情报学报　2017 年 1 月　第 36 卷　第 1 期
Journal of the China Society for Scientific and Technical Information, Jan. 2017, 36(1): 61-71

基于关键词链接网络分析方法的医学文献推荐服务研究

李树青[1]，徐　侠[2]，曹　杰[1]，庄光光[1]

（1. 南京财经大学信息工程学院，南京　210046；2. 南京邮电大学管理学院，南京　210046）

摘　要　本文提出了一种面向临床诊断决策支持服务的医学文献推荐方法，该方法首先对文献关键词从关键词列表、Mesh 标准医学词库和缩略词三个方面进行了规范化处理，并据此得到关键词和文献的完整对应关系。文章对相关处理方法的思路和细节都做了详细的说明。然后，利用已知的患者症状描述信息，根据关键词共现形式来获取目标句子集合和诊断相关关键词集合，并利用基于关键词共现形成的关键词链接网络，本文设计了一种测度重要关键词及其相关文献的查询方法。最后，文章对相关实验效果及其用户满意度评价都做了必要的说明。

关键词　文献推荐服务；关键词链接网络；临床诊断；决策支持服务

The Study of Medical Literature Recommendation Service Based on the Analysis of Keywords' Linking Network

Li Shuqing[1], Xu Xia[2], Cao Jie[1] and Zhuang Guangguang[1]

(1. *College of Information Engineering, Nanjing University of Finance & Economics, Nanjing*　210046, *China*;
2. *School of Management, Nanjing University of Posts and Telecommunications, Nanjing*　210046, *China*)

Abstract: This paper proposes a recommendation method of medical literature for clinical diagnostic decision supports. We get all the normalized keywords from the keywords lists in articles, Mesh lexicon and acronyms, then the full relation of keywords and articles are built with normalized keywords. The ideas and detailed processes are also introduced in this paper. Based on keywords co-occurrence analysis, the collections of target sentences and diagnosis-related keywords are abstracted according to existing description of symptoms. And the paper also lays out all the detailed process of measuring keywords and relevant articles with link analysis based on keywords co-occurrence. Finally, we report some related experiments and the results of user evaluations.

Key words: literature recommendation service; keywords' linking network; clinical diagnosis; decision support service

1　引　言

在海量医学文献中寻找有价值的目标文献，构成了现代文献信息检索一个重要的应用分支，也被多次列入信息检索技术评价测试任务，如从 2014 年 2016 年逐年都被列入美国 TREC 临床决策支持评测任务（TREC Clinical Decision Support Track，http://www.trec-cds.org）。它要求根据已有的患者症状描述信息，从近百万篇医学文献中自动获取与相关疾病诊断最为相关的文献列表。产生这个研究任务的主要原因在于两个方面：一是医学类文献的规模经过多年的积累和发展，已经形成了非常庞大的数据集合，从

收稿日期：2016-03-05；修回日期：2016-09-25
基金项目：国家社会科学基金项目“基于大数据分析的数字图书馆个性化服务模式创新研究”（16BTQ030），科技部科技支撑项目“外贸行业电子商务服务技术研究与应用”（BAH29F01）。
作者简介：李树青，男，1976 年生，教授，硕士生导师，主要研究领域为个性化服务、Web 挖掘和信息检索，E-mail：leeshuqing@163.com；徐侠，女，1977 年生，博士生，副教授，主要研究领域为科研管理；曹杰，男，1969 年生，博士生，教授，主要研究领域为推荐系统和商务智能处理；庄光光，男，1991 年生，在读研究生，主要研究领域为信息检索和文本挖掘。

中得到所需的有效文献变得愈发困难，同样，这个问题也存在其他类型的文献检索任务中；另一方面则来自于医学文献本身的特点，医学类文献中专业词汇聚集，相关关键词的构成结构和特点和一般文献具有明显的区别，比如大量含有希腊字母的医药表示方法，还有几乎所有的疾病专业术语都会提供对应的缩略形式，甚至很多文献通篇使用这些缩略词语而非原始对应词语，再如不同文献对于相同症状和疾病等术语的使用方法也存在较大的差异，同义词规范处理的必要性很强。

因此，如何在给定的医学文献集合中快速准确地找到所需的相关文献成为文献检索服务的一个重要研究领域。该项研究有很多不同的具体任务类型，如面向医学研究者的文献查询服务、面向临床治疗的文献辅助诊断服务等等。本文主要面向第二个方面，即根据临床观察获得的患者症状信息，在现有文献集合中找到与可能疾病诊断相关的文献信息，从而为临床医生提供决策支持。

2 文献回顾

计算机技术在临床决策方面的应用主要有医疗信息管理（Tools for Information Management）、诊疗提醒辅助[1]（Tools for Focusing Attention）和诊断决策（Diagnosis），其中诊疗提醒辅助还包括对药物互相作用的风险提醒[2]，以及在药物选择及剂量建议方面能给出建议[3][4]等。临床诊断决策（Clinical Diagnosis）是指结合患者的症状，从已有的数据分析中给出病理解释。与此相关的还有检验（Test），有时也指诊断过程分析（Diagnostic Process），即判断该进一步询问哪些问题，进行哪些检查，进行哪些治疗步骤，并结合预期的结果来决定可能的风险和成本[5]。这也是一种典型的循证医疗（Evidence-based Medicine）方法，也是近十几年发展较快的一门新兴临床学科[6]。

