基于Single-Pass的网络舆情热点发现算法

格桑多吉[1]，乔少杰[2]，韩 楠[3]，张小松[4]，杨 燕[2]，元昌安[5]，康 健[2]

(1. 西藏大学藏文信息技术研究中心 拉萨 850000; 2. 西南交通大学信息科学与技术学院 成都 610031;
3. 西南交通大学生命科学与工程学院 成都 610031; 4. 电子科技大学大数据研究中心 成都 611731;
5. 广西师范学院科学计算与智能信息处理广西高校重点实验室 南宁 530023)

【摘要】考虑网络事件的时间距离，基于半结构化网页中不同位置特征项重要程度的不同，提出改进的single-pass文本聚类算法single-pass*，优势在于对Web文本不同位置特征项的加权处理，仅需计算新文档与同类别种子文档间的相似度。实验结果表明，相比single-pass，改进算法极大减少了漏检率和错检率，降低了由于新文本流内文档进行相似度计算导致系统性能的下降，平均提高Web文本聚类效率40%。将聚类后的Web文本应用于网络舆情分析，进行主题关注度分析和话题热度特性分析。

关 键 词 舆情分析; single-pass; 文本聚类; 话题发现
中图分类号 TP312 **文献标志码** A doi:10.3969/j.issn.1001-0548.2015.04.021

An Internet Public Opinion Hotspot Detection Algorithm Based on Single-Pass

GESANG Duoji[1], QIAO Shao-jie[2], HAN Nan[3], ZHANG Xiao-song[4],
YANG Yan[2], YUAN Chang-an[5], and KANG Jian[2]

(1. Tibetan Information Technology Research Center, Tibet University Lasa 850000;
2. School of Information Science and Technology, Southwest Jiaotong University Chengdu 610031;
3. School of Life Science and Engineering, Southwest Jiaotong University Chengdu 610031;
4. Big Data Research Center, University of Electronic Science and Technology of China Chengdu 611731;
5. Science Computing and Intelligent Information Processing of Guangxi Higher Education Key Laboratory,
Guangxi Teachers Education University Nanning 530023)

Abstract By considering the time interval of Internet events as well as the importance of different feature items from semi-structured Web documents in different locations, an improved single-pass text clustering algorithm called single-pass* is proposed. The advantage is that it assigns the weight value to different feature items from different locations on the Web pages, and only needs to calculate the similarity between the new document and its seed document. Experimental results show that, compared to the single-pass algorithm, the improved algorithm can reduce the missing rate, the error detection rate, and the degradation of system performance caused by computing the topic similarity of documents in new Web data stream, and improve the clustering efficiency at an average rate of 40%. The clustered Web texts can be used to analyze the Internet opinion including the topic relevant degree and the hot degree.

Key words public opinion analysis; single-pass; text clustering; topic detection

话题发现和跟踪是指新闻专线和广播新闻等来源的新闻数据流中自动地发现话题并把话题相关的内容组织到一起的技术。通过增量的文档聚类的方法，信息流被聚集到有限的话题类簇中，类内高度相似，不同的类间相似度较低，以此进行海量数据的融合。热点舆情话题是话题舆情中受关注度最大，影响也较为突出的舆情，旨在从半结构化海量Web数据中获取相应的主题并进行整合，以新的热点事件分析并了解热点话题事件的发展。热点话题分析对舆情分析具有较大的实际意义，可以及时向网络监控部门提供网民关注焦点，辅助网络舆情分析。

随着网络舆情及预警机制研究的广泛深入和迫

收稿日期：2014 – 11 – 07；修回日期：2015 – 05 – 13
基金项目：国家自然科学基金(61100045, 61165013)；高等学校博士学科点专项科研基金(20110184120008)；中国博士后科学基金特别资助项目(201104697)；教育部人文社会科学研究青年基金(14YJCZH046)；中央高校基本科研业务费专项资金(2682013BR023)；科学计算与智能信息处理广西高校重点实验室开放课题资助(GXSCIIP201407)；四川省教育厅资助科研项目(14ZB0458).
作者简介：格桑多吉(1972 –)，男，副教授，主要从事藏文信息处理、Web文本挖掘方面的研究.

