

中文社交媒体谣言统计语义分析

刘知远^{①②③*}, 张乐^{①②③}, 涂存超^{①②③}, 孙茂松^{①②③}

① 智能技术与系统国家重点实验室, 北京 100084

② 清华信息科学与技术国家实验室(筹), 北京 100084

③ 清华大学计算机系, 北京 100084

* 通信作者. E-mail: liuzy@tsinghua.edu.cn

收稿日期: 2015-09-27; 接受日期: 2015-10-20; 网络出版日期: 2015-12-09

国家自然科学基金(批准号: 61202140)、国家社会科学基金重大招标项目(批准号: 13&ZD190)和国家重点基础研究发展计划(973计划)(批准号: 2014CB340501)资助

摘要 本文以新浪微博为例, 针对收集的谣言信息进行了不同角度的分析统计. 在统计分析方面, 我们研究分析了谣言的影响力及其产生与消亡的特点. 在语义分析方面, 根据不同谣言的语义内容, 将微博谣言划分为5大类, 并且建立了基于微博谣言的文本分类器. 在时序分析方面, 根据转发趋势进行了谣言的划分, 分析了不同类别谣言的传播特点. 最后, 本文面向社交媒体提出了一套综合机器智能和群体智能的自动辟谣框架.

关键词 社交媒体 谣言 语义分析 时序分析 辟谣

1 引言

谣言是社会心理学等领域关注的重要社会现象^[1~4]. 虽然谣言并没有公认的确切定义, 但是大多数学说都认同谣言是在人和人之间传播的, 含有公众关心信息的一种特殊陈述, 而其真实性不能很快得到证明或是得不到证明^[5].

近年来, 随着在线社交媒体的迅速发展, 大量不可靠的信息得以快速和广泛地在人群中传播. 社交媒体上谣言泛滥可能导致人们难以从纷繁的信息中甄别得到可信的信息, 进而影响人们正常的生活秩序. 特别是在面临突发公共事件(自然灾害、事故灾难、公共卫生事件、社会安全事件、经济危机等)时, 广泛传播的谣言可能会具有极大的破坏性. 因此, 深入定量分析社交媒体中的谣言性质意义重大.

社交媒体中的谣言泛滥不仅发生在我国, 也成为 Twitter 等英语世界社交媒体的痼疾, 因此近年来引起国内外学术界和产业界的广泛关注, 涌现了大量针对社交媒体谣言的定量分析研究和谣言检测. 然而这些研究大都是针对英语世界 Twitter 等社交媒体^[6~11], 主要利用了社会网络结构, 评论等信息展开分析. 也有少数工作是面向中文社交媒体(如新浪微博)的^[12~14], 但要么是针对有限的案例进行分析(如郭美美事件^[12]), 要么仅用少数语言学和网络特征进行谣言预测^[13, 14], 无法了解社交媒体谣言的全貌.

本文以新浪微博为研究对象, 收集了大规模社交媒体谣言数据, 对中国社交媒体谣言进行了比较全面的定量统计语义分析. 新浪微博是中国最大的微博服务平台, 截至2012年12月底, 新浪微博注



图 1 (网络版彩图) 微博社区管理中心不实信息处理公示页面

Figure 1 (Color online) The web page of Weibo misinformation processing platform

册用户数已超过 5 亿, 日活跃用户达到 4620 万, 用户每日发博量超过 1 亿条. 由于谣言泛滥带来的负面后果日趋严重, 新浪微博先后采取多种措施来应对微博谣言. 在 2012 年 5 月底, 新浪微博颁布实施《新浪微博社区公约》及一整套管理体系, 明确微博用户权利, 用户行为规范及社区管理机制. 相应地, 新浪微博专门设立了社区管理中心和举报处理大厅, 受理用户对各类不良信息的举报, 在人工查证后对相关违规微博进行公示和处理. 该中心将不良信息划分为: 不实信息、泄露他人隐私、人身攻击、内容抄袭、冒充他人、骚扰他人等类别. 从中心公布的数据来看, 这些信息中不实信息的举报和处理比例最大, 可见谣言的泛滥之广与危害之深. 我们利用新浪微博社区管理中心公布的举报处理网页, 可以便捷地获取微博谣言数据. 管理中心在公示时会提供被认定为谣言的微博内容、发布人、举报人、判定证据等信息, 图 1 展示了被判定为不实信息的处理页面.

