

数据挖掘技术在患者病情识别及管理中的研究进展

张燕¹, 彭伶俐², 梁玲玲³, 罗贞¹

Review of data mining technology in identification and management of patients' clinical conditions Zhang Yan, Peng Lingli, Liang Lingling, Luo Zhen

摘要: 数据挖掘通过逻辑回归、贝叶斯、决策树模型、人工神经网络等算法在海量医疗健康数据中挖掘出相关预测值或某病/某种情况的概率,以助医护人员识别潜在危重患者、评估疾病严重程度、预测病死率等,是有效的疾病管控方法。目前国内数据挖掘还处于起步阶段,需在预警模型、护理记录标准化、引进数据集管理人员等方面加强建设,使数据挖掘用于医疗护理工作中,预测病情发展趋势,辅助临床决策,从而维护患者安全。

关键词: 数据挖掘; 大数据; 疾病; 病情变化; 病情识别; 预警模型; 患者安全; 综述文献

中图分类号: R471; TP393 **文献标识码:** A **DOI:** 10. 3870/j. issn. 1001-4152. 2020. 11. 017

病情恶化指患者经历的一种损害血流动力学稳定性的动态变化过程,以生理失代偿为特征,伴或不伴有主观感觉异常或客观参数异常^[1]。临床住院过程中存在着大量的潜在危重患者^[2]。该类患者如果不早期识别及干预,有可能在短期内病情迅速恶化,从而发生严重的不良结局事件^[3],如心脏骤停、非预期死亡等。严重不良事件的发生不仅给患者及家属带来巨大打击,同时也增加医院负担,容易导致医疗纠纷。护士与患者接触频繁,在病情观察与治疗中发挥着重要的作用。如果患者的病情恶化没有被及时识别,将造成不恰当的护理或干预延迟^[4]。

为帮助护士早期识别患者的病情恶化,国外学者构建了多种基于患者生理参数的病情预警系统,如早期预警评分(Early Warning Score, EWS)^[5],改良早期预警评分(Modified Early Warning Score, MEWS)^[6],国家早期预警评分(National Early Warning Score, NEWS)^[7]等。尽管病情预警评分得到广泛的应用,但仍然存在不足,如评估工具包含的生理参数有限,且多是单一时点的评分,缺乏连续性,不能全面反映患者病情变化趋势等。而数据挖掘技术的兴起及其在病情识别方面的应用在很大程度上解决了上述难题。本研究对数据挖掘技术在病情恶化早期识别中的应用进展及问题进行综述,旨在为护理工作提供启示和参考。

1 数据挖掘及常见算法

随着医院信息系统的普及,产生了大量的医疗健康数据。医疗领域的数据挖掘(Data Mining, DM)是从大量、不完全、有噪声的、模糊的、随机的医疗健康数据信息中挖掘潜在的、有效的信息,从中找出有意义的关系、模式和趋势等的过程^[8],涉及统计学、数

学、机器学习方法、人工智能、数据可视化等多种技术。数据挖掘的实现依赖多种数据挖掘方法,应用于病情恶化早期识别中的数据挖掘算法有逻辑回归(Logistic Regression, LR)、贝叶斯(Bayesian Network, BN)、决策树模型(Decision Tree Model, DTM)、随机森林(Random Forest, RF)、人工神经网络(Artificial Neural Network, ANN)和支持向量机(Support Vector Model, SVM)等机器学习方法。

1.1 逻辑回归 逻辑回归是一种研究因变量为二分类(可扩展到多分类)变量与多个影响因素(自变量)之间关系的非线性回归方法,通过筛选有意义的自变量建立模型,根据模型预测在不同的自变量情况下,发生某病或某种情况的概率^[9]。

