

## 可视化驱动的交互式数据挖掘方法综述

马昱欣<sup>1)</sup>, 曹震东<sup>1)</sup>, 陈 为<sup>1,2)\*</sup>

<sup>1)</sup>(浙江大学 CAD&CG 国家重点实验室 杭州 310058)

<sup>2)</sup>(浙江大学赛博(CYBER)协同创新中心 杭州 310058)  
(chenwei@cad.zju.edu.cn)

**摘要:** 数据挖掘是一种从大量数据中发现信息的过程, 其大量依赖自动算法的特质, 使得用户难以对数据和算法过程本身直观地进行理解、探索和优化。近年来, 随着可视化领域的蓬勃发展, 有很多工作开始探究如何使用可视化方法辅助数据挖掘过程, 使用户更加直观地理解数据, 并对数据和算法和进行探索。文中首先对数据挖掘和可视化在知识提取流程进行比较分析, 并从可视化增强的通用数据挖掘方法和面向应用场景的方法 2 个方面对近年相关技术进行梳理总结, 并依托一些相关主题的国际会议内容指出需要进一步探索的方向。

**关键词:** 可视化; 数据挖掘; 交互式分析  
**中图法分类号:** TP391.41

## A Survey of Visualization-Driven Interactive Data Mining Approaches

Ma Yuxin<sup>1)</sup>, Cao Zhendong<sup>1)</sup>, and Chen Wei<sup>1,2)\*</sup>

<sup>1)</sup>(State Key Laboratory of CAD&CG, Zhejiang University, Hangzhou 310058)

<sup>2)</sup>(Innovation Joint Research Center for Cyber-Physical-Society System, Zhejiang University, Hangzhou 310058)

**Abstract:** Data mining is dedicated to retrieve knowledge from massive datasets by utilizing automated algorithms. However, due to the characteristic of automation processes, current data mining approaches can hardly allow the user to visually understand, explore and optimize the datasets and the computation process. Recently an increasing number of researchers in the field of visualization have been focusing on visualization-based interactive data mining approaches. With the assistance of visualization, users can gain insight and perform exploration from datasets and the mining results intuitively. In this paper we compare data mining and visualization process from the aspect of knowledge discovery. Additionally we classify the recent works into two main categories: 1) visual-enhanced general data mining approaches, and 2) application-based approaches. Additionally we propose a set of future challenges according to recent related conferences.

**Key words:** visualization; data mining; interactive analysis

### 1 引言

随着大数据时代的来临, 各种常见数据分析

方法已在研究界和工业界得到广泛研究和应用。其中, 数据挖掘方法作为一类主要的方法, 在近几十年来一直是研究者们所关注的热点之一<sup>[1]</sup>。

作为一种基于人类视觉通道进行数据分析的

收稿日期: 2015-10-29; 修回日期: 2015-11-10. 基金项目: 国家“九七三”重点基础研究发展计划项目(2015CB352503); 国家自然科学基金重点项目(61232012); 国家自然科学基金(61422211); 浙江省自然科学基金(LR13F020001); 中央高校基本科研业务费专项资金资助。马昱欣(1989—), 男, 博士研究生, 主要研究领域为信息可视化、可视分析; 曹震东(1993—), 男, 本科生, 主要研究领域为信息可视化; 陈 为(1976—), 男, 博士, 教授, 博士生导师, 论文通讯作者, 主要研究方向为科学可视化、信息可视化、可视分析等。

手段, 可视化在很多领域都发挥着帮助用户理解数据、发现数据特征, 进而辅助推理决策的作用<sup>[2-4]</sup>。近年来, 伴随可视化技术的发展, 基于可视化的分析方法开始在诸多领域发挥重要作用, 例如文本分析、社交网络分析、时空数据分析等。在数据挖掘领域, 由于用户需要对数据挖掘结果或过程进行理解、检测和验证, 可视化开始成为一种辅助数据挖掘过程的手段, 逐步发展出一系列基于可视化的数据挖掘方法<sup>[3,5-6]</sup>, 并用于辅助增强数据挖掘的可理解性、可交互性, 甚至是算法精度等。

### 1.1 数据挖掘与可视化的流程对比

图 1<sup>[6]</sup>展示了数据挖掘和可视化 2 种知识提取方式的流程。数据挖掘通常利用自动算法, 根据已有数据构建出相应模型。通过对模型进行解释和验证, 用户可以从其中提取出分析结果或蕴含的知识。可视化方法则通过视觉映射, 将输入数据映射为一定的视觉表达形式, 用户通过视觉通道从可视化结果中发现和提取符合需求的特征和模式, 以生成针对问题的假设。对于交互式可视化方法, 视觉模型还可接受用户的交互反馈, 并根据反馈生成新的可视化结果。