切性，话题发现和跟踪的研究已经成为当前的研究热点。卡内基梅隆大学采用经典的single-pass算法识别新闻中的事件[1]。文献[2]结合新闻要素提出了基于动态进化模型的新闻事件话题发现算法，应用基于时间距离的相似度计算模型自动对新闻资料进行组织，生成新闻专题。文献[3]提出了利用single-pass对新闻事件在线聚类进而实现话题发现的算法。文献[4]提出了一种基于multi-agent的思想single-pass聚类，使用分散的自底向上和自组织策略对相似的数据点进行分类。文献[5]提出基于词共现图的识别中文微博新闻话题的方法，综合相对词频和词频增加率这两个因素抽取微博数据中的主题词。文献[6]基于文本重建的网络话题发现模型，用主题区域发现话题并将其应用于整个文档中以区分子话题。文献[7]提出了一种中文微博热点话题发现方法，不足之处在于仅进行了中文话题发现，不支持多语言的话题发现与跟踪。本文研究的不同点在于：1)基于事件时间的先后引入了时间距离特性的相似度计算模型；2) 所提算法支持中、英、藏文等不同字符集的话题发现；3) 在话题发现基础上进行网络舆情分析工作。

1 话题发现与跟踪

1.1 文本特征提取

文本的表征有诸多方法，如布尔模型、向量空间模型、概率模型等，其中向量空间模型是在应用中广为采用的模型，首先被用于信息检索系统。通常，文档被表示为向量，每一维均对应独立的词。每篇文档，均可以表示为规范化的特征向量：

$$\boldsymbol{d} = \{(t_1, w_1), (t_2, w_2), \cdots, (t_n, w_n)\} \tag{1}$$

式中，t_i表示第i个特征项；w_i表示特征项t_i在文本$\boldsymbol{d}$中的权重，所有的文本向量构成文本集的一个特征向量。文本向量中权重值的求取最为有效的方法是使用tf-idf模型，tf称为词频，计算该词描述一篇文档内容的能力。其中，idf称为逆文档频率，计算该词区分文档的能力。

1.2 加权词频因子tf

tf-idf是词频和逆文档频率两项的乘积，有多种方法用于获取两种统计词频的精确值。使用在某篇文档中的原始词频是最简化的选择，如词t在文档$\boldsymbol{d}$中出现的次数。已知$f(t, \boldsymbol{d})$表示t的频率，那么tf的计算公式是$\mathrm{tf}(t, \boldsymbol{d}) = f(t, \boldsymbol{d})$。值得注意的是，本文结合了出现在文档中不同位置的词的特性，如meta中keyword、title和description等关键词在文档中的权重，因此tf的计算公式表示为：

$$\mathrm{tf}(t) = w_1 w_2 w_3 f(t, \text{ body}) + w_4 f(t, \text{ meta}) \tag{2}$$

式中，f(body)表示特征词t在Web文档的body标签位置出现的次数；f(meta)是在文档标题与描述中特征词出现的数目；w_1、w_2、w_3是权重系数，取值分别表示某个事件的关键信息，即事件名称、地点及组织这3个特征词。为了保持一致，本文采用文献[8]中权值的设置方法：

$$w_1 = \begin{cases} 3 & \text{单词表示事件名称} \\ 1 & \text{其他} \end{cases} \tag{3}$$

事件地点和组织的设置方法同事件名称。w_4 = 3，表示网页中meta中的keyword、title和description的权重。

1.3 逆文档频率idf

如果一个词在很多文档中出现过，则通过这个词来区分文档的区分度越小，可以用逆文档频率idf来度量，表示包含某个词的文档数目：

$$\mathrm{idf}(\mathrm{term}) = \log(n/m + 0.01) \tag{4}$$

式中，n代表文档的数量；m表示出现特征词的文档数量；0.01是为了防止n/m=0时对数值为1。综上所述，特征词的tf-idf值的计算公式如下：

$$\mathrm{weight}(t) = \mathrm{tf}(t) \times \mathrm{idf}(t) \tag{5}$$

1.4 话题模型和相似度计算

通常话题模型包含质心向量方法和中心向量方法。不准确中心向量的选择极易导致后续增量聚类结果的错误。对于一篇新文档，需要遍历在某指定类别中的所有文档，这样随着文档数量的增加，算法的运行效率会降低。为此，本文提出了种子话题的概念，即在一个文档类中，选择若干文档代表某一话题。此外，在文本相似度计算中，本文仅需计算新文档和种子文档间的余弦相似度。

