中国科学技术大学

“科学与社会”新生研讨课研究报告

报告题目：网络多媒体谣言检测技术调研

小组组长：陈文博

小组成员：陈文博，李杉杉，陈康，吴舜钰

导师姓名：张勇东

2018 年 5 月 21 日

工作安排：

1.研究小组成员及其承担的主要工作

学号	姓名	所在学院	在研究和报告撰写中承担的主要工作
PB17050948	陈文博	信息科学技术学院	主要的报告编辑和任务分配，撰写报告中“视觉特征“、”网络数据采集“、”谣言检测“部分。
PB17071396	李杉杉	信息科学技术学院	撰写报告“文本特征”和“特征融合”部分、“社交特征”、“篡改性特征”、“参考文献”。
PB17030851	陈康	信息科学技术学院	主要的 PPT 编辑整理，背景调查，撰写报告“摘要”、“引言”、“社交特征”部分。
PB17071401	吴舜钰	信息科学技术学院	制作整理调查问卷，撰写报告“谣言传播模型及基于网络结构的谣言源检测”，“通过传播树建模的基于传播核的谣言判断”和“民意调查”部分

2.进度安排

时间	进度
2018 年 1 月 25 日	确定课题，分配工作
2018 年 1 月 25 日~2 月 25 日	搜索收集论文资料，拟定项目架构，进一步分配工作
2018 年 2 月 25 日~3 月 6 日	各组员按分配的任务进行调研整理
2018 年 3 月 7 日	组员集会，讨论交换意见，并提出改进建议
2018 年 3 月 7 日~4 月 15 日	PPT 制作
2018 年 4 月 15 日~5 月 21 日	论文制作
2018 年 5 月 21 日~6 月 20 日	后期优化

摘要

近年来，随着互联网与多媒体技术的蓬勃发展，越来越多由用户生成的信息在网上传播，人们也逐渐接受这种由互联网多媒体社交平台上传播的信息。但由于互联网的开放性和更新速度极快，各种谣言也顺势而生，严重损害社会的稳定与和谐，造成巨大危害。由于传统人工谣言检测工作量大、实时性差，如何利用计算机技术自动鉴别多媒体谣言成为当下一个研究的热点。本文总结多媒体谣言检测的基本背景（包括多媒体谣言的传播模型与社会危害）、技术实现（包括数据采集、特征提取、特征融合、分类检测）以及用户们对该技术的看法（主要调查百姓对多媒体谣言检测技术的信任程度）。由此对多媒体谣言检测技术有着更加全面深入的了解。

关键词：谣言检测、多模态、机器学习、网络多媒体

Abstract

In recent years, with the rapid development of Internet and multimedia technology, more and more information generated by users is spread on the Internet, and people are gradually accepting this kind of information spread on the Internet multimedia social platform. However, due to the openness and rapid updating of the Internet, all kinds of rumors are also coming into being, which seriously damages the stability and harmony of the society and causes great harm. Due to the large workload and poor real-time performance of traditional artificial rumor detection, how to use computer technology to automatically identify multimedia rumors has become a hot topic of research. This paper summarizes the basic background of multimedia rumor detection (including the propagation model of multimedia rumors and social harm), and the technical realization (including data collection, feature extraction, feature fusion, classification detection) and the views of the users on the technology (mainly investigating the trust degree of the common people to the multi-media rumor detection technology). Thus, we have a more comprehensive and in-depth understanding of multimedia rumor detection technology.

Keywords: Rumor Detection, Multimodal, Machine Learning, Network Multimedia

目 录

一、 引言

多媒体（Multimedia）是多种媒体的综合，一般包括文本，声音和图像等多种媒体形式。在计算机系统中，多媒体指组合两种或两种以上媒体的一种人机交互式信息交流和传播媒体。使用的媒体包括文字、图片、照片、声音 、动画和影片，以及程式所提供的互动功能。

多媒体的应用领域已涉足诸如广告、艺术，教育，娱乐，工程，医药，商业及科学研究等行业。利用多媒体网页，商家可以将广告变成有声有画的互动形式，可以更吸引用家之余，也能够在同一时间内向准买家提供更多商品的消息。

利用多媒体作教学用途，除了可以增加自学过程的互动性，更可以吸引学生学习、提升学习兴趣、以及利用视觉、听觉及触觉三方面的反馈来增强学生对知识的吸收。

多媒体技术是一种迅速发展的综合性电子信息技术,它给传统的计算机系统、音频和视频设备带来了方向性的变革，将对大众传媒产生深远的影响。多媒体计算机将加速计算机进人家庭和社会各个方面的进程，给人们的工作、生活和娱乐带来深刻的革命。

多媒体还可以应用于数字图书馆、数字博物馆等领域此外，交通监控等也可使用多媒体技术进行相关监控。

作为信息时代的一大信息传播方式，影视、广播、图片、各大社交网络平台（如微博）无时无刻都在向大众传递着各式各样的信息。与此同时，多媒体也成为舆情爆发升温的聚集地。尤其是当今社交媒体的发展与普及，信息传播迅速，传播量可以在短时间内到达指数级增长;信息传播广泛，通过转发等活动，信息可以大范围的扩散，使信息在短时间内家喻户晓;信息传播不可控，人的转发意识是由思想控制的，但思想是不受外界因素的干扰，我们不能控制用户的行为。正是由于这种传播的便捷性、迅速性，发帖的自由性及个性化，传播者的平民化及多样化,关于热点话题的消息在传播的过程中,逐渐被放大夸张化,偏离正轨,错误的思想观点充斥着网络。很容易地逐层扩散,达到故意传播谣言,涣散人心,恶意抹黑他人的行径,或者利用悲惨的消息骗取网友同情心乃至造成金钱的损失。

目前，为了能够快速的粉碎谣言，各大媒体平台都采取相关措施打击造谣、传谣行为，例如，集合有丰富经验的人员，对核心话题及传播广泛的内容开展全天监控，鼓励网友参与谣言的举报、对疑似谣言的信息进行求证、实地考察。但是这些方法需要投入有经验的人员，并且耗时较长，需要具备专业的背景知识。我们希望能够通过谣言信息与真实信息在语言学上的差异，以及传播学等方面的差异性，构建模型，自动挖掘谣言，辅助预警、预防、监控、治理等谣言清除工作，净化我们的文化环境。

本研究主要针对以下几个方面进行调研：

1、多媒体谣言传播模型。（以新浪微博平台为对象）

2、多媒体谣言检测的基本流程与方法。

a) 数据采集

b) 特征提取

c) 特征融合

d) 谣言检测

e) 评估

3、多媒体谣言检测的民意调查

二、 相关调研

1. 多媒体谣言传播模型简介

谣言检测的最终目的是及时并有效地阻断未经证实消息的传播，防止其可能产生的不良社会影响。其中，谣言源的识别与控制至关重要，它能有效地找到谣言传播的根结所在，并能最高效地控制其进一步传播。基于网络结构的谣言检测聚焦于谣言源的检测，且隶属于信息源推断问题，它以图的形式抽象地描述社会网络拓扑结构，同时抽象出信息在社会网络中的传播模型，然后依据感染传播子图快照构建谣言源节点估计器，从而使估计的准确率最大。

1.1 在线社会网络（Online social networks,OSNs）上的信息传播

1.1.1 OSNs 结构抽象

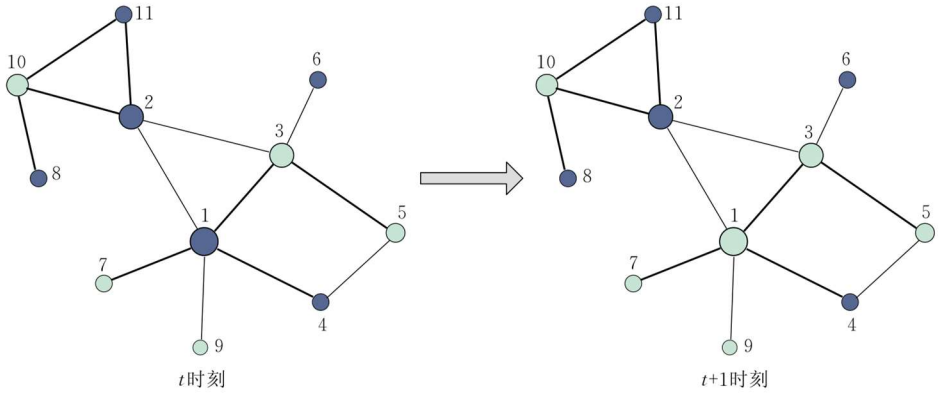


图 2　信息扩散示例

将一个社会化网络用图结构进行抽象，图由节点集合和传播边的集合构成。初始时刻网络中有一些激活节点（浅色）和未激活节点（深色），节点上的数字代表权重，对应用户影响力，传播边的节点代表用户关系强度。网络中的节点在受到活动节点影响时可以由未激活状态转为激活状态。此图称为传播图。

同时定义一个有向图，该图为上文所定义图结构的子图，由参与转发信息的节点集合和参与传播的边集合构成。此图称为传播路径图。

1.1.2 传播过程描述

信息在 OSNs 上的传播可描述为，在初始时刻信息源的节点集合发送信息。此时传播路径图中只有信息源集合。下一时刻与信息源直接相连的节点集合以一定概率被激活，往传播路径图中加入这些节点和信息传播边。对再下一个时刻进行相同操作，直至参与转发信息的节点结合不再变换，得到传播路径图。

1.2 传播模型分类

1.2.1谣言传播研究早期：

D-K 模型（谣言传播数学模型）

Daley 和 Kendall 于 20 世纪 60 年代提出了谣言传播的数学模型，后来的研究者以 Daley 和 Kendall 的名 字称之为 D—K 模型。该模型是借助随机过程的方法来分析谣言问题的，它把受众按照谣言传播效果分成了 3 类，并假定其中两类人之间角色转换的概率满足一定数学分布。

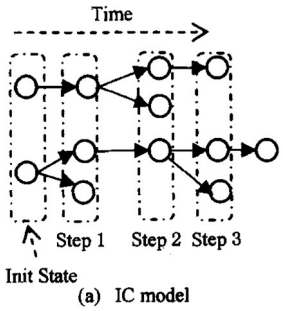
缺点：不完全符合谣言传播过程，但在一定近似条件下是合理的。

1.2.2网络结构模型

结合对信息传播路径图的介绍，根据研究对象的不同，将近几年提出的OSNs上的信息传播的模型分为：基于传播路径图节点、基于传播路径图规模两大类传播模型。

1.2.2.1基于传播图节点的传播模型：

（1）独立级联模型（independent cascades model, IC）：



(a) IC model

独立级联模型中传播图每条边对应一个传播概率，初始时刻激活状态的节点称为种子节点。任意时刻传播过程中，由上一时刻已激活节点以相应传播概率激活它的邻居节点，不同节点的激活过程相互独立，互不干涉。没有被激活的节点没有记忆性，下一次还可以被邻居节点以一定概率激活。

（2）线性阈值模型(Linear Threshold Model)

线性阈值模型考察接收节点，每个接受节点有一个影响阈值，当它的所有已激活邻居节点对它的影响之和超过阈值时，该接收节点被激活。没有被激活的接收节点有记忆性，本次影响力与下一次影响累计，达到阈值即被激活。

线性阈值模型充分考虑了信息接收者的特性和信息传播具有记忆效应。

小结：

基于传播路径图节点模型的实质是分析节点用户在多重因素作用下是否参与转发信息或转发信息的概率.

