



上海交通大学

Shanghai Jiao Tong University

学位论文开题报告登记表

Dissertation Proposal Form

学号/ Student ID:	117037910020
姓名/ Name:	丁如江
学生类别/Degree Program:	<input type="checkbox"/> 博士/Doctor <input type="checkbox"/> 学术硕士/Academic Master <input checked="" type="checkbox"/> 专业硕士/Professional Master
学习形式/Student Status:	<input checked="" type="checkbox"/> 全日制/Full-time <input type="checkbox"/> 非全日制/Part-time
导师/ Supervisor(s):	李国强
论文题目/ Dissertation title:	基于可满足性求解的辅助诊断内核的设计与实现
专业/ Major:	软件工程
学院/ School:	电子信息与电气工程学院
开题日期/ Date:	2019 年 1 月 7 日
开题地点/ Venue:	电信三号楼 318

填 报 说 明

Instruction

1. 开题报告为 A4 大小，于左侧装订成册。各栏空格不够时，请自行加页。

This form should be printed with A4 papers and bound together on the left. If the space left is not enough, please feel free to add extra pages.

2. 博士生导师可以根据博士生学位论文选题情况自行确定是否进行开题查新，博士学位论文开题查新报告应由查新工作站提供。

The supervisor should decide, based on the proposed topics, whether a novelty assessment report is needed or not, which should be conducted by an authorized novelty assessment department.

3. 开题报告通过后，分别由研究生、导师、学科各存档一份。

Upon passing the dissertation proposal, three copies of this form should be prepared, one for each of the doctoral candidate, the supervisor, and the academic discipline.

4. 医学院研究生如果以函评形式开题，开题地点请填写“函评”，专家组组长签名由导师签名。

For students in the School of Medicine, if the dissertation proposal is conducted via peer review, the “Proposal venue” shall be filled with “peer review” and the “Signature of Committee Chair” shall be signed by the supervisor.

<p>学位论文题目 Dissertation title</p>	<p>基于可满足性求解的辅助诊断内核的设计与实现</p>
<p>课题来源 Source of Project</p>	<p>国家自然科学基金</p>
<p>1、课题意义及国内外研究现状综述/Significance of the research topic and overview of its current status worldwide.</p> <p>课题意义：</p> <p>在当前我国医疗资源较为匮乏的情况下，一些简单的诊疗功能（分诊，常见病诊断）依然消耗了较多的医疗资源，比如很多三甲医院依然要处理例如感冒发烧等常见病的诊疗任务，这就造成了高端医疗资源的浪费。因此智能诊疗系统就显得尤为必要，如果可以通过机器智能批量的处理相对简单的诊疗任务，那么高端的医疗资源就可以集中去处理相对较难的诊疗任务。虽然现在已经有一些智能诊疗的系统出现，比如上世纪 70 年代的 mycin，科大讯飞出的导诊机器人，以及 IBM Watson 等，但是由于医学结构化数据的缺乏，医学知识的表示困难，以及基于深度学习人工智能可解释性的匮乏等原因，导致了现有的智能诊疗系统都难以实际应用。而且很多都是基于英文的诊疗系统，在我国难以适用。我们要做的就是一款中文辅助诊断内核，使用自然语言处理技术获取医学知识并结构化，然后将结构化医学知识表达成逻辑公式，通过组合逻辑公式构建可满足性问题的实例，最后通过求解该实例来完成诊断。虽然上世纪很多医疗专家系统已都采用基于逻辑的方式来进行诊断，但是由于结构化医学数据的缺乏导致了最终失败，而我们的内核会首先使用自然语言处理技术批量生成结构化的医学数据，这就弥补了之前专家系统的不足。同时由于我们的诊断内核是基于逻辑推理，而不是像现在很多智能治疗系统只是基于病历数据的统计，我们的诊疗内核可以对其诊断结果做出基于医学知识的解释，从而避免了缺乏可解释性的问题。</p> <p>国内外研究现状：</p> <p>1) 知识表示</p> <p>医学知识的结构化表示一直是业界难题，上世纪的专家系统大部分采用产生式规则（也称为 IF-THEN 表示）的方式来表达知识，它表示一种条件-结果形式，是一种比较简单表示知识的方法。IF 后面部分描述了规则的先决条件，而 THEN 后面部分描述了规则的结论。规则表示方法主要用于描述知识和陈述各种过程知识之间的控制，及其相互作用的机制。这样的表示方法编写困难，检索匹配性能低下，逐渐被业界抛弃。现在主流的都是采用基于 RDF 三元组（实体，关系，实体）知识图谱的方式来表达知识，图谱的表示简洁，自动化程度和检索效率都较高。</p> <p>2) 诊疗专家系统</p> <p>其实我们要做的就是一款诊疗专家系统，不过实现方式和传统的专家系统有所区别。传统的专家系统是通过专家编写产生式规则来表达知识，然后通过患者的交互来获取事实，通过事实与产生式规则进行推理诊断。上世纪 70 年代斯坦福大学 mycin 系统就是诊疗专家系统就是典型代表，它主要用于帮助医生对住院的血液感染患者进行诊断和选用抗菌素类药物进行治疗，但是由于通过专家手动编写产生式规则太过低效，所以 mycin 并没有太多的实际应用。而 IBM 的 Watson 主要用于为肿瘤患者提供治疗方案，由于给出治疗方案相比于做出诊断难度更高，需要从大量病历中学习出各式各样的治疗方案，Watson 在处理罕见病时会显得力不从心，最近也遭受到了业界的质疑。</p>	

