文章编号：2096-1472(2018)-03-27-03
DOI:10.19644/j.cnki.issn2096-1472.2018.03.007

聚类算法及其在护理管理中的应用研究

降 惠
(长治医学院算机教学部，山西 长治 046000)

摘 要：在传统的医疗过程中，医护人员随机分配，患者被动接受治疗的护理管理模式不仅会浪费有限的医疗资源，使患者承受较高医疗费用，有时还会延误最佳治疗时机，引发并发症，给患者和其家庭带来更多的痛苦和困扰。文章通过将几种常见的聚类算法在护理管理中的应用进行比较，最终选择了先验假设较少的层次聚类算法，基于R语言探讨了层次聚类算法分类护理患者的实现过程，说明了层次聚类算法在患者与护理资源的最优匹配中的应用方法，为医院开展科学护理管理提供一定的参考依据。

关键词：聚类算法；护理管理；层次
中图分类号：TP399 **文献标识码**：A

A Study on the Application of Clustering Algorithm on Nursing Management

JIANG Hui
(*Department of Computer Teaching,Changzhi Medical College,Changzhi* 046000,*China*)

Abstract: In the process of traditional healthcare,health care providers are randomly assigned and patients are passive to accept various treatments.Besides the high expense of medical treatment and the waste of limited medical resources,it would delay the best treatment time and cause complications,bringing great suffering to patients and their families.By comparing clustering algorithms,this paper chooses the hierarchical clustering algorithm with fewer priori assumptions.Based on R Language,this paper elaborates on the implementation process of hierarchical clustering algorithm in clustering patients,and the methods of using hierarchical algorithm to optimally match the patients and nursing resources,thus providing certain reference and basis to the scientific nursing management.

Keywords:clustering algorithm;nursing management;hierarchical

1 引言(Introduction)

护理管理是医疗机构以改善和提高医疗水平，合理利用医疗设备，最大程度发挥医护人员护理能力为目标的过程。然而，在传统的医疗护理管理过程中，医护人员随机分配，患者被动接受治疗方案，不仅会浪费医疗资源，使患者承受较高医疗费用，有时还会延误治疗，引发并发症，给患者和其家庭带来更多痛苦和困扰。随着信息化“智慧医疗”时代的到来，医院信息系统的普及，医疗数据采集、整理、分析不再受到传统人工操作在时间、效率、成本和疗效等方面的制约，取而代之的是高效、便捷、智能、科学的大数据分析处理过程。“互联网+医疗护理”的模式已悄然走近我们身边。许多学者开始尝试通过计算机算法精准识别患者、最优分配护理医生、优化患者转诊流程等方式来提高护理效果和降低医疗费用。在识别患者和分配医生时，为了更好地进行分类识别，聚类算法受到了广泛关注。

2 聚类算法(Clustering algorithm)

聚类就是按照“物以类聚”的思想，将抽象数据集划分成若干簇的过程，其中在每一簇中数据间具有高度的相似性[1,2]。聚类分析是当前数据挖掘中的重要手段。迄今，研究者已提出了多种聚类算法[3]。常用的聚类算法有基于划分、密度、网

格、模型和层次的方法等[4]。

2.1 基于划分的聚类方法

基于划分的聚类方法需要提前设定聚类的数目，如K均值法和K中心点法等。这类方法分析时需要首先在数据集中随机选择聚类数目的对象，每个对象作为该聚类的中心或平均值，通过距离远近等方法将其余对象逐步聚类到每个类中。基于划分的聚类分析结果与初始值关系非常密切，不同的初始值往往会导致不同的聚类结果，通常会呈现出局部最优解。基于这种情况，基于划分的聚类方法在实际应用中往往要求穷举所有可能的划分。

2.2 基于密度的聚类方法

基于密度的聚类分析依据数据是否属于相连的密度域将数据对象进行归类。常用的基于密度的聚类算法为DBSCAN。这种聚类方法可以剔除噪声点，同时可以发现任意形状的聚类而不局限于球状的类[5]。但DBSCAN必须指定邻域半径和最少点数这两个参数，同时聚类的结果对这两个参数的依赖性很强。

