

基于用户兴趣的微博溯源算法

杨潇^{1,2}, 陈秀真^{1,2}, 马进^{1,2}, 梁浩喆^{1,2}, 李生红^{1,2}

(1. 上海交通大学网络安全技术研究院, 上海 200240;
2. 上海市信息安全综合管理技术研究重点实验室, 上海 200240)

摘要: 微博信息溯源通过分析在平台采集的话题数据集, 挖掘相关话题的真正源头, 即发布时间较早且影响力大的微博集合, 实现网络舆论的管控与引导。提出一种基于用户兴趣的微博溯源算法, 该算法根据博主的兴趣计算博主影响力, 同时根据评论人、转发人的兴趣计算评论人、转发人的影响力, 结合博主关注度和发表时间等因素, 利用网页排序算法对微博评分, 根据微博得分进行排序溯源。实验结果表明, 该算法相较于传统溯源算法在查全率上提升了约 21%。

关键词: 信息溯源; 微博; 兴趣; 影响力; 关注度

中图分类号: TP391

文献标识码: A

doi: 10.11959/j.issn.2096-109x.2020086

User interests-based microblog tracing algorithm

YANG Xiao^{1,2}, CHEN Xiuzhen^{1,2}, MA Jin^{1,2}, LIANG Haozhe^{1,2}, LI Shenghong^{1,2}

1. Institute of Cyber Science and Technology, Shanghai Jiaotong University, Shanghai 200240, China
2. Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, Shanghai 200240, China

Abstract: Microblog information tracing refers to finding the source set of microblog topics according to the analysis of crawled microblog texts and it's of great significance in the aspect of public opinion control and information security. A user interests-based tracing method (ITM) was proposed. The proposed method calculates the influence of the blogger based on the interest of the microblog blogger, and also calculates the influence of the commentators based on the interest of the commentators. The ranking algorithm was used to score the blogs according to publication time, notability and influence, and the source of the blogs was traced according to the blog score rank. Experimental results show that the accuracy of the proposed algorithm improved about 21% compared with the traditional tracing algorithms.

Key words: information tracing, microblog, interest, influence, notability

收稿日期: 2019-11-20; 修回日期: 2020-09-23

通信作者: 陈秀真, chenxz@sjtu.edu.cn

基金项目: 国家重点研发计划 (2016YFB0801003); 国家自然科学基金 (61562004, 61431008)

Foundation Items: The National Key R&D Program of China (2016YFB0801003), The National Natural Science Foundation of China (61562004, 61431008)

论文引用格式: 杨潇, 陈秀真, 马进, 等. 基于用户兴趣的微博溯源算法[J]. 网络与信息安全学报, 2020, 6(6): 164-173.

YAGN X, CHEN X Z, MA J, et al. User interests-based microblog tracing algorithm[J]. Chinese Journal of Network and Information Security, 2020, 6(6): 164-173.

1 引言

微博作为国内最大的自媒体平台之一, 往往存在各类谣言、敏感话题等相关信息。微博信息的溯源, 不仅对维护信息安全有重要意义, 在信息传播研究以及在线社交平台分析上有较多应用。

在信息溯源方面, 有学者做过相关研究。时国华等^[1]根据微博的发表时间、原创性和中心性, 同时结合微博的转发关系对微博进行溯源。信息传播中, 用户特征有一定参考性, 基于此, 基于用户特征的算法被提出。刘荣叁^[2]等根据用户粉丝数、评论数等信息计算用户影响力, 同时结合排序算法计算出微博源头。米昂、张跃伟、刘岗等通过博主的发博频率, 微博的原创性系数、转发量、评论量、转发关系等因素计算得到微博源头^[3-5]。Wang 等^[6]利用用户的特征信息构建信息矩阵来判断用户之间的传播关系。吴信东等^[7]通过构建信息传播结构以及分析用户行为来确定微博信息传播的源头。近几年, 一些基于网络结构与模型的溯源算法被广泛研究。Sadikov 等^[8]通过构建 K 树模型来得到信息传播路径, 从而对信息溯源。杨静等^[9-10]将网络传播模型 AN 与微博参与者的数量结合起来, 利用公式递推微博源头。尹熙成等^[11]通过构建微博信息传播中的主路径来计算各个传播节点的权值从而找到影响力最大的节点作为源头。Leskovec^[12]通过提取出信息传播路径中的主干路径来寻找信息源头。何冰心等^[13]在原有病毒传播模型的基础上, 利用信息社区划分以及人群结构特征等属性提出了一种群行为特性划分的信息传播模型 (CAPIR) 对信息进行溯源。社交网络用户之间, 往往存在社会性拓扑结构。基于此, 王梦迪等^[14]提出一种基于克罗内克网络模型的初始邻接矩阵生成算法来迭代求得信息源头。Xiao 等^[15]从多消息传播网络入手, 引入话题传播树的传播模型, 利用信息传播网络来构建话题传播树, 从而确定传播路径并且溯源。部分学者将数理分析的方法用来分析信息溯源。Soheil 等^[16]根据时间、内容以及关联性等因素构建数学传播概率模型

来判断源头信息。INUI 等^[17]基于 twitter 平台, 将推文进行分割, 并结合相关内容发布次数建立了一种新闻源头溯源模型。

当前对于微博溯源的研究方法主要分为两类。一类是根据微博文本相似度, 或者利用微博的评论数、粉丝数等参数进行简单计算来溯源。另一类是根据转发关系以及传播的拓扑结构等信息利用数学模型迭代求得路径源头。但是两类方法都没有考虑微博的文本内容、其他用户对该微博所评论的内容以及转发内容所产生的影响。针对该问题, 本文提出了一种基于用户兴趣的微博溯源算法 (ITM, interests-based tracing algorithm)。该算法根据博主以前微博的内容计算出博主的兴趣, 然后根据博主兴趣计算博主影响力, 同时计算微博评论人、转发人的兴趣, 求得评论人、转发人的影响力。最后结合博主关注度以及微博发表时间等因素利用网页排序算法对所有微博进行评分, 根据微博得分进行排序溯源。与现有的两类方法相比, 本文所提算法考虑到了微博信息的文本内容对信息传播带来的影响, 从而更准确地分析用户的影响力。该算法利用用户兴趣, 挖掘计算出博主对于拥有类似兴趣的用户社区的影响, 从而更准确地定位信息源头。

2 用户兴趣的计算

微博用户的行为往往可以反映用户的一些兴趣爱好。例如, 某个用户喜欢娱乐方面的内容, 该用户发表、评论以及转发的微博会倾向于娱乐方面的信息。因此通过用户以前的微博信息, 可以分析出用户的兴趣倾向。

通过对大量微博数据的分析, 用户兴趣可以分为以下 5 类: 文娱、经济、科教、政治、军事。一条微博的内容与某个兴趣的相关程度越高, 那么这条微博中的关键词与该兴趣的相关度越高。因此, 本文通过提取用户以前博客信息中的关键词, 并计算其与上述某个兴趣关键词在知网词林中的距离得到微博与该兴趣的相关程度。

用户的某条博客信息与某个兴趣 i 的相关程度计算公式如下。

$$\text{BlogIntSim}(i) = \sum_{j=1}^m \text{KeyWordWeight}(j) \quad (1)$$

$$\text{HowNetDis}(\text{KeyWord}(j), \text{IntWord}(i))$$

其中, $\text{BlogIntSim}(i)$ 表示用户的一条博客信息与兴趣 i (文娱、经济、科教、政治、军事之一) 的相关程度。 $\text{KeyWord}(j)$ 表示从该条微博文本中利用汉语词法分析系统 (ICTCLAS) 提取的关键词, $\text{KeyWordWeight}(j)$ 表示该关键词对应的权值, $\text{IntWord}(i)$ 表示兴趣 i 的中文词汇。 HowNetDis 函数用于计算两个词在知网词林中的距离^[18-19]。

知网词林是一种分层次体系的词典, 利用义项和义原来定义词汇。义项是对一个词的一种描述, 一个词可以有多个义项, 而义原就是描述一个义项的基本单位。在知网词林中, 义项的基本结构如图 1 所示。

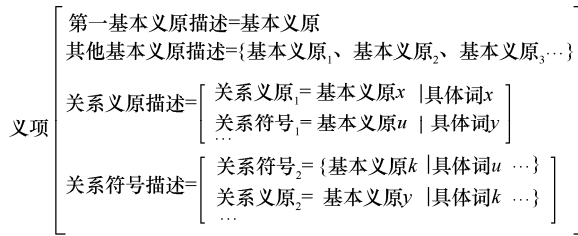


图 1 义项结构
Figure 1 Sememe structure

词语之间的距离使用义项来计算, 如果有两个词语 W_1, W_2 , 其中 W_1 的义项为 $S_{11}, S_{12}, S_{13}, \dots, S_{1n}$, 而 W_2 的义项为 $S_{21}, S_{22}, S_{23}, \dots, S_{2m}$, 则 W_1, W_2 的距离计算公式如下。

$$\text{HowNetDis}(W_1, W_2) = \max_{i=1, \dots, n, j=1, \dots, m} \text{Sim}(S_{1i}, S_{2j}) \quad (2)$$

