一种基于情感倾向分析的网络团体意见领袖识别算法

肖 宇 许 炜 夏 霖
(华中科技大学电信系互联网技术与工程研究中心 武汉 430074)

摘 要 意见领袖对网络舆情的产生和发展有着重要的指引作用,挖掘和识别网络社区中的意见领袖有重要的现实意义。结合聚类算法和分类算法的优势,提出一种基于话题内容分析的兴趣团体发现方法,以有效识别出兴趣团体。并通过分析用户回帖情感倾向来计算用户间链接的权重。在此基础上,提出了一种新的LeaderRank意见领袖发现算法,通过实验证明该算法能有效提高意见领袖挖掘的准确度。

关键词 社交网络,兴趣聚类,情感分析,意见领袖
中图法分类号 TP393.0 **文献标识码** A

Networking Groups Opinion Leader Identification Algorithms Based on Sentiment Analysis

XIAO Yu XU Wei XIA Lin
(ITEC, Department of Electronics and Information Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract Opinion leaders are core users in online communities, which can guide the direction of public opinion. We proposed a method to find the interest group based on topic content analysis, which combines the advantages of clustering and classification algorithms. Then we used the method of sentiment analysis to define the authority value as the weight of the link between users. On this basis, an algorithm named LeaderRank was proposed to identify the opinion leaders in BBS, and experiments indicate that LeaderRank algorithm can effectively improve the accuracy of leaders mining.

Keywords Social network, Insterest cluster, Sentiment analysis, Opinion leader

1 引言

BBS社区网络可被认为是社会网络的一种形式,属于复杂网络的研究范畴[2,3]。随着复杂网络理论体系的不断完善,人际关系网络的识别及规律研究已成为复杂网络中一个重要的研究方向。BBS网络社区中,人们可以在网络上依据自己的兴趣进行类似小组讨论的自由交流,更有主题性和目的性,而且成员有着固定的身份,在一定时间内成员保持稳定[1]。和现实社会一样,用户通过公众对其言论的认可,能不断提高权威度,形成公认的意见领袖。

BBS中的用户通常根据自己的兴趣发起或者回复话题,且兴趣相同的用户间的互动要多于兴趣不同的用户。BBS中的用户有天然的聚集性,兴趣相投的用户间的讨论更多。因此,BBS实际上可以划分为多个独立的兴趣领域。通过以上对BBS特征的分析,建立了BBS网络模型,该模型图如图1所示。方块代表兴趣,圆圈代表话题,用户层中用户间的虚线代表回复关系,用户层中椭圆表示用户组成的兴趣团体。为了在真实BBS中建立该网络模型,本文以话题层为切入点,首先将话题进行聚类,得到用户的兴趣,然后根据兴趣分类提取相关用户,得到用户兴趣团体。

在一定时期内,群体中存在一个以上的群体成员成为中心人物,其有一定数量的追随者,能够不断提出引人注目的观点,并积极参与讨论,这些成员被称为"意见领袖"。客观上这部分人群具有大量的上网时间,主观上在论坛中发起或参与广受瞩目的话题,吸引大量的成员参与讨论,对周围其他成员产生强烈的影响力。提出了一种新的LeaderRank意见领袖发现算法,它有助于更为准确地识别和挖掘网络社区中的意见领袖。发现意见领袖不仅有助于掌握BBS中用户的意见趋势,还可以将其运用到消息传播控制中,所以发现意见领袖具有重要意义。

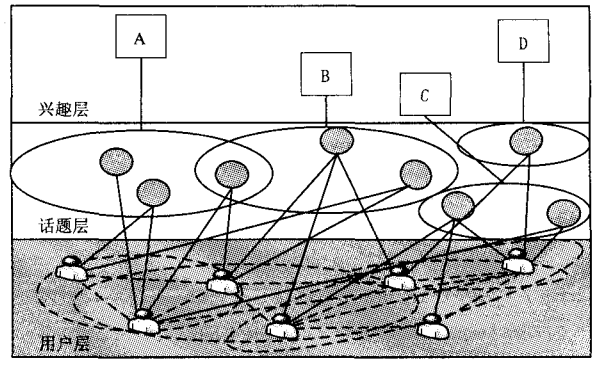

图1 BBS中用户、话题、兴趣结构模型

2 相关研究

国内外学者针对社区人际关系网络中意见领袖的识别方

到稿日期:2011-04-11 返修日期:2011-07-15 本文受"十二五"科技支撑计划重点项目(2011BAK08B01)资助。
肖 宇(1979—),男,博士生,主要研究方向为复杂网络,E-mail:xiaoyu@mail.hust.edu.cn;许 炜(1977—),男,主要研究方向为数据挖掘、协同推荐;夏 霖(1987—),女,硕士生,主要研究方向为社会网络、信息集成。

