

电子病历自然语言处理测评发展

简 哲 李 燕

(哈尔滨医科大学医学信息学系 大庆 163100)

〔摘要〕 分析自然语言处理在医学领域应用存在障碍的原因, 提出电子病历自然语言处理测评的方法, 介绍历年来有关电子病历自然语言处理测评内容及其发展情况, 包括文本检索会议、医学自然语言处理测评、SHARe/CLEF 测评、I2B2 测评等。

〔关键词〕 电子病历; 自然语言处理; 测评

〔中图分类号〕 R-056 〔文献标识码〕 A 〔DOI〕 10.3969/j.issn.1673-6036.2016.12.002

The Development of EMR Natural Language Processing Evaluation JIAN Zhe, LI Yan, Department of Medical Informatics, Harbin Medical University, Daqing 163100, China

〔Abstract〕 The paper analyzes the reasons for barrier of the application of natural language processing in the medical field, proposes the methods for evaluation on Electronic Medical Records (EMR) natural language processing, and introduces the contents and development situations relevant to the evaluation on EMR natural language processing over the years, including text retrieval conference, medical natural language processing evaluation, SHARe/CLEF evaluation, I2B2 evaluation, etc.

〔Keywords〕 Electronic Medical Records (EMR); Natural language processing; Evaluation

1 引言

1.1 电子病历自然语言处理

随着医学信息学的发展, 电子病历在医院中的应用逐渐普及。电子病历中含有大量的诊疗数据和患者的健康信息, 对其进行充分分析和挖掘, 可辅助临床决策和疾病研究, 已成为临床决策支持系统和构建用户健康数据库的基础。但当前的电子病历系统中大量的医疗记录以自由文本形式存储, 这种形式在进一步的数据处理中会出现歧义和非标准化描述, 因此, 需要使用自然语言处理技术将非结构化的医疗文本转化为含有重要医疗信息的结构化文

本。同时, 电子病历通常涉及患者的隐私信息, 获取到相关的临床数据比较困难, 而且大量专业医学术语使得标注语料成本太高, 这些都阻碍了临床数据进一步开发和利用。发起和组织电子病历自然语言处理测评促进了开放语料库的建立, 提供对临床数据标准化的评估方法, 同时吸引全世界不同领域的工作者参与到临床数据自然语言处理中, 促进了医学大数据的进一步发展。

1.2 医学领域自然语言处理落后的原因

自然语言处理已经广泛应用于网络信息检索等领域中。针对医学领域的自然语言处理研究开始于 20 世纪 60 年代, 但其研究始终落后于其他领域^[1]。究其原因有以下几点: (1) 缺少共享数据。出于担心泄露患者隐私和不良医疗记录, 医疗机构往往不愿意向外界公开有关的医疗记录, 同时缺乏有效的

〔修回日期〕 2016-07-19

〔作者简介〕 简哲, 本科生, 发表论文 2 篇。

去隐私工具也阻碍了开发语料库的建立。(2) 标记语料成本过高。由于医疗记录涉及众多的专业的医学术语, 对医疗记录语料的标注需要熟悉诊疗过程的医学专业人员的指导, 使得标注语料库的任务繁重且周期较长。传统的自然语言处理使用的文本处理方法在处理医疗数据时效果不佳, 研究人员需要重新对医疗记录的数据进行预处理才能进一步使用自然语言处理工具。(3) 忽略医疗机构的实际需求。尽管电子病历自然语言处理领域的研究逐渐增多, 但自然语言处理的工具却没有考虑到一线医务人员实际需求, 很少在临床实践中应用。研发和应用中出现脱节, 在研究阶段一个好的自然语言处理工具在应用中却出现许多问题, 同时在现有的医院信息系统中利用自然语言处理的方法也是实际应用中的一大障碍。

