

定位关节点的方法具有更好的适用性.

笔者提出的人体关节点定位方法,还有较大的改进空间,比如角点检测算法的效率有待提高等.

参考文献:

[1] SHOTTON J, KIPMAN A, et al. Real-time human pose recognition in parts from single depth images[J]. Communications of the ACM, 2013, 56 (1): 116 - 124.

[2] 肖秦琨,赵艳,高嵩. 基于 RGB 彩色和深度信息的人体关节点定位[J]. 国外电子测量技术, 2015 (2): 27 - 30.

[3] 余涛. Kinect 应用开发实战用:最自然的方式与机器对话[M]. 北京: 机械工业出版社, 2012.

[4] KO S J, LEE Y H. Center weighted median filters and their applications to image enhancement [J]. IEEE transactions on circuits and systems, 1991, 38 (9): 984 - 993.

[5] JACKWAY P T, DERICHE M. Scale-Space properties of the multiscale morphological dilation-erosion [J]. IEEE transactions on pattern analysis & machine intelligence, 1996, 18(1): 38 - 51.

[6] 孙少林,马志强,汤伟. 灰度图像二值化算法研究 [J]. 价值工程, 2010(5): 142 - 143.

[7] ZHANG T Y, SUEN C Y. A fast parallel algorithm for

thinning digital patterns [J]. Communications of the ACM, 1984, 27(3): 236 - 239.

[8] SHI J, TOMASI C. Good features to track [C] // 1994 Computer Vision and Pattern Recognition. Seattle: IEEE, 2002:593 - 600.

[9] NOUAR O D, ALI G, RAPHAEL C. Improved object tracking with camshift algorithm [C] // 2006 IEEE International Conference on Acoustics, Speech and Signal Processing. Toulouse: IEEE, 2006: 1121 - 1127.

[10] SOLEIMANIZADEH S, MOHAMAD D, SABA T, et al. Recognition of partially occluded objects based on the three different color spaces (RGB, YCbCr, HSV) [J]. 3D Research, 2015, 6(3): 1 - 10.

[11] 姚红革,郝重阳,雷松则. 序列图像中彩色目标跟踪的加权颜色分布方法 [J]. 中国图象图形学报, 2009(1): 99 - 105.

[12] COMANICIU D, MEER P. Meanshift: a robust approach toward feature space analysis [J]. IEEE transactions on pattern analysis and machine intelligence, 2002, 24(5): 603 - 619.

[13] GARY B, ADRIAN K. 学习 OpenCV(中文版) [M]. 北京: 清华大学出版社, 2012.

[14] WANG L, ZHANG Y, FENG J. On the euclidean distance of images [J]. IEEE transactions on pattern analysis & machine intelligence, 2005, 27(8):1334 - 1339.

A Method of Positioning the Human Body Joints Based on RGB and Depth Information

ZHANG Qingjian, HAN Jianping

(School of Computer, HangzhouDianzi University, Hangzhou 310016, China)

Abstract: In this paper, a method of positioning the human joints based on RGB-D was proposed. The human body skeleton line was obtained by the deep image filtering, dilating and thinning processing of the kinect acquisition, the end and turning points of the human skeleton line were obtained according to the corner detection algorithm. Then tracked and positioned the human's face in the color image, at last the 15 joints of human body were positioned combining with the proportion of the rigid parts of the human body. The results showed that the accuracy of positioning joints of this method was improved by 7% compared with the method of positioning the joints directly according to the proportion of the rigid parts of the human body, and the accuracy of positioning the joint of head was improved by 42% .