法进行了广泛的研究，提出了很多算法来识别虚拟社区中的团体和意见领袖。PageRank 和 HITS 等重要的网络超链接结构分析算法作为重要的理论模型基础，被借鉴到基于文本交流的网络人际关系识别中[4]。Jun Zhang 等人将 Java 论坛这种问答网络作为研究对象，采用 HITS 等多种算法来评价用户的权威度[5]。Hengmin Zhou 等人研究了意见网络中的情感倾向性问题，并给出了基于情感分析的意见领袖识别算法，但并没有考虑意见领袖影响范围的团体限制性[6]。Zhongwu Zhai 等人研究了 BBS 网络领袖识别问题，基于正向有权网络假设，提出一种基于兴趣的页面等级算法，并采用水木清华论坛数据进行实测，比较多种算法的适用性，发现基于兴趣的 PageRank 算法最能准确识别意见领袖，但是并没有详细证明社区团体聚类识别方法的准确性，也未考虑回帖者的主观情感倾向性[7]。

以上研究中，将人际关系网络模型简化为回帖者对发帖者的指向关系，存在以下问题：

1）较少考虑用户之间的多次网络交互行为而产生的用户关联权重问题。大部分的研究工作都是基于无权网络（unweighted network），即不考虑网络节点之间的作用强度。这是不符合现实观察的。

2）节点之间的权重不能简单地视为回复次数的叠加，若忽略了可能出现的负面评价，则需要进行回帖情感倾向性判别，分析回帖者的每次言论正负面评价结果，最终产生用户间的总体综合印象。

3）意见领袖有着天然的领域性。某领域的专家不一定在另一个领域同样受欢迎。因此意见领袖是一个相对概念，他的权威性受制于他积极参与的兴趣领域范围。

本研究将对用户产生的话题进行文本分析，通过聚类算法指导下的分类算法进行话题汇聚，从而识别社区兴趣团体，并在社区兴趣团体中考虑回帖者情感倾向性，获得用户间的权威评价矩阵，形成基于兴趣的意见权重人际网络，并提出一种基于情感倾向分析的 LeaderRank 识别算法来获得意见领袖，以有效地提高意见领袖识别的准确性。

3 兴趣团体发现及情感分析研究

3.1 兴趣团体的发现和识别

人际关系网络是一个典型的复杂网络，它的一个重要特征是网络中所呈现出的社区结构。大量实证研究证明，人际关系网络是异构的，并且由许多类型相同的节点组合在一起。相同节点之间存在较多连接，而不同类型节点之间连接相对较少。这些由同一类型节点及这些节点之间的边所构成的子图称为网络中的社区（Community）。与社区发现相关的理论包括图论以及模式识别等。复杂网络社区发现的研究起源于社会学的研究工作[8]，Wu 和 Huberman[9] 以及 Newman 和 Girvan[10] 的研究工作使得社区发现成为近年来复杂网络一个重要研究方向。

对 BBS 兴趣团体的识别是通过对用户发帖中是否包含相似兴趣来进行判别的，是一个典型的社区发现问题。用户通过发帖来进行互动，因此兴趣群体识别本质上是数据挖掘中的文本聚类过程。由于无法预知海量话题中应该存在多少种话题兴趣中心，采用自动文本聚类的方法能够有效地归并相似文章主题，找到话题聚集中心。理论上采用聚类算法就可以识别出持有相同兴趣的用户团体，但在实践中存在很大问题，例如某文章中提到了“马拉多纳”“河床队”等高区分度的典型关键词，人可以立刻判别出这是关于足球的文章，但是文中从未出现过“足球”或类似词汇。单纯以文本分词为基础的聚类无法对文本及其上下文背景语义进行深入的智能分析，导致多聚类中心的出现。这是聚类算法在处理类似问题上本身难以克服的缺陷。

