



在线中文商品评论可信度研究*

孟美任 丁晟春

(南京理工大学信息管理系 南京 210094)

【摘要】对在线中文商品评论中可信度较低的评论信息进行过滤,为消费者提供对制定购买决策有帮助的评论。在深入分析在线中文商品评论特点的基础上,结合相关研究成果,通过问卷调查进行可信度影响因素的实证分析。根据实证结果,选取内容完整性、情感平衡性、评论时效性以及发布者身份明确性4类特征,采用CRFs模型进行评论可信度4级分类,并进行特征组合实验,得到最佳特征组合。实验效果显著,分类模型正确率均在75%以上。该研究成果可以用于改善现有的“人工效用评价”方式,为在线评论的优化过滤提供一种新的方法与思路。

【关键词】在线商品评论 可信度 CRFs模型 影响因素 效用评价

【分类号】G353.1

Research on the Credibility of Online Chinese Product Reviews

Meng Meiren Ding Shengchun

(Department of Information Management, Nanjing University of Science & Technology, Nanjing 210094, China)

【Abstract】This paper aims at filtering the lower credible online Chinese product reviews to offer valuable reviews for consumers' purchase decision. Based on the deep analysis of the online Chinese product reviews' characteristics, also with some related works, the authors make an empirical analysis on the credibility factors through questionnaires. According to the results of the empirical analysis, the authors select content integrity, emotional balance, review timeliness and clarity of the identity of the publisher as four features, use CRFs as reviews credibility's classification model, and conduct feature combination experiments to get the best feature combination. The experiments achieve significant results, and the correct rates of the classification model are all above 75%. The research results of this paper can improve the existing artificial effectiveness evaluation method, thus offering new methods and thoughts for optimized filtering of the online reviews.

【Keywords】Online product reviews Credibility CRFs model Affecting factor Effectiveness evaluation

1 引言

在线商品评论可以作为消费者购买决策的重要参考依据、企业最直接的用户反馈,还可以指导各级工商管理部门进行商品监管。但是目前存在大量恶意操纵在线商品评论从中牟利的行为。本文定位于解决在线中文商品评论可信度这一关键问题,根据问卷调查实证分析结果,选取可信度特征,使用机器学习的方法,得到效果好的可信度分类模型。该研究成果可以有效地改善目前“人工效用评价”方式,更好地为消费者筛选出对制定购买决策有帮助的商品评论。

收稿日期:2013-06-19

收修改稿日期:2013-07-28

* 本文系国家自然科学基金项目“基于文本语义挖掘的商品评论信息可信度分析研究”(项目编号:71103085)的研究成果之一。

2 相关工作

近年来,在线商品评论已逐渐成为学术界的研究热点。研究内容集中于评论挖掘工作,主要包括评价对象抽取、情感极性分析等。但是对于在线商品评论可信度的研究相对较少,相关研究主要围绕商品评论可信度影响因素、商品评论质量特征以及商品评论质量评价方法三个方面。

2.1 在线商品评论可信度影响因素相关研究

在线商品评论可信度影响因素研究中,研究者主要从信源(评论发布平台)、信息传播者(评论发布者)、信息(评论内容、结构等)三个方面研究哪些因素会影响在线商品评论的可信度,部分具有代表性的研究成果如表 1 所示:

表 1 在线商品评论可信度影响因素部分代表性研究成果

作者	时间(年)	影响因素及结论
陈江涛等 ^[1]	2012	评论的详尽度
郝媛媛等 ^[2]	2010	情感倾向、正负情感混杂度 主客观表达混杂度、较长的平均 各句长度
Otterbacher ^[3]	2009	评论内容的质量
李念武等 ^[4]	2009	同嗜性、传播者专业性、网站信心 度、信任倾向
Liu 等 ^[5]	2008	评论者的经验、评论者写作风格 以及评论时效性
Ghose 等 ^[6]	2007	观点句的主观性