Key words: RGB-D; kinect; image thinning; corner detection; facing tracking; positioning joints

文章编号:1671-6833(2018)05-0039-06

基于多元媒体数据的教育舆情情绪可视化

梁翊涛,王长波

(华东师范大学 计算机科学与软件工程学院,上海 200062)

摘 要:教育舆情研究在危机公关、舆情引导等领域具有重要作用.目前,网络已经成为民众发表文章和评论的主要阵地.然而,如何识别网络教育舆情中的情绪倾向与情绪传播模式,并且在众多媒体平台中对比分析教育舆情情绪的特点,是当前舆情研究面临的挑战.针对该挑战,我们设计实现了教育舆情情绪可视化系统.首先,根据与目标用户的讨论定义了教育舆情情绪分析的需求;其次,利用情绪识别算法识别文本情绪,并根据情绪传播模型提出了情绪传播算法;然后,设计了多个可交互视图,允许用户进行多媒体平台的对比分析.最后,基于 2015 年网络教育舆情数据进行案例分析,表明本文设计的可视化系统能够满足用户需求,并能支持舆情引导、危机公关中的媒体平台选择.

关键词:教育舆情;可视分析;情绪分析;网络媒体数据;情绪传播

中图分类号:TP391 文献标志码:A doi:10.13705/j.issn.1671-6833.2018.05.002

0 引言

舆情是社会民众作为主体,发表对于时事、新闻、政策等社会态度的总和,在社会生活中具有重要意义.随着网络的发展,截至 2016 年 12 月,我国网民数量已经达到 7.13 亿^[1].因此,网络舆情已经成为舆情研究的主要阵地.研究网络教育舆情有助于引导公众舆论^[2],危机公关,并可以辅助政府、企业进行决策^[3].民众态度是舆情研究的核心领域,情绪作为民众态度倾向的直观反映,研究它能够分析网络用户、研判教育形势^[4]提供切实依据.然而,大量的网络舆情数据源,包括门户网站、论坛、博客、自媒体等,让人们难以选择合适的媒体平台获取信息、发表观点,或者进行危机公关等舆情管理活动.

因此,笔者设计开发了一个可视化交互系统,帮助用户进行教育舆情情绪的网络媒体对比研究.系统允许用户从两个层面对教育舆情进行比较研究.在宏观层面,系统揭示了教育舆情事件主题和舆情事件的情绪倾向;在微观层面,对舆情事件情绪进行多平台的对比分析.

1 需求分析

为了充分研究网络教育舆情,我们曾与教育领域的专家合作进行了网络教育舆情项目的研究^[5].相比普通网络舆情,教育领域的网络舆情具有下列特点^[6]:①意见主体的隐匿性,即民众可以通过匿名的方式在网上表达自己的观点;②教育客体的特殊性,即教育对于个体和社会的发展具有重要意义;③舆情传播的即时性与互动性.由于舆情参与者可以自由交互,因此教育事件更容易表现出“一边倒”的舆情倾向,更易产生群体性事件.

1.1 数据描述

首先,通过分布式爬虫技术,从腾讯网、搜狐网、凤凰网和新浪网 4 个门户网站上爬取了自 2015 年 1 月到 2015 年 12 月共计 2 744 篇文章和与之对应的 34 730 条评论.

1.2 用户参与

为了深入分析需求,我们选择了两类用户:教育领域专家和无专业知识但关心教育舆情的普通用户,如教师、家长、网络意见领袖等.我们将领域专家与普通用户组织成为一个志愿者小组,并与

收稿日期:2018-02-19;修订日期:2018-04-03
基金项目:上海市科委软科学基金资助项目(17692104400)
通信作者:王长波(1976—),男,湖北随州人,华东师范大学教授,博导,CCF 会员,研究方向为计算机图形学、数字媒体、信息可视化、数据分析、虚拟现实,E-mail:cbwang@sei.ecnu.edu.cn.

小组完成了 3 次线上访谈。

首先,由用户描述他们感兴趣的问题,为第一次访谈定义了在当前教育舆情分析中现存的问题和挑战,然后,为其提供基础的可视化形式,如条形图、折线图等,辅助用户进行问题的初步研究,并记录反馈。第二次访谈,展示根据用户需求所设计的可视化系统原型,测试是否符合用户需求,并收集反馈以改进系统。第三次访谈展示系统的最终版本,并指导用户使用。