我们收集了发布时间在 2011 年 8 月到 2013 年 5 月间的 9079 条谣言, 建立微博谣言数据集, 其中大部分是在 2012 年 5 月《公约》实施以后发布的, 这些谣言发布时间的分布情况如图 2 所示. 接下来, 我们首先介绍对该数据的预处理过程, 然后面向谣言数据分别进行了统计、语义和时序等几个方面的定量分析. 最后, 基于这些分析结果, 我们提出了一套综合机器智能和群体智能的社交媒体自动辟谣框架.

2 数据处理

为了对微博谣言进行统计和语义分析, 我们需要对其进行预处理. 我们将每条微博作为一篇文档, 采用信息检索和自然语言处理中常用的文档表示方法词袋模型 (bag-of-words model), 用微博中的词项及其权重代表这篇文档, 不考虑词项在微博中出现的顺序.

词袋模型中词项权重有很多计算方法, 其中最简单的是采用词项在微博中出现的频率作为词权重计算方法, 即对于词项 t , 将 t 在文档 d 中出现的频度称为词项频率 (term frequency), 记为 $TF_{t,d}$. 然而, 某些词项对于相关度计算而言区分能力很小. 例如, 有非常多微博会包含“转发”这一词项, 这时候“转发”的区分能力就很小, 无法很好地体现一条微博的主题. 为此, 我们引入文档频率 (document

frequency) DF_t 的概念, 表示词项 t 在所有文档 (微博) 中出现的频率. 假设所有的微博数为 N , 我们定义词项 t 的逆文档频率 (inverse DF, IDF) 为 $IDF_t = \log \frac{N}{DF_t}$. 这样, 高频的没有区分性的词的 IDF 值相对比较低. 因此, 我们用 TFIDF 来衡量每条微博 d 中词项 t 的权重, 即 $TFIDF_{t,d} = TF_{t,d} \times IDF_t$.

首先, 我们采用分词工具 ICTCLAS^[15] 对每一条微博谣言进行分词, 然后去除“的”, “着”等停用词. 这样, 每条微博表示为一个向量, 其中每个分量都对应词典中的一个词项, 分量值为该词项的 TFIDF 权重. 当某词项在一条微博中没有出现时, 对应的分量值就为 0. 这样, 所有微博都在同一向量空间中表示, 称为向量空间模型 (vector space model, 简称 VSM). 对于数据集中任意两个微博的向量 V_x 与 V_y , 我们可以通过余弦相似度 (cosine similarity) 计算其语义相关度: $SIM(V_x, V_y) = \frac{V_x \cdot V_y}{|V_x| |V_y|}$. 两条微博间的余弦相似度越高, 它们就越可能是同一话题的微博.

在谣言数据集中, 有大量重复出现的谣言, 这些谣言虽然有细微变化 (如地名、人名等变化), 但主题和内容相同. 为了避免某条谣言重复出现对语义分析造成的偏置影响, 在谣言的语义分析中, 我们希望将这些重复谣言聚类后进行分析. 我们基于余弦相似度对数据集中微博谣言进行聚类, 将同一话题的微博划分到一起, 具体算法如下: (1) 设定阈值 T , 初始聚类集合 R 为空; (2) 对于数据集中的每条微博 V_i , 计算其与聚类集合 R 中每一个聚类 c 的代表向量 V_c 的余弦相似度 $SIM(V_i, V_c)$, 若相似度大于阈值 T , 则将其加入到 c 中; 若所有聚类均不满足条件, 则新建一个类 c' , 并将 V_i 加入 c' , 将 c' 添加到聚类集合 R 中; (3) 计算添加 V_i 的类 c 的新代表向量 V_c , 即计算 c 中各向量的中心点. 我们利用该方法不断调整阈值, 保证每个聚类的谣言都是关于同一个话题的重复谣言, 最终将 9079 条微博聚成 1798 类. 我们将在对谣言进行语义分析时, 利用每个类中的代表性微博进行定量分析.

3 谣言的定量统计分析

我们首先从谣言的影响力, 首次举报时间以及谣言的发布者与举报人等方面对谣言进行定量统计分析.

3.1 微博谣言的影响力

由于微博谣言的话题, 发布时间和发布人等各方面的差异, 每条微博谣言的影响力都不同. 我们可以通过一条微博谣言被转发和评论情况来近似地衡量该微博的影响力, 即微博谣言被转发和评论次数越高, 说明其影响力越大. 数据集中所有微博的转发和评论数统计如图 3 所示.

