基于可满足性问题求解的辅助诊断系统的设计与实现

摘　要

近年来，“智慧医疗”概念逐渐兴起。本文主要针对智慧医疗中的辅助诊断场景，即机器可以根据患者的症状信息对其所患疾病自动地做出预测。现在市面上的辅助诊断产品采用的解决方案基本都是大数据 + 深度学习的方案，即采用深度学习的方法从大量的电子病历中学习出疾病和症状的关联模型，然后使用模型根据患者的症状信息对其所患疾病做出预测。这种方案的优势在于数据量大，几乎可以包含所有的常见疾病及症状。但是其问题也很明显，首先就是严重依赖病历，而病历可能无法覆盖罕见疾病。其次就是深度学习的预测结果缺乏可解释性。最后是数据的可靠性问题，病历中的数据可能存在错误，如何甄别这些错误是大数据方案无法解决的。

针对大数据方案存在的问题，本文提出了新的辅助诊断解决方案：首先从可信的医学资源 (比如医书、专业的医学知识库) 中抽取医学知识，并对医学知识进行审核验证保证可靠性。然后将医学知识编码成命题逻辑公式，组合这些逻辑公式构建可满足性问题的实例。最后采用 SMT 求解器去求解该实例，根据求解器的解来进行诊断。因为我们主要从医书等可信的知识源中抽取知识并且有审核过程，所以数据的可靠性和覆盖面都优于病历。而且因为我们从医书中提取包含因果关系的医学知识，并采用逻辑推理的方式模拟诊断过程，所以我们的诊断结果都有医书中的知识作为证据支撑，保证了诊断结果的可解释性。此外我们还提出了知识纠错的方案，可以对知识抽取过程中引入的错误进行纠正。

本文的主要贡献如下：

(1) 医学知识库的构建

为了实现本文的辅助诊断系统，我们首先构建了符合我们需求的

医学知识库。根据预先设计好的医学知识模型，从可信知识源中抽取医学知识，最终表达成医学知识图谱并存储到图数据库 Neo4j 中，供后续的算法使用。

(2) 辅助诊断内核

基于已经构建好的知识库，本文实现了辅助诊断的内核。辅助诊断内核可以根据患者的主述信息从知识库中提取相关的医学知识并编码成逻辑公式，然后组合这些逻辑公式构建可满足性问题的实例，最后使用 SMT 求解器求解其可满足性来进行诊断。

(3) 知识纠错方案

虽然我们从可信知识源抽取知识并进行审核，但由于人工或者机器自动抽取算法的缺陷，所以最终获取的知识还是可能存在错误。因为本文采用的是基于医学知识进行诊断而非大数据方案，错误的知识会很大程度影响最终诊断的准确性，所以必须对知识库中的错误知识进行纠正，本文就基于可满足性问题中的不可满足核 (unsatisfiable core) 技术，提出了知识库中矛盾知识的纠错方案。

(4) 实际应用的辅助诊断系统

本文的所有方案不仅仅是停留在理论原型层面，而是实现出了实际可用的辅助诊断系统，并且已经在合作医院中落地测试。

关键词：智慧医疗, 知识图谱, 知识纠错, 可满足性问题, SMT 求解器

DESIGN AND IMPLEMENTATION OF COMPUTER AIDED DIAGNOSIS SYSTEM BASED ON SATISFIABILITY PROBLEM SOLVING

ABSTRACT

In recent years, the concept of "intelligent healthcare" has gradually emerged. This paper focuses on the computer-aided diagnosis scenario in intelligent healthcare, that is when a computer can automatically diagnose disease based on patient's symptom information. For the time being, the issue of computer-aided diagnosis is solved with big data + deep learning method being involved, that is, deep learning solution is used to learn the association model of diseases and symptoms from a big data of electronic medical records, with further diagnosis based on the patient's symptoms. The distinctive feature of this method is that a large amount of data can contain almost all common diseases and symptoms. But the out coming issues are also very clear. First, this method relies heavily on medical records, which may not cover rare diseases. Second, the results received from the deep learning lack correct interpretation. And finally, there is a problem of data reliability as there may occur errors in the medical records. Unfortunately, big data cannot provide a solution to detect the errors from the correct data.

Focusing on the problems of big data solution, this paper proposes a new computer-aided diagnosis method: first, extract medical knowledge to create a valid knowledge base (KB) from trusted medical resources (such as medical books, professional medical knowledge bases), and verify such a knowledge base to ensure its reliability. Second, encode medical knowledge into propositional logical formulas, and combine these logical formulas to

build instances of satisfiability problem. Finally, the SMT solver is used to find plausible answers for these instances, and on the basis of the found solutions, diagnose a patient. Because we mainly form knowledge base from credible knowledge resources (such as medical books) and have a review process, the reliability and coverage of the received data are better than the one received from medical records. And due to the fact that the medical KB that we possess is based on causal relationships retrieved from the medical books and the fact that we use logical reasoning to simulate diagnosis procedure, our diagnosis results are supported by knowledge from the medical books, ensuring the possibility to interpret the diagnosis results correctly. In addition, we also propose a knowledge error correction scheme which can calibrate the results received in the process of knowledge extraction.

The main contributions of this paper are as follows:

(1) Creation of a medical knowledge base

In order to implement the computer-aided diagnosis system described in this paper, we first build a medical knowledge base that meets our requirements . According to the pre-designed medical knowledge model, the necessary data is extracted from a trusted knowledge source, then expressed as a medical knowledge graph and stored in the graph database Neo4j for subsequent algorithms usage.

(2) Computer-aided diagnosis core

Based on the well-built knowledge base, this paper implements the computer-aided diagnosis core. It can extract relevant medical knowledge from the knowledge base on the basis of the patient's symptoms and encode it into logical formulas. And it then combines these logical formulas to construct instances of the satisfiability problem. Finally, it uses the SMT solver to calibrate its satisfiability in order to perform diagnosis.

(3) Knowledge error correction scheme

Although we extract and review knowledge from trusted sources, due

to the disadvantages of manual or automatic extraction algorithms, there still may be errors in the final acquired knowledge. Because this paper uses medical knowledge-based diagnose methods rather than big data solution, erroneous knowledge will greatly affect the accuracy of the final diagnosis. Therefore, such an instance must be corrected. This paper is based on the unsatisfiable core of satisfiability problem to propose a way to correct the contradictory knowledge in the knowledge base.

(4) Computer-aided diagnosis system for practical application

All the schemes in this paper are not only theoretical, but are also implemented in computer-aided diagnosis systems and have been tested in cooperative hospitals.

KEY WORDS: Intelligent Healthcare, Knowledge Graph, Knowledge Error Correction, Satisfiability Problem, SMT Solver