

基于信息抽取的食品安全事件自动问答系统方法研究

陈 瑛 张晓强 陈昂轩 赵筱钰 董玉博

(中国农业大学信息与电气工程学院, 北京 100083)

摘要: 针对从海量食品安全事件新闻报道中很难提取出所需答案的问题,以食品安全事件语料库为研究对象,提出了一种基于信息抽取技术的自动问答系统。首先,利用深度学习模型 TextCNN 对用户输入的问题进行分类,得到其所属类型。其次,对于输入问题,借助 Lucene 搜索引擎找到其最佳匹配文档。再次,根据输入问题的类型,从食品安全事件数据库(采用信息抽取技术自动提取的一个结构化数据库)中筛选出该文档所包含的答案候选句集合。最后,利用深度学习模型 Bi-LSTM 及基于答案候选句上下文的特征提取方法构建一个答案抽取模型,该模型能从给定的答案候选句集合中提取出最终答案。为检查基于食品安全事件数据库的答案候选句筛选方式及基于答案候选句上下文的特征提取方式对整个自动问答系统性能的影响,进行了多种比较实验,结果表明含有基于食品安全事件数据库的答案候选句筛选方式和基于答案候选句上下文的特征提取方式的问答系统表现最佳,其回答准确率达到 44%。这相比于传统的问答系统,具有明显的优势。

关键词: 食品安全事件; 问答系统; 问题分类; 信息抽取; 深度学习

中图分类号: TP182; TS201.6 **文献标识码:** A **文章编号:** 1000-1298(2020)S2-0442-07

QA System for Food Safety Events Based on Information Extraction

CHEN Ying ZHANG Xiaoqiang CHEN Angxuan ZHAO Xiaoyu DONG Yubo

(College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China)

Abstract: To solve the problem of extracting the answer for a question from massive food safety incident news reports, a question answering (QA) system was proposed. Firstly, a deep learning method TextCNN was used to classify the question provided by users. Secondly, a search engine method Lucene was used to find the best matching report for the question. Thirdly, based on a food safety event database (a structured knowledge base which was automatically constructed by using the information extraction technology) and the type of the input question, a set of answer sentence candidates was selected from the best matching report. Finally, based on the deep learning model of bi-directional long short-term memory (Bi-LSTM) and a feature extraction method which can extract effective information from the contexts of the answer candidate sentences, an answer extraction model was constructed, which can automatically extract the final answer from the given set of answer candidate sentences. To evaluate the impact of the selection method of answer candidate sentences based on the food safety event database and the feature extraction method based on the contexts of answer candidate sentences, different experiments were conducted. The results showed that the QA system using the structured knowledge database and the context-based feature extraction achieved the best performance (44% in accuracy), which significantly outperformed over traditional QA systems.

Key words: food safety events; QA system; question classification; information extraction; deep learning

0 引言

食品安全一直是国民关注的焦点问题^[1],随着

网络媒体的普及,网络上的有关食品安全事件的报道层出不穷^[2],这些报道纷繁复杂,不利于政府监管部门对食品安全事件的监管。因此,有效的信息检索

方法对于食品安全事件的监管具有重要的意义。

在目前食品安全事件研究领域中,数据主要来源于网络媒体报道的食品安全事件相关新闻。由于新闻报道众多、语句繁杂等原因,导致寻找问题答案效率低下^[3]。为此本文构建一个食品安全事件自动问答系统(Question answering system, QA 系统),其能根据用户输入的问题给出包含答案的相关语句。

问答系统是指用户输入以自然语言形式描述的提问,机器系统从大量的数据中查找出准确、简洁、人性化的回答,并反馈给用户^[4]。与搜索引擎系统相比,问答系统返回的答案包含语义信息,能更好地满足用户需求。文献[5]首次提出可以把问答系统看成一个翻译过程,并建立了 SMT (Statistical machine translate) QA 模型。近年来,计算机的飞速发展发展为基于神经网络的问答系统打下了坚实的基础。文献[6]提出了一个基于神经网络的 NRM (Neural responding machine) QA 模型。相比 SMT QA 模型,其在性能上有很大的提升^[7]。文献[8]针对大规模知识库问答系统的特点,提出采用别名词典结合 LSTM(长短期记忆)语言模型进行命名实体识别,再使用双向 LSTM 模型结合两种不同注意力机制进行实体消歧和答案选择的方法。文献[9]提出一种基于规则的关系词提取方法,将关系词与知识库中的谓词进行相似度计算,得到答案句的集合。文献[10]采用 BERT 结合 Bi-LSTM - CRF (双向长短期记忆-条件随机场)网络用于提取问句中的命名实体,最终使用阈值选择策略选取符合要求的答案集合,在减少人工干预的同时,保证了问答质量。文献[11]提出了一种端到端的无监督 KB - QA (知识库问答)框架,并使用动态规划算法支撑全局的优化与决策,解决了目前问答系统缺乏统一化的建模与全局优化策略的问题,尤其在解决多跳问题上有所表现,为问答系统的发展提供了新思路。目前这些研究大多是针对问答系统的整体架构或其中的某个模块(例如,问题分类、答案句抽取等)进行改进,对问题类型与答案抽取的有效结合缺乏相应的研究。另外,在答案句抽取时,没有对答案句的上下文信息进行有效的利用,限制了问答系统准确率的提升。因此,本文针对答案句抽取模块,探讨其如何有效结合问题类型信息并融入答案句的上下文信息。

