

基于评论图的虚假产品评论人的检测

王 琢 李 准 徐 野 宋 凯

(沈阳理工大学信息科学与工程学院 沈阳 110159)

摘 要 由于网络产品评论信息可以极大地影响产品的销售,因此很多产品评论人故意捧抬或诋毁特定产品来达到其目的。Wang G 等人利用评论图中店铺、评论、评论人之间的相互关系,通过迭代计算得出评论、评论人和店铺的信誉度,从而发现虚假评论人。针对网络中无店铺的购物环境,提出了用产品替代店铺的新评论图结构,设计了一种逐步淘汰评论人及其评论的 ICE 算法,它极大地提高了迭代收敛速度。同时通过改进评论、评论人和产品的评分函数,进一步提高了基于评论图方法检测虚假评论人的准确度。实验表明,ICE 算法不但收敛速度更快,而且具有更高的准确度。

关键词 虚假评论,评论图,观点挖掘

中图分类号 TP181 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2014.10.062

Detecting Product Review Spammers Based on Review Graphs

WANG Zhuo LI Zhun XU Ye SONG Kai

(School of Information Science and Engineering, Shenyang Ligong University, Shenyang 110159, China)

Abstract Online product reviews can significantly affect product sales, resulting in a large number of reviewers who promote and/or demote target products by writing untruthful product reviews. Wang G et al proposed review graphs which reveal the relationships of reviews, reviewers and stores to calculate the reputations of reviews, reviewers and stores by convergent iterative computation, which can capture fake reviewers. To handle the storeless shopping environment, we proposed a new review graph structure by replacing stores with products, and designed a novel Algorithm ICE to fasten the iteration process by eliminating a certain portion of reviewers and reviews during each iteration. Meanwhile, by exploiting new scoring criteria for reviews, reviewers and products, the precision for identifying fake reviewers is also improved. Experiments show that the proposed Algorithm ICE not only performs faster but also more accurately than previous method.

Keywords Fake review, Review graph, Opinion mining

1 引言

网络在线产品销售环境中常常允许用户对产品质量或服务进行评论,这些评论将极大地左右未来顾客对该产品的选择^[1]。受利益驱使,某些评论人故意杜撰不真实的评论,目的是夸大或者诋毁某个产品,使自身利益最大化,甚至存在某些厂家雇佣网络水军,在某个时期(或定期)集中对其产品撰写虚假评论(fake review),亦称评论作弊(review spam)。这些虚假评论不但影响了消费者对产品的正确认识,而且还影响企业的生产销售,对市场造成严重冲击。通过对虚假评论人检测可以将评论集中的虚假评论人及其评论过滤,从而给消费者一个真实的网络购物环境。自从 Liu B 等人首次提出网络虚假评论检测问题以来^[2],各种针对虚假评论人的检测方法应运而生,概括起来主要有以下 4 种。一是基于机器学习的方法对垃圾评论或恶意评论进行标注,然后利用分类器对评论人进行分类^[2-4]。例如 Jindal N 等利用重复评论及其

特征作为训练集建立 logistic 回归分类器,从而预测异常评论^[2];Ott M 等人利用自然语言理解的方法获取评论内容的语言特征,然后基于分类学习的方法,对评论进行分类^[3]。二是基于评论人的行为模式或特征,比如打分(rating)的偏离度、多次评论、评论内容的相近度等,使用评分函数刻画评论的作弊程度,经过综合计算后发现高可疑虚假评论人^[5]。三是发现团体协作作弊的评论人,比如 Mukherjee A 等利用关联规则挖掘找到评论人同时出现的频繁项集,然后利用群体作弊行为模式去考察每一个频繁集,从而确定群体作弊行为^[6]。四是其他检测方法^[7,8],例如 Jindal N 等利用在评论数据集中发现的非预期规则去发现异常评论模式,从而确定作弊嫌疑人^[8]。

Wang G 等人提出了基于评论图(review graph)的网络产品虚假评论人的检测方法^[7]。该评论图通过评论、评论人和店铺(store)之间的相互影响关系建立一个互相制约的评论图结构。该图中定义了 3 种衡量作弊程度的度量指标:评论

到稿日期:2013-12-03 返修日期:2014-02-28 本文受国家自然科学基金(61373159)资助。

王 琢(1969—),男,硕士,副教授,主要研究方向为数据挖掘、信息检索,E-mail:zhuowang69@gmail.com;李 准(1987—),男,硕士生,主要研究方向为数据挖掘、信息检索;徐 野(1976—),教授,博士,主要研究方向为复杂网络建模与信息处理技术;宋 凯(1964—),教授,博士生导师,主要研究方向为图像信息处理、智能系统。