通过分析发现,该领域的研究者还未能通过影响因素分析结果进一步筛选出可信度较高的评论。但研究成果可以为可信度特征选取提供重要理论支持。

2.2 在线商品评论质量特征相关研究

在线商品评论质量特征的研究成果也可以作为可信度特征选取的重要依据之一。现有研究大致把质量特征分为文本特征以及元数据特征,如表 2 所示:

表 2 关于在线商品评论质量特征的研究成果

在线商品评论质量特征	细分特征	特征描述
文本特征 ^[7]	语法特征 ^[9]	专有名词、数字、情态动词、 感叹词等
	语义特征 ^[10]	情感极性、评价对象等
	文体特征 ^[11]	词汇和结构属性
元数据特征 ^[8]	商品评论发布者与读者 之间的互动属性 ^[9]	评论“有用”票数、评论发 表时间、商品评分等

目前,学术界对于什么特征能够准确表征商品评论的质量,还没有达成统一的观点,需要设计科学有效

的方法对质量特征选择的可靠性和准确性进行验证。

2.3 商品评论质量评价方法研究

近年来,商品评论质量评价问题已经从理论研究上升到方法的探讨,并逐渐趋于实践。Hu 等^[12]证明了目前存在部分出版商、作者以及卖家有意操纵评论信息。Hu 等^[13]在管理图书在线评论的研究中,开发出一个全权操作代理器,执行结果表明大量欺诈性在线评论对商品的销售造成了严重的影响。Liu^[14]首次提出了观点欺诈问题,对欺诈者的目标、行为、隐蔽技巧进行分析,并从评论、评论者、服务器三方面提出欺诈检测思路。Jindal 等^[15]将评论分为不真实的评论、无关的评论以及非评论三种类型,使用回归模型识别后两种垃圾评论,使用 Shingle 算法进行第一种评论的识别,发现实验结果存在遗漏,召回率较低。Wu 等^[16]利用同一商品下用户发表的唯一评论所占比例以及时间聚集程度来发现操纵评论的可疑行为。李霄等^[17]从评论、评论者和被评论的商品三个方面选择 11 个特征,使用 SVM 对垃圾评论进行了二分类。

目前,商品评论存在造假现象已经得到证明,学者试图解决这一问题,但仍存在以下几方面问题:

(1) 目前研究几乎都是基于英文语料。由于语法、语言结构等诸多方面的差异,其研究成果不能直接用于中文语料,本文方法致力于解决中文商品评论的可信度分类问题;

(2) 与国外相比,国内在该领域的研究还主要集中于对影响因素以及特征的探讨。少量的垃圾评论识别研究,虽然剔除了一部分垃圾评论,但是非垃圾评论数量仍然庞大,存在信息过载的现象,本文对在线评论进行可信度 4 分类,为消费者提供更加精华的评论信息;

(3) 在特征选取上现有研究并没有考虑商品评论独有的特点,本文在特征选取时以商品评论的自身特点作为重要依据。

3 研究方法 with 方案设计

条件随机场(Conditional Random Fields, CRFs)^[18]是由 Lafferty 等提出的,其模型思想的主要来源是最大熵模型以及马尔可夫模型。CRFs 模型是典型的判别式模型,在给定标记的观察序列的条件下,计算整个标记序列的联合概率。设 $X = (X_1, X_2, \dots, X_N)$ 为待标注的观察数据序列上的随机变量, $Y = (Y_1, Y_2, \dots,$

