的取值代表用户发帖行为规律性的程度，具体取值及判定条件见表 1。

表 1 用户发帖行为规律性的程度及判定条件

取值	规律性程度	判定条件
5	高度	$\overline{\sigma_t} \leqslant 10$，或 $\varepsilon_6 \geqslant 5$
4	中高度	$10 < \overline{\sigma_t} \leqslant 30$，或 $\varepsilon_5 \geqslant 5$
3	中度	$30 < \overline{\sigma_t} \leqslant 60$，或 $\varepsilon_4 \geqslant 10$
2	中低度	$60 < \overline{\sigma_t} \leqslant 120$，或 $\varepsilon_4 \geqslant 5$
1	低度	其它

备注：若判定条件有多个同时满足，则取值选最大值

从表 1 中我们可以看出，如果一个用户的发帖时间间隔非常有规律，即 $\overline{\sigma_t} \leqslant 10$，或者 24 小时都在发帖的天数在 5 天以上，即 $\varepsilon_6 \geqslant 5$，则该用户的发帖行为将被判定为具有高度规律性。

2.3 模型自变量的参数估计及显著性检验

为简化计算过程，我们利用 IBM SPSS Statistics 19.0 软件中的二元 Logistic 回归分析模块对模型进行训练，训练集采用 1.3 节中已获取的 5140 个用户数据，计算时我们将变量的选择方法设置为“进入”，分类标准值设定为 0.5，其它设置均为默认值，计算结果见表 2。

表 2 方程中的变量

变量说明	B	S.E.	Wals	df	Sig.	Exp(B)
用户与关注亲近率（ufr）	32.089	4.978	41.554	1	.000	8.62E13
用户与粉丝亲近率（ufa）	3.717	1.806	4.236	1	.040	41.120
互动率（Ir）	6.844	.972	49.535	1	.000	938.072
平均时间开销（$\overline{T_{wb}}$）	11.654	1.634	50.847	1	.000	1.15E5
行为与反馈相关度（r_u）	-2.454	.464	27.968	1	.000	.086
发帖行为规律性（Lt）	-.455	.087	27.107	1	.000	.635
常量（β_0）	-.863	.264	10.672	1	.001	.422

从表 2 中我们可以看出所有变量的 Sig 值均小于临界值 0.05，说明模型是可靠的，每一个变量都可以作为模型的最终变量。B 值即为各变量的系数，由此我们可以得出模型的参数估计值见表 3。

表 3 参数估计值

参数	β_0	$\beta_1(ufr)$	$\beta_2(ufa)$	$\beta_3(Ir)$	$\beta_4(\overline{T_{wb}})$	$\beta_5(r_u)$	$\beta_6(Lt)$
估计值	-0.863	32.089	3.717	6.844	11.654	-2.454	-0.455

从参数的估计结果来看 $\beta_1 \sim \beta_4$ 均为正，β_6 为负，说明亲近率、互动率和平均时间开销越大，发帖行为规律性越低，其置信值越高，符合之前的分析结果。然而 β_5 为负，说明行为相关度越低，其置信值越高，这与之前的分析预测正好相反。对样本数据进行分析可知，这是由于数据中有大量人工虚假账户的微博被很有规律的转发，使得其行为相关度反而普遍都高于了真实用户，从而导致参数估计值为负值，因其 Sig＝0.000，说明变量的显著性很高，故不影响在模型中的使用。

2.4 模型特殊情况的处理

在 2.2 节中变量的取值都是基于用户的微博计算的，很显然用户的微博数越多，其计算结果越准确。但是对于新注册的用户，以及围观型（只看微博，从不发微博）的用户其微博数并不多，有些甚至为零，运用上述方法来计算变量的取值显然不合理，因而模型的最终计算结果也不够准确。因此，本文将模型作如下改进：

当用户的微博数 $w \leqslant 10$ 时，设 T_x 为用户最后一条微博的时间与当前时间的间隔天数，如果 $w = 0$，则 T_x 为用户注册时间与当前时间的间隔天数，T_a 为用户的活跃天数，此时用户的置信值 CM 可以表示为一个值域在［0，1］之

间与 T_x 负相关、与 T_a 正相关的幂函数，计算公式如下

$$CM = CM_1 + CM_2 = \frac{1}{e^{T_x/100}} + \left(\frac{2}{1+e^{-T_a/100}} - 1\right) \quad (10)$$

其中，CM_1 与 T_x 负相关，值域为 [1，0]、CM_2 与 T_a 正相关，值域为 [0，1]，时间 T_x 和 T_a 除以 100 是对单位时间的作用进行放大。由于 T_x 和 T_a 不存在相关关系，故可能会出现 $CM > 1$ 的情况，此时我们取 $CM = 1$。可以看出，一个新注册的用户其 T_x 和 T_a 均为 0，置信值 $CM = 1$，即新注册用户默认为真实用户，随着时间的推移，如果用户仍然没有发帖操作，且不登录微博，则置信值将会不断降低，反之，用户虽然不进行发帖，但是经常登录微博，则其置信值将不断增加。