优秀的临床诊断决策支持需要三个前提，分别是准确的数据、合适的知识和有效的解决方法，其中关于解决方法的研究一直都是医学文献推荐服务领域的重要研究方向[7]。早期的研究可以完成差异化诊断和提示利用哪些信息可以不断加强对诊断结果判断的准确性，如 DXplain[8-9]和 QMR[10-11]等。有的系统可以按照一种方便临床医师理解的方式来将不同时间点上的患者记录进行总结[12]。从总体来看，医学类文献和患者记录都呈现比较规范的写作模式和方法，所以有利于从中提取各种定制化的结果[13]。但是已有研究也表明，临床诊断决策系统对于医生决策效果（Practitioner Performance）的支持有效性较高，但对于患者有用的结果信息（Patient Outcomes）则较少，其中的原因在于解读这些信息往往需要结合就诊时所收集的患者症状信息[14]。同时，也有学者指出使用者对于临床诊断决策系统的接受程度直接影响着系统的有效利用和使用效果[15]，还有学者强调管理和制度建设对于临床诊断决策系统的发展有着重要影响[16]。

所以，现有的诊断决策功能主要集中于药物选择支持和处方支持等方面，实现较为复杂的智能专家诊断决策支持尚不多见[17]。这也极大促进了当前相关研究的广泛开展。近年来在应用方面取得较大成就的往往多为一些针对特定病症的诊断决策支持服务。借助已有的标准诊断方案和治疗方法，可以使用计算机系统将人工操作进行程序化从而实现自动诊断，如有学者提出的对小儿哮喘临床决策支持系统，它按照美国国家哮喘教育与预防项目的标准，可以在患者就诊时按照症状标准自动评估严重等级和给出建议治疗步骤[18]。还有学者通过实验证明临床诊断决策支持系统可以降低急性肺栓塞患者中使用肺血管造影术的比重达 20%，而且结合肺血管造影术的判断准确度也提高了 69%[19]。也有学者说明临床诊断支持系统可以显著降低那些低风险患者中使用心电图的比重，但也强调总体上没有看出明显区别[20]。

具体到方法而言，目前已有多种方法可以实现相关应用，如自动问答系统、基于规则库的分析方法以及网络分析方法等。

随着自动问答（Question Answering）系统的快速发展，越来越多的学者将其利用到临床诊断决策支持服务上，常见方法是将复杂问题分解为以事实查询为基础的问答（Fact-seeking Questions）形式或者将其映射为其他相似的简单问题[21-22]，更为有效的方法往往结合语义领域模型（Semantic Domain Model），如 PICO 框架，它将诊断问题映射为四个主要子类型，分别是问题（Problem）、处置（Intervention）、比较（Comparison）和结果（Outcome），以此实现问题的规范化表达，并利用自然语言理解技术和已有的医学知识库完成文档查询，并再次利用聚类方法完成答案的自动抽取和提炼[22]。不过，该项技术非常依赖于正确有效的利用 PICO 框架来描述患者症状信息[23]。

关于基于规则库的分析方法，如有学者选取刺

参典型病例，根据提炼出的刺参疾病诊断规则建立起规则库，构建刺参疾病诊断推理机，并结合 BP 神经网络方法实现了辅助海参疾病诊断的自动化[24]。还有学者实现了基于 BP 神经网络模型的呼吸系统疾病诊断仿真系统[25]。这些系统都需要比较典型的训练数据，所以通常适用于对特定疾病的自动诊断，并不适合于一般性的通用型临床诊断决策服务。

相对于前两种方法而言，网络分析方法应用面极广，方法的具体种类也很多。常见的方法有两大类：第一类是利用二分网络投影方法[26]，将疾病和特征分别归入二分网络的两个不同节点内容中，并利用迭代算法计算出彼此的相关性。如有学者使用二分网络来判断遗传性疾病与致病基因之间的相关度，并据此探究遗传性疾病的发病机制[27]。值得注意的是，通过我们的实验观察，这种方法对于辨析度不高和语义量有限的关键词单元而言，查询相关有效文献的实验效果往往不好，相反对于辨析度较高且信息量较大的基因片段却较为有利；第二类为基于网络链接的节点权值算法，此类方法在诸如网页推荐等领域中得到了广泛的应用。我们在前期的研究中，也多次使用该方法在图情学科领域本体自动构建[28]和学术文献关键路径自动识别[29]方面进行了理论研究。实验表明，该方法在利用关键词来分析文档相关性的过程中能够起到非常有效的作用。对于临床诊断决策中的文献推荐服务而言，该种方法无需对患者症状的复杂表示方法和各种复杂的先验规则库，而利用描述症状信息的已有关键词，直接在医学文献数据集中查询相关度较高的推荐文献结果，这也构成了本文的主要研究特点。

3　标准关键词集合的获取

由于不同文献写作方式的差异和表述的不同，医学文献的相关关键词必须经过专门的处理才能直接使用，这包含多项具体任务。

3.1　关键词获取范围的选择

3.1.1　文献关键词

这部分关键词来源于文献本身的关键词列表，也是内容规范度最差的一类，主要问题有：

（1）由于不同文献数据库格式的差异，有的医学文献使用逗号区分不同关键词，但也有的使用分号，甚至有的使用反斜杠符号。因此在数据解析时，必须根据当前文献关键词的分割符号来进行不同的处理。

（2）混杂大量的无效关键词，除了常见的停用词外，不少文献还有纯数字形式的关键词，甚至将“Keyword”等词语也作为关键词，而且就这一个无效关键词，不同文献还有多种不同的写法，如“Keywords”“Key terms”“Key indexing terms”“Key indexing terms”等，更何况还有很多明显错误的关键词。因此，我们必须对得到的关键词进行规范性处理。我们采用了以词频统计为基础的判断方法，基本思想如下所述：

- 得到全部文献的全部关键词
- 去除停用词和无效关键词（纯数字形式、字段名称、长度异常大的关键词等）
- 对得到的所有关键词再次统计它们在每篇文献标题、摘要和正文中的词频
- 去除低频关键词
- 得到每篇文献新的关键词列表

