

基于句类模型的人机对话系统语句库设计

梁子琳^{1,2}, 丁润伟², 刘宏²

(1 山东大学威海分校 机电与信息工程学院, 威海 264209; 2 北京大学 信息工程学院, 深圳 518055)

摘要: 针对目前人机对话系统智能性较低、语句库领域受限的现状, 本文设计了面向日常生活常用话题的机器人回复语句库, 用于日常生活聊天. 语句库以汉语句子类型中的句类为模型, 构建了涵盖日常生活常用的 28 个话题范围的 900 句语句库. 该语句库通过关键词语的替换可以扩展出更多的语句, 从而用于更广泛领域的人机对话交流, 增加了语句库的完善性, 尽量避免出现了答非所问的情形. 同时本文的构建方法降低了人工构建语句库的工作量, 而且由于语句库较小, 确保了系统实现的实时性. 人机交互对比实验表明, 该语句库用于人机对话系统, 实现了使用较少的语句就基本可以进行日常话题的对话, 满意度高于对比网络聊天机器人小 i, 从而使得人机交互更加友好.

关键词: 人机对话; 聊天机器人; 句类; 语句库

中图分类号: TP391

文献标识码: A

文章编号: 1006-7043 (2006) xx-xxxx-x

Design of Human-computer Dialogue Statement Library System based on Sentence Category

Zilin Liang^{1,2}, Runwei Ding², Hong Liu²

(1 School of Mechanical, Electrical & Information Engineering, Shandong University at Weihai, Weihai, 264209;

2 School of information engineering, Peking University, Shenzhen, 518055, China)

Abstract: A statement library of common topics was constructed in this paper for the problems which the low intelligence of human-machine dialogue system and the statement library was limited in some special filed at present. The statement library based on sentence category of Chinese sentence types covered twenty-eight common topics, which was composed of 900 sentences. In addition, the statement library can generate much more sentences by taking place of keywords for much broad fields of human-machine interaction. The statement library was more complete, improved the intelligence of human-machine dialogue and reduced the workload of constructing statement library. The real-time can be achieved because the statement library was small. The experiment for human-machine interaction showed that the statement library presented in this paper is better than that of chat robot-little i. The statement library can improve the intelligence and real-time of a chat robot, so that made human-machine interaction more friendly.

Key words: Human-computer dialogue; chat robot; sentence category; statement library

随着网络的发展, 聊天机器人越来越受到人们的关注. 1950年, 图灵提出了著名的图灵测试^[1]: 如果一台计算机能像人一样对话, 它就能像人一样思考, 具有了人类智能. 最近几十年, 聊天机器人得到了飞速的发展. 1995年诞生的机器人Alice^[2], 三次通过图灵测试, 并一度被认为是最聪明的聊天机器人. 国内的聊天机器人发展也有长足的进步, 例如网络聊天机器人小i^[15], 腾讯的小q以及中国移动的飞信助手等等, 都是目前比较活跃的网络聊天机器人, 并且都得到了很好的应用. 聊天机器人是一种问答系统. 现有的中文问答系统在人机交互输出答案时主要有基于问句检索的方法^[3]和基于模式匹配的方法, 但基于模式匹配方法相对比较困难^[4].

聊天机器人的实现通常是把答句放入语句库中, 根据分析的输入语句的结果提取库中的语句作为回答, 所以语句库的完备性是决定聊天机器人回答贴切与否的一个重要因素. 目前大部分的聊天机器人都是面向特定专业领域的^[5-7], 用于日常生活交流的并不多, 而且通过人机交互实践发现很多的聊天机器人经常会答非所问, 其主要原因之一就在于语句库不够完备、包含的话题不够丰富. 语句库的构建是一项费时费力的工作, 而语句库越大, 机器人检索处理的时间越长, 时间复杂度越高, 进而会影响机器人的实时性, 降低了人们对机器人的满意度.

针对以上提到的问题我们设计了面向日常生活常用话题的人机对话语句库. 我们的语句库是由

收稿日期 2011-06-25.

作者简介 梁子琳(1988-), 女, 硕士研究生, E-mail: liangzilin@gmail.com.