而基于文本分类的算法则能有效避免上述问题，但是分类的前提条件是必须预先设定好分类中心及其判别阈值。因此分类算法是一种基于预先知道领域分类的方法。于是本文提出了一种基于自动聚类指导的分类算法，其将聚类算法和分类算法优势相结合，能够有效解决兴趣中心及其社区团体发现问题。本研究采用的聚类算法参照 1NN 文本聚类算法，核心思想参照文献[11]，其描述如下：

1）通过聚类算法自动进行文本特征项群的识别和分类。针对所有文本建立 VSM（向量空间模型），对每个文本分配 TFIDF 向量，进行 1NN 文本聚类，文本距离矩阵初始阈值设为 0.655。找到初始文本中心，及每篇文档的特征项词群排序。

2）第 1）步得到的聚类中心很多，人工进行聚类中心的合并，并合并新中心对应的特征项群，按照 IDF 进行排序，得到训练后的分类中心。

3）以新兴趣中心及其特征项群为基础，采用 SVM 支持向量机分类算法，自动进行海量文章的话题分类。类别设定归属阈值 S=3.0。

通过上述步骤能够有效对海量文本进行话题区分，算法运行效率高，结果准确。唯一美中不足之处就是需要人工指导，将聚类结果进行兴趣领域归并。由于聚类后兴趣中心已经大为减少，因此归并工作量相对很小。在工程实践上，它不失为极为有效的方法。兴趣团体发现流程如图 2 所示。

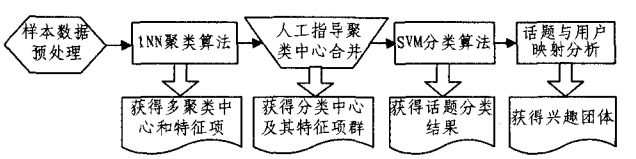

图 2 兴趣团体发现流程

由于每个话题群与用户群进行关联，因此可以方便地获取用户兴趣团体。本研究中暂不考虑意见领袖出现在回帖者中的可能性，我们假设只有发帖者才有资格成为意见领袖。

3.2 回帖情感倾向性分析

情感倾向性分析主要是针对文本内容的观点极性的分析，即通常所说的正面、负面或者中立。在人机交互领域，情绪对反应的影响已被人们重视。在 BBS 的话题回复中如何判定回帖者的情感倾向性是一个难题。沈阳通过构建情感词典的方式确定语料所表达的情绪。针对微博客情感挖掘测试的结果经交叉判定正确率达到 80.6%[12]。本文将在该研究基础上结合 BBS 的特征进行算法调整。

保留算法结构的核心部分，构建态度词权值词典，并引入自定义的否定词典、程度词典和感叹词典。其中通过对 BBS 回复情感词汇的分析，增加 514 个情感词条，达到 1852 条情

感词，并定义词条的极性和强度，将文本情感权值归一到[−1,1]之间。

由于微博限制在140字内，明显不符合BBS回帖文字量，因此在每段文字的首尾句增加权重的基础上，回帖中若存在多段文字，则首段和尾段增加情感权重。

A用户对B用户如果发生多次回帖行为，则A对B情感倾向性定义为多次情感倾向平均值WAB。物理上可以理解为A对B的综合权威评价。

将原C#代码版本改写为Java版本。

为了验证调整后情感倾向性分析算法在BBS语言环境下的有效性，在某部属高校大型BBS论坛上，集中抽取3天论坛回复数据3128条作为测试样本，过滤掉广告、链接等无效回复数据49条，余下3079条回复数据作为测试样本。对5名大学生表现出来的情感倾向进行人工标注：正面、负面、中立。并将3079条回复信息分别载入本文情感分析算法进行计算，将计算结果与人工判断结果进行对比，监测计算正确性。情感倾向计算正确率如表1所列。

表1　情感倾向计算正确率

天	情感类别	人工判定	正确计算条目	正确率(%)
Day1	正面	577	436	75.5
	负面	435	359	82.5
	中立	129	82	63.6
Day2	正面	731	566	77.4
	负面	334	262	78.4
	中立	243	167	68.7
Day3	正面	429	313	72.9
	负面	117	89	76.1
	中立	84	52	61.9