2.3 基于网格的聚类方法

基于网格的聚类方法将空间分成有限数目的多维网格，利用网格结构实现聚类。如CLIQUE算法和STING算法。CLIQUE算法主要适用于处理高维数据，同时它综合了密度算法，可以发现其中的密集聚类。STING算法是一种将空间划分成直方形网格单元，不同方形对应不同分辨率，以实现聚类分析的方法[6]。

2.4 基于模型的聚类算法

基于模型的聚类算法是通过寻找数据对每个类预先设置好的模型的最佳拟合而实现的。最典型的基于模型的聚类算法是COBWEB算法。COBWEB算法不需要用户提供参数，可以自动划分簇的数目[7]。

2.5 基于层次的聚类算法

基于层次的聚类算法对数据集按照凝聚或分裂的层次方式进行聚类。凝聚型将每个对象看作一个单独的类，然后通过合并相近的类实现聚类。分裂型将所有数据归到一个类中，然后通过迭代，分层分裂成小类。常见的基于层次的聚类算法有CURE、ROCK和Chamelemon。Chamelemon算法在合并类的过程中，可以综合考虑类的内在特征、近似度和互连性，可以构造任意大小的聚集簇[8]。

2.6 几种聚类算法比较分析

基于模型的聚类算法要求对象在每个属性具有独立的概率分布，但在实际生活中，这种假设是不存在的。基于网格的聚类算法需要选择合适大小的单元数目，基于密度的聚类算法和基于划分的聚类算法需要用户设置一定的参数来产生可接受的聚类结果。在基于层次的聚类算法中，执行合并或分裂后无法回退修正，其质量受到一定的影响，但由于层次聚类算法无需提前知道最终所需的集群数量，其先验假设较少，因此通用性很强。目前，层次聚类算法已广泛应用于各个领域。

3 护理管理中聚类算法的应用研究(The application of clustering algorithm in Nursing management)

3.1 算法选择

在护理管理中，聚类可以帮助科研人员从医院信息管理系统中区分患者，聚类分析作为数据挖掘的重要分析方法，可以发现每类患者更深层次的信息，归纳总结出每类患者的个体特征和临床检验特征，有针对性的为患者提供精准治疗，减轻患者心理和财力方面的负担，缩短就医疗程，提高患者的生活质量。根据患者个体参数合理聚类患者往往不同于其他领域聚类分析。首先，患者的个体特征属性并不满足独立的概率分布。患者某项指标的偏离往往依赖于其他体征。其次，护理管理初期通过设置一定的参数来产生可接受的聚类结果，需要具备丰富的医学经验和较高的计算机理论水平，实际应用中难度很大。再者，患者病情差异较大，选择合适大小的单元数目可操作性较低。综合分析几种聚类算法，层次聚类算法需要的先验假设更少，对患者的聚类集群更科学，更适合当前大数据背景下医院开展科学、高效的护理管理，因此本文基于凝聚型层次聚类算法探讨聚类算法在护理管理中的应用。

3.2 数据预处理

护理管理中我们收到的患者数据往往具有异构性、冗余性、复杂性等特征，在对数据进行分析时首先需要进行集成、清理和归约[9]。

数据集成：患者个体数据来源主要有几个方面：(1)医学检验中心；(2)体检中心；(3)急诊科室；(4)门诊科室；(5)住院科室；(6)医学影像诊断中心；(7)社区卫生服务中心；(8)县级医院；(9)专科医院。不同的医疗机构对数据的存储格式、方

法和实体的命名规则不同，在数据集成过程中需要考虑统一实体、删除冗余数据、处理冲突数据等。集成过程中可以考虑可视化分析和相关性分析等方法实现。

数据清理：由于医疗数据的特殊性，临床检验数据获取时间较长，人工填补缺失值困难大，而替代邻近值处理噪声数据又可能导致分析结果的不准确性，因此清理数据主要通过识别异常值并删除相应元组的方法实现。

数据归约：随着大数据时代的到来，可获得的患者个体数据越来越多，对患者属性进行归约处理可以有效地降低分析的时间复杂度，提高分析质量。因此，在分析过程中可以通过属性约简等方法归约处理数据。