表2 基于传播路径图节点的传播模型的比较

Table 2 Comparison of models based on nodes of propagation paths graph

小类	考虑因素	是否考虑网络结构	模型关键点
LT 模型	邻居节点的综合影响	是	阈值 θ_i、综合影响 $\varphi_i(t_i)$
IC 模型	单个邻居节点的影响	是	激活概率 $p(v_j,v_i)$

1.2.2.2 基于传播图规模的传播模型：

基于传播路径图规模的传播模型以传播路径图的节点规模为研究对象，其实质是分析参与转发信息的用户的规模。目前基于传播路径图规模的动态模型有：SIR模型及其改进模型/马尔可夫模型/场强模型/神经网络模型；基于传播路径图规模的静态模型有回归模型，这里只介绍SIR模型。SIR模型的理论方法是传染病理论。

由于传染病扩散和信息扩散类似，可以将传染病理论运用到信息扩散研究中。目前研究广泛的传染病模型是 SIS 模型和 SIR 模型。在 SIS 模型中节点有易受感染状态和已受感染状态两种状态。某一时刻，一个接收节点周围有一个或多个已受感染节点时，该接收节点会以一定概率变成已受感染状态。在 SIR 模型中，还有第三种免疫状态。

（1）SIR 模型：

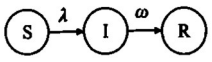


图2 SIR 状态转移图

SIR(Susceptible-Infective-Removal)模型最早是Kermack 等人研究黑死病的传播规律时提出的动力学模型.信息传播领域的 SIR 模型将节点用户集合按照状态不同划分为三个状态，分别为未接受信息并可能转发信息状态、已接受并转发信息状态、已接受信息但不转发信息状态，模型认为节点用户总数不变。

未接受信息并可能转发信息状态的节点和已接受并转发信息状态的节点交互后会以一定概率接受并转发信息，变为已接受并转发信息状态。已接受并转发信息状态的节点也会以一定概率对信息失去兴趣变为已接受信息但不转发信息

状态。状态转移路径如上图所示，模型满足一个动力学方程。

（2）SPNR模型

虽然传染病模型以及改进模型被广泛应用于谣言传播过程的描述 ，但通过对实际谣言传播过程的分析，发现传染病模型并不能完全表征谣言传播过程中用户的所有状态，由于谣言本身具有的未经证实性以及社交网络中用户生活、知识背景的差异，导致对同一谣言存在“相信谣言”及“不相信谣言”两种不同的状态， 即谣言传播过程中个体存在对谣言的两种不同态度。

根据这一特点，将谣言传播过程中个体对于谣言的感染被分 为两种截然不同的状态，一种是相信谣言状态，也称为对谣言的正向感染状态；而另一种是不相信谣言状态， 称为对谣言的负向感染状态。

由于谣言传播过程中存在两种不同的感染状态，且两种感染状态的变化趋势不同， 经典的 SIR 模型无法对谣言传播过程进行准确描述。因此，结合谣言传播过程的实际情况，对经典 SIR 模型进行了改进，将一般的感染状态细分为正向感染和负向感染两种感染状态，改进后的谣言 传播模型包含以下四种状态 ：易感状态（Susceptible）、正向感染状态（Positive Infected）、负向感染状态（Negative Infected）以及免疫状态（Recovered），称此谣言传播模型为 SPNR 模型，

模型中各状态的转移情况如图所示。

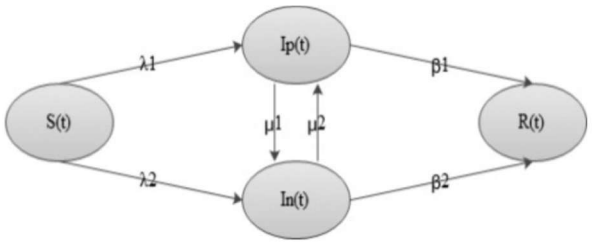


图1 SPNR谣言传播模型的状态转移图

- SPNR 模型中四种状态的转移情况

1）当一个易感状态个体与正向感染状态个体接触时， 易感状态个体以λ1的概率转变为正向感染状态个体；当一个易感状态个体与负向感染状态个体接触时，

易感状态个体以λ2的概率转变为负向感染状态个体；

2）当一个正向感染状态个体与负向感染个体接触时， 正向感染状态个体以μ1的概率转变为负向感染状态个体；当一个负向感染状态个体与正向感染个体接触时，负向感染状态个体以μ2的概率转变为正向感染状态个体；

3）当一个正向感染状态个体与免疫状态个体接触时，正向感染状态个体以β1的概率转变为免疫状态个体；当一个负向感染状态个体与免疫状态个体接触时，负向感染状态个体以β2的概率转变为免疫状态。

基于上述对谣言传播规则的描述， 提出 SPNR 谣言传播算法。

```
/*输入：
 *    正向感染率λ1，负向感染率λ2
 *    正向免疫率β1，负向免疫率β2
 *    正向转移率μ1，负向转移率μ2
 *    处于易感状态的节点：state（1，i） = 0；
 *    处于正向易感状态的节点：state（1，i）= 1；
 *    处于负向易感状态的节点：state（1，i）=-1；
 *    处于免疫状态的节点：state（1，i）= 2；
 *输出：
 *在t时间后，所有节点标识：state
 *生成无标度网格：G = {V,E}，领接矩阵为A
 *初始化：将各个节点初始时刻的状态进行标识，state = {1, -1, 0, ..., 0}
 */
while 时间间隔小于t  do
      for i = 1, 2, ..., n
            switch state(1,i)
                  处于易感状态的节点：
                        以概率λ1转变为正向感染状态
                        以概率λ2转变为负向感染状态
                  处于正向感染状态的节点：
                        以概率μ1转变为负向感染状态
                        以概率μ2转变为免疫状态
      处于易感状态的节点：
                        以概率μ1转变为正向感染状态
                        以概率μ2转变为免疫状态
            end switch
      end for
end while
//得到所有节点状态state
```

1.3 基于网络结构的谣言源检测方法

1.3.1基于传播子图快照的检测方法

基于传播子图快照的检测方法：

1． 一次或多次获取全部或部分节点是否收到谣言消息的状态子图

2． 对某一网络拓扑属性度量进行特征估计，

3. 推算出网络中最大可能成为谣言源的节点。

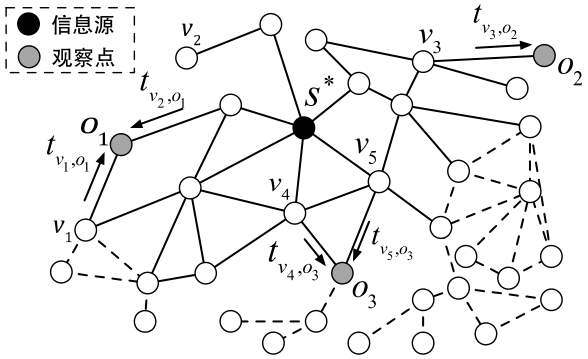
（其中传播子图的网络拓扑属性度量特征以 Shah 和 Zaman 等提出的谣言中心性最为常见，即基于组合数最大似然估计的源点估计量。）

缺点：

1. 很难一次观察到整个网络的状态，且观察到的网络感染状态不一定是真实的感染状态，存在部分节点被感染但未被表达出来的情况。
2. 研究假设每个节点是谣言源的概率是相同的，然而事实上只有部分节点才可能是谣言源。
3. 基于静态网络检测，未考虑时间属性特征。
4. 假设谣言传播仅存在单一的谣言源，未考虑多谣言源。
5. 假设底层的网络传播模型是固定也已知的，应用范围被限制。

1.3.2 基于部署节点的检测方法

由于 OSNs 网络的规模巨大且复杂，传播图中节点的真实感染状态很难完整获取。同时，在 OSNs 网络中，不同节点重要性不同，其实不需要对网络中所有节点都进行研究,考虑重要性较大的节点可以减少计算量。所以对于给定的网络，在不了解网络节点感染状态和节点关系的情况下,可以选取适量数量和重要位置的节点作为整个网络的观察点进行研究。基于部署节点的检测方法依据这一思想，在 OSNs 中部署少量的观察点，记录它们首次收到邻居节点发送来的消息时的时间和方向，然后通过统计计算推断出网络的谣言源。



基于部署节点的检测方法在具体操作过程：

1. 部署观察点
2. 非候选源点为根构建生成树
3. 计算候选源点估值

4. 候选源点估值排序

5. 将估算值最大的候选源作为当前网络的谣言源

优点：仅需要获取少量观察点反馈的传播信息而无需窥探整个网络传播状态，在实际应用中的可行性高。

缺点：该方法检测的准确度和算法开销取决于观察点在网络中所部署的位置和数量。

一个共识：在网络中对信息传播影响力越大的节点，其在信息传播过程中接收谣言信息的可能性越大，记录的谣言传播信息也越有效。

1.4 基于传播结构核学习的谣言检测

首先利用传播树对谣言扩散过程进行建模，为原始消息随时间推移的传播和发展趋势提供有价值的线索。然后提出了一种称为传播树的基于内核的方法，它通过评估传播树结构之间的相似性来捕获区分不同类型谣言的高阶模式。并利用传播结构判断一个消息是否是谣言。

1.4.1 传播树核建模

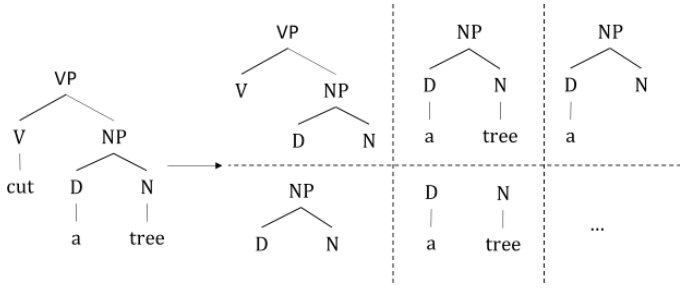
任务：给定一个源传播树，判断其类型（标签）

1.4.1.1 树核背景

树核旨在通过计算相应解析树之间常见子树的数量来计算两个自然语言句子之间的句法和语义相似度个节点都与语法生成规则相关联。

给定一个句法解析树，其孩子的每个节点都与语法生成规则相关联。

下图说明了cut a tree及其子树的句法分析树。子树被定义为具有多于一个节点的任何子图，限制是必须包括整个（非部分）规则生成。



评估公共子树方法：

1. 给定两个解析树和核函数，每棵树有由它的节点集合和生成规则

2. 递归计算，分为两颗解析树上的节点生成规则不同，前端节点生成规则相同和上述两条都不满足三种情况。

1.4.1.2 PTK 模型

为了对传播树进行分类，可以先计算出树之间的相似性，这些树根据结构、语言和时间属性来反映不同类型的谣言和非谣言的区别。但现有的树核还不能应用于传播树，因为：

1）与分析树不同，其中节点由可枚举的标称值（例如，词性标签）表示，传播树节点被给定为连续数字的向量表示，表示节点的基本属性；

2）两个解析树的相似度是基于共同子树的数量，通过检查相同的生成规则和相同子项是否与两个子树中的节点相关联来评估子树的共性，而在我们的上下文相似度应该被定义为软函数，因为来自不同传播树的两个节点很难是相同的。

为了解决上述问题，可以定义一个新的函数来评估两个节点之间的相似度，该函数与用户相似性和内容相似度有关。用户相似性由欧式距离定义，内容相似性由 Jaccard 系数测量。

当给定两个传播树，PTK 旨在基于枚举所有最相似的子树对来迭代计算两个传播树之间的相似性。

1.4.1.3 上下文感知的 PTK 扩展模型

PTK 的一个缺点是它忽略了子树外的线索，例如信息如何从源文件传播到当前的子树。

上下文感知的 PTK（cPTK）：

考虑从树根到子树根的传播路径，计算相似度，其想法与上下文感知的树核一致。

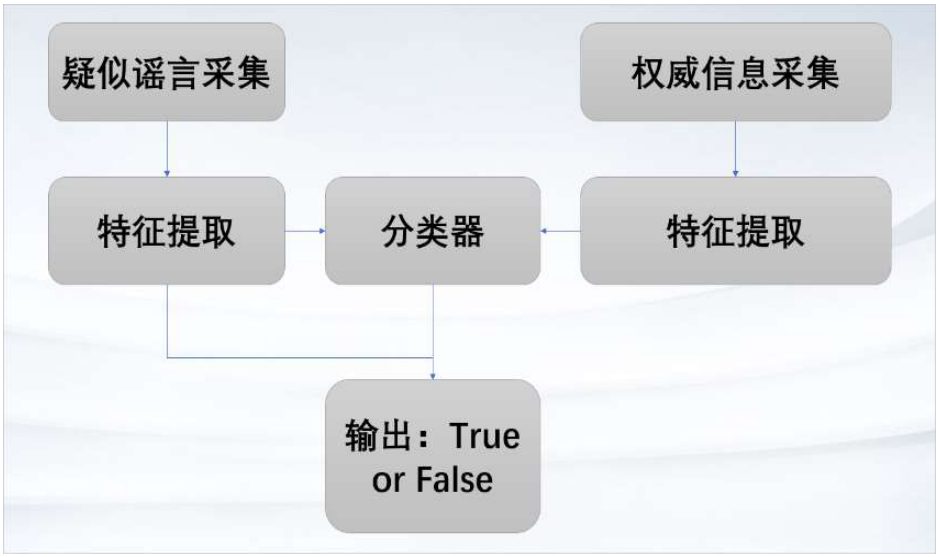
1.4.2 基于核学习的谣言检测

将所提出的树核函数，利用基于核的 SVM 分类器合并到监督学习框架中，将每个树视为一个实例，其所有训练实例的相似度值作为特征空间。

优点：避免复杂特征。因为核函数可以在计算两个对象之间的相似性时探索隐式特征空间。

2. 多媒体谣言检测技术

2.1. 基本流程



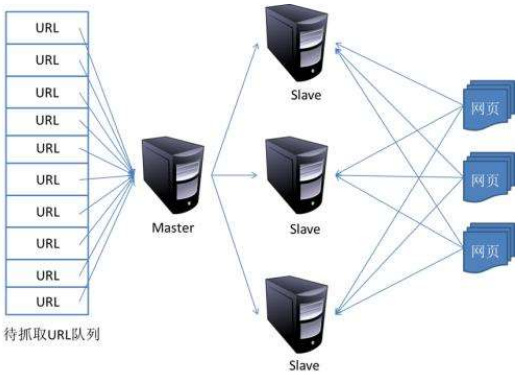
图一 谣言检测流程图

2.2. 网络数据采集（Network Data Acquisition）

进行多媒体谣言的自动检测，我们首先需要获取到疑似谣言信息作为鉴定对象以及与之相关的权威信息作为参考对象。传统的信息数据采集是通过相关工作人员通过搜索引擎进行检索的。随着技术的发展，API 与网络爬虫（Web Crawler）技术给我们提供了高效的网络数据采集方法。

现代网络数据采集基本方法包括网络爬虫、分词系统以及任务与索引系统等技术组成。

网络爬虫是一种按照一定规则，自动抓取网络信息的程序。基本流程是：URL→请求资源→解析网页→存储信息。按照按系统结构和实现技术可分为：通用网络爬虫（General Purpose Web Crawler）、聚焦网络爬虫（Focused Web Crawler）、增量式网络爬虫（Incremental Web Crawler）、深层网络爬虫（Deep Web Crawler）。按抓取的