根据问答系统处理的数据格式,可以将问答系统的发展分为基于答案对、基于自由文本数据、基于结构化数据库这 3 种类型^[12]。不同于之前的食品安全事件自动问答系统^[13],本文采用信息抽取技术

将基于自由文本数据的问答系统转化成基于结构化数据库的问答系统。本系统以爬取的食品安全事件相关新闻为语料基础,利用信息抽取技术构建一个食品安全事件结构化数据库,然后在该结构化数据库的基础上进行答案抽取。具体而言,在给定食品安全事件结构化数据库的基础上,本文的问答系统分 3 个步骤:问题分析、答案候选句集合提取和答案抽取^[14]。问题分析理解用户输入的问题,主要是对输入的问题进行分类;答案候选句集合提取是从食品安全事件数据库中抽取答案候选句集合;答案抽取是从答案候选句集合中抽取最终答案。本文结合深度学习技术和自然语言处理技术对食品安全事件自动问答系统中各个部分进行研究。

1 材料与方 法

1.1 总体架构

本文构建的食品安全事件自动问答系统主要包含 4 个部分:基于信息抽取技术的食品安全事件数据库构建、基于 TextCNN 的问题分类、基于食品安全事件数据库的答案候选句集合提取和基于 Bi-LSTM 的答案抽取。食品安全事件自动问答系统总体架构图如图 1 所示。

首先,基于信息抽取技术的食品安全事件数据库构建是采用自然语言处理技术中的信息抽取技术为本文的食品安全事件新闻语料库自动构建一个食品安全事件数据库^[15]。其次,基于 TextCNN 的问题分类^[16-19]根据用户输入的问题快速给出该问题所属的类型。再次,基于 Lucene 的信息检索系统^[20-22]从给定的语料库中抽取与输入问题匹配度最高的一篇文章,结合输入问题的类型和食品安全事件数据库,从该文档中提取出答案候选句集合。最后,基于 Bi-LSTM 的答案抽取^[23-26]从答案候选句集合中提取出最终答案。其中,基于 TextCNN 的问题分类、基于食品安全事件数据库的答案候选句集合提取及基于 Bi-LSTM 的答案抽取是本文的研究重点。

1.2 食品安全事件问答语料及食品安全事件数据库

1.2.1 食品安全事件问答语料的人工标注

本文采用之前的食品安全事件自动问答系统^[13]所使用的中文食品安全事件新闻语料库,其包含约 7 000 篇食品安全事件新闻文档。考虑到基于深度学习的模型训练需要大量的人工标注语料,本文对中文食品安全事件新闻语料库中的一些新闻做进一步的人工标注,增加数据类型和数据量。