实验数据表明情感判定正确率平均为：正面75.3%，负面79.0%，中立64.7%。相比之下，负面情感的判定较为准确，这和情感词库中负面词库覆盖较好有很大的关系，通过不断训练和完善情感词库能够进一步提高判别准确性。中立情感的判别相对正确率不高，可通过适当扩大情感阈值的上下界来弥补。本文中根据实测数据，将中立情感阈值上下界扩大0.1。情感倾向性判定算法调整后效果能够满足本研究所需。

用户A对用户B的权威度评价是A、B之间多次交往好、恶印象之综合，这是符合社会一般性观察的。本文中用户A对用户B的综合权威度评价指标是用户A对用户B多次回帖情感倾向性分析结果值的平均值WAB。最终用户两两之间都会算出对应的权威评价值，形成用户个体之间的权威评价矩阵W，定义如下：

$$W=\begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{n1} & \cdots & W_{m} \end{bmatrix} \tag{1}$$

$$W_{ij}=\frac{\sum e_{ij}}{t_{ij}} \tag{2}$$

式中，e_{ij}为节点j对节点i在一个帖链中回帖的情感倾向性值，$\sum e_{ij}$表示节点j对节点i的多个帖链的情感系数和，t_{ij}表示节点j对节点i共同参与话题的次数。

4　意见领袖的识别

意见领袖不能脱离自身兴趣领域而存在。例如足球领域专家，不一定是军事领域的专家。识别意见领袖，首先就要根据用户话题所呈现的兴趣领域进行识别和划分，通过上文基于聚类算法指导的分类算法进行兴趣群体的识别，一个用户可以属于多个群体。然后在已经划分的兴趣群体内根据上文的情感倾向性分析算法获得群体用户之间的权威评价矩阵，建立人与人之间的网络关联。最后通过基于情感权重的LeaderRank算法获得意见领袖排名。实验过程如图3所示。

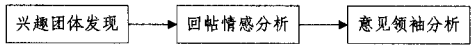

图3　意见领袖识别流程

4.1　数据集分析

本文的研究数据来源于某部属高校大型BBS论坛2007年1月至2009年12月间所有发帖及回复留言。为了准确研究BBS人际社区网络，必须清洗样本数据。例如将发贴者自我回复行为进行屏蔽；如果是基于回复的回复则需要指向真实被回复人，而不是原始发帖人。

表2中对研究样本的人际关系网络特征进行了统计。其中“度”指与该节点相连的其他节点的数目，是描述网络中个人影响力的重要指标。所有节点度值平均值为节点平均度，其样本中该值达到52.71，说明这是一个讨论热烈、人气旺盛的论坛。

表2　BBS网络特性统计

统计项目	特征值
注册用户数(U)	49902
实际发帖人数(N)	12779
总版块数(G)	120
总话题数(W)	906633
节点平均度(K)	52.71
平均聚集系数(C)	0.94
平均路径长度(L)	3.07

采用上文中基于聚类结果指导的分类算法进行话题兴趣领域分析。为了便于研究，采用BBS中最大的校园话题版块数据进行后续的实验和分析。该版块共计主贴19687个，用户2215人。根据初步聚类结果，产生聚类中心128个，通过人工归并后，自动合并每个分类下的关键词群及阈值，作为分类算法的基础。通过分类算法计算分类结果，共计9大类，各类所占帖数如表3所列。

表3　话题兴趣分类结果

类别	帖数	占总帖数比例
教师	11844	49.66
学费	3413	14.3
教育	1841	6.88
学校	1326	5.53
学习	347	1.43
招生	288	1.17
考试	264	1.09
就业	184	0.76
其他	176	0.71

人工将自动聚类结果进行归并，特别是聚类算法产生的高区分度关键词及其词频统计是分类算法的核心基础。通过话题分类结果，将话题和用户进行映射，其中回帖者也包括在用户群中。通过上述步骤完成了兴趣团体发现的全过程。为了更为直观地分析兴趣团体的形态，将教师类别兴趣团体人际关系网络图形化，可以更加清晰地看出网络是异构的，核心节点之间存在较多的链接，周围节点之间的连接相对较少，团体形态明显。兴趣团体人际关系网络结构如图4所示。

