

垃圾评论检测

王雪鹏 刘康

中国科学院自动化研究所

1. 引言

随着社交媒体的快速发展，来自网络上的评论观点，越来越多地影响组织或个人进行购买决策制定、选举投票和市场产品设计等事宜。对于商业和个人，正面的评论观点常常意味着更高的利润和更好的口碑。有研究表明：在Yelp.com网站上的，增加额外的半颗星评分，能帮餐馆增加约19%的销量 (Anderson and Magruder, 2012)，平均评分增长一颗星，将带来5-9%的利润增长 (Luca, 2011)。基于对于利润和市场的追逐，在网络中也出现了越来越多的虚假评论或虚假观点 (opinion spam)。通过发布这种虚假的观点或评论，来达到推销或诋毁一些目标产品、服务、组织或个人的目的。这样的个人或组织称为垃圾评论发布者 (opinion spammer)，他们的活动称为垃圾评论发布 (opinion spamming)。

目前，垃圾评论已经十分普遍，例如，有调查研究表明在Yelp.com网站上有接近25%的评论是欺骗性的评论 (<http://www.bbc.com/news/technology-24299742>)。相比于商业评论，针对社交和政治事件的垃圾评论发布活动具有更大的危害性，它们可能会歪曲观点来将民众调动到法律和道德的对立面上。因此，为了使得社交媒体继续作为公开观点的可信源，而不是充斥着越来越多的欺骗、谎言与伪造观点，垃圾观点检测具有着极其重要的意义和研究价值。

2. 垃圾评论的类型

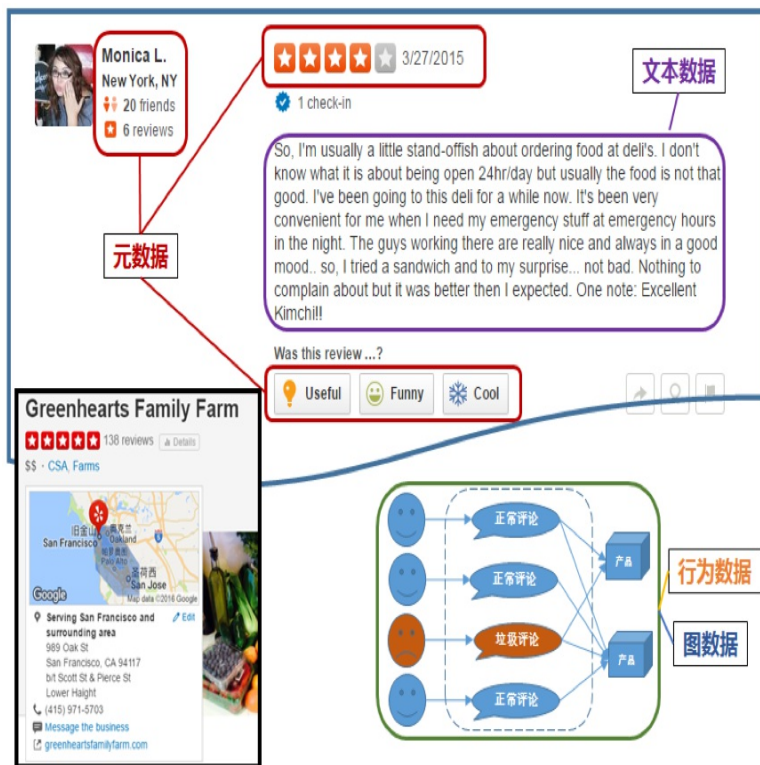
Jindal and Liu (2008) 经过分析指出，商业网站上目前主要有三种类型的垃圾评论：

1. **类型1（虚假评论）** 这些都是没有基于评论者针对真实产品或服务实际使用经历和体验的非真实评论。其中都蕴含着隐藏的动机。他们通常针对目标实体（产品或服务）发表不应得的正面评价，以促销实体；或用不公正的虚假负面评论来诋毁目标实体的声誉。
2. **类型2（仅关于品牌的评论）** 这些评论没有如期望地评论具体的产品或服务，而是对品牌或产品的制造商发表评论。虽然他们可能是真实的，但因为它们没有针对具体的产品，而且往往是有偏见的，因而被视为垃圾评论。例如，一个关于特定型号的HP打印机评论说：“我讨厌HP。我从来不买他们的任何产品。”
3. **类型3（非评论文本）** 这类垃圾评论往往都不是评论，包含有两种亚型：（1）广告；（2）不含观点的（例如，问题，答案和随机文本）不相关文本。严格地说，它们不被当做评论，因为它们没有提供用户对于产品和服务意见和评论。

