

基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制研究*

吴石松¹ 林志达²

(1. 南方电网数字电网研究院有限公司, 广州 510000;

2. 中国南方电网有限责任公司, 广州 510663)

摘要: 针对传统的聊天机器人对话生成机制词识别率低的情况, 电网企业设计一种基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制。采用最大匹配分词算法对语义匹配, 由于中文领域中存在词语具有多种意思的情况, 根据语义匹配结果, 寻找对话中的相似词语, 对聊天机器人的对话关键词进行拓展。在此基础上, 利用 seq2 seq 模型对词向量编码和分解, 结合词向量编码和分解结果与语义匹配结果, 生成聊天机器人对话中的特征向量, 并利用 Attention 模型查找聊天对话中相似成分和相异成分, 根据相似度最高的问题对应的答案反应出去, 进行对话, 以此完成基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制的研究。实验对比结果表明, 此次设计的对话生成机制比传统的生成机制词的识别率高, 能够识别正确的语句, 保证在实际对话中做出正确的反应。

关键词: seq2 seq; Attention 模型; 聊天机器人; 对话; 生成机制; 拓展

中图分类号: TP391

文献标识码: A

DOI 编码: 10.14016/j.cnki.1001-9227.2020.07.186

Research on dialogue generation mechanism of chat robot based on Seq2 seq and Attention model

WU Shisong¹, LIN Zhida²

(1. Southern Power Grid Digital Grid Research Institute Co., Ltd., Guangzhou 510000, China;

2. China Southern Power Grid Co., Ltd., Guangzhou 510663, China)

Abstract: Aiming at the low recognition rate of words in traditional chat robot dialogue generation mechanism, a chat robot dialogue generation mechanism based on seq2 seq and Attention model is designed. According to the results of semantic matching, similar words in conversation can be found, and the key words of chat robot can be expanded. On this basis, seq2 seq model is used to encode and decompose word vectors. Combining word vector coding and decomposition results with semantic matching results, feature vectors in chat robot dialogue is generated, and Attention model is used to find similar and different components in chat dialogue, and corresponding questions according to the highest similarity are answered to dialogu. The dialogue generation mechanism of chat robot based on seq2 seq and Attention model is completed. The experimental results show that the designed dialogue generation mechanism has a higher recognition rate than the traditional generation mechanism, and can recognize the correct sentences to ensure the correct response in the actual dialogue.

Key words: seq2 seq; Attention model; chat robot; dialogue; generation mechanism; expansion

0 引言

目前, 市场上出现了不同类型的机器人, 其中客服机器人因有效降低客服人工成本被电网企业引入信息化建设规划中, 这些机器人的引入, 极大地帮助电网企业提高用户体验, 增加企业效益, 增强电网企业人工智能建设具有重要的作用。

其中聊天机器人是一种特殊的自动问答系统, 能够模仿人的语言习惯, 通过模式匹配的方式寻找问题的答案。在聊天过程中, 会出现各种各样的问句, 存在语句

包括各种词性和句中一些语义不同的情况。这种情况导致聊天机器人对于一些便于回答的问题, 能够准确回答, 对于一些不确定问题, 采用猜测和转移话题的方式回答问题, 导致不能给出正确的答案。基于深度学习的生成式聊天机器人算法首先对生成式聊天机器人领域主流框架进行实验, 在此基础上加入 attention 注意力模型机制来提高生成效果, 为了避免大量的无意义的回复, 保证对话的自然流畅。但是, 该传统聊天机器人存在词识别率低的情况, 对于对话不能做出正确的回应, 因此, 对聊天机器人对话生成机制进行研究, 提出了一种基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制, 使聊天机器人聊天对话更具有真实意义。

seq2 seq 模型也称为序列到序列模型, 主要由 2 部分构成, 包括编码器和解码器, 编码器对应的是问句, 解码器对应的是回复的语句。深度学习里的 Attention model 其实模拟的是人脑的注意力模型, 用来处理预序列相关

收稿日期: 2019-11-07

* 基金项目: 中国南方电网有限责任公司科技项目 (No.090000 KK52170124)