1.3 提取用户需求

对于整体教育舆情情绪以及舆情事件的研究,不同类型的用户需求不同。如教师、家长希望分析他们关注的单个舆情事件,帮助他们决定哪种媒体平台更适合获取信息;意见领袖希望了解情绪从文章到评论的传播形式,从而帮助他们更有效吸引读者关注;领域专家与教育从业者(如辅导机构创业者等)则想要全面了解教育舆情,从而帮助他们制定决策或商业计划。

我们将用户系统需求总结为以下几点:①从不同层次研究网络媒体数据,例如从所有事件到单个事件;②对比不同媒体平台上文章到评论的情绪传播模式;③针对不同用户,总结最适于其发布、获取教育信息的平台类型。

2 多维情绪传播分析

2.1 情绪识别及分类

为了分析舆情文本情绪倾向,笔者采用累计情绪词在文本中出现次数的方法,为文档进行打分。根据情绪词典^[7]将情绪分为 7 个类别,并在领域专家的指导下,将 7 类情绪按照从正面到负面的顺序进行排布:乐、好、惊、哀、惧、恶、怒。此外,为了提高情绪得分的准确度,引入副词的权重计算。文档的 7 维情绪如公式(1)所示:

$$d = \{e_i = \sum_{j=1}^n score \times weight \mid D_1 \leq i \leq D_{max}, i \in N^+ \}.$$
 (1)

式中: d 为一篇文档(文章或者评论)的 7 维情绪得分向量; D_1 与 D_{max} 为要计算的情绪类别索引, $D_1 = 1, D_{max} = 7$,表示要计算所有 7 种情绪类别; e_i 为某一特定情绪类别的得分,并被标准化; j 为文档中某一类情绪词的数量; $score$ 为一个情绪词在情绪词典中的分值,根据副词的类型和情感激烈程度,将副词分为 $adv = \{most, very, more, insufficiently, ish, inverse\}$,并在多次试验后将其权重设定为 $weight = \{2.0, 1.75, 1.5, 1.2, 0.5,$

$-1.0\}$ 。无副词时 $weight$ 默认为 1.0。

文章表示为 $A = \{d_a, d_c\}$ 。 d_a 为文章 A 的情绪向量, d_c 为与 A 对应评论的情绪向量。一个舆情事件 $E = \{A_1, A_2, \dots, A_n\}$ 。在公式(2)中, E_1, E_2 为两个事件, D 为应用欧几里得距离公式计算两个舆情事件之间的情绪距离,用 D 衡量事件之间的情绪相似度,

$$D(E_1, E_2) = \sqrt{\sum_{i=D_1}^{D_{max}} (e_{E_{1i}} - e_{E_{2i}})^2}.$$
 (2)

2.2 情绪传播

目前,舆情情绪传播模型主要关注群体情绪对于群体性事件发展的影响力^[8],研究包括基于群际情绪理论的情绪层级模型^[9];基于系统动力学的负面情绪研究^[10]等。笔者集中关注不同网络媒体平台上文章到评论的情绪传播,根据与领域专家的讨论,上述传播具有明显的链式结构。因此,基于信息系统生态链模型^[11],笔者将网络媒体平台上的情绪传播定义为:情绪从媒体所发表的文章向读者传播,最后以评论的方式展示。研究单个事件中的情绪传播,需要定义情绪传播的类型和衡量情绪传播的效率 P ,

$$P = \left\{ P_i = \frac{e_i}{\sum_{j=D_1}^{D_{max}} e_j} \mid D_1 \leq i \leq D_{max}, i \in N^+ \right\}.$$
 (3)

笔者将情绪传播定义为两类:①同类情绪传播:文章与其评论的主情绪类型一致;②异类情绪传播:文章与其评论的主情绪类型不一致。若文章的主情绪得分小于评论的主情绪得分,则为增量传播,反之为减量传播。

同类情绪传播的效率 μ 计算公式如下:

$$\mu_{同} = |P_{a[main_a]} - P_{c[main_c]}|.$$
 (4)

异类情绪传播的效率计算公式如下:

$$\mu_{异} = [(|P_{a[main_a]} - P_{c[main_a]}|) + (|P_{a[main_c]} - P_{c[main_c]}|)] / 2,$$
 (5)