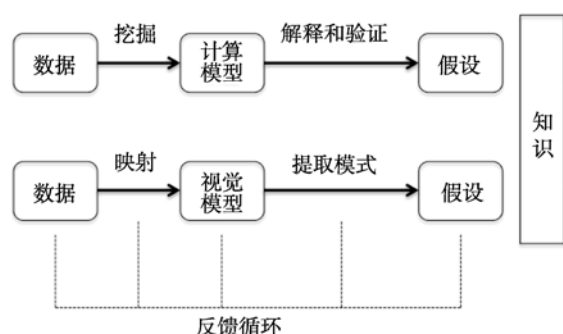


图 1 数据挖掘与信息可视化的流程对比

### 1.2 基于可视化的交互式数据挖掘方法分类

目前已有文献提出了针对数据挖掘与可视化结合方法的分类策略: 文献[6]提出了以可视化和数据挖掘的参与程度为分类标准的方法分类, 其中包括: 1) +V 型方法, 指以数据挖掘为主要的知识发现和提取方法, 可视化作为数据挖掘的辅助方法; 2) +M 型方法, 指以可视化方法为主, 其中数据挖掘方法帮助完成可视化中的数据处理、可视映射部分。从原始数据到知识表达的整个过程中, 不同方法在各个阶段都有其对应的主要任务, 如在从数据转化为可视表达的过程中, 数据挖掘方法可以帮助生成数据点在二维平面上的投影。

文献[7]提出了基于用户参与形式(types of user

involvement, TUI)的分类策略。TUI 用于描述基于用户的可视化探索参与数据挖掘算法的方式, 分别由以下 2 个维度组成:

#### 1) 信息流动方向

算法本身产生的过程描述信息形成算法对用户的反馈; 用户对算法的信息输入形成对算法的控制。

#### 2) 目标实体

信息交换阶段, 可以是算法运行过程时进行信息交换, 或是在算法运行完毕时, 通过可视化方法查看和理解算法结果。

以上 2 篇综述性文章主要从知识发现和信息抽取角度对数据挖掘和可视化结合的数据分析方法进行总结。本文以数据挖掘方法为主要分析对象, 根据可视化所增强的目标对象, 将基于可视化的交互式数据挖掘方法分为 2 大类:

#### 1) 可视化增强的通用数据挖掘方法

可视化增强的通用数据挖掘方法以数据挖掘算法和模型为主, 并不针对具体应用场景或数据类型。可视化作为辅助手段, 帮助更加直观地展示输入数据和计算结果, 或是直接深入模型内部, 使用户直接看到模型的核心数据和结构, 并进行交互式调整。

#### 2) 面向应用场景的方法

近年来, 数据挖掘与可视化方法的应用场合日趋广泛, 有不少文献基于特定应用场景提出了一些面向应用的方法。本文将按照数据或应用类型进行分类, 并对相关方法进行总结。

## 2 可视化增强的通用数据挖掘方法

可视化方法与通用数据挖掘方法的组合方式大致分为“黑盒方式”和“白盒方式”2种。

### 2.1 基于黑盒方式的可视化增强方法

通常, 用户都需要一些辅助查看算法的输入/输出等信息的手段。在基于黑盒方式的可视化增强方法中, 可视化设计者仅根据数据挖掘算法所面向的任务进行设计, 并不考虑具体算法的内部机制, 即把算法本身当作“黑盒”。例如在聚类任务中, 算法输出通常是输入数据点所属的类别, 因此针对聚类结果的可视化方法一般仅考虑算法所输出的类别信息, 而忽略算法本身是怎样运作的。该类方法的分析对象主要有以下几个类别:

#### 1) 面向输入数据的可视化方法

用户在使用数据挖掘方法之前, 可先使用

可视化方法对数据进行预探索. 这种方式有助于帮助用户获得输入数据的大致分布特征, 可以指导其后续的数据清洗、参数设定等工作. 文献[8]提出了一种可视化方法, 用于辅助子空间聚类算法. 该方法使用多个散点图并列排布的方式展示数据点所在的原始空间中各个维度的相关性, 用户通过观察散点图中的数据点相关度来筛选出冗余维度, 以得到不同子空间下的聚类结果.

### 2) 面向算法结果的可视化方法

数据挖掘算法的输出结果是用户最关心的内容, 可视化方法通常用于结果分析、模型验证等方面. 在分类任务方面, 文献[9-11]等工作主要针对分类结果的可视化; 文献[12]提出了基于分类混淆矩阵 ManiMatrix 系统, 使用户交互式地查看混淆矩阵中的内容和分类边界. 对于聚类任务, DICON 系统<sup>[13]</sup>使用基于面积分割的视觉编码展示聚类中的数据点在各个属性上的分布. 除单纯的结果展示之外, 文献[14]提出了一种“即时标记(just-in-time annotation)”策略, 辅助用户在聚类结果的探索过程中对数据分布特征进行语义标注. 文献[15]构建了一种用于展示话题模型中不同话题之间的相关程度的框架, 帮助用户发现重复或意义相近的冗余话题.

