

引用格式:徐桂忠,李起成.问题诊断领域的智能对话系统研究[J].中国传媒大学学报(自然科学版),2023,30(06):44-49.
文章编号:1673-4793(2023)06-0044-06

问题诊断领域的智能对话系统研究

徐桂忠^{1,2*},李起成³

(1. 中国传媒大学信息与通信工程学院,北京 100024;2. 中国电子科技集团公司信息科学研究院,北京 100041;3. 南开大学计算机学院,天津 300350)

摘要:在人工智能领域,智能对话系统有着重要的研究意义和应用价值。智能对话系统是一种通过自然语言与人进行沟通交流的人机交互系统,它被定义为未来各种服务的入口。然而针对问题诊断领域的智能对话系统研究相对较少。本文设计并实现了应用于问题诊断领域的任务型智能对话系统,采用了管道方法(Pipeline Method)的架构设计了包括自然语言理解、对话管理、知识库和自然语言生成的对话流程,并提出了一套新颖的知识库模型和对话管理方法;基于问题症状和问题根因的二分图知识组织方式和基于概率分布的问题根因对话策略。该系统作为远程诊断的一个解决方案,利用人工智能打通医疗全科领域的关键技术点,形成可落地、可用、可工程化的全科辅助诊断对话系统。针对从公开出版物上随机选取的23种常见疾病的病例数据,本系统的召回率得到了较好的结果,其中前四召回率接近专业医生的诊断结果。本系统达到了缩短响应时间,降低成本,提升客户满意度的效果。

关键词:智能对话系统;对话管理;对话策略;二分图

中图分类号:TP391.1 **文献标识码:**A

Research on intelligent dialogue systems in the field of problem diagnosis

XU Guizhong^{1,2*}, LI Qicheng³

(1. School of Information and Communication Engineering, Communication University of China, Beijing 100024, China; 2. Information Science Academy of China Electronics Technology Group Corporation, Beijing 100041, China; 3. College of Computer Science, Nankai University, Tianjin 300350, China)

Abstract: In the field of AI, intelligent dialogue systems have important research significance and application value. Intelligent dialogue system is a human-computer interaction system that communicates with people through natural language, and is defined as the entrance to various future services. However, there is relatively little research on intelligent dialogue systems in the field of problem diagnosis. In this article a task-based intelligent dialogue system applied in the field of problem diagnosis was designed and implemented, using the Pipeline Method architecture to design a dialogue process that includes natural language understanding, dialogue management, knowledge base, and natural language generation. And a novel knowledge base model and dialogue management method were proposed: a bipartite graph knowledge organization method based on problem symptoms and root causes, and a problem root cause dialogue strategy based on probability distribution. This system, as a solution for remote diagnosis, utilized artificial intelligence to connect key technical points in the field of medical general practice, forming a practical, usable, and engineering oriented general practice auxiliary diagnostic dialogue

system. For the randomly selected case data of 23 common diseases from public publications, the recall rate of this system has achieved good results with the top four recall rates close to the diagnostic results of professional doctors, and this system achieves the effects of shortening response time, reducing costs, and improving customer satisfaction.

Keywords: intelligent dialogue systems; dialogue management; dialogue strategy; bipartite graph

1 引言

让机器具备与人交流的能力是人工智能领域的一项重要工作,同时也是一项极具挑战的任务。智能对话系统作为人工智能领域的核心技术,即将成为新的和谐人机交互方式,具有重大的研究意义和应用价值。

在学术界,过去很长一段时间关于对话系统的研究主要针对限定领域具体任务开展,是口语对话系统的重要分支。这些系统严格定义和限制接收的输入,并针对特定任务设计相应的规则、逻辑和回复语句。虽然这些方法在人机交互上取得了很大的进展,但受手工设定规则等方面的影响,其稳健性、可扩展性和领域适应性都有缺陷,也不适用于开放领域。近年来,随着互联网上社交数据的快速增长,数据驱动的开放领域对话系统逐渐成为了学术界关注的热点,深度学习的飞速进展,使多层神经网络得以在计算机视觉和语音识别领域取得突破性进展,深度学习的研究成果拓展到更广泛的研究领域,包括自然语言处理(Natural Language Processing, NLP)。人机对话系统作为服务的角色开始渗透到众多领域。

工业界近年来投入大量人力到对话系统的研发工作中。对话系统商业应用不断涌现,在限定领域对话系统中,以苹果Siri为代表的语音助手已通过手机被人们广泛使用;另外,以亚马逊Echo、百度Duer、谷歌Home和天猫精灵为代表的虚拟助手式智能音箱也已走进千家万户。在开放领域对话系统方面,微软开发了聊天机器人Xiaoice,IBM开发了聊天机器人Watson Assistant。但是针对问题诊断领域的智能对话系统的研究相对很少,本文将作者对问题诊断领域的智能对话系统的研究工作做具体阐述。