目前，类型2和3的垃圾评论比较罕见，而且很容易用基于监督学习的统计模型或者给定模板加以检测 (Jindal and Liu 2008)。即使没有检测到它们，也不成大问题，因为人类读者可以在阅读过程中很容易发现它们。因此现有的大量研究工作也主要关注于类型1的垃圾评论检测研究，即**虚假评论检测**。

3. 可观察到的信息

垃圾评论活动本身具有很高的隐蔽性，垃圾评论者在评论网站上留下的搜索浏览记录、IP迁移变更速率、客户端登陆记录等等信息，对于检测垃圾评论十分有效，但这些信息对于评论网站本身也是极为保密且有价值的信息。虽然 dianping.com、yelp.com等网站开发了高效的垃圾评论过滤算法，但其算法细节与所用数据信息都是保密的。相比之下，对于学术界而言，所能获取的数据信息就比较有限了，主要就是普通用户能在网站页面上浏览到的信息。但学术界也作出了大量的研究贡献并得出了很多有价值的结论。图1就展示了学术界目前所能利用的数据信息（以yelp.com数据为例）。



** 图1. 可观察到的信息**

如图1所示，对于产品的评论页面而言包含有元数据和文本数据。- 元数据：主要指的是用户发表评论的基本信息，包括：用户名、用户所在城市、用户朋友数目、用户发表过的评论数、该条评论的评分、发布日期、所获得的有帮助反馈数等时间、地理、社交维度方面的数据，以及包括对于被评价实体（产品、宾馆或餐厅）页面包含的元数据有实体的平均分、实体所在地、实体所获得评论的数量等元数据；- 文本数据：该条评论的文本内容；- 行为数据：在整个评价系统中，用户、评论、被评价实体三者所也构成了一种网络图结构。在这个图中包含了用户评价的行为踪迹，这种用户的行为踪迹被视为关于用户评论的行为数据。

元数据、文本数据、行为数据都是可以用来挖掘有效特征的数据信息。

4. 数据集

巧妇难为无米之炊。由于垃圾评论活动本身的隐蔽性和难标注性，一个合理的数据集对于研究工作起到了很大的支撑作用。

- **Amazon数据集** 垃圾评论检测的问题由（Jindal & Liu,2008）第一次提出。他们收集了来自Amazon.com的数百万条评论数据，他们假设垃圾评论发布者为了节省精力而重复利用已有评论，进而将重复性的评论当做垃圾评论数据。但由于缺乏标注数据集，该工作只是简单地将重复性的评论当做垃圾评论数据来训练模型，因而会错过分析很多非重复性的伪装性很高的垃圾评论。
- **Resellerratings数据集** （Wang et al.2011）第一次提出了应用graph model来解决垃圾评论检测问题，也是第一次关于店铺评论检测的研究工作。该数据集来自www.resellerratings.com上的店铺评论数据，采用人工方式标注数据。（Mukherjee et al.2012）同样利用了www.resellerratings.com上的数据，并进行了垃圾评论用户组的研究工作。
- **Tripadvisor数据集** 由于缺乏人工标注的数据集，（Ott et al.2011）借助Amazon Mechanical Turk，运用众包的方法，即在AMT网站上发布有偿任务，雇佣他人写关于宾馆或餐馆的伪造评论，并提出了第一个垃圾评论检测的gold-standard数据集。对于非垃圾评论数据的收集，（Ott et al.2011）认为具有大量评论且平均分较高的宾馆或餐馆，对于采用负面评论诋毁拉低评分的垃圾评论攻击活动已无攻击价值，因为基于大量评论的平均评分很难被少量负面评价拉低。于是（Ott et al.2011）在tripadvisor网站上收集了这样的若干五分评论作为非垃圾评论。（Ott et al.2013）通过添加负面伪造评论，对（Ott et al.2011）的数据集进行了扩充。（Jiwei Li et al.2014）将（Ott et al.2011）的数据集扩展到了多领域，并加入了来自专家的伪造评论进行实验。但是，该数据集采用众包的方式收集到的垃圾评论并非来自于专业垃圾评论者，缺乏商业上的真实性与实战性。（Mukherjee et al.2013）运用（Ott et al.2011）得到的训练好的模型，在来自Yelp.com的真实商业性的数据集上测试，只得到了68%的准确率，并统计观察发现（Ott et al.2011）的数据集中真伪评论的词分布差异性，相比真实商业性数据要大很多。Mukherjee在文中表明，来自Amazon Mechanical Turk的评论撰写者，并不一定是商业网站中专业的或有经验的垃圾评论发布者，且受到的经济利益驱动也不大（AMT上发布的任务报酬及收集评论数都很少），因而他们写的评论更容易从文本角度被检测出来；同时没有保留用户的行为信息，也为（Ott et al.2011）的模型应用于真实任务造成了困难。
- **Yelp数据集** 这是第一个兼具商业性与ground truth的数据集。（Mukherjee et al.2013b）是第一个分析由商业网站过滤后的real-life fake 评论的工作。Yelp.com是一个商业性的评论收集网站，从2005年开始致力于可疑或伪造评论的过滤工作，其算法成熟，准确率高，获得了业界研究的认可（Weise,2011）。但由于是商业秘密性，Yelp.com并没有公布其算法细节，但公布了其对评论过滤的结果。作者在文中将在（Ott et al.2011）数据集上用监督学习方法训练好的模型（只利用评论的文本特征），在来自Yelp.com的真实商业性的数据集上测试，只得到了68%的准确率，远低