1.2 贝叶斯 贝叶斯定理最初由英国学者 Thomas Bayes 提出,而后被发展应用为一种统计推断方法,其基本方法是将样本信息与未知参数的先验信息结合,依据贝叶斯定理得出后验信息后进行统计推断^[10]。其计算公式为 $P(A|B) = P(B|A)P(A)/P(B)$, $P(A|B)$ 是在 B 发生的情况下 A 发生的概率。根据先验概率 $P(A)$ 和条件概率 $P(B|A)$ 最终得到后验概率 $P(A|B)$ 。如已知某种疾病的发病率 $P(A)$,有一种检验技术诊断此病的误诊率为 5%,假设一个人的检验结果显示有病,推测这个人确实患病的概率(后验概率)。

1.3 决策树模型 决策树是主要用于解决分类问题的数据挖掘算法,它是一种类似树形结构的流程图,一棵决策树一般包括一个根节点、若干个内部节点和叶节点。其中每个内部节点表示一个属性上的判断,每个分支代表一个判断结果的输出,最后每个叶节点代表一种分类结果^[11]。如根据 Braden 评分(根节点)、翻身计划的落实情况和皮肤损伤情况及有无促进压力性损伤发生的特殊情况(内部节点)筛选难免性压力性损伤的高危患者(叶节点)^[12]。其目的是从中挑选出最有意义的分类变量和分类数,形成简单易读的分类规则。

1.4 随机森林 随机森林是将多个决策树整合成一个分类器的集成算法^[13],不同的决策树构成森林。

作者单位: 1. 中南大学湘雅护理学院(湖南 长沙, 410013); 2. 中南大学湘雅医院骨科; 3. 新疆医科大学第六附属医院骨科

张燕:女,硕士在读,护士

通信作者:彭伶俐, pll98124@126.com

科研项目:新疆维吾尔自治区 2017 年区域协同创新专项(科技援疆计划)(2017E0278)

收稿:2020-01-28;修回:2020-03-03

当进行分类时,新的输入样本进入,就让森林中的每个决策树进行判断和分类,每个决策树会生成一个分类结果,将分类结果进行汇总,哪种分类结果最多便挑选其作为最终的分类结果。

1.5 人工神经网络 人工神经网络是一种应用类似大脑神经元突触联接的结构进行信息处理的数学模型,由大量的节点(或称神经元)之间相互联接构成,包括输入层、隐含层和输出层。输入层接受外部世界的信号与数据;经过隐含层、输出层的计算得到预测值,预测值为输出层的输出结果^[14]。

1.6 支持向量机 支持向量机是 20 世纪 90 年代中期由 Vapnik 等^[15]根据结构风险最小化理论提出的一种二分类模型,其基本思想是利用核函数将输入向量映射到一个高维的特征向量空间,并在该特征空间中构造最优分类面^[16],以正确划分训练数据集区分不同类别的样本。

2 基于数据挖掘算法的病情恶化预警模型的构建及应用

2.1 帮助医护人员识别潜在危重患者,减少严重不良事件发生 由于 ICU 床位的数量限制,部分被转入到普通病房的潜在危重患者容易因监护不到位、临床决策失误等原因而造成严重的不良事件^[17]。据报道,因病情恶化未被识别使患者转入 ICU 延迟、需要接受心肺复苏所致的病死率高达 67%^[18]。早期识别潜在危重患者,对减少严重不良事件、改善患者临床结局意义重大^[19]。Ghosh 等^[20]收集 11 362 例普通病房患者的生命体征,使用朴素贝叶斯算法形成了病情恶化早期预警指标评分(Early Deterioration Indicator, EDI)用于识别死亡或需转入 ICU 或过渡 ICU (Progressive Care Unit, PCU)的患者,结果发现该评分比常用的 MEWS 评分、NEWS 评分的预测准确性更高,且发现病情恶化的时间从恶化前 7 h 提前至恶化前 24 h。另有研究显示,利用多层感知器神经网络对普通病区的患者发生非计划性转入 ICU 进行预测,该模型能将预测窗口提前至转 ICU 前的 16 h^[21]。这类预警模型能较好地预测普通病房患者严重不良结局的发生,从而有助于护理人员对可能发生不良事件的高危人群实施积极干预,降低致残率、病死率等^[22]。