由式(2)可知, 词语之间的距离是两个词各个义项之间相似度的最大值, 而义项是由义原定义的, 所以义项之间的相似度由义原计算。由图 1 可知, 义原呈树形结构, 通过考虑义项的义原层次树中 4 类义原描述之间的距离, 可以得到其义项间的相似度, 如下。

$$\text{Sim}(S1, S2) = \sum_{j=1}^4 \mu_j \text{Sim}_j(S1, S2) \quad (3)$$

其中, $\text{Sim}_j(S1, S2)$ 为义项 $S1, S2$ 之间 4 类义原描述之间的距离, 通过考虑义原层次树的深度、密度来计算, μ_j 为对应 4 类义原描述间距离的权

重。 $\text{Sim}_j(S1, S2)$ 的计算公式如下。

$$\text{Sim}_j(S1, S2) = \frac{\gamma}{d + \gamma} \quad (4)$$

其中, d 为 $S1, S2$ 之间第 j 类义原在知网层次体系结构中的路径距离, γ 为可调参数。

利用知网词林可以计算每条微博与某个兴趣 i 之间的相关度, 根据相关度可以评估用户对 i 的感兴趣程度。用户的兴趣在短期时间内变化不会太大, 所以本文取近期一个月内用户发布的微博与兴趣 i 的相关度之和作为该用户对 i 的感兴趣程度, 公式如下。

$$\text{Interest}(i) = \sum \text{BlogIntSim}(i) \quad (5)$$

其中, $\text{Interest}(i)$ 是用户对 i 的感兴趣程度, $\text{BlogIntSim}(i)$ 是用户近期一个月内的一条博客信息与兴趣 i 的相关程度, 由式(1)计算。

3 基于用户兴趣的微博溯源算法

微博话题的溯源是指在给定关于某个事件的微博数据集的情况下, 根据一定指标和方法, 从中找出信息传播的源头。

基于对各类话题下源头微博的研究, 本文认为发布时间较早并且对社会大众拥有较大影响力的微博为源头微博。影响力较大微博的博主一般具有一定的兴趣倾向, 并且博主与粉丝有相似的兴趣, 博主发布一条微博时其粉丝可以获得该微博的相关推送。所以博主的粉丝数量越多, 发布的微博与博主兴趣越相似, 其对微博社区产生的影响越大。同理, 微博的评论内容和转发会推送给对应用户的粉丝, 评论人、转发人的粉丝数越大, 并且评论人及转发人的兴趣与原微博内容越相关, 影响力越大。如果博主近期发布的其他微博的关注度较高, 也会为新发布的微博带来一定的关注度, 从而为新微博的传播贡献一定的影响力。同时, 微博发布时间是判断源头的重要参数之一, 发布越早的微博能越快得到人们的关注, 其成为源头微博的可能性越大, 但微博的源头不一定是发布时间最早的微博, 因为有些微博可能是经由一些影响力大的人转发才被关注。本文所提 ITM 算法综合考虑博主影响力, 评论人、转发人的影响力、关注度、时间,

通过计算得分来对微博进行排序溯源, ITM 算法流程如图 2 所示。

对应的算法伪代码如算法 1 所示。

算法 1 Interest-based Tracing Method (ITM)

- 1) Input: Topic blogs Blogs = {Blog₁, Blog₂, ..., Blog_M}, source blogs Sources = {}, blog score set Scores = {}, number of sources N;
- 2) Output: Source blogs Sources;
- 3) for k = 1: M do
- 4) Calculate user influence: BloggerInf = User Influence Calculation(Blog_k);
- 5) Calculate comment influence: CommentInf = Comment Influence Calculation (Blog_k);
- 6) Calculate user attraction Attra;
- 7) Standardize time T;
- 8) Calculate blog score: Score = Blog Score Calculation(BloggerInf, CommentInf, Attra, T)
- 9) Add blog score Score in Scores: AddinScores(Score)
- 10) end for
- 11) Pick the N blogs with highest scores in

Scores into Sources;

12) return Sources

3.1 基于兴趣的博主影响力计算

博主所发的微博内容能反映博主的兴趣爱好, 其粉丝一般与其有着相同或者相似的兴趣, 所以博主所发微博的内容与其兴趣相似程度越高, 对其粉丝和大众的影响力越大。同时博主的粉丝数越多, 微博的影响力越大, 但是当微博粉丝数超过一定数量级之后(如千万级), 粉丝数贡献的影响力会趋紧饱和, 其影响力的增长趋于平缓。

基于以上分析和式(1)、式(5), 博主的微博在兴趣 i 方面(文娱、经济、科教、政治、军事之一)的影响力计算公式如下。

$$\text{BloggerIntInf}(i) = \text{Interest}(i) \cdot \text{BlogIntSim}(i) \cdot \log_{\beta}(\text{Fans} + 1) \quad (6)$$

其中, $\text{Interest}(i)$ 是博主对兴趣 i 的感兴趣程度, 由式(5)计算。 $\text{BlogIntSim}(i)$ 是微博内容与兴趣 i 的相关程度, 用式(1)计算。 $\log_{\beta}(\text{Fans} + 1)$ 表示粉丝数的影响力, 使用 $\log$ 函数计算, 该影响力会随着粉丝数 Fans 的增长变大, 但当粉丝数超过一

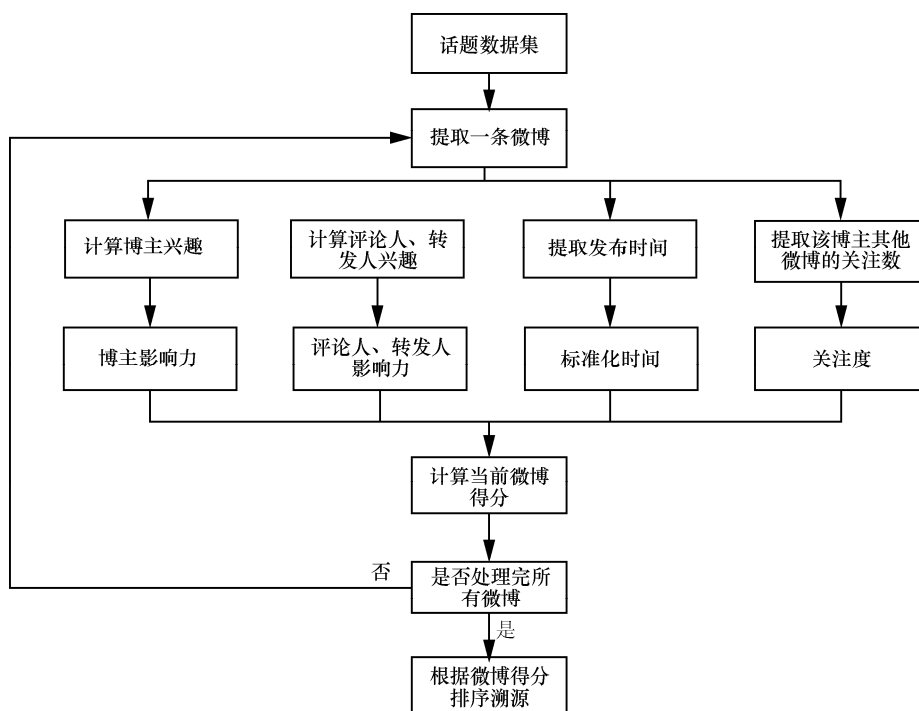


图2 算法流程
Figure 2 Algorithm flowchart

定数量级后,增长会变缓。由式(6)可以看出,博主影响力与 $\text{Interest}(i)$ 、 $\text{BlogIntSim}(i)$ 以及 $\log_{\beta}(\text{Fans}+1)$ 分别呈正相关关系。

在各个兴趣下的博主影响力之和为总的博主影响力,计算公式如下。

$$\text{BloggerInf} = \sum \text{BloggerIntInf}(i) \quad (7)$$

其中, BloggerInf 表示博主影响力, $\text{BloggerIntInf}(i)$ 表示博主在兴趣 i 方面的影响力,由式(6)计算。

博主影响力计算的算法2如下。

算法2 User Influence Calculation

- 1) Input: Topic blog Blog_k , interest fields $I = \{i_1, i_2, i_3, i_4, i_5\}$, user influence $\text{BloggerInf} = 0$;
- 2) Output: User influence BloggerInf ;
- 3) for $k = 1: 5$ do
- 4) Employ Blog_k to calculate user influence under a certain interest $\text{BloggerIntInf}(i_k)$ by Eq.(6);
- 5) Add $\text{BloggerIntInf}(i_k)$ in total influence BloggerInf : $\text{BloggerInf} += \text{BloggerIntInf}(i_k)$;
- 6) end for
- 7) return BloggerInf

3.2 基于兴趣的评论人、转发人影响力计算

用户可以在微博下方发表评论,评论可以被其他人点赞和留言,并且评论人的粉丝会被推送该评论的相关信息,所以评论人可以为其评论的微博贡献一定的影响力。评论人与其粉丝一般有相似的兴趣,评论人的兴趣和微博的话题越相似且评论人粉丝越多,影响力越大。同理,转发可以附加评论,也能为原微博贡献一定的影响力。