基金项目 国家自然科学基金(No.60875050), 广东省自然科学基金(NO.9151806001000025), 深圳市科技计划及基础研究项目(JC200903160369A)

以句类为模型的900句句所组成,并且具有可替换性,可扩展到更广泛的领域,提高了机器人对话的智能性和实时性,使的人机交互更加友好.

1 句类模型

人工智能一个不可或缺的环节就是自然语言处理^[8-10],在自然语言的表述和处理模式方面,源于印欧语系的语法学和句法分析一直居于主导地位^[11].处理的主要内容主要包括:分词、词性标注等,但是分词的准确性有待提高,并且分词的错误会直接影响后面分析的准确性.句法分析一般就是分析句子成分或者是句子结构,例如动宾还是主谓,但是分析准确度仍然受限.范晓的《汉语的句子类型》一书中提出,汉语句子可以分为“三个平面”:句型.句模.句类^[12].

不同于以往都是根据句法结构或语义结构进行自然语言分析的方法,本文在构建聊天机器人语句库的时候创新性地采用句类进行句子分析.在《汉语的句子类型》一书中,根据句子的表达用途以及句子传达信息时的主题情形把句子分别分成了5类和4类,本文结合日常生活中聊天机器人的实际,把句子分成9类:叙述句、描记句、诠释句、评议句、祈使句、疑问句、选择比较、礼貌用语、口语.其中诠释句又分为判断句和领有^[12],评议句又分为个人看法、是非和赞同反对,祈使句又分为为祝福类、劝告类和请求类.

2 语句库

语句库是聊天机器人系统不可或缺的组成部分,本文的语句库的建立过程分为四步:

- 确定日常人机对话的话题分类.
- 建立基于句类模型和话题分类的人机对话语句库.
- 通过调研和测试数据分析得到语句使用频率.
- 确定900句机器人语句库.

2.1 确定机器人日常对话话题分类

本文所要建立的是面向日常生活中常用话题的机器人语句库.确定话题范围的资料来源主要有四个方面:《新版英语900句》^[13]、与网络机器人小i等的聊天、报纸杂志、现实聊天.从而确定了28个话题分类,每类话题下还有细分,基本上可以覆盖日常生活聊天常用内容.

2.2 语句库的建立

语句库的语句在构建过程中遵循以下原则:

a. 语句库中的语句少而精.在保证语句库完备的前提下,尽量避免语句库中语句的冗余,有效的减少语句数量,从而做到少而精.

b. 语句库中的语句具有可扩展性.语句库中的语句是以句类为模型得到的,这样可以通过语句中个别关键词的替换得到更多的语句,使得900句的语句库可以扩展成上千句.

c. 机器人回答更智能.要使与机器人聊天的过程更真实,就要使机器人的回答更智能、更巧妙,因此构建语句库的过程中我们加入反问句或者是一些可以转移话题的句子,进一步的明确用户所表达的意思,使得交互过程更加友好,机器人更加智能,回复更加多样化.

根据以上原则,结合之前总结出的28类话题,构建出聊天机器人用来回复日常生活常见话题的语句库.然后通过调研、人机对话实验,根据语句的使用频率和用户满意程度,最后确定出900句作为机器人语句库.

3 系统实现

在实现过程中,首先把语句库中的语句分成了3级:第一级为常用语句、第二级为规则语句、第三级为后备语句.

第一级常用语句主要是包括了常用问候类语句、礼貌用语以及口语化的语句.例如“你好”、“谢谢”等等.第二级规则语句是语句库的主体,包含了全部针对28类常用话题的回复语句.第三级后备语句主要包括了识别不出给出的回答语句.当由于输入语句不在28类话题范围内或者是输入处理错误而造成机器人识别不了,进而无法给出正确的回答时,就调用第三级后备语句.后备回复语句是比较通用型的语句,是对前面两级语句的一个补充.后备语句的回答可能不是最令人满意的回答,但是可以做到不答非所问,通过反问、主动提问、转移话题等方式进行进一步的交互,可以提高机器人的智能性,得到较高的满意度.