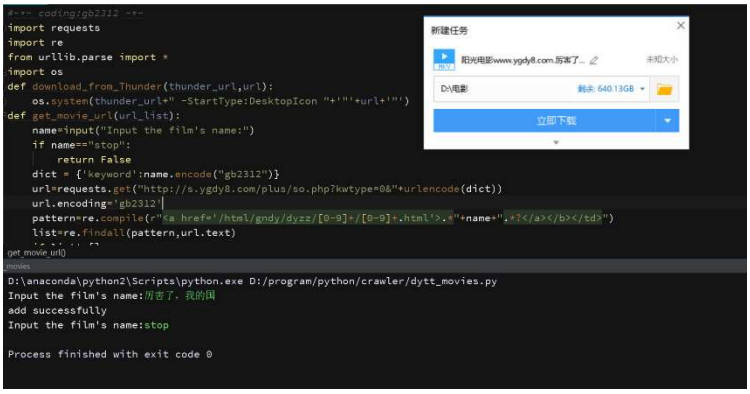
图二 爬虫原理图

目标可分为基于目标网页特征、基于目标数据模式、基于领域。

常用的搜索策略：深度优先(Depth First Search, DFS)、广度优先(Breadth First Search, BFS)、最佳优先(Best First Search)。常用的文本分析算法有超文本和纯文本的分析和聚类算法。

小结：

本节介绍了一些网络数据采集的基本方法，实现机器数据采集代替人工数据采集，极大提高了检测效率，为后面的检测提供了数据基础。



图三 图为调研过程中小组成员编写的一个简单的基于爬虫的电影下载器

2.3. 信息特征提取(Feature Extraction)

在获取所需要的信息后，我们需要对所得信息进行处理方便我们进一步检测。传统人工检测谣言需要有专业人员将收集到的信息进行多方调查比对，其中离不开决定了信息真伪的关键特征，在机器自动检测谣言技术中，如何更全面更准确的提取这些决定性特征是谣言检测技术中的一大重点。我们以微博作为多媒体平台进行研究。

微博是一个多用户社交媒体平台，离不开用户与信息，其中信息一般包括文字、图片和视频。其中文字和图片是信息传播的一大重要载体。下面我们将信息特征分为文本特征、视觉特征以及社交特征进行分析。

2.3.1. 文本特征(Text Feature)

文本是消息传播最常见的途径，简短的文本常常含有丰富的内容，许多谣言伪信息也潜藏在字里行间中，由此，本文介绍几种可以从文本信息中获取的特征来作为我们谣言检测的原材料：

（1） 词法特征(morphological features)：单个字级别的或单个词级别的语言特征，包括总字数、总词数、不同单词个数、

每个词平均长度等。

（2） 句法特征(syntax features)：句子级别的语言特征，包括关键词频数、标点符号类型和数目，以及词性标注等。

（3） 主题特征(Thematic characteristics)：主题级别的语言特征，例如对整个文档集构建主题模型、还有提取的消息话题特征、消息的情感倾向特征等。

其中，我们提到消息的情感倾向特征，文献【2】中提到一种基于情感分析的谣言检测方法，接下来简单介绍一下：

1）界定高低质量源

主要通过两种方法或两种方法补充使用：

a. 直接评价法一般通过建立指标评价体系的方法, 对不同信息来源媒介每一项指标进行打分, 综合各项指标对 信息源进行评价。

b. 间接评价法通过信息用户来评价信息源, 以调查表 的方式调查用户对信息源的需求和利用情况, 其评价较为客观, 但是工作量大, 需要信息用户的高度配合。

2）预处理文本

主要是对句子进行分词以及过滤无关词汇。

3）计算情感值

文献采用的计算情感值的方法是基于情感字典，基本算法如下：

遍历疑似谣言文本词表，如果词语匹配到专有名词词典中的词汇。

①寻找该专有名词前后的词语，查找修饰该词的正向情感动词或负向情感动词，情感得分为 *Score*。

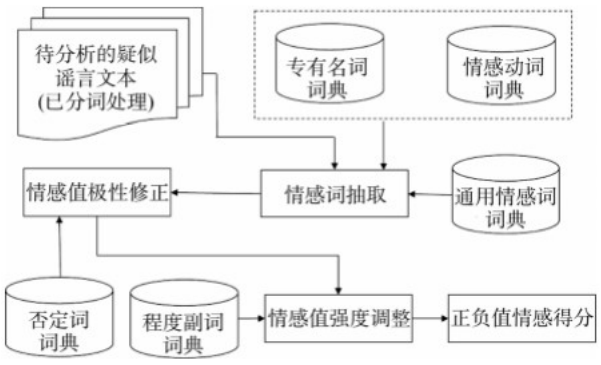
②若满足条件①，则继续寻找并统计该专有名词词汇前面 5 个词语及该词汇后两个词语的范围内否定词的个数 $reverseTime_{SpeNoun}$，如果 $reverseTime_{SpeNoun}$ 是奇数，修正情感极性，令 $reverse_{SpeNoun}$=-1，否则不需要修正情感极性。

③寻找该专有名词前 5 个词语及该专有名词后两个词语范围内的程度级别词汇，用 *degree* 乘以匹配程度词语在程度级别词典中的情感强度系数，如公式(3)所示。

$$degree = degree \times DegDic_{DegWord} \quad (3)$$

④分别累加对应的正、负面情感得分，规则如公式(4)所示，当 $Score \times reverse_{SpeNoun} > 0$，情感得分累计到 *positive*，否则情感得分累计到 *negative*。

$$SentimentValue = \sum Score \times degree \times reverse_{SpeNoun} \quad (4)$$



图四 情感值计算流程

4）谣言识别

通过文本情感值计算模块,由评分结果判断信息真伪。

例：“牛奶致癌”，“牛奶”一词在高质量文本中常与“补钙”“对人体有益”等积极情感相关联，累积得分为正分，而“致癌”情感得分为负，出现差异，判断“牛奶致癌”为谣言的可能性更大。

但情感分析只能用于判断较为简单直白的谣言，且对已有权威资料依赖度高，方法不适用于对预言性谣言的判断，对从前尚未涉及的领域也无法给出判断结果。

2.3.2. 视觉特征(Visual Feature)

图像视频具有简洁客观反映信息的能力，是媒体信息传播中的一个重要信息传播媒介，加上人们更愿意相信由摄像机所捕获的“真实图景”而非由作者复述的文字，图像视频常常也用于提高信息的真实性。然而随着数字图像处理的日趋成熟，照片信息也是可以根据人的意愿而改变的。而所谓的视觉特征就是图像视频中存在的能够一定程度反映其对该疑似谣言信息的真伪性的一种特征。

根据特征提取的角度不同，可以将视觉特征分为三大类：统计特征、内容特征、深度学习特征：

1. 统计特征(Statistics Feature)

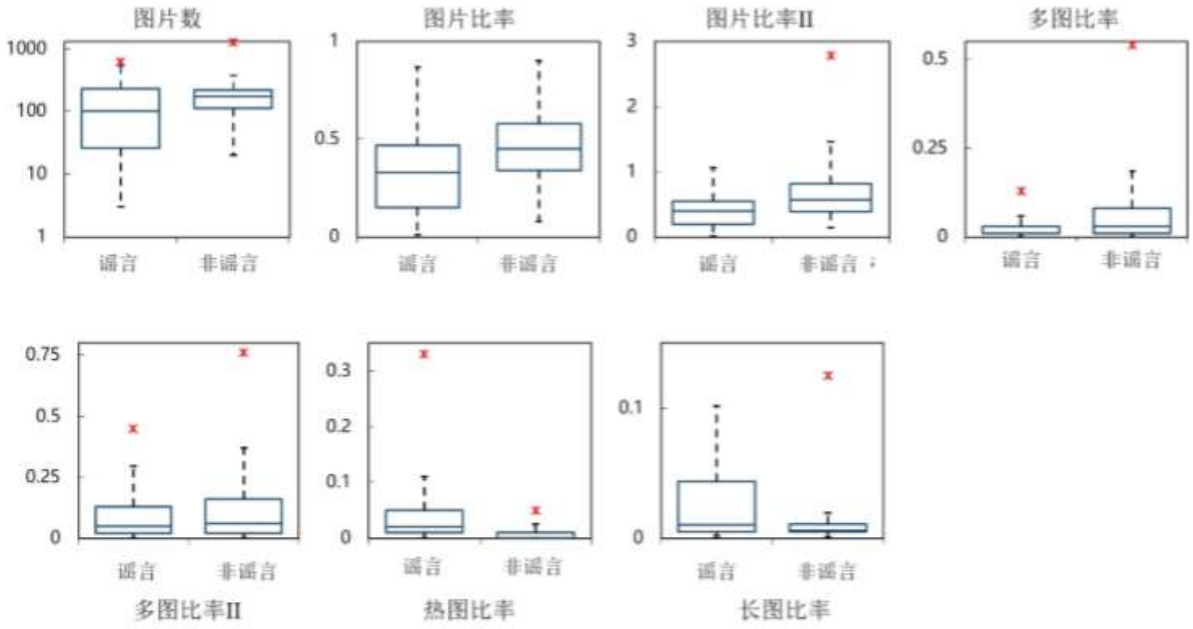
统计特征是对图像本身附带的属性特征进行分析，不对具体的图像内容作要求。例如：制造疑似谣言用户是否设置了头像（有无设置头像一定程度决定了用户的可信度）；疑似谣言信息是否具备图片(拥有图片一定程度会降低作为的谣言概率)等等。

（匿名用户怀疑度会提高）

特征	描述
图像数	事件包含的事件总数
图像数比率 I	包含图像的消息所占比率

表一 视觉统计特征基本类别

图像数比率 II	总图像数除以总消息数
多图比率 I	包含多图的消息所占比率
多图比率 II	包含多图的消息占所有包含图像消息的比率
热图比率	最热门的图像占去重后所有图像的比率
长图比率	包含长图的消息占所有包含图像消息的比率



图五 统计特征数据集分布

2. 内容特征(Content Feature)

基于内容的图片语义特征从视觉语义角度描述传统的基于内容的图片视觉特征从视觉语义的角度描述了图片内容，针对谣言检测这一任务，我们通常并不关心图片是否描述了某一特定对象或者场景，我们需要从区分谣言事件的角度分析图片在真假事件中不同的分布特点，通过观察真假不同事件中的热门图片，可以发现，真新闻里的图片更多，差异性更大,原创性强，而假新闻里，图片多样性更差，有时甚至包含合成伪造图片。因此，在视觉特性上，我们总结出下列 5 种可以用于描述图片内容特点的特征：

(1) 清晰度特征(Visual Clarity Score，VCS)

图像信息的原创性由两个图像集的分布差异来衡量。其中一个是指定新闻事件中的图片集（事件集），另一个是包含所有图片的全集。基于真实事件中包含大量原创图片的假设，如果一个事件集和全集中的图片分布差距很大，那么这个事件很有可能是真实事件。可以通过构建 2 个语言模型来计算这一特征，即分别对事件集和全集构建视觉词汇语言模型。该特征定义为两个视觉词汇语言模型之间的 KL 散度(Kullback－Leibler divergence)$D(p||q)=\sum p(x_i)log\frac{p(x_i)}{q(x_i)}$。

(2) 一致度特征(Visual Coherence Score, VCoS)

视觉一致度特征描述了同一事件中的图片是否具有一致性。相关的图片通常会具有相似的视觉外观，通过计算视觉一致度，能够量化出同一事件中的图片管理程度。这里定义视觉一致度为事件内任意图片对相似度的平均值 $\sigma=\frac{1}{C_n^2}(\sum_{\substack{0<i\le n\\ i<j\le n}}Similarity(Img(i),Img(j)))$

其中，以下列举 5 种常见的相似度计算方法：

a) 直方图匹配（Histogram Matching）

b) 感知哈希算法(Perceptual hash algorithm, PHA)

c) 尺度不变特征变换匹配（Scale Invariant Feature Transform，SIFT）

d) 结构相似性（Structural Similarity Index，SSIM）

e) 峰值信噪比（Peak Signal to Noise Ratio，PSNR）

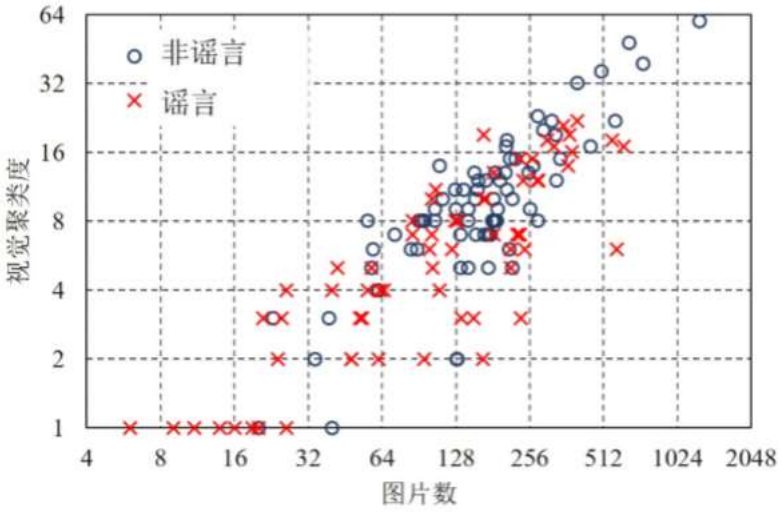
(3) 多样性特征(Visual Diversity Score)