的可信度、评论人的忠实度和店铺的可靠度,其中每个指标可以通过自身或其他指标计算得出。然后利用节点增强方法(node reinforcement method)^[7]对各个指标进行迭代计算,直到其稳定为止。根据稳定后的指标确定虚假评论人的排名。该算法的时间复杂度为 $O(k(N_r + N_v + N_s))$,其中 N_r 、 N_v 和 N_s 分别为评论人、评论和店铺个数, k 为迭代次数。由于 N_s 远远小于 N_r 和 N_v ,因此计算店铺信誉度的时间基本可以忽略不计。然而,现实中网上评论的产品很少通过店铺进行交易,而是直接在电子商务网站(如亚马逊、京东商城等)上直接销售。此时,产品的数量 N_p 远远超过店铺的数量 N_s ,所以当以产品代替评论图中的店铺时,算法的计算量将急剧增加,严重影响算法的可行性。本文研究基于评论、评论人及产品构成的评论图,逐步淘汰真实的评论人及其评论,从而更快收敛到虚假评论人。另外,通过使用产品而非店铺,我们可以引入更多关于评论作弊模式的特征来刻画评论和评论人的虚假性,从而进一步提高基于评论图的虚假评论人检测精度。

2 基于评论、评论人和产品的评论图

2.1 评论图的结构

在文献[7]所提出的评论图基础上,我们重新定义了基于评论、评论人和产品的评论图。该评论图是由评论、评论人和产品作为节点的有向图。每个评论人节点具有指向其所写评论节点的边,每个评论节点具有指向其所评论产品节点的边,产品节点通过评论节点与评论人节点建立联系。其中每个节点可以附加一系列特征数据,比如评论节点具有评分等级(rating)、与其所评产品的其他评论间的差异度、其他评论人对该评论的投票数(votes)及该评论有帮助的投票数(helpful)等,产品节点具有该产品的平均打分值(average rating),评论人节点具有该评论人进行多次评论的产品数,以及在多次评论的产品中该评论人的评论数量等。评论图的结构如图1所示,我们为图中各类型节点定义了其信誉度的度量指标。

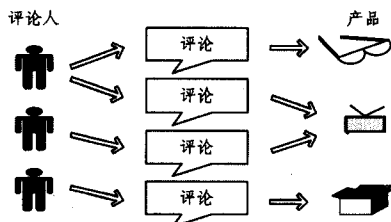


图1 评论图节点的相互关系

定义1(评论的忠实度, Honesty of review) 对于评论 v , 其忠实度表示该评论的忠实程度得分, 记为 $H(v)$ 。为便于迭代计算, 限定 $H(v)$ 的取值范围为 $[-1, 1]$ 。

定义2(评论人的可信度, Trustiness of reviewer) 对于评论人 r , 其可信度为该评论人的总体可信任程度得分, 记为 $T(r)$ 。限定 $T(r)$ 的取值范围为 $[-1, 1]$ 。

定义3(产品的可靠度, Reliability of product) 对于产品 p , 其可靠度表示所有评论人对其质量的认可程度得分, 记为 $R(p)$ 。其取值范围限定为 $[-1, 1]$ 。

各节点之间具有互相制约的复杂关系: 一个评论人的可信度取决于其所写评论的忠实度, 以及该评论人所写的异常评论的数量; 一个产品的可靠度取决于对其评论的所有可靠

评论人的打分; 一个评论的忠实度取决于该评论与其周围评论(指该评论所评论产品的其他评论)的一致度、该评论的帮助度和该评论所评论产品的可靠性。由“节点增强方法”可知, 在给定合理的初始值, 经过多次迭代后, 各节点的信誉度必将收敛到稳定值, 系统达到最终平衡。最后可以认定得分较低的评论人为虚假评论撰写者。

2.2 评分函数

本文假定评论系统使用5级打分制(1—5分)。文献[7]认为各个分值分布是均匀的, 并直接根据分值之差表示打分的差异度。我们观察到实际数据集中打分的分布是极不均匀的, 图2是实验中采用的两个数据集各个评分的分布直方图。我们发现, 1—3分评论比例很小, 一共只占到评论集的20%左右, 4、5分的比例明显偏多, 因此各个分值之差不能反映评分的差异度。为了更准确地描述打分的相近程度, 我们将评分分为两组, 规定分值不超过3的为低分组, 分值是4、5的为高分组, 这样更有利于打分的公正比较。由于评论内容的分析需要额外增加计算量, 而且评论内容往往与打分呈正相关^[7], 因此这里不考虑评论内容存在的欺诈信息。