结合 2.3 节的参数估计结果，我们可以得出 WUREM 模型置信值 CM 的最终计算公式为

$$CM = \begin{cases} \dfrac{1}{1+e^{-(-0.863+32.089\cdot ufr+3.717\cdot ufa+6.844\cdot lr+11.654\cdot \overline{T_{wb}}-2.454\cdot r_u-0.455\cdot Lr)}} & w > 10 \\ \dfrac{1}{e^{T_x/100}} + \left(\dfrac{2}{1+e^{-T_a/100}} - 1\right) & w \leqslant 10 \end{cases} \quad (11)$$

3 实验结果与分析

3.1 模型评价指标

对于分类器的分类结果，通常采用如表 4 所示的混淆矩阵表示，性能评价指标通常有如下几个[12]：

正确率 T = 判定正确的样本数/样本总数
$= (a+d)/(a+b+c+d) \times 100\%$

查全率 R = 命中的正样本数/所有正样本数
$= a/(a+b) \times 100\%$

查准率 P = 命中的正样本数/所有判定为 1 的样本数
$= a/(a+c) \times 100\%$

漏检率 M = 遗漏的正样本数/所有正样本数
$= b/(a+b) \times 100\%$

表 4 分类器分类结果

实际类别	预测结果	
	类别 A(1)	类别 B(0)
类别 A(1)	判定正确 a	判定错误 b
类别 B(0)	判定错误 c	判定正确 d

3.2 测试结果

测试数据选择 1.3 节中已经人工分好类的 5139 个用户信息，分类标准值（阈值）设为 0.5，即置信值 $CM \geqslant 0.5$ 则判定为真实用户，测试结果见表 5、表 6。

表 5 WUREM 模型预测结果

实际类别	预测结果		
	真实用户	虚假用户	预测准确率/%
真实用户	2607	270	90.62
虚假用户	187	2075	91.73
总计准确率	—	—	91.11

表 6 评价指标结果/%

正确率 T	查全率 R	查准率 P	漏检率 M
91.11	90.62	93.31	9.38

从表 5 的预测结果可知，测试集中共有真实用户 2877 个，被判定为真实用户的有 2607 个，预测准确率为 90.62%。共有虚假用户 2262 个，被判定为虚假用户的有 2075 个，预测准确率为 91.73%，说明模型对每一类的预测能力都比较强，且对虚假用户的判定性能要高于对真实用户的判定。从表 6 的结果中我们可知，WUREM 模型的正确率 T、查全率 R、查准率 P 都超过了 90%，而漏检率 M 则低于 10%，说明模型的性能良好，可靠性强。

3.3 类别区分能力测试

为了更好的验证模型对于不同类别用户的区分能力，我们另外随机采集了微博中常见的 4 类虚假用户“机器广告、人工活粉、机器活粉、僵尸粉”以及真人广告和真实用户各 500 名进行测试，运用 WUREM 模型对其置信值 CM 进行计算，并将计算结果进行排序，其概率曲线如图 2 所示。

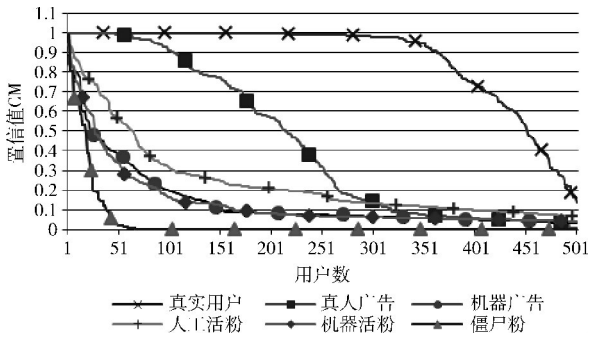

图 2 不同类别用户的置信值 CM

从图 2 中可以看出 WUREM 模型不仅对区分真实用户和各种类型的虚假用户有良好的性能，而且对于各类虚假用户之间也有较好的区分能力。但是模型对于真人广告型的用户区分能力较差，分析可知这是由于各真人广告型用户的发帖方式、登录情况、互动程度等都不尽相同，没有明显的规律性，所以导致真人广告型用户的置信值 CM 在图 2 中略成直线，因此对于真人广告型用户我们不能简单的把它归结为真实用户或者虚假用户。同时，模型对于机

器广告型和机器活粉型虚假用户的区分能力也较差，分析可知这是因为这两类用户的行为方式高度相似，只是发帖的内容不同。

另一方面，从图 2 中还可以看出虽然各类虚假账户的置信值 *CM* 基本在 0.5 以下，但其 *CM* 值所属的范围明显不同，且有：真人广告＞人工活粉＞机器广告、机器活粉＞僵尸粉，这说明用户的行为越接近真实用户，其置信值就越高，验证了文章之前的分析，说明 WUREM 模型用于衡量用户的可信度是可靠的。