视觉多样度度量了指定新闻事件图片集中的视觉上的差异程度．和视觉一致度相比，这个特征直接计算了图片的多样性分布特点，而且更加强调代表性的图片．我们定义一个图片的多样度为该图片到排在其之前的图片中的最小的距离．视觉一致度计算的是整个图片集上相似度的算术平均，而视觉多样度计算的是不相似度的加权平均即 diff=1-sim．在社会多媒体网络上，

可通过图片获得的转发量来排序图片．因此视觉多样度打分能够加重这些代表性图片的权重，减少事件中噪音图片的干扰。

(4) 聚类度特征(Visual Clustering Score, VClS)

视觉聚类度从图像聚类的角度衡量了事件 中图像的视觉分布特点。它被定义为图像集中聚类得到的类簇的个数。 本文采用分层聚合聚类算法（hierarchical agglomerative clustering, HAC）自底向上地将相似图像聚集成类。相比于其他聚类算法，如 K-means， 该算法不需要事先指定聚类个数，而能根据数据分布特点自动聚集出若干个类 簇。设定相同的参数下，该算法能够揭示出图像集的多样性特点。在聚类时，采用单连接策略来计算两个簇之间的相似度,即两个簇的距离定义为它们之间最近的两个对象的距离。基于图像的 GIST 特征，利用欧式距离计算图像的距离。移除掉大小小于 3 的小类簇后，本文将剩下的簇的个数记做视觉聚类度。



图六 事件图像数与视觉聚类度的关系

图七 当年引发争议的“和平鸽广场”获奖照片伪造案件

(5) 篡改性特征

随着图像合成处理技术的发展和成熟，让照片按照人的意愿的修改已经不是问题。对于照片的伪造性可以通过图像取证技术进行检测，其主要方法有：遍历搜索法、图像块自相关矩阵法、图像块匹配法等等

遍历搜索结合图像块匹配判断照片真实性：

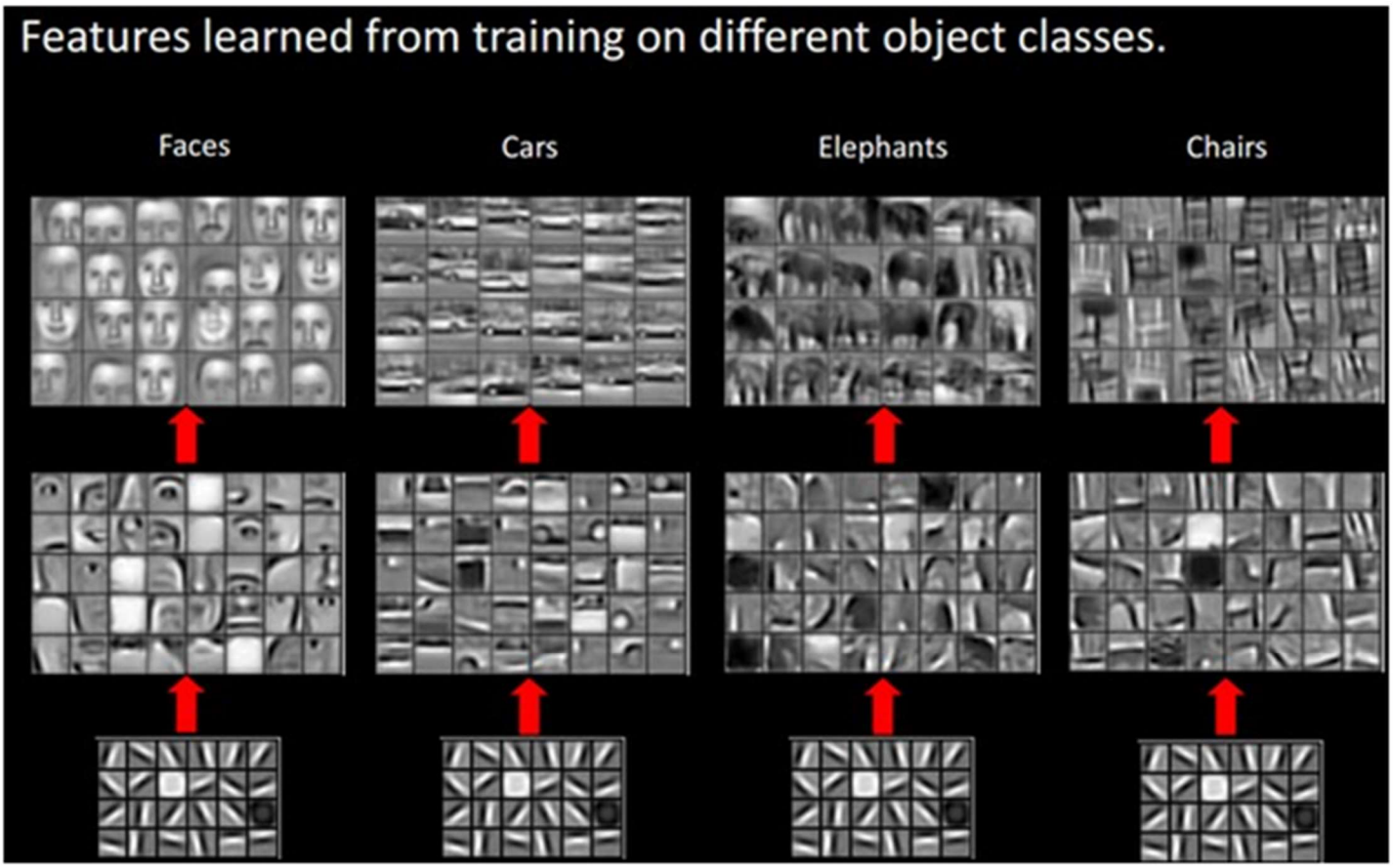
在网络上自动搜集与待判断照片相似度高的图像，对比该照片与其他图像，在关键细节上寻找篡改痕迹。

图像自相关判断原理

对构成待测图像的各个组成元素进行识别，并测试各元素之间的相关性。例如 ps 过的图片中同时有埃及金字塔和埃菲尔铁塔，判断埃及金字塔与埃菲尔铁塔地理位置存在空间上距离，判断两者出现在同一张照片的可能性极低，为篡改过的照片的嫌疑较大。

3. 深度学习(Deep-Learning)特征

随着近几年人工神经网络理论的提出和深度学习的快速发展，以卷积神经网络（CNN）为代表的一系列人工神经网络算法在视觉检测、分类上有着广泛的应用。CNN 通过一系列卷积滤波操作，提取图像中轮廓、角点、线条等特征，其效果和效率都要高于传统手工提取图像特征的方法。设有两个 m*n 函数 f(x, y)和 g(x,y)，其离散卷积（Discrete convolution）表示为 $f(x,y) * g(x,y) =$

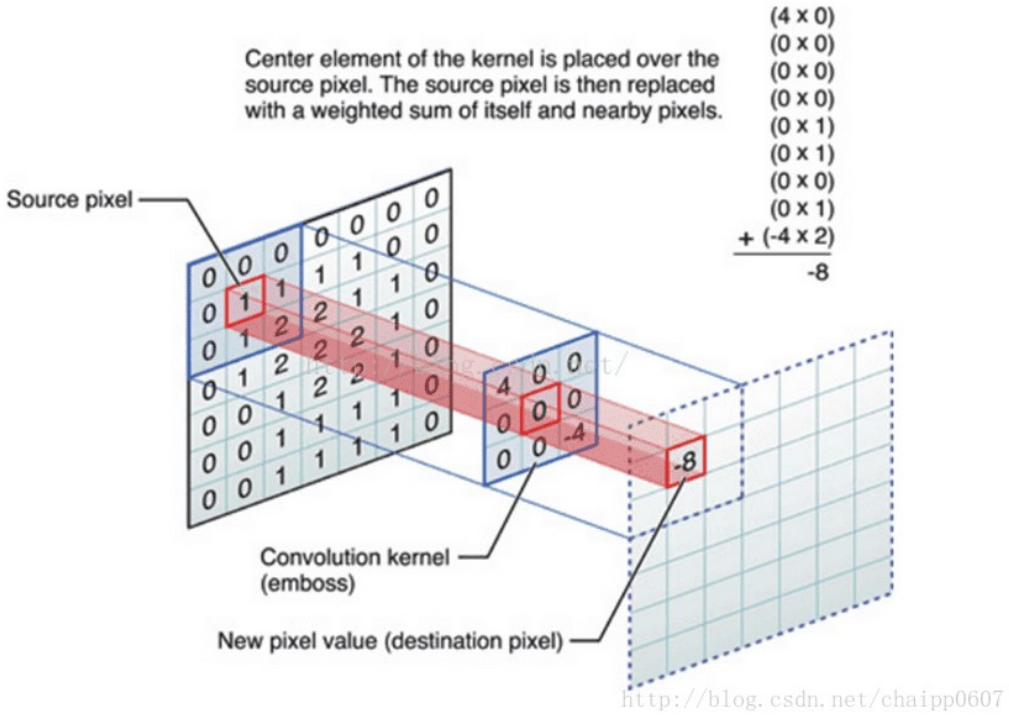


图八 卷积神经网络特征简图

$\frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1} f(i,j)g(x-i,y-j)$。

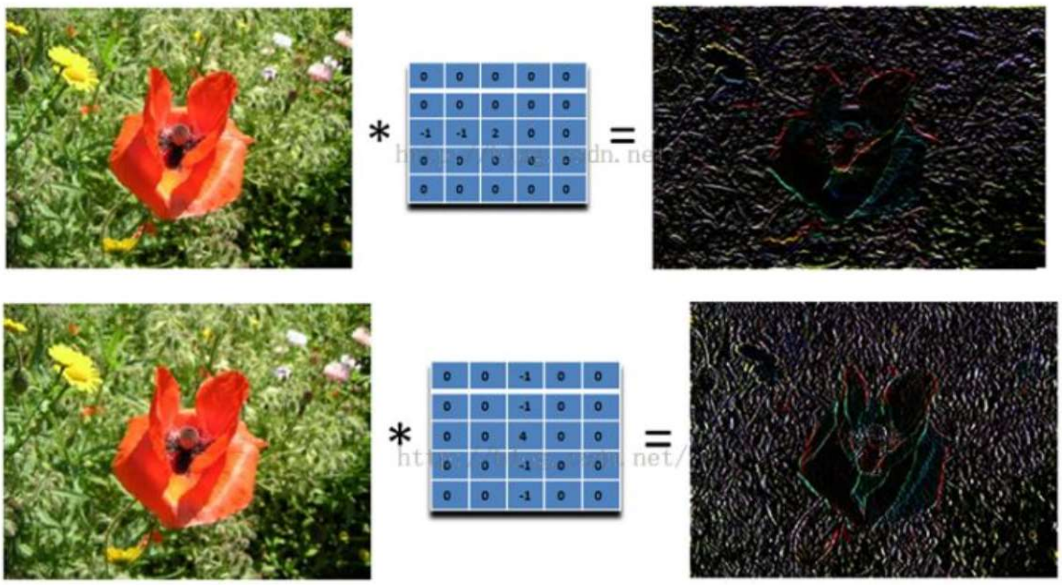
数字图像是一个二维的离散信号，对数字图像做卷积操作其实就是利用卷积核（卷积模板）在图像上滑动，将图像点上的像素灰度值与对应的卷积核上的数值相乘，然后将所有相乘后的值相加作为卷积核中间像素对应的图像上像素的灰

度值，并最终滑动完所有图像的过程。



图像卷积运算示意图

利用不同的卷积核进行卷积操作就可以得到所要的不同特征。例如：



图九 边缘特征提取

以上两图均为边缘检测的例子，用于提取边缘特征。

2.3.3. 社交特征

广泛的互联性是社会多媒体的一大特点，其中主要有三方面的互动：一是用户与用户之间的互动，包括“加好友”、“群聊”等等；二是媒体内容的互动，不同媒体信息通过超链接串联在一起，构成一个错综复杂的信息网络；三是用户与媒体内容的互动，包括“转发”、“点赞”等一系列操作。通过这些互联网络，我们也可以得到用于检测谣言的丰富特征信息，本文主要概括为用户特征和传播特征：

1.用户特征(User Features)

用户特征是指从用户社交网络抽取的或者是从单个用户信息抽取的特征。谣言传播过程中，可能存在大量“水军”推波助澜，或者一些恶意账户故意捏造、传播。不同类型的账户对大众具有不同的可信度，例如“官方认证用户”、“媒体”相比普通用户更有信赖感。因此抽取基于用户的特征能够帮助提高谣言检测结果。基于用户的社交特征是指描述用户在社交网络中传播信息时展现出来的特点。从特征的粒度看，这些特征可以分为两大类：个体特征和群组特征。