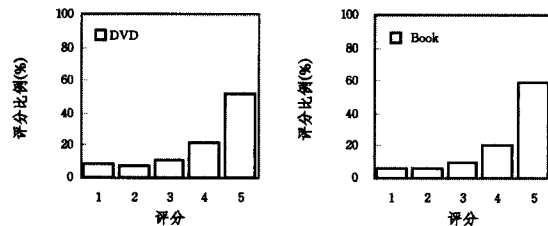


图2 DVD和Book数据集中各评分比例

2.2.1 评论人的可信度计算

评论人的可信度取决于其所有评论的忠实度之和。令评论人 r 的所有评论的忠实度之和为:

$$H_r = \sum_{i=1}^{n_r} H(a_i^r) \quad (1)$$

其中, n_r 为 r 所写评论的数量, a_i^r 为 r 的第 i 个评论, $H(v)$ 表示评论 v 的忠实度(稍后定义)。为了使 $T(r)$ 落在区间 $[-1, 1]$, 而且避免评论人的可信度过度受到评论人的所有评论的忠实度之和的影响, 文献[7]采用 logistic 曲线对 $T(r)$ 进行平滑, 得到:

$$T(r) = \frac{2}{1 + e^{-H_r}} - 1 \quad (2)$$

$T(r)$ 越大, 说明该评论人越可信。进一步地, 若 $T(r)$ 为正, 说明该评论人可信, 否则说明该评论人不可信。式(1)、式(2)说明评论人所写评论越真实, 该评论人越可信。

然而我们发现式(2)存在以下问题: 当评论人发表的所有评论的忠实度之和大于 3.66 时, 其可信度将大于 0.95, 从而该评论人被判定为可信评论人。例如, 如果评论人发表了 52 条忠实度为 1 的评论, 同时发表了 48 条忠实度为 -1 的评论, 则该评论人将被视为可信评论人。显然式(2)难以发现这类虚假评论人。故我们引入了更多揭示评论人作弊行为模式的指标对可信度进行矫正。

假定某顾客在买到一个产品后, 觉得该产品很差, 并给出很低的评分, 那么该顾客再次购买该产品的可能性将会很低。如果该评论人再次甚至多次对该产品进行低分差评, 那么该评论人将是极度可疑的。另外, 像 DVD 和图书之类的商品,

一般不需要多次购买,所以多次对同一种商品反复打分也属于可疑的作弊行为。

为了对上述情形建模,将评论人对同一产品多次评论作为影响评论人可信度的一个因素。令评论人 r 对某产品多次打分的产个数为 $N_L(r)$, r 的总评论产品数为 $N_P(r)$, 则定义评论人多次打分的度量指标 $Dup(r)$ 为:

$$Dup(r) = \frac{N_L(r)}{N_P(r)} \quad (3)$$

于是定义新的评论人可信度指标为:

$$T'(r) = T(r) - k_0 Dup(r) \quad (4)$$

其中, k_0 为权重系数,用于强调 $Dup(r)$ 的重要性。实验中我们取 $k_0 = 1$ 。如果 $T'(r) < -1$, 则 $T'(r)$ 取 -1 。通过对 $T(r)$ 的矫正,可以得到更为精确的评论人可信度,进而可以更加有效地检测虚假评论人。

2.2.2 评论的忠实度计算

评论的忠实度取决于多个方面。虚假评论者受利益驱使趋向于做出与大众相反的评价。一方面,如果一个评论与其周围的评论(对同一产品的临近时间窗口内发表的评论)差别较大,常常是虚假评论的信号。另一方面,如果一个评论的打分与产品的可靠度(即大众对产品的认识)偏差过大,则有理由怀疑该评论的真实性。设评论 v 所评论的产品为 p , 则在 v 评论的时间窗口 Δt 内, p 的所有评论集合 U 按评论的打分为 v 的相似集 S 和非相似集 US 。文献[7]定义 v 的相似集 S 和非相似集 US 分别为:

$$\begin{aligned} S &= \{i \mid |\gamma - \gamma_i| \leq 1, i \in U\} \\ US &= U - S \end{aligned} \quad (5)$$

其中, γ 为评论 v 的评分, γ_i 表示评论集合 U 中评论 i 的评分。然而,我们发现 $\gamma = 3$ 时,其相似分数为 $2-4$, 不相似集为 1 和 5 。显然将 1 分和 5 分视为同一类评论不合理。所以,我们不是按照 v 与周围评论打分差的绝对值来确定相似集,而是将打分为高分和低分两大类,根据评论 v 所属的类别确定其相似集,将相似集和非相似集重新定义为:

$$\begin{aligned} S' &= \{i \mid i \text{ 与 } v \text{ 属于同一分组}, i \in U\} \\ US' &= U - S' \end{aligned} \quad (6)$$

令

$$A(v, \Delta t) = \sum_{i \in S'} T(K(i)) - \sum_{j \in US'} T(K(j)) \quad (7)$$