1.3 电子病历自然语言处理测评

电子病历自然语言处理测评在一定程度上解决了上述问题。共享测评能够最大程度地标准化语料库, 确保不同年份的数据有相同的语料标注和相同的输入输出参数, 能使用一致的方法进行评估, 减少评估误差。开放的语料库共享测评能解决不可再现性和数据过度拟合问题, 研究人员能够用已标注语料重复训练系统来提高系统的可靠性和再现性。测评通常是针对某一个特定疾病或者诊疗过程, 通过测评能使医疗机构认识到自然语言处理在电子病历数据处理中的作用, 促进研发出的技术真正应用到临床实践中去。各个国家的不同组织都能参与到测评中, 注重了不同学科背景研究人员之间的合作, 为医学信息学的发展注入了生机和活力。

2 电子病历自然语言处理测评的发展

2.1 文本检索会议

2.1.1 目标 第 1 个有关临床数据自然语言的共享测评是 2000 年举办的文本检索会议 (Text Retrieval Conference, TREC) 的文本过滤项目^[2]。由于医学文献的数据量呈爆炸式增长, 如何即时、有效地检索出给定的疾病或症状描述的相关文献成为

本次测评的关注焦点。该项目目标是实现一个即时的信息过滤系统, 在用户需求固定的情况下, 能够过滤动态变化文本流中符合用户需要的信息, 实时地向用户推荐相关文档, 根据用户的反馈信息进一步自学习用户的需要来提高系统的性能。TREC 在 2011 年和 2012 年^[3]关注了基于语义的非结构化文本电子病历检索技术的研究, 目的是开发利用电子病历中的患者和诊疗数据。由于电子病历中有大量的由医护人员书写的非结构化自然文本, 而这些文本内容无法有效准确地被搜索和规范化, 本次测评的目标是将去隐私化的临床报告进行结构化和规范化处理以供高效检索。

2.1.2 效果 作为文本检索领域影响力最大的测评, TREC 对临床数据关注的侧重点主要在对疾病和症状描述的高效检索和规范化上。在大数据的浪潮下, 开发高效、便捷的检索系统是提高临床数据应用和知识发现效率必不可少的环节。TREC 测评所应用的语料集达到了 35 万篇医学文本, 很好地模拟了在海量文本中检索时可能遇到的问题。TREC 测评项目通常根据自然语言处理在各个场景的应用进行更新, 包括微博检索、Web 检索、法律检索等方面, 将电子病历作为测评项目也反映了自然语言处理技术在临床数据处理方面的巨大潜力。

2.2 医学自然语言处理测评

由美国技术医学中心组织的医学自然语言处理测评分别在 2007 年和 2011 年举办。在 2007 年^[4]该项测评关注了放射性报告的自动编码。2011 年的测评^[5]目标是对遗书中的情绪进行分类。研究者发现 10% ~ 43% 的自杀者会留下遗书, 遗书中往往有愤怒、自责、无助和恐惧等情绪和自杀的动机。至少 15% 的首次自杀未遂者往往会尝试第 2 次自杀, 但由于患者的反抗情绪、记忆障碍和矛盾情绪等因素, 如何确定患者再次自杀的可能性是精神病治疗的难点。这两届测评进一步拓宽了自然语言处理在医疗领域的应用范围, 将焦点从疾病命名实体识别转移到了放射性报告和自杀者遗书上来。其中 2011 年对自杀者遗书中的情绪分类测评引起了广泛的关注, 在临床实践中棘手的自杀未遂者的治疗和心理

辅导问题在自然语言处理技术的帮助下开拓了新的研究思路。从自杀者的遗书中识别出情绪因素可以帮助医护人员有针对性地对可疑自杀者进行预防和治疗,也可以让研究者更好地从个人、家庭和社会的角度重新认识自杀者的过激情绪。借助自然语言处理的方法可以突破诊治中瓶颈,更好地为临床实践服务。