1）个体特征

个体特征是指针对单个用户的各项统计指标中抽取出来，用来描述该特定用户的一系列特征。主要包括从用户个人信息抽取的特征，如注册时间、用户名类型、年龄和性别等，以及从用户的历史交互抽取的特征， 如粉丝数、关注数和发布微博数等，同时也包括两个描述用户发消息行为的特征：“客户端”（“client”）这个特征记录了用户发布消息时使用的是哪种类型的客户端平台：手机或是电脑；“地点”这个特征记录了消息发布的地点是否与消息描述的事件地点一致。除此之外，用户的历史信用记录可以反映用户的信誉程度，影响该用户微博信息的可信度，可以作为检测微博谣言的特征。用户的个人信息首页记录了该用户的信用信息。

图一显示了当前用户的信用积分为 80 分、信用等级为正常，但从信用历史记录中可以发现该用户曾经发布过多条不良微博。但是必须考虑到新浪微博制定的信用恢复规则，即当用户因发布不良微博扣除信用积分后，在之后的若干天时间内如若没有违规则恢复信用积分，所以不能根据当前的信用积分作为用户历史信用特征值，但是可以通过统计该微博发布之前出现不良历史记录的条数作为用户历史信用特征值，该特征值影响该微博发布的可信度。

图 用户的个人信用信息

计算用户历史信用特征值的一个具体方法如下所示。

算法 计算用户历史信用特征值的算法流程：

输入：微博用户的信用信息记录。

输出：用户历史信用特征值 CreditHistory。

```
1 for eachHistory in Historys:
2 if eachHistory 时间小于发布该微博的时间:
3 if eachHistory 因发布不良微博:
4 CreditHistory + = 1
```

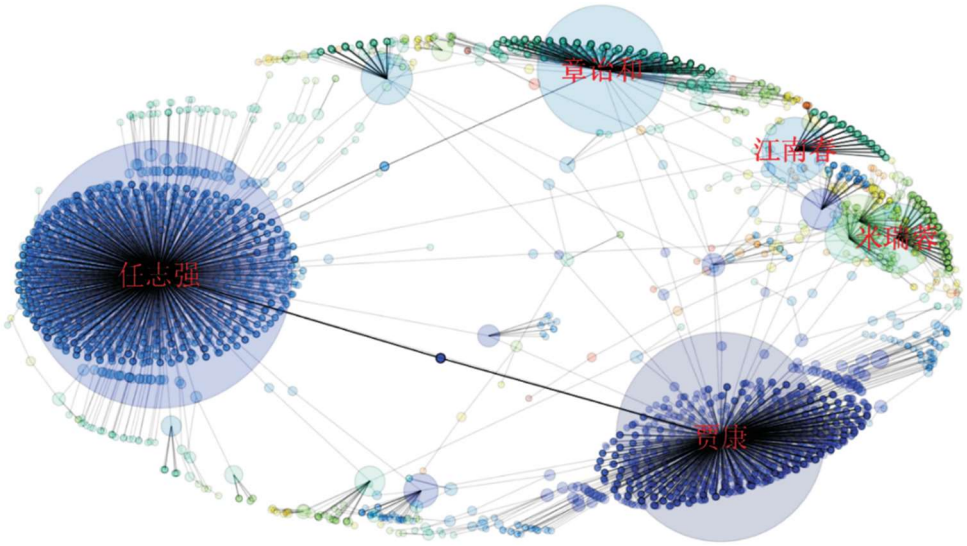
2）群组特征

群组特征描述的是在信息传播过程中具有相似性的某个用户群体的整体特征。抽取该类特征时的一个基本的假设就是传播谣言的社区和传播真实消息的社区各不相同并且有不同的特点。群组特征通常是从个人特征聚合而来，例如认证用户的比例、平均粉丝数、用户平均注册时间、用户平均年龄等。

2.传播特征(Propagation Fratures)

传播特征指的是信息在传播过程中产生的能够用来衡量信息真实性的特征。例如：评论数与当月所有微博平均评论数之差、转发数与当月所有微博平均转发数之差、评论频率(评论数/微博生命周期)、转发频率(转发数/微博生命周期）、用户参与度(评论数/(评论数+转发数））、意见领袖传播影响力等等。

1）意见领袖传播影响力特征



图二 寻找被拐儿童谣言微博传播过程图

图二是一张寻找被拐儿童谣言微博传播过程图。从图中可以看出，谣言发布者贾康发布该微博后，意见领袖任志强转发了该微博，导致该微博的热度呈现爆发式增长的趋势。故把微博用户分为实名认证的用户和普通用户两类。实名认证用户往往是具有一定知名度或影响力的个人或者企业。通过计算由实名认证用户传播给用户的微博数与该微博的转发总数的比值作为意见领袖在传播过程中的影响力。具体算法如下。

算法 1　计算微博传播树特征值的算法

输入：微博传播树。

输出：意见领袖传播影响力特征值即经过实名认证用户转发微博的比例 AuthenticatedForwardRatio。

```
for eachNode in nodes:
    计算 eachNode 的父节点以及 eachNode 的子节点数
for i in max_depth:
//从顶层开始
    for eachNode in levelNodes [i]:
    //第 i 层节点
        if eachNode 为实名认证用户:
            将 eachNode 的子节点数加入到 AuthenticatedFor-
            wards
            从 levelNodes [i + 1] 中剔除 eachNode 的子节点
AuthenticatedForwardRatio = AuthenticatedForwards / 总节点数
```

2）评论数差值和转发数差值特征

通过大数据分析，当用户发布一条谣言微博时，谣言微博的评论数和转发数在通常情况下回远远高于其前一个月所有微博的平均评论数和转发数，由此可以说明谣言微博更容易引发用户的关注和评论，并迅速地蔓延起来。评论数和转发数虽然不能作为检测谣言的特征，但是可以根据微博的评论数和转发数与用户历史微博的评论数和转发数的差值作为谣言检测的特征，具体算法如下。其中：commentDiff 表示微博评论数差值；Ncom 表示每条微博的评论数；Ncom 表示历史微博的平均评论数；forwardDiff 表示微博转发数差值；Nfor 表示每条微博的转发数；Nfor 表示历史微博的平均转发数。当特征的数值范围太大时，即使进行归一化后再用于分类，也会影响分类器识别的准确率，所以当微博的评论数和转发数与历史微博的评论数和转发数差值较大时，通过 log 函数来降低差值的范围。

$$\text{commentDiff}=\begin{cases}\left|\dfrac{N_{\text{com}}-\overline{N}_{\text{com}}}{N_{\text{com}}}\right| & \left|\dfrac{N_{\text{com}}-\overline{N}_{\text{com}}}{N_{\text{com}}}\right|<1\\ \log\left(\left|\dfrac{N_{\text{com}}-\overline{N}_{\text{com}}}{N_{\text{com}}}\right|\right) & \text{其他}\end{cases}\tag{7}$$

$$\text{forwardDiff}=\begin{cases}\left|\dfrac{N_{\text{for}}-\overline{N}_{\text{for}}}{N_{\text{for}}}\right| & \left|\dfrac{N_{\text{for}}-\overline{N}_{\text{for}}}{N_{\text{for}}}\right|<1\\ \log\left(\left|\dfrac{N_{\text{for}}-\overline{N}_{\text{for}}}{N_{\text{for}}}\right|\right) & \text{其他}\end{cases}\tag{8}$$

2.4. 特征融合(Feature Fusion)

前面我们已经详细介绍数据采集和特征提取，我们还需要对各式各样的特征进行整合，称为特征融合，特征融合是谣言检测过程中的一个重要环节，通过融合确定各特征对预测函数的贡献权重，使每个特征能够更好发挥其作用，达到特征之间的互补。主流的多媒体谣言检测特征融合算法有两类：一类是简单融合包括前融合和后融合；另一类是基于神经网络的多模态特征融合方法：

1. 简单特征融合算法

（1） 前融合(Early Fusion)

概念： 前融合是在特征级上对不同模态的特征进行直接融合，即在学习检测算

法之前，对各种特征组合形成统一的特征表示后，再输入分类检测算法进行学习。现有的谣言检测研究中通常采用前融合的方式进行特征融合，常用的前融合方法包括基于串联的方法、基于降维的方法和基于核的方法。

优点： 融合方式简单，能够融合来自多个不同模态的特征，而且只需要学习一次分类器模型就可以实现融合，避免了针对各模态特征的多次学习。

缺点： 由于组合后的特征通常是一个维数更高的特征，容易导致学习检测模型时出现维数灾难的问题。简单的融合方式通常不能够发现不同模态之间的关系，导致最终的检测模型倾向于在某一类特征上起作用。由于不同特征描述所在的特征空间不同，直接的前融合很难协调好不同模态特征、充分发挥每种特征的作用。

（2） 后融合(Late Fusion)

概念： 后融合是在模型输出级对不同特征的检测结果进行融合，即对各组特征分别学习检测模型，最后对各组特征的结果进行融合。常用的后融合算法可以分为：有监督的后融合和无监督的后融合。 有监督的后融合方法以各检测模型的输出结果作为特征，再训练一个检测模型（如 SVM，逻辑回归等）来融合各个分类器的输出结果。这种方法的优点是能够发现不同特征组训练的模型之间的非线性关系。但是，由于融合的检测模型的性能取决于训练数据需要独立于训练各个特征组模型的训练数据，否则会产生 “过拟合”（over-fitting）问题。无监督的后融合方法采用简单的算术运算来完成 对各模型结果的融合。常用的方法包括：最大值融合、最小值融合、算术平均融 合和几何平均融合等。

优点： 避免维数灾难问题、发挥每组特征作用。

缺点： 但是需要进行多次学习，计算消耗较大。

2. 基于神经网络的特征融合算法

（1） 视觉问答特征融合算法（VQA）

视觉模型：利用卷积神经网络学习和抽取图像的视觉特征。即原始图像经过预训练的 VGG 网络得到原始的图像特征后，再 经过连接的两个全连

接视觉网络层得到最终的视觉特征表示。

语言模型：利用卷积-递归神经网络（CNN-RNN）来学习文本特征。即对原文本语句中的每个词项做原始的 one-hot 特征表示，经过一层神经网络变换后，依次输入 CNN-RNN 网络，该网络中的卷积神经网络（CNN）部分对词组单元建模，递归神经网络（RNN）部分对词项序列进行建模，最 终得到表示整句话的语言特征表示。

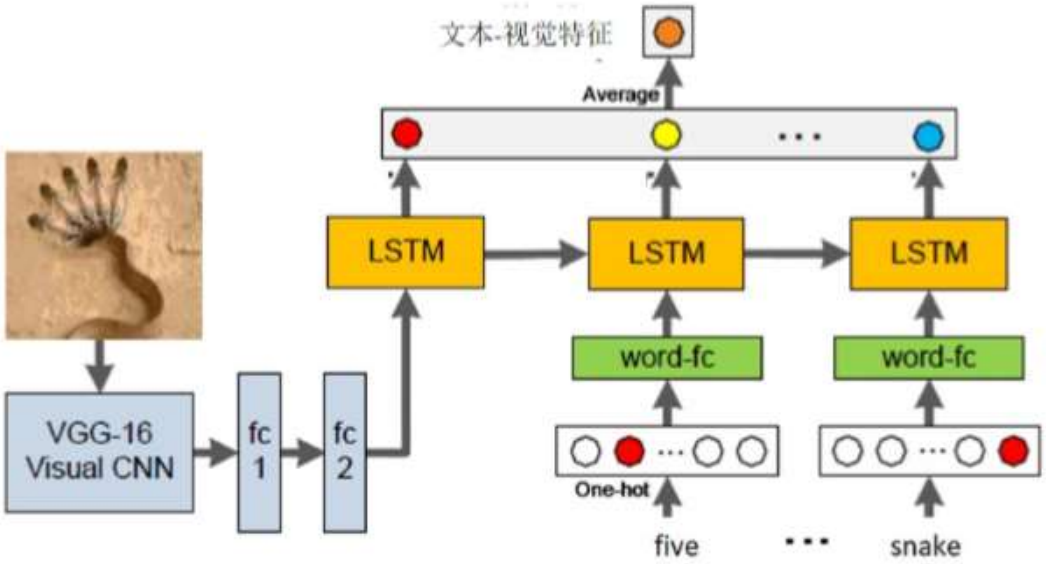
特征融合：将视觉特征和语言特征进行元素级别的相乘，得到一个表示文本-视觉相匹配的特征值，最终实现了文本-视觉特征的对齐融合。