因为在食品安全事件问答系统当中,用户输入的问题有很大一部分是与实体相关的,因此本文根据输入问题的相应答案的实体类型对输入问题进行

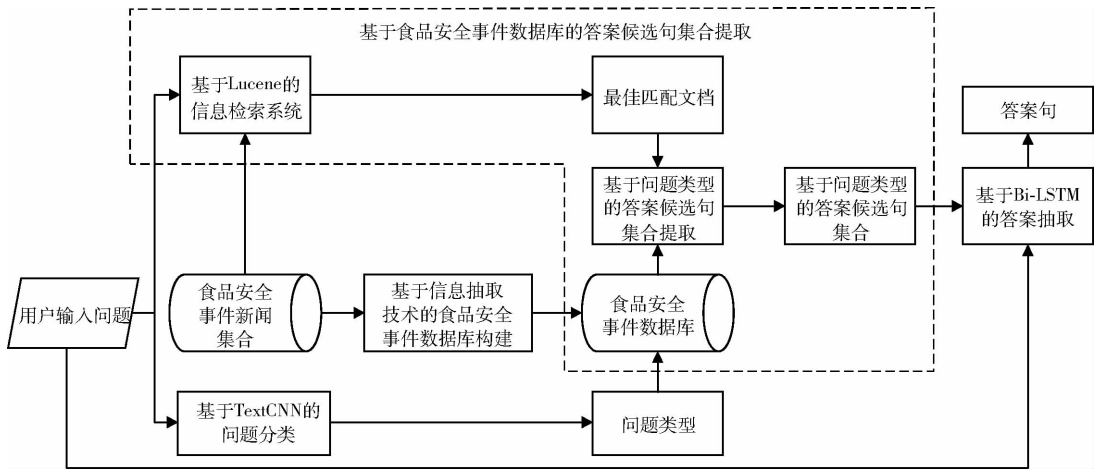


图1 问答系统总体架构

Fig.1 Overall architecture of question and answer system

分类。本文选择了时间、地点、责任主体、食品成分或种类、流通环节、事件原因、程度描述、应对措施这8种实体类型,并制定了相应的8种问题类型。经过对人工标注的约24 000条数据(问题-问题类型)进行统计,发现90%以上的问题隶属于时间、地点、责任主体、食品成分或种类。另外,其他问题类型(即不隶属于这4种问题类型)存在分类模糊、回答标准不清等问题。因此,本文在后续研究中,仅考虑时间、地点、责任主体、食品成分或种类这4种问题类型。另外,本文将之前的食品安全事件自动问答系统^[13]标注的300篇食品安全事件新闻文档扩展到872篇,并从中提取出1 500条问题-答案对,形成一个高质量的人工标注的食品安全事件问题答案对话料库。

1.2.2 基于信息抽取的食品安全事件数据库构建

根据本文所关注的4种实体类型(时间、地点、责任主体、食品成分或种类),采用自然语言处理技术中的信息抽取技术,针对中文食品安全事件新闻语料库中的每篇文档自动提取出其包含的这4种实体的实例,这些实例构成了一个结构化食品安全事件数据库。给定一篇食品安全事件新闻文档,该食品安全事件数据库提供其包含的各个实体实例,例如:食品名称,金枪鱼;地点,北京等。

1.3 基于TextCNN的问题分类

自动问答系统中的问题分析部分主要包括问题分类、关键词提取和关键词扩展,其中问题分类是其中的主要工作。问题分类^[27]处于自动问答系统对用户输入问题的语句处理的最前沿,其分类准确性对于后续最终答案的提取起着重要的作用。

在分类模型选择方面,因为CNN模型能够有效地提取文本的局部特征,所以本文采用了基于TextCNN的神经网络对用户输入的问题进行分类。TextCNN是快速高效的文本分类模型,其包括多个

卷积层和一个线性预测层。

在特征提取方面,本文采用词嵌入的表示方式^[28-30]对用户输入的问题进行向量表示,作为TextCNN的输入。另外,经统计显示本文所研究的4个问题类型通常会有一些常用的提示词(如表1所示)。比如,与地点相关的问题往往含有“哪里”,“什么地方”等提示词。因此,本文在输入问题的词向量表示后部拼接了一个one-hot向量,作为附加语义信息。该向量表示用户输入问题是否含有相关提示词。拼接形式为:[问题的词向量表示]||[0,0,1,0],其中,[0,0,1,0]表示输入问题中存在class=2,即存在与责任主体相关的提示词。

表1 提示词-问题类型的统计

Tab.1 Keywords statistics for question types

高频提示词序号	问题类型			
	时间	地点	责任主体	食品成分或种类
1	时候	哪里	品牌	成分
2	年份	地方	公司	东西
3	时间	哪儿	哪家	物质
4	何时	何地	企业	原料

1.4 基于食品安全事件数据库的答案候选句集合提取

1.4.1 基于Lucene的信息检索

类似之前的食品安全事件自动问答系统^[13],本文采用Lucene倒排索引检索技术构建一个搜索引擎,从而可以通过用户输入的问题在给定的海量候选文档库进行快速匹配查询。当把用户输入的问题传入基于Lucene的信息检索系统时,系统就会从中文食品安全事件新闻语料库中自动抽取与输入问题匹配度最高的候选文档。