2 相关工作

对话系统按技术发展历史可以分为三种类型:基于规则和模版的对话系统;基于统计机器学习

的对话系统和基于深度学习的对话系统。基于规则和模版的对话系统主要依赖专家制定的人工规则,这种方法的优点是易于理解,但是需要大量的人力制定规则,所以很难跨领域而且扩展性非常差。基于统计机器学习的对话系统是利用数据挖掘的方法从数据中提取规则从而降低对话系统的手工成本,这种方法具有一定的学习能力,但仍然需要手工参与。近年来,随着深度学习的兴起,利用深度学习方法构造对话系统^[1]成为目前学术研究的热门方向,但目前完全基于深度学习方法非人工干预的对话系统只能应用于某些具体的场景。

针对不同的应用场景,对话系统可以分为两种类型^[2]:任务型对话系统(Task-Oriented Dialogue System)和非任务型对话系统(Non-Task-Oriented Dialogue Systems)。任务型对话系统面向垂直领域,通常是使用尽量少的对话轮数帮助客户完成特定任务,例如预定餐馆等。非任务型对话系统面向开放领域,通常为闲聊机器人,由于话题不确定,所以对对话系统的知识库要求广泛。而实际系统的语料大多是通过爬取获得,导致非任务型对话系统难以保持对话上下文信息的逻辑性。因此,非任务型对话系统离真正意义上的商用还有距离。

对话系统按架构实现方法分为两类:管道方法(Pipeline Method)和端到端方法(End to End Method)。管道方法的各个模块独立、相互依赖且任务明确。管道方法架构如图1所示。

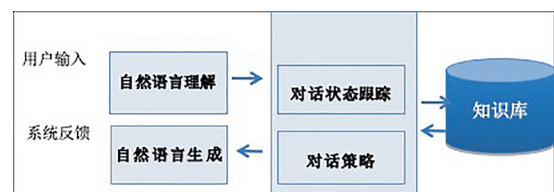


图1 管道方法架构图

端到端架构的模型一般使用Seq2Seq^[3],根据用户的输入直接生成相应的回复,这种架构简单直接便于扩展到更多领域。对特定领域的支持需要

辅以领域知识库。而对话过程中对知识库的查询检索一直是难点问题,如何根据对话的信息进行逻辑推理出知识库的查询语句需要将对话的自然语言转换为符号语言,由于自然语言理解中推理能力的限制,这个难题还没有得到很好的解决。目前的端到端架构的对话系统主要还是限定在开放域的对话问题。

本文的研究领域为问题诊断学(Diagnositics)^[4],最早的诊断开始于医疗诊断,随着人类文明的进步和科技的发展,诊断活动覆盖到越来越多的领域,包括企业诊断、环境诊断、设备诊断以及IT诊断^[5]等等。得益于互联网技术的迅速发展,基于互联网的远程诊断技术逐渐走向市场。尤其在疫情期间,人员往来受到限制的情况下远程诊断技术成为主要的问题诊断途径。而本文提到的智能对话系统是应用于远程诊断中的一个解决方案,利用智能对话系统以虚拟客服、微信聊天助手、机器人等多种形式提供服务,从而缩短响应时间,降低服务成本,提升客户满意度。

3 针对问题诊断领域的对话系统架构

首先针对问题诊断领域的对话系统是一个典型的任务型对话系统,这个系统的目的就是发现问题的根源,从而根据根因找到问题的解决方案。应用场景是该对话系统作为一个智能顾问辅助问题领域专家或者直接面对客户帮助解决具体的问题。此类对话系统面对具体的垂直领域,例如设备故障诊断、IT系统问题解决或者疾病诊断等。为了保证系统的扩展性,采用深度学习方法更好地理解用户的需求意图。因为问题诊断领域对准确性的要求,所以采用了管道方法(Pipeline Method)来构造对话系统,包括自然语言理解(Natural Language Understanding, NLU)、对话管理(Dialogue Management, DM)、知识库(Knowledge Base, KB)和自然语言生成(Natural Language Generation, NLG)^[6]四部分组成。