（2） 图像自动标题生成算法（NeuralTalk）

视觉模型：同样也利用卷积神经网络 VGG-16 学习和抽取图像的视觉特征。

语言模型：利用递归神经网络 LSTM 来学习文本特征。

特征融合：将视觉特征作为第一项输入 LSTM 与文本语句一起建模，这种方式

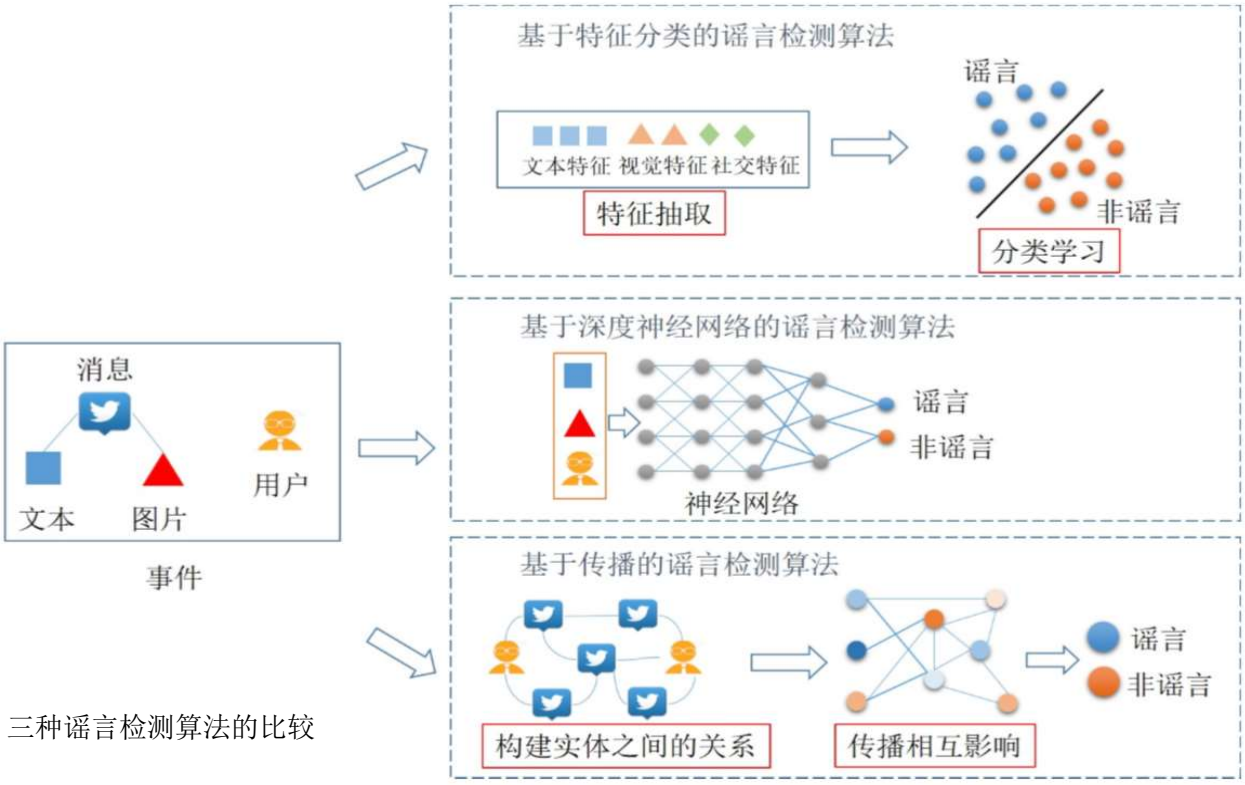


图十 NeuralTalk 网络结构示意图

得到的特征既包含原有的文本序列特征又包含了第一项的视觉特征。最终的融合特征取 LSTM 每步的输入特征取平均值得到。

2.5. 检测算法

谣言检测事实上是一个二分类问题，谣言预测函数可表示为给定事件 e 及其包



图十一 三种谣言检测算法的比较

含的消息集合 M 和用户集合 U，谣言检测的目标是判定该事件是否能够被证实是真实事件或者是虚假事件，即学习到一个预测函数 F(e) →{0,1} 满足以下要求： F(e) ={if 1 e 是谣言,if 0 e 是非谣言}。本文介绍的谣言检测算法有基于特征分类、基于深度神经网络、基于传播模型三种，下面简单介绍一下这三种方法：

1. 基于特征分类的谣言检测算法

这种方法主要通过传统机器学习算法，通过人工提取特征，利用监督学习算法，实现二分类。主要流程为：

(1) 特征提取：每个类别分别提取特征向量。

(2) 训练模型：利用已标注的数据集，训练基于特征提取的分类器。

(3) 模型预测：通过训练好的模型，预测新数据并评估算法预测效果。

当今传统机器学习的算法种类繁多，其中使用较多的有决策树、支持向量机、K 邻近算法、贝叶斯网络、逻辑回归还有强化分类器的 AdaBoost 等，下面简单介绍一下几个传统机器学习分类的例子：

（1） 决策树（Decision Tree）

概念：在已知各种情况发生概率的基础上，通过构成决策树来求取净现值的期望值大于等于零的概率，评价项目风险，判断其可行性的决策分析方法，是直观运用概率分析的一种图解法。

算法流程：

a. 特征选择

b. 决策树生成

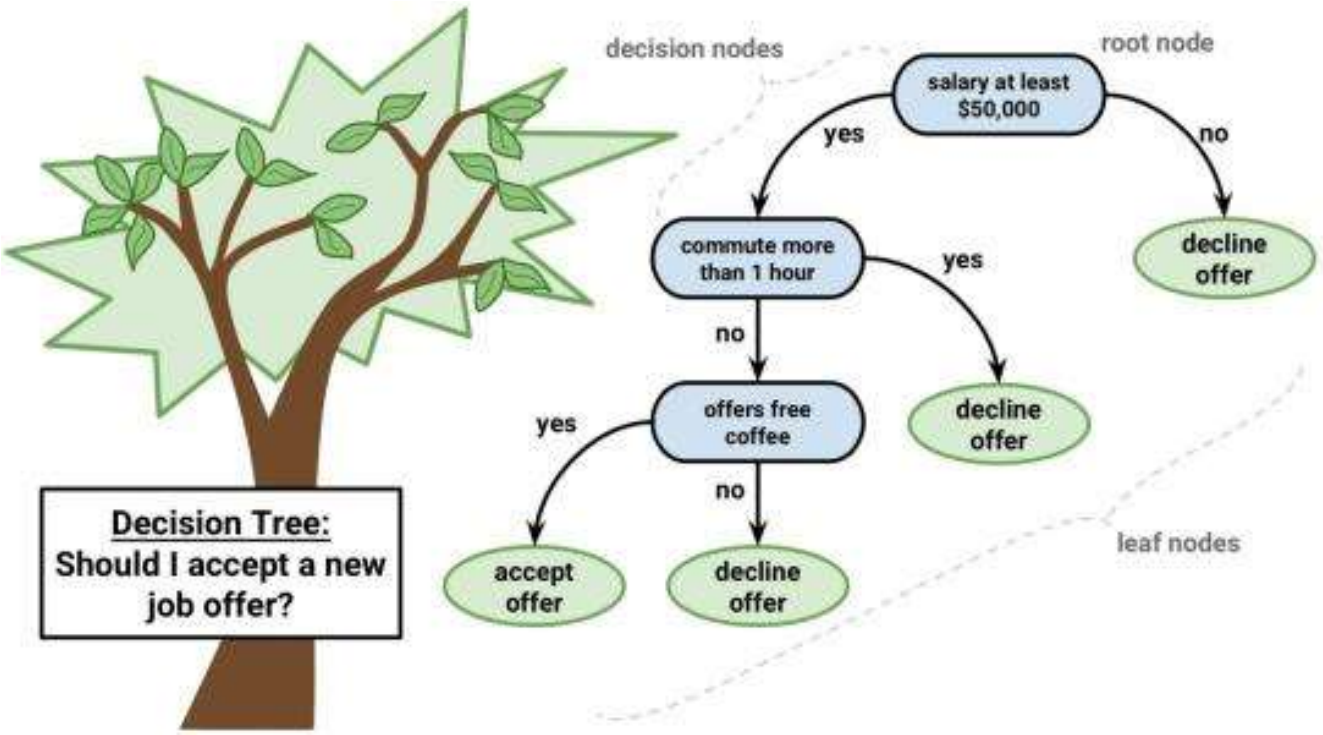
c. 决策树的修剪

优点：易于理解和实现；能够同时处理数据型和常规型属性，在相对短的时间内能够对大型数据源做出可行且效果良好的结果；易于通过静态测试来对模型进行评测，可以测定模型可信度。

缺点：对连续性的字段比较难预测；对有时间顺序的数据，需要很多预处理的工作；当类别太多时，错误可能就会增加的比较快；一般的算法分类的时候，只是根据一个字段来分类。

主要适用范围：决策过程应用较多。

示意图：



（2） 支持向量机（Support Vector Machine，SVM）

概念：建立在统计学习理论的 VC 维理论和结构风险最小原理基础上的，根据有限的样本信息在模型的复杂性（即对特定训练样本的学习精度）和学习能力（即无错误地识别任意样本的能力）之间寻求最佳折中，以求获得最好的推广能力。（简单将就是将低维线性不可分的数据映射到高维空间中使其变成线性可分）

算法流程：

a. 原始问题的分析与变形：将原始问题转化成凸优化问题。
b. 原始问题对偶化：构建拉格朗日函数并选择核函数，将原始函数对偶化。
c. 利用 KKT 条件求解决策函数

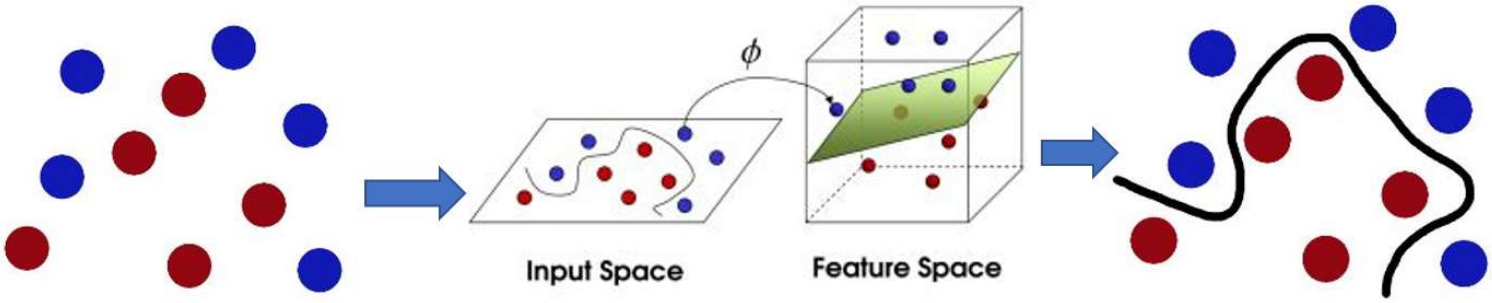
优点：避开求解高维空间的复杂性，直接用此空间的内积函数（既是核函数），再利用在线性可分的情况下的求解方法直接求解对应的高维空间的决策问题. 当核函数已知，可以简化高维空间问题的求解难度。而且支持向量机比神经网络具有较好的泛化推

广能力。

缺点：求解核函数的困难。即使确定核函数以后，在求解问题分类时，需要求解函数的二次规划，这就需要大量的存储空间。

主要适用范围：二分类问题。

示意图：



（3） K 邻近算法(k-Nearest Neighbour，KNN)

概念：如果一个样本在特征空间中的 k 个最相邻的样本中的 大多数属于某一个类别，则该样本也属于这个类别，并具有这个类别上样本的特性。

算法流程：

a. 实验数据预处理：包括特征提取、选用合适的数据结构存储训练数据和测试元组以及设定参数。

b. 训练数据：

维护一个大小为 k 的按距离由大到小的优先级队列，用于存储最近邻训练元组。随机从训练元组中选取 k 个元组作为初始的最近邻元组，分别计算测试元组到这 k 个元组的距离，将训练元组标号和距离存入优先级队列。遍历训练元组集，计算当前训练元组与测试元组的距离，将所得距离 L 与优先级队列中的最大距离 Lmax 进行比较。若 L>=Lmax，则舍弃该元组，遍历下一个元组。若 L < Lmax，删除优先级队列中最大距离的元组，将当前训练元组存入优先级队

列。遍历完毕，计算优先级队列中 k 个元组的多数类，并将其作为测试元组的类别。

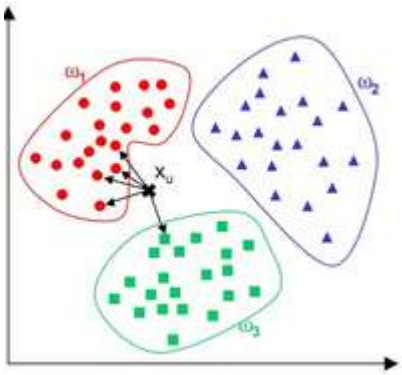
c. 评估：测试元组集测试完毕后计算误差率，继续设定不同的 k 值重新进行训练，最后取误差率最小的 k 值。

优点：简单，易于理解，易于实现，无需估计参数，无需训练；适合对稀有事件进行分类；特别适合于多分类问题(对象具有多个类别标签)， kNN 比 SVM 的表现要好。

缺点：计算量较大、可理解性差、样本不平衡时会导致准度降低。

主要适用范围：聚类分析、多分类问题。

示意图：



（4） 贝叶斯网络(Bayesian network)

概念：一个贝叶斯网络是一个有向无环图，由代表变量节点及连接这些节点有向边构成。节点代表随机变量，节点间的有向边代表了节点间的互相关系(由父节点指向其子节点)，用条件概率进行表达关系强度，没有父节点的用先验概率进行信息表达。

主要适用范围：有条件地依赖多种控制因素的决策，可以从 不完全、不精确或不确定的知识或信息中做出推理。

（5） 逻辑回归(logistic regressive)