式中: P_a 为文章情绪百分比向量; P_c 为评论情绪百分比向量; $main_a, main_c$ 分别为文章和其评论的主情绪类型索引值。

3 系统设计与可视分析

可视化系统包含了 3 个模块:数据处理模块、情绪分析模块、可视化模块。系统结构如图 1 所示。数据预处理模块:从网页上爬取文本数据,使用 LDA 算法^[12]识别舆情事件。情绪分析模块:识别文本情绪,对情绪进行相关性分析,并应用

MDS 算法^[13]进行降维显示,分析文章到评论的情绪传播.可视化模块:通过情绪花朵图、情绪传播图 etc 视图对教育舆情情绪进行交互对比分析.

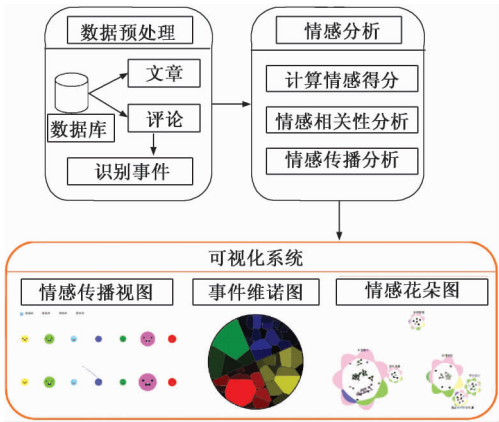


图1 系统结构图

Fig. 1 The structure of system

3.1 数据预处理

笔者应用 LDA 算法从文档中识别出 91 个教育舆情事件,并根据与领域专家的讨论,总结出 6 个舆情事件主题:校园暴力、高考腐败、禁止中小学生补课、学生创业、乡村教育与留守儿童.

3.2 可视化设计

可视化系统如图 2 所示,系统包含 5 个可交互视图,允许用户在多个层次上对网络教育舆情进行探索研究.

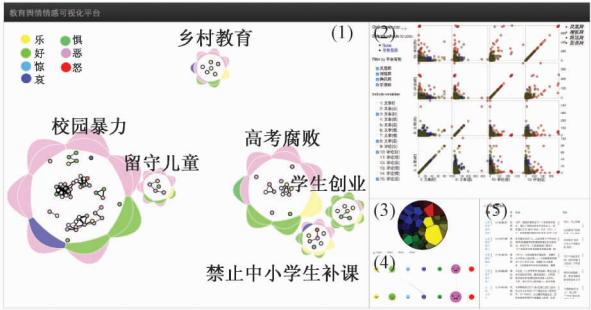


图2 可视化系统图

Fig. 2 The overview of system

3.2.1 情绪花朵图

如图 2 中(1)所示,情绪花朵图展示了所有舆情事件的情绪倾向.一朵花表示一个舆情主题,花朵的大小表示主题中舆情事件的数量.花瓣的数量表示每个主题中不同情绪类型百分比,因为情绪分为 7 种类型,所以每朵花有 7 片花瓣.花瓣分为内外两个部分,内侧表示文章情绪,外侧表示评论情绪,内外对比展示出舆情主题文章和评论情绪的差异.为了表示 7 种情绪类型,笔者应用了 Plutchik 情绪轮模型^[14],如图 2 中(1)左上角的图

例所示,黄、青、蓝、紫、绿、粉红、红分别代表乐、好、惊、哀、惧、恶、怒 7 种情绪.

以导向布局作为花心,展示舆情主题中事件的情绪相似度聚类.每个节点表示一个舆情事件,节点的颜色表示其主情绪类型.同时,我们使用 2.1 节描述的方法度量事件之间的情绪相似度,如果相似度大于阈值,就在两个事件节点之间添加一条边,由此形成主题中的事件情绪子聚类.通过多次试验,笔者将阈值设置为 0.1.