作者简介: 吴石松 (1986-), 男, 湖北松滋人, 硕士研究生, 中级工程师, 主要研究方向为人工智能。

的数据,每生成一个单词,找到源句子中与其相对应的单词,生成准确的翻译,具有直观、易于理解的特点。

此次设计的基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制主要从语义匹配、对话关键词扩展、词向量编码和语义相似计算 4 方面完成了聊天机器人对话生成机制的研究。实验对比结果表明,此次设计的聊天机器人对话生成机制比传统机制词识别率高,能够正确识别对话内容,以做出准确的回答。

1 语义匹配

电网企业采用最大匹配分词算法对语义匹配,首先建立一个中文词语较多的词典,将待匹配的语句正向扫描,从左到右,按照字符匹配规则查找待切分的字符串。如果找到对应的字符串,则将字符串作为成功切分的一个词,同时保证待切分词的长度^[2]最大,保证匹配性。在上述步骤完成后,进行下一步匹配,重复此操作,将所有待切分的语句全部切分完成,若匹配失败,更改当前的字符串^[3]重新查找,重复上述过程,直到所有的语句被切分完成,算法流程图如图 1 所示。

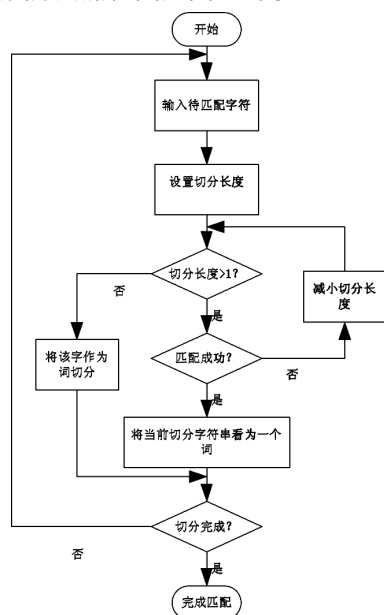


图1 最大匹配匹配算法流程

按照上述算法进行匹配,计算公式如下所示。

将待切分的字符串 V 设置为:

$$X = (x_1, x_2, x_3, \dots, x_n) \quad (1)$$

预先建立的字典 B 为:

$$B = (b_1, b_2, b_3, \dots, b_n) \quad (2)$$

公式(2)中,词典 B 中最大词语长度为 \maxlength , 并保证切分获得的词串长度尽量最大。在开始切分时,首先设置字符长度,设置为 \maxlength , 目的是在最大长度的切分下进行匹配。如果当前长度切分不成功,就减少切分长度,再进行匹配,直到匹配成功,为聊天机器人对话生成机制提供基础依据。

2 对话关键词拓展

在上述语义匹配的基础上,拓展聊天机器人^[4]对话

关键词。因为在中文领域^[5]中,一个词语会有不同的释义,这些同义词会影响聊天机器人的对话情况,因此对聊天机器人对话中的重要关键词进行扩展^[6]。

关键词拓展是指经过语义匹配后获得的词串,首先对关键词预处理,计算过程如下所示:

$$TF-IDF = \frac{V}{TF \times IDN} \quad (3)$$

公式(3)中, TF 为语句在整个词串中出现的频率, IDN 代表文档集中的文件频率^[7], V 代表经过分词取出后的词串。

假设聊天机器人 Ψ 共有 W 个对话数据,其中 w 各文档中含有词语 v_j ,词串 V 中共有 n 个词语,其中 n 个词语 v_j 则:

$$\begin{cases} TF = \frac{n}{W} \\ IDF = \log\left(\frac{W}{w}\right) \end{cases} \quad (4)$$

在上述计算完成后,取 $TF-IDF$ 值较大的前 k 个词进行下一步关键词拓展。

寻找目标词语中 x_i ,寻找与其语义相近的词语,将这些词语插入到字符串中 X 中。概念之间的相似度^[8]最终转化为:

$$Sim(S_i, S_j) = \frac{c}{dis+c} \quad (5)$$

公式(5)中, dis 为两个词汇的最短距离, c 为计算中的调节参数^[9]。

按照上述过程,完成聊天机器人对话关键词的扩展,为聊天对话生成提供基础。

3 制定聊天机器人对话生成机制

在上述语义匹配和对话关键词拓展的基础上,制定聊天机器人对话生成机制^[10]。在制定对话生成机制之前,首先对对话中的词向量编码^[11],利用 seq2 seq 模型对词向量编码和解码^[12],过程图如图 2 所示。

