软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2016,27(Suppl.(2)):69−75

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

对话意图及语音识别错误对交互体验的影响*

杨明浩, 高廷丽, 陶建华, 张大伟, 孙梦伊, 李 昊, 巢林林

(模式识别国家重点实验室(中国科学院 自动化研究所),北京 100190)

通讯作者: 杨明浩, E-mail: mhyang@nlpr.ia.ac.cn

摘 要: 在自然人机对话中,由于环境噪声、方言口音等因素带来的语音识别错误以及语义分析的不充分等原因,计算机在理解用户交互意图时出现偏差,使得计算机对要反馈的话题出现错误,造成人机对话进程的断裂.以面向咖啡为主题的漫谈式人机对话为例,将对话中断分为 3 种情况:话题反馈不当引起中断、话题正确情况下的模糊反馈不当和精确反馈不当引起中断.根据用户与计算机对话的记录分析比较上述 3 种情况下人机对话进程断裂情况.统计数据结果表明,话题反馈不当带来的对话中断最为明显,在对话进程断裂情况中达到了 60.1%的比例;在话题反馈正确情况下,模糊回答不当和精确回答不当带来的话题中断比例分别为 22.2%和 21.6%;在语音识别错误情况下,语义分析会带来数量更大的反馈错误.实验数据分析结果表明,在语音识别错误情况下,根据上下文信息提高计算机对用户话题反馈的准确率,能够有效降低人机对话的中断,提高人机对话的自然度.该工作为自然人机对话的意图分类重要性提供了数据分析和实验论证.

关键词: 意图分析;话题中断;语音识别

中文引用格式: 杨明浩,高廷丽,陶建华,张大伟,孙梦伊,李昊,巢林林.对话意图及语音识别错误对交互体验的影响.软件学报, 2016,27(Suppl.(2)):69−75. http://www.jos.org.cn/1000-9825/16021.htm

英文引用格式: Yang MH, Gao TL, Tao JH, Zhang DW, Sun MY, Li H, Chao LL. Error analysis of intention classification and speech recognition in human-computer dialog. Ruan Jian Xue Bao/Journal of Software, 2016,27(Suppl.(2)):69−75 (in Chinese). http://www.jos.org.cn/1000-9825/16021.htm

Error Analysis of Intention Classification and Speech Recognition in Human-Computer Dialog

YANG Ming-Hao, GAO Ting-Li, TAO Jian-Hua, ZHANG Da-Wei, SUN Meng-Yi, LI Hao, CHAO Lin-Lin

(National Laboratory of Pattern Recognition (Institute of Automation, The Chinese Academy of Sciences), Beijing 100190, China)

Abstract: In the natural human-computer dialogue system, environmental noises, accents and some other factors may cause the speech recognition errors which leads to computers' error responses to human. The dialogs are often interrupted by the system's bad responses. Three types of human computer interruptions are considered in this paper: improper feedback for topic, improper response for a vague user query, and improper feedback for an exact user query. According to the records of the user and computer dialogue analysis, the interruptions caused by three situations above are compared and used to analyze the importance of intention classification in human-

* 基金项目: 国家重点研发计划(2016YFB1001404); 国家高技术研究发展计划(863)(2015AA016305); 国家自然科学基金(61425017, 61403386, 61305003, 61332017, 61375027, 61273288, 61233009, 61203258); 中国科学院战略性先导科技专项(XDB02080006); 广西云计算与大数据协同创新中心、广西高校云计算与复杂系统重点实验室资助项目(YD16E11); 广西可信软件重点实验室研究课题(kx201601)

Foundation item: National Key Research & Development Plan of China (2016YFB1001404); National High-Tech R&D Program of China (863) (2015AA016305); National Natural Science Foundation of China (61425017, 61403386, 61305003, 61332017, 61375027, 61273288, 61233009, 61203258); Strategic Priority Research Program of the Chinese Academy of Sciences (XDB02080006); Program of Guangxi Cooperative Innovation Center of cloud computing and Big Data, Guangxi Colleges and Universities Key Laboratory of cloud computing and complex systems (YD16E11); Program of Guangxi Key Laboratory of Trusted Software (kx201601)

收稿时间: 2015-06-01; 采用时间: 2016-01-05

computer conversation. The statistical data find that the dialogue interruption caused by the inappropriate topic feedback is the most obvious problem, amounting to 60.1%. Under the correct feedback of the topic, the interrupt ratio of the subject caused by accurate answer and fuzzy answer is 22.2% and 21.6% respectively. In the case of error speech recognition, semantic analysis can bring more feedback error to the error of speech recognition. The analysis of experimental data shows that the speech recognition errors, can effectively reduce the man-machine conversation interrupt and improve the naturalness of human-computer dialogue system according to the context information to improve the accuracy of the computer on the topic of user feedback,. This paper provides the importance of intention classification in human machine dialogue, which helps to improve the performance of human-computer dialogue system.
Key words: intention analysis; topic interruption; speech recognition

人机对话的一个重要目标是实现人与计算机之间的自然交流.目前的大多数自然人机对话系统通常从用户输入中抽取关键信息进行意图判断,根据意图判断的结果,确定反馈话题,并从知识库抽取满足用户需求的答句返回给用户.因此,用户意图理解和准确的话题判断是人机对话中的重要环节.