$$\mathrm{sim}(d_i, d_j) = \cos\theta = \frac{\sum_{k=1}^{M} w_{ik} w_{jk}}{\sqrt{\sum_{k=1}^{M} w_{ik}^2 \sum_{k=1}^{M} w_{jk}^2}} \tag{6}$$

$$\overline{\mathrm{sim}(d_i, d_j)} = \frac{1}{k} \sum_{j=1}^{k} \mathrm{sim}(d_i, d_j) \tag{7}$$

式(6)和式(7)中，d_i表示新文档的特征向量；d_j表示某个话题的第j个种子话题的特征向量；M表示特征向量的维度；w_{ik}表示新文档i的特征向量的第k个权重；w_{jk}表示第j个种子话题特征向量的第k个权重；$\mathrm{sim}(d_i, d_j)$表示新文档和一个类别中某一种子的相似度；$\overline{\mathrm{sim}(d_i, d_j)}$表示新文档特征向量和某类中第$j$个种子

话题特征向量的平均相似度。

1.5 网络事件的时效性

考虑到相似的事件可能在不同时间段发生的情况，本文引入报道时间距离的概念，对新闻报道与话题的相似性计算利用时间距离进行综合衡量。对大量新闻报道研究发现，时间相距较远的两篇内容相似报道中出现的特征词往往非常相似。如果不考虑这两篇Web报道的时间距离，single-pass聚类算法会将不相关话题聚为同一类，因此引入时间距离来进一步区分文档类别，时间距离计算方法如下：

$$\mathrm{dis}(d,c)=\frac{2t_d-t_{\mathrm{ch}}-t_{\mathrm{ce}}}{2} \tag{8}$$

式中，t_{d}表示报道d出现的时间；t_{ch}是与话题c相关第一篇报道时间；t_{ce}是话题c最近一篇报道时间。改进后报道d和话题c间相似度计算公式如下：

$$\mathrm{sim}(d,c)=\alpha\mathrm{sim}(d,c)-\beta\,\mathrm{dis}(d,c) \tag{9}$$

式中，sim(d, c)利用式(6)计算；dis(d, c)由式(8)求取。改进的文本相似度计算方法既考虑了文档内容相似度的影响，又考虑了时间因素的影响，α和β是对这两种因素所赋予的权值，其中，$\alpha+\beta=1$。

2 基于single-pass算法的话题发现

2.1 single-pass聚类算法

single-pass算法采用增量聚类的方式将文本向量与已有话题内的报道进行比对，计算文本相似度进行匹配。若与某个话题类别匹配，则把该文本归入该话题，若该文本域所有话题类别的相似度均小于某一阈值，则将该文本表示成新的种子话题。single-pass聚类算法步骤如下：1) 输入新文档$\boldsymbol{d}$；2) 计算$\boldsymbol{d}$与已有话题分类中每篇文档的相似度，获取与$\boldsymbol{d}$相似度最大的话题并得到相似度值T；3) 若T大于阈值θ，则文档$\boldsymbol{d}$被分类到已知的话题类别，否则作为一个新的话题类别；4) 聚类过程结束。

2.2 single-pass*聚类算法

通过引入种子话题和在网页中不同位置的文本信息要素加入权重，本文提出了一种改进的single-pass*聚类算法，区别在于：1) 引入了种子话题；2) 计算网页不同位置的特征项权重，仅计算新文档和类别种子文档间的相似度。算法如下：

