

电子病历辅助临床诊断系统研究

王庆阳

(清华大学深圳研究生院生物
医学工程研究所 深圳 518055)

张书旭

(广州医科大学附属肿瘤
医院放疗中心 广州 510095)

白彦灵 王业伟

(哈尔滨医科大学附属肿瘤
医院放疗中心 哈尔滨 150081)

袁克虹

(清华大学深圳研究生院生物
医学工程研究所 深圳 518055)

[摘要] 为共享医生诊疗经验,从相似电子病历检索的角度探讨辅助诊断系统的建设,介绍系统功能模块组成和理论算法,包括主诉相似度、诊疗模式相似度、图像相似度、综合相似度计算算法。此方法能够较为全面地根据电子病历的内容计算相似度,检索出的电子病历为医生制订诊疗计划提供参考。

[关键词] 电子病历;辅助诊断;相似度计算;深度学习

[中图分类号] R-056 **[文献标识码]** A **[DOI]** 10.3969/j.issn.1673-6036.2016.07.001

Research on the Auxiliary Clinical Diagnosis System Based on Electronic Medical Records WANG Qing-yang, Research Center of Biomedical Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China; ZHANG Shu-xu, Center of Radiotherapy, Affiliated Tumor Hospital of Guangzhou Medical University, Guangzhou 510095, China; BAI Yan-ling, WANG Ye-wei, Department of Radiation Oncology, The Cancer Hospital of Harbin Medical University, Harbin 150081, China; YUAN Ke-hong, Research Center of Biomedical Engineering, Graduate School at Shenzhen, Tsinghua University, Shenzhen 518055, China

[Abstract] In order to share doctors' experience in diagnosis and treatment, the paper studies the construction of the auxiliary diagnosis system from the perspective of similar Electronic Medical Records (EMR), and introduces the system's functional modules and theoretic algorithms, including computational algorithms for the similarity of complaints, similarity of diagnosis and treatment modes, image similarity and comprehensive similarity. This method can compute the similarities comprehensively according to contents of EMR. The EMR retrieved can serve as reference for doctors to make a plan for diagnosis and treatment.

[Keywords] Electronic Medical Records (EMR); Auxiliary diagnosis; Similarity computation; Deep learning

1 引言

[收稿日期] 2016-03-28

[作者简介] 王庆阳,硕士,发表论文2篇;通讯作者:袁克虹,博士,副教授,发表论文30余篇。

电子病历(Electronic Medical Records, EMR)是医生诊疗经验和模式的载体,它的核心价值体现

在辅助诊断方面, 为医生提供决策支持^[1]。目前国内外有很多学者、机构以及企业致力于基于 EMR 的辅助诊断系统的研究, 其领域可以涉及完整的医疗过程, 在优化工作流程、提高工作效率、降低医疗差错、提高医疗质量等方面具有重要作用^[2]。

国内基于 EMR 的应用研究一方面是针对 EMR 系统的研发^[3], 另一方面是基于 EMR 的临床路径优化和相似 EMR 搜索^[4-7]。有关文献研究分别如下: 文献^[4-5]分别利用语义分析技术和模糊聚类分析算法从临床路径优化的角度构建辅助诊断系统。文献^[6]的研究不是直接提供辅助诊疗建议, 而是通过检索相似 EMR 从而为医生提供诊疗参考。文献^[7]较为详细提出一种辅助诊断系统的理论框架, 其核心在于数据库的构建。国外的科研机构和企业在这一领域的研究的发展较快。文献^[8]使用时间序列分析算法和语义知识库分析 EMR 中的显型因素和隐型因素, 以达到提高治疗水平的目的; 除了用于治疗, 文献^[9]验证了通过使用基于 EMR 的辅助诊断系统, 能够提高老年人的用药安全系数。另外, IBM 公司研发的 Watson 计算机能够根据患者 EMR、医学知识等内容利用大数据技术为肺癌患者提供治疗方案^[10]。EMR 辅助诊断系统不仅有利于个人的治疗, 还有助于公共卫生管理, 辅助地方公共卫生管理政策的制定^[11]。未来辅助诊断系统的发展不会仅仅局限于电子病历, 文献^[12-14]提出并实验了将 EMR 数据与基因组数据相结合实现精准医疗的可行性和研究方法。