3.3 基于R语言的层次聚类算法及其在护理管理中的应用

在数据预处理的基础上，将每个患者视为一类，利用欧几里得距离、曼哈顿距离、切比雪夫距离等方法计算出每类之间的相似度。接着，将相似度最高的合并为一个类，重复计算，直至相似度大于某个终止条件值时结束聚类。简言之，层次聚类算法是通过计算每个类中的数据点与所有数据点之间的距离来确定其相似性，距离越小，相似度越高[10]。在R语言中，具体实现步骤如下：

(1)计算距离矩阵：利用dist()函数计算患者间的距离矩阵。

d<−dist(x,method=”euclidean”)

其中，x为患者数据集，method表示计算距离的方法，常用的方法有euclidean、manhattan、maximum等。方法的选择可根据患者具体数据而不同。

(2)聚类分析：在距离矩阵的基础上，使用hclust( )函数合并聚类簇。

hc<−hclust(d,method=“ward.D2”)

其中，d为患者数据的距离矩阵，method表示合并的方法，常用的方法有complete、ward、median、average、centroid等。

(3)剪枝操作：在数据分析中，患者数据的层次聚类结果有时很复杂，cutree()函数可以实现对聚类结果的剪枝操作，分析时可综合层次聚类结果和医生数选择合适的聚类数。

clusterCut<−cutree(hc,k)

其中，k为层次聚类结果的指定聚类数。最佳聚类数的选择是层次聚类剪枝中最复杂的一个问题，可以使用mclust、Nbclust、factoextra等包计算出最优聚类数。

4 护理管理分析(The analysis of Nursing management)

在实际护理管理中，广义上讲护理管理人员涵盖的范围不仅包括医院护士、家庭护理人员[11]，还包括专家和医生。传统的护理管理中，专家、住院医生、护士随机分配和组合，往往会产生不同的治疗效果和医疗费用。为了改善这一现状，部分学者开始探讨优化护理流程、改善护理模式来提高护理质量。但这些方法虽然在某种程度上改善了护理质量，但护理方法的针对性较差。因此，在护理管理分析中，可以考虑事先仿真不同医生对每位患者的治疗效果。使用聚类分析方法，将患者进行科学分类，综合聚类结果、患者最终的治愈情况和医疗费用，发现每类患者的最佳护理者，从而为医院开展科学护理管理提供一定的参考依据。

5 结论(Conclusion)

随着“互联网+医疗护理”的发展，信息化技术在医学中的运用一定会越来越普及，越来越智能。我们探讨基于R语言的层次聚类算法在护理中的应用，就是要充分发挥层次聚类算法的先验假设较少的优点，可以直接对患者进行科学分类，达到患者与护理资源的最优匹配，既有效提高临床护理效率和效果，又能节省有限的医疗人力、物力、财力，最终达到医患关系的和谐融洽，相信随着互联网时代的到来，人工智能、计算机算法必将渗透到医疗护理的每一个环节，深入学习探讨各种算法在临床护理实际中的运用，将具有较大的科研和现实意义。

参考文献(References)

[1] 王有为,吴迪.聚类技术在医学数据中的应用[J].河南教育学院学报(自然科学版),2016,2:32−35.

[2] 李雪梅,张素琴.数据挖掘中聚类分析技术的应用[J].武汉大学学报,2009,42(3):396−399.

[3] 陈黎飞,姜青山,王声瑞.基于层次划分的最佳聚类数确定方法[J].软件学报,2008,19(1):62−72.

[4] Jiawei Han,MichelineKamber.范明,孟小峰,译.数据挖掘概念与技术[M].北京:机械工业出版社,2004:1−262.

[5] 杨海斌.一种新的层次聚类算法的研究与应用[D].兰州:西北师范大学,2011:8.

[6] 张鑫.层次聚类算法的研究与应用[D].赣州:江西理工大

学,2008:19.

[7] 段明秀.层次聚类算法的研究及应用[D].长沙:中南大学,2009:9.

[8] 毕鹏.改进的Chameleon层次聚类算法在目标分群中的应用研究[D].杭州:浙江大学,2009:17.

[9] 尤婷婷.健康大数据预处理技术及其应用[D].成都:电子科技大学,2017:6-16.