于 (Ott et al.2011) 90%的准确率。作者在Yelp.com数据集上, 综合利用文本特征和用户行为特征来检测垃圾评论, 获得了较好的实验效果。(Mukherjee et al.2013c) 做了类似的工作, 发现在Yelp.com网站上, 为了避免在评论中重复使用特殊词汇, 垃圾评论发布者都试着以相似频率使用在真实评论中出现的词汇。但同时垃圾评论发布者为了使评论更可信, 过度伪造了评论。基于Yelp.com过滤算法的业界认可度和多年开发的成熟性, 该数据集是目前在商业真实性与ground truth性之间, 做出较好平衡的一个数据集。

5. 垃圾评论检测方法

已有方法多把垃圾评论检测问题看成一个分类任务, 其研究的核心问题是如何从评论网站上可观测的信息中抽取有效的特征, 然后输入分类器当中, 构建有效地垃圾评论检测器。

根据可观测的信息类型以及特征表示方法, 已有方法可以分为: 1. 基于文本内容的垃圾评论检测方法; 2. 基于用户行文的垃圾评论检测方法; 3. 基于表示学习的垃圾评论检测方法。

5.1 基于文本内容的垃圾评论检测方法

基于文本内容的垃圾评论检测方法, 顾名思义就是仅从文本内容中抽取指示垃圾评论的特征。其中, Ott et al. (2011)将心理语言学线索特征应用到了垃圾评论检测上。Harris (2012)探索了写作风格方面的特征。Feng et al. (2012a)调研了文体学方面的文本特征。Li et al. (2013)将主体模型应用到了垃圾评论检测中。Li et al. (2014b)又进一步研究了垃圾评论与正常评论之间语言使用上的差异性。Li et al. (2014a)基于unigram与bigram特征进行了垃圾评论检测中单类学习的相关研究。Kim et al. (2015)提出了基于框架特征的深度语义分析。表1展示了一些可以适用于垃圾评论检测任务中的文本特征。

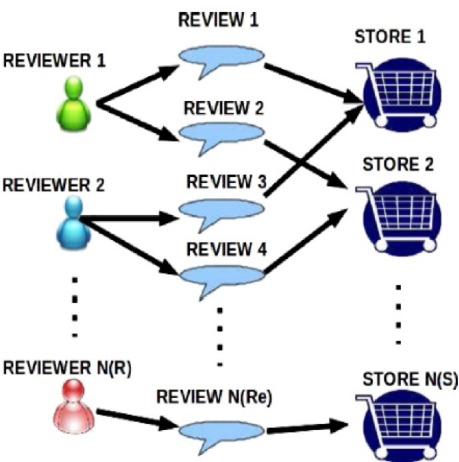
表1：文本特征

特征类别 | 特征 ---|--- 一般文本特征 | N-Grams: Unigrams, Bigrams, Trigrams 基于心里语言学线索 | Linguistic Inquiry, Word Count 评论体裁 | The frequency of each POS tag