评论人、转发人在各个兴趣下的影响力的和即评论人、转发人总的影响力,计算公式如下。

$$\text{CommentInf} = \sum \text{ComIntInf}(i) \quad (8)$$

其中, CommentInf 为总的评论人、转发人影响力, $\text{ComIntInf}(i)$ 为评论人、转发人在兴趣 i 方面的影响力,其计算公式如下。

$$\text{ComIntInf}(i) = \sum \text{SinComIntInf}(l) \quad (9)$$

其中, $\text{SinComIntInf}(l)$ 为单个转发人、评论人在兴趣 i 方面的影响力。结合前面的分析,微博评

论人、转发人对兴趣 i 感兴趣程度越高,且微博内容与兴趣 i 的相似度越高,其影响力较大;转发人、评论人的粉丝越多,影响力越大,但是当微博粉丝数大到超过一定数量级之后,粉丝数所造成影响趋近饱和,其增长趋于平缓;评论内容的情感度与影响力也有一定关系,具有正面感情色彩的评论内容往往更具有号召力,其影响力较大,反之,负面感情的内容影响力较小;评论的点赞数表示评论人的被认同度,能体现一定的影响力,点赞数量越多则影响力越大,但是点赞数超过一定数量级时,影响力趋于饱和,增长趋于平缓。基于以上分析,单个评论人、转发人在兴趣 i 方面的影响力的计算公式如下。

$$\text{SinComIntInf}(l) = \text{ComInterest}(i) \cdot \text{BlogIntSim}(i) \cdot \log_{\beta}(\text{ComFans}+1) \cdot E \cdot L \quad (10)$$

其中, $\text{ComInterest}(i)$ 是单个转发人、评论人对兴趣 i 的感兴趣程度,由式(5)计算。 $\text{BlogIntSim}(i)$ 是微博内容与兴趣 i 的相关程度,由式(1)计算。 $\log_{\beta}(\text{ComFans}+1)$ 表示该评论人、转发人的粉丝数影响力, ComFans 是对应的粉丝数。 E 表示评论内容的感情度,利用语言分析组件 snowNLP 计算,0 表示最负面的情感,1 表示最正面的情感,0.5 表示中性的情感^[20]。 L 表示该条评论或转发的点赞数对影响力的贡献,使用 $\log$ 函数计算,公式如下。

$$L = \log_{\lambda}(\text{Like}+1) \quad (11)$$

其中, Like 表示该评论或转发的点赞数。

从式(8)~式(10)可以看出,评论人、转发人影响力与 $\text{ComInterest}(i)$ 、 $\text{BlogIntSim}(i)$ 、 $\log_{\beta}(\text{ComFans}+1)$ 、 E 以及 L 分别呈正相关。

评论人、转发人的影响力计算过程使用算法描述如下。

算法3 Comment Influence Calculation

- 1) Input: A Topic blog Blog_k 's comments $\text{Com} = \{\text{com}_1, \text{com}_2, \dots, \text{com}_N\}$, interest fields $I = \{i_1, i_2, i_3, i_4, i_5\}$, comment influence $\text{CommentInf} = 0$;
- 2) Output: Comment influence CommentInf ;
- 3) for $k = 1: 5$ do

```

4)   ComIntInf( $i_k$ ) = 0;
5)   for  $i = 1: N$  do
6)       Calculate single comment influence
under a certain interest SinComIntInf( $i_k$ ) by Eq.(9);
7)       Add all single comment influences
in total influence under a certain interest: ComIntInf( $i_k$ ) += SinComIntInf( $i_k$ );
8)       Add all total comment influence under
different interests together to get the final influence:
CommentInf += ComIntInf( $i_k$ )
9) end for
10) return CommentInf

```

3.3 关注度

博主近期的其他微博被关注的程度越高, 其他用户在浏览这些微博时越有大概率看到话题微博, 并且微博用户倾向于浏览关注度较高的博主所发的微博, 这也会为当前话题微博带来一定的关注度, 从而为话题微博的传播贡献一定的影响力。所以, 计算微博得分时, 应该考虑博主关注度。

微博博主的关注度, 由最近一个月内除话题微博外的其他微博的点赞、转发和评论的数量来计算, 数量越多则关注度越大。本文认为用户近期发布的微博带来的关注度比较影响力, 所以取一个月内发布的微博来计算关注度, 其计算公式如下。

$$Attr_a = \sum (LN + RN + CN) \cdot TW \quad (12)$$

其中, $Attr_a$ 表示博主的关注度, LN 、 RN 、 CN 分别表示博主一个月内发布的一条微博的点赞、转发、评论数量, TW 是根据微博的持续时间来计算对应权重参数, 通过文献[21]中基于粗糙集理论确定属性权重的方法获得。该方法通过对比属性之间的依赖度的差来计算属性的重要性, 即当集合去掉某些属性后, 整个集合的分类域以及决策属性受到影响的大小来确定属性权重。令微博数据集表示为 A , 微博持续时间为 h , 取数据集的属性分类为 $R = \{R_1, R_2, R_3\} = \{h < 2\text{Days}, 2\text{Days} \leq h < 4\text{Days}, 4\text{Days} \leq h < 7\text{Days}, h \geq 7\text{Days}\}$, 取微博的 LN 、 RN 、 CN 之和 sum 为 R_i 的子属性, 则构建属性分类如下。

$$\begin{aligned}
A / R_1 &= \{sum_1 < 10, 10 \leq sum_1 < 100, sum_1 \geq 100\} \\
A / R_2 &= \{sum_2 < 10, 10 \leq sum_2 < 100, sum_2 \geq 100\} \\
A / R_3 &= \{sum_3 < 10, 10 \leq sum_3 < 100, sum_3 \geq 100\}
\end{aligned}$$

所有微博都根据持续时间划分进入某个 R_i 大类, 再根据 sum 大小被划分进某个子类。最后取微博是否为源头微博为决策属性, 通过对比样本数据 52 条源头微博中不同属性与决策属性之间的依赖度, 确定 TW 定义如下。

$$TW = \begin{cases} 0.39, & h < 2\text{Days} \\ 0.33, & 2\text{Days} \leq h < 4\text{Days} \\ 0.18, & 4\text{Days} \leq h < 7\text{Days} \\ 0.10, & h \geq 7\text{Days} \end{cases} \quad (13)$$

从式(13)可以看出, 近期发布的微博权重较高, 其关注度造成的影响较大, 而较早以前发布的微博权重较小, 其关注度造成的影响较小。

3.4 时间

微博发布时间是判定微博源头的重要指标之一, 发布时间越早的微博成为微博源头的概率越大。

微博发布时间可以转化为相对应的持续时间 (以小时为单位), 本文所提算法所有的微博持续时间需要进行标准化, 本文使用的标准化函数如下。

$$T = \frac{\text{time}}{\text{time}_{\max}} \quad (14)$$

其中, time 表示微博的持续时间, $\text{time}_{\max}$ 表示所有溯源微博中最大的 time 值。

3.5 微博综合得分计算

ITM 算法综合考虑博主影响力, 评论人、转发人的影响力、关注度、时间, 利用 Hacker News 排序算法对每条微博进行评分, 根据得分进行排序, 将得分最高的 N 条微博 (N 是根据专家经验预设的可能源头数量) 输出为源头微博^[22], 微博得分算法如下所示。

算法 4 Blog Score Calculation

1) Input: A Topic blog Blog_k 's user influence BloggerInf , comment influence CommentInf , user attraction Attr_a , blog Standardized time T ;

2) Output: Blog score Score ;

3) Calculate blog score: $\text{Score} = \text{hacker news}(\text{BloggerInf}, \text{CommentInf}, \text{Attr}, T)$

4) return Score

Hacker News 排序算法是一种网络社区平台的文章排序算法, 对于每篇文章, 其他用户可以选择为这篇文章投上一票 (其他用户阅读、评论文章会为文章投上一票), 后台管理器会根据文章的得票和发布时间对所有文章进行评分排序。但是该排序算法会将发布时间较晚的文章排在较前面, 从而对新的文章进行推广, 防止得分很高的文章霸榜。Hacker News 排序算法如式(15)所示。

$$\text{Score} = \frac{P-1}{(T+2)^G} \quad (15)$$

其中, P 表示文章投票数, 票数越多值越大; T 为文章的持续时间, T 越大, 得分越小; G 为重力因子常数, 决定了得分随时间增长的速度, G 越大, 得分增长越慢。