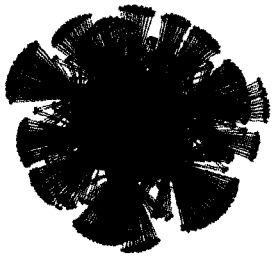

图 4　兴趣团体人际关系网络结构

4.2　基于情感权重的 LeaderRank 算法

PageRank 算法是用来衡量网络中节点重要程度的经典算法，在社会网络分析研究中得到了广泛的应用。本文借鉴 PageRank 算法思想提出了基于用户权威评价权值的 LeaderRank 算法。用户节点网络中，将通过多次交互而形成的权威评价值作为节点之间的边权，通过 LeaderRank 算法计算出每个用户在群体中的领袖值。假设用户 A 在社区网络中与用户 $U_1 \cdots U_n$ 发生过交互行为，定义用户 A 的领袖值(LeaderRank，LR)如下：

$$LR(u)=(1-d)+d\times\sum_{v\in B_u}\frac{LR(v)\times W_{uv}}{C(v)} \tag{3}$$

$$C(V)=\sum_{k\in T_v}|W_{vk}| \tag{4}$$

式中，$LR(u)$是节点 u 的 LeaderRank 值，B_u 是指向 u 的节点的集合，T_v 是 v 指向的节点的集合，w_{uv}是 v 对 u 的权威印象值，$C(V)$是节点 u 的链出的权重绝对值之和，d 是阻尼系数，可设定在(0,1)之间，本文取值为 0.85。将所有节点的 LeaderRank 初始值设为 0.1，通过迭代到收敛为止，可以得到所有用户的 $LR(u)$，LR 值最高的用户即为意见领袖。

同时，为了方便比较，本文研究拟采用另外 5 种传统方法得到的用户权威度排名作为相关对比衡量指标。文献[5,7]中的实测数据说明，简单的统计学方法在对意见领袖权威度识别上也能够得到较好的实验结果。

人度(Indegree)：指向本节点的邻居数之和。

传统 PageRank(Global PageRank)：不区分兴趣领域，对全样本数据直接执行标准 PageRank 算法。

基于兴趣 PageRank(Interest based PageRank)：考虑兴趣领域差别，但是不考虑情感倾向性。

在线时长(Online Time)：用户注册后累计在线时长。

经验值(experience value)：论坛通过简单发文奖励规则给出用户经验值评分。

4.3　评价指标

由于并不存在一个绝对的精确社区意见领袖的评价指标体系，因此我们根据意见领袖的定义：其核心性体现在与之交互的节点数量多寡以及节点间交互的频繁程度，提出节点核心率(CoreRadio)作为评价指标，定义如下：

$$CR(i)=\frac{\sum_{j=1}^{N}a_{ij}W_{ij}}{\sum_{i=1}^{N}\sum_{j=1}^{i}a_{ij}W_{ij}} \tag{5}$$

式中，W_{ij}为节点 i,j 之间的权重，本文中即为用户综合印象，a_{ij}表示邻接矩阵中描述与加权网络相对应的无权网络(顶点 i,j 之间有边存在时 $a_{ij}=1$，否则 $a_{ij}=0$)。核心率体现了权重网络中，节点在整个网络中的重要程度和影响能力。为了比较 LeaderRank 算法和其他算法，在每个兴趣团体中进行核心率计算，然后计算平均核心率。