这个最终的关键词列表具有如下两个优点：一是全面，对于一些没有能在关键词列表罗列完整的文献而言，通过这种方式可以极大地扩张有效关键词数量；二是准确，即使一些文献在自己的关键词列表中错误标注了关键词，但是通过在当前文献其他字段中获取的正确关键词，而且这些正确关键词的数量也明显高于错误关键词数量，从而为我们判断文献的正确关键词提供了基础。事实上，我们对于词频较低的关键词和文档频率较低的关键词都做了直接去除的处理方式。

3.1.2　医学词库的标准关键词

目前已有很多质量较高的医学词库，如 Mesh 和 UMLS 等。我们使用的是 Mesh，它将全部收录的医学关键词分为三个等级，第一等级为“描述符（Descriptor）”，第二等级为“概念（Concept）”，第三等级为“术语（Term）”，每一个概念都只属于一个描述符，每一个术语也都只隶属于一个概念。其中，概念是 Mesh 词库中最为标准的关键词内容，与同一概念对应的术语可能有多种不同的写法但是具有相同的语义，从词语字面形式来看，处于最底层的术语包含了所有的描述符和概念，而概念则包含了所有的描述符。

同时，Mesh 词库还提供一组同义词映射关系，即给出了语义相近的概念映射关系，从而为相关术语的同义词映射提供了查询基础，相关联系如图 1 所示。

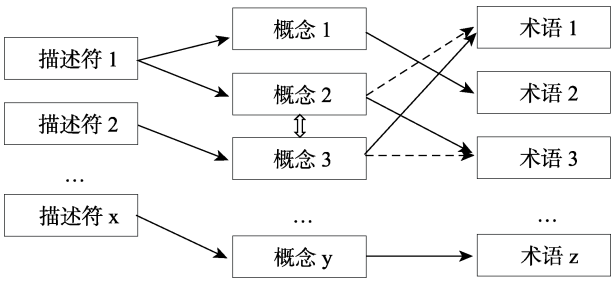

图 1 Mesh 词库中不同层级间词语联系示意图

由于描述符、概念和术语之间的联系都为一对多，因此词语联系图从左向右呈现一个树状结构，术语层构成了叶节点。但是，利用概念之间的同义联系，可以在概念和术语层之间建立语义更为丰富的网状联系。假设图 1 中“概念 2”和“概念 3”存在同义联系，则我们可以给“术语 1”和“术语 3”建立如虚线表示的新概念映射。因此利用这种结构对应关系，我们可以扩展文献的同义关键词数量，从而有效提高了结果的查全率。

3.1.3 缩略词

虽然缩略词本身很少直接出现在文献的关键词列表中，但是在描述症状和疾病时，它却是最为常见的有效词语。目前能够直接获取到的医学缩略词表并不多，而且内容也不全，因此我们探索了从文献集合中直接获取的新方法。

值得注意的是，不是所有的医学文献都会在给出缩略词的同时，也给出原始词语的对应形式，而且医学文献中的缩略词写法种类极其多样，并非总是取连续多个单词的首字母大写。所以，我们采用的启发式判断方法建立在一个较为灵活的假设基础之上，那就是文献中出现的缩略词及其原始词语对应形式遵循一个固定的出现格式，形如“A*B*（AB*）”，其中的*字符表示任意其他多个字母或者空格。这里只考虑了前两位大写字母的出现形式，原因在于部分缩略词甚至使用了小写字母和数字。

该方法的基本思路是直接从文献正文中搜寻所有可能的缩略词及其原始词语，所依据的特征是若干大写字母开头的关键词组并且后面紧跟着位于括号中的缩略词。具体方法描述如下：

（1）对于每一篇文献，查询所有正文出现的前大括号。

（2）判断紧跟大括号的至少两个连续字母为大写字母。

（3）如是，则查找大括号前出现这两个大写字母的位置，如果两个字母出现顺序与括号后顺序一致，则截取从括号前第一个大写字母开始到括号前的所有关键词，并判断是否包含诸如句号、逗号等之类的无效字符。同时，以非字母和非数字为结束字符，截取括号后的缩略词，形成一组缩略词及其原始词语的对应关系。

（4）对于得到的全部缩略词及其原始对应词语，进行必要的验证，包括：

- 去除长度不大于缩略词的原始词语及其缩略词
- 相同原始词语对应多个不同长度的缩略词，如一缩略词为另一缩略词的一部分，则去除该缩略词表示
- 检查全部缩略词，依此判断每个原始词语是否按序出现了对应缩略词的每个字母。如不是，则去除

（5）对于所有缩略词，统计出现词频，将词频较低的去除，因为这些可能都是误判，即使不是，也没有实际意义。实验表明，利用词频统计，可以非常简单和准确的去除无效和误判缩略词。

4 关键词文档关系的抽取

4.1 基本抽取算法

为了方便后续的实验分析，本文对每篇文献的关键词进行了三种层次的关键词提取，即句子（以句号分割）、段落（以分段符分割）和文档三个层次。对于上述三类关键词，分别采用不同的抽取方法。另外全部文献数据都做了小写统一转换，多个空格置换为一个空格，同时鉴于医学文献的特点，保留了希腊字母。

由于关键词数量和文献数量都很庞大，直接循环遍历每篇文献中出现的每个关键词极其低效。本文采用了一种以最大关键词长度匹配的单次遍历算法，对每篇文献从前往后单向快速遍历，基本思路如下所述：