在式(15)中, 较晚发布的信息得分会高于早发布的信息, 然而在微博溯源中, 较早发布的信息影响力一般越大。通过对 Hacker News 算法公式进行优化, 使发布时间较早、得票较高的文章排名靠前, 优化后的微博评分公式如下。

$$\text{Score} = \frac{\text{Inf}-1}{\left(\frac{1}{T}+2\right)^G} \quad (16)$$

其中, T 为微博发布到现在的持续时间, 由式(14)计算, T 越大, 得分越高; G 为重力因子常数; Inf 表示微博综合影响力, 且 Inf 越大, 微博得分越高。 Inf 由博主影响力, 评论人、转发人影响力以及关注度计算, 本文使用综合影响力 Inf 替代投票数 P 来更准确地反映微博对于网络群众造成的影响, 在微博平台上, 一条微博主要是通过自己的粉丝以及评论、转发和自身关注度来扩散微博信息, 从而造成相应的影响, 所以利用 Inf 来替代 P 具有一定合理性。微博综合影响力 Inf 的计算公式如下。

$$\text{Inf} = k_1 \cdot \text{BloggerInf} + k_2 \cdot \text{CommentInf} + k_3 \cdot \text{Attr} \quad (17)$$

其中, k_1 、 k_2 、 k_3 分别为各项权重, BloggerInf 为博主影响力, 由式(7)计算, CommentInf 为评论人、转发人影响力, 由式(8)计算, Attr 为博

主关注度, 由式(12)计算。

式(17)中各项得分的权重 k_1 、 k_2 、 k_3 可采用 AHP (层次分析法) 得出相应的参数。AHP 是将与决策有关的元素分解成目标、准则、方案等层次, 在此基础上进行定性和定量分析的决策方法。

AHP 通过对比两两参数之间相对重要性, 确定某个参数相对另一参数所占的比重, 从而构建判断矩阵, 之后计算其最大特征向量, 以该特征向量代表各个参数的权重。最后通过判断矩阵的特征值计算一致性比率进行一致性检验, 如果一致性比率小于阈值, 则认为计算出来的权重值是较为合理的。

综合各项因素, 通过专家评判得到判断矩阵, 如表 1 所示。

表 1 判断矩阵
Table 1 The judgement matrix of AHP

	k_1 (博主影响力)	k_2 (评论人、转发人影响力)	k_3 (关注度)
k_1 (博主影响力)	1.00	2.00	2.00
k_2 (评论人、转发人影响力)	0.50	1.00	2.00
k_3 (关注度)	0.50	0.50	1.00

计算得到最大特征向量, 确定各个权重值为 $k_1=0.493$, $k_2=0.311$, $k_3=0.363$, 计算一致性比率为 5.6%, 小于阈值 10%, 所以计算所得权重值较为合理。

在计算每条微博的得分之后将微博按照得分进行排序, 把得分最高的 N 条微博返回作为源头微博。

4 实验结果

实验选用新浪微博作为数据平台, 新浪微博作为现今为止国内用户最多的社交平台之一, 其数据具有较好的代表性。利用爬虫软件, 从微博平台采集“李飞飞离职”“高锟去世”“美国重启登月计划”“雷克萨斯将实现国产”和“国科大星命名”5 个事件的微博数据用于测试本文所提算法 ITM, 采集时间为 2018 年 9 月 11 日到 10 月 12 日。实验采集的微博共计 12 110 条, 包含具有评论、转发或点赞的微博以及相关的博主信息、

评论人信息,过滤了评论、转发和点赞都为0的微博,这主要考虑到这些微博没有对信息传播造成实质的影响。同时使用文献[1]所提出的基于文本中心度的微博文本溯源(OR)算法以及文献[2]所提出的基于用户影响力的微博溯源算法(UITA)进行对比。之所以使用OR算法和UITA算法作为对比,是因为它们分别是通过拓扑结构和参数计算来进行溯源的方法(已有的两类方法),具有一定代表性,且算法效果较好。使用人工标注的方法对微博的源头进行标注,对比两种算法找到源头正确的个数,并计算相应的查全率。查全率是指从文档数据集中检索出关联文档成功率的一种指标,即检索出来的关联文档占所有关联文档数据的百分比,实验采用的查全率计算公式如下。

$$\text{RecallRatio} = \frac{\text{TracedBlogs}}{\text{SourceBlogs}} \times 100\% \quad (18)$$

其中, TracedBlogs为成功检索的源头微博数量, SourceBlogs为所有源头微博的数量。

同时查看两种算法的溯源结果在相应话题下对应的热门微博个数,热门微博是微博官方在搜索提示中标记的热度较大的微博,并且往往对公众造成较大的影响,源头微博往往是热门微博之一,因此溯源结果中的热门微博个数越多,其准确性越高。

实验参数设置以及结果如表2和表3所示。

从表3可以看出,总体来说本文所提ITM算法在溯源微博正确个数上优于OR算法以及UITA算法,并在查全率上分别提升了17.9%和25%,平均提升了21.45%。在事件1李飞飞离职和事件5国科大星命名等事件上,OR算法以及UITA算法没有考虑到受众具有明显的兴趣倾向,

所以本文所提ITM算法在溯源准确性上有一定提升。在事件3美国重启登月计划上,人工标注源头微博中影响力较大的几个微博的文本内容非常相似,而ITM算法和基于文本相似中心性的OR算法的溯源正确数相同,所以即使是在微博文本相似度较大的情况下,ITM算法也能保证溯源的准确性。

表2 实验参数设置
Table 2 Experimental parameter settings

参数	公式	值
$\mu_1, \mu_2, \mu_3, \mu_4$	3	0.5, 0.2, 0.17, 0.13
γ	4	1.6
β	6	1.2
λ	10	1.2
G	16	1.4
k_1, k_2, k_3	17	0.493, 0.311, 0.363

在热门微博个数方面,由于OR算法只考虑文本的相似度和中心度,并未考虑微博在话题各方面的影响,所以ITM算法相对于OR算法准确性较高。UITA算法只是单纯基于微博参数来计算影响力,未考虑微博文本的影响,所以相对于UITA而言,ITM算法的准确性较高。

选择上述的事件1李飞飞离职,ITM算法的微博溯源结果以及微博基本信息如图3和表4所示(为便于分析,将微博得分换算为百分制)。

从图3可以看出,每日经济新闻和新浪科技等博主的微博得分较高,根据表4中的信息,发现得分较高微博发布时间较早,在点赞、评论、转发数量上相对较多,并且拥有较多的粉丝数量,通过对比人工标注的源头和话题热门微博,发现溯源结果较为准确。全景网发布的微博虽然有较多关注,但由于发布时间较晚所以得分不高,结果

表3 实验结果
Table 3 Experiment results

微博事件	标注源头数 N	准确数			查全率			热门微博个数		
		ITM	OR	UITA	ITM	OR	UITA	ITM	OR	UITA
1	7	6	4	4	85%	57%	57%	5	4	4
2	6	5	4	4	83%	66.7%	66.7%	5	4	3
3	5	3	3	2	60%	60%	40%	3	3	2
4	7	5	4	3	71.4%	57.5%	42.9%	4	4	3
5	3	2	1	1	66.7%	33.3%	33.3%	2	1	0
总计	28	21	16	14	75%	57.1%	50%	19	16	12

表 4 李飞飞离职事件微博基本信息
Table 4 Blog information about Li Feifei's resignation event

博主	发表时间	点赞数	评论数	转发数	粉丝数
每日经济新闻	9.12 08:10	135	37	38	3 481 万
新浪科技	9.11 05:51	134	29	55	1 510 万
第一财经日报	9.11 15:19	42	15	33	1 067 万
全景网	9.12 12:30	108	23	29	114 万
电商行业	9.20 08:30	202	28	92	264 万
财新网	9.12 11:10	27	27	13	383 万
中国妇女报	9.11 15:10	51	2	14	116 万
胡舒立	9.12 12:17	29	16	12	515 万
财经杂志	9.12 10:24	20	6	21	213 万
爱肉丸的兔纸	9.15 12:15	39	11	8	5 万

符合实际情况。中国妇女报发布的微博虽然时间比较早,但由于没有收到太多的关注并且粉丝数较少所以造成的影响比较小,结果符合实际情况。

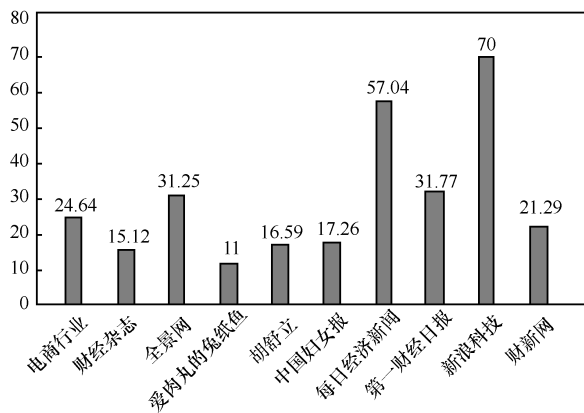


图 3 李飞飞离职事件微博的得分
Figure 3 Scores of blogs for Li Feifei's resignation event

5 结束语

本文主要研究微博平台中的信息溯源问题,并提出一种 ITM 算法对信息进行溯源。ITM 算法利用用户兴趣计算博主影响力以及评论人、转发人影响力,同时结合关注度和时间参数对所有微博进行评分,从而对微博进行排序溯源。相比以前基于文本相似度和转发关系的溯源算法,ITM 算法考虑到了微博文本以及其他用户的评论转发对信息传播带来的影响。实验结果表明,ITM 算法在溯源准确率上相对传统溯源算法有一定的提升。