2.2 帮助医护人员寻找病情恶化的早期预警指标 病情恶化的预警指标即预测患者发生病情恶化的预测因子,通常为与病情恶化相关的各种临床变量。数据挖掘技术通过从众多临床变量中遴选有意义的变量建立预警模型,能帮助医护人员寻找病情恶化的危险因素,辅助临床决策。Churpek 等^[23]回顾了 269 999 例住院患者的人口学资料、实验室检查及生命体征数据,比较不同数据挖掘算法建立的预警模型预测心脏骤停、非计划转 ICU 和死亡的准确性,结果发现随机森林模型的预测准确性最高,其中呼吸频率、心率、年龄、收缩压是模型中最重要的预

测变量。Zhai 等^[24]运用逻辑回归方法从 36 个临床变量(包括生命体征、意识水平、疼痛、呼吸做功)中筛选出 29 个变量,建立儿童非计划转 ICU 的预警模型,结果显示该模型的预测效能较现有的儿童早期预警评分^[25]更佳。模型中早期预警指标异常对患者的病情恶化有着指示作用,是护士进行病情观察的重要指标。

2.3 评估疾病严重程度,预测 ICU 住院病死率 ICU 患者多为器官或系统功能衰竭的危重患者,可同时患有多种疾病,并发症的发生率和病死率极高^[26]。患者医疗费用高,护士的工作量大,通过评估患者病情严重程度和死亡风险能避免过度治疗,同时提升护士处理危重患者突发事件的应对能力,提高 ICU 住院患者的总体生存率^[27]。

近年来,学者们相继开发了多种基于数据挖掘技术的 ICU 死亡预警模型。谢俊卿等^[28]基于国外的大型重症医疗数据库 MIMIC-III (Medical Information Mart for Intensive Care),提取第二代简化急性生理评分模型(Simplified Acute Physiology Score, SAPS-II)^[29]中的 17 个变量,应用随机森林算法建立 ICU 患者住院死亡风险预测模型,发现其曲线下面积高达 0.855,其预测的准确性较高,这与 Ng 等^[30]的研究结果一致。潘昌霖等^[31]收集来自心脏、内外科及创伤等 ICU 的 4 000 例患者的生理生化数据,运用贝叶斯算法建立 ICU 患者住院病死率的预测模型,该模型包含体温、呼吸、心率、收缩压及血糖、白蛋白等共 25 个变量,结果显示其能很好地区分出住院死亡的患者。综上,利用数据挖掘算法建立的预警模型可能比常用的病情评分工具预测 ICU 病死率更准确,能帮助护理人员对患者病情作出正确评价,从而加强对危重患者的监测和实施紧急干预,提高患者的生存率。

2.4 帮助急诊医护人员进行分流,保证急诊运营效率 急诊科就诊的患者病情严重程度不一,优先救治危重患者,避免医疗资源浪费,评估患者病情以合理分诊,是急诊科工作的重要环节,也是缓解急诊科拥挤现象,减少漏诊和误诊,避免医疗纠纷的关键。Ong 等^[32]以心脏骤停和死亡为结局事件,收集 925 例急诊科患者的前瞻性队列数据,以患者入院 72 h 内心脏骤停和死亡为预测结局,利用支持向量机算法形成基于心率变异性指标、年龄、生命体征的预测评分系统,以最佳截断值作为临界点,通过计算得分将急诊患者分为低、中、高危组。根据这一风险分层工具,护士能区分不同风险的 cardiac arrest 和死亡患者,从而进行合理分流,在医疗资源有限的条件下指导优先救治危重患者,使医疗资源利用最大化。

3 数据挖掘技术应用于病情恶化早期识别领域的问题及启示

2016 年,国务院印发的《“健康中国 2030”规划纲要》提出要加强医疗大数据应用体系建设,推进医疗健康大数据开放共享、深度挖掘和广泛应用^[33]。但

目前国内医疗数据挖掘还处于起步阶段,且存在较多问题。

3.1 自身条件限制

3.1.1 许多预警模型基于国外人群,外推性较差,未进行临床应用 许多数据挖掘算法建立的预警模型依赖国外人群的数据集,存在人种差异,能否应用于国内的目标人群存在争议;大多数研究是基于回顾性的数据集开展的单中心研究,存在数据缺失、记录错误等问题,模型的外推性较差;同时,关于模型效能验证的前瞻性研究较少,没有进行实际的临床应用。因此医护人员在选择使用此类预警模型时应谨慎,今后有必要开展多中心、前瞻性的研究对模型的效能进行验证。