5　结果与分析

为了更好地观察研究结果，实验中去除了只有发贴没有回帖的孤立节点。考察不同算法下 Top K 节点用户群的核心率，实验结果如图 5 所示。

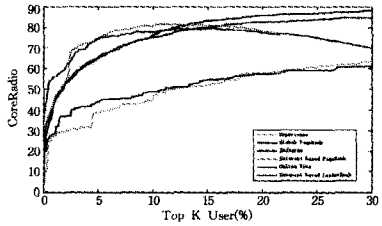

图 5　6 种算法和核心率对比

通过对图 5 数据的分析发现，LeaderRank 算法、单纯基于兴趣的 PageRank 算法、传统 PageRank 算法、Indegree 4 种评价算法具有较好相关性，并且明显优于经验值和在线时长等简单统计方法排序结果，这充分说明通过网络结构特征能够非常有助于发现和识别意见领袖。并且前 4 种方法有着共同的特征：曲线斜率在 10%左右开始趋于平缓，这个现象说明 10%的用户基本上覆盖了 80%的互动关系，也就是说意见领袖只存在于不到 10%的用户中，并对论坛起到了核心推动作用。同时我们发现两种基于兴趣的算法都明显优于其他算法，这说明通过兴趣领域的划分能够有效地让意见领袖凸显出来，防止出现某些意见领袖在传统 PageRank 算法被“人为矮化”的现象。这与文献[7]中的研究结果是一致的。本文指出，LeaderRank 算法在前 5%时表现最为优异，能够迅速让意见领袖凸显出来，达到实验预期效果。用户经验值这种统计学方法由于只考虑了用户的发文数量，并没有考虑节点之间的回复关系，因此并不能作为准确的用户权威度值。

结束语　本文对网络社区的形态进行了深入分析，提出了基于兴趣类聚的网络模型结构；提出一种基于聚类指导的分类算法，它能实现话题分类，发现和识别用户兴趣群体；在此基础上通过用户交互行为中产生的情感倾向性分析，形成用户之间的综合评价权威指标，并作为用户网络节点之间的权重，提出一种基于情感倾向的 LeaderRank 算法；通过对比主流的意见领袖识别算法，发现兴趣聚类和情感倾向分析能够更为有效地发现和识别意见领袖。

参 考 文 献

[1]　彭小川，毛晓丹. BBS 群体特征社会网络分析[J]. 青年研究，2004,4:39-44

[2]　Peng Xiao-chuan, Mao Xiao-dan. Social Network analysis of BBS community characteristics[J]. Youth Studies, 2004, 4:39-44

[3]　Albert R, Barabasi A L. Statistical Mechanics of Complex Networks[J]. Rev. Mod. Phys., 2002, 74:47-97

[4]　Dorogovtsev S N, Mendes J F F. Evloution of networks [J]. Adv. Phys., 2002, 51:1079-1187

[5]　Matsumura N, Ohsawa Y, Ming, et al. Influence Diffusion Model in Text-Based Communication[C]//WWW02. 2002

[6]　Zhang J, Ackerman M, Adamic L. Expertise networks in online communities: structure and algorithms[C]//WWW '07. 2007

(下转第 46 页)

4.3 与相关检测系统的比较

有关网络异常检测的研究很多，这里将采用 KDD CUP 1999 数据集进行测试的几个异常检测系统与 MRF 做了比较，包括 KDD CUP1999 竞赛的冠军 EMERALD 在内，在误报率为 10%的情况下，结果如表 2 所列。

表 2 MRF 与其它算法的比较

Systems	Detection Methods	Amount of detected attacks/Total attacks	Detection Rate
EMERALD	Expert System	85/201	42%
LERAD	Learning goog rules from training set	114/190	60%
PHAD	Packets head anomaly detection	54/201	27%
ALAD	Using well know ports	60/201	30%
NETAD	PHAD and ALAD	132/201	66%
FAD	D-S theory	119/201	59%
MRF	Based on Hilbert-Huang transform and D-S theory	171/201	85.1%

实验结果表明，MRF 性能超过了当年 KDD CUP 的冠军 EMERALD 以及采用动态学习规则的 LERAD。仅仅使用 D-S证据理论的 FAD 检测系统与其它方法比较并没有太大的性能提高，并且与 NETAD 比较性能还有所降低，但是在与 HHT 相结合使用后，D-S 证据理论检测有了较大的提高，因为网络流特征信号是非线性和不平稳的，并且正常流量特征表现为自相似性，HHT 将信号进行不同时间尺度的分解去除了趋势等不平稳分量，并将多尺度分量作为证据，有效区分了突发流和 DoS 流。

结束语 HHT 是一种自适应的时频局部化多尺度分析方法，适合处理非线性、非平稳信号；D-S 证据理论作为不确定推理理论之一，已经广泛用于多传感器信息融合等领域。最近，该理论被引入到网络异常检测中，在多特征融合方面取得了一定效果，但其检测率仍然不理想。本文将 HHT 与其结合，提出了 MRF 异常检测方法，它从多尺度的角度检测异常，并使用 KDD CUP 1999 数据集进行验证。实验表明，该系统在保证一定误报率的情况下提高了检测率。