EMR 是医生经验的载体, 如何从中挖掘完整的医生经验是辅助诊断的基础。病历中的数据可以大致分为文本型、数字型和图像型 3 类, 单独研究某一类数据都无法完整地继承医生经验。因此一个完善的辅助诊断系统必然要将这 3 种数据作为研究对象。基于上述考虑, 本文提出了一种基于 EMR 相似度计算的辅助诊断系统的构建方法, 依靠检索出的相似病历辅助医生制定诊疗计划。

2 系统概述

本系统是基于 Matlab 平台开发, 其目的是通过检索相似电子病历从而辅助医生诊断, 系统的框架, 见图 1。其中数据库模块的构建在本课题组已

有的成果中已被详细描述^[7], 本文的主要内容是从如何计算电子病历相似度的角度描述系统的功能。本系统处理的数据对象是的患者主诉数据 (文本型)、化验结果 (数字型) 和医学图像 (图像型), 首先系统分别计算出 3 类数据的相似度, 然后依据层次分析法计算出 3 种数据的权重系数, 通过加权计算即可获得 EMR 的综合相似度从而检索出相似度高的 EMR 辅助医生制订诊疗计划。本系统与其他辅助诊断系统的不同之处是综合了电子病历中的文本型数据、数值型数据和图像数据来计算 EMR 相似度, 通过返回相似度高的 EMR 为医生提供参考。

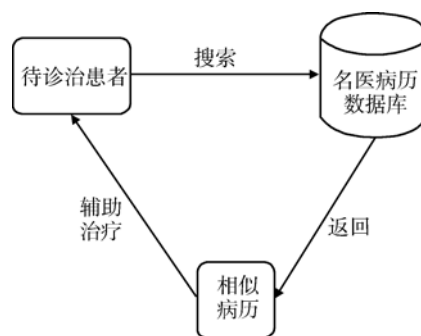


图1 系统框架

3 系统组成

3.1 概述

除数据库模块, 本系统还包括另外两个部分: 交互界面和后台功能模块。系统模块构成和工作流程, 见图 2。

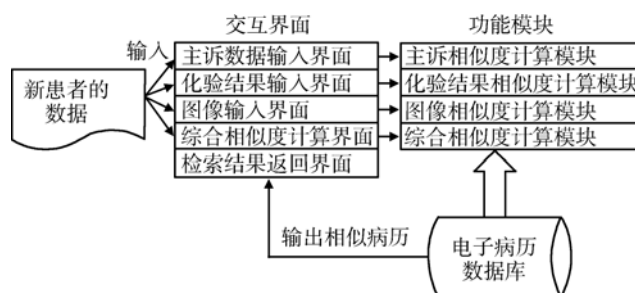


图2 系统模块构成和工作流程

3.2 交互界面

交互界面的设计用于数据的输入和返回结果的

查询。3 种数据：患者主诉、化验结果和医学图像，分别对应 3 个数据输入界面。图像的输入较容易理解，而另外两种输入界面的设计要依据便于处理的原则。患者主诉是文本数据，主要描述患者的症状，其表达形式以词语或者短语为主。因此本系统中患者主诉的输入方式为各种症状的独立勾选形式。对于化验结果，其本身就是结构化、数字化的表达，因此输入方式与患者主诉相同。

3.3 后台功能模块

本系统的后台功能模块由 4 部分组成，其相互关系，见图 3：患者主诉相似度计算模块、化验结果相似度计算模块、图像相似度计算模块和综合相似度计算模块。前 3 者相互独立，分别计算出相应的相似度数值后，经由综合相似度计算模块转化成电子病历的综合相似度。