3.1.2 某些模型中预警指标的权重未知,不能指导提供针对性的护理措施 依靠传统算法如逻辑回归建立的模型能明确知道模型中每个指标的权重,而某些数据挖掘算法如神经网络方法、随机森林算法是一个黑匣子(BlackBox)^[34-35],得出的结论无法进行解释,只能作为软件应用程序来运行以达到预测的目的。由于可解释性差,护理人员无法知晓需要着重干预的影响因素,不能为实施针对性护理措施提供指导。

3.2 客观条件限制

3.2.1 护理记录未标准化导致数据处理困难 护理记录是护理病历的重要组成部分,目前国内各个医疗机构内部及医疗机构之间对于护理记录的书写没有统一标准。护理记录存在数据缺失、表达不清、未包含某些数据如交班记录、压疮的图片等问题^[36],给数据处理造成了很大困难。为此明尼苏达大学护理学院举办了利用大数据和科学改变医疗共识会议,旨在推进电子病历中标准化护理术语的整合^[37]。因此实现护理信息共享,有必要由相关组织牵头开展类似活动,同时借鉴国际上现存的标准化护理术语集来发展符合中国国情的标准化护理术语集,推动护理术语逐步实现标准化。

3.2.2 缺乏管理和开发大型临床数据集的专业人员

数据挖掘算法非常复杂,需要计算机技术的辅助。临床医护人员具备丰富的专业知识和实践经验,但对于数据挖掘方法、计算机技术等并不精通,因此还需培养精通护理信息学方面的专业人才或者与专业人士协作。国外的护理信息学教育起源于 20 世纪 80 年代后期,英美等发达国家陆续开设了护理信息学专业课程^[38]。我国的护理信息学教育才刚起步,今后应在本科、研究生课程中加入护理信息学的课程,培养精通护理与信息学的复合人才,推动临床护理、护理科研、护理教育的发展。

3.2.3 实时处理和分析数据存在缺陷 早期预警需要护理人员动态的、连续的监测,从而实时了解其病情变化趋势,提高预测的准确性^[39]。Hackmann 等^[40]采用电子病历系统和实时传感器系统来识别病情恶化,具体流程包括无线传感器网络将患者实时的生命体征

数据导入到电子病历系统中,不良事件监测装置分析导入的数据并在患者出现病情变化时自动通知医护人员。彭伶俐等^[41]将 MEWS 评分表植入到电子病历系统中,护士输入生命体征数值后系统可自动计算 MEWS 分值,但缺少关于自动监测和通知系统的报道。开发信息系统以实现床旁监测设备数据向电子病历系统的实时传输,将预警模型嵌入电子病历系统评估患者的病情,将是未来研究的重点和难点。

4 小结

病情恶化的早期识别是管理临床危重症患者或潜在危重症患者的有效手段,通过早期的风险感知,帮助医护人员在病情恶化早期采取针对性的治疗护理,从而减少住院严重不良结局事件的发生,维护患者安全,降低医疗成本。

多种数据挖掘技术被应用于病情恶化早期识别的模型构建中,且显示了较高的准确性,但目前缺少对此类模型的效能验证以及与传统病情评估工具的比较分析,今后有必要开展多中心、前瞻性的研究验证模型的有效性。护理记录术语尚未实现标准化,护理信息教育学还处于起步阶段,阻碍了数据挖掘技术在医疗数据中的发展,应在借鉴国外先进经验的基础上建立适合我国国情的标准化护理记录术语集,大力开展护理信息学教育。另外,数据挖掘还有赖医学、统计学、计算机科学等学科的交叉和融合,开发人工智能化系统进行实时数据处理和分析将是未来的研究方向,有助于医护人员更精准地预测病情发展趋势、辅助临床决策,维护患者安全。