其中, $K(i)$ 表示评论 i 所属的评论人。显然 A 代表了评价 v 与周围评价的一致度。文献[7]定义评论 v 的忠实度为:

$$H(v) = |R(p)| \left(\frac{2}{1 + e^{-A(v, \Delta t)}} - 1 \right) \quad (8)$$

其中, $R(p)$ 为 v 所评论的产品 p 的可靠度,用于加强评论的忠实度,因为如果产品可靠度的绝对值过小,则评论的可靠性很难把握。

为了更加准确地描述评论的虚假性,我们引入了评论打分 γ 与产品的平均打分 $\mu(p)$ 的差异度 $D(v)$, 以及每条评论的帮助度系数 $Help(v)$ 。差异度 $D(v)$ 定义为:

$$D(v) = \frac{|\gamma - \mu(p)|}{\max(\gamma)} \quad (9)$$

其中, $\max(\gamma)$ 为评分系统的最大评分(1-5级打分制中为5)。显然, $D(v)$ 越大,说明评论的忠实度越差。帮助度系数 $Help(v)$ 定义为:

$$Help(v) = \frac{helpful(v)}{votes(v)} \quad (10)$$

特别地,如果该评论尚未有读者投票,我们规定其取值为0.5。最终定义评论 v 的忠实度为:

$$H'(v) = H(v) - k_1 D(v) + k_2 Help(v) \quad (11)$$

其中, k_1 为 $D(v)$ 的权重,用于强调 $D(v)$ 的重要性,实验中我们取 $k_1 = 0.5$ 。 k_2 为评论帮助度的权重,取 $k_2 = 0.1$ 。若 $H'(v) > 1$, 则 $H'(v)$ 取 1 ; 若 $H'(v) < -1$, 则取 -1 。

2.2.3 产品的可靠度计算

产品的可靠性取决于可信评论人对产品做出的评价,而不取决于不可靠评论人的评价。令

$$B(p) = \frac{\sum_{v \in U_p, T(K(v)) > 0} T(K(v))(\gamma - \mu)}{\sum_{v \in U_p, T(K(v)) > 0} T(K(v))} \quad (12)$$

其中, U_p 代表产品 p 的所有评论集合, γ 为评论 v 的打分值, μ 为打分系统的中值,如在5级分制中可取3。类似文献[7]用 logistic 曲线对其标准化,得到产品的可靠度为:

$$R(p) = \frac{2}{1 + e^{-B(p)}} - 1 \quad (13)$$

为进一步刻画产品的可靠度,我们又考虑了产品的平均打分因素,最终得到:

$$R'(p) = R(p) + k_3 \frac{\mu(p)}{\max(\gamma)} \quad (14)$$

其中, k_3 为权重,用于强调产品平均打分因素的重要性,实验中我们取 $k_3 = 0.1$ 。若 $R'(p) > 1$, 则取 1 。显然,产品的可靠度依赖于该产品的评论人的可信度及其每条评论的打分值。

2.3 迭代算法

文献[7]提出了一个基于评论、评论人和店铺的评论图的虚假评论人迭代计算框架算法。在考虑基于产品而非店铺的评论图时,由于产品可靠度的计算相当耗时,因此在涉及大量产品的评论系统中,该算法不具备可行性。为了加快算法的速度,我们提出了基于评论人及其评论逐步淘汰的方法,其在保持检测准确度的情况下,极大地提高了算法收敛速度。原算法没有给出迭代次数的确定方法,因此迭代次数依赖于经验值。我们根据各次迭代后评论人可信度的平均残差平方和 ARSS(Average Residual Sum of Squares) 决定算法的终止条件,从而更准确地确定各节点的稳态值,其定义为:

$$ARSS^{(i)} = \frac{1}{n} \sum_{r \in R^{(i)}} [T(r)^{(i)} - T(r)^{(i-1)}]^2 \quad (15)$$

其中, $R^{(i)}$ 表示第 i 次迭代完成时未被淘汰的评论人集合, $T(r)^{(i)}$ 表示第 i 次迭代完成时评论人 r 的可信度。我们提出的 ICE 算法如算法 1 所示。

算法 1 Iterative Computation with Elimination(ICE)

输入: 产品集 P , 评论人集 R , 评论集 V , 时间窗口 Δt , 要寻找的虚假评论人数目上限 N , 淘汰率 ρ , 残差阈值 δ , 迭代最大次数 M
输出: 不超过 N 个虚假评论人(按可信度升序排列)

//初始化

1. 将产品的可靠度、评论人的可信度初始化为 1 ;
2. 标记所有评论人为未淘汰;