（1）循环遍历每一篇文献。

（2）对于每一篇文献，从前往后，首先截取已有关键词集合中最大长度的词语序列，并判断是否匹配已有关键词。

（3）如是，则从匹配结束处的下一字符开始继续步骤 2 的操作。

（4）如不是，则对已截取的词语序列，从后往前，以空格为结束单位，逐一获取长度渐短的词语序列，并依此判断是否匹配已有关键词，一旦判断

成功，则从匹配结束处下一字符开始继续步骤 2 的操作。如果全部判断不成功，则从文献现有起始位推移一个关键词，继续步骤 2 的操作。

4.2　不同类型的关键词抽取和处理

对于三类关键词集合，我们采用了三种不同的抽取和处理方法。

对于文献关键词，直接从文献正文采用 4.1 算法抽取。

对于 Mesh 医学词库的标准关键词，除了直接抽取关键词外，如果当前关键词存在与其他关键词相同的上级概念，则也将那些其他关键词标注在当前文献中，词频相同。

对于缩略词，我们采用了两步处理的方法。首先，直接扫描所有缩略词的原始词语形式，采取和抽取一般文献关键词相同的策略；其次，再扫描所有文献中的缩略词（保留大写形式进行判断）。由于同一缩略词可能对应多个不同的原始词语形式，因此必须再次判断当前文献中缩略词所对应的原始词语是哪一个。我们仍然采用了以词频统计为基础的算法，即如果一篇文献出现某一缩略词，则查询该文献是否出现与该缩略词对应的所有原始词语，并将出现频次最高并且长度最大的那一个原始词语作为最终的有效对应原始词语。也就是说，最终我们得到的文档和关键词对应关系中并不包含缩略词，只包含缩略词所对应的原始词语。

4.3　关键词整理

最后我们可以得到每篇文献及其出现关键词的对应关系，考虑到医学类关键词主要为名词形式，再次对单复数表达形式进行规范化处理。最终，我们可以得到以句子、段落和文档三个单位的关键词词频形式，并可以据此计算出不同单位上的关键词词频和文档频率。这些关键词数据形成了后续研究的数据基础。

5　临床诊断辅助文献的自动判断方法

对于临床医生，通常在接受患者的时候，只能通过观察和询问的方式，来收集有关患者身体条件、当前状况及其相关症状信息。此时就需要通过临床医生自身的经验和知识，来大致判断可能的疾病以及需要哪些进一步的化验方法。利用已有的医学文献，其实也可以提供一种更好的决策支持辅助服务，很多医学文献都记载了相关患者诊断的具体过程，而且还提供较为详细的治疗方案。

对于临床医生所收集到的患者症状信息，通常表现为一段文字性的描述。因此要想实现临床诊断决策支持文献的自动推荐，我们必须完成下面两个基本步骤：一是利用患者症状信息直接查询与此相关的目标文献；二是判断该目标文献集合中哪些文献更为相关，最终以一个排序的文献列表形式提供给临床医生。下面分别予以介绍。

5.1　获取反映患者症状的特征关键词信息

首先利用已有的关键词集合，直接查询患者症状描述信息的相关出现情况。由于描述信息通常较短，大部分关键词词频都只为 1，因此描述信息中的词频信息可以忽略。但是，由于我们得到的标准关键词集合数据庞大，直接匹配后命中的关键词数量会很多，因此需要进一步得到真正与诊断有关的关键词。这分为两个步骤：一是利用文档频率，我们可以对相关关键词排序，对于高频关键词可以做忽略处理；二是不同的关键词组合通常会反映不同的疾病，因此我们利用以句子为单位的划分单元，得到这些关键词的共现组合对。这种共现组合对中的关键词数量在实际实验中大部分为 2，也有部分为 1 或者 3。值得说明的是，这部分工作需要用户介入，人工辅助选择最为恰当的关键词共现组合对，可以极大地提高结果的有效性。

5.2　获取相关句子集合

利用得到的标准关键词集合和从患者症状描述信息中抽取的重要关键词，我们建立了测度相关文献集合的基本数据准备。虽然最终需要获取的是文献，但为了提高相关度的计算准确性，我们采用了结合句子和段落单位的综合分析方法。对于关键词共现分析而言，单纯使用句子为单位，虽然精度较高，但是所丢失的内容也较多，而单纯使用段落，又在一定程度上引入了太多相关度不高的共现对。所以，在本文所采取的方法中，对于在患者症状描述信息中位于同一句子的重要关键词对，我们逐篇计算文献哪些段落存在具有相同重要关键词对的句子，如果有并且出现的句子数量大于事先定义好的阈值，则将当前段落加入目标段落集合，据此可以得到反映相关文献集合的目标段落集合。同样，再次以段落为单位，判断段落中是否出现足够多的不同句子来过滤段落，最终就可以得到位于有效段落集合中的全部句子。这种方法兼顾了共现有效性和

结果覆盖面两个方面，实验表明效果良好。

5.3 获取诊断相关关键词集合

所谓诊断相关关键词，是指与当前症状信息有关的疾病、治疗等相关关键词。它们虽然没有出现在患者症状描述信息中，但是从理论上看，包含重要关键词对的句子所在段落通常也会含有此类诊断相关关键词。因此，我们必须能够从这些目标句子集合中，计算出具有较高权值的诊断相关关键词。

为此，我们抽取目标句子集合中所有关键词作为节点，将它们在同一句子中的共现作为链接，采用类似于 PageRank 的网络链接权值计算方法，来得到每个关键词的最终权值。有两点需要补充：一是由于链接权值算法要求有向链接，由于逆文档频率反映了关键词本身的词语辨析度和价值，因此我们对每个共现对中的关键词，按照逆文档频率高的关键词指向逆文档频率低的关键词建立了链接；二是为了减少计算复杂度和减少无关关键词的干扰，我们对那些逆文档频率高于既定阈值的关键词再次做去除处理。

传统的 PageRank 方法是一种计算网页重要度权值的有效迭代算法，具有较高权值的网页通常都是那些处于入度较大的核心网页。类似于此，作为能够反映最终疾病诊断信息的关键词，也应该在上述以逆文档频率为基础建立的关键词链接集合中，具有较大的入度和相应的权值，它的假设基础在于很多相关文献会在谈及相关患者症状的同时，都会说明相关疾病的名称。

具体关键词权值计算方法如式 1 所示：

$$weight_{n+1}(keyword_k) = (1-\alpha) + \alpha \times \sum \frac{idf(inKeyword_i) \times weight_n(inKeyword_i)}{C(inKeyword_i)} \quad (1)$$