1.4.2 基于问题类型的答案候选句集合提取

候选文档需要对其全文内容做进一步的筛选,

提取一个答案候选句集合,这些答案候选句会传递给后续的答案抽取模块进行处理。答案候选句的精准有效提取可以最大程度地缩小后续的答案抽取的工作范围,提高其准确率。

由于问题分类给出了其关注的实体类型,本文使用食品安全事件数据库进行答案候选句的提取。给定问题类型(比如“地点”),利用数据库提取出相对应的实体实例(比如“北京”)。这些提取出来的实体实例所在的句子构建成一个答案候选句集合。此外,对于每个候选句,加入其在原文档中的前后句,为后续的答案抽取提供更多的语义信息。

1.5 基于 Bi-LSTM 的答案抽取

类似之前的食品安全事件自动问答系统^[13],其采用 LSTM^[31-33]进行答案抽取,本文采用 Bi-LSTM 模型进行答案抽取。LSTM 是对循环神经网络(Recurrent neural network, RNN)的改进,引入了“细胞状态”、“遗忘门”、“输入门”和“输出门”的概念来进行状态的传递和控制,从而较好地解决了序列中的长期依赖问题。但是,在句子建模的过程中,LSTM 无法获取文本反序列的词依赖关系,而通过 Bi-LSTM 可以更好地捕捉这种双向的语义依赖。另外,本文采用词嵌入的表示方式对候选句及其前后句进行向量表示,作为答案抽取模型的输入。

2 实验结果与分析

2.1 实验设置

2.1.1 数据设定

本文的食品安全事件自动问答系统包含多个模块,其最终效果会受到各个模块的影响。为此,本文分别进行如下 3 个测试:基于 TextCNN 的问题分类测试、基于 Lucene 的信息检索系统测试及基于 Bi-LSTM 的答案抽取测试。

(1)基于 TextCNN 的问题分类测试:采用的数据集是人工标注的约 24 000 条问题-问题类型数据,按照 8:1:1 的比例将全部的数据分为训练集、交叉验证集和测试集。

(2)基于 Lucene 的信息检索系统测试:选取的测试集是人工标注的 192 篇文档,该测试集内容为问题-对应文档名称。

(3)基于 Bi-LSTM 的答案抽取(食品安全事件问答系统)测试:基于 Bi-LSTM 的答案抽取是整个食品安全事件问答系统的最后一个模块,所以基于 Bi-LSTM 的答案抽取测试也是食品安全事件问答系统的测试,代表了本文的食品安全事件问答系统的

最终效果。人工标注的食品安全事件问题答案对语料库提供了问题-答案对(1 500 条)。但是,答案抽取模块的训练还需要问题-非答案对,这些问题-非答案对可以由如下 2 种选取策略得到:全文抽取(即候选文档中所有句子)和数据库抽取(即使用基于问题类型的答案候选句集合提取方法抽取答案候选句)。最后,整个答案抽取数据集中包含答案的句子和不包含答案的句子控制在 1:1 左右。另外,按照 8:1:1 的比例将全部的数据分为训练集、交叉验证集和测试集。

2.1.2 评价标准

在实验过程中,本文对食品安全事件自动问答系统性能的评价主要采用精确率和 F_1 分数。精确率表征预测正确的样本在总体样本中的比例。 F_1 分数是统计学中用来衡量分类模型精度的一种指标,它可以同时兼顾分类模型的精确率和召回率,召回率表征人工标注的样本被系统预测正确的比例。以分类类别 i 为例,其相应的精确率、召回率和 F_1 分数的计算公式分别为

$$A = \frac{t_p}{t_p + f_p} \times 100\% \quad (1)$$

式中 A ——精确率

t_p ——两个集合(人工标注的类别是 i 的样本;系统预测的类别是 i 的样本)交集集中的样本数目

f_p ——两个集合(人工标注的类别不是 i 的样本;系统预测的类别是 i 的样本)交集集中的样本数目

$$B = \frac{t_p}{t_p + f_n} \times 100\% \quad (2)$$

式中 B ——召回率

f_n ——两个集合(人工标注的类别是 i 的样本;系统预测的类别不是 i 的样本)交集集中的样本数目

$$F_1 = \frac{2AB}{A+B} \quad (3)$$

2.2 实验结果

2.2.1 基于 TextCNN 的问题分类实验分析

本文对基于 TextCNN 的问题分类进行了相关测试,其评价效果如表 2 所示。其中, F_1 分数是 4 种问题类型各自的 F_1 分数的平均值;“问题”是把用户输入问题的词向量表示作为模型输入;“问题+提示词”把输入问题的词向量表示与相关提示的词向量进行拼接得到的向量作为模型输入。从表 2 可以得出,本文基于 TextCNN 的问题分类取得了非常好的分类效果。