2.3 SHARe/CLEF 测评

不同于以往测评更多关注医疗机构的需要,2013 年举办的 SHARe/CLEF 测评^[6]将重点放在如何更好地服务患者上。随着个人健康记录的普及,患者越来越多地会接触到自己的出院小结、临床报告等文本,但是由于这些文本中包含大量的医学术语、缩略词、书写错误等,患者在阅读和理解时会出现很多困难。本届测评的任务就是利用自然语言处理技术来识别和纠正在电子病历中书写错误、不规范的缩写以及缩略词以及其他阻碍患者理解的模糊信息。可以看到,SHARe/CLEF 测评首次关注患者在使用电子病历中可能出现的困境,强调患者的用户体验。在搜索引擎广泛应用的背景下,越来越多的患者和家属倾向于使用搜索引擎来检索疾病和症状,此次测评试图为患者在文本中的医学术语上建立词条链接,方便患者在互联网上获得更多关于疾病和症状的信息,也防止电子病历中出现的错误所造成的患者理解障碍。这次测评反映了患者在使用电子病历中出现的需求,如何更好地利用电子病历帮助患者理解病情和提高电子病历的可读性引起了研究者和电子病历制造厂商的关注。

2.4 I2B2 测评

2.4.1 发展历程 从 2006 年 I2B2 开始组织针对电子病历中临床数据的自然语言处理的评测,主要围绕利用自然语言处理技术对电子病历进行实体识别和关系抽取,推动电子病历的临床数据研究。到 2014 年 I2B2 一共组织了 6 次评测,历届 I2B2 测评中用到的语料库可以在其网站上下载,供研究和商业用途使用。(1) I2B2 于 2006 年电子病历测评^[7]的两项任务分别是去隐私信息和识别患者吸烟状

态,去隐私信息任务要求将涉及患者的隐私信息进行替换,而且要求尽最大程度地保护电子病历的完整性以供进一步研究使用。在去隐私信息过程中要解决歧义性和外来词两个问题,歧义性是指隐私信息和非隐私信息出现重叠,外来词问题是指由于隐私信息在字典中无法识别而错把这些信息当成非隐私信息。在识别患者吸烟状态任务中,要求根据患者临床记录对患者的吸烟状态进行分类,来建立疾病和吸烟状态的联系。(2) 2008 年的评测任务^[8]是利用已经去隐私化出院总结识别患者的肥胖及综合征的状态^[3]。2009 年的测评任务^[9]关注了药品信息的识别,主要包括药名、剂量、给药途径、给药频率、持续时间和给药原因等。本次测评要求参赛者抽取的信息按照上述标准将药品信息进行分门别类,不同类别之间不能存在重复信息。(3) 2010 年的测评任务^[10]关注从电子病历中抽取医学概念,对抽取出的医学概念进行分类和识别医疗问题、检查和治疗之间的关系 3 大临床路径中的核心问题。本次测评应该引起特别重视,将最重要的临床治疗信息作为关注焦点,目标是解决识别概念、修饰和概念之间关系,形成了完整的病人治疗的临床路径,从而为以后疾病谱和专家咨询系统的建立提供了所患疾病、检查和治疗信息。测评结果表明,开发的系统能有效地对医学概念进行分类,而抽取医学概念和识别关系上表现平平。究其原因,难以确定概念的边界成为抽取医学概念的障碍,而识别医疗问题、检查和治疗的关系需要的不仅仅是对文本的语义分析,更需要建立在对上下文和领域专业知识的理解上。(4) 2011 年的测评任务^[11]是共同引用消解,目的是识别出电子病历中等价的概念信息。(5) 2012 年的测评任务^[12]是识别电子病历中的时间信息,目的是要识别临床事件(检查、治疗、出入院等)发生的时间,建立时间轴和事件和时间之间的关系,用时间信息来分析患者的健康状况、病因和治疗情况、治疗效果和预后等。这次测评反映出电子病历实际使用中出现问题,相关概念的共同引用情况阻碍了关键医疗事件的识别,而不规范的时间书写形式使得形成时间轴有一定困难。(6) 2014 年的 I2B2 测评^[13]包括两部分,一是去病程记

录纵向中的隐私信息和识别糖尿病患者电子病历中的冠心病危险因素。本次测评关注去隐私信息化的一个难点,即电子病历中独立的信息可能不会泄露患者的隐私信息,但该患者多项信息结合之后却有泄露隐私的风险。第二部分要求识别出每一个患者冠心病危险因素指标是出现在该病历记录之前、治疗期间还是结束治疗之后,危险因素包括吸烟史、家族史、高血压、高血脂等及这些因素的危险程度。