//迭代过程

3. while 残差 $ARSS > \delta$ and 迭代次数 $< M$ do
4. 根据式(11)计算未被淘汰的评论人所写评论的忠实度 $H(v)$;
5. 根据式(4)计算每个未被淘汰的评论人的可信度 $T(r)$;
6. 根据式(14)计算每个产品的可靠度 $R(p)$;
7. 计算未被淘汰的评论人可信度残差 $ARSS$;

8. 以 $T(r)$ 值降序顺序, 从未被淘汰评论人中淘汰比例为 ρ 的评论人, 并将其 $T(r)$ 值置 1;

9. end

10. 按 $T(r)$ 升序返回前 N 个评论人及其可信度值。

算法 1 中, 在淘汰一定比例的评论人的过程中需要对未被淘汰评论人的 $T(r)$ 值排序。为了加快排序过程, 我们使用了寻找第 k 小元素位置的算法^[9], 该算法可以在 $O(n)$ 内找到第 k 小元素对应的值 T_k , 然后我们在所有未被淘汰的评论人中依次淘汰 k 个 $T(r) > T_k$ 的评论人。该优化可以有效提升算法速度。

由于算法 1 逐步淘汰评论人, 每轮迭代时 $H(v)$ 和 $T(r)$ 的计算量将逐步减少, 因此有望快速进入收敛状态。算法总的迭代次数取决于具体的数据集, 这里我们给出迭代次数的上限值估计。如果不考虑收敛, 设评论人由 N_r 个淘汰为只剩下 N 个所需要的次数为 k , 则有:

$$N_r(1-\rho)^k = N$$

于是

$$k = \frac{\ln(N/N_r)}{\ln(1-\rho)} \quad (16)$$

例如取 $N/N_r = 0.01, \rho = 0.05$, 则 $k = 89$ 。实际上, 因被淘汰的评论人的可信度已经稳定, 且其所写评论不再更新忠实度, 故实际中算法的迭代次数远小于 k 。

3 实验结果

3.1 数据集和评测指标

我们使用了 2006 年从 Amazon.com 爬取的产品评论数据集 (<http://snap.stanford.edu/data/amazon-meta.html>) 作为实验数据集, 并从中提取了 DVD 和 Book 两个数据集。每个数据集中过滤掉评论数少于 6 条的评论人及相关评论后, 所含各项数据的个数如表 1 所列。

表 1 数据集的记录个数统计

	评论人数	产品数	评论数
DVD	19308	16439	455097
Book	86076	205987	2726775

其中评论所包含的属性有: 评论编号、所评论产品编号、评论人编号、评分值、评论时间、投票数、认为有帮助的投票数。评论人包含的属性有: 评论人编号、多次评论的产品数量、评论产品总数。产品包含的属性有: 产品编号、产品的平均打分。由于在两个数据集上的实验结果基本相似, 这里仅以 Book 数据集为例进行说明。

对于给定 N , 算法 1 将返回 N 个虚假评论人的结果集。为了比较不同方法返回结果的一致性, 我们引入了不同结果集的重合率和相似度的定义。

定义 4(重合率, overlap) 在两个均包含 N 个评论人的结果集 A 和 B 中, 其重合率 $O(A, B)$ 定义为:

$$O(A, B) = \frac{|A \cap B|}{N} \quad (17)$$

定义 5(相似度, similarity) 在两个均包含 N 个评论人的结果集 A 和 B 中, $a \in A, loc_A(a)$ 表示评论人 a 在结果集 A 中的位置(由 1 开始编号), 其中 A 中评论人 a 相对于 B 的距离 dis_a 为:

$$dis_a = \begin{cases} |loc_A(a) - loc_B(a)|, & a \in B \\ N, & a \notin B \end{cases}$$

则结果集 A, B 的相似度定义为:

$$s(A, B) = 1 - \frac{\sum_{a \in A} dis_a}{N^2} \quad (18)$$

3.2 算法效率与一致性

为了评估在 ICE 算法中逐步淘汰评论人及其评论的有效性, 我们将 ICE 算法与不淘汰评论人的算法(IC 算法)的效率进行了比较。实验中, 我们取最大迭代周期 $round = 100$, 时间窗口 $\Delta t = 30$, 平均残差平方和的阈值 $\delta = 10^{-7}$ 。图 3 是对于不同的淘汰剩余率 er (即 $1-\rho$) 迭代收敛到 δ 所需迭代次数的比较。图中 $er=1$ 代表 IC 算法所需要的迭代次数。可以看出, ICE 算法能明显加快收敛速度, er 越小, 收敛速度越快; 当 er 低于 0.94 时, 收敛次数趋于平缓。