表2 TextCNN 问题分类评价

Tab. 2 Evaluation of classification by TextCNN %

特征选取	评测指标	
	精确率	F_1 分数
问题	87.9	86.8
问题 + 提示词	92.5	94.0

另外,问题分类模型预测会不可避免地出现预测错误的情况。表3是分类模型的典型错例。前两句在客观上分类不清,导致模型错判;最后一句则是明显的人工标签错误,而模型预测是正确的。

表3 问题分类模型典型错例

Tab. 3 Typical error examples of TextCNN-based question classification model

问题	自动分类	人工标注
浙江哪儿的超市发生过食品霉变事件?	责任主体	地点
安徽泾县什么小学曾出现过食物中毒事件?	地点	责任主体
什么时候出现了毒蘑菇中毒事件?	时间	地点

表4 基于 Lucene 的信息检索系统典型错例

Tab. 4 Typical error examples of Lucene-based information retrieval system

问题	人工标注	人工标注	检索输出	检索输出
	文档编号	文档标题	文档编号	文档标题
2018年什么地方曾经查出假食盐?	41608	西安一黑窝点里查获15吨假食盐 母亲带儿造假	160144	成都10平米作坊查出4吨假盐
2018年什么地方曾查获过硫磺熏制的生姜?	41676	湖北宜昌查获近千公斤硫磺熏制生姜	41973	湖北宜昌工商部门近日查获近1吨用硫磺熏制的生姜

2.2.3 基于 Bi-LSTM 的答案抽取实验分析

本文根据是否使用上下文进行特征提取及是否使用数据库进行候选句抽取,构建了4组食品安全事件问答系统实验:候选句+上下文-数据库抽取、候选句+上下文-全文抽取、候选句-数据库抽取和候选句-全文抽取。其中,“全文抽取”和“数据库抽取”是答案候选句提取策略,前者是把候选文档中所有句子当作答案候选句,后者是使用食品安全事件数据库抽取的答案候选句;“候选句”和“候选句+上下文”是特征提取方式,前者是仅将候选句作为基于 Bi-LSTM 的答案抽取的输入,后者是将候选句及其前后句作为基于 Bi-LSTM 的答案抽取的输入。

针对问答系统的特性,效果评测仅考虑基于 Bi-LSTM 的答案抽取模型输出结果中的正样本的 F_1 分数。即,如果输出的正样本对应的答案句中包含人工标注的答案词,则可以认为系统回答正确,否则,视为回答错误。最终的实验结果:候选句+上下文-数据库抽取方式下系统回答准确率为44%,候选句+上下文-全文抽取方式下系统的回答准确

2.2.2 基于 Lucene 的信息检索系统实验分析

针对基于 Lucene 的信息检索系统,本文采用准确率进行评估:输入问题文本,若输出与测试集原文档相同,则为检索正确;否则,为检索错误。对192篇文档进行实验测试得到的结果是:检索正确的文档数为137篇,检索错误的文档数为55篇,检索准确率71.4%。

基于 Lucene 的信息检索系统不可避免会出现一些检索错误的情况,如表4所示。检索错误主要分为两种情况:

问题一:检索系统找到了相同类型的其他食品安全新闻报道(如西安和成都的两起售卖假盐事件)。

问题二:检索系统找到了对于同一食品安全事件的不同媒体的新闻报道(如均为湖北宜昌的生姜事件)。这种情况在本文采用的数据集中较为常见。因此,可以认为基于 Lucene 的信息检索系统的实际准确率应高于测试得到的准确率(71.4%)。

率为19%,候选句-数据库抽取方式下系统的回答准确率为42%,候选句-全文抽取方式下系统的回答准确率为25%。从最终实验结果可以得出,基于食品安全事件数据库的答案候选句集合提取可以有效提高食品安全事件问答系统的 F_1 分数。同时,上下文提供了丰富的语义信息,能够进一步提高食品安全事件问答系统的准确率。另外,候选句-全文抽取是之前的食品安全事件自动问答系统^[13]所采用的实验设置。最终实验结果表明本文采用的候选句+上下文-数据库抽取可以有效改进问答系统的效果。