2.4.2 经验与展望 历届的 I2B2 测评可以看出电子病历领域自然语言处理的发展历程。从最基本的对电子病历进行去隐私化处理逐步深化到对患者疾病因素的识别、药品信息、疾病与检查治疗的关系识别、时间识别等临床实践的各个方面。笔者总结了 I2B2 发展历程中可供参考的一些经验和展望。

(1) 对去隐私性的关注。如何识别并合理替代隐私信息是首先摆在研究者面前的一大问题。I2B2 测评于 2006 年和 2014 年两届都关注了在电子病历中去隐私信息,如今在测评中发展出的去隐私化方法已经能够应用到电子病历语料库建立中。在海量临床数据的今天,利用有效的去隐私化工具能够帮助研究人员获得更多的电子病历,减少在利用过程中存在的隐私泄露及其潜在的伦理学问题,真正实现健康医疗大数据。(2) 对特定疾病危险因素的抽取。I2B2 测评多次关注了特定疾病患者的电子病历并从中识别其疾病发生因素和治疗对策,包括 2006 年的吸烟状态识别、2008 年的肥胖状态识别和 2014 年的糖尿病患者冠心病危险因素识别。电子病历中包含有病人的基本信息、家族史和既往疾病史等信息,在大量电子病历文本中建立疾病和致病因素之间的联系能有效识别疾病的关键危险因素,可以展望着通过自然语言处理更多疾病的危险因素能够识别出来,从而促进临床医学和公共卫生领域的研究。(3) 语料库建立的必要。I2B2 测评中建立的一系列开发语料库为今后的自然语言处理提供了宝贵的语料资源,也保证开发出的自然语言处理技术有可靠性和可再现性。建立开放的语料库丰富电子病历的语料资源,历届 I2B2 测评中用到的语料库可以下载,可供研究和商业用途使用。但现在仍然没有成型的中文电子病历语料库, I2B2 测评中语料库建立

的方法和语料标注的经验值得借鉴。

3 结语

随着医学信息化不断发展,电子病历在医院中逐渐普及,积累的海量数据如何应用到临床实践中成为研究的热点问题。以非结构化的自然语言存储电子病历中的病历信息成为进一步数据挖掘和分析临床数据的障碍,电子病历领域里的自然语言处理测评推动电子病历中的命名实体识别和临床数据的结构化,提高对临床数据的重视和利用效率,推动自然语言处理研究领域与医院信息化的结合,向医学信息学注入了生机和活力。

参考文献

- 1 Chapman W W, Nadkarni P M, Hirschman L, et al. Overcoming Barriers to NLP for clinical Text: the role of shared tasks and the need for additional creative solutions [J]. Journal of the American Medical Informatics Association, 2011, 18 (5): 540-543.
- 2 Robertson S, Hull D: The TREC - 9 - Itering track - nal report [C]. Gaithersburg: The 9th Text REtrieval Conference (TREC 9), 2000.
- 3 Diaz N C, Córdoba J M, Ma? a M J, et al. Medical - Miner at TREC 2011 Medical Records Track [C]. Gaithersburg: Text REtrieval Conference - TREC Medical Records Track, 2011.
- 4 Pestian J, Brew C, Matykiewicz P, et al. A Shared Task Involving Multi - label Classi cation of Clinical Free Text [J]. Association for Computational Linguistics, 2006, (6): 97-104.
- 5 Pestian J, Matykiewicz P, Linn - Gust M, et al: Sentiment Analysis of Suicide Notes: a shared task [J]. Biomedical Informatics Insights, 2012, 5 (Suppl 1): 3-16.
- 6 Kelly L, Goeuriot L, Suominen H, et al. Overview of the ShARe/CLEF eHealth Evaluation Lab 2013 [J]. Information Access Evaluation, Multilinguality, Multimodality and Interaction, 2015, (2885): 172-191.
- 7 Uzuner O, Luo Y, Szolovits P. Evaluating the State - of - the - art in Automatic De - identification [J]. Journal of the American Medical Informatics Association, 2007, 14 (5): 550-563.

