面向消化内科辅助诊疗的生成式对话系统

桑粉玲，翟卓玲，李婷婷

云南省第二人民医院消化内科 云南昆明

【摘要】随着经济的快速提升，人们的生活和工作压力逐渐增加，受社会环境和自身状况的影响，许多人们的生活规律都发生了较大的变化，尤其是饮食方面，多数人的饮食呈现出不规律和不健康的现象，这种情况下，消化系统疾病的发生率呈现出逐年上升的趋势。许多人发现健康出现异常时，第一反应不是入院就医，而是通过网络寻求答案。就消化内科疾病而言，其具有多样性的特点，仅仅依靠网络搜索，难以获取疾病的相关信息[1]。因此，为了提高患者获取信息的准确性，应该对传统的搜索方式进行改进。在信息检索系统中，对话系统相对较为高级，用户通过该系统将自己的需求输入，对话系统的另一方根据文字描述，将有效信息及时反馈，通过这种方法能够提高用户获取信息的准确性。本研究对消化内科中的生成式对话系统进行探索，利用向量机的学习功能和分类功能，从医学网站中将有关于消化内科的专业问诊语句进行获取，同时通过统计方法将消化内科中的各种知识制作成专业词典，并对分词效果进行改善和提升，在此基础上，将注意力机制加入多层结构和门控循环单元的结合方式中，并对训练方法进行加强，通过以上步骤，促进消化内科生成式对话系统的完善。通过研究表明，生成式对话系统与传统的对比方式，比较分词准确率，传统方式相对较低，由此可见生成式对话系统能够提高回答的准确性。

【关键词】消化内科；辅助诊疗；生成式对话系统

Generative dialogue system for auxiliary diagnosis and treatment of digestive medicine

Fenling Sang, Zhuoling Zhai, Tingting Li

Department of Gastroenterology, Second People's Hospital of Yunnan Province Kunming, Yunnan

【Abstract】With the rapid economic improvement, people's life and work pressure are gradually increasing. Affected by the social environment and their own conditions, many people's life laws have undergone major changes, especially in terms of diet, most people's diet appear irregular and unhealthy phenomena, in this case, the incidence of digestive system diseases shows a trend of increasing year by year. When many people find abnormal health, the first reaction is not to be admitted to the hospital, but to seek answers through the Internet. As far as gastroenterological diseases are concerned, it has the characteristics of diversity, and it is difficult to obtain disease-related information only by relying on network search [1]. Therefore, in order to improve the accuracy of information obtained by patients, the traditional search method should be improved. In the information retrieval system, the dialogue system is relatively advanced. The user enters his needs through the system. The other party of the dialogue system feeds back the effective information in time according to the text description. This method can improve the accuracy of the user's access to information. This study explores the generative dialogue system in Gastroenterology, using the learning function and classification function of the vector machine to obtain professional questioning statements about Gastroenterology from the medical website, and at the same time statistical methods Various knowledge is made into a professional dictionary, and the effect of word segmentation is improved. On this basis, the attention mechanism is added to the combination of multi-layer structure and gated circulation unit, and the training method is strengthened. Promote the improvement of the generative dialogue system of digestive medicine through the above steps. The research shows that the comparison between the

generative dialogue system and the traditional way compares the accuracy of word segmentation. The traditional method is relatively low, which shows that the generative dialogue system can improve the accuracy of the answer.

【**Keywords**】 Gastroenterology; Auxiliary Diagnosis and Treatment; Generative Dialogue System

就消化内科对话系统而言，其需要强大的数据支持，但是在公开的医学数据中，影响数据占比较大，关于问诊方面的数据少之又少[2]。此外，许多挖掘研究与医学语言方面的研究十分相似，首先就是分词，然而消化内科涉及的疾病类型较多，且不断的发现新疾病，其中涉及到较多内容，这种情况下，在对话式系统的构建中，常用的分词系统难以提供保障。

1 处理流程

消化内科想要构建出完善的生成式问诊系统，需要对其模板进行明确。语料获取和语料处理、构建词向量、对话等模块是主要消化内科生成式问诊系统的主要模块。

1.1 语料获取和语料处理

1.1.1 词库构建

在消化内科的文本中专业词汇、否定词汇、副词、方位词等是主要词汇表现，具有一定的特殊性。在分词的果从中使用结巴分词工具，无法有效的识别以上词汇，由此可见，该分词工具在消化内科中缺乏显著的分词效果，无论是对话模式的训练过程，还是建立词之间相关性的过程，均具有一定的要求，尤其准确率的要求较高。因此，为了提高准确率，需要将消化内科的词库自定义化处理，看分为专业和停用词库两方面。

通过当前医学网站获取的疾病、症状等相关名称，是消化内科词库的主要来源。其中涉及多个网站，比如：百度文库、寻医问药、有问必答等网站，同时还要通过医学字典获取的专业词语。

就停用词库而言，主要是指经常在问诊语料中出现的词汇，但这些词汇缺乏实用性。比如语气助词：请、谢谢等，还要部分副词，包括：的、了等。

1.1.2 预处理数据

文本的预处理是问题及分词操作前主要的环节之一。首先，要使用正确的方法进行文本中的副词部分，可将停用词库作为根据，对文本进行适当的处理。其次，在网络问答的过程中，多个标点符合的出现较为常见，需要对文本中重复标点去除。最后，消化内科的药品中，有部分药物不仅有别名，还有简称，此时需要对药名进行统计，来减少复杂性。