参考文献:

- [1] Padilla R M, Mayo A M. Clinical deterioration: a concept analysis[J]. J Clin Nurs, 2018, 27(7-8): 1360-1368.
- [2] 楼滨城. 对急诊医学的再认识[J]. 世界急危重病医学杂志, 2004, 3(1): 165-168.
- [3] Ludikhuize J, Brunsveld-Reinders A H, Dijkgraaf M G, et al. Outcomes associated with the nationwide introduction of rapid response systems in the Netherlands[J]. Crit Care Med, 2015, 43(12): 2544-2551.
- [4] Al-Moteri M, Plummer V, Cooper S, et al. Clinical deterioration of ward patients in the presence of antecedents: a systematic review and narrative synthesis[J]. Aust Crit Care, 2019, 32(5): 411-420.
- [5] Morgam R, Williams F, Wright M. An Early Warning Scoring System for detecting developing critical illness[J]. Clin Hatens Care, 1997, 8(1): 100-101.
- [6] Subbe C P. Validation of a Modified Early Warning Score in medical admissions[J]. QJM, 2001, 94(10): 521-526.
- [7] Royal College of Physicians. National Early Warning Score (NEWS): standardising the assessment of acute-illness severity in the NHS. Report of a working party[R]. London: Royal College of Physicians, 2012.
- [8] 殷蕾, 朱锦士. 试述护理领域中的数据挖掘技术: 2006 年全国医院信息管理学术交流会论文集[C]. 昆明: 中国医院管理杂志社, 2006.
- [9] 孙振球, 徐勇勇. 医学统计学[M]. 北京: 人民卫生出版