本文采用了基于关键字匹配的自然语言处理方法来验证所建立的语句库的完备性和回答语句的满意度,取得了较好的实验结果.系统流程首先是输入自然语言,匹配第一级语句,如果成功则返回库中语句输出,如果匹配失败则进入第二级匹配.在第二级匹配过程中,根据规则对输入语句进行多重匹配,这样可以提高回答的准确性,提高智能性.如果第二级匹配成功,则返回库中语句输出,如果匹配失败则进入第三级匹配,根据关键词匹配结果返回库中语句作为输出.

4 实验结果

为了证明语句库的可用性, 本文进行了对比试验, 对比对象为网络聊天机器人小*i*. 根据本文总结出的 28 类话题, 通过调研选取了 840 个问题作为输入语句, 每类话题 30 个问题, 然后对比针对同一问题机器人小*i* 以及本文构建的语句库的回答的满意度, 来判断语句库的好坏. 问题的话题范围具有完备性, 没有针对某一群体的倾向性. 满意度采取评分制进行量化: 1→5 分代表对回答语句的满意程度, 即不满意→满意. 而实验评分标准则依据回答语句的趣味性、切题度、实时性. 满意度得分来自 100 位被测试人员的调研评测. 样本按照随机抽样的方式在大学城范围内抽取.

小*i* 回答语句的获取来自小*i* 机器人的聊天网页^[14], 本文的聊天机器人命名为鹏鹏, 回答语句来自本文构建的语句库. 实验分为两组:

1. 通过本文所设计的人机对话系统得出的回答语句与小*i* 的回答语句的对比实验.
2. 鉴于本文的系统实现算法对输入语言的处理不够完善, 会影响鹏鹏的回答, 进而会影响语句库的评测, 所以第二组实验是假设输入处理完备的情况下, 人工选择输出语句. 并与前两组实验结果的对比. 第二组的输出语句记为鹏鹏 II.

实验结果如表 1 所示. 满意度取值为调研结果的均值.

表 1 满意度得分结果

	均值	方差
小 <i>i</i>	2.94	1.39
鹏鹏	3.02	1.49
鹏鹏 II	3.73	1.24

由表 1 可以看出, 鹏鹏的满意度与小*i* 的满意度略高一点, 但是鹏鹏回答的满意度波动比较大. 鹏鹏 II 的满意度比前两者有了显著提高, 方差鹏鹏 II 的相比小*i* 和鹏鹏的要低, 说明回答水平波动较小, 比较稳定. 由此可以看出本文所总结的语句库就日常生活聊天的常用话题来说, 总体是优于网络机器人小*i* 的. 下面我们对每类话题具体分析, 具体实验结果如表 2 所示. 从表 2 中可以看出以下几点:

- a. 在所有 28 类话题中鹏鹏 II 的回答一直优于小*i* 和鹏鹏的, 充分说明了语句库中的语句在这些话题的回答方面的优势.
- b. 在问候类话题中, 小*i* 的回答明显好于鹏鹏, 分析实验结果得出的主要原因是因为小*i* 有学习功能, 跟上千万网友聊过, 在问候方面语句库很丰富, 有很多很俏皮的语句, 所以满意度得分比较高.
- c. 在经济、军事、科技、教育、饮食、休息、

运动、医疗、交通、购物、穿着外貌、天气、颜色、智慧、动植物、自然等方面, 鹏鹏回答的满意度要略高于小*i* 的, 分析原因主要是因为这些方面比较专业, 很多输入小*i* 无法识别, 所以会答非所问, 而鹏鹏因为实现过程中只要关键词匹配准确就可以给出一个相应的回答, 所以满意度得分比小*i* 要高.