算法 1 基于single-pass的话题发现算法。

输入：Web文档集合T，话题种子文档集合S；

输出：聚类后的话题文档集合T'。

```
Initialize T and S;
for (Ti∈T) do{
  for (Sj∈S) do{
  if ((S= =null) && (T != null))
    {sim ←sim(Ti, Sj);
     if (sim >θ)
       {C.add(j, sim );}
    }
  }
  if (!C.isEmpty){
    sort(C);
    if (S.size()<l && k > ξ)
      {Sk←insertSeedDoc(Ti);}
    }
  }
  else{
    create a new topic t;
    t.setId(S.size()+1);
    t.setTopic(Ti);
    S.add(t);
  }
}
output( T′ );
```

算法基本思想为：1) 对文档进行向量空间模型规范化处理(第1行语句)，每篇文档都由一个<term, weight>对象集合组成，对新文档集合进行遍历，计算新文档与每个话题类中种子文档对象间的平均相似度(第2～5行语句)，若相似度大于已知的相似度阈值θ，将此新文档与当前文档类的平均相似度和类标加入到集合C中(第6～9行语句)，2) 根据平均相似度大小对C中对象进行排序，获取平均相似度最大值所对应的类标，将新文档加入到对应的种子文档中(第10～11行语句)。若当前类标的种子数目与原始话题种子话题数目的比值k大于阈值ξ，并且当前种子文档的数目小于l，将当前文档插入到本类文档列表中(第12～15行语句)；若C为空，说明此时没有相应的类别与新文档相似度大于阈值，新建文档分类(第16～17行语句)，并将新文档加入到此类的种子文档集合中，根据当前新文档对象列表循环迭代上述操作(第18～22行语句)。最后，输出聚类后的话题文档集合T'(第23行语句)。

3 实验及算法性能分析

3.1 实验评价标准

实验中采用的评测标准包括：漏检率M、错检

率F和错误识别代价Cost[9]。

表1 话题发现评价标准

类别	相关文档数	不相关文档数
被检测到的文档数	A	B
未被检测到的文档数	C	D

定义漏检率$M=C/(A+C)$，错检率$F=B/(B+D)$。类似于F-measure，话题检测与跟踪引入了耗费代价函数对结果进行综合评价，定义为：

$$\text{Cost} = w_m Mp + w_f F(1-p) \tag{10}$$

式中，w_m是漏检率系数；w_f是错检率系数；p是文档归属某一话题的先验概率。

3.2 实验数据

本文设计实现了一个网页抓取器，实验中中文Web数据来源于新浪和搜狐网站的专栏，藏文数据来源于藏文门户网站。其中，中文事件分为10类，藏文事件分为6类。

3.3 话题发现评价及分析

实验利用IKAnalyzer中文分词工具包对中文进行分词，利用文献[10]中使用的藏文分词算法对藏文文本进行分词，构建向量空间模型。single-pass算法中的阈值θ和ξ分别设置为0.02和0.15，取值依据是通过大量实验，参数调节得到的最优值。

1) 中文话题发现评价

从中文实验语料库中整理出10个话题，每个话题包括90～120篇报道。实验主要采用3.1节给出的话题检测与跟踪评价指标，结果如表2～表4所示。

表2 single-pass算法中文话题发现结果

话题类别	马航客机	云南昭通地震	韩国客轮沉没	叙利亚化武	山东招远打死女顾客	湖南军训教官与师生冲突	玉林狗肉节	广州公交车爆炸	越南打砸抢烧外国企业事件	多个不明飞行物坠入黑龙江境内
文档数	95	110	120	120	120	120	120	120	120	100
A	89	102	97	100	87	99	104	101	79	75
B	19	23	15	28	21	17	22	29	22	27
M	0.06	0.07	0.19	0.17	0.275	0.175	0.13	0.158	0.34	0.25
F	0.018	0.022	0.015	0.027	0.02	0.017	0.021	0.028	0.02	0.021

表3 single-pass*算法中文话题发现结果

话题类别	马航客机	云南昭通地震	韩国客轮沉没	叙利亚化武	山东招远打死女顾客	湖南军训教官与师生冲突	玉林狗肉节	广州公交车爆炸	越南打砸抢烧外国企业事件	多个不明飞行物坠入黑龙江境内
文档数	95	110	120	120	120	120	120	120	120	100
A	85	105	105	92	112	109	113	111	103	87
B	11	19	15	25	17	19	19	23	13	5
M	0.105	0.045	0.125	0.233	0.067	0.092	0.058	0.075	0.142	0.13
F	0.01	0.018	0.015	0.024	0.017	0.019	0.019	0.022	0.013	0.005