- 社,2014:244.
- [10] 韩明. 关于贝叶斯[J]. 中国统计,2014(9):32-33.
 - [11] Alonso S G, de la Torre-Díez I, Hamrioui S, et al. Data mining algorithms and techniques in mental health: a systematic review[J]. J Med Syst,2018,42(9):161.
 - [12] 杨青,王国蓉,江宾,等. 基于决策树的肿瘤患者难免性压疮风险预测模型研究[J]. 护理学杂志,2019,34(13):4-7.
 - [13] 曹正凤. 随机森林算法优化研究[D]. 北京:首都经济贸易大学,2014.
 - [14] Loy C, Rensl S E, Torng W, et al. Machine learning in chemoinformatics and drug discovery[J]. Drug Discov Today,2018,23(8):1538-1546.
 - [15] Vapnik V, Cortes C. Support vector networks[J]. Mach Learn,1995,20(3):273-297.
 - [16] 凤思苑,巩晓文,崔壮,等. 支持向量机和人工神经网络在冠状动脉旁路移植术后晚期静脉移植血管病患风险预测中的应用[J]. 中国卫生统计,2019,36(4):493-496.
 - [17] Hodgetts T J, Kenward G, Vlackonikolis I, et al. Incidence, location and reasons for avoidable in-hospital cardiac arrest in a district general hospital[J]. Resuscitation,2002,54(2):115-123.
 - [18] Nadkarni V M, Larkin G L, Peberdy M A, et al. First documented rhythm and clinical outcome from in-hospital cardiac arrest among children and adults[J]. JAMA,2006,295(1):50-57.
 - [19] Brilli R J, Gibson R, Luria J W, et al. Implementation of a medical emergency team in a large pediatric teaching hospital prevents respiratory and cardiopulmonary arrests outside the intensive care unit[J]. Pediatr Crit Care Med,2007,8(3):236-247.
 - [20] Ghosh E, Eshelman L, Yang L, et al. Early Deterioration Indicator: data-driven approach to detecting deterioration in general ward[J]. Resuscitation,2018,122:99-105.
 - [21] Wellner B, Grand J, Canzone E, et al. Predicting unplanned transfers to the intensive care unit: a machine learning approach leveraging diverse clinical elements[J]. JMIR Med Inform,2017,5(4):e45.
 - [22] Huang Z, Dong W, Duan H, et al. A regularized deep learning approach for clinical risk prediction of acute coronary syndrome using electronic health records[J]. IEEE Trans Biomed Eng,2018,65(5):956-968.
 - [23] Churpek M M, Yuen T C, Winslow C, et al. Multi-center comparison of machine learning methods and conventional regression for predicting clinical deterioration on the wards[J]. Crit Care Med,2016,44(2):368-374.
 - [24] Zhai H, Brady P, Li Q, et al. Developing and evaluating a machine learning based algorithm to predict the need of pediatric intensive care unit transfer for newly hospitalized children[J]. Resuscitation,2014,85(8):1065-1071.
 - [25] Parshuram C S, Hutchison J, Middaugh K. Development and initial validation of the Bedside Paediatric Early Warning System score[J]. Crit Care,2009,13(4):135.
 - [26] 王宇,马朋林. ICU 的形成与发展[J]. 医学与哲学(临床决策论坛版),2006,27(2):7-9.
 - [27] 阮一鸣,吴骋,贺佳. 患者病情分级系统在重症监护室护理工作中应用的现况与进展[J]. 解放军护理杂志,2019,36(9):42-44,74.
 - [28] 谢俊卿,蔺轲,李春晓,等. 随机森林模型在 ICU 患者住院死亡风险预测中的应用[J]. 中国数字医学,2017,12(11):86-89.
 - [29] Le Gall R, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score(SAPS II) based on a European/North American multicenter study[J]. JAMA,1993,270(24):2957-2963.
 - [30] Ng K, Steinhubl S R, de Filippi C, et al. Early detection of heart failure using electronic health records: practical implications for time before diagnosis, data diversity, data quantity, and data density[J]. Circ Cardiovasc Qual Outcomes,2016,9(6):649-658.
 - [31] 潘昌霖,何史林,应俊,等. 基于贝叶斯方法的 ICU 患者死亡概率预测研究[J]. 中国数字医学,2012,7(10):17-20.
 - [32] Ong M E, Lee Ng C H, Goh K, et al. Prediction of cardiac arrest in critically ill patients presenting to the emergency department using a machine learning score incorporating heart rate variability compared with the modified early warning score[J]. Crit Care,2012,16(3):R108.
 - [33] 中华人民共和国中央人民政府. 中共中央国务院印发《“健康中国 2030”规划纲要》[EB/OL]. (2016-10-25) [2020-01-27]. http://www.gov.cn/xinwen/2016-10/25/content_5124174.htm.
 - [34] Lamping F, Jack T, Nicole Rübsamen, et al. Development and validation of a diagnostic model for early differentiation of sepsis and non-infectious SIRS in critically ill children—a data-driven approach using machine-learning algorithms[J]. BMC Pediatr,2018,18:112.
 - [35] 王国强,阚红星,王宗殿. 综合多种数据挖掘技术的糖尿病诊断系统[J]. 电脑知识与技术,2012,8(23):5547-5551.
 - [36] 王振容,蒋晓莲. 数据挖掘在护理领域中的应用进展[J]. 中华护理杂志,2017,52(10):1262-1264.
 - [37] Westra B L, Latimer G E, Matney S A, et al. A national action plan for sharable and comparable nursing data to support practice and translational research for transforming health care[J]. J Am Med Inform Assoc,2015,22(3):600-607.
 - [38] 许燕. 国内外护理信息化实践现状[J]. 中国护理管理,2010,10(5):11-14.
 - [39] Van Loon K, Guiza F, Meyfroidt G, et al. Dynamic data analysis and data mining for prediction of clinical stability[J]. Stud Health Technol Inform,2009,150:590-594.
 - [40] Hackmann G, Chen M, Chipara O, et al. Toward a two-tier clinical warning system for hospitalized patients[J]. AMIA Annu Symp Proc,2011:511-519.
 - [41] 彭伶俐,李映兰,贺爱兰,等. 建立病情早期预警工作流程保障患者安全[J]. 中国护理管理,2014,14(12):1250-1254.