1.1.3 分词

将理解作为基础进行分词的方法、将规则作为基础进行分词的方法、将统计作为基础进行分词的方法，是鲜有的 3 中分词类型，为了能够构建完成的词典，本研究将结巴分词算法、逆向匹配分词算法作为基础，对将规则作为基础进行分词的方法、将统计作为基础进行分词的方法改进。

据相关研究表明，通过不同方法对句子进行切分，无论是正向切分，还是逆向切分，在比对的过程中发现两种切分方法存在较高的重合概率，和正确率，高达 90%，也就是说在 9%的概率中，必有一个正确结果。所以，本研究根据上述内容，将比较机制添加入问答和分词中，从而完成正、逆行切分，并进行两只两种切分结果的比较。如果比较中发现生词存在较大的差异，那么就选择少的部分为参考结果，若相同两种方法均可选择。

1.1.4 分类

消化内科的疾病类型较多，可通过系统分类的方法，将其分为五大类，比如胃肠病、肝病、胰胆病、内镜等。在模拟对话的过程中，需要将五大类型进行君合处理，以免出现拟合[3]。由于疾病的症状表现较多，在问答对中也可同时存在多种，而且不同类型之间影响不大，所以，在进行分类问题的求解中，要将多标签类型进行转化。

1.2 构建词向量

特征学习技术和语言建模在 Word 嵌入式自然语言理解的方式成为词向量，这也是自然语言通过计算机理解的一种方式。据相关研究表明，将这种词向量作为底层输入表示方式时，具有提高 NLP 任务的性能，尤其是情感、语言分析。为了能够明确词汇之间的相关性，需要在有限的数据集下进行获取。本研究词向量的选择包括 Word2Vec 向量、键值对向量。

1.2.1 键值对向量

统计分词后获取的初始语料，尤其是语料中词语的频次，要统计完全，按照由高至底的方法进行排序，并对合理的词语进行统计，并使用键值核对序号。对比选择完成的高频词[4]。若统计后的高频词和自定义词典中没有出现，此时在问答对中将未出现的词进行

增加，并对序号的范围进行调整。

1.2.2 Word2Vec

CBOW 及 Skip-Gram.CBOW 是 Word2Vce 的两种模型，对这两种模型的训练输入，其词向量是上线相关词语特征词的对应，一个特定词的向量称为输出。但在上下文的预测中，Skip-Gram.CBOW 起到了重要作用，然而在实际训练中，需要对事件序列进行充分的考虑。

1.3 对话模式

与机器翻译相比，生成对话场景与之有较多相似之处，两者都能对相同位置词的相关性进行翻译，同时两者还能够完成词的推算。在主模型不被改变的基础上，对神经元和编解码结构进行改变，在此果从中将注意力机制加入，能够获得良好的效果。就生成对话场景而言，其具有一定的特殊性，使用长短期记忆网络对其进行改变，并加入注意力机制能给个有效的减少对问题的依赖性。

表 1 分次评估结果（n/%）

指标	召回率	准确率	F 值
结巴分词	81.18	70.48	75.06
结巴+清华医学词库	83.37	73.77	78.15
结巴三加	98.56	97.26	97.87

2 讨论

自经济和科技高发发展以来，人们的生活质量得到了显著的提升，同时生活压力和工作压力的增加，许多人的生活规律和饮食规律均发生了巨大的变化，而且部分食品药品存在安全隐患问题，造成消化内科疾病的发生率逐渐上升。据相关数据显示，大多数人群习惯通过网络解答自身对疾病的疑惑，此类人群占总人数的 80%左右。然而在网络检索的过程中，往往需要通过关键词匹配，然而完成信息的获取，同时还可通过以往录入的内容中进行信息的匹配，以上方法存在一定的限制[5]。另外，不同的用户在语言表达方式上存在一定的差异，且许多网络信息存在一定的混乱现象，这种情况下，用户的部分意图无法在网络知识库中显示，从而导致用户的需求得不到满足。将自定义词典及前后选择机制融入到消化内科词汇中，能够有效的提高分词的准确率。在均衡数据的获取中，使用主动学习和支持向量结合的方法可以完成。通过以上方法的改进，能够促进对话系统的性能提升，同时为了使对话系统具有实用性效果，还需要相关人员对消化内科中的相关数据特征进行深入研究，从而获取新型的训练方法和模型结构。

做好饮食护理工作，消化系统疾病与饮食、生活习惯具有较大关系，良好的饮食习惯可以有效降低疾病发生率，对疾病恢复具有积极意义。日常饮食应该以清淡类食物为主，多食用高纤维类食物，可以有效促进胃肠蠕动，禁止服用辛辣、刺激性食物，从而减少对胃肠道的刺激。

参考文献

[1] 程梦卓.面向消化内科辅助诊疗的生成式对话系统研究[D].中国科学技术大学,2019.

[2] 王淑倩. 基于云平台的服务机器人个性化对话系统研究和设计[D].山东大学,2019.

[3] 陈晨,朱晴晴,严睿,柳军飞. 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(07):1439-1466.

[4] 曹东岩. 基于强化学习的开放领域聊天机器人对话生成算法[D].哈尔滨工业大学,2017.

[5] 吴承泽. 面向社区论坛多轮对话线索的连贯性评估方法研究[D].哈尔滨工业大学,2018.

收稿日期：2020 年 6 月 16 日
出刊日期：2020 年 7 月 17 日
引用本文：桑粉玲，翟卓玲，李婷婷，面向消化内科辅助诊疗的生成式对话系统[J]. 国际护理学研究, 2020, 2(4): 478-480.
DOI: 10.12208/j.ijnr.20200147
检索信息：中国知网、万方数据、Google Scholar