YN)为相应的标注序列上的随机变量,链式 CRFs 模型定义状态序列的联合条件概率为:

$$p(y|x) = \frac{1}{z(x)} \exp(\sum_i \lambda_k f_k(y_{i-1}, y_i, x, i))$$

$$z(x) = \sum_y \exp(\sum_i \lambda_k f_k(y_{i-1}, y_i, x, i))$$

其中, $z(x)$ 为归一化处理参数, $f_k(y_{i-1}, y_i, x, i)$ 是基于给定输入的特征函数, 参数 λ_k 为特征函数 f_k 的权重, 可从训练数据中估计。 λ_k 表示事件发生的可能性。由此, 对于给定观察序列找到相应最可能的标记序列的任务就转化为对于给定的输入观察序列 X , 找到一个输出标记序列 Y^* 使得 $P(Y|X, M)$ 最大。

$$Y^* = \operatorname{argmax} P(Y|X, M)$$

由于 CRFs 模型不对单个标记归一化, 避免了标记偏置问题。CRFs 模型已被广泛应用于序列标记、数据分割、组块分析等自然语言处理任务中。同时, 本课题组之前采用 CRFs 模型进行了中文微博观点句抽取、商品评论评价对象抽取、评论情感极性及其强度计算等研究, 均获得了较好的准确率和召回率。

鉴于此, 本文选取 CRFs 模型作为实验分类模型, 设计研究思路如图 1 所示:

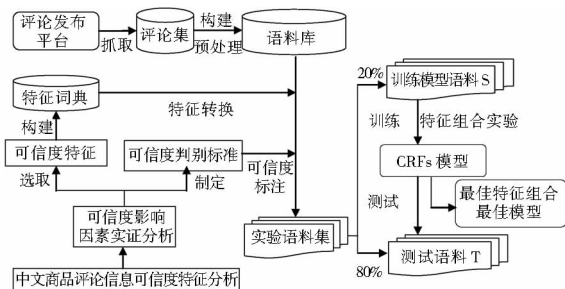


图1 商品评论可信度研究实验流程

(1) 编写自动爬取程序, 获取在线评论语料;

(2) 结合中文商品评论可信度已有研究成果, 深入分析在线中文商品评论可信度特征, 选取可信度影响因素, 设计调查问卷进行实证分析;

(3) 根据实证分析结果, 一方面, 选取实验特征, 构建特征库, 依照特征库对语料进行特征集自动标注; 另一方面, 制定可信度 4 级判别标准, 提出本文可信度标注方法, 对语料进行可信度标注, 得到最终的实验语料;

(4) 随机从语料集中选取 20% 语料作为训练集 (S), 剩余作为测试集 (T)。使用外部工具包 CRF++-0.53 对训练集 S 进行训练, 生成模型文件, 并且进

行特征组合和多种模板实验, 调整特征阈值, 选取最佳特征模板, 生成最佳的商品评论可信度判别模型。

4 中文商品评论可信度特征分析实验

依据 Liu^[14]对虚假评论发布者动机的分类, 以及本课题组的实地调研结果, 从造假评论发布者动机的角度分析商品评论可信度影响因素, 并进行实证分析, 根据实验结果, 选取可信度实验特征。

4.1 造假评论发布者动机分析

本文将造假评论发布者动机分为以下 4 种:

(1) 推销——为了提高自身销量, 对目标商品进行夸大的、不切合实际的正面评价。该类评论中的评价词几乎均为正面(褒义)词汇(A1), 并且往往会在短时间内(A2)对同一商品发布较多的评论(A3)。此类评论大多出现于“刚刚开店”(A4)、“店铺搞活动”(A5)、“调整价格”(A6)时期。

(2) 诋毁——撰写大量错误、负面的评论来诋毁竞争对手的商品。该类评论中的评价词几乎均为负面(贬义)词汇(B1)。商家往往会诋毁比较有实力的竞争对手, 所以诋毁评论的情感倾向往往与该店铺的信誉存在差异(B2), 并且评论中涉及与此商品同类的其他品牌产品信息(B3)。

以上两类评论反映出可信度较高的商品评论中的评价词往往既包含正面词汇也包含负面词汇(B4), 情感极性平衡。

(3) 干扰——评论中还存在大量广告(电话、QQ号)、链接(网址)式评论, 此类评论中往往出现大量英文字母(C1)以及数字(C2); 或者由于购买环节上的问题单纯发泄不满情绪的评论, 此类评论往往与商品本身毫无关系, 并不包含任何个人使用感受(C3)、产品属性(C4)等信息; 并且可信度高的商品评论的情感极性与其打分基本相符, 而为了误导简单检测算法的低可信度商品评论, 其对商品打分与评论的情感倾向性存在差异(C5)。