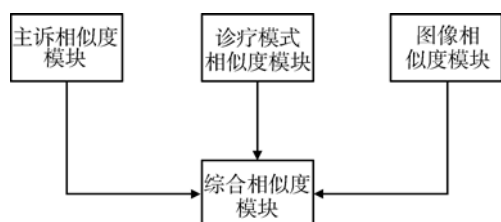


图 3 功能模块

4 关键算法

4.1 主诉相似度

主诉相似度计算的关键环节在于语义分析技术。患者主诉中的很多症状之间不是相互独立的，症状之间存在某种医学联系，而这种医学联系可以通过语义分析技术进行机器表达。由于患者主诉数据的表达形式和内容相对简单，因此本文使用两种语义关系表达症状之间语义联系：从属（A Kind of, AKO）关系和条件关系。另外，症状中存在特异性症状，其定义是患者具有某种症状时，临床上可以据此判断此患者患有某种特定疾病，那么此症状就成为特异性症状。根据特异性症状可以筛选检索范围。主诉相似度计算流程，见图 4。可以看出症状相似度计算规则是两个患者同时表现某一症状，则此症状的相似度为 1，反之为 0。之后利用统

计学方法计算各种症状在数据库中出现的频率作为相应的权重系数，以及相似度为 1 的症状在两个患者的症状总量之中比例系数，然后经加权运算求出患者主诉的相似度。

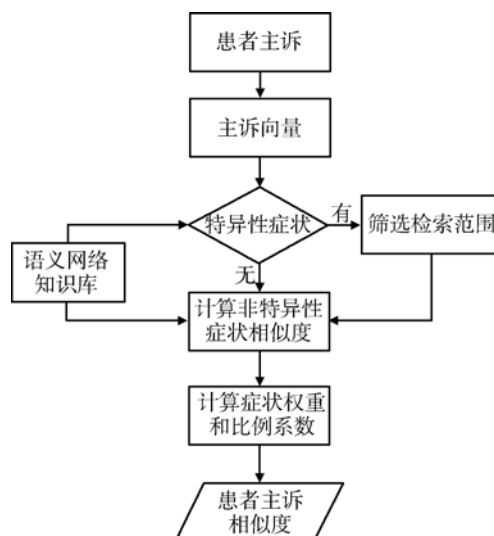


图 4 主诉相似度计算流程

4.2 诊疗模式相似度

所谓医生诊疗模式，即医生分析各种化验指标的严重程度以及各种化验之间的重要性，综合得出诊断结论的思维过程。医生的诊疗模式中凝聚着医生的行医经验，通过衡量医生诊疗模式的相似度，可以将名医的经验共享给其他水平一般的医生。化验指标具有层次性，即某一个化验项目会产生多个子指标的化验结果，如血常规项目中含有多项化验指标，因此在计算相似度时需要将此情况考虑在内。本文在研究医生诊疗模式的相似度计算方法时，以肺癌患者的化验结果为研究数据。算法流程，见图 5。可以看出诊疗模式相似度算法包含两个环节：单独的化验指标相似度计算和化验指标的权重计算。对于前者，首先将各种化验指标离散化，分别对应指标的正常、过低、过高或者隐性、阳性；对于表现相同的指标，赋予其相似度为 1，反之为 0；对于权重计算，本文根据双层次分析法，首先将医生对于每一个化验指标的重要性打分转化成每一项指标的权重，以此计算一个化验项目的相似度；同样，再次使用层次分析法求出每个化验