通过该图可以观察到, 大多数微博谣言的影响力都比较小, 转发和评论数在 500 次以下的微博占到整体的 84%. 只有极少量的微博谣言具有极广的传播范围和强大的影响力. 如果我们可以借鉴最近微博信息流行度预测的工作, 能够预测谣言的转发和评论数乃至传播范围, 将能够有效遏制微博谣言的传播范围和负面影响.

3.2 微博谣言的首次举报时间

谣言的首次举报时间指的是微博谣言从发布到第一次被举报所经过的时间间隔. 一条微博谣言的首次举报时间越短, 一定程度上可以反映出该微博的辟谣难度越小, 受关注的程度越高. 数据集中的微博谣言的首次举报时间分布如图 4 所示.

通过该图可以看到, 一条微博谣言通常需要一定的时间传播, 被一定数量的用户看到后才有可能被举报, 因此很少有微博谣言会在发布后的 10 分钟内被举报 (5.3%). 同时, 微博谣言由于其传播的速

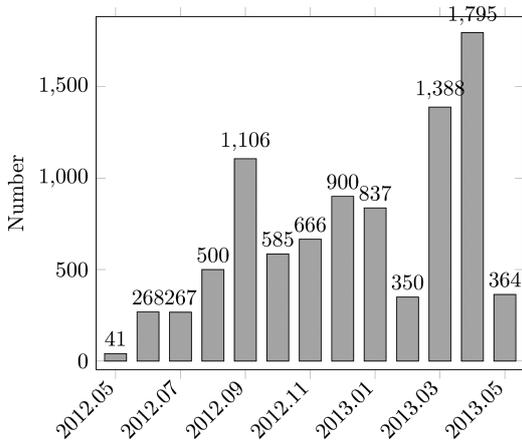


图 2 微博谣言发布时间分布情况

Figure 2 The distribution of Weibo rumors according to posting timestamps

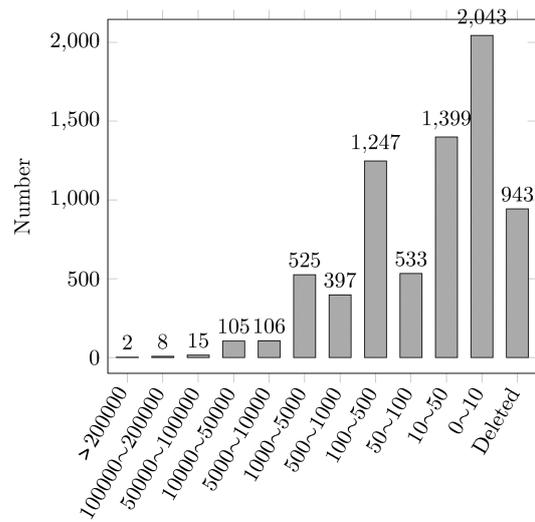


图 3 微博谣言按转发量的分布情况

Figure 3 The distribution of Weibo rumors according to forwarding numbers

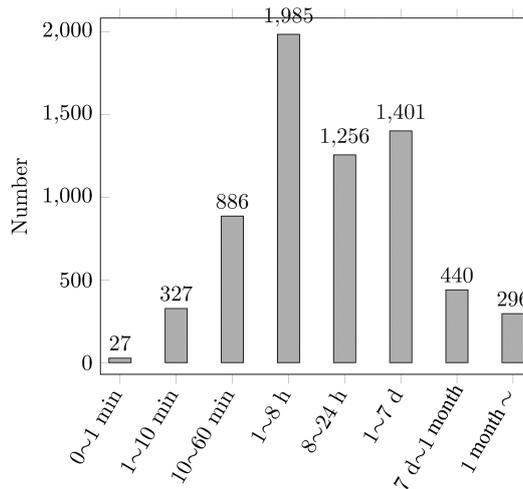


图 4 微博谣言按举报时间间隔的分布情况

Figure 4 The distribution of Weibo rumors according to report time intervals

度较快, 大部分的微博谣言会在其发布的一个周内被举报并辟谣 (88.9%).

从该图统计情况, 我们总结认为: (1) 由于谣言被用户举报的前提是微博谣言已经得到一定程度的传播, 因此利用计算机尽快地自动检测谣言并辟谣, 更具重要意义. 有句谚语称“当真相还在穿鞋, 谣言已经跑遍半个世界”, 我们希望通过计算机的自动检测与辟谣, 帮助“真相”把鞋穿得更快. (2) 大部分微博谣言会在一周内被举报并辟谣的现实, 要求计算机自动辟谣能够做到智能高效, 否则无法起到辅助人工辟谣的作用.