其中自然语言理解部分是将用户的文本输入映射、识别到诊断领域中的意图识别和语义槽。诊断领域识别和意图识别本质上是一个文本分类问题,通过收集来的语料来训练分类模型。传统的机器学习模型有支持向量机(Support Vector Machine, SVM)、朴素贝叶斯模型(Naive Bayesian Model, NBM)和K近邻(K-Nearest Neighbors, KNN)等方法。现在兴起的深度学习模型有卷积神经网络(Convolutional

Neural Networks, CNN)模型^[7]、循环神经网络(Recurrent Neural Network, RNN)模型^[8]等。语义槽填充可以通过规则的方法或者机器学习模型来识别,机器学习的方法本质上是一个序列标注的问题,通常采用现行统计方法包括条件随机场(Conditional Random Field, CRF)模型^[9]等,深度学习的模型一般使用长短期记忆网络(Long Short-Term Memory, LSTM)模型。而本文的对话系统使用的是BERT(Bidirectional Encoder Representation from Transformers)模型^[10],BERT提供了强大的预训练模型,用户只需要使用少量的训练数据进行微调(Fine-Tune)就能应用在具体领域中。BERT的下游任务包括分类任务和序列标注任务,正对应自然语言理解中的诊断领域识别、意图识别以及语义槽填充的需求。自然语言理解的一个例子如表1所示。

表1 自然语言理解示例

用户输入	网关的一个接口一直闪黄灯
诊断领域	IT设备故障领域
用户意图	网关异常症状
语义槽	IT设备黄灯

自然语言生成部分是将对话管理(DM)输出的抽象表达转化为自然语言句子。目前自然语言生成的模型表现还不够好,因此本文的系统采用模版的方法来生成回复,保证问题诊断领域的准确性和专业性。

对话系统中的核心模块是对话管理和知识库,相当于整个对话系统的大脑。知识库的研究热点是知识图谱,知识图谱本质上是一种语义网络,其结点代表实体(Entity),边代表实体之间的各种语义关系,目前基于知识图谱的多轮问答应用比较成熟。对话管理的任务是根据自然语言理解模块的输入,考虑历史对话中的上下文信息,决定对话系统要采取的相应动作,例如确认、追问或者陈述等等。对话管理中有两个子模块,即对话状态跟踪和对话策略。对话状态跟踪是以当前的对话输入、前一轮的对话状态和相应的对话系统动作作为输入,计算出当前这一轮的对话状态。而对话策略模块是根据对话状态跟踪模块计算出的对话状态,在候选对话动作集中选择对话系统动作。目前对话管理模块的对话状态很难建模因为场景内容因素繁杂,同时对话策略准确性难以保障。

4 问题诊断领域的知识模型和对话管理模块

本文针对问题诊断领域的对话系统提出了一套新颖的知识库模型和对话管理模块:

(1)基于问题症状和问题根因的二分图知识组织方式;

(2)基于概率分布的问题根因对话策略。

故障问题的整个生命周期由以下四个过程组成:问题发现、问题诊断、根因分析、问题解决。而问题诊断是整个流程中最重要的一步,其任务是不

断发现问题的表现症状,然后基于症状集合来缩小排查问题的范围,直至最后可以定位问题的根本原因。可以看到问题诊断中涉及的两个主要概念是症状和根因,症状是问题的表象,根因是触发问题的隐藏根源,症状和根因有着复杂紧密的关联,每个根因可能引起几种症状,而一个症状又可能对应多个根因。把每种症状作为结点放在一侧,每个根因作为结点放在另一侧,如果症状和根因之间有联系就建一条边,这样就形成了如图2所示的一个二分图。

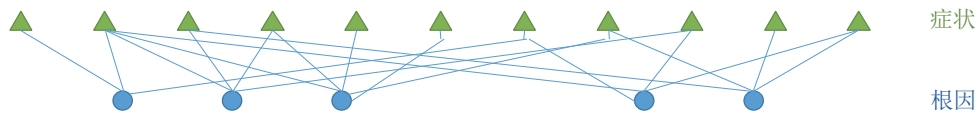


图2 症状和根因二分图结构

症状和根因二分图的构建过程可以基于历史的问题诊断记录实现自动化的构建。历史的问题诊断记录包含了问题的发现、诊断、根因分析以及解决方法整个过程,例如故障诊断过程中的工单或者疾病诊断过程中的病例。这些记录中都包含了症状和根因信息,如果症状和根因出现在同一记录之中就说明记录中的症状和根因之间有联系,根据症状和根因同时出现的频率则可以计算出症状和根因之间联

系的权重。对于每个症状,可以预设一个是否的问题来跟客户进行问答确认,通过用户的回答,系统可以获得有无该症状的答案。以计算机IT系统诊断为例,构建如图3所示的二分图。图3中第一列内容是预设的症状确认问题,第二列是症状确认问题的候选问题,第三列的结点是二分图的症状结点,第四列是二分图的根因结点,第三列和第四列之间的连线以及线上的权重表示症状与根因的联系以及概率。