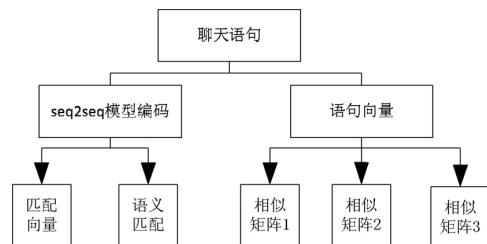


图2 词向量编码和解码过程图

按照上述词向量编码和解码过程对词汇编码和解码,具体步骤如下所示。

step1: 利用特定的词汇向量表示句子中的语句,其中每个语句向量维度一致,保证准确解决存在的词汇问题;

step2: 利用 seq2 seq 模型中的编码器使语句相互作用,使每个词都具有自身的特征向量,主要用来消除对话之间的差异性;

step3: 利用语义匹配向量,将对话中的词分解为 2 部分,解决词语之间的差异问题;

step4: 利用 seq2 seq 模型中解码器^[13],将回复语句

的相似和相异矩阵组合为一个特征向量;

step5: 最后结合 seq2 seq 模型 ,根据特征向量就算两个句子之间的相似度 ,特征向量越接近则代表 2 个语句的相似度越高。

将词向量编码和分解与第一步的语义匹配相结合 ,计算语句的特征向量 ,计算语句相似度过程如图 3 所示。

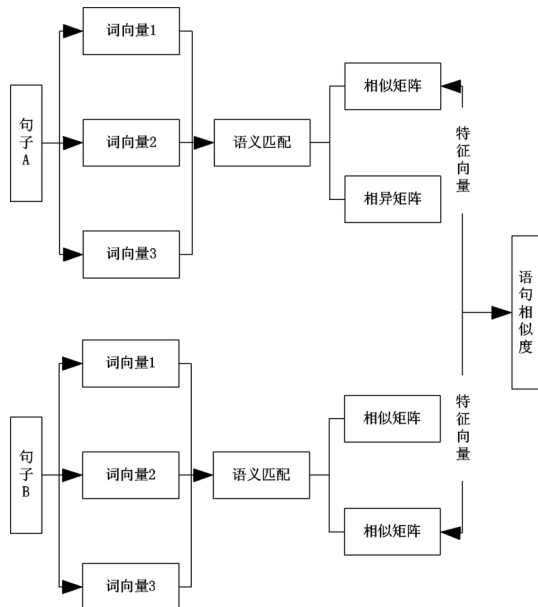


图 3 语义相似度计算过程

主要利用 Attention 模型^[14]处理机器人聊天数据 ,查找聊天对话中相似成分和相异成分 ,对语句进行更直观地解释。计算公式为:

$$S_e = \frac{f_{rig}}{s^+ \times s^-} \tag{6}$$

公式(6)中 s^+ 代表语句中的相似成分 s^- 代表语句中的相异成分 f_{rig} 代表语句中对应的特征向量。

经过上述步骤计算 ,利用特征向量^[15]求出语句之间的相似度 ,计算公式为:

$$\cosine(S, T) = \cos(\bar{S}, \bar{T}) \tag{7}$$

公式(7)中 S, T 代表语句中的特征向量 \bar{S}, \bar{T} 代表句子相似度 \cos 代表语句之间相似度的计算因子。

通过上述过程完成语句相似度的计算 ,根据计算结果计算目前语句和聊天机器人聊天库中哪个语句最相近 ,判断语句的相似度高低 ,输出相似度最高的语句 ,就可以实现聊天机器人对话 ,以此完成基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制的研究。

4 实验对比

4.1 实验环境

电网企业为证明此次设计的基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制的有效性 ,进行实验对比。此次实验使用 Google 开源的编程系统 Ehdsufi 实现实验设计 ,Ehdsufi 是一种基于数据流程的处理框架 ,主要表示运算节点之间的数据交互。具体的实验环境如表 1 所示。

表 1 实验环境

软件		硬件	
框架平台	Ehdsufi	处理器	2.7 GHz Core
运行系统	IOS	内存	8 GB
开发语言	Python3.7	显卡	Intel Iris Graphics 6100