(4) 敷衍——觉得麻烦或为了赢得积分的敷衍评论, 也称之为无意义评论。此类评论大多使用几个单字(D1)进行简单评价, 如“好!”、“挺!”、“不错!”等。此类评论虽然可能不存在造假但是并没有对商品进行较长的详细评价(D2), 所以对消费者并无帮助。

另外, 从评论者的角度, 可以将评论者级别作为一

项特征,认为匿名(E1)发布的评论往往没有信誉较高的评论者(E2)发布的评论可信。而“人工效用”评价模块的“有用”(F1)或“无用”(F2)评论也会对可信度造成影响。在内容时效性方面,认为发布时间越新(G1)的商品评论越能代表商品的当前情况,其可信度较高。

4.2 商品评论可信度影响因素实证分析

笔者对南京理工大学、武汉理工大学等高校发放问卷 200 份,回收 172 份,其中有效问卷 151 份,受访者年龄段集中在 18 岁-24 岁之间,是网上购物的主要群体。问卷设计为对 4.1 节所述因素进行详细描述的 5 级李克特量表,1 表示非常不信任,5 表示非常信任。对各影响因素变量进行描述性数据分析发现,均值 3.00 为商品评论信息可信度正面影响与负面影响的临界值。笔者认为因素标准差越大,说明该因素更偏离均值,更能显著表征商品评论的可信程度,排名前 11 的影响因素如表 3 所示:

表 3 各因素标准差排名

排名	影响因素	标准差
1	包含大量数字(C2)	1.149
2	包含大量字母(C1)	1.139
3	评价词褒贬俱有(B4)	1.129
4	发布时间较新(G1)	1.099
5	整体认为“有用”(F1)	1.098
6	评价词全为正面(A1)	1.037
7	涉及产品属性(C4)	1.033
8	涉及体验感受(C3)	1.028
9	极短时间间隔内发布(A2)	1.019
10	评价词全为负面(B1)	1.002
11	长度趋长(D2)	0.998

根据影响因素、评论本身及其他相关属性、可信度理论等多个方面,对商品评论可信度提出相关假设。通过信度和效度检验,发现变量间有公共因子存在,适合进行因子分析。总量表的 α 系数为 0.777,量表信度颇佳。经验证,假设均得到支持。

4.3 特征选取

以 4.2 节中因素的标准差以及成立的假设为依据,本文将影响因素分为内容完整性、情感平衡性、评论时效性以及发布者身份明确性 4 类。

(1) 内容完整性特征(Fa:F1-F4)

评论中涉及个人体验感受 and 对应产品属性(F1)时,对其可信度的影响是正向的;当内容中包含大量数字(F2)或字母(F3)信息时,对其可信度的影响是负面的。就其文本长度(F4)而言,在合理范围内,长度较

长的评论能获得较高的可信度。

(2) 情感平衡性特征(Fb:F5,F6)

评论的情感倾向极性越单一,其可信度越低。即正面评价词所占比例(该条评论正面评价词数量/该条评论所有评价词数量)(F5)、负面评价词所占比例(该条评论负面评价词数量/该条评论所有评价词数量)(F6)越高,其可信度越低。情感极性越平衡,其获得的可信度越高。

(3) 评论时效性特征(Fc:F7,F8)

发布时间(F7)越新的评论越能获得较高的可信度,同时间段内发布次数(F8)过高则会对评论的可信度评判产生负面影响。

(4) 发布者身份明确性特征(Fd:F9)

评论发布者的身份也对可信度构成影响,评论的发布者评级(F9)越高,身份越明确,该评论的可信度越高。

5 中文商品评论可信度分类实验

5.1 语料集预处理

笔者依据 Alexa 排名,选择淘宝网、新浪、太平洋电脑网以及中关村 4 个平台获取 5 094 条相机评论信息。由两位志愿者按照以下可信度判定标准对语料进行 4 级可信度标注:

(1) 可信度 I 级评论:评论中无任何广告等无用信息,包含详尽的产品属性评价、评论内容情感平衡,可以很好地帮助消费者进行商品选择;

(2) 可信度 II 级评论:评论中无任何广告等无用信息,提到 1-2 个产品属性,情感极性较为单一,对消费者进行商品选择较为有用;

(3) 可信度 III 级评论:评论中无任何广告等无用信息,但内容简短,如只有“不错!”、“好”等词语;或者评论发布时间较为集中,真实存在可疑,对消费者进行商品选择用处不大;

(4) 可信度 IV 级评论:评论中含有网址联系方式等的广告贴,或者只是单纯的情感发泄,无法对购买决策起到帮助。

参考文献[19]对中文观点句标注的方法,笔者提出了针对商品评论信息可信度的 4 级标注方法。要求两名志愿者分别用 I、II、III、IV 标注评论可信度,同时用数字 1,2,3,4 表示对标注的确认程度,1 为不确定,4

为非常确定。将确认程度 1 和 2,3 和 4 结合,两名标注者 J1、J2 对 4 级可信度进行标注的结果如表 4 所示:

表 4 J1 和 J2 的标注结果

J2 \ J1	I _{1,2}	I _{3,4}	II _{1,2}	II _{3,4}	III _{1,2}	III _{3,4}	IV _{1,2}	IV _{3,4}	Total
I _{1,2}	48	188	45	41	7	17	5	0	350
I _{3,4}	234	1 868	36	44	6	14	0	0	2 201
II _{1,2}	23	20	23	153	9	11	12	68	318
II _{3,4}	56	42	103	384	12	12	26	3	638
III _{1,2}	1	11	45	53	15	98	12	44	277
III _{3,4}	4	3	23	9	134	597	34	24	828
IV _{1,2}	0	0	12	35	67	117	6	18	255
IV _{3,4}	0	0	3	0	23	6	12	185	229
Total	366	2 130	290	717	273	870	107	341	5 094

从标注结果中去掉理解存在偏差的语料,首先去掉两人没有标为同一级的语料(如 J1 标为 I 级,J2 标为 II 级),再去掉所有至少 1 次被标注为不确定性为 1 或 2 的语料。最终得到判断比较一致的 I 级评论 1 868 条、II 级评论 384 条、III 级评论 597 条、IV 级评论 185 条。

5.2 特征标注规则

实验中特征标注规则如表 5 所示:

表 5 CRFs 模型下特征标记及描述

特征类别	特征标记	特征描述(单条评论)
评论中出现商品属性(F1)	SH/SL/N	包含 2 个以上 F1/包含 1-2 个 F1/不包含 F1
数字长度(字符数)(F2)	NUM/N	F2 < 8/F2 > 8
字母长度(字符数)(F3)	LT/N	F3 < 12/F3 > 12
评论内容长度(汉字数)(F4)	L/N	10 < F4 < 65 / F4 < 10 或 F4 > 10
正面评价词所占比例(F5)	OP/N	15% < F5 < 85% / F5 > 85% 或 F5 < 15%
负面评价词所占比例(F6)	NG/N	15% < F6 < 85% / F6 > 85% 或 F6 < 15%
评论发布时间(F7)	X/N	F7 为近 1 个月/F7 为 1 个月之前
同时间段评论发布次数(F8)	Y/N	F8 > 1/F8 = 1
评论者等级(F9)	VIP/O/N	高级用户 / 普通用户 / 匿名用户

(注:各特征的阈值是由统计分析得到。)

5.3 实验结果分析

使用外部工具包 CRF++ -0.53 进行实验。虽然前文选取的特征都可以表征商品评论的可信度,但由于评论中各特征存在不确定性,相互搭配存在随机性。所以本文对内容完整性(Fa)、情感平衡性(Fb)、评论时效性(Fc)以及发布者身份明确性(Fd)4 类特征进行