[10] 张良均,谢佳标,杨坦等.R语言与数据挖掘[M].北京:机械工业出版社,2017(8):204-206.

[11] Katie M.Dean,Laura A.Hatfield,et al.Preliminary Data on a Care Coordination Program for Home Care Recipients[J].The American Geriatrics Society,2016(64):1900-1903.

作者简介：

降 惠(1983-)，女，硕士，副教授.研究领域：医学计算机应用，数据挖掘.

(上接第36页)

中位置索引为倒数第5个，而且是<p>元素的子元素。

(2):nth-of-type()伪类

:nth-of-type()伪类的位置索引的编排规则是，从同类型兄弟元素中的第一个开始建立索引，元素类型由伪类前的元素选择器决定。例如，在图2中，选择器p:nth-of-type(3)，表示选择第3个<p>类型的兄弟元素。

(3):nth-last-of-type()伪类

:nth-last-of-type()伪类的位置索引的编排规则是，从倒数第一个同类型兄弟元素开始建立索引，元素类型由伪类前的元素选择器决定。例如，在图2中，选择器p:nth-last-of-type(3)，表示选择倒数第3个<p>类型的兄弟元素。

4.3.3 不带参数结构伪类

在匹配子元素的结构伪类中，有六个不带参数的结构伪类。本质上，它们是带参数结构伪类取特殊数值时的快捷写法[3]，只用来匹配一个子元素。这种写法更具语义，便于记忆。不带参数结构伪类的功能及与不带参数结构伪类的对应关系见表1。

表1 不带参数结构伪类的功能及与带参数结构伪类的对应关系

Tab.1 Corresponding functional relations between the structural pseudo-classes with and without parameters

不带参数结构伪类	功能	对应的带参数结构伪类
E:first-child	匹配是E元素，而且，是第一个子元素的元素	E:nth-child(1)
E:last-child	匹配是E元素，而且，是最后一个子元素的元素	E:nth-last-child(1)
E:first-of-type	匹配E类型子元素中的第一个	E:nth-of-type(1)
E:last-of-type	匹配E类型子元素中的最后一个	E:nth-last-of-type(1)
E:only-child	匹配是E元素，而且，是唯一子元素的元素	E:nth-child(1):nth-last-child(1)
E:only-of-type	匹配子元素中，是E元素，而且是唯一一个的元素	E:nth-of-type(1):nth-last-of-type(1)

5 结论(Conclusion)

CSS3结构伪类的引入，有效减少了HTML源代码中类名的使用，使源代码更加简洁，便于维护。CSS3设计结构伪类是基于文档树的根节点、空节点、子节点等结构信息。CSS3结构伪类可用来匹配根元素、空元素和子元素，其中匹配子元素的结构伪类采用位置索引来定位元素，可分为带参数和不带参数两类，不带参数结构伪类本质上是带参数结构伪类取特殊值时的语义写法。带参数结构伪类的参数可以是表达式an+b和关键字odd、even，它们都表示位置索引值的集合，通过位置索引值可定位到一个或多个子元素。

参考文献(References)

[1] W3C.Selectors Level 3 W3C Candidate Recommendation 30 January 2018[EB/OL].https://www.w3.org/TR/css3-selectors.

[2] Eric A.Meyer.CSS: The Definitive Guide[M].Published by O'Reilly Media, Inc.,June 2017:1127-11143.

[3] CSS魔法.CSS揭秘[M].北京:人民邮电出版社,2016:178-182.

[4] (英)弗雷恩(Frain,B.).田永强,译.响应式Web设计：HTML5和CSS3实战[M].北京:人民邮电出版社,2013:113-121.

[5] 大漠.图解CSS3核心技术与案例实战[M].北京:机械工业出版社,2014:312-357.

作者简介：

黄志刚(1965-)，男，本科，高级工程师.研究领域：Web技术，大数据技术.

(上接第59页)

教育研究,2017(3):20-26.

作者简介：

沈海波(1963-)，男，博士，教授.研究领域：网络与信息安全，软件工程.

周如旗(1971-)，男，博士，副教授.研究领域：人工智能，软件工程.

朱雄泳(1976-)，男，博士，讲师.研究领域：图形图像处理，软件工程.