3.3 谣言举报人和谣言发布者

微博谣言数据集中每条谣言都对应一个发布者和一个举报人, 数据集中的谣言共涉及到 7055 个谣言发布者和 4559 个举报人. 通过选取其中举报谣言条数较多的举报人和发布谣言条数较多的发布者进行分析, 我们可以发现一些规律:

(1) 大量举报谣言的用户, 所举报的谣言往往与自己相关. 例如, 微博用户“美汁源饮料”举报了大量关于“美汁源果粒橙”饮料含有农药的谣言, 张家界纪委书记汪业元举报了大量关于“汪业元发表‘对网络暴民杀无赦’的言论”的谣言, 等等.

(2) 大量发布谣言的用户, 往往带有网络水军的性质, 例如, 微博用户“猫大为美”于 2012 年 10 月 19 日上午 5 时左右, 仅在几分钟的时间里发布了几十条微博, 其中大部分是谣言, 之后该用户就再未发过微博. 对于类似网络水军的谣言发布者, 往往很难对谣言制造者给予应有惩罚, 从这个角度来看, 社交媒体实名制也许对微博谣言具有一定的抑制作用.

以上是对于举报谣言数量较多的举报者和发布谣言数量较多的发布者的经验性观察结果. 未来我们可以进一步地结合微博发布者和举报者的粉丝数、有无签名、签名内容、注册时间、认证类型等特征, 判断微博用户的可信度, 进而辅助判定微博信息的可信性.

4 谣言的语义分析

微博谣言涉及人类社会生活的方方面面, 话题多种多样. 与普通微博信息相比, 微博谣言的主题有比较明显的倾向性. 根据微博谣言的主题, 我们可以大致将其划分为以下 5 类:

(1) 政治类谣言. 带有明显意识形态或者政治斗争目的的, 与国际、国内政治话题相关的谣言. 例如, 某年月日, 钓鱼岛海域中日两国爆发海战的谣言.

(2) 经济类谣言. 涉及某些公司企业或者与经济贸易相关的谣言. 如三星赔偿苹果几十车硬币的谣言.

(3) 欺诈类谣言. 利用人们同情或是逐利的心理, 留下虚假的联系方式或是骗取人们大量关注的谣言, 多含有“求转发”、“转发有奖”等字. 例如, 四川藏区儿童需要御寒冬衣, 并留下了一个虚假的联系电话的谣言.

(4) 社会生活类谣言. 多为社会各界人物的花边新闻等. 例如, 六小龄童去世的谣言.

(5) 常识类谣言. 关于自然常识、历史常识、生活常识的谣言. 例如, 阿司匹林能治疗心脏病的谣言.

我们请两位标注人员对数据集中的微博谣言进行标注, 将这些谣言按主题分到上述五个主题类别中, 每条微博谣言都属于并且只属于上述 5 个类别中的一个. 各个类别的统计情况如图 5 所示. 大部分的微博谣言属于社会生活类和政治类谣言 (约占 70%), 其他 3 类谣言所占的比例大体相当. 从该图可以看出, 微博谣言有着较强的主题倾向性, 主要分布在网络用户比较关注的领域 (如政治类) 和与个体生活息息相关的领域 (社会生活类).

微博谣言的主题分类信息具有重要意义, 有助于提高微博不实信息自动辟谣的效率. 在上述带标注的谣言数据集的基础上, 我们建立了微博谣言的主题分类器, 能够将谣言自动划分到 5 个类别中. 我们采用第 2 节的词袋模型表示每条微博谣言, 利用 LibLinear 工具^[16]训练线性分类器. 分类器的分类结果如表 1 所示. 其中分类出错较多的情况是将社会生活类谣言错误地分为政治类或将政治类分为社会生活类, 出现这种情况的原因有两个:

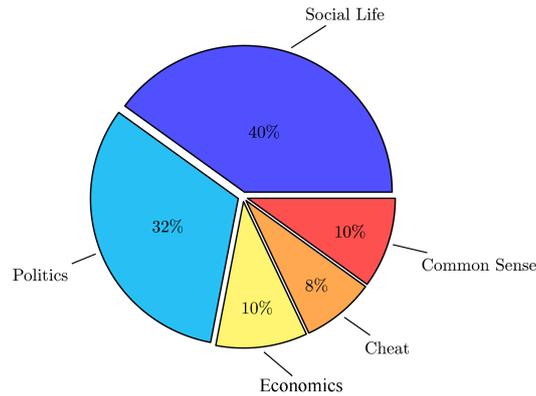


图 5 (网络版彩图) 微博谣言按主题类别的分布情况

Figure 5 (Color online) The distribution of Weibo rumors according to topic types