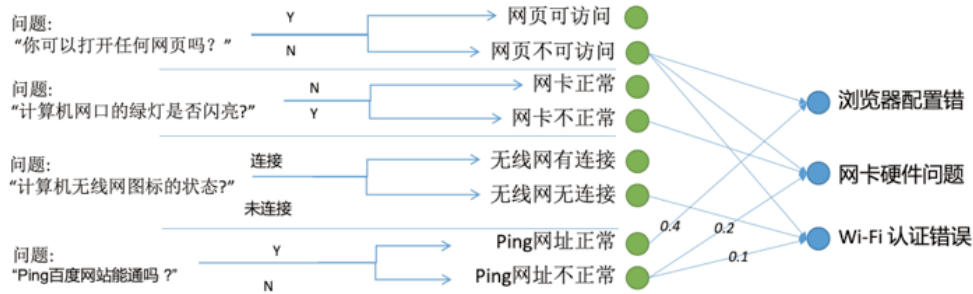


图3 计算机系统诊断二分图示例

问题诊断是一个排查的过程,通过逐渐明确问题的症状来缩小查找根因的范围,最后达到确定唯一的根因,而知道了根因就找到了解决问题的方法。问题诊断的对话系统实际上就是通过对话交互来不断明确问题的症状、最终确认根因的过程。然而在实际应用场景中症状的种类非常多,不可能对全部所有的症状都提问一遍。对话管理模块就需要充分利用二分图的结构实施对症状集合S的策略,使用尽可能少的对话轮数解决问题。因此对话管理模块

的对话策略就是如何选择下一个症状确认问题,采用信息增益的算法来实现对话策略,流程如下:

- (1)用户输入已知的症状集合 S_1 ;
- (2)根据症状和根因的二分图。找到已知的与症状集合有关联的根因集合 R ,并利用根因的概率计算出根因集合的信息熵 E ;
- (3)根据症状和根因的二分图,从根因集合 R 找到有关联的症状 S_2 ,用 S_2 减去 S_1 就得到候选的症状集合 S ;

(4) 针对候选的症状集合 S 中的每一个症状 S_i 都可以找到对应的确认问题, 根据用户对确认问题的不同种回复答案, 系统利用二分图结构将根因集合 R 分成几个子集合 R_i ;

(5) 对每个症状对应的根因子集合 R_i 计算熵 E_i , 然后计算对应的信息增益 $E_\Delta = E - E_i$;

(6) 系统取最大 E_Δ 对应的症状的确认问题发给客户;

(7) 重复以上的 1 - 6 步骤直到明确到具体的根因。

可以看到上述流程的主要计算是在熵的计算, 包括根因集合的信息熵 E 和每个症状对应的子集合 R_i 的熵 E_i 。根因集合的信息熵 E 的计算如式(1):

$$E = -\sum_{i=1}^n P_i \log P_i \quad (1)$$

其中 n 表示根因的个数, P_i 表示每个根因的概率。

每个症状对应子集的熵计算如式(2):

$$E_i = \sum_{j=1}^m \frac{|R_j|}{|R|} \text{entropy}(R_j) = -\sum_{j=1}^m \frac{|R_j|}{|R|} P_j \log P_j \quad (2)$$

其中 R 是根因子集, R_j 是子集中的一个根因, P_j 表示每个根因的概率。

通过以上的对话策略已经可以实现最快的找到问题的根因, 但当症状特别多而且根因很复杂的情况下, 还是需要询问确认很多个症状才能最终找到根因。因此本文根据根因概率分布的情况进一步优化对话策略。当前对话时根因概率分布情况可以总结为三种情况:

(1) 唯一高概率的根因: 仅有一个根因的概率超过阈值时, 则确定该根因, 让用户采取对应的解决方案;

(2) 多个高概率的根因: 有多个根因的概率超过阈值时, 则仅针对超过阈值的根因进行下一步的对话策略;

(3) 没有高概率的根因: 没有任何根因的概率超过阈值时, 则对所有的根因采取下一步的对话策略。

对话系统推断出的根因有可能出错, 需要用户进行确认是否问题得到解决。当用户根据根因没有解决问题时, 反馈给对话系统。对话系统会从根因集合中删除这个根因, 重新开始对话流程直到用户解决问题。