表2和表3分别为single-pass和single-pass*文本聚类算法的实验结果，话题文本流被聚为10类，如马航客机、云南昭通地震等。通过结果可以发现，single-pass*算法的性能明显优于single-pass。

表4 中文Web文本下算法性能比较

性能	single-pass	single-pass*
平均漏检率	0.182	0.107
平均错检率	0.022	0.016
错误识别代价	0.006	0.004

如表4所示，在实验参数和实验文本数据一致的情况下，single-pass*算法的平均漏检率和平均错检率均低于single-pass算法，漏检率平均减少41.2%，错检率平均减少27.3%。原因在于：对Web文本不同位置的特征词加权值使文档的属性标注更加准确，同时对固定维度的种子文档进行文本相似度计算使文档归类更加有效，从而在漏检率和错检率指标上都有所降低。

2) 藏文话题发现评价

从藏文实验语料库中每个话题包括86～1 441篇报道，实验结果如表5～表7所示。

表5 single-pass算法藏文文本发现结果

话题类别	地震等新闻	百家争鸣等文化	名胜古迹等旅游	论藏等经济	花园等音乐	闲暇等视频
文档数	1 441	160	199	185	86	147
A	1 125	131	154	159	79	121
B	36	62	49	30	33	34
M	0.219	0.181	0.226	0.141	0.081	0.177
F	0.046	0.03	0.024	0.015	0.015	0.016

表6　single-pass*算法藏文文本发现结果

话题类别	地震等新闻	百家争鸣等文化	名胜古迹等旅游	论藏等经济	花园等音乐	闲暇等视频
文档数	1 441	160	199	185	86	147
A	1 259	145	165	175	72	128
B	19	23	15	28	21	17
M	0.126	0.094	0.171	0.054	0.163	0.129
F	0.024	0.011	0.007	0.014	0.010	0.008

表7　藏文Web文本下算法性能比较

	single-pass	single-pass*
平均漏检率	0.171	0.123
平均错检率	0.025	0.012
错误识别代价	0.005 9	0.003 6

改进算法在处理藏文Web文本上优势依然明显，在漏检率和错检率上较single-pass算法都有较大的改善，原因与中文话题发现相同。

3.4　算法运行时间比较

本节讨论single-pass*与single-pass运行时间，实验结果如图1所示。可以发现single-pass*算法的时间消耗明显低于single-pass算法，平均降低40%。原因是改进后的算法仅需计算新文档与指定数目的代表事件类别的种子节点的相似度，不需要与包含所有事件的文档进行比较，减少了计算时间。

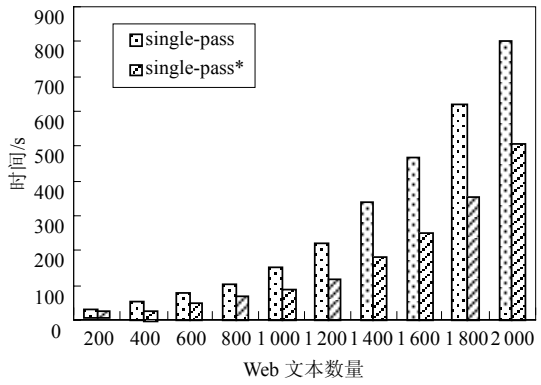

图1　算法运行时间对比

3.5　Web话题舆情分析

对Web文本应用single-pass*算法进行文本聚类的主要目的是在聚类产生的不同主题类中进行舆情分析。主题关注度是指过去某一时间段内，舆情主题被关注的程度。时间段t_1～t_2内关于舆情主题S的主题关注度：

$$R_S(t_1,\ t_2\ )=r_S(t_2)-r_S(t_1) \tag{11}$$

式中，$r_S(t)$表示关于某一个舆情主题S的相关页数随时间的变化。

以固定时间间隔作为统计周期，主题关注度用$P_i(i\in[1,\ 5])$表示，如：P_1={2009.4～2011.10}。表8显示5个不同时间段内部分舆情的主题关注度，图2显示热点话题主题关注度随周期的变化。