表 1 谣言主题类别分类结果 (五折交叉验证, 平均准确率为 88.2%)

Table 1 The classification results of rumor topic types

Be classified into =>	Cheat	Politics	Economics	Social life	Common sense	Recall (%)
Cheat	139	2	2	8	1	91.4
Politics	5	515	5	43	5	89.9
Economics	2	4	155	7	7	88.6
Social life	14	38	18	635	22	87.3
Common sense	5	4	6	14	142	83.0
Precision (%)	84.2	91.5	83.3	89.8	80.2	—

(1) 这两个类别谣言基数较大, 因此它们之间互相分错的可能性更大. 此外, 我们可以看到社会生活类谣言容易被错分为其他类型, 这也说明社会生活类的边界本身比较模糊.

(2) 在人工标注时, 我们要求每条谣言都属于且只属于一个类别, 但实际上谣言内容可能涉及多个主题和领域. 例如, “杨澜是美国国籍”的谣言, 在数据集标注过程中被分在社会生活类, 但该谣言也带有一定的政治目的, 如果被分为政治类谣言也可以理解. 因此, 分类器的实际性能应比 88.2% 的准确率要高.

5 谣言的时序分析

每条微博谣言都会经历发布、传播、高峰和消亡的过程, 不同的是这个过程有的很漫长, 也有的非常短暂, 属于同一话题的谣言还有可能会重复出现多次. 微博谣言的时序发展趋势可以在一定程度上反映该谣言的热度和辟谣难度等. 这里我们将对微博谣言进行时序分析.

从第 2 节谣言数据聚类结果中, 我们取出包含谣言数目最多的 100 个话题, 分析其总转发数的变化趋势. 按照微博谣言主题的转发数峰值出现的时间和峰值数量, 我们归纳出以下 4 种时序类型 (如图 6 所示):

- A1 谣言刚发布时即有较大峰值, 只有一个较大峰值;
- A2 谣言刚发布时即有较大峰值, 且有多个较大峰值;

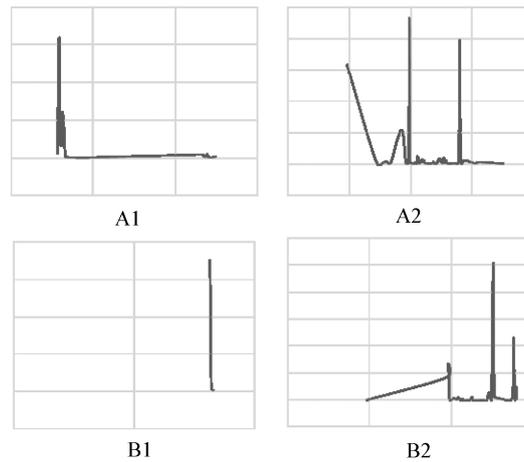


图 6 转发时序类型划分. A1. 《贞子 3D》内地上映的谣言; A2. 四川藏区需要捐衣服的谣言; B1. 中日钓鱼岛开战的谣言; B2. 新的交通法规实施的谣言

Figure 6 Four types of forwarding temporal patterns for Weibo rumors. A1. The rumor about the Japanese movie “Sadako 3D” to be shown in China; A2. the rumor about clothes donation required from Sichuan Tibetan region; B1. the rumor about Sino-Japanese War on the Diaoyu Island; B2. the rumor about the new traffic regulation to be implemented

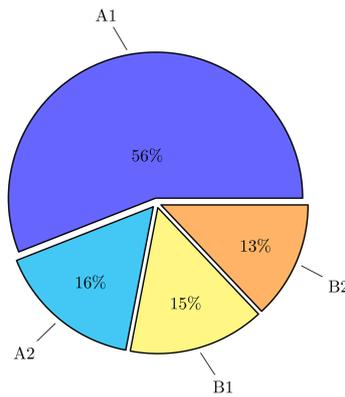


图 7 (网络版彩图) 微博谣言按时序类别的分布情况

Figure 7 (Color online) The distribution of Weibo rumors according to temporal patterns

- B1 谣言发布一段时间后才有较大峰值, 只有一个较大峰值;
- B2 谣言发布一段时间后才有较大峰值, 且多个较大峰值.