3 结论

(1)在基于 TextCNN 的问题分类模型中,将相关提示词向量与问题词向量表示进行拼接融合的方法可以有效提高问题分类的准确率。

(2)在整个食品安全事件问答系统中,答案候选句集合提取策略的选择和答案抽取模型中的特征提取方式的选择对系统的总体性能有较大影响。其中,最优组合是基于食品安全事件数

数据库的答案候选句集合提取 + 基于上下文的特征提取。并且, 相比基于上下文的特征提取, 基于食品安全事件数据库的答案候选句集合提取对系统整体性能的影响更大, 问答系统准确率提

升更显著。

(3) 通过对问答系统各个模块的实验测试, 表明本文的食品安全事件问答系统能较好地满足从海量数据中抽取问题答案的需求。

参 考 文 献

- [1] 刘彦华. 2017 中国综合小康指数: 83.5 最受关注十大焦点问题: 租售并举成为新课题[J]. 小康, 2017(34): 62 - 67.
- [2] LIU P, MA L. Food scandals, media exposure, and citizens' safety concerns: a multilevel analysis across Chinese cities[J]. Food Policy, 2016, 63(9): 102 - 111.
- [3] 吉丽颖. 法经济学视角下中国食品安全事件的规制维度[J]. 食品与机械, 2018, 34(12): 63 - 66. JI Liying. Regulation dimension of food safety events in China from the perspective of economics of law[J]. Food Machinery, 2018, 34(12): 63 - 66. (in Chinese)
- [4] XU B, ZHUGE H. The influence of semantic link network on the ability of question-answering system[J]. Future Generation Computer Systems, 2020, 108: 1 - 14.
- [5] RITTER A, CHERRY C, DOLAN W B. Data-driven response generation in social media[C] // Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, 2011.
- [6] SUTSKEVER I, ORIOLE V. Sequence to sequence learning with neural networks[M] // Advances in neural information processing systems. Neur IPS, 2014.
- [7] 岳世峰, 林政, 王伟平, 等. 智能回复系统研究综述[J]. 信息安全学报, 2020, 5(1): 20 - 34. YUE Shifeng, LIN Zheng, WANG Weiping, et al. Research on intelligent reply system: a survey[J]. Journal of Cyber Security, 2020, 5(1): 20 - 34. (in Chinese)
- [8] 周博通, 孙承杰, 林磊, 等. 基于 LSTM 的大规模知识库自动问答[J]. 北京大学学报(自然科学版), 2018, 54(2): 286 - 292. ZHOU Botong, SUN Chengjie, LIN Lei, et al. LSTM based question answering for large scale knowledge base[J]. Acta Scientiarum Naturalium Universitatis Pekinensis, 2018, 54(2): 286 - 292. (in Chinese)
- [9] 张芳容, 杨青. 知识库问答系统中实体关系抽取方法研究[J]. 计算机工程与应用, 2020, 56(11): 219 - 224. ZHANG Fangrong, YANG Qing. Research on entity relation extraction method in knowledge-based question answering[J]. Computer Engineering and Applications, 2020, 56(11): 219 - 224. (in Chinese)
- [10] 吴天波, 刘露平, 罗晓东, 等. 基于弱依赖信息的知识库问答[J/OL]. 计算机工程: 1 - 8 [2020 - 08 - 17]. <https://doi.org/10.19678/j.issn.1000-3428.0058312>. WU Tianbo, LIU Luping, LUO Xiaodong, et al. Knowledge based question answering based on weak dependency information [J/OL]. Computer Engineering: 1 - 8 [2020 - 08 - 17]. <https://doi.org/10.19678/j.issn.1000-3428.0058312>. (in Chinese)
- [11] 王玥, 张日崇. 基于动态规划的知识库问答方法[J]. 郑州大学学报(理学版), 2019, 51(4): 37 - 42. WANG Yue, ZHANG Richong. Question answering over knowledge base with dynamic programming[J]. Journal of Zhengzhou University (Natural Science Edition), 2019, 51(4): 37 - 42. (in Chinese)
- [12] 毛先领, 李晓明. 问答系统研究综述[J]. 计算机科学与探索, 2012, 6(3): 193 - 207. MAO Xianling, LI Xiaoming. A survey on question and answering systems[J]. Journal of Frontiers of Computer Science & Technology, 2012, 6(3): 193 - 207. (in Chinese)
- [13] 陈瑛, 陈昂轩, 董玉博, 等. 基于 LSTM 的食品安全自动问答系统方法研究[J/OL]. 农业机械学报, 2019, 50(增刊): 380 - 384. CHEN Ying, CHEN Angxuan, DONG Yubo, et al. Methods of food safety question answering system based on LSTM[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(Supp.): 380 - 384. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?flag=1&file_no=2019s058&journal_id=jcsam. DOI: 10.6041/j.issn.1000-1298.2019.50.058. (in Chinese)
- [14] 刘里, 曾庆田. 自动问答系统研究综述[J]. 山东科技大学学报(自然科学版), 2007, 26(4): 73 - 76. LIU Li, ZENG Qingtian. An overview of automatic question and answering system[J]. Journal of Shandong University of Science and Technology (Natural Science), 2007, 26(4): 73 - 76. (in Chinese)
- [15] 刘敬培, 李江, 季文平, 等. 面向文本的事件信息抽取方法的研究[J]. 计算机与现代化, 2012(7): 198 - 201. LIU Jingpei, LI Jiang, JI Wenping, et al. Research on text-oriented event information extraction method[J]. Computer and Modernization, 2012(7): 198 - 201. (in Chinese)
- [16] 明建华, 胡创, 周建政, 等. 针对直播弹幕的 TextCNN 过滤模型[J/OL]. 计算机工程与应用, <http://kns.cnki.net/kcms/detail/11.2127.TP.20200110.1720.012.html>.
- [17] GUO Bao, ZHANG Chunxia, LIU Junmin, et al. Improving text classification with weighted word embeddings via a multi-channel TextCNN model[J]. Neurocomputing, 2019, 363: 366 - 374.
- [18] QIN X, PENG S, YANG X, et al. Deep learning based channel code recognition using TextCNN[C] // 2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). IEEE, 2019.
- [19] SUN X, MA X, NI Z, et al. A new LSTM network model combining TextCNN[M] // CHENG L, LEUNG A C S, OZAWA S, Neural information processing. Cham: Springer International Publishing, 2018: 416 - 424.
- [20] HIRSCH L, BRUNSDON T. A comparison of lucene search queries evolved as text classifiers[J]. Applied Artificial Intelligence, 2018, 32(7 - 10): 768 - 784.
- [21] MASNIZAH M. Development of search engines using lucene: an experience[J]. Procedia-Social and Behavioral Sciences, 2011, 18: 282 - 286.
- [22] 高琰, 谷士文, 谭立球, 等. 基于 Lucene 的搜索引擎设计与实现[J]. 计算机技术与发展, 2004, 14(10): 27 - 30. GAO Yan, GU Shiwen, TAN Liqiu, et al. Design and implementation of search engine based on Lucene[J]. Computer Technology and Development, 2004, 14(10): 27 - 30. (in Chinese)