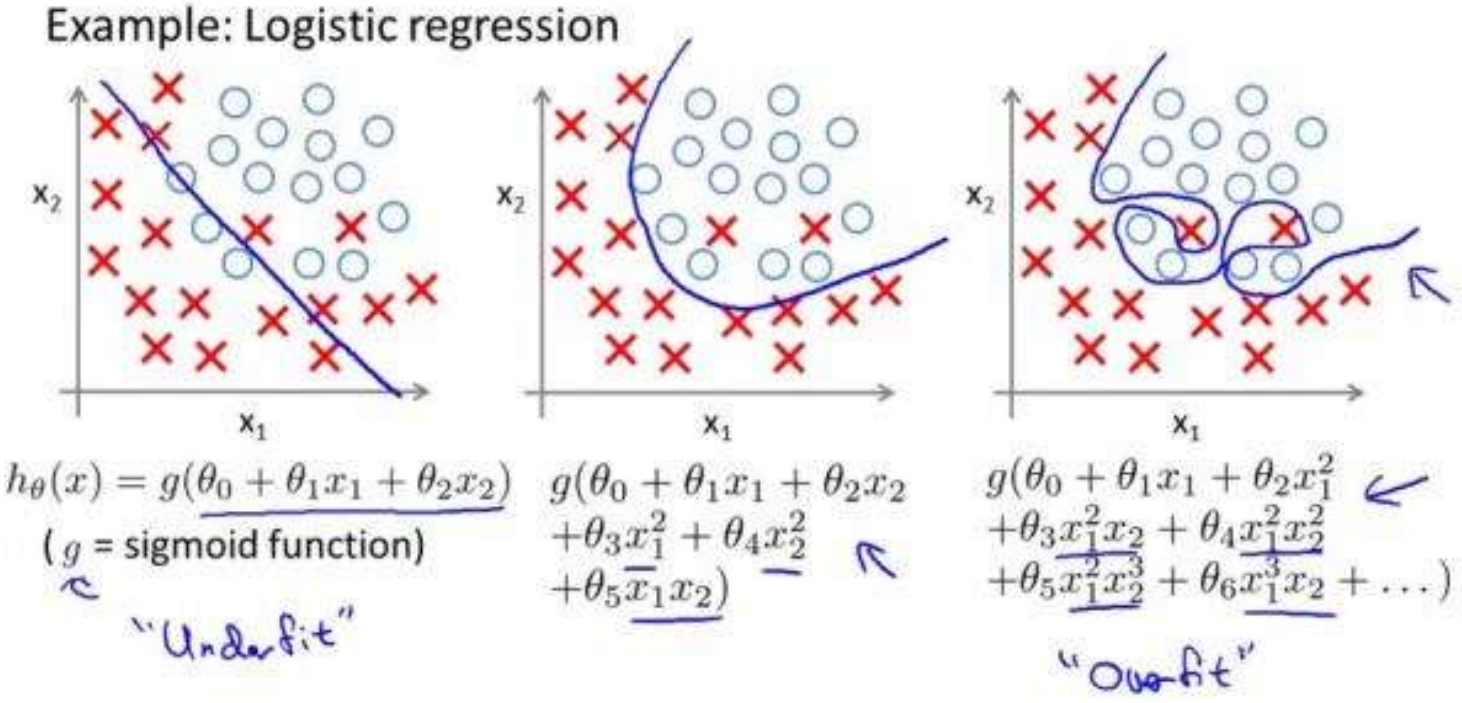
概念：广义的线性回归算法

优点：预测结果是界于 0 和 1 之间的概率；可以适用于连续　性和类别性自变量；容易使用和解释。

缺点：对模型中自变量多重共线性较为敏感。预测结果呈“S”型，因此从 log(odds)向概率转化的过程是非线性的，在两端随着 log(odds)值的变化，概率变化很小，边际值太小，slope 太小，而中间概率的变化很大，很敏感。导致很多区间的变量变化对目标概率的影响没有区分度，无法确定阀值。

主要适用范围：分类问题。

示意图：



（6） AdaBoost

概念：AdaBoost 是一种迭代算法，其核心思想是针对同一个训练集训练不同的分类器(弱分类器)，然后把这些弱分类器集合起来，构成一个更强的最终分类器(强分类器)。其算法本身是通过改变数据分布来实现的，它根据每次训练集之中每个样

本的分类是否正确，以及上次的总体分类的准确率，来确定每个样本的权值。将修改过权值的新数据集送给下层分类器进行训练，最后将每次训练得到的分类器最后融合起来，作为最后的决策分类器。

算法流程：

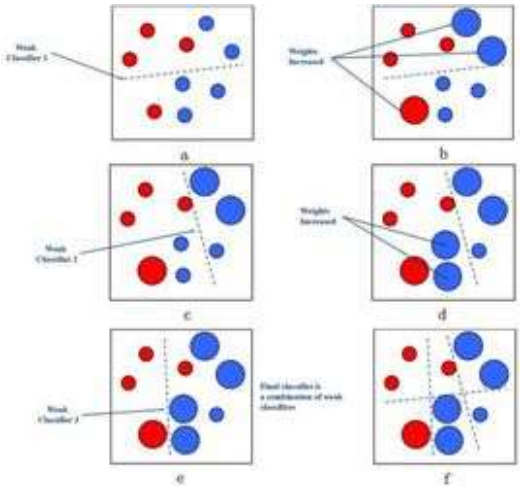
a. 先通过对 N 个训练样本的学习得到第一个弱分类器
b. 将分错的样本和其他的新数据一起构成一个新的 N 个的训练样本，通过对这个样本的学习得到第二个弱分类器 ；
c. 将 1 和 2 都分错了的样本加上其他的新样本构成另一个新的 N 个的训练样本，通过对这个样本的学习得到第三个弱分类器
d. 最终经过提升的强分类器。即某个数据被分为哪一类要由各分类器权值决定

优点：很好的利用了弱分类器进行级联；可以将不同的分类算法作为弱分类器；具有很高的精度。

缺点：AdaBoost 迭代次数也就是弱分类器数目不太好设定，；数据不平衡导致分类精度下降；训练比较耗时。

主要适用范围：分类问题。

示意图：

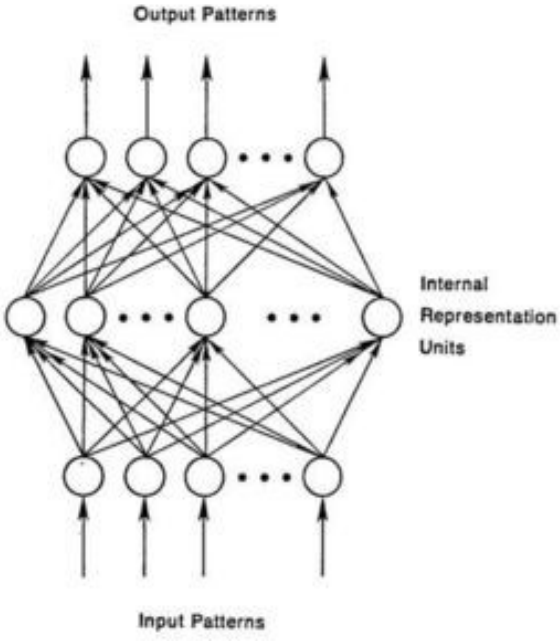


2. 基于深度神经网络的谣言检测算法

近年来随着神经网络算法的迅速发展，越来越多的研究转向深度学习方向，利用深度神经网络可以代替原始的人工提取特征。目前也有科学家将深度学习技术应用到多媒体谣言检查中来，现有的神经网络算法大致分为基于递归神经网络和基于卷积神经网络的两种检测算法：

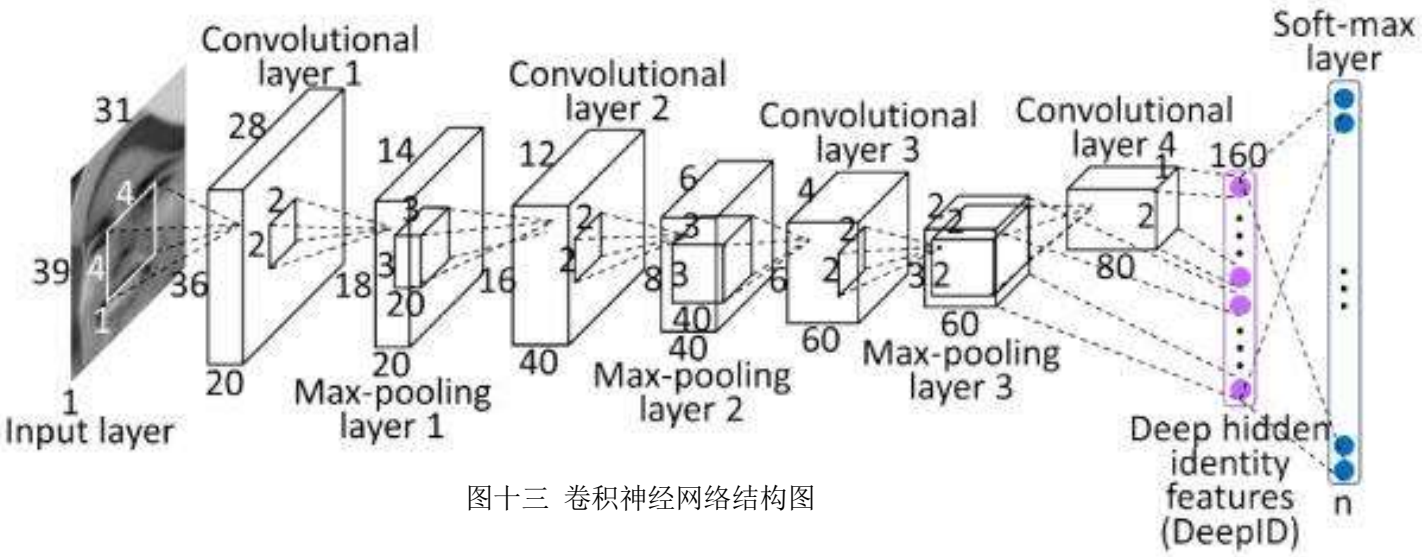
（1） 基于递归神经网络(RNN)

递归神经网络 (Recurrent Neural Network, RNN) 是一种前向传播神经网络模型，它能够对变长的序列信息进行建模。基于递归神经网络的谣言检测方法充分利用了递归神经网络对序列输入建 模的能力，首先将原始谣言数据根据时间、主题等特性转换为一个序列数据，再 输入递归神经网络进行建模学习。这类方法成功的关键在于，递归神经网络的一 个序列连接中的每个神经元保留了之前神经元的信息，这样整个网络就能尽可能 地提取到谣言传播过程中的动态变化的各种特征。



图十二 递归神经网络示意图

（2） 基于卷积神经网络(CNN)



图十三 卷积神经网络结构图

前面我们再视觉特征提取部分已经提到过卷积操作和卷积神经网络算法，卷积神经网络在谣言检测技术中也是十分有用。与一般神经网络模型相比，它利用数据的上下文关系 来减少需要学习的参数数目，从而提高一般后向反馈算法的训练性能；同时，通过共享网络权值的方法，极大的减少了自由参数学习的数量从而提高了学习的效率。深度卷积神经网络不仅能够从输入样例中自动地抽取到局部和全局的重要 特征，而且能够抽取出输入样例的高层特征。传统的递归神经网络倾向于受到最 新一个输入序列节点的影响，而且目标是抽取到在整个序列中保持不变的特征。 与之相比，卷积神经网络利用池化的结构能够灵活地抽取到分散在所有输入中的各类特征，而不是仅局限于全局的或当前的最新输入。

3. 基于传播模型的谣言检测算法

网络多媒体信息常常存在的联系，基于单个样例的特征往往不能充分利用这些联系进行谣言检测，基于传播模型的谣言检测算法就是从整体上评估各社交网络信息的可信度通过建模研究信息传播网络特点来反馈消息的可信度。该算法主要由两部分构成，一是网络模型的构建，二是传播模型的计算：

（1） 构建网络模型

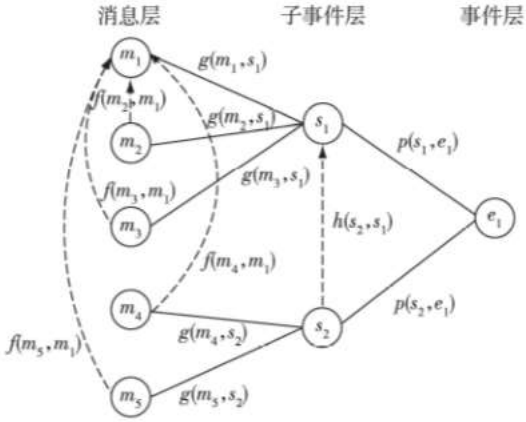
首先找出影响事件可信度的各种对象，例如消息、用户和话题等。根据它们 在社交网络上的交互关系和内容上的语义关系构建出彼此

之间的联系。例 如，消息与其发布者之间连边、消息与对应的话题进行连边。同时按照一定方法，如语义相似度，定义各连边的权值。网络中的每个节点都被赋予 一个可信度值。通常可以使用特征分类的结果来初始化网络可信度值。这 里可信度是对一个事件或消息属于谣言的程度的概率量化，即有多大可能性是属于非谣言，可信度越低的对象越有可能是谣言。

（2） 传播算法

基于某种针对网络结构或者问题定义的假设设计传播算法。可信度网络上的各种实体的可信度值在一定约束条件下，基于该算法彼此互相影响和传播，直至收敛产生最终的可信度评估值。不同实体的初始可信度值可以通过基于分类的方法学习得到，因此该算法可以看作是对单样例分类方法的优化改进，往往比简单的分类方法具有更好的认证 准确率和稳定性。