13 种特征组合实验,分别得到对 I、II、III、IV 识别效果较好的特征组合,结果如表 6 所示:

表 6 特征组合实验结果

实验结果	评价指标		
	Precision	Recall	F-measure
可信度 I 级识别最好结果 (Fa + Fb + Fd)	82.62%	67.50%	74.30%
可信度 II 级识别最好结果 (Fa + Fb)	76.81%	69.13%	72.77%
可信度 III 级识别最好结果 (Fa)	77.45%	63.56%	69.82%
可信度 IV 级识别最好结果 (Fa + Fd)	85.78%	72.56%	78.62%
所有特征组合平均值	77.67%	65.29%	72.77%

从表 6 可以发现,本文的识别效果显著,正确率均在 75% 以上。对实验结果分析如下:

(1) 特征组合 Fa + Fb + Fd 识别可信度 I 级商品评论的效果最好。该级别评论具有内容详尽以及情感平衡的主要特点,并且级别较高的评论者更愿意根据自身购买经验进行详细的评论,以更好地帮助其他消费者进行购买决策的制定。

(2) 对于识别可信度 II 级的商品评论,Fa + Fb 为最佳特征组合。一般的购买者通常会对商品的 1-2 个属性进行评论,与消费者等级并无太显著关系。另外,大部分出于推销和诋毁目的的造假者发布的商品评论都属于可信度 II 级,为避免被发现造假,内容往往与可信度 I 级评论较为相似。虽然本文的模型对于该级别评论的识别效果为 4 级中最低的,但是也达到了 76.81%,向消费者推荐时,可以在时间区间相同的条件下,排序在可信度 I 级评论之后。

(3) 单独使用 Fa 就可以很好地识别可信度 III 级商品评论。在可信度标注时就发现了大量只用单个字或几个字进行评论的现象。而 Fa 特征恰恰表征了评论的详尽程度,所以本文模型对于该级别评论的识别效果为 4 级中最高的,同时保持了较高的召回率。另外,发现本文选取特征 F8 存在缺陷,由于大量造假信息是由计算机批量发布,内容差别不大,甚至完全相同。所以应对时间段再次缩短,并同时比对短时间发布信息的内容差别性。

(4) 对于识别可信度 IV 级评论,Fa + Fd 的特征组合效果最好。因为 Fa 特征包括数字、字母的比例,这恰恰是广告贴的明显标志。但是模型漏掉了一些情感发泄等与主题毫无关系的商品评论,导致召回率偏低。主要原因是部分商品评论隐去了评价对象,只含有评价词,导致模型混淆了单纯的情感发泄与 II 级可信度

情感单一的商品评论,可以考虑通过加入本体来解决隐性评价对象识别问题。

6 结 语

本文选取内容完整性、情感平衡性、评论时效性以及发布者身份明确性 4 类特征,采用 CRFs 模型进行可信度 4 级分类。对 4 类特征进行特征组合实验,分别得到了对可信度 I、II、III、IV 级商品评论识别效果最好的特征组合。

另外,依据本文的研究成果可以改善目前的“人工效用评价”方式。首先,较新的评论往往没有足够时间获得更高的“有用”票数,所以“有用”票数不高的评论并不意味着对消费者毫无帮助。依据本文的研究方法,剔除掉可信度 III、IV 级商品评论,再按照可信度 I、II 级对商品评论进行排序,以此来替代根据“有用”投票数向用户推荐评论的方式;其次,“人工效用评价”方式认为用户打分为 4 星、5 星并且“有用”投票数越高的评论为“最有帮助的好评”。本文提出的情感平衡性特征可以更好地反映商品评论的情感极性,与可信度分类结果相结合,能够更为准确地向用户推荐“最有帮助的好评”以及“最有帮助的差评”。

参考文献:

- [1] 陈江涛,张金隆,张亚军. 在线商品评论有用性影响因素研究:基于文本语义视角[J]. 图书情报工作, 2012, 56(10):119-123. (Chen Jiangtao, Zhang Jinlong, Zhang Yajun. Impact Factors of Online Customer Reviews Usefulness: A Text Semantics Approach [J]. *Library and Information Service*, 2012, 56(10): 119-123.)
- [2] 郝媛媛,叶强,李一军. 基于影评数据的评论有用性影响因素研究[J]. 管理科学学报, 2010, 13(8):78-88,96. (Hao Yuanyuan, Ye Qiang, Li Yijun. Research on Online Impact Factors of Customer Reviews Usefulness Based on Movie Reviews Data [J]. *Journal of Management Sciences in China*, 2010, 13(8):78-88,96.)
- [3] Otterbacher J. "Helpfulness" in Online Communities: A Measure of Message Quality[C]. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2009:955-964.
- [4] 李念武,岳蓉. 网络口碑可信度及其对购买行为之影响的实证研究[J]. 图书情报工作, 2009, 53(22):133-137. (Li Nianwu, Yue Rong. An Empirical Study on Credibility of Online Word - of - Mouth and Its Effects on Consumers' Purchase Behavior [J]. *Library and Information Service*, 2009, 53(22):133-137.)
- [5] Liu Y, Huang X J, An A, et al. Modeling and Predicting the Helpfulness of Online Reviews[C]. In: *Proceedings of the 8th IEEE International Conference on Data Mining*. Washington: IEEE Computer Society, 2008:443-452.
- [6] Ghose A, Ipeirotis P G. Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews[C]. In: *Proceedings of the 9th International Conference on Electronic Commerce*. New York, NY, USA: ACM, 2007:303-310.
- [7] Kim S M, Pantel P, Chklovski T, et al. Automatically Assessing Review Helpfulness[C]. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*. Stroudsburg, PA, USA: Association for Computational Linguistics, 2006:423-430.
- [8] Weimer M, Gurevych I. Predicting the Perceived Quality of Web Forum Posts[C]. In: *Proceedings of Recent Advances in Natural Language Processing*, Bororets, Bulgaria. 2007:643-648.
- [9] Zhang Z. Weighing Stars: Aggregating Online Product Reviews for Intelligent E-commerce Applications [J]. *IEEE Intelligent Systems*, 2008, 23(5):42-49.
- [10] Abbasi A, Chen H, Salem A. Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums [J]. *ACM Transactions on Information Systems*, 2008, 26(3):1-34.
- [11] Ahmed A, Hsinchun C. Applying Authorship Analysis to Extremist - Group Web Forum Messages [J]. *IEEE Intelligent Systems*, 2005, 20(5):67-75.
- [12] Hu N, Liu L, Sambamurthy V. Fraud Detection in Online Consumer Reviews[J]. *Decision Support Systems*, 2011, 50(3):614-626.
- [13] Hu N, Bose I, Gao Y J, et al. Manipulation in Digital Word of Mouth: A Reality Check for Book Reviews [J]. *Decision Support Systems*, 2011, 50(3):627-635.
- [14] Liu B. Web Data Mining [M]. Berlin: Springer, 2009:316-317.
- [15] Jindal N, Liu B. Review Spam Detection[C]. In: *Proceedings of the 16th International Conference on World Wide Web*, Banff, Alberta, Canada. New York, NY, USA: ACM, 2007:1189-1190.
- [16] Wu G, Greene D, Smyth B, et al. Distortion as a Validation Criterion in the Identification of Suspicious Reviews[C]. In: *Proceedings of the 1st Workshop on Social Media Analytics*. Washington, DC, USA: ACM, 2010:10-13.
- [17] 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(1):63-68. (Li Xiao, Ding Shengchun. Research on Review Spam Recognition [J]. *New Technology of Li-*

brary and Information Service, 2013(1):63-68.)