(下转第 21 页)

治; 在家中向医生做健康咨询; 随时得到医生的健康提醒; 患者将真正成为医院服务的中心。(4) 居民健康档案可以直接、快速、准确地为突发性、传染性、多发性病提供资料。(5) 服务平台的应用可为社区居民的医疗、保健需求提供专业化、智慧化的医疗服务, 有效地缓解了老、幼、病、孕、亚健康等特殊居民“看病难、看病贵”等现实问题。(6) 平台的应用各级医院相互联系, 资源共享, 注重提高现代化管理水平和服务水平, 医院将不再仅仅是一个看病的场所, 而是发展成为集医疗、保健及健康咨询为一体的服务型机构。

4 结语

通过打造社区智慧医疗服务系统, 实现患者与医务人员、医疗机构、医疗设备之间的互动, 逐步达到数字化、信息化, 使患者用较短的时间、较少的费用, 就能够享受到安全、便利、优质的诊疗服务, 在一定程度上解决“看病难、看病贵”等问题, 真正做到“人人健康, 健康人人”。

参考文献

1 李建功, 唐雄燕. 智慧医疗应用技术特点及发展趋势

[J]. 医学信息学杂志, 2013, 34 (6): 2-7.

2 邵星, 王翠香, 孟海涛, 等. 基于物联网的社区智慧医疗系统研究 [J]. 软件, 2015, 36 (12): 45-48.

3 方媛. 智慧医疗应用探索 [J]. 医学信息学杂志, 2014, 35 (12): 2-7.

4 石永峰, 曾云锋, 徐菲. 智慧医疗推动就医模式创新 [J]. 中国公共安全, 2013, (11): 29-34.

5 胡新丽. 物联网框架下的智慧医疗体系架构模型构建—以武汉智慧医疗为例 [J]. 电子政务, 2013, (12): 24-30.

6 倪荣, 陈启岳, 楼毅. 智慧医疗背景下移动在线支付医疗模式创新应用 [J]. 医学信息学杂志, 2014, 35 (12): 8-12.

7 吴妮娜, 白波, 张晓. 北京城区居民健康状况及社区智慧医疗服务研究 [J]. 中国社会医学杂志, 2013, 30 (6): 409-410.

8 王琳华. 云计算在智慧医疗集成平台中的应用 [J]. 中国数字医学, 2014, (2): 64-65.

9 古锐. 基于 Web 服务的药学信息服务系统设计 [J]. 医学信息学杂志, 2014, 35 (1): 29-31.

10 陆伟良, 杜昱, 侯惠荣, 等. 智慧医疗的现状与发展 [J]. 中国医院建筑与装备, 2016, (3): 82-84.

11 吴艳, 艳唐源. 医院智能化建设及其在智慧医疗发展中的作用 [J]. 医学信息学杂志, 2015, 36 (3): 38-41.

(上接第 13 页)

8 Uzuner O. Recognizing Obesity and Comorbidities in Sparse Data [J]. Journal of the American Medical Informatics Association, 2009, 16 (4): 561-570.

9 Uzuner O, Solti I, Cadag E. Extracting Medication Information from Clinical Text [J]. Journal of the American Medical Informatics Association, 2010, 17 (5): 514-518.

10 Uzuner O, South B R, Shen S, et al. 2010 i2b2/VA Challenge on Concepts, Assertions, and Relations in Clinical Text [J]. Journal of the American Medical Informatics Association, 2011, 18 (5): 552-556.

11 Uzuner O, Bodnari A, Shen S Y, et al. Evaluating the State

of the Art in Coreference Resolution for Electronic Medical Records [J]. Journal of the American Medical Informatics Association, 2012, 19 (5): 786-791.

12 Sun W, Rumshisky A, Uzuner O. Evaluating Temporal Relations in Clinical Text: 2012 I2B2 challenge [J]. Journal of the American Medical Informatics Association, 2013, 20 (5): 806-813.

13 Stubbs A, Uzuner O. Annotating Longitudinal Clinical Narratives for De-identification: the 2014 i2b2/UTHealth corpus [J]. Journal of Biomedical Informatics, 2015, 58 (Suppl): S20-S29.