下一步的研究将主要着眼于信息跨平台传播的情况并且考虑微博中的图片内容来提升溯源的准确性。

参考文献:

- [1] 时国华. 微博信息溯源及传播面分析技术的研究与实现[D]. 长沙: 国防科学技术大学, 2012.
SHI G H. The research and implementation of microblogging initiator detection and dissemination analysis[D]. Changsha: National University of Defense Technology, 2012.
- [2] 刘荣叁, 张宇, 王星. 面向新浪微博的信息溯源技术研究[J]. 智能计算机与应用, 2017, 7(2):94-98.
LIU R S, ZHANG Y, WANG X. Research of information trace technology based on Sina micro-blog[J]. Intelligent Computer and Applications, 2017, 7(2):94-98.
- [3] 米昂. 结合影响力分析的微博舆情溯源研究[D]. 北京: 北京交通大学, 2015.
MI A. Research on source tracing of public opinion on micro-blogs combined with impact analysis[D]. Beijing: Beijing Jiaotong University, 2015.
- [4] 张跃伟. 基于微博客话题的热点预测及传播溯源[D]. 北京: 北京邮电大学, 2014.
ZHANG Y W. Hotspot prediction and analysis of propagating of topics based on microblog[D]. Beijing: Beijing University of Posts and Telecommunications, 2014.
- [5] 刘岗. 基于微博事件的话题溯源方法[D]. 哈尔滨: 哈尔滨工程大学, 2015.
LIU G. A method to track topics based on events of micro-blog[D]. Harbin: Harbin Engineering University, 2015.
- [6] WANG D, ZHOU W, ZHENG J X, et al. Who spread to whom? inferring online social networks with user features[C]//2018 IEEE International Conference on Communications (ICC). 2018: 1-6.
- [7] 吴信东, 李毅, 李磊. 在线社交网络影响力分析[J]. 计算机学报, 2014, 37(4): 735-752.
WU X D, LI Y, LI L. Influence analysis of online social networks[J]. Chinese Journal of Computers, 2014, 37(4): 735-752.
- [8] SADIKOV E, MONTSERRAT M, JURE L et al. Correcting for

- missing data in information cascades[C]// WSDM '11 Proceedings of the fourth ACM International Conference on Web Search and Data Mining. 2011:55-64.
- [9] 杨静, 董圆, 张健沛. 一种基于话题影响力的微博话题溯源方法[J]. 小型微型计算机系统, 2015, 36(9):1939-1942.
YANG J, DONG Y, ZHANG J P. Method for tracing the source of micro-blogging topic based on the topic influence[J]. Journal of Chinese Mini-Micro Computer Systems, 2015, 36(9):1939-1942.
- [10] 董圆. 一种基于话题影响力的微博话题溯源方法[D]. 哈尔滨: 哈尔滨工程大学, 2015.
DONG Y. Trace of microblogging topic initiator based on the influence of topic[D]. Harbin: Harbin Engineering University, 2015.
- [11] 尹熙成. 在线社交网络中信息传播主路径的识别与应用研究[D]. 南京: 南京邮电大学, 2017.
YIN X C. Research on the identification and application of the main paths of information diffusion in online social networks[D]. Nanjing: Nanjing University of Posts and Telecommunications, 2017.
- [12] LESKOVEC J, MCGLOHON M, FALOUTSOS C, et al. Patterns of cascading behavior in large blog graphs[C]//Proceedings of the 2007 SIAM International Conference on Data Mining. 2007: 551-556.
- [13] 何冰心. 社交网络中信息传播和溯源模型研究[D]. 大连: 大连理工大学, 2018.
HE B X. A study on information diffusion and source locating model in social network[D]. Dalian: Dalian University of Technology, 2018.
- [14] 王梦迪. 基于 OSN 的信息溯源问题研究[D]. 北京: 中国人民公安大学, 2018.
WANG M D. Research on information tracing based on OSN[D]. Beijing: People's Public Security University of China, 2018.
- [15] XIAO Y, YU H, LI Q, et al. MPURank: a social hotspot tracking scheme based on tripartite graph and multimessages iterative driven[J]. IEEE Transactions on Computational Social Systems, 2019, 6(4): 715-725.
- [16] FEIZI S, MEDARD M, QUON G, et al. Network infusion to infer information sources in networks[J]. IEEE Transactions on Network Science and Engineering, 2018.
- [17] INUI T, MASAKI S, MIKIO Y. Automatic news source detection in twitter based on text segmentation[C]//Proceedings of the 28th Pacific Asia Conference on Language, Information and Computing. 2014.
- [18] 葛斌, 李芳芳, 郭丝路. 基于知网的词汇语义相似度计算方法研究[J]. 计算机应用研究, 2010, 27(9):3329-3333.
GE B, LI F F, GUO S L. Word's semantic similarity computation method based on hownet[J]. Application Research of Computers, 2010, 27(9):3329-3333.
- [19] 刘群, 李素建. 基于知网的词汇语义相似度计算[J]. 中文计算语言学, 2002, 7(2): 59-76.
LIU Q, LI S J. Word similarity computing based on how-net[J]. International Journal of Computational Linguistics & Chinese Language Processing, 2002, 7(2): 59-76.
- [20] 周雪妍. 在线社会网络关键用户挖掘方法研究[D]. 哈尔滨: 哈尔滨工程大学, 2016.
ZHOU X Y. Research on key users mining of online social network[D]. Harbin: Harbin Engineering University, 2016.
- [21] 曹秀英, 梁静国. 基于粗集理论的属性权重确定方法[J]. 中国管理科学, 2002, 5(5): 98-100.
CAO X Y, LIANG J G. The method of ascertaining attribute weight based on rough sets theory[J]. Chinese Journal of Management Science, 2002, 5(5): 98-100.
- [22] SALIHEFENDIC A. How hacker news ranking algorithm works[EB].

[作者简介]



杨潇 (1993-), 男, 重庆人, 上海交通大学硕士生, 主要研究方向为社交网络数据分析、自然语言处理。



陈秀真 (1977-), 女, 山东聊城人, 博士, 上海交通大学副教授, 主要研究方向为社交网络数据分析、车联网信息安全、安全检测与评估。



马进 (1977-), 女, 山东滕州人, 博士, 上海交通大学高级工程师, 主要研究方向为大数据与人工智能应用、车联网信息安全、网络空间安全综合管理新技术。



梁浩喆 (1994-), 男, 广西柳州人, 上海交通大学硕士生, 主要研究方向为浏览器安全、分布式计算。



李生红 (1971-), 男, 辽宁绥中人, 博士, 上海交通大学教授、博士生导师, 主要研究方向为信息安全、人工智能。

网络与信息安全学报

2020年(第6卷)总目次

视 角

	期	页码
新形势下理工门类学科建设内涵与学科评估导向	李凤华, 李 晖	1 [2020014] (1)

综 述

网络层匿名通信协议综述	王良民, 倪晓铃, 赵 蕙	1 [2020006] (11)
深度学习中的对抗样本的构造及防御研究		
.....	段广晗, 马春光, 宋 蕾, 武 朋	2 [2020016] (1)
路径分支混淆研究综述	耿 普, 祝跃飞	2 [2020027] (12)
智能合约安全综述	孟博, 刘加兵, 刘琴, 王潇潇, 郑旭睿, 王德军	3 [2020030] (1)
基于深度学习的文本分类研究进展	杜思佳, 于海宁, 张宏莉	4 [2020010] (1)
网络钓鱼识别研究综述	付 溪, 李 晖, 赵兴文	5 [2020062] (1)

专 栏

面向跨域可信的泛中心化区块链 DNS 架构研究		
.....	雷 凯, 束方兴, 黄 磊, 章奇超	2 [2020024] (19)
ContractGuard: 面向以太坊区块链智能合约的入侵检测系统		
.....	赵淦森, 谢智健, 王欣明, 何嘉浩, 张成志, 林成创,	
	Ziheng Zhou, 陈冰川, Chunming Rong	2 [2020025] (35)
基于比特币区块链的公共无线局域网接入控制隐私保护研究		
.....	牛玉坤, 魏凌波, 张 驰, 张 霞, Gustavo Vejarano	2 [2020022] (56)
基于区块链的零知识位置证明系统设计	吴 炜, 刘儿兀, 杨昌鑫, 王 睿	2 [2020020] (67)
基于可信联盟的 P2P 网络身份认证机制	陈盈盈, 章 峰, 蒋文保, 周 旭	2 [2020021] (77)
基于区块链的分布式电能数据可信存储机制		
.....	李 瑾, 仵松颀, 张森林, 陆月明	2 [2020029] (87)
基于差分隐私保护技术的多方求和查询方法	何贤芒	3 [2020032] (14)
按需披露的区块链隐私保护机制	李少卓, 王 娜, 杜学绘	3 [2020033] (19)