此外, 可视化还可以用于提高结果的可解释性, 使得非数据挖掘领域的专家用户也能从算法结果中得出一定的结论. Explainer 系统<sup>[16]</sup>提出了一种分类结果的解释方法, 将线性分类平面的法向方向解释为一个新的有实际意义的维度, 并将数据点按照该法向维度进行排序.

### 3) 迭代式可视化方法

除了单纯的输入或输出数据可视化之外, 迭代式可视化方法帮助用户基于已有探索结果对输入参数或数据进行修改, 继而生成新的模型和输出结果, 并使用迭代策略引导用户对算法结果进行优化. 文献[17-19]等工作针对分类任务提出了多种迭代式可视化方案. 在距离学习方面, Dis-Function 系统<sup>[20]</sup>使用基于距离的投影可视化方法, 使用户能够交互式指定训练数据点之间的距离关系, 引导距离度量的生成, 并根据新的距离度量矩阵生成新的投影结果, 以达到渐进式优化的目的. PIVE 系统<sup>[21]</sup>提出了一种迭代式框架, 用于实时展示数据挖掘算法中的迭代优化过程.

## 2.2 基于白盒方式的可视化增强方法

除了 2.1 节中提到的黑盒方式外, 近年来随着

可视化界对数据挖掘方法的关注度升高, 有不少工作已经开始着手对算法过程本身进行展示, 使用户能够更好地理解计算结果与输入数据、参数之间的关系.

决策树因其可解释性较强的特性, 在可视化界有不少针对决策树模型的可视化案例<sup>[22-24]</sup>. BababView<sup>[25]</sup>是近年来比较综合的决策树可视化系统, 它提供了基于节点链接方法的树可视化视图和分类结果混淆矩阵的可视化视图, 以及大量的交互方法. 用户可以在查看训练生成的决策树结构时, 对决策树直接进行修改, 如增删分支、调整判断条件等. 然而, 大部分数据挖掘模型的可理解程度都较低, 对于非数据挖掘领域的用户来说造成了一定的理解困难. 可视化在这一方面为用户提供了一定程度上的辅助. 文献[26]针对支持向量机提出了基于投影的可视化方案. 文献[27]构建了 EasySVM 系统, 它除了帮助用户理解支持向量机模型的核心结构及分类结果外, 还提出了一种可视化辅助的规则抽取方法. 文献[28]提出了一种面向图像识别的深度学习神经网络可视化方法, 其使用可视化方式展示出不同层次神经元上的特征, 帮助用户进行参数调整等工作.

## 3 面向应用场景的方法

### 3.1 文本分析

数据挖掘及可视化方法在文本分析中占有非常重要的地位. 常见应用类型有文本分类、聚类及异常检测等.

对于非机器学习专家来说, 使用机器学习算法进行文本分析学习成本较高, 短时间内无法精通. Heimerl 等<sup>[29]</sup>提出了一种如图 2<sup>[29]</sup>所示的交互式文本分类器训练系统, 以辅助普通文本分析人员进行文本分类分析. 该系统借鉴机器学习中主动学习策略, 降低了用户标记文本标签的成本, 并加快分类器的训练过程. 同时, 作者设计了一系列可视化视图以帮助理解文本集的大致词频分布和特征, 以及简化用户对所需文本的搜索过程. ScatterBlogs2 系统<sup>[30]</sup>使用类似的交互式分类器训练方法, 帮助用户监控及过滤感兴趣的微博文本.

文献[31]设计了一种基于 LDA(latent Dirichlet allocation)的可视化文本探索方法. 针对有类别标签的文本数据, 该方法首先抽取所有文本的主题