项目的权重,从而计算出医生诊疗模式的相似度。

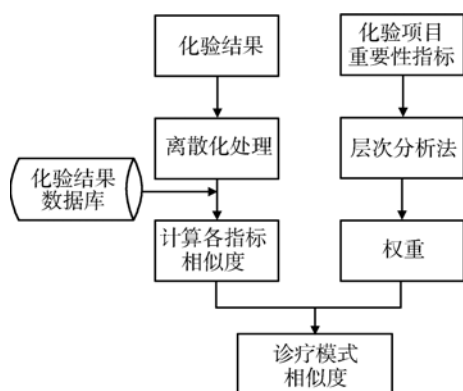


图5 诊疗模式相似度算法流程

4.3 图像相似度

图像相似度计算模块和传统的基于内容的医学图像检索一样,本文以来自于相近解剖层的肺癌患者 CT 数据为研究对象,采取两类相似度计算方法求取相似度:基于检索分类的方法和基于 ROI 特征的方法。另外图像预处理是必不可少的环节,本文首先去除胸腔壁以外以及大部分胸腔壁区域,并且对图像做尺寸和灰度归一化处理^[15]。使用了两种基于检索分类的方法,分别是 SIFT + Hash 算法和卷积神经网络算法。前者的处理流程是首先使用尺度不变特征转换 (Scale - invariant Feature Transform, SIFT) 算法^[16]提取图像 SIFT 描述子,然后对描述子进行 Hash 编码^[17],通过计算 Hash 编码间的汉明距离检索相似的肺癌 CT 图像,统计返回图像中正常和异常图像的比例用以确定被检索图像的分类,从而实现分类的目的。卷积神经网络 (Convolutional Neural Network, CNN)^[18]算法模拟人类视觉处理过程,在计算机视觉领域应用广泛。用于 CNN 处理的异常样本是从肺癌 CT 中截取的肿瘤图像,正常样本是从肺实质区域中随机截取的血管、支气管和其他肺实质图像。CNN 算法的核心在于建立一个多层卷积神经网络,能够直接以原始数据未处理对象并且实现分类的目的。基于分类算法的相似度计算策略是类别相同的 CT 图像间的相似度为 1,而类别不同的图像间相似度为 0。本文使用的第 3 种方法是基于 GMM 的 ROI 特征方法:使用 GMM 定位出异

常 CT 图像中的肿瘤区域,然后提取 ROI 区域的图形特征向量^[19],通过计算特征向量间的欧氏距离求出图像相似度。

4.4 综合相似度

此模块的算法基础是层次分析法,其计算流程,见图 6。如图 6 所示,综合相似度计算也包括两个环节:相似度归一化和计算权重。为了消除 3 种相似度数值差异的影响,首先将患者主诉、医生诊疗模式和 CT 图像的相似度在各自的范围内进行归一化处理,获得相同尺度的相似度;然后利用层次分析法将医生给出的 3 种相似度的重要性转化至各自的权重;通过加权运算即可获得最终的电子病历综合相似度。依据综合相似度可以检索相似度高的电子病历为医生提供参考。

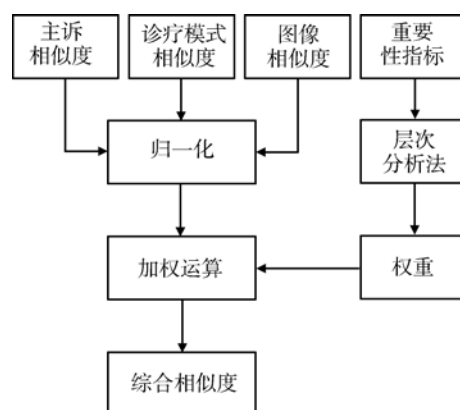


图6 综合相似度计算流程

5 结语

目前本系统尚处于实验阶段,对其性能的临床适用性还需进一步的验证。本系统从 3 种类型的数据计算电子病历的综合相似度,尽可能涵盖了目前电子病历中所包含的主要内容,从而更为全面地继承了医生的诊疗经验和诊疗模式。此系统对于 3 种数据的处理是并行操作的,既可以利用某一类数据进行相似病历检索也能够整合两种或者 3 种数据共同检索电子病历,因此本系统同样也适用于单一类型数据的辅助诊断。本系统能够通过计算电子病历内容的客观相似度,检索出具有可参考价值的电子病历用以辅助医生临床诊断,它的优点在于避免了