基于椭圆曲线的 ElGamal 型位置隐私保护方案			
..... 赖慧彬, 许 力, 曾雅丽, 朱 瑞	3	[2020031]	(30)
基于差分隐私的非等距直方图发布方法	杨 磊, 郑 啸, 赵 伟	3	[2020035] (39)
基于区块链的轻量化移动自组网认证方案	张 勘, 马 欣	4	[2020040] (14)
高效安全的可审计盲混币服务方案	乔 康, 汤红波, 游 伟, 李海涛	4	[2020043] (23)
基于改进 PBFT 算法的 PKI 跨域认证方案	钱思杰, 陈立全, 王诗卉	4	[2020042] (37)
基于区块链的 5G 物联网数据共享方案	乔 康, 游 伟, 王领伟, 汤红波	4	[2020041] (45)
基于智能合约的多微电网市场化交易模型及算法			
..... 李 芬, 李 瑾, 仵松颀, 张森林, 陆月明	4	[2020023]	(56)
基于日志的异常检测技术综述	张颖君, 刘尚奇, 杨 牧, 张海霞, 黄克振	6	[2020072] (1)
基于融合编译的软件多样化保护方法	熊小兵, 舒 辉, 康 绯	6	[2020075] (13)
分支混淆中的条件异常代码构造研究	耿 普, 祝跃飞	6	[2020061] (25)
基于卷积神经网络恶意安卓应用行为模式挖掘			
..... 张 鑫, 羌卫中, 吴月明, 邹德清, 金 海	6	[2020073]	(35)
Docker 组件间标准输入输出复制的 DoS 攻击分析			
..... 周天昱, 申文博, 杨男子, 李金库, 秦承刚, 喻 望	6	[2020074]	(45)
基于代码碎片化的软件保护技术	郭京城, 舒 辉, 熊小兵, 康 绯	6	[2020063] (57)
电子邮件安全扩展协议应用分析	尚菁菁, 朱宇佳, 刘庆云	6	[2020083] (69)

学术论文

基于通信特征的 CAN 总线泛洪攻击检测方法			
季一木, 焦志鹏, 刘尚东, 吴 飞, 孙 静, 王 娜, 陈治宇, 毕 强, 田鹏浩	1	[2020005]	(27)
基于边界值不变量的对抗样本检测方法	严 飞, 张铭伦, 张立强	1	[2020012] (38)
基于同态加密和区块链技术的车联网隐私保护方案			
..... 王瑞锦, 唐榆程, 张巍琦, 张凤荔	1	[2020011]	(46)
基于注意力机制的社交垃圾文本检测方法	曲 强, 于洪涛, 黄瑞阳	1	[2020002] (54)
银行账户交易网络中特定组织发现研究			
..... 吕 芳, 卢西婧, 王 巍, 黄俊恒, 王佰玲	1	[2020001]	(62)
基于 FPGA 的软件定义协议无关解析器	苗力心, 刘勤让, 汪 欣	1	[2020013] (70)
基于机器学习的 TLS 恶意加密流量检测方案	骆子铭, 许书彬, 刘晓东	1	[2020008] (77)
融合多特征的视频帧间篡改检测算法			
..... 肖 辉, 翁 彬, 黄添强, 普 菡, 黄则辉	1	[2020007]	(84)
基于流量分析的软件升级漏洞自动检测方法			
..... 腾金辉, 光 焱, 舒 辉, 张 冰	1	[2020004]	(94)

基于隔离等级的网络切片部署方法	潘启润, 黄开枝, 游 伟	2	[2020003] (96)
基于 AES 轮函数认证加密算法研究与设计	高国强, 李子臣	2	[2020028](106)
基于卷积神经网络的多尺度 Logo 检测算法	江玉朝, 吉立新, 高 超, 李邵梅	2	[2020026](117)
基于区块链的众包物流分级多层智能服务交易监管架构	余春堂, 韩志耕, 李致远, 王良民	3	[2020017] (50)
数字隐写模型及形式化描述	宋海涛, 汤光明, 孙怡峰, 王 洋	3	[2020036] (59)
基于深度学习的加密恶意流量检测研究	翟明芳, 张兴明, 赵 博	3	[2020034] (66)
基于区块链的智能网联车队协同轨迹预测系统	王仁生, 吴天昊, 张 琳, 朱孔林	3	[2020015] (78)
基于 CP-ABE 算法的区块链数据访问控制方案	邱云翔, 张红霞, 曹 琪, 章建聪, 陈兴蜀, 金泓键	3	[2020037] (88)
无须重注册的单向通信动态口令认证	鲁迁迁, 朱友文, 蒋 炎	3	[2020038] (99)
几个无证书签名方案的伪造攻击	王 菁, 李祖猛	3	[2020039](108)
Method for reducing cloud workflow completion time under the task interruption	Yawen WANG, Yunfei GUO, Wenyan LIU, Shumin HUO	3	[2020018](113)
敌对攻击环境下基于移动目标防御的算法稳健性增强方法	何 康, 祝跃飞, 刘 龙, 芦 斌, 刘 彬	4	[2020052] (67)
MLAR: 面向 IP 定位的大规模网络别名解析	袁福祥, 刘粉林, 刘 翀, 刘 琰, 罗向阳	4	[2020044] (77)
PPC 和 MIPS 指令集下二进制代码中函数参数个数的识别方法	尹小康, 刘 鑒, 刘 龙, 刘胜利	4	[2020047] (95)
基于 DeepLink 的社交网络去匿名方法	王 培, 贾 焰, 李爱平, 蒋千越	4	[2020045](104)
基于生成对抗网络的文本序列数据集脱敏	张 煜, 吕锡香, 邹宇聪, 李一戈	4	[2020046](109)
轻量级门限群签名方案的改进	竺智荣, 吕佳玉, 叶帼华, 姚志强	4	[2020048](120)
面向深度神经网络的安全计算协议设计方法	毕仁万, 陈前昕, 熊金波, 刘西蒙	4	[2020050](130)
灰度可视秘密共享方案分析与设计	刘 健, 张 益, 刘玉岭, 胡 浩	4	[2020049](140)
基于自对偶量子低密度校验码的量子对话协议	刘 芬, 贺振兴, 徐鹏翱, 马鸿洋	4	[2020051](148)
湖南大学信息安全专业本科培养计划修订调查分析及改进思路	胡 桥, 胡玉鹏, 廖鑫, 秦 拯, 孙建华, 张吉良	4	[2020053](153)
区块链存储: 技术与挑战	蔡振华, 林嘉韵, 刘 芳	5	[2020019] (11)
基于刚性内存的区块链协议改进	程 穗, 林宪正, 俞能海	5	[2020060] (21)

交叉口场景下基于区块链技术的匿名车辆身份认证方案		
..... 王 昊, 吴天昊, 朱孔林, 张 琳	5	[2020057] (27)
深度学习中的对抗攻击与防御	刘西蒙, 谢乐辉, 王耀鹏, 李旭如	5 [2020071] (36)
高安全属性价值设备威胁态势量化评估方法		
..... 韩 磊, 刘吉强, 王 健, 石 波, 和旭东	5	[2020070] (54)
基于深度神经网络的 Android 恶意软件检测方法		
..... 超 凡, 杨 智, 杜学绘, 孙 彦	5	[2020060] (67)
基于 LSTM-KF 模型的无人机抗 GPS 欺骗方法		
..... 孙 旸, 曹春杰, 赖俊晓, 于天娇	5	[2020069] (80)
基于广义共同邻居的有向网络链路预测方法		
..... 赵学磊, 季新生, 刘树新, 赵 宇	5	[2020059] (89)
基于 HRCA 的可重构 SM4 密码算法研究与实现	张 骁, 周清雷, 李 斌	5 [2020064] (101)
基于 Polar 码改进的 RLCE 公钥加密方案	李 喆, 韩益亮, 李 鱼	5 [2020068] (110)
基于强化学习的低时延车联网群密钥分配管理技术		
..... 徐堂炜, 张海璐, 刘楚环, 肖 亮, 朱珍民	5	[2020055] (119)
基于残差空洞卷积神经网络的网络安全实体识别方法		
..... 谢 博, 申国伟, 郭 春, 周 燕, 于 淼	5	[2020009] (126)
基于软件定义的多协议控制器架构设计		
..... 夏云飞, 张 丽, 杨 堃, 李沛杰, 许立明	5	[2020065] (139)
广域网出口流量调度 SDN 部署研究		
..... 李国春, 马 睿, 马季春, 李伯中, 刘惠明, 张桂玉	5	[2020067] (148)
政府开放数据的国家安全风险评估模型研究		
..... 王 标, 刘兴洋, 许 卡, 刘汪洋, 王克楠, 夏宇清	6	[2020077] (80)
基于跨域协同的网络空间威胁预警模式	熊 钢, 葛雨玮, 褚衍杰, 曹卫权	6 [2020078] (88)
低故障恢复开销的软件定义网络控制器布局算法	吴 奇, 陈鸿昶	6 [2020066] (97)
基于 TPCM 的服务器可信 PXE 启动方法	刘国杰, 张建标	6 [2020079] (105)
基于可分离卷积的轻量级恶意域名检测模型		
..... 杨路辉, 白惠文, 刘光杰, 戴跃伟	6	[2020084] (112)
基于多层模糊综合评估的隐私保护效果评估方法	俞艺涵, 付 钰, 吴晓平	6 [2020082] (121)
三接收方公钥密码系统	赵琉涛, 钟 林, 刘吉东, 王彩群, 吴 丹	6 [2020080] (128)
移动平台典型应用的身份认证问题研究	张效林, 谷大武, 张 驰	6 [2020081] (137)
基于循环神经网络的空载电动出租车的充电桩推荐方法		
..... 贾 鉴, 刘林峰, 吴家皋	6	[2020085] (152)
基于用户兴趣的微博溯源算法	杨 潇, 陈秀真, 马 进, 梁浩喆, 李生红	6 [2020086] (164)