5.2 基于用户行为的垃圾评论检测方法

Lim et al. (2010)研究了垃圾评论者打分为行为方面的特征。Jindal et al. (2010)分析了评论系统中可疑性非正常的评论模式。Feng et al. (2012b)分析了垃圾评论者的评论行为分布性。Xie et al. (2012)研究了不活跃用户时间维度上的非正常行为模式。Mukherjee et al. (2012)研究了群组性垃圾评论者的行为特征。Mukherjee et al. (2013a)提出了一种原则性方法来对评论者的垃圾攻击性建模。Fei et al. (2013)分析了垃圾评论者在评论爆炸中的共现性行为。Mukherjee et al. (2013c)证明了在垃圾评论检测中, 评论者的行为特征比文本特征更有效。Li et al. (2015)与KC and Mukherjee (2016)分析了垃圾评论活动中的时间和空间动态特征。

Wang et al. (2011)利用线上商铺评论分析了评论系统中的网络图特征, 并应用图模型来解决垃圾评论检测问题。作者提出了三个度量指标: 店铺的可靠度、评论用户的可信任程度、评论的诚实度。在此基础上考虑了店铺、评论发布者、评论之间的三角关系, 迭代计算三个指标来对三者分别计算排序, 而不是简单的只考虑单个评论者的行为特征。Akoglu et al. (2013)也在图特征方面做了类似的研究工作, 其跳出了以往了要么依赖于用户文本特征、要么依赖于用户行为特征来检测垃圾评论的思路, 其直接关注于评论者与产品之间的作用结构网络图特征, 提出了一种更为有效的方法: FRAUDEAGLE。图5 (Wang et al.2011) 展示了评论系统中基本的图结构, 从该图结构中, 我们可以利用图算法 (HITS、PageRank、RandomWalk等) 挖掘分析当前评论是否为一个虚假评论, 当前评论者是否为一个垃圾评论者。



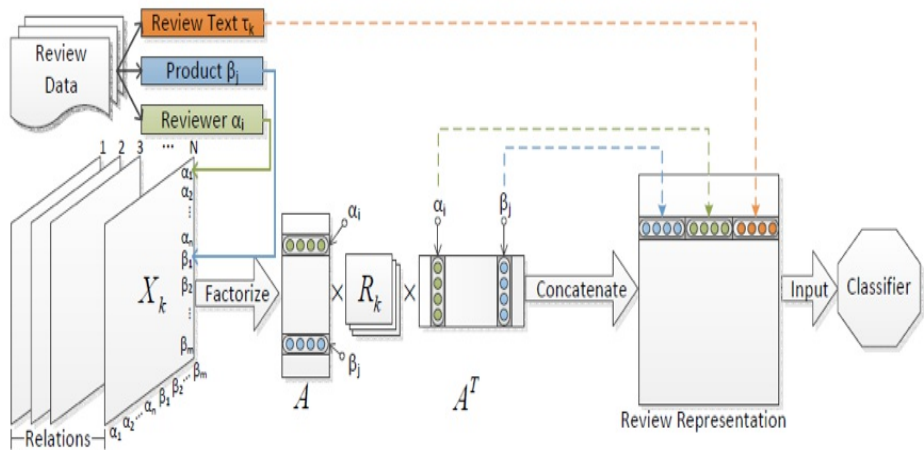
** 图2. 挖掘用户行为的图结构 (Wang et al. 2011)**

在此基础上, Mukherjee et al. (2013b)联合利用文本特征和行为特征来检测垃圾评论, 并利用Yelp.com数据集证明了, 在垃圾评论检测任务中, 特别是在真实的商业性数据集上, 他们发现: 行为特征比文本特征更有效。Rayana and Akoglu (2015)在Akoglu et al. (2013)工作的基础上, 第一次综合利用了从评论文本、评论者行为、评论系统图结构中挖掘出的线索特征来检测垃圾评论, 也得到了相同的结论。总体上来说, 行为特征比文本特征