4 结束语

本文从分析微博用户的行为特征出发，运用逻辑回归算法，提出了一个基于逻辑回归的微博用户可信度评价模型 WUREM。从实验结果来看，WUREM 模型不仅对传统的“僵尸粉”有较好的识别能力，而且对目前微博中流行的各种“活粉”也能较好的进行识别。同时，我们还可以根据用户的置信值 *CM* 来度量其为真实用户的可能性大小，而不是简单的将其分为真实或者虚假两类用户，这将在研究微博信息可信度、微博用户影响力、微博热度等方面提供必要参考[13,14]。但是模型对微博中的广告用户的区分能力并不理想，同时由于置信值 *CM* 是基于用户微博计算的，对于微博数量特别多或者微博被评论、被转发量特别大的用户，在原始数据的获取上效率比较慢，因此模型并不适用于这类用户的判别，这些都是下步工作需要解决的问题。

参考文献：

[1] WANG Peng. Sina-Weibo false users as much as hundreds of millions and mostly for internal staff to cultivate for profit [EB/OL]. [2013-02-22]. http://tech.ifeng.com/internet/detail_2013_02/22/22375816_0.shtml(in Chinese).[王鹏. 新浪虚假微博用户多达上亿多为内部人员培植以牟利[EB/OL].[2013-02-22]. http://tech.ifeng.com/internet/detail_2013_02/22/22375816_0.shtml.]

[2] YUAN Fuyong, FENG Jing, FU Qianqian, et al. A method to reduce the impact of zombie fans in micro-blog [J]. New Technology of Library and Information Service, 2012, 28 (5): 70-74 (in Chinese). [原福永，冯静，符茜茜，等. 一种降低微博僵尸粉影响的方法 [J]. 现代图书情报技术，2012，28 (5)：70-74.]

[3] FANG Ming, FANG Yi. A new intelligent recognition method of zombie fan [J]. Computer Engineering, 2013, 39 (4): 190-193 (in Chinese). [方明，方意. 一种新型智能僵尸粉甄别方法 [J]. 计算机工程，2013，39 (4)：190-193.]

[4] Zi Chu, Gianvecchio S, Wang Haining, et al. Detecting automation of twitter accounts: Are you a human, bot, or cyborg? [J]. Dependable and Secure Computing, 2012, 9 (6): 811-824.

[5] Baidu encyclopedia: Zombie fan [EB/OL]. [2013-09-14]. http://baike.baidu.com/view/4047998.htm(in Chinese).[百度百科：僵尸粉[EB/OL].[2013-09-14]. http://baike.baidu.com/view/4047998.htm.]

[6] Musa A B. Comparative study on classification performance between support vector machine and logistic regression [J]. International Journal of Machine Learning and Cybernetics, 2013, 4 (1): 13-24.

[7] Dou Huili, Wang Guohua, Guo Min. Algorithm of traffic state probability forecast based on logistic regression [C] //International Conference on Electronics, Information and Communication Engineering. New York: ASME, 2012: 99-104.

[8] WU Kai, JI Xinsheng, LIU Caixia. Modeling information diffusionbased on behavior predicting in microblog [J]. Application Research of Computers, 2013, 30 (6): 1809-1812 (in Chinese). [吴凯，季新生，刘彩霞. 基于行为预测的微博网络信息传播建模 [J]. 计算机应用研究，2013，30 (6)：1809-1812.]

[9] Gupta A, Kumaraguru P. Credibility ranking of tweets during high impact events [C] //Proceedings of the 1st Workshop on Privacy and Security in Online Social Media. New York: ACM, 2012.

[10] CHEN Ke, BIAN Xianhua. Research on reliability based on text semantic and social structure [J]. Application Research of Computers, 2013, 30 (8): 2334-2336 (in Chinese). [陈珂，卞先华. 基于文本语义和社会结构的可信度研究 [J]. 计算机应用研究，2013，30 (8)：2334-2336.]

[11] Gupta M, Zhao Peixiang, Han Jiawei. Evaluating event credibility on twitter [C] //Proceedings of the Twelfth SIAM International Conference on Data Mining. Anaheim: Omni Press, 2012: 153-164.

[12] FENG Guohe. Review of performance evaluation of text classification [J]. Journal of Intelligence, 2011, 30 (8): 66-70 (in Chinese). [奉国和. 文本分类性能评价研究 [J]. 情报杂志，2011，30 (8)：66-70.]

[13] JIANG Shengyi, CHEN Dongyi, PANG Guansong, et al. Research review of information credibility analysis on microblog [J]. Library and Information Service, 2013, 57 (12): 136-142 (in Chinese). [蒋盛益，陈东沂，庞观松，等. 微博信息可信度分析研究综述 [J]. 图书情报工作，2013，57 (12)：136-142.]

[14] ZHOU Jinyuan, ZHANG Shasha, LIU Guifeng, et al. Review of the research on microblog in China [J]. Journal of Intelligence, 2013, 32 (9): 46-51 (in Chinese). [周金元，张莎莎，刘桂锋，等. 国内微博研究综述 [J]. 情报杂志，2013，32 (9)：46-51.]