其中，$weight_n$ 表示第 n 次迭代运算时关键词的权值，α 为衰减因子，idf 表示关键词的逆文档频率值，C 表示关键词的出度，$inKeyword$ 表示指向当前关键词的链入关键词。和传统 PageRank 方法相比，该算法强调了关键词自身逆文档频率的影响力，显然一个关键词如果具有较高的入度，而且每个入度都来自于逆文档频率值较高而且权值较高的关键词，则当前关键词就具有较大的权值。

5.4 文献相关度值的计算

从上述得到的关键词及其权值集合中，我们就可以计算文献的相关度值。由于上述关键词都是以句子为单位，所以一篇文献如果有越多含有高权值关键词的句子，则它的相关度就越大，能够提供的有效诊断信息就越多。具体计算方法如式 2 所示：

$$weight(doc_k) = avg(weight(keyword_i^{dock})) \quad (2)$$

6 实验说明

6.1 实验环境准备

实验数据来自于 TREC 2014 临床决策支持评测任务中提供的文献资源，网址为 http://www.trec-cds.org/2014.html，其中总共包含 733328 篇医学临床类相关科研文献，每篇文献都是以 XML 文件格式存储，并且有一个唯一的文献号。

6.2 关键词获取的相关实验说明

6.2.1 文献关键词的获取

按照第 3 节所述方法，我们从文献关键词列表中抽取了有效关键词共 347999 个，过滤停用词使用了 Reuters-RCV1（Reuters Corpus Volume 1）提供的 25 个标准常见词语，文档频率最高的前 10 个关键词如表 1 所示。

表 1 文档频率最高的前 10 个关键词

关键词	文档频率
apoptosis	3472
breast cancer	3019
epidemiology	2922
inflammation	2639
cancer	2435
children	2196
obesity	2163
prognosis	1740
oxidative stress	1719
depression	1706

从中可以看出没有与临床医学无关的关键词，这为下一步的实验处理提供了良好的关键词数据基础。

6.2.2 Mesh 词库标准关键词的获取

从 Mesh 医学词库中，我们提供了 27149 个描述符，51525 个概念和 218985 术语，其中概念和术语的映射关系共 49412 个。部分 Mesh 词语及其关系如表 2 所示。

表 2 所示的第一大行反映了不同层次之间标准的词语映射关系，从中可以看出概念层对于词语规

范化处理的用途和价值。第二大行和第三大行则展示了利用概念之间的同义联系，在不同术语层之间建立同义词联系的方法，比如因为概念“Breath Tests”和“Breathalyzer Tests”为同义概念，所以相关下级术语分别增加了新的上级概念映射层，在图 2 中使用浅灰底色表示。实验证明，这种规范度较高的词库词语对于关键词规范化处理和提高结果查全率有着重要作用。

表 2　部分 Mesh 词库标准关键词及其相互关系

描述符号	描述符	概念号	概念	术语号	术语
D004417	Dyspnea	M0006931	Dyspnea	T364731	Shortness of Breath
				T364731	Breath Shortness
				T364731	Breath Shortnesses
D001944	Breath Tests	M0002911	Breath Tests	T005544	Breath Tests
				T005544	Breath Test
		M0002912	Breathalyzer Tests	T005544	Test, Breath
				T005544	Tests, Breath
D001944	Breath Tests	M0002912	Breathalyzer Tests	T005545	Breathalyzer Tests
				T005545	Breathalyzer Test
		M0002911	Breath Tests	T005545	Test, Breathalyzer
				T005545	Tests, Breathalyzer

6.2.3　缩略词的获取

实验从文献正文中获取缩略词及其原始词语的表达对共 289670 个，经过验证处理和去除较低词频的表达对，最终得到了 139888 个有效结果。我们以“AIDS”为例，展示了相关处理结果，如表 3 所示。

表 3　缩略词“AIDS”及其各种原始词语形式*

缩略词	原始词语形式	出现频率
AIDS	acquired immunodeficiency syndrome	975
AIDS	acquired immune deficiency syndrome	466
AIDS	acquired immuno deficiency syndrome	35
AIDS	acquired immunodeficiency disease syndrome	8
AIDS	autoimmune deficiency syndrome	8
AIDs	autoimmune diseases	6
AIDs	autoinflammatory diseases	4
AIDs	autoinflammatory disorders	3
AIDS	acquiredimmunodeficiency syndrome	3
AIDS	acquired immunodeficiencysyndrome	3

注：*请注意部分缩略词存在小写字母形式。

从中可以看出利用高频形式可以给出相关缩略词的正确表达形式，但是中频形式也提供了很好的同义表达形式。我们利用该数据，可以判断文献中出现的缩略词最可能的是哪一个原始词语形式，也可以利用最高词频的原始词语形式来规范化当前文献中其他相同缩略词的最终原始词语表达形式。

6.2.4　关键词整理

利用得到的全部关键词，在去重汇总处理后，得到了最终关键词集合，共计 311379 个。我们对每篇文献进行了以句子、段落和文档三个单位的统计处理，共计 148757694 个句子和 24054364 个段落，得到关键词和句子的出现关系对共计 881930831 个。

6.3　辅助文献自动判断方法的实验结果

TREC 2014 提供了 1 个带有推荐结果的测试用例和 10 个尚未公布答案的标准用例，我们分别进行了测试，下面结合测试用例来说明具体的实验情况。

该测试用例内容为：“A woman in her mid-30s presented with dyspnea and hemoptysis. CT scan revealed a cystic mass in the right lower lobe. Before she received treatment, she developed right arm weakness and aphasia. She was treated, but four years later suffered another stroke. Follow-up CT scan showed multiple new cystic lesions”。其对应的中文含义为“一名 35 岁左右的妇女患有呼吸困难（dyspnea）和咳血（hemoptysis），CT 检查表明在右肺下叶（right lower lobe）有囊性肿块（cystic mass）。治疗前患者已有右臂无力（right arm weakness）和失语（aphasia），治疗四年后，再次患中风（stroke），CT 检查表明新增多处囊性病变（cystic lesions）”。