更有效，联合利用文本特征、行为特征与图特征是目前最为有效的方法。

5.3 基于表示学习的垃圾评论检测方法

虽然已有工作在检测垃圾评论方面取得了很大的进步，但（Wang et al. 2016）发现其中依然存在着以下几点不足：- 已有工作依据人为做出的假设及专家知识来提取特征或对图结构进行分析，不仅费时费力，可靠性也有限。同时专家知识或先验知识也是有限的，在提取特征之前，做了大量的依据专家知识的统计。例如，Li et al. (2015)根据大众点评的数据统计分析得到了这样的先验知识：“垃圾评论行为是一种地理外包式的活动”、“一个用户的注册地也就是他平时所生活的地方”，所以真实用户更倾向于评论注册地附近的餐馆，而垃圾用户的餐馆领域的评论不受地理条件约束。当我们对Yelp数据集统计时，发现了相悖的结论，在餐馆领域，垃圾评论中72%来自于餐馆附近的地方，而真实评论中有64%来自于餐馆附近的地方，反而低于垃圾评论中比例。可见专家知识结论的可靠性并不高。同时，专家知识或先验知识也是有限性的，很难发现大量可疑线索。- 依赖于专家知识或先验知识的方法，在选取特征时也只人为地利用了系统中的部分信息。例如，传统的行为特征都是围绕着一个评论者个体进行统计的，未能考虑评论者在系统中与产品及其他评论者的交互信息；基于图结构提取信息表达评论的方法，虽然考虑了评论者、产品之间的交互信息，但是只考虑了同一产品评价页面下的评论者之间的交互信息，没有突破评论页面的限制。我们认为，在不同产品间，全局地比较地收集评论者与评论者、评论者与产品、产品与产品间的信息也具有很强的辅助作用。因此，针对上述已有方法不足，目前研究的关键问题主要集中于：我们是否能够直接从数据的层面，而不依赖于专家经验，自动学习得到用户的行为表示以及所评价的商品表示？



** 图3. 基于张量分解的用户和产品表示学习**

为了解决以上提出的以往工作中的方面不足，（Wang et al. 2016）提出了一种基于表示学习的垃圾评论检测方法。模型如图3所示，该方法不依赖于专家知识（或者说极大地减轻了对于专家知识的依赖性和受影响性），利用多关系全局信息进行联合学习，同时考虑了用户和产品信息，并且用隐含的方式表示评论，具有很好的鲁棒性。具体做法如下：

- 为了尽可能地摆脱对于专家知识的依赖，并尽可能全局性收集信息，捕捉用户行为特征，（Wang et al. 2016）在未作出任何垃圾嫌疑倾向性假设的前提下，定义了两类基础关系“两个实体（评论者或产品）的直接关联关系（一阶关系）”与“两个实体的属性间关系（二阶关系）”，在此基础上，他们分别从时间、空间、社交等维度记录了两个实体之间的关联信息，共衍生出11种具体的关系（如图4所示），并用一个张量记录了这11种关系下用户之间、产品之间以及用户和产品之间的关联。该方法突破了传统的评论者必须有评论交集的限制或者说页面内部局部限制，在每个关系中，将任意两两实体（评论者和评论者，评论者和产品，产品和产品）进行比较，在全局范围内来衡量一个评论者或产品在当前关系中的分布特点。

1	用户是否评价了某产品
2	用户给某产品的评分
3	用户A与用户B是否共同评价了某产品
4	用户A与用户B共同评价某产品的时间差
5	用户A与用户B共同评价某产品的评分差
6	用户A与用户B注册的时间差
7	用户A/产品M与用户B/产品N的平均分差
8	用户A与用户B的朋友数目差
9	用户A/产品M与用户B/产品N所在地是否相同
10	产品M与产品N拥有的共同评论者数量
11	产品M与产品N拥有的评论数量差

** 图4. 基于张量分解的用户和产品表示学习**

- 为进一步联合运用多关系信息，并用隐含的方式表示评论（评论者表示+产品表示），(Wang et al. 2016) 采用了基于上面提到的11种关系（非常基础且不易被垃圾评论者改变的关系）的张量分解方法（RESCAL），在多关系之间应用全局性的损失函数，来更充分地联合学习评论者与产品各自的信息表示。如(Nickel et al., 2011)所述，在全局性损失函数的学习中，“所有直接或间接的关系对于实体的表达都有影响”，也就是说最后联合学习到的实体隐含表示，融合了全局的关系信息。这些隐含表示不能被垃圾评论者所理解，有助于提升检测系统的鲁棒性。其中表示第种关系下用户和产品的行为数据，通过矩阵分解，我们可以得到用户和产品的表示矩阵。

□

- 为了能区分来自于同一用户的不同产品评论，我们将学习到的评论者与产品的表示（embedding）拼接起来，作为评论者发给该产品的评论的信息表示。以此得到融合全局性信息的、隐含式的、有产品区分度的评论表示。最后，我们将评论的表示输入到分类器中，完成垃圾评论的检测工作。