参 考 文 献

[1] Dempster A. Upper and lower probabilities induced by multivalued mapping[J]. Annals of Mathematical Statistics, 1967, 38(2): 325-339

[2] Mahoney M V, Chan P K. PHAD: Packet Header Anomaly Detection for Identifying Hostile Network Traffic[R]. Melbourne: Department of Computer Science, Florida Institute of Technology, 2001

[3] Mahoney M V, Chan P K. Learning Nonstationary Models of Normal Network Traffic for Detecting Novel Attacks[C]// Proc. of the Eighth International Conference on Knowledge Discovery and Data Mining. Edmonton: ACM, 2002: 376-385

[4] Porras P A, Neumann P G. EMERALD: Event Monitoring Enabling Responses to Anomalous Live Disturbances[C]// Proc. of the 20th National Information Systems Security Conference. Baltimore, 1997: 353-365

[5] Lippmann R, Haines J W, Fried D J, et al. The 1999 DARPA off-line intrusion detection evaluation[J]. Computer Networks: The International Journal of Computer and Telecommunications Networking, 2000, 34(4): 579-595

[6] Moayedi H Z, Masnadi-Shirazi M A. Arima model for network traffic prediction and anomaly detection[C]// Proc. of the International Symposium on Information Technology. 2008

[7] Li Zong-lin, Hu Guang-min, Yao Xing-miao. Detecting Distributed Network Traffic Anomaly with Network-Wide Correlation Analysis[J]. EURASIP Journal on Advances in Signal Processing, 2009

[8] Kline J, Nam S, Barford P, et al. Traffic Anomaly Detection at Fine Time Scales with Bayes Nets[C]// Proc. of the Third International Conference on Internet Monitoring and Protection. Washington: IEEE Computer Society, 2008: 1-10

[9] Huang N E, Shen Z, Long S R, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[C]// Proc. of the Royal Society of London. 1998, A454: 903-995

[10] 诸葛建伟，王大为，陈昱，等. 基于 D-S 证据理论的网络异常检测方法[J]. 软件学报，2006，17(3)：463-471

[11] Lakhina A, Crovella M, Diot C. Diagnosing Network-Wide Traffic Anomalies[C]// Proc. of ACM SIGCOMM. Portland: ACM, 2004

[12] Andeson D, Frivold T, Tamaru A. Next-generation intrusion detection expert system(NIDES)[R]. Software Users Manual, Beta-Update release. Menlo Park: Computer Science Laboratory, SRI International, 1994

[13] Mahoney M V, Chan P K. Learning Models of Network Traffic for Detection Novel Attacks[D]. Computer Science Department, Florida Institute of Technology, 2002

[14] Leland W E, Taqqu M S, Willinger W, et al. On the Self-similar Nature of Ethernet Traffic [J]. Transactions on Networking, 1994, 2(1): 1-15

[15] Silveira F, Diot C, Taft N, et al. ASTUTE: Detecting a Different Class of Traffic Anomalies[C]// Proc of SIGCOMM. 2010

(上接第 37 页)

[7] Zhou Heng-min, Zeng D, Zhang Chang-li. Finding Leaders from Opinion Networks[C]// ISI 2009. 2009: 266-268

[8] Zhai Zhong-wu, Xu Hua. Identifying opinion leaders in BBS[C]// IEEE Proceedings of Web Intelligence and Intelligent Agent Technology. 2008

[9] Scott J. Social Network Analysis: A Handbook[M]. Sage Publications, London, 2000

[10] Wu F, Huberman B A. Finding communities in linear time: A Physics approach[J]. Phys. J B, 2003, 38: 331-338

[11] Newman M E, Girvan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69: 026113

[12] Cover T M, Hart P E. Rates of convergence for nearest neighbor procedure[C]// Proc. HaWaii Int. Conf. on System Science. 1967: 413-415

[13] Yang Shen, Li Shu-chen, Zhen ling. Emotion mining research on micro-bolg[C]// SWS 2009. 2009: 71-75