按照第 5 节所述方法，我们以完整关键词列表扫描用例得到备选关键词列表，再从中获取了 7 个重要关键词及其以句子为单位的共现关系，分别是“dyspnea”和“hemoptysis”、“mass”和“lobe”、“weakness”和“aphasia”、“stroke”。

以段落中句子是否出现上述四组关键词共现对之一为标准，我们扫描全部文献，共获得 88512 个段落，相关文档数量为 30622 篇。再以至少出现两个命中句子为标准，我们再次过滤段落数量，最终得到相关段落为 58 个，相关句子为 625 个，涉及了 52 篇文档。

在句子集合中，我们得到了 1728 有效关键词，以所在句子共现关系得到的关键词链接对为 38456 个。按照 5.3 节关键词权值计算方法，最终得到了每个关键词的权值，其中具有最高权值的关键词如表 4 所示。

表 4 具有最高权值的 10 个关键词及其权值

关键词	逆文档频率	人度	关键词权值
aphasia	6.0157	656	7.3642E-4
functional ambulation category	10.8611	156	6.8603E-4
ergometry	7.7948	332	6.6700E-4
severe adverse event	10.7180	58	6.5213E-4
pulmonary arteriovenous malformation	10.6534	56	6.4324E-4
sporadic hemiplegic migraine	11.2288	66	6.4199E-4
hemianopsia	8.2844	91	6.3998E-4
kaolin clotting time	11.3466	83	6.3093E-4
semantic dementia	8.0144	43	6.3048E-4
neurologic symptom	9.4007	43	6.2866E-4

按照 5.3 节式（2）可以得到结果文献的权值，具有最高权值的文献如表 5 所示。

从表 5 中可以看出，前两篇文献主要介绍“肺动静脉畸形（Pulmonary ArterioVenous Malformation，PAVM）”，符合最终的患者症状描述特征和诊断结果。第三篇和第四篇虽然分别主要谈及“偏头疼（Migraine）”和“后脑动脉梗塞（Infarction in Posterior Cerebral Artery）”，但是较为全面的涉及了相似症状特征和诊疗判断方法，这对于临床医生进一步区分和决策提供了对比分析基础。最后一篇为“中风康复（Rehabilitation in Stroke）”，内容较少，相关度相对较低。

表 5 具有最高权值的 5 篇文献及其权值

文献号	文献标题	权值
3025345	A Case of a Pulmonary Arteriovenous Malformation With Ebstein's Anomaly	5.8679E-4
3148967	Stroke in hereditary hemorrhagic telangiectasia patients. New evidence for repeated screening and early treatment of pulmonary vascular malformations: two case reports	5.8396E-4
3420796	Sporadic Hemiplegic Migraine with Seizures and Transient MRI Abnormalities	5.8335E-4
3180463	Infarctions in the vascular territory of the posterior cerebral artery: clinical features in 232 patients	5.8289E-4
3287708	Maximising adherence to study protocol within pharmaco-rehabilitation clinical trials	5.8238E-4

在该测试用例所提供的 3 个推荐结果中，我们的方法只命中了其中的第二篇文献。之所以没有命中其他两篇推荐文献，主要原因在于第一篇推荐结果文献（文献号：2987927）在描述患者症状时篇幅较大，从而没有一个句子具有两个重要关键词，而第三篇推荐结果文献（文献号：3082226）却根本没有任何患者症状描述信息，这使得这两篇文献没有进入我们预选的 52 篇目标文献集合中。这也表明单纯使用关键词方法可能存在的问题。但是，我们实验结果给出的第一篇文献也正确指出了可能症状的名称和较为相似的患者症状描述，也是高度相关文献。

我们再次对 10 个尚未公布答案的标准用例进行了测试，下面结合其中一例说明实验效果。该标准用例内容为：“A 58-year-old African-American woman presents to the ER with episodic pressing/burning anterior chest pain that began two days earlier for the first time in her life. The pain started while she was walking, radiates to the back, and is accompanied by nausea, diaphoresis and mild dyspnea, but is not increased on inspiration. The latest episode of pain ended half an hour prior to her arrival. She is known to have hypertension and obesity. She denies smoking, diabetes, hypercholesterolemia, or a family history of heart disease. She currently takes no medications. Physical examination is normal. The EKG shows nonspecific changes”。其对应的中文含义为“一名 58 岁的非裔美国妇女因为两天前首次感觉不规律的前胸压痛灼烧感来急诊，疼痛是在走路时开始的，放射到背部，同时还伴有恶心、出汗和轻微的呼吸困难，但是没有随着吸气感觉加重。在到达前半小时疼痛消失了。她有高血压和肥胖症，不吸烟，无糖尿病和高胆固醇血症，也没有家族心脏病。目前没有使用任何药物，体检正常，心电图显示没有异常”。相关查询结果只有 4 篇，如表 6 所示。

表 6 具有最高权值的 4 篇文献及其权值

文献号	文献标题	权值
2994533	Primary pericardial malignant mesothelioma and response to radiation therapy	3.5544E-3
3681367	Medical emergencies on board commercial airlines: is documentation as expected	3.5278E-3
3258729	Epipericardial fat necrosis – a rare cause of pleuritic chest pain: case report and review of the literature	3.5189E-3
3490454	Does the patient with chest pain have a coronary heart disease? Diagnostic value of single symptoms and signs – a meta-analysis	3.5066E-3

第一篇文献主要介绍了“原发性心包恶性间皮瘤（Primary Pericardial Malignant Mesothelioma）”，文中患者症状与用例一致，相关度较高。第二篇主要介绍“飞行医疗紧急救治（In-flight Medical Emergencies）”的相关文献分析，内容不相关，但是由于为文献分