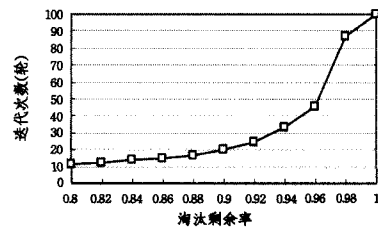


图 3 不同的淘汰剩余率与迭代次数的关系

图 4 是收敛到指定的 δ 所用时间随淘汰剩余率 er 变化的结果。可以发现, 即便是很小的淘汰率也可能极大地提高算法效率, 如当 $er=0.98$ 时, ICE 算法所用时间不足 IC 算法所需时间的 1/3; 当 er 低于 0.96 时, 时间变化趋于平稳。

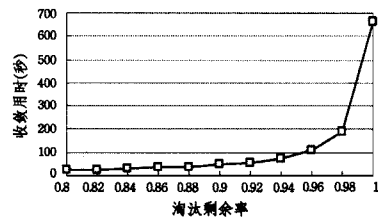


图 4 不同的淘汰剩余率与收敛时间的关系

为了比较 ICE 算法与 IC 算法返回结果的一致性, 我们使用了不同的淘汰剩余率, 对两种算法返回结果的重合率和相似度进行了比较。图 5 为不同的淘汰剩余率所对应的重合率和相似度, 其中 $er=1$ 对应于 IC 算法。所有运行结果中返回的虚假评论人人数 N 为 2000。可以发现, 重合率、相似度与淘汰剩余率基本呈线性变化。当 $er > 0.92$ 时, ICE 算法与 IC 算法结果已高度一致。由上述结果可以发现, 当 $er = 0.94$ 时, 无论是算法效率还是算法一致性都可以得到很理想的效果。

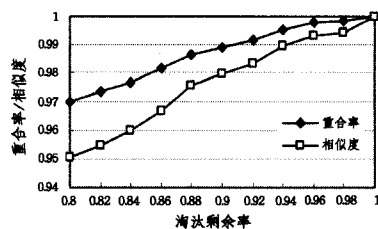


图 5 不同的淘汰剩余率与重合率和相似度的关系($N=2000$)

图 6 是不同的结果集大小 N 对于 ICE 算法和 IC 算法结果集间重合率和相似度的比较结果(其中 $er=0.94$)。可以看出, 对于较大 N 值, 两算法具有较高的重合率和相似度; 当 N 较小时, 重合率依然较高, 然而相似度差异较大, 但因 N 值较

小,故不影响虚假评论人的检测。

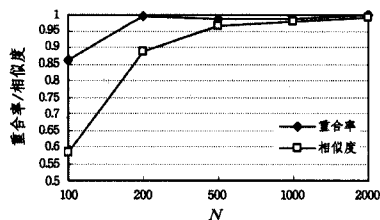


图6 不同的N值与重合率和相似度的关系($er=0.94$)

综合上述实验结果,ICE算法与IC算法相比不仅具有较高的一致性,而且可以极大地提高算法速度。

3.3 矫正指标对算法准确度的影响

为了评估对节点信誉度加入的矫正指标的有效性,我们比较了ICE算法与未加入矫正指标时的算法(即采用式(2)、式(8)、式(13)计算节点信誉度,以下称IC⁻算法)的重合率和相似度。图7为不同N值下两种算法的比较结果,其中ICE算法中取 $er=0.94$ 。可以发现其重合率和相似度整体偏低,说明两种方法结果并不一致。

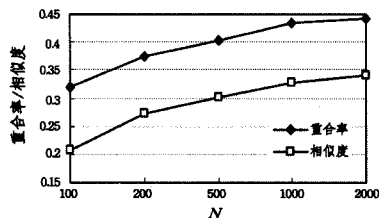


图7 ICE算法与IC⁻算法结果比较

为了验证ICE算法引进的矫正指标对虚假评论人检测的有效性,我们采用了与文献[7]相同的虚假评论人验证方法,即人工评估的方法,对两种算法返回的虚假评论人集合进行人工检测,然后比较人工检测结果与两种算法检测结果的一致性。我们安排3名具有丰富网购经验的硕士生独立地对算法给出的候选人进行作弊认定。

为了减小评估人主观性对人工检测过程的影响,我们制定了一些评估指标作为人工检测过程中的参考。为了方便人工评判、减小工作量,我们参照文献[5]中的做法开发了一个评论评估软件,它可以迅速浏览评论人的统计信息,比如打分模式等汇总信息。并且评估人还可以通过评估软件在amazon.com中查看被检测虚假评论候选人的个人信息及其在线评论内容,为进一步识别虚假评论人提供更多依据。

为了进一步减少人工评估工作量,分别从两算法结果中各取前100名虚假评论候选人,然后过滤掉两个结果集中都出现的33名候选人,将剩余的134人作为虚假评论候选人进行人工作弊认定。为了保证人工评估真实性、准确性,将剩余的134名虚假评论候选人进行随机排序,使人工评估人不知道每个候选人是由哪种算法产生的。