这 4 种时序类型的微博谣言话题所占的比例如图 7 所示. 通过该图可以发现:

(1) 多数谣言话题在刚发布时就会出现转发峰值, 与发布一段时间后才出现峰值的话题的比例约为 7:3. 多数谣言话题只有一个较大转发峰值, 与多个峰值的谣言话题的比例也约为 7:3.

(2) 微博谣言话题的内容主题类别与其时序类型存在一定联系. 例如, 常识类谣言由于受众广, 辟谣难度较大, 往往会反复被人们提及, 出现多次爆发, 约 70% 的常识类谣言通常有多个转发峰值. 而关于名人或知名机构的谣言, 由于关注人数众多, 辟谣难度较小, 因此发布之初就会出现较大转发峰值, 但很快会被辟谣, 约 60% 的此类谣言会在一个周内消亡.

6 面向社交媒体的自动辟谣框架

我们已经对社交媒体进行了定量统计、语义分析和时序分析. 进行这些分析的最终目的是实现对谣言的自动识别. 已有分析表明大部分谣言 (70% 以上) 只出现一个转发峰值, 而且大部分在刚发布就会出现转发峰值 (70% 以上). 通过分析疑似谣言的评论信息和转发模式的方法^[10] 需要以大量的转发数据为基础, 很有可能错过谣言转发峰值, 丧失自动辟谣, 降低谣言影响的最佳时机.

为了实现更快速及时的自动辟谣, 我们首先需要了解人们为什么会相信谣言. 通过对新浪微博中广泛传播的 859 条谣言的经验性分析, 我们按照谣言被相信的原因将谣言分为两类^[17]:

(1) 知识受限谣言, 即缺乏专业知识而导致误信或无法辨认的谣言. 例如, 阿司匹林可以治疗急性心脏病.

(2) 时空受限谣言, 即由于地域和时间限制无法辨认的谣言. 例如, 有谣言称“杭州上城区一妇女喝了 3 罐可乐, 两天后离开了这个世界. 验尸结果是她死于细螺旋体病, 发病原因是直接用嘴对可乐罐饮用”. 统计发现, 知识受限谣言通常关注科学技术的话题, 时空受限谣言则更多地关注政治和社会的话题, 并会更多的提及特定的人名、地名和机构名.

目前, 自然语言处理技术还无法根据微博内容自动判定是否为谣言. 因此, 我们提出一种结合机器智能和群体智能的思想, 通过分析微博用户的专长和可疑谣言的主题, 利用专家发现技术找到最有可能判定该谣言的专家. 我们将所有微博用户构成一个候选专家集合, 记作 E . 对于一条可疑微博 m , 可以通过下面的公式来估算集合 E 中的用户 e 是关于这条微博 m 的专家的概率:

$$\Pr(e|m) = \frac{\Pr(m|e) \Pr(e)}{\Pr(m)} \propto \Pr(m|e) \Pr(e), \quad (1)$$

其中 $\Pr(m)$ 是微博 m 的先验概率, 对于所有的用户来说是一个定值. $\Pr(e)$ 是专家 e 的先验概率, 可以作为用户 e 的权威度的测度. 而 $\Pr(m|e)$ 则衡量了用户 e 关于信息 m 的知识水平. 新浪微博允许每个用户给自己标注标签, 这些标签代表了用户们感兴趣的领域. 因此, 我们可以采用微博用户的社交标签来模拟他们的知识水平, 估计 $\Pr(m|e)$ 值.

对于知识受限谣言, 我们可以直接使用上述公式进行专家发现. 但对于时空受限谣言, 我们需要添加一些约束条件 (微博中出现的相关权威部门或当地知情人士) 来找出可能的专家集合. 在知识受限谣言和时空受限谣言的专家发现方面, 我们提出的算法分别达到 61.5% 和 76.0% 的准确率, 详细评测过程和结果可参考文献 [17].

基于上述微博谣言的特征和专家发现的方法, 我们提出一种面向社交媒体的自动辟谣框架如图 8 所示, 主要包括 3 个阶段工作:

(1) 谣言发布早期, 通过用户举报和对可疑用户的监控建立疑似谣言的集合. 一方面, 将疑似谣言和谣言库中的谣言进行比对, 以此来发现一些每隔一段时间就会反复出现的谣言. 另一方面, 对于在谣言库中没有匹配内容的谣言, 通过查询该领域的专家库, 推荐若干专家对该疑似谣言进行鉴别, 判定该信息是否为谣言.