- [23] MA X, HOVY E. End-to-end sequence labeling via bi-directional LSTM – CNNs – CRF[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, 2016: 1064 – 1074.
- [24] CROSS J, HUANG L. Incremental parsing with minimal features using bi-directional LSTM[C]//Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). Berlin, Germany: Association for Computational Linguistics, 2016: 32 – 37.
- [25] YAO Y, HUANG Z. Bi-directional LSTM recurrent neural network for Chinese word segmentation[M]//HIROSE A, OZAWA S, DOYA K, et al. Neural information processing. Cham: Springer International Publishing, 2016: 345 – 353.
- [26] RUI Z, RUQIANG Y, JINJIANG W, et al. Learning to monitor machine health with convolutional bi-directional LSTM networks[J]. Sensors, 2017, 17(2):273.
- [27] 张冲. 基于 Attention – Based LSTM 模型的文本分类技术的研究[D]. 南京:南京大学, 2016.
- [28] CHURCH K W. Word2Vec[J]. Natural Language Engineering, 2017, 23(1):155 – 162.
- [29] WOLF L, HANANI Y, BAR K, et al. Joint word2vec networks for bilingual semantic representations[J]. International Journal of Computational Linguistics and Applications, 2014, 5(1):27 – 42.
- [30] ZHANG D, XU H, SU Z, et al. Chinese comments sentiment classification based on word2vec and SVMperf[J]. Expert Systems with Applications, 2015, 42(4): 1857 – 1863.
- [31] GERS F A, SCHMIDHUBER J, CUMMINS F. Learning to forget: continual prediction with LSTM[J]. Neural Computation, 2000, 12(10):2451 – 2471.
- [32] SHI X, CHEN Z, WANG H, et al. Convolutional LSTM network: a machine learning approach for precipitation nowcasting [M]//CORTEZ C, LAWRENCE N D, LEE D D, et al. Advances in neural information processing systems 28. Curran Associates, Inc, 2015: 802 – 810.
- [33] GERS F A, SCHMIDHUBER E. LSTM recurrent networks learn simple context-free and context-sensitive languages[J]. IEEE Trans. Neural Netw., 2001, 12(6):1333 – 1340.