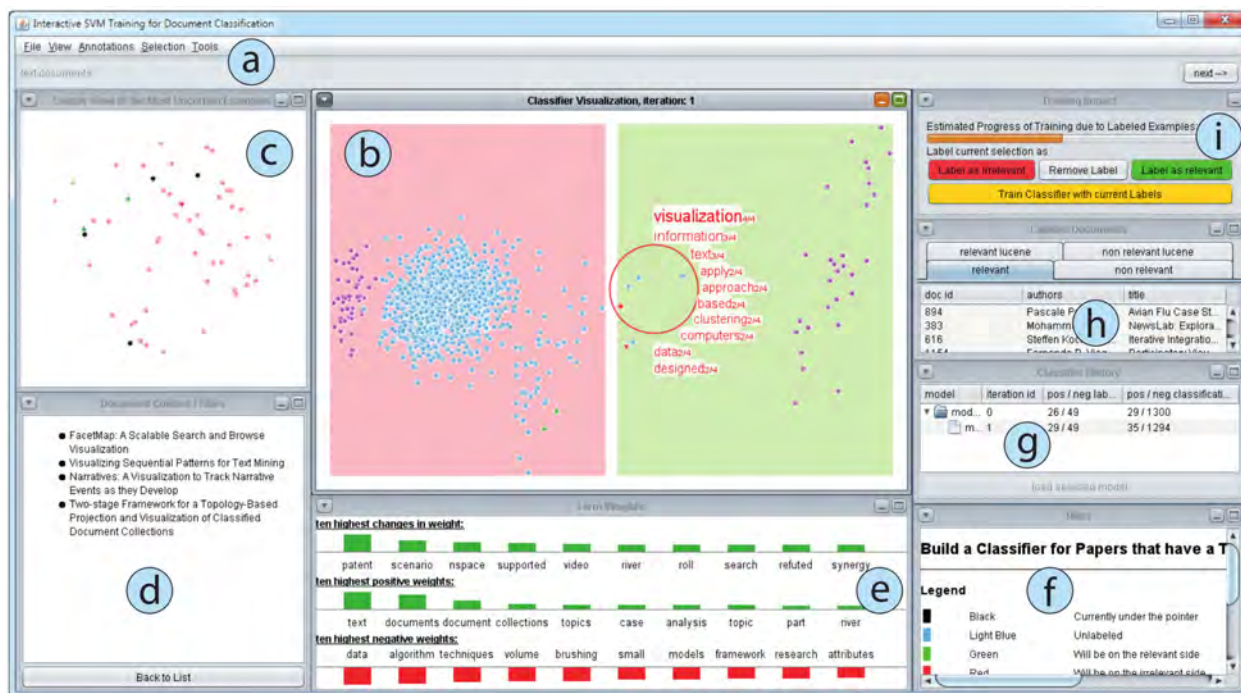


图 2 交互式文本分类器训练系统

分布, 并计算不同文本类别间的主题相似性, 最终使用基于散点图设计的地形视图(landmark view)展示类别间的相似性. 作者还提出了可视化文本分析中模型选取的 2 个原则: 1) 可解释性, 即用户可通过分析模型就能大致数据分布特征; 2) 可信度, 即可视化能够真实并准确反应数据特征. iVis-Clustering 系统<sup>[32]</sup>同样使用基于 LDA 模型的交互式可视化方法, 用于进行文本聚类分析. 该系统主要利用节点-链接图和平行坐标方法分别展示 LDA 模型的聚类结果和文档的主题分布.

除了上述针对主题抽取结果进行可视化的工作外, UTOPIAN 系统<sup>[33]</sup>对传统 LDA 模型进行改进, 提出一种基于交互式非负矩阵分解的用户驱动式主题抽取方法. 该方法允许用户通过该模型提供的接口交互式调整模型参数, 并根据先验知识对主题生成进行干预.

Zhao 等<sup>[34]</sup>提出的 FluxFlow 交互式可视分析系统用于探测和分析社交媒体文本流中的异常信息传播. 该系统利用单类条件随机场(one-class conditional random fields, OCCRF)作为核心机器学习算法, 用户可通过基于线索时间线的可视化方法发现当前文本流中的异常传播模式, 并深入探索相关文本内容.

### 3.2 图像分析

如图 3<sup>[35]</sup>所示, Hoferlin 等提出了针对视频图

像数据的交互式主动学习(inter-active learning)策略<sup>[35]</sup>. 在训练图像分类器时, 该策略将专家用户对图像的标记判断融入分类过程中. 专家用户可通过搜索或查看分类结果的方式找出最需要人工标记的图像, 这些标记后的训练图像能够使分类器的训练效率最大化.

如图 4<sup>[36]</sup>所示, 加州大学伯克利分校的 deepViz 系统<sup>[36]</sup>用于可视化面向图像分类任务的深度神经网络. 该工作的出发点是卷积网络的神经元都可直接表达为图像形式, 用户可以直接看到模型内部的中间图像结果. 通过查看每层神经元所代表的特征图像, 用户可在训练过程中大致了解神经元上的特征分布, 据此对相应模型参数进行调整, 以完成模型训练快速迭代.

谱聚类方法在图像分析中有着广泛的应用. 文献<sup>[37]</sup>基于传统的谱聚类方法, 提出了一种白盒式可视化增强的聚类系统. 该系统面向 CT, MRI 等三维医学图像数据, 支持用户利用专家知识, 通过交互式可视化方式调整谱聚类方法所需的参数、聚类数等变量.