Chinese Journal of Network and Information Security

Contents List Vol.6 (2020)

Perspective

	No.	Page
Construction of academic discipline of science and engineering and orientation for their ranking under the new situation..... <i>LI Fenghua, LI Hui</i>	1	[2020014] (1)

Comprehensive Reviews

Survey of network-layer anonymous communication protocols..... <i>WANG Liangmin, NI Xiaoling, ZHAO Hui</i>	1	[2020006] (11)
Research on structure and defense of adversarial example in deep learning <i>DUAN Guanghan, MA Chunguang, SONG Lei, WU Peng</i>	2	[2020016] (1)
Review of path branch obfuscation..... <i>GENG Pu, ZHU Yuefei</i>	2	[2020027] (12)
Survey of smart contract security..... <i>MENG Bo, LIU Jiabing, LIU Qin, WANG Xiaoxiao, ZHENG Xurui, WANG Dejun</i>	3	[2020030] (1)
Survey of text classification methods based on deep learning <i>DU Sijia, YU Haining, ZHANG Hongli</i>	4	[2020010] (1)
Survey on phishing detection research..... <i>FU Xi, LI Hui, ZHAO Xingwen</i>	5	[2020062] (1)

Special Columns

Research on cross-domain trustable blockchain based decentralized DNS architecture <i>LEI Kai, SHU Fangxing, HUANG Lei, ZHANG Qichao</i>	2	[2020024] (19)
ContractGuard: defend Ethereum smart contract with embedded intrusion detection..... <i>ZHAO Gansen, XIE Zhijian, WANG Xinming, HE Jiahao, ZHANG Chengzhi, LIN Chengchuang, Ziheng ZHOU, CHEN Bingchuan, Chunming RONG</i>	2	[2020025] (35)
Privacy-preserving access control for public wireless LAN utilizing the bitcoin blockchain <i>NIU Yukun, WEI Lingbo, ZHANG Chi, ZHANG Xia, Gustavo Vejarano</i>	2	[2020022] (56)
Design of blockchain based zero-knowledge proof of location system <i>WU Wei, LIU Erwu, YANG Changxin, WANG Rui</i>	2	[2020020] (67)
P2P network identity authentication mechanism based on trusted alliance <i>CHEN Yingying, ZHANG Feng, JIANG Wenbao, ZHOU Xu</i>	2	[2020021] (77)
Trusted storage mechanism of distributed electric energy data based on blockchain <i>LI Jin, WU Songqi, ZHANG Senlin, LU Yueming</i>	2	[2020029] (87)
Multi-party summation query method based on differential privacy <i>HE Xianmang</i>	3	[2020032] (14)
Privacy protection mechanism of on-demand disclosure on blockchain <i>LI Shaozhuo, WANG Na, DU Xuehui</i>	3	[2020033] (19)
Elliptic curve ElGamal encryption based location privacy-preserving scheme <i>LAI Huibin, XU Li, ZENG Yali, ZHU Rui</i>	3	[2020031] (30)
Non-equal-width histogram publishing method based on differential privacy <i>YANG Lei, ZHENG Xiao, ZHAO Wei</i>	3	[2020035] (39)
Lightweight mobile Ad Hoc network authentication scheme based on blockchain <i>ZHANG Xu, MA Xin</i>	4	[2020040] (14)

Efficient and safe auditable mixed-coin service scheme based on blind signature	<i>QIAO Kang, TANG Hongbo, YOU Wei, LI Haitao</i>	4	[2020043] (23)
PKI cross-domain authentication scheme based on advanced PBFT algorithm	<i>QIAN Sijie, CHEN Liquan, WANG Shihui</i>	4	[2020042] (37)
Data sharing scheme for 5G IoT based on blockchain	<i>QIAO Kang, YOU Wei, WANG Lingwei, TANG Hongbo</i>	4	[2020041] (45)
Multi-microgrid marketized transaction models and algorithms based on smart- contracts	<i>LI Fen, LI Jin, WU Songqi, ZHANG Senlin, LU Yueming</i>	4	[2020023] (56)
Survey on anomaly detection technology based on logs	<i>ZHANG Yingjun, LIU Shangqi, YANG Mu, ZHANG Haixia, HUANG Kezhen</i>	6	[2020072] (1)
Method of diversity software protection based on fusion compilation	<i>XIONG Xiaobing, SHU Hui, KANG Fei</i>	6	[2020075] (13)
Research on construction of conditional exception code used in branch obfuscation	<i>GENG Pu, ZHU Yuefei</i>	6	[2020061] (25)
Mining behavior pattern of mobile malware with convolutional neural network	<i>ZHANG Xin, QIANG Weizhong, WU Yueming, ZOU Deqing, JIN Hai</i>	6	[2020073] (35)
Analysis of DoS attacks on Docker inter-component stdio copy	<i>ZHOU Tianyu, SHEN Wenbo, YANG Nanzi, LI Jinku, QIN Chenggang, YU Wang</i>	6	[2020074] (45)
Software protection technology based on code fragmentation	<i>GUO Jingcheng, SHU Hui, XIONG Xiaobing, KANG Fei</i>	6	[2020063] (57)
Analysis of security extension protocol in e-mail system	<i>SHANG Jingjing, ZHU Yujia, LIU Qingyun</i>	6	[2020083] (69)

Papers

CAN bus flood attack detection based on communication characteristics	<i>JI Yimu, JIAO Zhipeng, LIU Shangdong, WU Fei, SUN Jing, WANG Na, CHEN Zhiyu, BI Qiang, TIAN Penghao</i>	1	[2020005] (27)
Adversarial examples detection method based on boundary values invariants	<i>YAN Fei, ZHANG Minglun, ZHANG Liqiang</i>	1	[2020012] (38)
Privacy protection scheme for internet of vehicles based on homomorphic encryption and block chain technology	<i>WANG Ruijin, TANG Yucheng, ZHANG Weiqi, ZHANG Fengli</i>	1	[2020011] (46)
Attention-based approach of detecting spam in social networks	<i>QU Qiang, YU Hongtao, HUANG Ruiyang</i>	1	[2020002] (54)
Research on the method of discovering specific organization structure in bank account transaction network	<i>LYU Fang, LU Xijing, WANG Wei, HUANG Junheng, WANG Bailing</i>	1	[2020001] (62)
Software-defined protocol independent parser based on FPGA	<i>MIAO Lixin, LIU Qinrang, WANG Xin</i>	1	[2020013] (70)
Scheme for identifying malware traffic with TLS data based on machine learning	<i>LUO Ziming, XU Shubin, LIU Xiaodong</i>	1	[2020008] (77)
Video inter-frame tampering detection algorithm fusing multiple features	<i>XIAO Hui, WENG Bin, HUANG Tianqiang, PU Han, HUANG Zehui</i>	1	[2020007] (84)
Automatic detection method of software upgrade vulnerability based on network traffic analysis	<i>TENG Jinhui, GUANG Yan, SHU Hui, ZHANG Bing</i>	1	[2020004] (94)
Network slicing deployment method based on isolation level	<i>PAN Qirun, HUANG Kaizhi, YOU Wei</i>	2	[2020003] (96)