(2) 谣言发布中期, 通过自然语言处理技术分析疑似谣言的评论信息, 通过社会网络分析技术分析疑似谣言的传播模式, 判定该信息是否为谣言.

(3) 谣言发布后期, 对于判定为谣言的信息, 将其加入谣言库; 对该信息的发布人进行可信性分析, 确定其信用等级, 将信用等级低于一定阈值的用户加入可疑用户库, 在一段时间内对其发布的微博内容进行监控; 对该信息的举报人和评论人进行专家发现, 充实和更新该信息相关的知识领域的专家库.

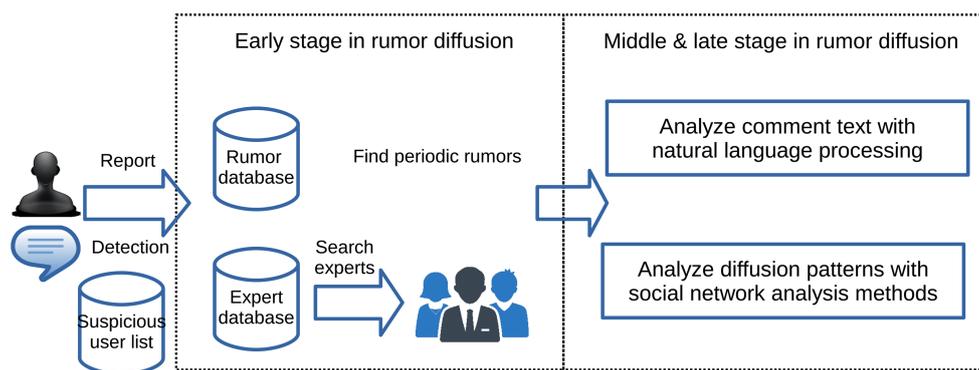


图 8 (网络版彩图) 面向社会媒体的自动辟谣框架

Figure 8 (Color online) The framework of refuting misinformation on social media

7 总结

本文以中文社交媒体新浪微博为研究平台, 从定量统计、语义分析和时序分析等角度, 对收集的 9 千余条谣言信息进行了分析研究, 并面向社交媒体提出一套综合机器智能与群体智能的自动辟谣框架, 实验证明其具有较好的有效性和可行性。

据作者了解, 这是首次面向中文社交媒体大规模谣言数据比较全面的定量分析研究。社交媒体谣言还有很多方面亟待深入探索, 例如信息发布者与信息的可信性度量手段、不实信息和可信信息的传播模式、微博评论内容的观点分析与挖掘、专家发现与匹配等等。总之, 谣言分析与信息可信性的定量分析已经成为当前社会的重要问题, 其高度复杂性要求社会学、传播学和计算机科学的研究者们齐心协力, 共同探索出一条解决之道。

致谢 感谢清华大学新闻与传播学院金兼斌教授等专家对本工作提出的建议和意见。本工作成果曾应邀在由中国社会科学院语言研究所、北京语言大学、商务印书馆和人民网舆情监测室主办的“2015 中青年语言学沙龙”上报告, 感谢与会专家提出的建议和意见。

参考文献

- 1 Cai J. Rumor: Social Transmission in the Shadow. Beijing: China Radio and TV Press, 2008 [蔡静. 流言: 阴影中的社会传播. 北京: 中国广播电视出版社, 2008]
- 2 Zhou Y Q. A Study of the Rumors in the internet of Contemporary China. Beijing: the Commercial Press, 2012 [周裕琼. 当代中国社会的网络谣言研究. 北京: 商务印书馆, 2012]
- 3 Allport G W, Postman L. The Psychology of Rumor. Oxford: Henry Holt, 1947
- 4 Kapferer J-N. Rumeurs: le Plus Vieux Média du Monde (in French). Paris: Le Seuil Editions, 1987
- 5 Peterson W A, Gist N P. Rumor and public opinion. American J Soc, 1951, 57: 159-167
- 6 Budak C, Agrawal D, Abbadi A El. Limiting the spread of misinformation in social networks. In: Proceedings of the 20th International Conference on World Wide Web, Byderabad, 2011. 665-674
- 7 Castillo C, Mendoza M, Poblete B. Information credibility on twitter. In: Proceedings of the 20th International Conference on World Wide Web, Byderabad, 2011. 675-684
- 8 Nguyen D T, Nguyen N P, Thai M T. Sources of misinformation in online social networks: who to suspect? In: Proceedings of Military Communications Conference. Orlando: IEEE, 2012. 1-6
- 9 Okazaki N, Nabeshima K, Watanabe K, et al. Extracting and aggregating false information from microblogs. In: Proceedings of the Workshop on Language Processing and Crisis Information, Nagoya, 2013. 36-43