同时 ,为了完成语言模型的创建 ,训练以及生成对话的进行 ,需要大量的聊天语句。因此 ,采用电影中的字幕文件 ,选择爬虫外文影片网站获取语料库 ,下载大量语言数据。将电影文件作为实验数据 ,实验过程中在利用两种对话机制对电影字幕文件识别。

4.2 实验方案设计

为了使此次实验更具说明性 ,将基于深度学习的聊天机器人对话机制与此次设计的基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制进行对比 ,主要对比两种方法的词识别率。词识别率评价是 2 个单词序列的不一致程度 ,将 2 个句子中一个作为模式序列 ,一个座位匹配序列 ,词正确率公式为:

$$Word = \frac{N - D - S - I}{N} \times 100\% \tag{8}$$

式中 N 代表词汇中的单词总数 D 代表删除的错误总数 S 代表单词的替代错误总数 I 代表插入错误总数。

通过词识别率判断 2 种方法的有效性 ,词识别率高才能保证对话的准确性。

4.3 实验结果分析

由于此次实验采用的实验语句为电影语句 ,与其他类型的实验采用的问题集不同 ,测试集与测试集之间不具备可比性。所以 ,采用评测法 ,处理这些实验数据 ,计算公式如下所示。

$$Q_i = \log \frac{d+h}{m_i \times x_i} \tag{9}$$

式中 m_i, x_i 代表分别代表数据中的问题集和回答集 d, h 分别代表实验中不同语句 \log 代表评测因子。

在上述实验数据处理完成的基础上 ,进行实验 ,实验结果如图 4 所示。

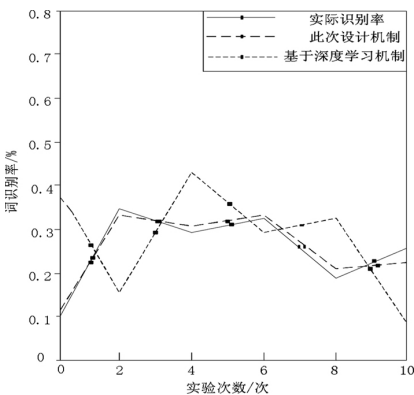


图 4 实验结果

分析上述实验结果 ,可以得到以下结论:

第一 ,按照统计结果生成的曲线 ,即使存在词识别率与实际识别率相差较大的情况 ,但是整体曲线的词识别率都与实际的识别率相近 ,都能够显示权重信息 ,说明能够在一定程度上提高正确语句的识别率;

第二:此次设计的实验数据都是随机选取的,具有一定的普遍性,不考虑语句的权重信息情况下,此次设计的聊天机器人对话生成机制识别率更高,与实际的词汇相差小,而基于深度学习的对话生成机制词识别率与实际识别率相差较大,识别不准确。

综上所述,此次设计的基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制比基于深度学习方法词识别率高,证明此次设计的聊天机器人对话生成方法对大多数问句都能够有效识别,具有一定的实际应用意义。

5 结束语

针对传统聊天机器人对话生成机制词识别率低的情况,电网企业设计了一种基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制。首先采用最大匹配分词算法对语义匹配,然后根据匹配结果对对话关键词进行拓展,在此基础上,应用 seq2 seq 和 Attention 模型对对话进行编码、解码和求解特征向量,最后计算语句之间的相似度,将与答案相似度大的语句输出,完成对话,以此完成了基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制的设计。实验对比结果表明,此次设计的基于 seq2 seq 和 Attention 模型的聊天机器人对话生成机制比传统机制词识别率高,能够为聊天机器人正确对话提供基础依据,具有一定的实际应用意义。

目前聊天机器人的深度技术仍处于发展初期,不管从技术和系统性能上看,都有很大的发展空间,系统在以后的研究工作中,实现一个更佳的聊天机器人。

参考文献

- [1] 孙立茹,余华云.基于深度学习的生成式聊天机器人算

法综述[J].电脑知识与技术,2018,14(23):227-228.