### 3.3 用户行为分析

用户在同一网站上的页面浏览顺序形成用户在该网站的点击流. 针对用户点击流进行分析, 可以刻画出用户在网页间的浏览行为. 文献<sup>[38]</sup>使用自组织映射(self-organizing map)方法对用户点击

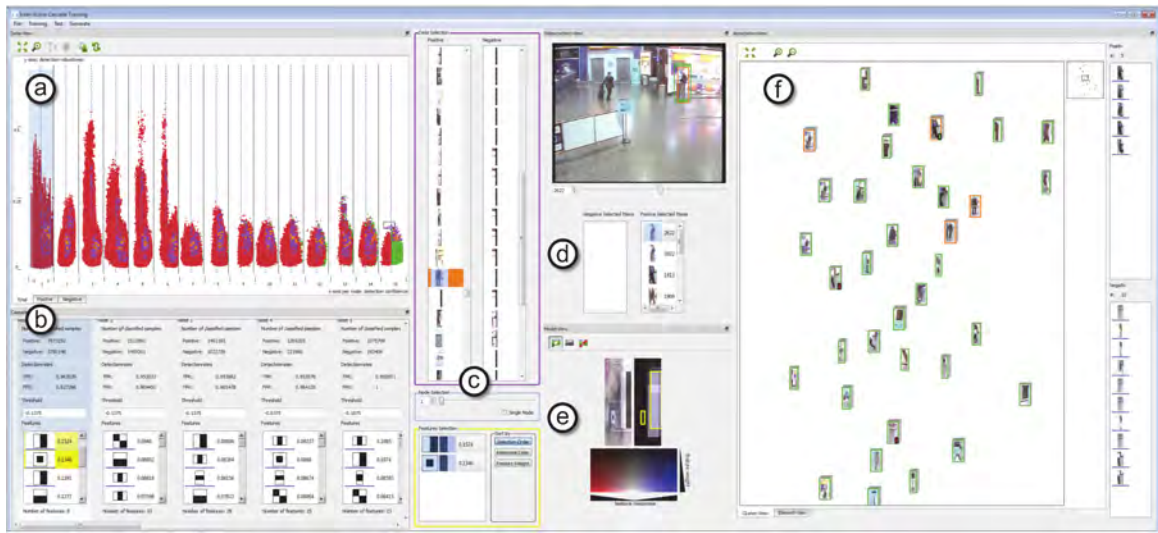


图 3 基于交互式主动学习策略的图像分类器训练系统

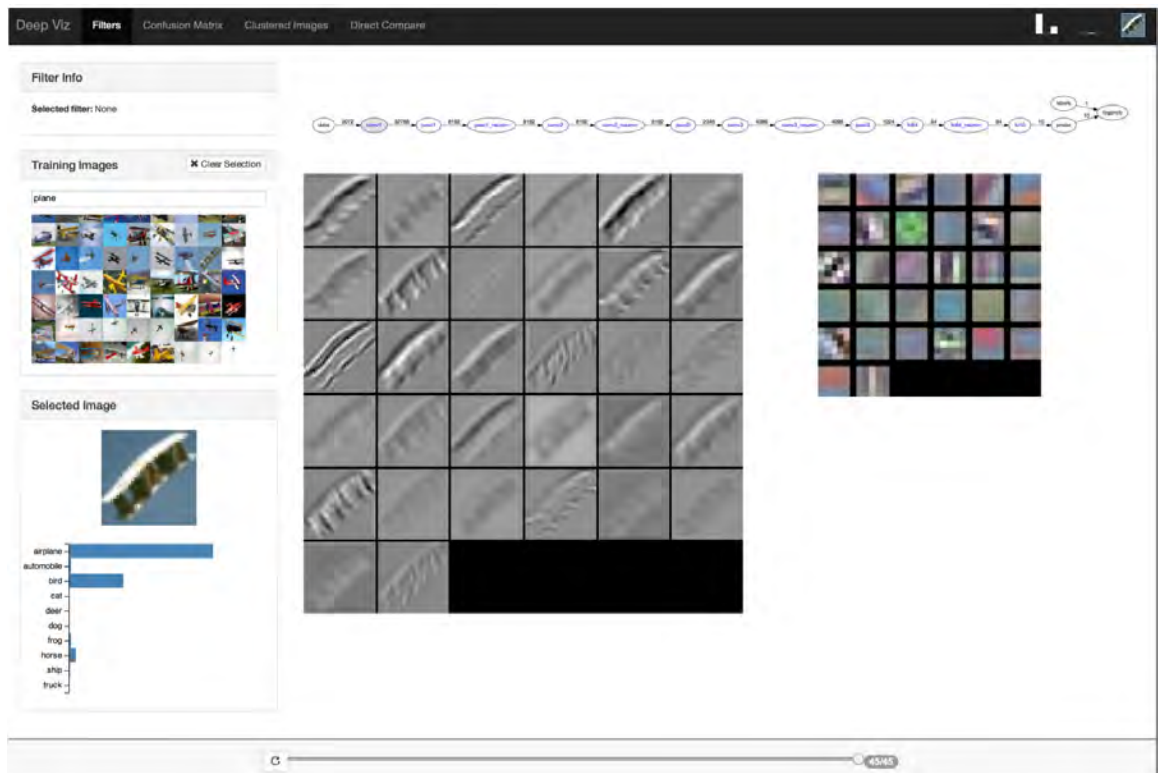


图 4 deepViz 系统界面

流进行投影分析, 使用色条的大小和在二维平面上的位置分布表示点击流的出现频率, 以及点击流之间的相似关系. 用户在投影视图上进行选中等交互操作, 可以详细查看某些点击流的具体属性分布. 通过对 eBay 网站的真实数据集进行实际案例分析, 用户能够在复杂的点击流中找到相关特征模式, 并对不同的典型模式进行比较分析.