3.2.2 事件维诺图

为了对比同一事件在不同媒体平台上文档数量、情绪得分的差异,笔者应用了事件维诺图^[15].如图 2 中(3)所示,从媒体平台、文章、评论 3 个层次展示.第一层通过颜色划分不同的媒体平台,凤凰网、搜狐网、腾讯网、新浪网分别用红、绿、蓝、黄表示,以颜色亮度表示情绪得分的大小;第二层表示在相应平台上发表的文章;第三层表示文章对应的评论.

3.2.3 情绪传播图

如图 2 中(4)所示,情绪传播图展示同一事件在不同媒体平台上,从文章到评论的情绪传播情况.第一行 7 个圆展示文章的情绪得分,第二行展示评论的情绪得分.情绪传播的效率通过两行圆之间的连线表示.用户可以通过按钮切换不同的媒体平台,对比其情绪传播情况的差异.此外,圆中卡通标记能让用户直观理解不同情绪类型,强化颜色和情绪类型的对应关系.

3.2.4 矩阵散点图

用户可以通过矩阵散点图从所有事件文章与单个事件文章两种层次来分析情绪相关性,如图 2 中(2)所示.在矩阵散点图中,每一个点都代表一篇文章,点击文章点可以在文本视图中查看具体文章与评论.用户可以通过设置矩阵散点图的维度选择想要观察的情绪类型,媒体平台的颜色编码与事件维诺图相同.

3.2.5 文本视图

系统中设计了文本视图,用以展示文章与评论的原始数据,如图 2 中(5)所示.花朵图、事件维诺图、矩阵散点图均可与文本视图进行交互.

4 案例分析

笔者采用案例分析法^[16]对 2015 年全年舆情数据进行分析.首先,向用户展示如何使用系统.然后,根据用户提出的感兴趣的研究问题,选择合适案例进行分析.最后,记录用户的反馈并讨论用

户的分析结果.

4.1 2015 年网络教育舆情情绪分析

2015 年教育舆情花朵图如图 3 所示,其中校园暴力主题在 2015 年相关的舆情事件最多,总体情绪偏向恶.但文章和评论情绪仍然有不同,文章里面有 28.57% 好的情绪,哀占 14.28%;但评论中恶占 85.71%,好占 14.28%,没有哀伤情绪.对于上述情况,领域专家认为出现这种情况的原因是:“媒体平台的表达必须客观中立,因此情绪表达较为克制,但用户能够自由评论,对于校园暴力这样的负面话题就体现出恶评.”此外,图中有一个显著倾向于好的话题,即“学生创业”,因为 2015 年,国家出台一系列政策鼓励学生创业,社会各界对于创业都持积极态度.

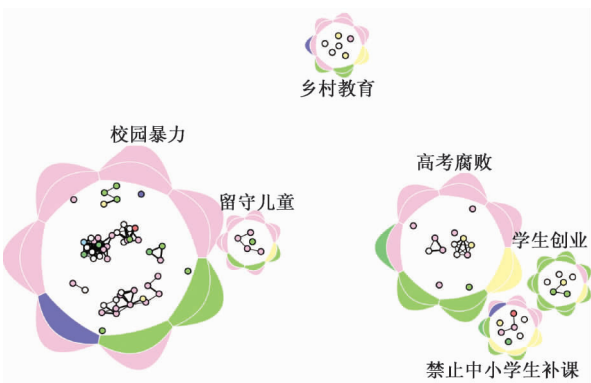


图 3 舆情花朵图

Fig.3 The flower view of public opinion

图 4 为分平台展示的舆情文本矩阵散点图.腾讯网的文章(散点)数量最多,且分布较其他 3 个平台平均,并无明显的高情绪得分文章.从情绪来说,腾讯网上的“喷子”(不关心事实,只发泄情绪的网民)较多,情绪倾向于负面.

