**清 华 大 学**

**综 合 论 文 训 练**

**题目：社交网络中的谣言检测系统**

|  |  |
| --- | --- |
| 系 别： | 软件学院 |
| 专 业： | 计算机软件 |
| 姓 名： | 钟仰新 |
| 指导教师： | 刘世霞 副教授 |

2016年 月 日

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签 名： 导师签名： 日 期：

中文摘要

近年来，社交网络（如推特、微博）被互联网用户广泛使用。但同时此类平台中存在大量不真实谣言，其传播造成的危害巨大。由于社交网络消息数目庞大，使得人工排查谣言成本很高，因此对谣言的自动检测技术具有重大意义。

本文描述了一个谣言检测系统，此系统首先对社交网络中的消息进行谣言特征匹配，检测出疑似谣言的消息；然后对这些候选消息进行聚类，生成疑似的谣言话题；接着对候选话题进行特征抽取，利用监督学习技术对其进行可疑度排名，最终检测出最有可能为谣言的话题。

在对消息进行聚类时，本文采取了基于相似度矩阵的聚类方法，研究了包括发布时间、消息标签、命名实体、文本单词、用户群体的相似度度量，聚类上采用了k均值、层级聚类以及谱聚类等方法，并对比了不同度量和聚类算法之间的优劣。实验表明，此系统能有效地将属于同一话题的不同消息归为一类，降低人工排查谣言的成本。

在对候选话题进行特征抽取时，本文总结相关工作，选取了45种不同的特征，并引入了名为过滤器和包装器的两种特征选择技术，对比两种技术并将其结合使用，最终选择出有效的特征子集。实验证明，经过选择的特征子集能有效提高系统对谣言的识别准确度。

在对话题进行可疑度排序时，系统选取了决策树和朴素贝叶斯两种分类器算法，并引入多分类器组合投票技术。实验表明，引入组合投票技术能有效提高排名系统的实用性，显著降低人工排查的成本。

**关键词**：社交网络；谣言检测；聚类；特征选择；监督学习

ABSTRACT

In recent years, social network (e.g., Twitter and Weibo) has been widely used on the Internet. But at the same time, there are a large number of false rumors in social network. Their spread may have harmful effect on individuals and society. The large amount of messages on social network make the cost of manual checking extremely high. As a result, automatic rumor detection techniques are of great significance.

This thesis describes a rumor detection system. It first filter out suspected rumor tweets by pattern matching, and then cluster these candidate tweets into topics. After that, it extract features of the candidate topics and rank them by the likelihood of false rumor through supervised learning technique. In this way, the most likely rumor topics can be detected automatically.

The system applies clustering techniques based on similarity matrix. The thesis studies similarity metrics that consider different factors including published time, hashtags, named-entities, terms and users. The system employs k-means, hierarchical and spectral clustering algorithms. Experiments demonstrate that the system can lower the cost of rumor checking by effectively clustering similar messages into same topics.

For feature extraction, the thesis chooses 45 different features and introduces two feature selection techniques, Filter and Wrapper, to select effective feature subset for the training of next step. A method of combining these two feature selection techniques have been proposed in this thesis. Experiments show that the new method and the step of feature selection can effectively improve the accuracy of rumor detection.

To rank and find most likely rumor topics, the system employs Decision Tree and Naive Bayes as classifiers and a multiple classifiers combining technique by voting has been introduced. Experiments demonstrate that the combining method can effectively improve practicability of the detection system.

**Keywords:** Social network; rumor detection; clustering; feature selection; supervised learning

目录

[第1章 引言 1](#_Toc452476646)

[1.1 社交网络 1](#_Toc452476647)

[1.2 社交网络中的谣言 1](#_Toc452476648)

[1.3 谣言检测 3](#_Toc452476649)

[插图索引 4](#_Toc452476650)

[表格索引 5](#_Toc452476651)

[参考文献 6](#_Toc452476652)

[致 谢 8](#_Toc452476653)

[声 明 9](#_Toc452476654)

[附录 A 外文资料的调研阅读报告 10](#_Toc452476655)