析类内容，因此在文字匹配度上较高，从而获得了较大的权值提升，这也从一个侧面说明了单纯使用关键词匹配可能存在的问题。第三篇介绍“心膜周围脂肪坏死（Epipericardial Fat Necrosis）”，文中患者症状与用例一致，相关度较高。最后一篇为介绍“胸痛（Chest Pain）”和“冠心病（Coronary Heart Disease）”联系的文献回顾类文章，提供了大量有价值的对比数据和分析。

最后我们利用用户评价方法对 1 个带有推荐结果的测试用例和 10 个尚未公布答案的标准用例进行了用户满意度评价，我们选择 5 位具有医学类专业知识和 6 位具有信息检索评价方面知识的测试用户，由不同测试者独立评价其中 2 个查询的前 5 个最高权值文献结果，因此每个查询结果只被 2 个用户评价，每个结果评价层次为 5 个级别，5 为最优，1 为最差。最后统计用户对上述功能的评价情况。结果如表 7 所示。

表 7　用户满意度评价结果

文献序号	第 1 篇文献		第 2 篇文献		第 3 篇文献		第 4 篇文献		第 5 篇文献	
用户 1	5	5	4	5	3	3	3	2	1	3
用户 2	5	5	2	1	4	5	5	4		
用户 3	5	4	5	5	4	3	3	5	5	5
用户 4	5	5	5	4	4	5	4	5	5	5
用户 5	5	5	2	3	1	1				
用户 6	4	4	5	5						
用户 7	4	3	3	4	3	3	4	5	5	5
用户 8	3	4	3	2	5	5	5	5		
用户 9	5	5	5	5	3	4	3	3	4	3
用户 10	5	5	5	5	4	3	3	3	3	4
用户 11	1	2	3	4	5	5	5	5	5	5

我们从两个方面对这部分用户评价数据进行了有效性测度。一是利用评价结果的序列一致性，我们认为对于排序输出的记录而言，最理想的序列应该是第一条记录具有最高等级 5，而最后一条记录应该具有最低等级 1。因此，我们对每个用例的两人评价结果进行，评价判断和理想序列一致的设定为 R，不一致设定为 N，重叠度的计算指标为：RR/(NR+RR+RN)。完整的一致性统计表格数据如表 8 所示。

二是利用评价结果的直接一致性，我们只考虑在一个用例的一个输出结果上，两人的评价等级是否一致，该种方法没有考虑整体序列的持续，该出点发在于认为有时部分结果可能会有多条高质量的输出记录结果。如果一致，则相应输出结果值为 1，否则为 0。完整的直接一致性统计表格数据如表 9 所示。

表 8　用户满意度评价的序列一致性统计结果

	NN	NR	RR	RN	重叠度
1	0	1	2	2	0.4
2	3	0	1	0	1
3	3	1		1	0
4	3	1	1		0.5
5	2		1		1
6	2				
7	3	1	1		0.5
8	4				
9	3		1	1	0.5
10	3	1	1		0.5
11	4	1			0
平均值					0.489

表 9　用户满意度评价的直接一致性统计结果

	1	2	3	4	5
1	1	0	1	0	0
2	1	0	0	0	
3	0	1	0	0	1
4	1	0	0	0	1
5	1	0	1		
6	1	1			
7	0	0	1	0	1
8	0	0	1	1	
9	1	1	0	1	0
10	1	1	0	1	0
11	0	0	1	1	1
平均值			0.5		

两种方法测算结果都显示一致性比较理想。进一步分析可以看出，因为部分查询返回结果数量较少，空白表示没有该项查询文献。总体平均满意度为 3.98，高出平均值 33%，显示出较好的查询效果。如果只考虑前几个返回文献结果，可以发现查询准确度会随着查询结果数量的增加而逐渐下降，随后基本稳定而略有上升。具体数据如表 10 所示。相关变化趋势如图 2 所示。

表 10　只考虑前几个返回文献结果时的平均用户满意度

文献结果数量	平均用户满意度
只考虑前 1 个文献结果	4.27
只考虑前 2 个文献结果	4.07
只考虑前 3 个文献结果	3.94
只考虑前 4 个文献结果	3.95
只考虑前 5 个文献结果	3.98

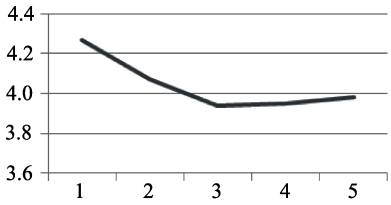

图 2 平均用户满意度随查询结果数量的变化趋势图

7 结 语

本文提出了一种面向临床诊断决策支持服务的文献推荐方法，该方法利用规范化的文献关键词，结合基于关键词共现形成的关键词链接网络，探索了重要关键词及其相关文献的查询方法。从实验效果来看，初步实现预期的设计目标，同时还具有易于实现等特点。不过该方法也存在很多需要进一步研究和改进的地方，主要有以下两点：一是目前受限于有限的患者症状信息描述，我们只能利用关键词来查询相关推荐文献结果，部分实验数据也说明由于文献写作风格的差异，可能会导致文献的误判和漏检，因此，进一步结合更多的患者症状特征来加强推荐结果的有效性；二是推荐文献结果本身只是提供了一种供临床医生参考的基本素材，我们准备在下一步的研究中，探索将推荐文献结果再次进行加工整理，使用文本总结和提炼有效成分的方法，以一种用户友好的可视化界面予以呈现，以加强用户使用的方便度和应用价值。

参 考 文 献

[1] McCoy A B, Waitman L R, Lewis J B, et al. A framework for evaluating the appropriateness of clinical decision support alerts and responses[J]. Journal of the American Medical Informatics Association, 2012, 19(3): 346-352.