[18] Lafferty J D, McCallum A, Pereira F C N. Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data[C]. In: *Proceedings of the 18th International Conference on Machine Learning*. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2001:282-289.

[19] 叶强, 张紫琼, 罗振雄. 面向互联网评论情感分析的中文主观

性自动判别方法研究[J]. *信息系统学报*, 2007, 1(1):79-91. (Ye Qiang, Zhang Ziqiong, Luo Zhenxiang. Automatically Measuring Subjectivity of Chinese Sentences for Sentiment Analysis to Reviews of the Internet[J]. *China Journal of Information Systems*, 2007, 1(1):79-91.)

(作者 E-mail: memory8.8nmr@gmail.com)

NISO 计划为补充计量制定标准和推荐做法

美国国家信息标准协会(NISO)宣布了一项新的两阶段项目,研究并开发基于社区的补充计量(Altmetrics)标准或是推荐做法。学术评价是科研工作中很重要的一环,影响深远,包括哪个项目得到资助,谁应该晋升,哪个出版物更加优秀,等等。从20世纪60年代 Eugene Garfield 的首创性工作开始,科研评价一直是基于引用数据的。引用数据是一个很有用的指标,但是无法跟上时代发展的步伐,无法有效对在线读者行为、网络交互情况、社交媒体、在线内容管理等行为进行测量。受一些新的创新平台的影响,如 ImpactStory,业界开始了一项新的运动,即开发更为强大的补充标准——称之为补充计量学(Altmetrics)——以补充传统的引用指标。NISO 会先举办一些现场的或虚拟的会议以确定亟需制定补充计量标准和推荐做法的关键问题,然后组成一个工作组,制定一致的标准和推荐做法。整个项目由 Alfred P. Sloan 基金会资助,资助金额为 207 500 美金。

“引文分析无法对更新且更为普遍的文献产生影响的方式进行衡量,比如,通过像 Twitter, Facebook 或博客等社交网络工具产生影响。”NISO 项目部副主任 Nettie Lagace 解释,“而且,新形式的学术成果,如数据集、软件工具、算法或分子结构等,已经很常见,但是不容易,甚至无法通过传统的引用指标来衡量。这是补充计量领域最为关注的两个问题。”

“如果补充计量要走出现在的测试和概念验证阶段,学界必须开始将注意力集中在给出一套公认的定义、计算方法和数据共享实践上。”NISO 执行董事 Todd Carpenter 表示,“组织和科研人员如果想利用这些指标,就应该充分地理解它们,确保其一致性和意义统一性,而且应该有方法来审计其准确性。我们必须就一些基本问题达成共识,包括:应该测量哪些指标,衡量指标质量好坏的标准,这些指标应该在何种粒度上进行收集和分析,补充计量所测量的时间跨度应该是多少,社交媒体在补充计量中的角色,要交换这些数据需要什么样的技术基础设施,以及哪个补充计量指标是最有价值的。补充计量标准和最佳做法的制定将会有利于学界对补充计量的信任和采纳,有利于补充计量被更广泛的接受,进而确保补充计量能够准确地地在出版商之间和平台之间进行比较和交换。”

该项目的第一阶段将召集两组特邀的补充计量、传统出版、文献计量及教师评估等领域的专家,进行实地访谈,目标是确定补充计量的一些关键问题,以及亟需制定标准和推荐做法的几个方面,这将作为下一步两个虚拟会议的基础。其他社会各界人士的意见将通过一系列领域内会议和活动收集。总结报告将会给出 NISO 需要在哪些方面制定标准和推荐做法。项目第二阶段,工作组将会根据报告制定相关的标准和推荐做法。整个项目预计为期两年。有关会议和参与方法的通知将会在 NISO 官网的每月实事通讯中发布(<http://www.niso.org/publications/newsline/>)。

(编译自:http://www.niso.org/news/pr/view? item_key = 72efc1097d4caf7b7b5bd9c54a165818399ec86)

(本刊讯)