# 引言

## 社交网络

社交网络是近十年来新兴起的一类互联网社交平台，其在国外的代表有推特、脸书（Twitter, Facebook），在中国的代表有微博、人人网等。社交网络，顾名思义是将人们日常的社交活动，推广拓展到了互联网中，或者说是由互联网公司提供一个平台，给互联网用户提供在线交流、分享、交友等服务。

近年来，这类平台的理念、模式日趋成熟，在发展上取得了巨大成功，甚至逐步成为现代人生活的一部分。究其成功的原因，是这类平台将社交这种活动的门槛降低，同时拓展了其形式、主体，大大满足了人类喜爱社交的心理：任何人都能很简单地在此类平台上注册，然后通过简单的操作来发布消息、接收消息、推广消息，用户也能利用平台添加生活中认识的朋友，或者结识素未谋面的新朋友，通过将社交活动移植到互联网，社交网络成功让人们不出门就能完成社交；正是由于社交网络能通过很低的成本拉近人与人的距离，很多文体明星、政治名人（如歌坛巨星、美国总统）都在平台上注册，人们从此可以通过关注他们的账号，与这些原本在生活中很难相遇的名人进行交流、互动，而名人从此也能通过平台很容易地与自己的粉丝分享生活的点点滴滴、对社会事件的看法等等，为自己赢得更高的知名度和更多的人气；同时，社交网络的注册不仅限于个人，任何社会团体（如商业公司、公益组织）也都能注册账号，发布与自己相关的信息，通过平台进行宣传、推广，这就将社交的主体从人与人拓展到了人与团体，甚至团体与团体。

正是由于社交网络的壮大发展，使得这类平台中的信息规模以指数形式爆发增长，每天有越来越多的信息在不同的平台用户间流动、传播，产生各种影响和价值。因此这几年社交网络获得了学术界的广泛关注，越来越多的学者开始研究其中的信息流动模式，或发展相关技术来自动检测有价值的信息和用户，希望能通过研究更好地发挥社交网络的作用，使其产生正面的、更有价值的社会影响。

## 社交网络中的谣言

在现实的社交生活中，存在着一些不真实或真实性有待确定，但却被社交圈中的人们广泛讨论、传播的“小道消息”，学术界称之为“流言”或“谣言”(rumor)。与真正社交活动相似，社交网络中也存在大量的谣言，这些谣言一般分为两种：误传和虚假消息。

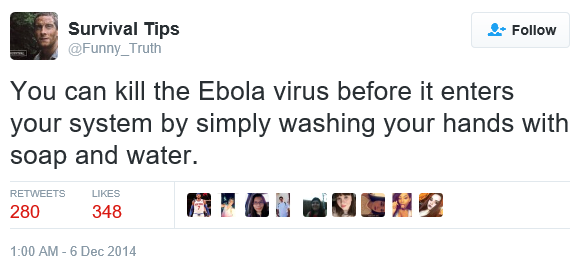
误传消息原本多为源头正规的真实消息，但却在信息的传播过程中被部分用户误解，导致消息在传播中逐渐走样、变形，成为不真实的消息。如图 1.1是推特中的一则误传消息，消息发布于埃博拉病毒盛行的2014年，其大意是我们可以通过用肥皂洗手，简单地杀死手上那些还未进入体内的埃博拉病毒。之所以说这是一则误传消息是因为消息本身并无恶意，是想鼓励更多的人勤洗手来预防埃博拉病毒，事实上这也是世界卫生组织一直提倡的；但是此消息的内容却不具有真实性，原因是用肥皂洗手可以减少手上附着的病毒数目，却不一定能简单地杀死埃博拉病毒（参考信息来源：http://www.killebolavirus.com/does-hand-sanitizer-kill-the-ebola-virus/），而世界卫生组织也从未提到一般的肥皂能有效杀死埃博拉病毒。