凤凰网的某些点评论(恶)的情绪得分显著较高,其他点都集中在左下角,即除了个别文章受大众广泛关注外,其他文章情绪倾向并不突出.新浪网的分布与腾讯网类似,但相比腾讯网,新浪网的文章(好)、文章(恶)的情绪得分均偏低,即新浪网不倾向于发表情绪倾向强烈的文章.最后,搜狐网的文章情绪得分平均,但评论的得分多为 0,表示在搜狐网上网民的参与度相比其他 3 个平台低.“虽然新浪网上的报道文章较少,但其文章态度明显中立,对我来说是更好的获取信息的平台.”一位用户表示,“搜狐网的参与度太低,即使上面发表评论也不会得到什么反馈.”

4.2 “江西高考替考案”事件分析

江西高考替考案是 2015 年关注热度最高的

舆情事件,搜狐网、新浪网、腾讯网、凤凰网共计有 40 篇文章对此进行了报道.

2015 年 6 月 7 日中午,江西省教育厅接到有人组织替考的举报.经查,外省替考组织在网上招揽高校学生,串通招考办及医院有关工作人员为外省籍考生在江西违规报名、体检,从而实施替考的有组织、有预谋的高考舞弊案件.

通过文本视图对高情绪得分的文章进行分析,发现用户评论共计 529 条,且大部分都是对高考替考行为的批判.然而,虽然评论的情绪恶得分很高,评论的正面得分也较高.其原因首先是存在一些反讽的评论,如“记者装着一副大义凛然的感觉,觉得自己真的很伟大吗?”;其次,也有对于替考卧底记者勇于揭露行为的赞扬,如“这是正能量崛起,有更多这种利国利民的揭发,说明了正义之花处处开放.”其中,“大义凛然”“伟大”“利国利民”,均为正面词汇.因此,评论的正面情绪较高可以总结为以下两个原因:①虽然高考替考案为负面事件,勇于揭发的行为亦值得赞扬;②部分民众评论时惯于使用反讽,其正面词汇在情绪词典方法中会造成误判.

如图 5 所示,情绪传播图展示了不同媒体平台从文章到评论的情绪传播情况.除搜狐网的文章、评论太少,无法体现特征以外,凤凰网、腾讯网、新浪网的文章情绪情况相似:虽然恶的情绪占比较高,但也有一定比例的乐与好.文章到评论的情绪传播却体现出了一定的差异性,凤凰网的评论情绪比文章情绪更加激烈,同时一部分文章(乐)的情绪向评论(哀)传播;腾讯网和新浪网的文章情绪得分均高于评论情绪,显示在这两个平台上用户评论的情绪并不高.

4.3 案例分析总结

总体而言,2015 年教育舆情呈现出负面倾向.除负面事件频发的原因外,无论是从文章发布还是网民评论来看,民众对于负面事件的参与度都比正面事件更高.对媒体平台而言,观点多样化和意见参与度高的网络媒体是更理想的信息发表、获取平台.意见领袖等舆论引导者如果希望获得更多关注,他们会选择凤凰网.想获取信息的用户则偏爱民众参与度更高,情绪倾向分布也更平均的腾讯网.