表 2 各类话题满意度得分

	小 <i>i</i>	鹏鹏	鹏鹏 II
问候	3.74	3.07	4.12
政治	2.88	2.71	3.41
经济	3.73	2.96	3.57
军事	2.47	2.73	3.35
科技	2.38	2.94	3.35
生活设施	2.59	2.70	3.41
教育	2.91	3.06	3.41
属性	3.16	2.89	4.10
饮食	3.14	3.45	3.98
休息	3.39	3.72	4.27
运动	3.03	3.18	3.91
健康	2.98	2.89	3.87
医疗	2.50	2.58	3.85
交通	2.55	2.95	4.03
购物	2.44	3.03	3.65
穿着外貌	2.88	3.18	3.93
工作职业	3.25	3.08	3.85
娱乐	3.28	3.09	3.77
电影音乐	3.29	2.76	3.34
电子数码	2.69	2.55	3.68
天气	3.09	3.49	3.81
颜色	3.51	3.57	3.83
语言	3.11	2.99	3.59
恋爱婚姻	3.50	3.22	4.05
智慧	3.03	3.18	3.64
动植物	2.69	2.90	3.73
货币	2.47	3.09	3.52
自然	2.79	2.85	3.43

d. 在政治、属性等剩余方面小*i* 回答的满意度要略高于鹏鹏, 分析原因首先是因为在这些方面小*i* 的语句库比较丰富, 而鹏鹏由于语句库语句比较少, 个别回答很单一, 所以满意度不如小*i* 的高. 其次是输入语言的随意性使得鹏鹏无法识别某些关键词, 所以做不出很满意的回答. 鹏鹏 II 的回答满意度高于小*i* 这一结果证明了该原因.

本文总结了面向日常生活常用 28 类话题, 并以句类为模型构建了人就对话常用语句库. 而且语句库具有良好的可扩展性, 可用于广泛领域的人

机对话领域，提高了语句库的完善性以及机器人的智能性，同时降低了人工构建语句库的工作量，并且确保了系统实现的实时性，使得人机交互更加友好。人机交互对比实验结果表明本文构建的语句库总体测评满意度较好，优于对比聊天机器人。由于系统实现过程采用了基于关键词匹配的算法，因此还需要改进对输入语句的自然语言处理算法，增加语义分析提高机器人的识别能力。同时输出语句的智能合成也是下一步需要研究的问题。

参考文献

- [1] A Turing. Computing Machinery and Intelligence [J]. Mind, 1950, 236(59):433-460.
- [2] R S. Wallace, The Anatomy of A.L.I.C.E, <http://www.alicebot.org/anatomy.html>.
- [3] R D. Burke, Kri J. Hammond, V. Kulyukin, et al. Question answering from frequently asked question files: experiences with the FAQ finder system[J]. AI Magazine, 1997, 18:57-66.
- [4] W. Lide, X. Huang, Y. Zhou, et al. FDUQA on TREC2003 QA Tasks[C] //The 12th Text Retrieval Conference, Gaithersburg, Gaithersburg: NIST,2003.
- [5] D. Bobbert, M. Wolska. Dialogue OS: an extensible platform for teaching spoken dialogue systems[C] // Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue. Trento. Ron Artstein and Laure Vieu, 2007:159-160.
- [6] Y. Xu, J. Li, S. Seneff. Mandarin Language Understanding in Dialogue Context[C] // International Symposium on Chinese Spoken Language Processing (ISCSLP 2008). Kunming: ISCA Archive, 2008:1-4.
- [7] G. Tur, D. Hakkani-Tur, L. Heck. What is left to be understood in ATIS[C] // Spoken Language Technology Workshop (SLT). California: Springer-Verlag, 2010:19-24.
- [8] J. Allen, Natural Language Understanding[M]. New Jersey: Addison-Wesley, 1995.
- [9] P J. Hayes and J G. Carbonell, A Tutorial on Techniques and Application for Natural Language Processing[M]. New Jersey: Carnegie-Mellon University, 1983.
- [10] A. Barnard. The Nursing Profession: Implications for AI and Natural Language Processing[C] //Natural Language Processing and Knowledge Engineering International Conference. Beijing: Springer-Verlag, 2007:497-501.
- [11] 黄曾阳. HNC理论概要[J]. 中文信息学报,1997,4:11-20.
- [12] 范晓. 汉语的句子类型[M]. 北京: 书海出版社, 1998.
- [13] 牛诚义. 新版英语900句[M]. 北京: 世界图书出版公司, 2009
- [14] 赢思软件. <http://i.xiaoi.com/>
- [15] 赢思软件. <http://www.xiaoi.com/>