图 1.1 推特中的误传消息

（图片来源：https://twitter.com/funny\_truth/status/541155156264382465）

虚假消息，则是另一种更加普遍、潜在危害更大的谣言。这种谣言的特点是其发布时并无可靠的消息来源或有力的证据进行佐证，通常是人为臆测或刻意捏造产生的，其内容不具有真实性且常带有恶意或中伤。但通过包装，此类消息能获得很高的隐蔽性，虚假性很难被人们一眼辨识出来；通过包装，此类消息也能具有很强的轰动性，使得很多用户不顾其是否真实，盲目地争相转发。由于社交网络的用户量庞大、用户网络连通性强，虚假消息很容易被广泛传播：从一开始仅有的零星几个源头，发展成被成千上万人知晓、推送只需要短短几个小时。一旦这类消息被广泛传播，将有可能对某些个人或者团体的利益造成巨大损害，甚至引起社会的恐慌，产生非常负面的社会影响。如图 1.2这则推文谈论的就是一则虚假消息：当时谣言盛传美国摇滚歌手Akon在刚果共和国的一场庆典中将自己包裹在一个巨型的泡泡球内进行表演，其目的是为了避免自己在与群众接触时感染到当时在非洲盛行的埃博拉病毒。这则谣言当时在推特上疯传，部分用户对谣言表示怀疑、不确定（如图 1.2），也有Akon的粉丝站出来为他说话，但也有很多用户因此对Akon的为人表示愤怒、谴责和质疑，这使得其名声受到了很大的冲击。但后续的相关报道证实这只是一则虚假消息，Akon这种表演形式只是为了更好地和观众互动，而且远在几年前，埃博拉还未爆发时，他就已经用过同样的形式在世界各地进行表演了（参考信息来源：http://www.independent.co.uk/arts-entertainment/music/news/akon-didnt-perform-in-a-bubble-in-dr-congo-because-of-ebola-he-was-just-having-a-fun-time-for-9772004.html）。

由于社交网络每天都产生并传播着大量的消息，其中混杂了不少诸如此类的谣言，如果不加以监督管制，尽早发现危害性高的谣言，及时对其进行辟谣并将谣言散播者绳之于法，那么社交网络将成为虚假消息传播的温床，将对个人、团体以及社会造成不可估量的负面影响。

图 1.2 推特中的虚假消息

（图片来源：https://twitter.com/realalexjones/status/517808892437204992）

## 谣言检测

因为社交网络存在不少拥有负面影响的谣言，所以必须找到一种有效的监管机制。这就面临着一个很重要的问题：如何找到社交网络中的谣言？

一种朴素的想法是设立“消息审查员”的职位，利用人力去审核推特上的消息，找到其中疑似的谣言消息。但是社交网络的用户规模、消息数量巨大，若要对消息进行逐条的人工审核，那么需要的审查员数量也将极大，其人力成本将极高。因此人工地进行逐条审查并不现实，这不是一个合理有效的解决方案。

正是因为人工检测成本太大，所以研究学者们转而研究开发能自动识别、检测出谣言的技术。这类技术通常会利用计算机强大的处理能力分析海量消息，分析其文本、发布者或传播路径等等的特点，自动找到可疑度高、疑似谣言的消息。但由于机器识别的准确率有一定的限制，所以机器识别出来的可疑消息通常还需要送与审查员进行人工复查，确认为谣言才能采取后续的辟谣和法律相关工作。

以上是当前被广泛认可的，监管社交网络中谣言传播的解决方案之一，学术界称之为谣言检测（rumor detection）。目前还存在一些其他的解决方案，如设立辟谣公众账号，一旦群众用户发现疑似谣言的消息，就举报给公众账号，由公众账号选择举报度高的消息进行审查、辟谣；但本文只专注于谣言检测而略过其它。