6. 实验和结论

我们在Yelp数据集上进行试验，对于上述三类方法进行比较，数据集详细情况如表2所示：

Domain | Hotel | Restaurant ---|--- fake | 802 | 8368 no-fake | 4876 | 50149 #reviews | 5678 | 58517 #reviewers | 5124 | 35593

表2: Yelp评测数据集 实验结果如图5所示。

Method	C.D.	P	R	F1	A	P	R	F1	A
SPEAGLE+(80%)	50:50	75.7	83.0	79.1	81.0	80.5	83.2	81.8	82.5
	N.D.	26.5	56.0	36.0	80.4	50.1	70.5	58.6	82.0
Mukherjee_BF	50:50	82.4	85.2	83.7	83.8	82.8	88.5	85.6	83.3
	N.D.	41.4	84.6	55.6	82.4	48.2	87.9	62.3	78.6
Mukherjee_BF+Bigram	50:50	82.8	86.9	84.8	85.1	84.5	87.8	86.1	86.5
	N.D.	46.5	82.5	59.4	84.9	48.9	87.3	62.7	82.3
Ours_RE	50:50	83.3	88.1	85.6	85.5	85.4	90.2	87.7	87.4
	N.D.	47.1	83.5	60.2	85.0	56.9	90.1	69.8	85.8
Ours_RE+PE	50:50	83.6	89.0	86.2	85.7	86.0	90.7	88.3	88.0
	N.D.	47.5	84.1	60.7	85.3	57.4	89.9	70.1	86.1
Ours_RE+PE+Bigram	50:50	84.2	89.9	87.0	86.5	86.8	91.8	89.2	89.9
	N.D.	48.2	85.0	61.5	85.9	58.2	90.3	70.8	87.8

(a) Hotel

(b) Restaurant

** 图5. 实验结果**

其中“50: 50”表示训练数据中正负例的比例为50比50。而“N.D.”则表示我们不专门对于数据的分布进行调整。其中SPEAGLE是利用文本特征进行垃圾评论滤除，MukherjeeBF是利用用户行为特征进行这一任务。MukherjeeBF+Bigram是把前两类（文本特征和用户行为特征）相结合的结果。而OursRE是利用张量分解对于用户进行表示学习之后所得结果作为特征，然后进行分类的结果，即在表示矩阵中，将评论者所对应的向量抽取出来作为特征(Wang et al. 2016)；OursRE+PE实在OursRE的基础上加入对于产品的表示学习结果作为特征的结果，即把表示矩阵中所有的向量都作为特征；OursRE+PE+Bigram是在Ours_RE+PE的特征基础上结合文本特征（Bigram）的结果。

从这一结果中，我们可以看出：1）针对垃圾评论检测任务来说，用户行为比文本特征更加有效（MukherjeeBF vs. SPEAGLE）；2）通过表示学习，能够有效地学习到用户行为的特征表示，相对于传统的行为特征抽取，表示学习能够直接从用户的基础数据出发，学到有效的行为特征和所评论产品的特征（OursRE vs. MukherjeeBF）；3）将用户行为特征与文本特征相结合能够达到State-of-the-arts的性能（OursRE+PE+Bigram取得最好的结果）。

本文对于垃圾评论检测这一任务进行了简单综述，介绍了垃圾评论检测任务详情，包括可用的数据集、任务特点以及目前的研究分类和进展。特别介绍了基于表示学习的垃圾评论检测方法，其重点在于利用基于张量分解的表示学习方法，自动学习评论者以及所评论商品的特征表示，并将评论者、所评论商品以及所评论文本的特征相结合，在Yelp数据集上取得了最好的结果。

参考文献

Leman Akoglu, Rishi Chandy, and Christos Faloutsos. 2013. Opinion fraud detection in online reviews by network effects. ICWSM, 13:2–11.

Michael Anderson and Jeremy Magruder. 2012. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database*. The Economic Journal, 122(563):957–989.

Geli Fei, Arjun Mukherjee, Bing Liu, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013. Exploiting burstiness in reviews for review spammer detection. In ICWSM. Citeseer.

Song Feng, Ritwik Banerjee, and Yejin Choi. 2012a. Syntactic stylometry for deception detection. In *Proceedings of the 50th ACL: Short Papers-Volume 2*, pages 171–175. ACL.