[2] Saverno K R, Hines L E, Warholak T L, et al. Ability of pharmacy clinical decision-support software to alert users about clinically important drug-drug interactions[J]. Journal of the American Medical Informatics Association, 2012, 19(7): 114.

[3] Kuperman G J, Bobb A, Payne T H, et al. Medication-related clinical decision support in computerized provider order entry systems: a review[J]. Journal of the American Medical Informatics Association, 2007, 14(1): 29-40.

[4] Moxey A, Robertson J, Newby D, et al. computerized clinical decision support for prescribing: provision does not guarantee uptake[J]. Journal of the American Medical Informatics Association, 2010, 17(1): 25-33.

[5] Musen M A, Middleton B, Greenes R A. Clinical decision-support systems[M]. London: Biomedical Informatics, Springer, 2014: 643-674.

[6] 郑荣佩. 在《中图法》中增设循证医学类目的探讨[J]. 中国图书馆学报, 2007, 33(4): 100-103.

[7] 牟冬梅, 张艳侠, 黄丽丽, 等. 基于 SNOMED CT 和 FCA 的医学领域本体构建研究[J]. 情报学报, 2013, 32(6): 653-662.

[8] Elkin P L, Liebow M, Bauer B A, et al. The introduction of a diagnostic decision support system (Dxplain) into the workflow of a teaching hospital service can decrease the cost of service for diagnostically challenging diagnostic related groups (DRGs)[J]. International Journal of Medical Informatics, 2010, 79(11): 772-777.

[9] Barnett G O, Cimino J J, Hupp J A, et al. DXplain: an evolving diagnostic decision-support system[J]. Journal of the American Medical Association, 1987, 258(1): 67-74.

[10] Heckerman D. A tractable inference algorithm for diagnosing multiple diseases[J].Machine Intelligence and Pattern Recognition, 1990, 10: 163-171.

[11] Shwe M A, Middleton B, Heckerman D E, et al. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base[J]. Methods of Information in Medicine, 1991, 30(4): 241-255.

[12] Klimov D, Shahar Y. iALARM: an intelligent alert language for activation, response, and monitoring of medical alerts[M]//Process Support and Knowledge Representation in Health Care 2013. London: Springer International Publishing, 2013: 128-142.

[13] Elhadad N, Kan M Y, Klavans J L, et al. Customization in a unified framework for summarizing medical literature[J]. Artificial Intelligence in Medicine, 2005, 33(2): 179-198.

[14] Jaspers M W, Smeulers M, Vermeulen H, et al. Effects of clinical decision-support systems on practitioner performance and patient outcomes: a synthesis of high-quality systematic review findings[J]. Journal of the American Medical Informatics Association, 2011, 18(3): 327-334.

[15] Seidling H M, Phansalkar S, Seger D L, et al. Factors influencing alert acceptance: a novel approach for predicting the success of xlinical decision support[J]. Journal of the American Medical Informatics Association, 2011, 18(4): 479-484.

[16] Wright A, Sittig D F, Ash J S, et al. Governance for clinical decision support: case studies and recommended practices from leading institutions[J]. Journal of the American Medical Informatics Association, 2011, 18(2): 187-194.

[17] Wright A, Sittig D F, Ash J S, et al. Development and evaluation of a comprehensive clinical decision support taxonomy: comparison of front-end tools in commercial and internally developed electronic health record systems[J]. Journal of the American Medical Informatics Association, 2011, 18(3): 232-242.

[18] Hoeksema L J, Bazzy-Asaad A, Lomotan E A, et al. Accuracy of a computerized clinical decision-support system for asthma assessment and management[J]. Journal of the American Medical

Informatics Association, 2011, 18(3): 243-250.

[19] Raja A S, Ip I K, Prevedello L M, et al. Effect of computerized clinical decision support on the use and yield of CT pulmonary angiography in the emergency department[J]. Radiology, 2012, 262(2): 468-474.

[20] Romano M J, Stafford R S. Electronic health records and clinical decision support systems: impact on national ambulatory care quality[J]. Archives of Internal Medicine, 2011, 171(10): 897-903.

[21] 欧石燕. 基于文本蕴涵的受限领域自动问答方法研究[J]. 情报学报, 2011, 30(5): 540-547.

[22] 游斓, 周雅倩, 黄萱菁, 等. 基于最大熵模型的 QA 系统置信度评分算法[J]. 软件学报, 2005, 16(8): 1407-1414.

[23] Demner F D. Complex question answering based on a semantic domain model of clinical medicine[D]. OCLC's Experimental Thesis Catalog, College Park, Md.: University of Maryland (United States), 2006.

[24] Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a Knowledge Representation for Clinical Questions[C]//AMIA Annual Symposium Proceedings, American Medical Informatics Association. Bethesda, 2006: 359-363.

[25] 李凡, 韩胜菊, 张丹. 刺参疾病诊断推理机的构建[J]. 计算机应用与软件, 2012, 29(12): 211-213.

[26] 黄肇明, 钟诚, 黎小如. 基于 BP 神经网络的呼吸系统疾病诊断仿真研究[J]. 合肥工业大学学报: 自然科学版, 2012, 35(3): 347-349.

[27] 李树青, 徐侠, 许敏佳. 基于读者借阅二分网络的图书可推荐质量测度方法及个性化图书推荐服务[J]. 中国图书馆学报, 2013, 39(3): 83-95.

[28] 陈文琴, 陆君安, 梁佳. 疾病基因网络的二分图投影分析[J]. 复杂系统与复杂性科学, 2009, 6(1): 13-19.

[29] 李树青. 基于引文关键词加权共现技术的图情学科领域本体自动构建方法研究[J]. 情报学报, 2012, 31(4): 371-380.

[30] 李树青, 徐侠, 钱钢, 等. 基于振荡算法和领域本体的学术文献关键路径自动识别和可视化展示方法[J]. 情报学报, 2012, 31(7): 676-685.

（责任编辑　车　尧）