谣言检测技术的研发并不容易，其面临了许多问题和挑战，以下是其中两个主要的挑战：识别准确度不高，检测重复率过高。

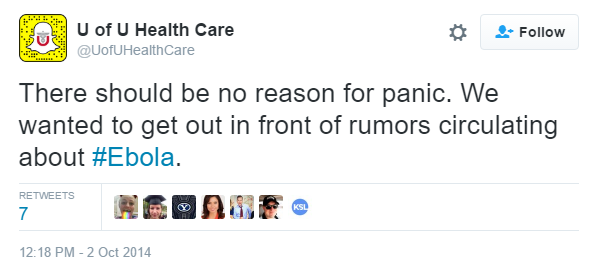
识别准确率不高，是指很多非谣言的消息被机器错误地识别为谣言，或者许多真正的谣言被机器漏过，误认为是普通消息。造成这个问题的因素很多，最主要的一个是分析自然语言难度很高，目前的技术通常只能对文本中的词或者短词组进行语义、情感分析，却很难将文本形成一个整体去分析句子或段落的真正意义。如图 1.3就是一个例子，在这则推文中，存在着如恐慌（panic）、谣言（rumors）和流传（circulating）这些词语，如果仅仅分析它们的语义和情感导向，机器很容易将这则推文误认为是谣言相关的消息。但事实上在恐慌这个词前面还存在没有理由（no reason）这个词组，在谣言和流传这些词前面还存在走出（get out）这个词组。只有将这些词组都串联起来考虑，这个句子的真正意义才能被分析出来：它不是在讲一引起恐慌的谣言，而是在讲我们不应当恐慌，而应该走出来面对那些传播的谣言，这是一则普通消息。

图 1.3 非谣言例子：自然语言分析面临的困难

（图片来源：https://twitter.com/UofUHealthCare/status/517755407100420096）

除此之外，即便自然语言的语义能被机器分析得十分透彻，也不一定能准确地判定一则消息是不是谣言，因为谣言的判定要求对消息的来源、内容的论据进行深入的考证，这就涉及到一定的专业知识和技巧，这不是单一的文本分析就能完成的事情。如图 1.4中的推文称感染了埃博拉病毒的护士Nina Pham的男朋友也被检测出埃博拉病毒的症状，在推文后面作者还附上了佐证的新闻网址，点开后也的确是一则符合推文内容的网页新闻。现在你能告诉我这是一则新闻推广还是一则谣言吗？或许有的读者会认为是前者，因为其附上了消息来源，是一个正规的新闻网站。但事实上我们并不能这么简单地下结论，因为即使是正规的新闻网站也会由于监管不力或为了博取点击率等原因出现一些不实的报道。很遗憾，这则新闻正是这样的报道，后续已有大量媒体对这则谣言进行了辟谣（参考信息来源：http://www.ibtimes.com/ebola-nurse-nina-phams-boyfriend-rumored-admitted-hospital-ebola-symptoms-alcon-releases-1707586）。这个例子说明，光是文本分析并不能完成谣言检测的任务。

图 1.4 谣言例子：谣言判定面临的困难

（图片来源：https://twitter.com/Jody\_Arrington/status/523633560779907074）

因此为了提高谣言检测的准确度，许多研究学者不仅考虑文本语义方面的因素，还综合考虑了如消息发布者信誉度、消息来源可靠度、消息传播路径拓扑结构等等的一些特征，在多元特征上加之以复杂的分析模型（如规则模型、概率模型），最终形成较为鲁棒、识别准确率较高的谣言检测系统。但这其中面临的问题就更多了，例如：如何找出那些有价值的、需要被考虑的特征？不同的分析模型各有优劣，它们之间能否形成互补？本文在接下来的几个章节中也在尝试解答以上的两个问题。

检测重复率过高，是谣言识别技术面临的另一个挑战。现在，假设我们的谣言检测系统的识别准确率达到了很高，甚至到达100%（即便这是不可能的），即便如此，但因为社交网络中的消息数目极大，经常会有如果成千上万个用户在讨论同一个话题，假设他们都在讨论、推送同一则谣言，那么就会产生成千上万则谣言相关的消息，而这些消息都能被我们卓越的系统成功地检测出来；但当我们将检测出来的谣言交给审查员进行人工复查，或者交给相关部门进行辟谣和法律维权时，问题就出现了：这么多的谣言消息，审查人员和相关部门根本看不过来，其中还混杂着一些传播不广泛的小谣言；最重要的是，有大量的消息都是关于同一个谣言的，重复去审查、处理这些消息根本毫无意义。要想利用有限的人力资源审查更多不同的谣言