Research and design of authenticated encryption algorithm based on AES round function	GAO Guoqiang, LI Zichen	2	[2020028](106)
Multi-scale Logo detection algorithm based on convolutional neural network	JIANG Yuchao, JI Lixin, GAO Chao, LI Shaomei	2	[2020026](117)
Blockchain-based hierarchical and multi-level smart service transaction supervision framework for crowdsourcing logistics	YU Chuntang, HAN Zhigeng, LI Zhiyuan, WANG Liangmin	3	[2020017] (50)
Digital steganography model and its formal description	SONG Haitao, TANG Guangming, SUN Yifeng, WANG Yang	3	[2020036] (59)
Survey of encrypted malicious traffic detection based on deep learning	ZHAI Mingfang, ZHANG Xingming, ZHAO Bo	3	[2020034] (66)
Cooperative trajectory prediction system for intelligent and automated connected convoy based on blockchain	WANG Rensheng, WU Tianhao, ZHANG Lin, ZHU Konglin	3	[2020015] (78)
Blockchain data access control scheme based on CP-ABE algorithm	QIU Yunxiang, ZHANG Hongxia, CAO Qi, ZHANG Jiancong, CHEN Xingshu, JIN Hongjian	3	[2020037] (88)
One-time password authentication for one-way communication without re-registration	LU Qianqian, ZHU Youwen, JIANG Yan	3	[2020038] (99)
Forgery attacks on several certificateless signature schemes	WANG Jing, LI Zumeng	3	[2020039](108)
Method for reducing cloud workflow completion time under the task interruption	Yawen WANG, Yunfei GUO, Wenyan LIU, Shumin HUO	3	[2020018](113)
Improve the robustness of algorithm under adversarial environment by moving target defense	HE Kang, ZHU Yuefei, LIU Long, LU Bin, LIU Bin	4	[2020052] (67)
MLAR: large-scale network alias resolution for IP geolocation	YUAN Fuxiang, LIU Fenlin, LIU Chong, LIU Yan, LUO Xiangyang	4	[2020044] (77)
Function argument number identification stripped binary under PPC and MIPS instruction set	YIN Xiaokang, LIU Liu, LIU Long, LIU Shengli	4	[2020047] (95)
De-anonymization method for networks based on DeepLink	WANG Pei, JIA Yan, LI Aiping, JIANG Qianqian	4	[2020045](104)
Differentially private sequence generative adversarial networks for data privacy masking	ZHANG Yu, LYU Xixiang, ZOU Yucong, LI Yige	4	[2020046](109)
Improvement of lightweight threshold group signature scheme	ZHU Zhirong, LYU Jiayu, YE Guohua, YAO Zhiqiang	4	[2020048](120)
Design method of secure computing protocol for deep neural network	BI Renwan, CHEN Qianxin, XIONG Jinbo, LIU Ximeng	4	[2020050](130)
Analysis and design of visual secret sharing scheme for gray image	LIU Jian, ZHANG Yi, LIU Yuling, HU Hao	4	[2020049](140)
Quantum dialogue protocol based on self-dual quantum low density parity check codes	LIU Fen, HE Zhenxing, XU Pengao, MA Hongyang	4	[2020051](148)
Investigation, analysis and improvement mentality of the revision on information security's training program to undergraduates in Hunan University	HU Qiao, HU Yupeng, LIAO Xin, QIN Zheng, SUN Jianhua, ZHANG Jiliang	4	[2020053](153)
Blockchain storage: technologies and challenges	CAI Zhenhua, LIN Jiayun, LIU Fang	5	[2020019] (11)

Improvement of the blockchain protocol based on memory-hard function		
.....CHENG Sui, LIN Xianzheng, YU Nenghai	5	[2020060] (21)
Anonymous vehicle authentication scheme based on blockchain technology in the intersection scenario.....		
.....WANG Hao, WU Tianhao, ZHU Konglin, ZHANG Lin	5	[2020057] (27)
Adversarial attacks and defenses in deep learning		
.....LIU Ximeng, XIE Lehui, WANG Yaopeng, LI Xuru	5	[2020071] (36)
Quantitative threat situational assessment method for higher secure attribute value device		
.....HAN Lei, LIU Jiqiang, WANG Jian, SHI Bo, HE Xudong	5	[2020070] (54)
Android malware detection method based on deep neural network.....		
.....CHAO Fan, YANG Zhi, DU Xuehui, SUN Yan	5	[2020060] (67)
AntiGPS spoofing method for UAV based on LSTM-KF model.....		
.....SUN Yang, CAO Chunjie, LAI Junxiao, YU Tianjiao	5	[2020069] (80)
Link prediction methods based on generalized common neighbor in directed network		
.....ZHAO Xuelei, JI Xincheng, LIU Shuxin, ZHAO Yu	5	[2020059] (89)
Research and implementation of reconfigurable SM4 cryptographic algorithm based on HRCA		
.....ZHANG Xiao, ZHOU Qinglei, LI Bin	5	[2020064](101)
Improved RLCE public key encryption scheme based on Polar codes		
.....LI Zhe, HAN Yiliang, LI Yu	5	[2020068](110)
Reinforcement learning based group key agreement scheme with reduced latency for VANET.....		
.....XU Tangwei, ZHANG Hailu, LIU Chuhuan, XIAO Liang, ZHU Zhenmin	5	[2020055](119)
Cyber security entity recognition method based on residual dilation convolution neural network		
.....XIE Bo, SHEN Guowei, GUO Chun, ZHOU Yan, YU Miao	5	[2020009](126)
Architecture design of multi-protocol controller based on software definition		
.....XIA Yunfei, ZHANG Li, YANG Kun, LI Peijie, XU Liming	5	[2020065](139)
Research on SDN deployment practice for WAN egress traffic scheduling		
.....LI Guochun, MA Rui, MA Jichun, Li Bozhong, LIU Huiming, ZHANG Guiyu	5	[2020067](148)
Research on national security risk assessment model of open government data		
.....WANG Biao, LIU Xingyang, XU Ka, LIU Wangyang, WANG Kenan, XIA Yuqing	6	[2020077] (80)
Model of cyberspace threat early warning based on cross-domain and collaboration		
.....XIONG Gang, GE Yuwei, CHU Yanjie, CAO Weiquan	6	[2020078] (88)
Low failure recovery cost controller placement strategy in software defined networks		
.....WU Qi, CHEN Hongchang	6	[2020066] (97)
TPCM-based trusted PXE boot method for servers		
.....LIU Guojie, ZHANG Jianbiao	6	[2020079](105)
Lightweight malicious domain name detection model based on separable convolution		
.....YANG Luhui, BAI Huiwen, LIU Guangjie, DAI Yuewei	6	[2020084](112)
Evaluation method of privacy protection effect based on multi-layer fuzzy comprehensive evaluation		
.....YU Yihan, FU Yu, WU Xiaoping	6	[2020081](121)
Triple receiver public key encryption cryptosystem		
.....ZHAO Liutao, ZHONG Lin, LIU Jidong, WANG Caiqun, WU Dan	6	[2020082](128)
Issues of identity verification of typical applications over mobile terminal platform		
.....ZHANG Xiaolin, GU Dawu, ZHANG Chi	6	[2020080](137)
Charging pile recommendation method for idle electric taxis based on recurrent neural network.....		
.....JIA Jian, LIU Linfeng, WU Jiagao	6	[2020085](152)
User interests-based microblog tracing algorithm.....		
.....YANG Xiao, CHEN Xiuzhen, MA Jin, LIANG Haozhe, LI Shenghong	6	[2020086](164)

《网络与信息安全学报》

第二届编辑委员会

主 任 委 员：方滨兴

常务副主任委员：李风华

副 主 任 委 员：李 晖 俞能海 Kui Ren[USA] 刘建伟

邹德清 杜学绘 刘华鲁

委 员：（按姓氏笔画排序）

丁 勇	马建峰	王 标	王 滨	王布宏
王连海	王良民	王国胤	王明生	王美琴
牛 犇	仇洪冰	田有亮	田俊峰	史国振
付章杰	冯 涛	朱建明	延志伟	仲 盛
向 涛	刘 哲	刘 渊	刘西蒙	刘哲理
闫 峥	孙东红	杨宏宇	肖 甫	肖 亮
吴 铤	吴晓平	邱卫东	何道敬	沈 剑
沈 超	张小松	张云勇	张红旗	张宏莉
张建标	张振江	陆月明	陈立全	陈兴蜀
陈性元	陈雷霆	苗春雨	林 鹏	易东山
赵 波	赵琉涛	咸鹤群	禹 勇	姚志强
秦 拯	贾 焰	殷丽华	翁 健	黄欣沂
曹春杰	彭国军	蒋文保	韩伟力	韩劲松
程 光	阚海斌			

David Naccache [FRA] Gang Qu [USA] Jinjun Chen [AUS]

Willy Susilo [AUS] Yier Jin [USA]

收录声明

本刊对发表的文章,拥有出版电子版、网络版版权,并拥有和其他网站交换信息的权利。本刊支付的稿酬中已经包含上述费用。

Chinese Journal of Network and Information Security has the copyright to publish electronic edition, online edition of the published articles, and has the right to exchange information with other sites. The expenses have been included in the fee paid by editorial department.

道德声明

本刊发表的论文是作者独立取得的原创性研究成果,无一稿多投;论文内容不涉及国家机密;未曾以任何形式用任何文种在国内外公开发表过;论文内容不侵犯他人著作权和其他权利。若发生一稿多投、侵权、泄密等问题,论文作者将承担全部责任。

The authors of *Chinese Journal of Network and Information Security* guarantee that their submitted articles are original and contain nothing confidential. The said article is only submitted to *Chinese Journal of Network and Information Security*. The said article has not been published before and has not been submitted elsewhere for print or electronic publication consideration. The said article is no way whatever a violation or an infringement of any existing copyright or license from the third party. Otherwise, the authors of the said article shall take the blame for the violation or infringement of the related copyright and the leakage of secrets.

网络 Chinese Journal of
Network and Information Security
信息安全学报



2020年12月15日出版 定价：60.00元

ISSN 2096-109X



9 772096 109202

1.2