设计可靠、合理的传播算法是基于传播的谣言检测算法的关键。不同对象的可信度初值在内容网络上的传播过程可以看作是一种半监督的网络学习模型。作为一种有效的图优化学习方法，半监督图学习的理论已被广泛的研究和应用。该类算法的目标是在保持已有标注数据和网络结构一致性的前提下，预测未标注数据的类别。构建内容网络、在网络上进行可信度传播，能够弥补传统基于特征的单例分类方法的不足。



图十四 三层内容网络

本文简单介绍选自参考文献【1】中所提到的分层可信度传播网络：

对于一个新闻事件来说，一个分层的内容网络由 3 层网络(消息层、子事件层和事件层)以及它们之间的边组成.如图 6 所示，该网络中有 3 种在上节中定义的实体:消息 m、子事件 s 和事件 e。以及 4 种类型的边:消息到子事件之间的边、子事件到事件之间的边、消息之间互联的边以及子事件之间互联的边。各边的权重都定义为

该边 2 个定点的函数.通过子事件聚类，消息连接到对应的子事件。

该网络中各类型的边权重计算方法如下：

1）消息-消息：利用 Jaccard 系数计算 2 条消息之间的情感极性，定义不同情感倾向的消息之间的边权值为 0，相同情感倾向的消息之间的边权值正比于 2 条消息的内容相似度。

2）子事件-子事件：2 各聚类中心之间的余弦距离即子事件的关联度。

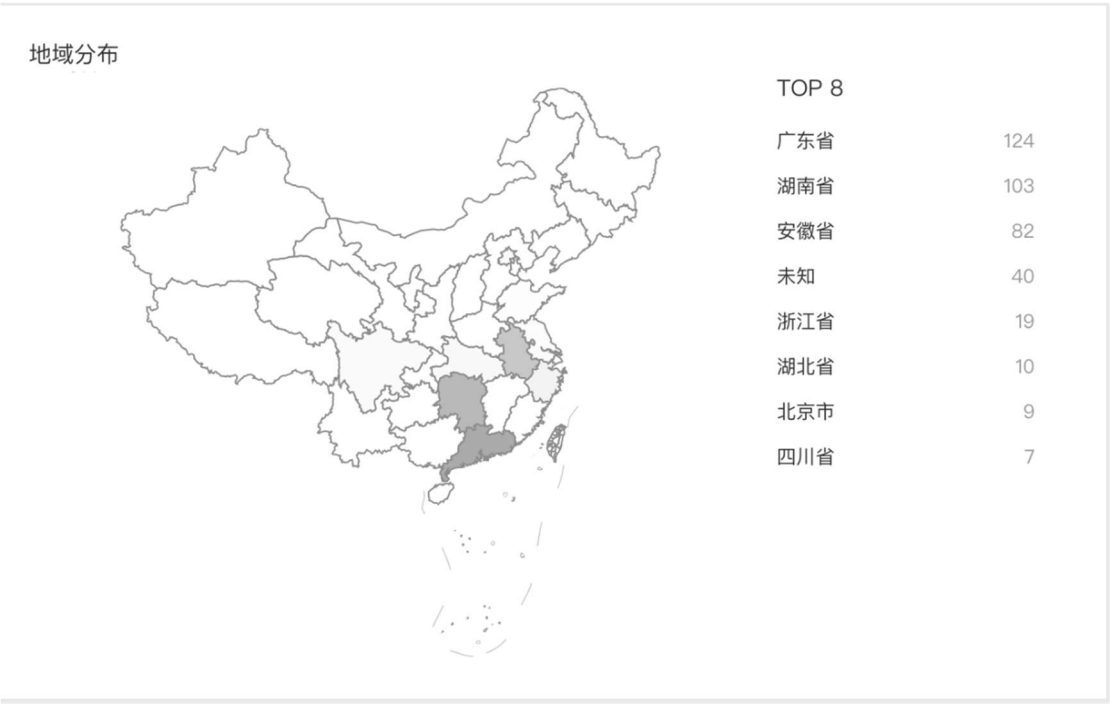
3）消息-子事件：定义一条消息对所在子事件的影响来自 2 个方面：一是消息与子事件的一致程度，可以通过文本相似度来刻画；二是消息在子事件中的重要程度，可以通过媒体转发量来刻画。

4）子事件-事件：同“消息-子事件”，由相似度与重要程度来刻画。

通过把不同实体在该分层网络上的可信度传播过程定义为一个图优化问题，定义损失函数后，利用梯度下降法解得函数的迭代解，从而得到各实体的最终可信度值。

3. 多媒体谣言检测技术的民意调查情况

（1） 数据基本来源



（2） 基本调查内容及分析

4.您怎样看待网络谣言？		
选项	小计	百分比%
宁可信其有，不可信其无，防范于未然	52	11.2
大部分群众科学素养较低，对谣言缺乏判别能力	226	48.8
不是群众的错，政府和媒体没有及时尽到破除谣言的责任，让群众缺乏安全感	143	30.9
谣言止于智者	282	60.9
我对谣言真假有良好的分辨能力	161	34.8
谣言在一定程度上代表了民意	56	12.1
其他___	8	1.7

5.您了解多媒体谣言检测技术吗？		
选项	小计	百分比%
完全不了解	304	65.7
听说过	145	31.3
比较了解	12	2.6
很了解	2	0.4

6.您能否接受自己在以微博为代表的社会媒体上发布的文章图片（原创）等被提取检测？		
选项	小计	百分比%
能	190	41
不能	65	14
视内容而定	208	44.9

7.您能否接受自己在以微博为代表的社会媒体上的行为（例如转发/评论/点赞等）被提取检测？		
选项	小计	百分比%
能	331	71.5
不能	132	28.5
8.您能否接受微博（或其他社交媒体）实名制？		
选项	小计	百分比%
能	357	77.1
不能	106	22.9
9.您是否愿意在保证个人信息不外泄的情况下告诉谣言检测机构真实的职业及受教育程度等信息？		
选项	小计	百分比%
愿意	262	56.6
不愿意	50	10.8
不相信信息保密性	151	32.6
10.您能否接受社交媒体上的好友/关注用户状况（即用户关系网）被提取检测？		
选项	小计	百分比%
能	243	52.5
不能	220	47.5
11.您能否接受个人信用与不负责任的谣言传播挂钩（举报增加谣传降低）？		
选项	小计	百分比%
能	272	58.8
不能	191	41.3

15.您是否知道传播网络谣言要负法律责任？		
选项	小计	百分比%
是	450	97.2
否	13	2.8
16.您认为机器判定的结果能不能作为法律上判定传谣的依据？		
选项	小计	百分比%
能	139	30
不能	324	70
17.您是否愿意相信基于不断发展的计算机技术而给出的谣言断定？		
选项	小计	百分比%
是	329	71.1
否	134	28.9

调查结果显示，超过半数的人还是愿意配合多媒体谣言检测的，但为了获取更丰富的样本，与保障公民的信息安全，相关机构必须有可靠的技术与信用防止信息泄漏，人们对于机器能力的信任确实在日益增长，其证据第 17 题显示，大部分人愿意相信科学的谣言判定，但同时法律与技术，社会与科学，情感与理智的矛盾并存。人们在愿意相信机器的同时并不愿意交出法律的审判权。也许多媒体技术的谣言检测除了技术上的困难之外，于人情上也还有很长一段路要走。

三、 总结

本文通过大量资料整理当今主流的多媒体谣言自动检测技术的相关信息，随着机器学习和神经网络算法的迅速发展，多媒体谣言检测技术也迎来一系列技术上的突破，数字图像处理与计算机视觉的应用丰富了媒体信息可提取的特征，多模态特征融合解决了单个孤立特征的局限性，基于传播模型的检测算法弥补了特征提取分类算法忽视谣言内在联系的局限性，基于可信度传播模型等一系列算法又对现有的分类传播算法进行优化。同时，调研中也发现谣言检测的一些研究展望：

1. 提高模型的解释性，更加细致地对谣言建模、找出其文本、图像模式。
2. 对不确定的谣言检测问题建模，实现全面、可靠的谣言检测。
3. 如何在对抗性环境下不断增强谣言检测的能力。

四、参考文献

[1] 金志威,曹娟,王博,王蕊,张勇东.融合多模态特征的社会多媒体谣言检测技术研究[J].南京信息工程大学学报(自然科学版),2017,9(06):583-592. (Jin Zhiwei, Cao Juan, Wang Bo, Wang Rui, Zhang Yongdong. Rumor Detection on Social Media with Multimodal Feature Fusion[J].Journal of Nanjing University of Information Science & Technology.2017.6:583-592.)

[2] 首欢容,邓淑卿,徐健.基于情感分析的网络谣言识别方法[J].数据分析与知识发现,2017,1(07):44-51. (Huan Rong, Deng Shuqing, Xu Jian. Online Rumor Recognition Method based on Sentiment Analysis [J]. Data Analysis and Knowledge Discovery, 2017,1 (07): 44-51.)

[3] 毛二松,陈刚,刘欣,王波.基于深层特征和集成分类器的微博谣言检测研究[J].计算机应用研究,2016,33(11):3369-3373. (Mao Ersong, Chen Gang, Liu Xin, Wang Bo. Research on detecting micro-blog rumors based on deep features and ensemble classifier [J]. computer application research, 2016,33 (11): 3369-3373.)

[4] 杨文太,梁刚,谢凯,杨进,许春.基于突发话题和领域专家的微博谣言检测方法[J].计算机应用,2017,37(10):2799-2805. (Yang Wentai, Liang Gang, Xie Kai, Yang Jin, Xu Chun. Rumor detection method based on burst topic detection and domain expert discovery [J]. computer application, 2017,37 (10): 2799-2805.)

[5] 程亮,邱云飞,孙鲁.微博谣言检测方法研究[J].计算机应用与软件,2013,30(02):226-228+262. (Cheng Liang, Qiu Yunfei, Sun Lu. Research on Detecting Microblogging Rumours, 2013,30 (02): 226-228+262.)

[5] 任文静.面向微博谣言的检测方法研究[D].哈尔滨工业大学,2017. (Ren Wenjing. Research on Detecting Methods on Micro-blog Rumors [D]. Harbin Institute of Technology, 2017.)

[6]刘政,卫志华,张韧弦.基于卷积神经网络的谣言检测[J].计算机应用,2017,37(11):3053-3056+3100. (Liu Zheng, Wei Zhihua, Zhang Tsun. Rumor

Detection Based on Convolutional Neural Network[J]. computer application, 2017,37 (11): 3053-3056+3100.)

[7] 祖坤琳,赵铭伟,郭凯,林鸿飞.新浪微博谣言检测研究[J].中文信息学报,2017,31(03):198-204.(Zu Kunlin, Zhao Mingwei, Guo Kai, Lin Hong Fei. Research on The Detection of Rumor on Sina Weibo[J]. Chinese Journal of information, 2017,31 (03): 198-204.)

[8] 路同强,石冰,闫中敏,周珮.一种用于微博谣言检测的半监督学习算法[J].计算机应用研究,2016,33(03):744-748.(Lu Tong Qiang, Shi Bing, Yan Zhong min, Zhou Jie. Semi-supervised Learning Algorithm Applied to Microblog Rumors Detection[J].computer application research, 2016,33 (03): 744-748.)

[9] 马晶. 微博网站的谣言检测方法研究[D].北京邮电大学,2016.(Ma Jing. Detect Rumors on Microblogging Website[D]. Beijing University of Posts and Telecommunications, 2016.)

[10] 冈萨雷斯(Rafael C. Gonzalez)等著.数字图像处理.阮秋琦、阮宇智译.第三版.北京.电子工业出版社. 2017.5

[11] 西奥多里蒂斯(Sergios Theodoridis)等著.模式识别.李晶皎译.第四版.北京.电子工业出版社. 2010.2

[12] 周琳娜、王东明著.数字图像取证技术.北京.北京邮电大学出版社. 2008

[13] 海金(Haykin,S)著.神经网络与机器学习.申富饶等译.北京.机械工业出版社. 2011.1

[14] Jing Ma, Kam-Fai Wong, Wei Gao. Detect Rumors in Microblog Posts Using Propagation Structure via Kernel Learning. ACL 2017 Long Papers

[15] 张芳，司光亚，罗批. 谣言传播模型研究综述. 国防大学信息作战与指挥训练教研部，第二炮兵青州士官学校

[16] 李栋，徐志明，李生，刘挺，王秀文. 在线社会网络中信息扩散.《计算机学报》

[17]周东浩，韩文报，王勇军。基于节点和信息特征的社会网络信息传播模型[1](国防科学技术大学计算机学院 长沙 410073);[2](数学工程与先进计算国家重点实验室 郑州 450002);[3](解放军信息工程大学 郑州 450002)

[18]陈燕方，李志宇，梁循，齐金山.在线社会网络谣言检测综述,（中国人民大学 信息资源管理学院 北京 100872） 2)（中国人民大学信息学院计算机系 北京 100872),《计

算机学报》

[19] 海沫，郭庆. 在线社交网络信息传播模型研究. （中央财经大学信息学院）. 小型微型计算机系统

（注*：部分图片来自互联网）

致谢

感谢给本次研究活动提供指导与宝贵意见的张勇东老师

&

劳力工作的全体组员

&

所有对本次研究提供宝贵研究参考成果的研究人员

导师评审意见

导师签字： **日期：**