# 插图索引

[图 1.1 推特中的误传消息 2](file:///D:\rumor\DataProcess\综合论文训练-钟仰新.docx#_Toc452484983)

[图 1.2 推特中的虚假消息 3](file:///D:\rumor\DataProcess\综合论文训练-钟仰新.docx#_Toc452484984)

[图 1.3 例子：自然语言分析面临的困难 4](file:///D:\rumor\DataProcess\综合论文训练-钟仰新.docx#_Toc452484985)

# 表格索引

[表 4.1 兔子布偶场景中两种渲染方法的效率比较 14](#_Toc451804965)



# 参考文献

[1] Cao, N., Shi, C., Lin, S., & Lu, J. (2015). TargetVue: Visual analysis of anomalous user behaviors in online communication systems. IEEE Transactions on Visualization and Computer Graphics, 22(1), 280-289.

[2] Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on twitter. Proceedings of the 20th International Conference on World Wide, 675-684.

[3] Karamchandani, N., & Franceschetti, M. (2013). Rumor source detection under probabilistic sampling. International Symposium on Information Theory, 2184-2188.

[4] Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). Prominent features of rumor propagation in online social media. IEEE 13th International Conference on Data Mining, 1103-1108.

[5] Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q. (2011). Rumor has it: Identifying misinformation in microblogs. Proceedings of the Conference on Empirical Methods in Natural Language Processing, 1589-1599.

[6] Seo, E., Mohapatra, P., & Abdelzaher, T. (2012). Identifying rumors and their sources in social networks. Proceedings of the Society of Photographic Instrumentation Engineers, , 8389

[7] Sun, S., Liu, H., He, J., & Du, X. (2013). Detecting event rumors on sina weibo automatically. Proceeding of the Web Technologies and Applications, 120-131.

[8] Takahashi, T., & Igata, N. (2012). Rumor detection on twitter. Joint 6th International Conference on Soft Computing and Intelligent Systems and 13th International Symposium on Advanced Intelligent Systems, 452-457.

[9] Wu, K., Yang, S., & Zhu, K. Q. (2015). False rumors detection on sina weibo by propagation structures. IEEE 31st International Conference on Data Engineering, 651-662.

[10] Yang, F., Liu, Y., Yu, X., & Yang, M. (2012). Automatic detection of rumor on sina weibo. Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics, Article No. 13.

[11] Zhao, Z., Resnick, P., & Mei, Q. (2015). Enquiring minds: Early detection of rumors in social media from enquiry posts. Proceedings of the 24th International Conference on World Wide Web, 1395-1405.

# 致 谢

感谢张慧老师对我论文的知道。感谢黄经纬同学和我进行的有关光场的讨论，给我的工作拓展了思路。

声 明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签 名：\_\_\_\_\_\_\_\_\_\_\_\_\_ 日 期：\_\_\_\_\_\_\_\_\_\_\_\_

# 附录 A 外文资料的调研阅读报告

In 1996, Marc Levoy and Pat Hanrahan first brought the concept light field to the computer science community, in their siggraph paper Light Field Rendering. The advantage of light field compared to other image-based rendering (IBR) techniques is that no depth information or correspondence knowledge about images is required.

Light field is defined as the radiance at a point in a given direction, which can be described using 5D plenoptic function. But because the radiance along a line is constant (if there is no blocking object), this function can be reduced to 4D function to remove the redundancy.

In their paper, the authors recommended to use two planes (called light slab) for the representation of the light field. The coordinate system on the first plane is (u, v) and on the second plane is (s, t). The radiance along an arbitrary line passing through these two planes is represented by L (u, v, s, t).

In practice, we usually call the uv plane camera plane and the st plane focal plane. Light field can then be represented by 2D array (camera plane) of images (focal plane). So we can use an array of cameras to capture light fields.

When reconstructing image at an arbitrary viewpoint, there are two steps in the process: step 1 is to compute the line parameters (u, v, s, t), and step 2 is to resample the radiance at the line parameters. The first step can be easily done using texture mapping. As for the second step, they simply interpolate the 4D function from the nearest samples in their implementation.

