Physica A 459 (2016) 42–54

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Clinical-decision support based on medical literature: A complex network approach

Jingchi Jiang [a], Jichuan Zheng [b], Chao Zhao [a], Jia Su [a], Yi Guan [a,*], Qiubin Yu [c]

[a] School of Computer Science and Technology, Harbin Institute of Technology, Harbin 150001, China
[b] RICOH Software Research Center (Beijing) CO., LTD, Beijing 100044, China
[c] Medical Record Room, The 2nd Affiliated Hospital of Harbin Medical University, Harbin 150086, China

HIGHLIGHTS

- We constructed a medical literature network (MLN) based on retrieved literature.
- The MLN improves the relevance retrieval result for clinical-decision support.
- We also proposed a re-ranking model to sort all retrieved literature by relevance.
- Our clinical-decision method based on the MLN yields higher scores in TREC 2015.
- Our study results confirmed that the MLN can facilitate the investigation of CDS.

ARTICLE INFO

Article history:
Received 28 December 2015
Received in revised form 7 April 2016
Available online 28 April 2016

Keywords:
Small-world
Scale-free
Complex network
Medical literature network
Clinical decision support

ABSTRACT

In making clinical decisions, clinicians often review medical literature to ensure the reliability of diagnosis, test, and treatment because the medical literature can answer clinical questions and assist clinicians making clinical decisions. Therefore, finding the appropriate literature is a critical problem for clinical-decision support (CDS). First, the present study employs search engines to retrieve relevant literature about patient records. However, the result of the traditional method is usually unsatisfactory. To improve the relevance of the retrieval result, a medical literature network (MLN) based on these retrieved papers is constructed. Then, we show that this MLN has small-world and scale-free properties of a complex network. According to the structural characteristics of the MLN, we adopt two methods to further identify the potential relevant literature in addition to the retrieved literature. By integrating these potential papers into the MLN, a more comprehensive MLN is built to answer the question of actual patient records. Furthermore, we propose a re-ranking model to sort all papers by relevance. We experimentally find that the re-ranking model can improve the normalized discounted cumulative gain of the results. As participants of the Text Retrieval Conference 2015, our clinical-decision method based on the MLN also yields higher scores than the medians in most topics and achieves the best scores for topics: #11 and #12. These research results indicate that our study can be used to effectively assist clinicians in making clinical decisions, and the MLN can facilitate the investigation of CDS.

* Correspondence to: School of Computer Science and Technology, Harbin Institute of Technology, Comprehensive Building 803, China. Tel.: +86 186 8674 8550.
E-mail address: guanyi@hit.edu.cn (Y. Guan).

http://dx.doi.org/10.1016/j.physa.2016.04.026
0378-4371/

1. Introduction

With the advancing age of populations worldwide, people have paid more attention to health problems. Each year, the numbers of deaths caused by cardiovascular diseases (CVDs) and hypertension are estimated to be 17.5 million and 7.1 million, respectively [1]. A World Health Organization (WHO) report on CVDs shows that 80% of all CVD deaths are due to heart attacks and strokes, which represent approximately 31% of all deaths globally [2]. Moreover, yearly figures from the WHO revealed that a person dies of diabetes and diabetic complications about every 10s. Diabetes is directly responsible for 1.5 million deaths in 2012 and 89 million disability adjusted life years [3]. The population of diabetic adults is expected to reach 300 million by 2025 [4]. Finding an effective auxiliary means to assist clinicians making a more correct clinical decision has become a critical problem to upgrade the clinician knowledge and reduce mortality.

Clinical-decision support (CDS) [5–9] which is a health information technology, provides physicians, patients, and other health professionals with knowledge and person-specific information, that is intelligently filtered and retrieved at appropriate times, to enhance patient health and health care [10]. In making clinical decisions, clinicians often seek out medical literature on how to best care for their patients. Medical literature can answer the three most common generic clinical questions faced by clinicians on a daily basis: "What is the patient's diagnosis?", "What tests should the patient receive?" and "How should the patient be treated?". However, given the volume of existing literature and the rapid pace at which new research is published, locating the most relevant and timely information for a particular clinical need can be a daunting and time-consuming task [11].

The Text Retrieval Conference (TREC) 2015 CDS track [12], which is similar to the goal of TREC 2014 [13–15], is designed to retrieve relevant medical literature to answer generic clinical questions on 30 actual patient records. The patient record typically describes three types of challenging medical cases and consists of 10 records per type, including diagnosis, test and treatment. Each record mainly contains two sections: description (detailing the patient condition) and summary (extracting meaningful information from the description based on the experience of doctors). The corpus for the retrieval task is the Open Access Subset of PubMed Central (PMC) on January 21, 2014, which contains a total of 733,138 literature [11]. According to the summary of a patient record, its description, or both, participants are challenged to retrieve a ranked set of 1000 papers at most, which are likely to support the decision of a physician on appropriate patient care.

The important aspects of the CDS according to medical literature have been discussed, and many valuable research ideas have been proposed. Garcia-Gathrighta et al. [16] adopted the vector space model using term frequency-inverse document frequency similarity and a unigram language model with Jelinek–Mercer smoothing. Mourao et al. [17] proposed multiple information retrieval techniques: retrieval functions, re-ranking, query expansion and classification of medical articles. Xu et al. [18] demonstrated the efficiency of the Johns Hopkins University HAIRCUT retrieval engine using character n-grams as an indexing term. Choi et al. [19] proposed an external tagged knowledge-based query expansion method for relevance ranking. Moreover, a machine-learning classifier-based text categorization method was used for the task-specific ranking.

Although many experts and scholars have explored the issues of clinical decision in different perspectives and realized a series of achievements, the use of complex network approach to solve the clinical decision problem has not yet been studied. In the present study, we focus on helping clinicians make better clinical decisions by retrieving relevant medical literature. In summary, we conducted our investigation in the following manner:

(1) We proposed a method of building a medical literature network (MLN). Then, we further analyzed the topological structure and characteristics of the MLN, which refers to the features of a complex network.
(2) According to the MLN and some analytical methods of complex networks, we adopted two strategies to mine potential literature, which can also assist clinicians making clinical decisions in addition to the basic literature retrieval.
(3) Combining the relevance factor of a search engine with the structural factor of the MLN, we further proposed a re-ranking model to sort all retrieved literature.
(4) From the comparison with those of other participants in TREC 2015, we numerically found that our approach can better improve the normalized discounted cumulative gain (NDCG) indicator than the median scores in most topics.

The rest of this paper is organized as follows: in Section 2, we introduce the structural characteristics of a complex network and the MLN. In Section 3, the potential-literature-mining algorithm and re-ranking model are proposed on the basis of a complex network approach. In Section 4, we further evaluate the validity of the potential-literature-mining and the accuracy of the re-ranking model. Finally, we conclude this paper and discuss directions for future work in Section 5.

2. Construction of the MLN

2.1. Process of relevant literature retrieval

To search some indicative literature to help clinicians make clinical decisions, we first need to retrieve some relevant literature using a search engine. Some classical retrieval techniques are adopted, including index building, query construction and literature retrieval.

The corpus from the PMC is given as a set of XML files. Therefore, an XML parser is employed to extract the PMC IDs, keywords, titles, abstracts, full texts and references. If an abstract is not available, the conclusion section will be used as a