人工评估结果显示,在134名虚假评论候选人中,被3名评估人都判定为虚假评论人的共45人,其中属于IC⁻算法的有13人,属于ICE算法的有32人,这表明ICE算法相比于IC⁻算法具有更高的准确度。

表2、表3列出了3名评估人所认定的作弊评论人在两

种算法结果集中所占的人数。例如在ICE算法给出的67名虚假评论人中,被评估人1认定为虚假评论人的有39人;同时被评估人1和评估人2认定的有35人;同时被评估人1和评估人3认定的有33人。对比表3可以看出,被人工评估人认定的虚假评论人中大约2/3来自ICE算法的结果,说明ICE算法能够更准确地发现虚假评论人。为了验证3个评估人的一致性,我们计算出3个评论者之间的Kappa系数^[10]为61.2%,该值表明3个评估人评判结果具有高度的一致性^[11]。

表2 ICE算法人工评价结果

	评估人1	评估人2	评估人3
评估人1	39	35	33
评估人2		45	38
评估人3			40

表3 IC⁻算法人工评价结果

	评估人1	评估人2	评估人3
评估人1	21	15	15
评估人2		24	18
评估人3			24

3.4 案例分析

为了进一步确认ICE算法发现的虚假评论人的准确性,我们从ICE算法发现(但未被IC⁻算法发现)的虚假评论人中取两名被3名评估人同时认定的虚假评论人进行具体分析。

评论人 Shelli Treely¹⁾:该评论人在一个月内3个不同的日期共发表评论90条,其中1分的评论有57条,4分5条,5分28条。该用户在同一天期的评论中对一部分产品多次打1分,对另一些产品多次打5分或4分,该行为明显有悖于正常的购买者。通过到amazon.com中查看该用户的个人资料,90条评论中仍能在amazon.com中查询到的只剩4条,其中一条评论打分为1分,但评论内容却对产品大加赞誉。该用户总的有帮助投票数只有1票,且自2000年5月8日以来再未发表任何评论。

评论人 Vaughan Smith²⁾:该评论人在同一天内共发表评论42条,且42条评论均为1分。在amazon.com的个人资料信息中仍剩余3条评论,3条评论不仅评分均为1分,而且评论内容完全重复。所有评论均发表于2004年4月27日,自此后该评论人未发表任何评论。

结束语 本文提出了基于评论、评论人和产品的评论图,用于检测网络产品评论系统中的虚假评论人。提出了新的更为有效的节点信誉度评分准则,以提高算法识别虚假评论人的精确度。利用ICE算法在每轮迭代中淘汰一些可信度高的评论人及其评论,大大提高了算法的收敛速度,并且显著缩短了算法的执行时间。amazon.com 2006数据集的实验表明,相比原算法ICE算法不仅速度更快,而且通过对算法结果的人工评估发现,它可以更准确地检测虚假评论人。

参考文献

[1] 孙升芸,田莹. 产品垃圾评论检测研究综述[J]. 计算机科学, 2011,38(10A):198-201

(下转第305页)

¹⁾ <http://www.amazon.com/gp/pdp/profile/A3EJAWYC9CL6N9>

²⁾ <http://www.amazon.com/gp/pdp/profile/A2KELIXFXS3KIG>

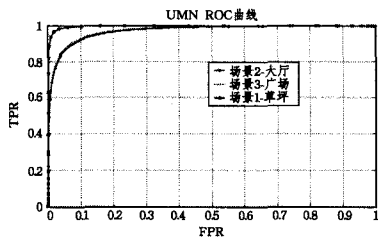


图7 UMN 3个场景异常事件检测 ROC 曲线

结束语 线性动态系统模型描述提取的视频时空块,稀疏编码进行异常事件检测。然而,线性动态系统属于非欧氏空间模型,无法直接应用传统的稀疏编码。输入视频时空小块的每帧图像可以由训练纹理的图像线性重构,重构系统与输入系统是等价的,它们之间可以通过非奇异线性变换矩阵进行转化。因此,算法的优化学习任务即是学习编码 α 和相似性变换矩阵 P 。由于线性变换空间是无限的,使得学习 P 无法实现。本文压缩线性变换空间,在正交子空间内寻找 P ,将训练事件集合分为若干子集合,在每个子集中将用于重构次数最多的若干样本作为子集的字典,子集字典的集合构成算法的字典。实验结果表明,本文所提出的方法具有较好的检测结果。由于优化过程得到局部最优解,参数初值的选择至关重要。寻找到最佳的初值使识别率更高,或者寻找到获得全局最优解的优化方法,是本文以后的研究方向。