- [2] 宋皓宇,张伟男,刘挺.基于 DQN 的开放域多轮对话策略学习[J].中文信息学报,2018,32(7):104-113.
- [3] 赵宇晴,向阳.基于分层编码的深度增强学习对话生成[J].计算机应用,2017,37(10):2813-2818.
- [4] 马萍,宋颖慧.CAD/CAM 人机接口中对话结构生成器的研究[J].哈尔滨工业大学学报,2016,22(3):131-135.
- [5] 冯升.聊天机器人系统的对话理解研究与开发[D].北京邮电大学,2017,34(11):34-55.
- [6] 柳琳,罗军.基于 Android 的图灵聊天机器人设计[J].电脑知识与技术,2017,12(17):169-171.
- [7] 朱晶晶,韩立新.基于 RNN 句子编码器的聊天机器人[J].计算机与现代化,2018,23(1):32-35.
- [8] 戴开宇,张申生,王森.分布式虚拟环境中的聊天机器人的研究[J].计算机工程与应用,2017,38(7):13-16.
- [9] 蒲巧.双向交互机器人的语言自动生成仿真[J].计算机仿真,2019,36(4):310-314.
- [10] 许智超,胡祥培,李玉敏,等.电子商务订单拟人化实时处理的人机对话软件机器人研究[J].哈尔滨工业大学学报,2016,38(2):274-278.
- [11] 谢云,杨宜民.自主机器人足球系统的通信机制及通信语言设计[J].微电子学与计算机,2017,21(7):38-40.
- [12] 田行斌,张海名,黄真.对话型轮式移动机器人的开发[J].燕山大学学报,2016,27(3):269-271.
- [13] 刘超,冯东宇.智能问答的聊天机器人系统的设计与实现[J].信息技术,2017,45(5):176-177.
- [14] 赵冬梅,李雅,陶建华,顾明亮.基于协同过滤 Attention 机制的情感分析模型[J].中文信息学报,2018,32(8):128-134.
- [15] 汪敏.基于环境地图的多移动机器人协同机制研究[D].江苏:江苏科技大学,2017,45(1):34-44.

(上接第 185 页)

参考文献

- [1] 朱梦梦.直流电子式电压互感器测试方法研究与应用[J].云南电力技术,2018,46(02):69-72.
- [2] 李伟,樊博,陈勉舟,等.220 kV 电子式电压互感器带电校验系统研究[J].电测与仪表,2017,54(21):122-128.
- [3] 李振华,李春燕,张竹.基于 Q-ARMA 的电子式电压互感器误差状态预测[J].中国科学:技术科学,2018,48(12):1401-1412.
- [4] 沈鑫,束洪春,曹敏,等.基于法拉第磁光理论的光学电子电压互感器[J].电子器件,2019,42(02):286-290.
- [5] 刘翔,童悦,胡蓓,等.电子式电压互感器谐波准确度试验系统的建立[J].高电压技术,2018,44(03):835-840.
- [6] 鄧娇娇,吴小钊,王坤,等.10 kV 电子式电压互感器试验分析及探究[J].电气时代,2018,(4):105-108.
- [7] 杨桂平,郭毅.阻容分压型电子式电压互感器研究与设计[J].电力电容器与无功补偿,2018,39(03):96-101.
- [8] 刘罡,常源,江若嘉,等.110 kV 精密电压互感器主变比较准线路分析及不确定度评定[J].东北电力技术,

2019,40(01):34-37+62.

- [9] 曾光辉,薛峰,谢培成,等.电容式电压互感器谐波测量数据校正方法研究[J].黑龙江电力,2018,40(06):503-506.
- [10] 章鹿华,郭琨,段晓明,等.GIS 内电压互感器现场校验技术研究及设备研制[J].自动化仪表,2017,38(11):71-74.
- [11] 徐思恩,冯宇,汪本进,等.特高压电容式电压互感器误差在线同级比对技术研究[J].电力系统保护与控制,2018,46(03):133-137.
- [12] 石吉银.特高压变电站 TPY 级电流互感器暂态特性现场校核方法研究[J].陕西电力,2017,45(05):83-87.
- [13] 童悦,刘翔,胡蓓,等.电子式电流互感器谐波准确度整体校准试验研究[J].变压器,2018,55(12):33-38.
- [14] 齐小波,刘海波,程伟,等.电流互感器在线校准方案设计[J].中国计量,2018,(12):106-107.
- [15] 程昱舒,韩海安.基于北斗同步的特高压直流电流互感器现场校准装置的设计[J].电气技术,2018,19(09):65-68.