表8　话题主题关注度分析

话题主题	P_1	P_2	P_3	P_4	P_5
地震等新闻	12	360	490	564	15
智慧等文化	12	60	33	32	23
名胜等旅游	99	32	12	12	44
论藏等经济	48	49	32	25	31
花园等音乐	37	14	11	32	12
闲暇等视频	25	83	9	18	12

图2　藏文新闻主题关注度趋势分析

为了进一步验证舆情主题关注度算法的性能，观察随着文本数量的增加，主题关注度分析方法的时间性能的变化如图3所示。通过图3可以发现，算法运行时间近似呈线性增长，与式(11)的定义吻合。

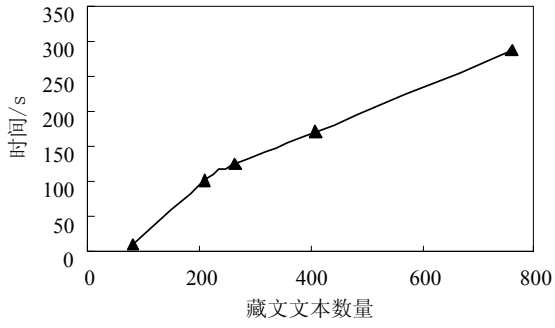

图3　不同文本数量下主题关注度算法运行时间

4　结 束 语

IT技术与互联网的迅猛发展，数据存储量爆炸性地增长，并行处理与计算已经成为越来越重要的数据挖掘的问题。从大数据中发现有价值的知识需

要各种高效的数据挖掘算法。single-pass聚类算法能够高效地发现话题，通过引入话题种子的概念，本文提出了改进的single-pass聚类算法。实验结果表明，本文提出的Web文本聚类算法不仅能够提高聚类的质量，即在漏检率、错检率和时间开销等方面均有所改善，而且对网络舆情分析的研究具有较好的应用价值。

参 考 文 献

[1] BAEZA-YATES R, RIBEIRO-NETO B. Modern information retrieval[M]. Boston, USA: Addison Wesley, 2000.

[2] 贾自艳, 何清, 张海俊, 等. 一种基于动态进化模型的事件探测和追踪算法[J]. 计算机研究与发展, 2004, 41(7): 1273-1280.
JIA Zi-yan, HE Qing, ZHANG Hai-jun, et al. A news event detection and tracking algorithm based on dynamic evolution model[J]. Journal of Computer Research and Development, 2004, 41(7): 1273-1280.

[3] GONG Z, JIA Z, LUO S, et al. An adaptive topic tracking approach based on single-pass clustering with sliding time window[C]//Proceedings of the 2011 International Conference on Computer Science and Network Technology. Washington DC, USA: IEEE Computer Society, 2011: 1311-1314.

[4] FORESTIERO A, CLARA P, GIANDOMENICO S. A single pass algorithm for clustering evolving data streams based on swarm intelligence[J]. Data Mining and Knowledge Discovery, 2013, 26(1): 1-26.

[5] 赵文清, 侯小可. 基于词共现图的中文微博新闻话题识别[J]. 智能系统学报, 2012, 7(5): 444-449.
ZHAO Wen-qing, HOU Xiao-ke. News topic recognition of Chinese microblog based on word co-occurrence graph[J]. CAAI Transactions on Intelligent Systems, 2012, 7(5): 444-449.

[6] ZHU Z, WANG P, JIA Z, et al. Network topic detection model based on text reconstructions[J]. Informatica, 2013, 37(4): 367-372.

[7] YANG C, YANG J, DING H, et al. A hot topic detection approach on Chinese microblogging[C]//Proceedings of the International Conference on Information Engineering and Applications (IEA) 2012. London: Springer, 2013: 411-420.