社交网络是近年来数据挖掘和可视化方向的

研究热点之一, 其中用户之间的交互可用于刻画用户交往行为. 文献[39]面向社交网络中的社团发现任务, 提出了一种如图 5<sup>[39]</sup>所示基于可视分析的多上下文移动社交网络社团发现方法. 该方法使用上下文相关模型对社交网络中的上下文属性进行建模, 并设计了基于平行坐标的上下文属性可视化方案, 以帮助用户结合上下文属性探索网络中的社团结构.

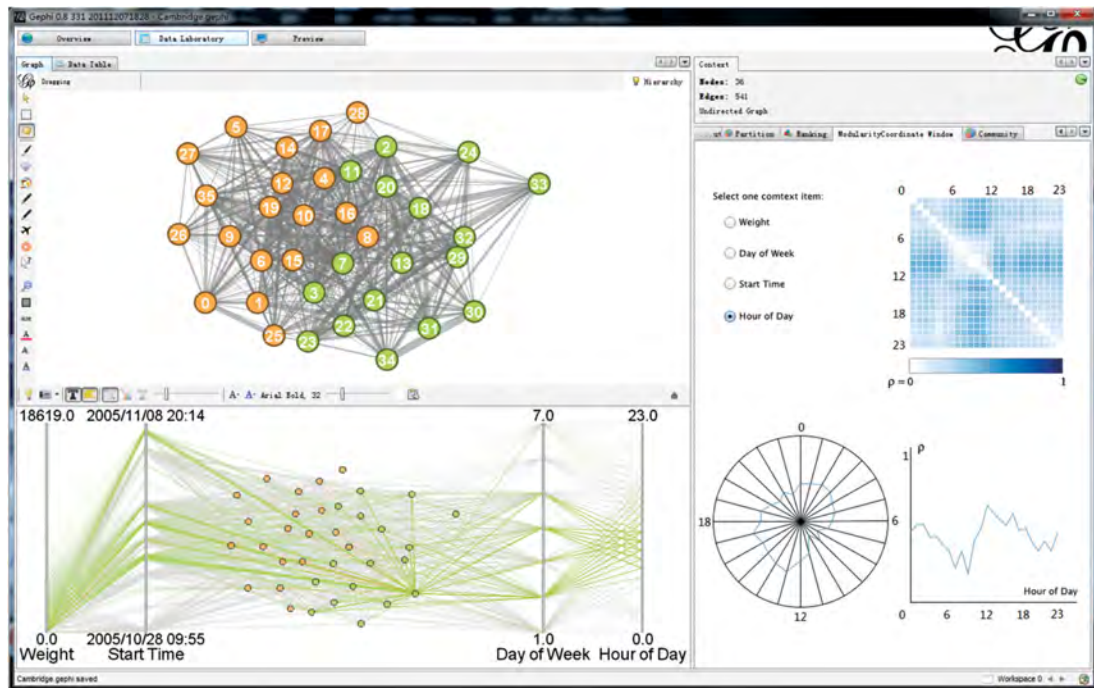


图 5 基于可视分析的多上下文移动社交网络社团发现方法

### 3.4 时空数据分析

时空数据挖掘重点关注带有时间和空间属性的数据。在可视化方面,同样有很多时序数据和地理空间数据的方法来帮助展示数据挖掘算法的结果和算法过程。针对微博或其他社交平台上带有时间戳和地点信息的文本流数据,文献[40]基于 LDA 模型和 STL(seasonal-trend decomposition procedure based on loess smoothing)方法分别提取某时间段内的文本主题和异常关键词,并可视化相关文本信息的地理位置和时序波动信息。用户可以通过多种交互方式渐进式探索和发现异常文本信息,并通过相关文本提取出事件。

对于非统计领域的专家来说,时序数据分析模型的模型选择工作是一件烦琐和复杂的任务。TiMoVA 系统<sup>[41]</sup>针对这一问题提出了一种基于可视分析方法的交互式模型选择流程。用户可借助这一系统对某一模型的分析结果进行验证,并根据验证结果选择重新训练或是接受该模型。在与时序分析模型的交互过程中,用户可以同时对时间序列数据进行探索,并从多次模型生成结果中抽取数据特征。

文献[42]提出了一种启发式空间数据聚类方法,将基于计算几何的空间聚类算法与交互式可视化结合在一个同一框架下。用户可从某个算法输出的启发式结果出发,对数据的聚类特征和离