3) 逻辑推理

早期的专家系统也有很多通过逻辑的方式来表达知识，虽然逻辑表达严谨，而且一阶谓词逻辑的表达能力也足以表达现有的医学知识，但是一阶逻辑的复杂度较高，其上的可满足性问题不可判定，这就导致了基于一阶谓词逻辑的知识表示在进行推理时性能较为低下。后来学术界又提出描述逻辑的概念，它的表达能力介于一阶谓词逻辑和命题逻辑之间，而且其可判定，这就相当于找到了一个不错的折中方案。但是后来的研究与实践却表明，描述逻辑在应用起来有诸多困难，直到现在也没有太多的应用。而我们的内核主要采用命题逻辑的方式，本质上是结合大量的数据与简单的逻辑推理来完成诊疗任务。虽然命题逻辑的表达力相对较弱，但是对于我们的诊疗内核已然足够。同时也由于其简单，使得自动化生成命题逻辑公式和在其上做推理相对容易高效。

2、课题研究目标、研究内容和拟解决的关键性问题 / **Research objectives, main contents and key issues to be solved**

研究目标：使用求解可满足性问题的方式来进行诊断

研究内容：

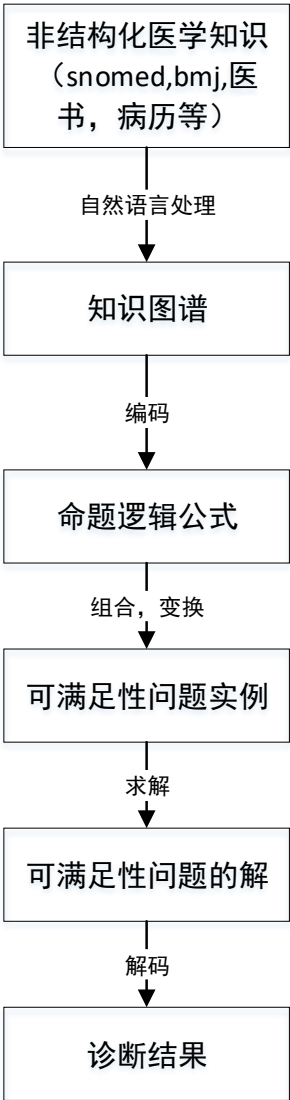
- 1) 将结构化的医学知识表达成命题逻辑公式
- 2) 命题逻辑公式构建可满足性问题实例
- 3) 将可满足性问题的解解码成诊断结果
- 4) 基于上述研究内容，设计并实现一款辅助诊断内核，可以完成基本的分诊，预诊功能
- 5) 对系统进行测试调优

拟解决的关键性问题：

- 1) 如何将结构化的数据有效合理的表达成命题逻辑公式
- 2) 如何根据可满足性问题实例的解进行诊断
- 3) 如何通过逻辑公式来找出数据中的错误

3、拟采取的研究方法、技术路线、试验方案及其可行性分析 / **Research methods, technical route, experimental scheme to be adopted and feasibility analysis.**

技术路线：



可行性分析：

- 1) 自然语言处理技术现在虽然有较多局限,但是提取我们内核辅助诊断所需的医学知识已然足够,我们只需要一些例如“疾病 A 导致症状 B”,“疾病 A 执行检查 B”类似的医学知识,现有的自然语言处理技术完全可以自动化提取出这些知识。
- 2) 将知识图谱内的医学知识表达成命题逻辑公式是一个难点,如何有效完备的表示也是我们课题的重点所在。
- 3) 命题逻辑的满足性问题虽然是 NP 完备问题,但是现有很多求解器 (minisat,z3) 对于绝大多数实例已经可以快速求解。
- 4) 将可满足性问题的解解码成为诊断结果其实是将结构化医学知识编码到逻辑公式的逆过程,只要我们解决了 2),这一步就不会成为难点。

4、课题的创新性/Novelties of the proposed topic.

- 1) 传统的专家系统采用专家手工编写规则表示知识，效率低下。而我们采用自然语言处理技术批量获取知识并结构化，可以使得知识表示的过程更为高效。
- 2) 现有的基于机器学习的专家系统的核心技术大多是深度学习，主要是基于大量病历数据的统计，但由于深度学习缺乏可解释性，这类专家系统在做出诊断后无法给出做出该诊断的原因，使得诊断结果难以使人信服，而且出错后也难以进行追踪调试。而我们的内核是基于逻辑推理，内部存储了大量的医学知识而非仅仅是病历上的经验，所以我们内核的可解释性也较强，可以给出做出诊断的原因。

5. 计划进度、预期成果/Research schedule, and expected outcomes

2019. 1-2019. 3	知识表示，逻辑推理，sat 求解器理论知识学习
2019. 4-2019. 7	根据已有的根据自然语言处理技术提出的医学知识，实现内核的初版，进行一些实验
2019. 8-2019. 10	获取更大量的数据，并根据上一阶段的实验结果对内核进行迭代。针对具体的情况分析，改进具体的算法
2019. 11-2019. 12	基于中期的研究，撰写学位论文，准备答辩

6、与本课题有关的工作积累、已有的研究工作成绩/Prior experience and accomplished achievements related to the proposed topic.

- 1) 根据 BMJ Best Practice 数据提取了近 5000 条三元组，构建了初步的医学知识图谱。
- 2) 根据现有的知识图谱实现了诊疗内核的第一个版本，实现了基本的分诊，预诊功能。

本人承诺：开题报告中的内容真实无误，若有不实，愿承担相应的责任和后果。I hereby declare and affirm that the details provided in this Form are valid and accurate. If anything untruthful found, I will bear the corresponding liabilities and consequences.

学生签字/Signature of Student: 丁如江

日期/Date: 2019. 1. 6