Song Feng, Longfei Xing, Anupam Gogar, and Yejin Choi. 2012b. Distributional footprints of deceptive product reviews. In *ICWSM*. C Harris. 2012. Detecting deceptive opinion spam using human computation. In *Workshops at AAAI on Artificial Intelligence*.

Nitin Jindal and Bing Liu. 2008. Opinion spam and analysis. In *Proceedings of the First WSDM*, pages 219–230. ACM.

Nitin Jindal, Bing Liu, and Ee-Peng Lim. 2010. Finding unusual review patterns using unexpected rules. In *Proceedings of the 19th CIKM*, pages 1549–1552. ACM.

Santosh KC and Arjun Mukherjee. 2016. On the temporal dynamics of opinion spamming: Case studies on yelp. In *Proceedings of the 25th International Conference onWorldWideWeb*, pages 369–379. International World Wide Web Conferences Steering Committee.

Seongsoon Kim, Hyeokyeon Chang, Seongwoon Lee, Minhwan Yu, and Jaewoo Kang. 2015. Deep semantic frame-based deceptive opinion spam analysis. In *Proceedings of the 24th CIKM*, pages 1131–1140. ACM.

Fangtao Li, Minlie Huang, Yi Yang, and Xiaoyan Zhu. 2011. Learning to identify review spam. In *IJCAI Proceedings*, volume 22, page 2488.

Jiwei Li, Claire Cardie, and Sujian Li. 2013. Topicspam: a topic-model based approach for spam detection. In *ACL (2)*, pages 217–221. Huayi Li, Bing Liu, Arjun Mukherjee, and Jidong Shao. 2014a. Spotting fake reviews using positive-unlabeled learning. *Computación y Sistemas*, 18(3):467–475.

Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. 2014b. Towards a general rule for identifying deceptive opinion spam. *ACL*.

Huayi Li, Zhiyuan Chen, Arjun Mukherjee, Bing Liu, and Jidong Shao. 2015. Analyzing and detecting opinion spam on a large-scale dataset via temporal and spatial patterns. In *Ninth International AAAI Conference on Web and Social Media*.

Ee-Peng Lim, Viet-An Nguyen, Nitin Jindal, Bing Liu, and Hady Wirawan Lauw. 2010. Detecting product review spammers using rating behaviors. In *Proceedings of the 19th CIKM*, pages 939–948. ACM.

Bing Liu. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press.

Michael Luca. 2011. Reviews, reputation, and revenue: The case of yelp. com. *Com (September 16, 2011)*. Harvard Business School NOM Unit Working Paper, (12-016).

Arjun Mukherjee, Bing Liu, and Natalie Glance. 2012. Spotting fake reviewer groups in consumer reviews. In *Proceedings of the 21st WWW*, pages 191–200. ACM.

Arjun Mukherjee, Abhinav Kumar, Bing Liu, Junhui Wang, Meichun Hsu, Malu Castellanos, and Riddhiman Ghosh. 2013a. Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD*, pages 632–640. ACM.

Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie Glance. 2013b. Fake review detection: Classification and analysis of real and pseudo reviews. Technical report, Technical Report UIC-CS-2013-03, University of Illinois at Chicago.

Arjun Mukherjee, Vivek Venkataraman, Bing Liu, and Natalie S Glance. 2013c. What yelp fake review filter might be doing? In *ICWSM*.

Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. 2011. A three-way model for collective learning on multi-relational data. In *Proceedings of the 28th ICML*, pages 809–816.

Myle Ott, Yejin Choi, Claire Cardie, and Jeffrey T Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 49th ACL: Human Language Technologies-Volume 1*, pages 309–319. ACL.

Shebuti Rayana and Leman Akoglu. 2015. Collective opinion spam detection: Bridging review networks and metadata. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 985–994. ACM.

Guan Wang, Sihong Xie, Bing Liu, and Philip S Yu. 2011. Review graph based online store review spammer detection. In *Proceedings of the 11th ICDM*, pages 1242–1247. IEEE. Xuepeng Wang, Kang Liu, Shizhu He and Jun Zhao. Learning to Represent Review with Tensor Decomposition for Spam Detection. In *Proceedings of EMNLP 2016*.

Sihong Xie, Guan Wang, Shuyang Lin, and Philip S Yu. 2012. Review spam detection via temporal pattern discovery. In *Proceedings of the 18th KDD*, pages 823–831. ACM.