Because 4D light field array is very large, to make creation and transmission possible, they also proposed a compression method. In their application, they first applied vector quantization on the light field data. Typically, they used 4D tiles of the light field, yielding 48-dimensional vectors, and generated codebook using mean-squared error(MSE). Then they further compressed the data using entropy coding, which can be easily done using gzip. Typical compression rate is 100:1.

In Marc et al.’s work, the cameras have to be placed with fixed intervals on a fixed plane, which restricted its application. In 2001, Chris et al. first proposed a real-time rendering method that does not require the input camera to be restricted to a plane, or any specific manifold.

When reconstructing a desired ray, Chris et al. proposed to use source image rays with the most similar angle to the desired ray, instead of the closest ray in (u, v, s, t) parameter space.

The most important concept proposed in their paper is camera blending field. It describes how each source camera is weighted to reconstruct a given pixel.

Some notations: given desired ray , intersects with the object at some frontmost point p. is the center of each camera. is the ray between p and .

The difference between and is defined as:

accounts for resolution difference and accounts for resolution difference.

The larger the difference, the smaller the weight should be:

Because rays may fall outside of the source camera, we have

is a function that continuously goes to zero as approaches the edge.

Find k largest . Finally normalizing over i yields the blending weight:

To achieve real time rendering, in their implementation, they only compute the camera blending at discrete set of points, and interpolate the camera blending over the image.

In 2012, Ade et al. were the first to design a system that allowed user to capture light field using mobile phones. Although in Chris et al.’s work, unstructured lumigraph rendering had been proposed, it is still very challenging for users to capture light fields. Because users usually take images by scanning the subject in a back-and-forth manner, it is crucial to guide the user to make sure the scanline starts at the correct offset from previous one. Without guidance, the distance between the camera and subject can vary discontinuously, which usually will give rise to artifacts. Also it is better to give users an idea how much has been covered.

Ade et al. designed a capturing process and used the augmented reality library SLAM to solve the above issues. In their system, the user first points the camera to the scene and selects a subject to capture. The system then records new images whenever it determines that it is viewing an under-sampled region of light field. They guide the user with a viewpoint coverage map to help them achieve dense coverage.

One of the main contributions of their work is to propose a coverage criterion, called reprojection error. Consider a captured view s, and a new view n. We need then to determine when the view n is not covered by s. Consider a cone of ambiguity associated with a pixel V\_s of s, which means all the object points in this cone may contribute to the pixel color of V\_s. Then consider the intersection part between the core and the object sphere. The reprojection error is then defined as the longest distance (in this case AB) projected to the new view n. If this value is larger than a threshold, we determine that n is no longer determined by s and a new image is recorded. This new criterion is sensitive to both parallax error and resolution difference.

Another contribution of their work is a fast rendering technique that can achieve piecewise-bicubic at the limit. Here we only introduce the piecewise-linear reconstruction because piecewise-bicubic reconstruction is a simple modification of piecewise-linear one by extending the projection to 2-ring and apply the Loop subdivision rules. They first build Delaunay triangulation on the viewpoints. Then instead of finding the k nearest cameras, they intersect the desired ray with the triangle mesh and interpolate using barycentric coordinates. To accelerate this process, they first project the triangle mesh to the target view. For a given vertex, it has a stored view with it. At the vertex, the weight is 1 and at the neighboring vertex, the weight is 0. This is to say a given stored view can only affect a triangle fan. So they can efficiently render the view using texture mapping and alpha blending.

综合论文训练记录表

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **学生姓名** |  | **学号** |  | **班级** |  |
| **论文题目** |  | | | | |
| **主要内容以及进度安排** | **指导教师签字：**  **考核组组长签字：**  **年 月 日** | | | | |
| **中期考核意见** | **考核组组长签字：**  **年 月 日** | | | | |
| **指导教师评语** | **指导教师签字：**  **年 月 日** | | | | |
| **评阅教师评语** | **评阅教师签字：**  **年 月 日** | | | | |
| **答辩小组评语** | **答辩小组组长签字：**  **年 月 日** | | | | |

**总成绩：**

**教学负责人签字：**

**年 月 日**