5 结论

笔者首先通过以用户为中心的设计流程,总结了教育舆情情绪分析领域特定的目标和设计原

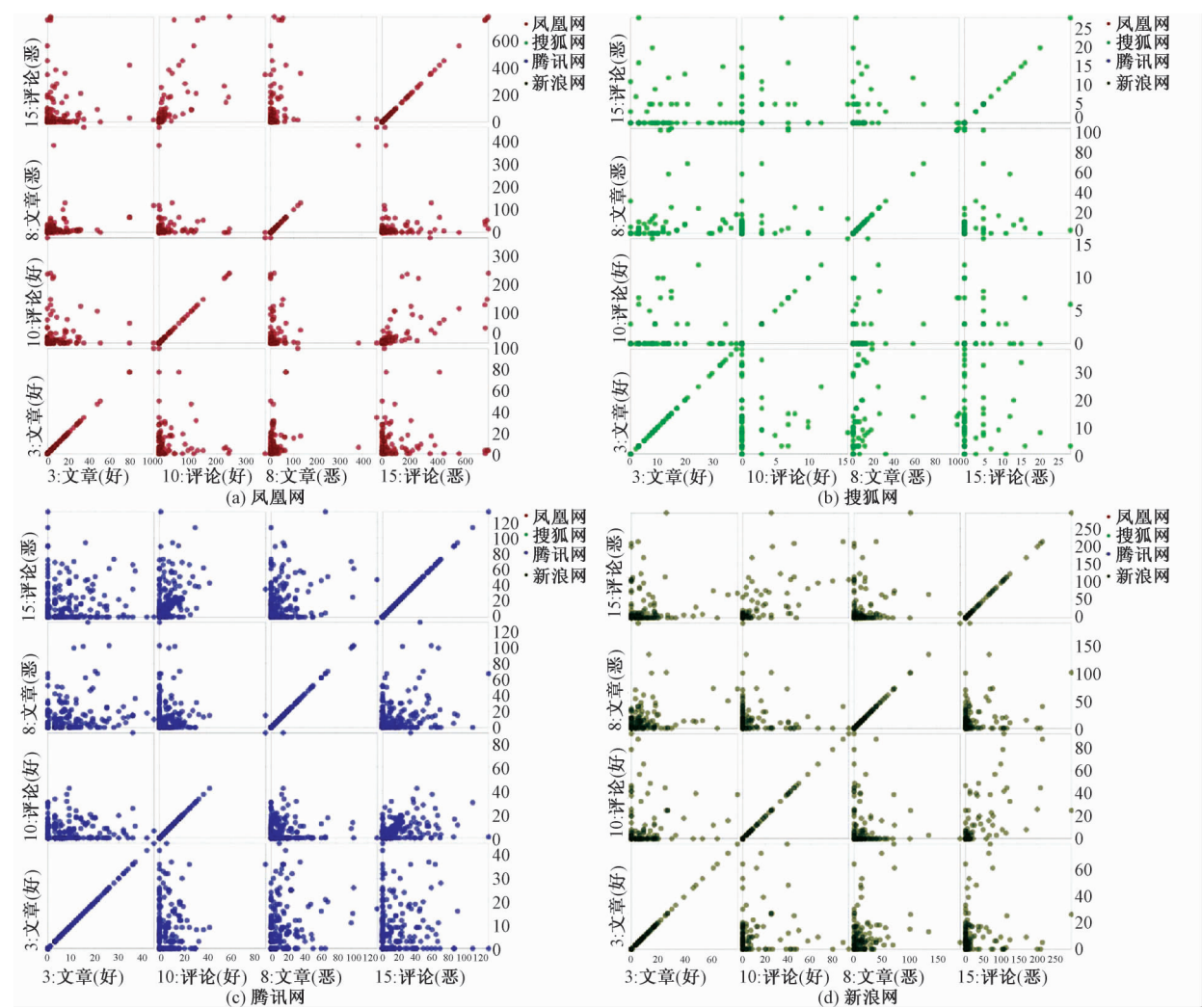


图4 凤凰网、搜狐网、腾讯网、新浪网情绪矩阵散点图

Fig. 4 The emotion matrix views of Ifeng, Sohu, Tencent, and Sina

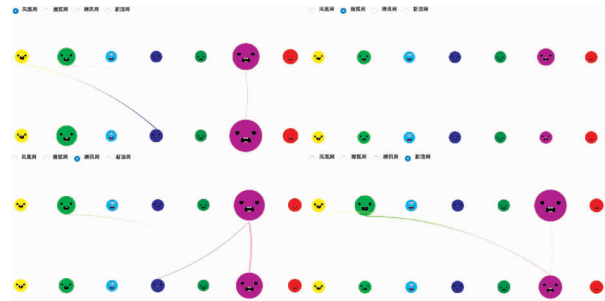


图5 事件情绪传播图

Fig. 5 The emotion propagation view.