参 考 文 献

[1] Shi Ying-huan, Gao Yang, Wang Rui-li. Real-Time Abnormal Event Detection in Complicated Scenes[C]//International Conference on Pattern Recognition. 2010

[2] Dollar P, Rabaud V, Cottrell G, et al. Behavior recognition via sparse spatio-temporal features[C]//VS-PETS. October 2005

[3] Soatto S, Doretto G, Wu Y N. Dynamic textures [J]. IJCV, 2003, 52(2):91-109

[4] Ravichandran A, Chaudhry R, Vidal R. View-invariant dynamic texture recognition using a bag of dynamical systems[C]//CVPR. 2009:1651-1657

[5] Ravichandran A, Chaudhry R, Vidal R. Categorizing Dynamic

Textures using a Bag of Dynamical Systems[J]. IEEE Transaction on Pattern Analysis and Machine Intelligence, 2013, 35(2): 342-353

[6] Coviello E, Mumtaz A, Chan A B, et al. Lanckriet. Growing a Bag of Systems Tree for Fast and Accurate Classification[C]//IEEE CVPR. 2012

[7] Wright J, Yang A Y, Ganesh A, et al. Robust face recognition via sparse representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2):210-227

[8] Ghanem B, Ahuja N. Sparse Coding of Linear Dynamical Systems with an Application to Dynamic Texture Recognition[C]//Proc. of the International Conference on Pattern Recognition (ICPR). 2010:987-990

[9] Cong Yang, Yuan Jun-song, Liu Ji. Sparse reconstruction cost for abnormal event detection[C]//CVPR. 2011:3449-3456

[10] Zhao Bin, Li Fei-fei, Xing E P. Online Detection of Unusual Events in Videos via Dynamic Sparse Coding[C]//CVPR. 2011: 3313-3320

[11] Lee H, Battle A, Raina R, et al. Efficient sparse coding algorithms[C]//NIPS. 2006

[12] Jimenez N D, Afsari B, Vidal R. Fast Jacobi-type algorithms for computing distances between linear dynamical systems[C]//IECC. 2013

[13] Mahadevan V, Li W X, Vasconcelos N. UCSD dataset [OL]. [2012-01- 11]. <http://www.svcl.ucsd.edu/projects/anomaly/>

[14] UMN dataset [OL]. [2012-01- 11]. <http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi>

[15] Adam A, Rivlin E, Shimshoni I, et al. Robust real-time unusual event detection using multiple fixed-location monitors[J]. TPA-MI, 2008, 30(3):555-560

[16] Mahadevan V, Li W, Bhalodia V, et al. Anomaly detection in crowded scenes[C]//CVPR. 2010

[17] Wu S, Moore B, Shah M. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes[C]//CVPR. 2010

[18] Mehran M S R, Oyama A. Abnormal crowd behavior detection using social force model[C]//CVPR. 2009

(上接第 299 页)

[2] Jindal N, Liu B. Opinion spam and analysis[C]//Proceedings of the International Conference on Web Search and Web Data Mining. Palo Alto, California, USA; ACM, 2008:219-230

[3] Ott M, Choi Y, Cardie C, et al. Finding deceptive opinion spam by any stretch of imagination[C]//Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics. Portland, Oregon; 2011: 309-319

[4] 李霄, 丁晨春. 垃圾商品评论信息的识别研究[J]. 现代图书情报技术, 2013, 229:63-68

[5] Lim E P, Nguyen V A, Jindal N, et al. Detecting product review spammers using rating behaviors [C]//Proceedings of the 19th ACM International Conference on Information and Knowledge Management. Toronto, ON, Canada; ACM, 2010:930-948

[6] Mukherjee A, Liu B, Wang J, et al. Spotting fake reviewer groups in consumer reviews[C]//Proceedings of 21st Interna-

tional Conference on World Wide Web. New York, NY, USA; ACM, 2012:191-200

[7] Wang G, Xie S, Liu B, et al. Review graph based online store review spammer detection[C]//Proceedings of the 11th International Conference on Data Mining. Vancouver, BC, Canada; IEEE, 2011:1242-1247

[8] Jindal N, Liu B, Lim E P. Finding unusual review patterns using unexpected rules [C]// Proceedings of the 19th International Conference on Information and Knowledge Management. Toronto, ON, Canada; ACM, 2010:1549-1552

[9] 《编程之美》小组. 编程之美[M]. 北京:电子工业出版社, 2008: 141

[10] 栾建安, 苏炳华. 多类别多评估人的 Kappa 分析[J]. 中国卫生统计, 1995, 12(6):20-22

[11] Landis J R, Koch G G. The measurement of observer agreement for categorical data[J]. biometrics, 1977, 33(1): 159-174