至此本文设计实现了针对问题诊断的对话系统, 问题诊断的对话流程如图4所示。

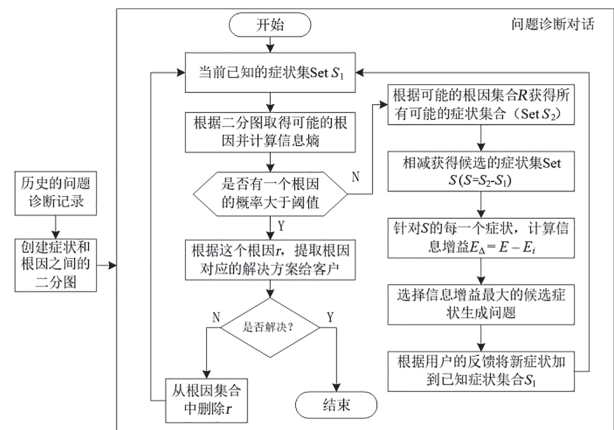


图4 问题诊断对话流程

5 实验结果

问题诊断领域最直接的应用就是疾病诊断。本文的方法是利用人工智能打通医疗全科领域的关键技术点, 形成可落地、可用、可工程化的全科辅助诊断对话系统, 包括了呼吸和消化科常见病共60种。对话系统的知识库采用了二分图的结构, 构造过程参考了权威诊断学书籍^[11]和数名全科医生的专家意见, 实现了症状与病因的关联以及权重。对话系统针对用户的初始输入, 利用自然语言理解技术支持主述、现病史、既往史、家族史、过敏史和检查报告。

实验方案是从公开出版物上随机选取了23种疾病案例, 分别由本文对话系统和专业医生进行问诊最后给出确诊疾病。本文的对话系统的平均对话轮数为9轮左右, 并获得如表2所示的准确率, 其中前四召回率接近专业医生的诊断结果。

表2 疾病案例的实验结果

		本文系统的诊断结果	专业医生的诊断结果
消化科疾病(12种)	第一召回率	41.70%	68.33%
	前四召回率	66.70%	77.88%
	未召回率	33.30%	22.12%
呼吸科疾病(11种)	第一召回率	54.50%	63.64%
	前四召回率	63.60%	72.73%
	未召回率	36.40%	23.27%

6 结论

本文针对问题诊断领域的对话系统提出了一套基于二分图结构的知识库模型的对话策略流程, 重点介绍了采用信息增益的算法来实现对话策略, 以及具体实现的架构和流程, 并根据本文的对话模型

实现和验证了疾病诊断领域的对话系统。

现实中,问题诊断的交互过程中常常混合使用各种类型的数据,即文本、语音、图像甚至增强现实等媒体数据,例如疾病诊断中的X光照片,设备诊断中的三维模型等数据。人工智能技术在语音、图像、文本的单一领域内的应用研究较多,但将这些技术综合应用的多模态技术还有待深入研究。为了将对话系统投入实际问题诊断使用中,发挥有效价值,建立基于多模态的问题诊断领域对话系统是当前工业界关注的方向。

参考文献 (References):

- [1] 陈晨,朱晴晴,严睿,等.基于深度学习的开放领域对话系统研究综述[J].计算机学报,2019,42(7):28.
- [2] 赵阳洋,王振宇,王佩,等.任务型对话系统研究综述[J].计算机学报,2020(10):1862-1896.
- [3] Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks [C]// NIPS'14: Proceedings of the 27th International Conference on Neural Information Processing Systems, 2014, 2:3104-3112.
- [4] 吴今培,肖健华.智能故障诊断与专家系统[M].北京:科学出版社,1997.
- [5] Liu R, Li Q, Li F, et al. Big data architecture for IT incident management [C]// Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics, 2014:424-429.
- [6] Salazar V L, Cabeza E M E, Peña J L C, et al. A case based reasoning model for multilingual language generation in dialogues [J]. Expert Systems with Applications, 2012, 39(8):7330-7337.
- [7] Kim Y. Convolutional neural networks for sentence classification [C]// Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2014:1746-1751.
- [8] Sordoni A, Galley M, Auli M, et al. A neural network approach to context-sensitive generation of conversational responses [C]// Proceedings of the 2015 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, 2015:196-205.
- [9] Ma X, Hovy E. End-to-end sequence labeling via bidirectional LSTM-CNNs-CRF [C]// Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2016:1064-1074.
- [10] Devlin J, Chang M W, Lee K, et al. BERT: pre-training of deep bidirectional transformers for language understanding [C]// Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), 2019:4171-4186.
- [11] 万学红,卢雪峰.诊断学[M].北京:人民卫生出版社, 2003.

编辑:赵志军