则。其次,对网络舆情文本进行情绪分析,定义了文章到评论的情绪传播类型。再次,设计开发了基于教育舆情情绪的可视化平台,展示教育舆情情绪倾向及不同媒体平台的情绪传播模式,帮助用户选择合适的媒体平台发表、获取信息。最后,案例分析证明了系统的可用性,给舆情情绪分析,媒体比较分析和可视化三者的交叉领域提出了一种新的可能。后续研究将考虑提升文本情绪识别的算法精度,并实时分析显示舆情情绪。

参考文献:

[1] 马捷,胡漠,张世良,等. 网络舆情危机等级评价模型构建及其应用:以廉政网络舆情为例[J]. 情报资料工作,2017(4):36-42.

[2] 姜胜洪. 网络舆情热点的形成与发展、现状及舆论引导[J]. 理论月刊,2008(4):34-36.

[3] 肖文涛,许小美. 新媒体时代的网络舆论现状与引导对策[J]. 行政论坛,2012,19(6):79-84.

[4] 肖宇,许炜,夏霖. 一种基于情绪倾向分析的网络团体意见领袖识别算法[J]. 计算机科学,2012,39(2):34-37+46.

[5] 何晓丰,朱益明,王长波,等. 2015年中国教育网络舆情分析报告[M]. 上海:华东师范大学出版社,2016.

[6] 左明章,国桂环,曾睿. 网络教育舆情的内涵及特点[J]. 中国教育信息化,2011(6):8-11.

[7] 徐琳宏,林鸿飞,潘宇,等. 情绪词汇本体的构造[J]. 情报学报,2008,27(2):180-185.

[8] 汪明艳,陈梅. 社交媒体网络舆情传播影响力研究

- 综述[J]. 情报科学, 2017,35(5):171-176.
- [9] 刘志明, 刘鲁. 面向突发事件的群体情绪监控预警[J]. 系统工程, 2010,28(7):66-73.
- [10] 李从东, 洪宇翔, 谢天. 基于系统动力学的群体负面情绪动力机制研究[J]. 情报杂志, 2013, 32(5):38-42.
- [11] 李美娣. 信息生态系统的剖析[J]. 情报杂志, 1998, 17(4):3-5.
- [12] PAN C C, MITRA P. Event detection with spatial latent Dirichlet allocation[C] // In Proceedings of the 11th Annual International ACM/IEEE, Joint Conference on Digital Libraries, ACM, 2011: 349-358.
- [13] CAMBRIA E, SONG Y, WANG H, et al. Semantic multidimensional scaling for open-domain emotion analysis [J]. IEEE intelligent systems, 2014, 29(2): 44-51.
- [14] PLUTCHIK R. The nature of emotions human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice [J]. American scientist, 2001, 89(4): 344-350.
- [15] HENRY P, VINES P. Voronoi treemaps in D3 [EB/OL]. <https://cse512-14w.github.io/fp-plvines-djpeter/final/paper-plvines-djpeter.pdf>. (2014-2-13).
- [16] 焦宝臣, 陈诗明, 刘振昌, 等. 研究生管理信息系统应用效果评价研究[J]. 郑州大学学报(工学版), 2017,38(2):9-12.

Visualization of Public Opinion Emotion in Education Based on Multiple Media Data

LIANG Yitao, WANG Changbo

(Department of Computer Science and Software Engineering, East China Normal University, 200062, China)

Abstract: The research of public opinion on education had important values in the field of crisis management, public opinion guidance and so on. At present, Internet has become the main channel for people to express their opinions. However, distilling the emotion trend of public opinion and propagation mechanisms of the emotion, and analyzing the characteristics of emotion on multiple media platforms were still challenges. Therefore, a visualization system of emotion on education public opinion was designed. Firstly, the requirements of emotion analysis of public opinion were defined after the discussions with the users. Secondly, the emotion recognition algorithm was used to recognize emotion of text, and the emotion propagation algorithm was proposed based on the emotion propagation model. Thirdly, multiple interactive views were designed, which allows users to do comparative analysis on different online media platforms, and analyzed propagation mechanisms of emotion. Lastly, case studies on network data of education in 2015 showed that the system could meet the requirements of users and effectively support users' selection of media platforms in public opinion guidance and crisis management.

Key words: public opinion in education; visual analysis; emotion analysis; network media data; propagation of emotion