(上接第 322 页)

- [14] VAN P D, BARROS A, BÉRENGUER C, et al. Dynamic grouping maintenance with time limited opportunities [J]. Reliability Engineering and System Safety, 2013, 120(12): 51 – 59.
- [15] 苏春, 陈武. 基于滚动窗口方法的风力机动态机会维修优化[J]. 机械工程学报, 2014, 50(14): 62 – 68.
SU Chun, CHEN Wu. Dynamic opportunistic maintenance optimization for wind turbine system based on rolling horizon approach[J]. Journal of Mechanical Engineering, 2014, 50(14): 62 – 68. (in Chinese)
- [16] DO P, VU H C, BARROS A, et al. Maintenance grouping for multi-component systems with availability constraints and limited maintenance teams[J]. Reliability Engineering and System Safety, 2015, 142(10): 56 – 67.
- [17] HAI C V, DO P, BARROS A, et al. Maintenance planning and dynamic grouping for multi-component systems with positive and negative economic dependencies[J]. IMA Journal of management Mathematics, 2015, 23(2): 145 – 170.
- [18] GENG J, AZARIAN M, PECHT M. Opportunistic maintenance for multi-component systems considering structural dependence and economic dependence[J]. System Engineering and Electronics, 2015, 16(3): 493 – 501.
- [19] DO P, VU H C, BARROS A, et al. Opportunistic maintenance based on multi-dependent components of manufacturing system [J]. CIRP Annals-Manufacturing Technology, 2016, 65(1): 401 – 404.
- [20] HAI C V, DO P, BARROS A. A stationary grouping maintenance strategy using mean residual life and the Birnbaum importance measure for complex structures[J]. IEEE Transactions on Reliability, 2016, 65(1): 217 – 234.
- [21] 苏春, 陈武. 考虑部件经济相关性的风力机系统状态维修优化[J]. 东南大学学报(自然科学版), 2016, 46(5): 1007 – 1012.
SU Chun, CHEN Wu. Optimization of condition based maintenance for wind turbine system considering economic dependence among components[J]. Journal of Southeast University (Natural Science Edition), 2016, 46(5): 1007 – 1012. (in Chinese)
- [22] 张根保, 郭书恒. 基于竞争威布尔模型的加工中心可靠性评估[J]. 计算机集成制造系统, 2015, 21(1): 180 – 186.
ZHANG Genbao, GUO Shuheng. Reliability evaluation of machining tool center based on competing Weibull model[J]. Computer Integrated Manufacturing System, 2015, 21(1): 180 – 186. (in Chinese)
- [23] 魏泉, 岳高峰, 柳剑, 等. 基于改进竞争威布尔模型的联合收割机可靠性评估研究[J]. 农机化研究, 2019, 41(3): 13 – 20.
WEI Xiao, YUE Gaofeng, LIU Jian, et al. Research on reliability evaluation of combine harvester based on improved competitive Weibull model[J]. Journal of Agricultural Mechanization Research, 2019, 41(3): 13 – 20. (in Chinese)
- [24] 胡文泽, 何珂, 金诚谦, 等. 基于模糊综合评判的农业机械 FMECA 方法研究[J/OL]. 农业机械学报, 2018, 49(增刊): 332 – 337.
HU Wenzhe, HE Ke, JIN Chengqian, et al. FMECA method based on fuzzy comprehensive evaluation[J/OL]. Transactions of the Chinese Society for Agricultural Machinery, 2018, 49(Supp.): 332 – 337. http://www.j-csam.org/jcsam/ch/reader/view_abstract.aspx?file_no=2018s044&flag=1. DOI:10.6041/j.issn.1000-1298.2018.S0.044. (in Chinese)
- [25] 徐浩, 张韬, 李晓松, 等. Logistic 回归中不同 Pearson 残差估计方法的探讨[J]. 四川大学学报(医学版), 2015, 46(1): 129 – 132.
XU Hao, ZHANG Tao, LI Xiaosong, et al. Calculating Pearson residual in Logistic regressions: a comparison between SPSS and SAS[J]. Journal of Sichuan University (Medical Science Edition), 2015, 46(1): 129 – 132. (in Chinese)