人为因素的不确定性影响,同时完整保留了电子病历中的信息供医生参考。未来的工作是借助于大量的电子病历数据验证本系统的性能,同时扩大数据覆盖面和改进相似度计算算法以提高其性能。

参考文献

- 1 李敬华,崔蒙,卜宪峥. 未来电子病历的发展与技术探讨 [J]. 中国数字医学, 2012, 7 (6): 8-10.
- 2 马锡坤,杨国斌,于京杰. 国内电子病历发展与应用现状分析 [J]. 计算机应用与软件, 2015, 32 (1): 10-12.
- 3 张小光. 支持医疗过程管理的电子病历系统研发及有意义应用研究 [D]. 杭州: 浙江大学, 2012.
- 4 王华琼. 基于语义技术的个性化临床路径研究 [D]. 杭州: 浙江大学, 2015.
- 5 张焕君,杨小宁. 基于模糊聚类分析的临床路径决策研究 [J]. 控制工程, 2013, (6): 31.
- 6 赵洋,李万龙,白杰英. 基于本体的电子病历检索系统研究 [J]. 计算机技术与发展, 2010, 20 (3): 211-213.
- 7 李怡雪,韩薇,王庆阳,等. 医院信息化: 临床决策支持系统构建 [J]. 医学信息学杂志, 2015, 36 (11): 6-10.
- 8 Post A R, Kure T, Cholleti S, et al. The Analytic Information Warehouse (AIW): a platform for analytics using electronic health record data [J]. Journal of Biomedical Informatics, 2013, 46 (3): 410-424.
- 9 Alagiakrishnan K, Wilson P, Sadowski C A, et al. Physicians' Use of Computerized Clinical Decision Supports to Improve Medication Management in the Elderly - the Seniors Medication Alert and Review Technology intervention [J]. Clinical Interventions in Aging, 2016, (11): 73.
- 10 IBM. IBM Watson for Oncology [EB/OL]. [2015-8-15]. <http://www.ibm.com/smarterplanet/us/en/ibmwatson/watson-oncology.html>.
- 11 Paneth-Pollak R, Schillinger J A, Borrelli J M, et al. Using STD Electronic Medical Record Data to Drive Public Health Program Decisions in New York City [J]. American Journal of Public Health, 2010, 100 (4): 586-590.
- 12 Wei W Q, Denny J C. Extracting Research - Quality Phenotypes from Electronic Health Records to Support Precision Medicine [J]. Genome Med, 2015, 7 (1): 41.
- 13 Tate A R, Beloff N, Al-Radwan B, et al. Exploiting the Potential of Large Databases of Electronic Health Records for Research Using Rapid Search Algorithms and an Intuitive Query Interface [J]. Journal of the American Medical Informatics Association, 2014, 21 (2): 292-298.
- 14 Newton K M, Peissig P L, Kho A N, et al. Validation of Electronic Medical Record - based Phenotyping Algorithms: results and lessons learned from the eMERGE network [J]. Journal of the American Medical Informatics Association, 2013, 20 (1): 147-154.
- 15 Yuan Kehong, Tian Zhen, Zou Jiying, et al. Brain CT Image Database for Medical Diagnosis Using Content - based Image Retrieval [J]. Information Processing & Management, 2011, 47 (2): 176-185.
- 16 Lowe D G. Object Recognition from Local Scale - invariant Features [C]. The Proceedings of the Seventh IEEE International Conference, 1999: 1150-1157.
- 17 Zhang X, Liu W, Dundar M, et al. Towards Large - scale Histopathological Image Analysis: Hashing - based image retrieval [J]. IEEE Transactions on Medical Imaging, 2015, 34 (2): 496-506.
- 18 LeCun Y, Bottou L, Bengio Y, et al. Gradient - based Learning Applied to Document Recognition [J]. Proceedings of the IEEE, 1998, 86 (11): 2278-2324.
- 19 Zhen Tian Z Y, Yuan K. Automatic Detection of Abnormal Regions Using Gaussian Mixture Model [J]. ICIC Express Letters, 2009, 3 (4): 921-926.