- 10 Qazvinian V, Rosengren E, Radev D R, et al. Rumor has it: identifying misinformation in microblogs. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Stroudsburg: Association for Computational Linguistics, 2011. 1589–1599
- 11 Ratkiewicz J, Conover M, Meiss M, et al. Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th International Conference Companion on World Wide Web, Byderabad, 2011. 249–252
- 12 Liao Q Y, Shi L. She gets a sports car from our donation: rumor transmission in a chinese microblogging community. In: Proceedings of the 2013 Conference on Computer Supported Cooperative Work. San Antonio: ACM, 2013. 587–598
- 13 Sun S Y, Liu H Y, He J, et al. Detecting event rumors on Sina Weibo automatically. Web Technologies and Applications. In: 15th Asia-Pacific Web Conference (APWeb). Sydney: Springer, 2013, 7808: 120–131
- 14 Yang F, Liu Y, Yu X H, et al. Automatic detection of rumor on Sina Weibo. In: Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, Beijing, 2012. 13
- 15 Zhang H P, Yu H K, Xiong D Y, et al. Hhmm-based chinese lexical analyzer ictclas. In: Proceedings of the 2nd SIGHAN Workshop on Chinese Language Processing. Stroudsburg: Association for Computational Linguistics, 2003. 184–187
- 16 Fan R E, Chang K W, Hsieh C J, et al. Liblinear: a library for large linear classification. J Mach Learn Res, 2008, 9: 1871–1874
- 17 Liang C, Liu Z Y, Sun M S. Expert finding for microblog misinformation identification. In: Proceedings of International Conference on Computational Linguistics, Mumbai, 2012. 703–712

Statistical and semantic analysis of rumors in Chinese social media

LIU ZhiYuan^{1,2,3*}, ZHANG Le^{1,2,3}, TU CunChao^{1,2,3} & SUN MaoSong^{1,2,3}

1 *State Key Laboratory of Intelligent Technology and Systems, Beijing 100084, China;*

2 *Tsinghua National Laboratory for Information Science and Technology, Beijing 100084, China;*

3 *Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China*

*E-mail: liuzy@tsinghua.edu.cn

Abstract In this study, we collect a large number of rumors from Sina Weibo, and analyze them from different perspectives. We study the influence of rumors from a quantitative perspective, as well as their beginning and ending. From a semantic analysis perspective, we manually classify the rumors into five categories based on the semantic information. Then, we train a classifier to classify the rumors automatically. From a timing perspective, we manually classify the rumors into four categories based on the forward trends. Then, we find the propagation tendency of rumors. Finally, we propose a rumor-recognition system, which combines machine intelligence and swarm intelligence.

Keywords social media, rumor, semantic analysis, temporal analysis, rumor recognition



LIU ZhiYuan was born in 1984. He received the Ph.D. degree in Computer Science from the Department of Computer Science and Technology, Tsinghua University, Beijing, in 2011. Currently, he is an assistant researcher in the Department of Computer Science and Technology, Tsinghua University. His research interests include natural language processing and social computation.



ZHANG Le was born in 1992. He received the B.E. degree in Computer Science from the Department of Computer Science and Technology, Tsinghua University, Beijing, in 2010. Currently, he is a Ph.D. student at the Tsinghua University. His research interests are cloud databases, database security, and privacy protection.



TU CunChao was born in 1990. He received the B.E. degree in Computer Science from the Department of Computer Science and Technology, Tsinghua University, Beijing, in 2013. Currently, he is a Ph.D. student in the Department of Computer Science and Technology, Tsinghua University. His research interests are user representation and social computation.



SUN MaoSong was born in 1962. He received the B.E. and M.E. degrees in Computer Science from the Department of Computer Science and Technology, Tsinghua University in 1986 and 1988, respectively, and received the Ph.D. degree in 2004 from the Department of Chinese, Translation and Linguistics, City University of Hong Kong. Currently, he is a Professor in the Department of Computer Science and Technology, Tsinghua University. His research interests are natural language processing, Chinese computing, web intelligence, and computational social sciences.