[8] 税仪冬, 瞿有利, 黄厚宽, 等. 周期分类和Single-Pass聚类相结合的话题识别与跟踪方法[J]. 北京交通大学学报, 2009, 33(5): 85-87.
SHUI Yi-dong, QU You-li, HUANG Hou-kuan, et al. A new topic detection and tracking approach combining periodic classification and single-pass clustering[J]. Journal of Beijing Jiaotong University, 2009, 33(5): 85-87.

[9] 张晓燕, 王挺. 话题发现与追踪技术研究[J]. 计算机科学与探索, 2009, 3(4): 347-357.
ZHANG Xiao-yan, WANG Ting. Research of technologies on topic detection and tracking[J]. Journal of Frontiers of Computer Science and Technology, 2009, 3(4): 347-357.

[10] 康健, 乔少杰, 格桑多吉, 等. 基于群体智能的半结构化藏文文本聚类算法[J]. 模式识别与人工智能, 2014, 27(7): 663-671.
KANG Jian, QIAO Shao-qie, GESANG Duoji, et al. A semi- structured Tibetan text clustering algorithm based on swarm intelligence[J]. Pattern Recognition and Artificial Intelligence, 2014, 27(7): 663-671.

编 辑 蒋 晓

(上接第598页)

[7] CARVAJAL A. Quantitative comparison between the use of 3D vs 2D visualization tools to present building design proposals to non-spatial skilled end users[C]//9th International Conference on Information Visualisation. Washington DC, USA: IEEE Computer Society, 2005: 291-294.

[8] STOTT D T, GREENWALD L G, KREIDL O P, et al. Tolerating adversaries in the estimation of network parameters from noisy data: a nonlinear filtering approach[C]//Military Communications Conference, 2009. Boston, MA, USA: IEEE, 2009: 1-7.

[9] LAU S, RED C, NIMDA B, et al. The magazine archive includes every article published in communications of the ACM for over the past 50 years[J]. Communications of the ACM, 2004, 47(6): 25-26.

[10] WARE C. Information visualization: Perception for design[M]. Waltham, MA: Elsevier, 2013.

[11] KOIKE H, OHNO K. SnortView: Visualization system of snort logs[C]//Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security. New York, USA: ACM, 2004: 143-147.

[12] COCKBURN A. Revisiting 2D vs 3D implications on spatial memory[C]//Proceedings of the 5th Conference on Australasian User Interface. Sydney, Australia: Australian Computer Society Inc, 2004: 25-31.

[13] HUBONA G S, WHEELER P N, SHIRAH G W, et al. The relative contributions of stereo, lighting, and background scenes in promoting 3D depth visualization[J]. ACM Transactions on Computer-Human Interaction (TOCHI), 1999, 6(3): 214-242.

[14] SHIRAVI H, SHIRAVI A, GHORBANI A A. A survey of visualization systems for network security[J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(8): 1313-1329.

[15] NUNNALLY T, CHI P, ABDULLAHK, et al. P3D: a parallel 3D coordinate visualization for advanced network scans[C]//2013 IEEE International Conference on Communications (ICC). Budapest, Hungary: IEEE, 2013: 2052- 2057.

[16] NUNNALLY T, ULUAGAC A S, COPELAND J A, et al. 3DSVAT: a 3D stereoscopic vulnerability assessment tool for network security[C]//2012 IEEE 37th conference on Local Computer. FL: IEEE. 2012: 111- 118.

[17] NAEDELE M. Standards for XML and Web services security[J]. Computer, 2003, 36(4): 96-98.

[18] SHOEMAKE K. Animating rotation with quaternion curves[C]//ACM SIGGRAPH computer graphics. New York, USA: ACM, 1985, 19(3): 245-254.

[19] Vacommunity.VAST challenge homepage in vacommuity [EB/OL]. [2014-10-11]. http://www. vacommuity.org/VAST, 2013.

[20] BOSTOCK M. D3 Example[EB/OL]. [2014-10-11]. https://github.com/mbostock/d3/wiki/Gallery

[21] BODUROV V. VectorVisualizer[EB/OL]. [2014-10- 11]. http://www.bodurov.com/VectorVisualizer/.

编 辑 蒋 晓