群点进行探索式分析,并调整算法参数以进行迭代式分析。

### 3.5 其他应用场景

除上述应用场景外,可视化方法还应用于其他相关领域。文献[43]提出了结合用户反馈的聚类方法,用户可根据动物学专业知识对数据点的类别进行人为限制,以分析蝙蝠耳朵形状和声学特征。在高维数据分析方面,INFUSE 系统<sup>[44]</sup>借助交互式可视化方法对特征选取任务进行辅助和增强。通过展示特征在多种特征评价方法上的排名,用户可以看到特征的排名分布,以及不同的特征集合在分类算法上的测试效果,进而决定最终需要选出的特征。在回归分析方面,HyperMoVal 系统<sup>[45]</sup>和文献[46]均提出了新的可视化设计,以展示回归分析中的数据分布和模型的训练结果。

## 4 总结与未来发展

本文对基于可视化的交互式数据挖掘方法进行了综述,并针对现有研究成果,介绍了可视化增强的通用数据挖掘方法和面向应用场景的方法。在很多场景下,2 种类型的方法可以相互借鉴,如针对某种数据挖掘任务的通用方法可经过改进后为一些特定应用场景和数据类型服务。

近年来,随着可视化方法与数据挖掘方法的

不断结合, 基于可视化的交互式数据挖掘方法开始逐步展现其强大的数据分析能力和交互效率. 在 IEEE VIS 2014 Workshop on Visualization in Predictive Analysis 上, 可视化领域的专家提出了该方向的一系列主要挑战, 包括训练数据和测试数据选择、特征选择与测试、模型比较与选择、模型验证、模型预测质量控制以及可视化辅助的复杂问题预测等. 这些将成为该研究方向未来需要攻克的问题.

## 参考文献(References):

- [1] Han J, Kamber M, Pei J. Data mining: concepts and techniques[M]. 3rd edition. Burlington: Morgan Kaufmann. 2011: 3-5
- [2] Chen W, Guo F, Wang F Y. A survey of traffic data visualization[J]. IEEE Transactions on Intelligent Transportation Systems, 2015
- [3] De Oliveira M C F, Levkowitz H. From visual data exploration to visual data mining: a survey[J]. IEEE Transactions on Visualization and Computer Graphics, 2003, 9(3): 378-394
- [4] Chen Wei, Sheng Zheqian, Tao Yubo. Data visualization[M]. Beijing: PublishingHouse of Electronics Industry. 2013: 20-23 (in Chinese)  
陈 为, 沈则潜, 陶煜波. 数据可视化[M]. 北京: 电子工业出版社. 2013: 20-23
- [5] Ma K L. Machine learning to boost the next generation of visualization technology[J]. IEEE Computer Graphics and Applications, 2007, 27(5): 6-9
- [6] Bertini E, Lalanne D. Surveying the complementary role of automatic data analysis and visualization in knowledge discovery[C] //Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration. New York: ACM Press, 2009: 12-20
- [7] Muhlbacher T, Piringer H, Gratzl S, et al. Opening the black box: Strategies for increased user involvement in existing algorithm implementations[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1643-1652
- [8] Tatu A, Maas F, Farber I, et al. Subspace search and visualization to make sense of alternative clusterings in high-dimensional data[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2012: 63-72
- [9] Becker B, Kohavi R, Sommerfield D. Visualizing the simple Bayesian classifier[J]. Information Visualization in Data Mining and Knowledge Discovery, 2001, 18: 237-249
- [10] Wang J, Yu B, Gasser L. Classification Visualization with Shaded Similarity Matrix[R]. Champaign: University of Illinois at Urbana-Champaign, 2002
- [11] Alsallakh B, Hambury A, Hauser H, et al. Visual methods for analyzing probabilistic classification data[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1703-1712
- [12] Kapoor A, Lee B, Tan D, et al. Interactive optimization for steering machine classification[C] //Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2010: 1343-1352
- [13] Cao N, Gotz D, Sun J, et al. Dicon: Interactive visual analysis of multidimensional clusters[J]. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(12): 2581-2590
- [14] Kandogan E. Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2012: 73-82
- [15] Chuang J, Gupta S, Manning C, et al. Topic model diagnostics: Assessing domain relevance via topical alignment[C] //Proceedings of the 30th International Conference on Machine Learning. Madison: Omnipress, 2013: 612-620
- [16] Gleicher M. Explainers: Expert explorations with crafted projections[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2042-2051
- [17] Talbot J, Lee B, Kapoor A, et al. EnsembleMatrix: interactive visualization to support machine learning with multiple classifiers[C] //Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2009: 1283-1292
- [18] Behrisch M, Korkmaz F, Shao L, et al. Feedback-driven interactive exploration of large multidimensional data supported by visual classifier[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2014: 43-52
- [19] Paiva J G S, Schwartz W R, Pedrini H, et al. An approach to supporting incremental visual data classification[J]. IEEE Transactions on Visualization and Computer Graphics, 2015, 21(1): 4-17
- [20] Brown E T, Liu J, Brodley C E, et al. Dis-function: Learning distance functions interactively[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2012: 83-92
- [21] Choo J, Lee C, Park H. PIVE: a per-iteration visualization environment for supporting real-time interactions with computational methods[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2014: 241-242
- [22] Ankerst M, Elsen C, Ester M, et al. Visual classification: an interactive approach to decision tree construction[C] //Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 1999: 392-396
- [23] Ankerst M, Ester M, Kriegel H P. Towards an effective cooperation of the user and the computer for classification[C] //Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2000: 179-188
- [24] Ware M, Frank E, Holmes G, et al. Interactive machine learning: letting users build classifiers[J]. International Journal of Human-Computer Studies, 2001, 55(3): 281-292.
- [25] van den Elzen S, Van Wijk J J. Baobabview: interactive construction and analysis of decision trees[C] //Proceedings of

- IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2011: 151-160
- [26] Caragea D, Cook D, Honavar V G. Gaining insights into support vector machine pattern classifiers using projection-based tour methods[C] //Proceedings of the 7th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2001: 251-256
- [27] Ma Y X, Chen W, Ma X H, *et al.* EasySVM: a visual analysis approach for open-box support vector machines[C] //Proceedings of IEEE VIS Workshop on Visualization for Predictive Analytics. Los Alamitos: IEEE Computer Society Press, 2014: 20
- [28] Erhan D, Bengio Y, Courville A, *et al.* Visualizing higher-layer features of a deep network[R]. Montreal: Université de Montréal, 2009
- [29] Heimerl F, Koch S, Bosch H, *et al.* Visual classifier training for text document retrieval[J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(12): 2839-2848
- [30] Bosch H, Thom D, Heimerl F, *et al.* Scatterblogs2: Real-time monitoring of microblog messages through user-guided filtering[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2022-2031
- [31] Chuang J, Ramage D, Manning C, *et al.* Interpretation and trust: Designing model-driven visualizations for text analysis[C] //Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 2012: 443-452
- [32] Lee H, Kihm J, Choo J, *et al.* iVisClustering: an interactive visual document clustering via topic modeling[J]. Computer Graphics Forum, 2012, 31(3pt3): 1155-1164
- [33] Choo J, Lee C, Reddy C K, *et al.* Utopian: User-driven topic modeling based on interactive nonnegative matrix factorization[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 1992-2001
- [34] Zhao J, Cao N, Wen Z, *et al.* #FluxFlow: Visual analysis of anomalous information spreading on social media[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1773-1782
- [35] Hoferlin B, Netzel R, Hoferlin M, *et al.* Interactive learning of ad-hoc classifiers for video visual analytics[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2012: 23-32
- [36] Bruckner D, Rosen J, Sparks E. deepViz: visualizing convolutional neural networks for image classification[OL]. [2015-11-10]. <https://github.com/bruckner/deepViz>
- [37] Schultz T, Kindlmann G L. Open-box spectral clustering: applications to medical image analysis[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2100-2108
- [38] Wei J, Shen Z, Sundaresan N, *et al.* Visual cluster exploration of web clickstream data[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2012: 3-12
- [39] Ma Y X, Xu J Y, Peng D C, *et al.* A visual analysis approach for community detection of multi-context mobile social networks [J]. Journal of Computer Science and Technology, 2013, 28(5): 797-809
- [40] Chae J, Thom D, Bosch H, *et al.* Spatiotemporal social media analytics for abnormal event detection and examination using seasonal-trend decomposition[C] //Proceedings of IEEE Conference on Visual Analytics Science and Technology. Los Alamitos: IEEE Computer Society Press, 2012: 143-152
- [41] Bogl M, Aigner W, Filzmoser P, *et al.* Visual analytics for model selection in time series analysis[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2237-2246
- [42] Packer E, Bak P, Nikkila M, *et al.* Visual analytics for spatial clustering: Using a heuristic approach for guided exploration[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2179-2188
- [43] Hossain M S, Ojili P K R, Grimm C, *et al.* Scatter/gather clustering: Flexibly incorporating user feedback to steer clustering results[J]. IEEE Transactions on Visualization and Computer Graphics, 2012, 18(12): 2829-2838
- [44] Krause J, Perer A, Bertini E. INFUSE: Interactive feature selection for predictive modeling of high dimensional data[J]. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1614-1623
- [45] Piringner H, Berger W, Krasser J. HyperMoVal: interactive visual validation of regression models for real time simulation[J]. Computer Graphics Forum, 2010, 29(3): 983-992
- [46] Muhlbacher T, Piringner H. A partition-based framework for building and validating regression models[J]. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 1962-1971