最早期的用户对话目的理解出现在机票航班信息交互系统中[1].该项目侧重于建立旅游领域的口语意图理解系统,要处理用户的口语机票预订和查询语句.一些商用的针对领域的对话系统,如 Ultra Hal 系统[2],在处理查询意图理解过程中通过模糊匹配的方式来避开识别结果的某些错误,在意图理解可靠性较低的情况下采用主动对话,使谈话集中在上一个比较可靠的意图领域.美国人工智能大师 Richard 博士设计了 AI 系统,开发了聪明的爱丽丝(Alice)网络聊天机器人[3].该系统具有很强的学习能力,能够迅速从历史意图预测用户下一交互目的,并结合用户当前输入,准确判断用户意图.凭借着强大的意图理解功能,该系统于 2000 年和 2001 年两次通过了图灵测试.斯坦福大学的 CYC 自然语言理解系统的目标是通过知识工程的手段建立起和人一样的常识系统,并把这样的系统应用于对话系统中的意图理解中,以提升对话系统的可接受性.中国科学院的 K-Talk 少儿对话系统可以以自然的方式与一个学龄前儿童进行交谈,该系统通过对学龄前儿童心理特征和知识水平等特点对儿童的意图进行预测,交谈过程中能够准确获取用户的意图,因此具有较好的体验感.美国 DARPA 计划、欧盟框架计划、日本 JSPS 计划以及我国国家高技术研究发展计划(863)和国家自然科学基金长期以来均在用户意图分析领域设立了相关研究项目.

近年来,随着移动互联网的迅速崛起,自然人机对话系统在移动聊天平台和社交网络上有着巨大的用户群体,这类网络机器人通常能够很好地回答用户提问,然而在语音识别错误或用户输入语句比较模糊的情况下,会返回答非所问的句子,使得聊天中断.

计算机的反馈通常是针对对话主题展开的,如传统的基于填充槽[4−6]以及基于统计模型如 POMDPO[7,8]的对话管理方法中,聊天过程建立在对话状态(本文称为话题)的基础上.本文以面向咖啡为主题的漫谈式人机对话为例,将对话中断分为 3 种情况:话题反馈不当引起中断、话题正确情况下的模糊反馈不当和精确反馈不当引起中断.实验根据人机对话的聊天记录统计不同情况下的中断比例,数据分析结果表明:话题反馈不当带来的对话中断最为明显.另外,在语音识别错误情况下,语义分析会带来数量更大的反馈错误,分析结果表明,在语音识别错误情况下,根据上下文信息提高计算机对用户话题反馈的准确率,能够有效降低人机对话的中断,提高人机对话的自然度.

本文第 1 节首先介绍面向自然人机对话的多通道人机对话系统,该系统为本文实验提供了数据基础.第 2 节首先分析在语音识别正确的情况下,话题回答不当引起对话断裂、模糊回答不当引起对话断裂、精确回答不当引起对话断裂的情况;然后进一步分析在语音识别错误的情况下,同话题回答不当引起话题断裂、不同话题回答不当引起话题断裂的情况.第 3 节进行总结.

1 自然交互的多通道人机对话系统的框架

本文的实验数据来自多通道自然人机对话系统[9].该系统是一个面向实用的多模态自然人机语音交互对话模型,它采用有限状态转换图模型实现对话管理模型,将用户的多模态交互行为方式(包括用户的语音信息、情感信息和姿态信息)融合到多模态人机对话模型中.其系统结构如图 1 所示.其中,对话管理模块采用有限状态转

换图来实现.文献[9,10]介绍了基于以上框架设计的多模态智能人机对话系统.文献[9]的系统主要是针对交通路况信息查询进行的,这些查询大多属于精准信息问答.本文需要同时考虑模糊回答和精确回答的对话内容,因此,本文基于文献[9]的对话结构,重新设定了对话话题及话题跳转有限状态自动机,整个对话中的状态对应于有限自动机的 4 个节点:谈论咖啡、谈论茶饮、聊天气、漫谈.图 2 给出了人工智能咖啡厅多模态对话系统的状态转换图.其中,系统中语音识别采用百度语音识别 API 接口[11].后续的实验就在该对话管理模型下产生的日志文件上进行统计比较和分析.

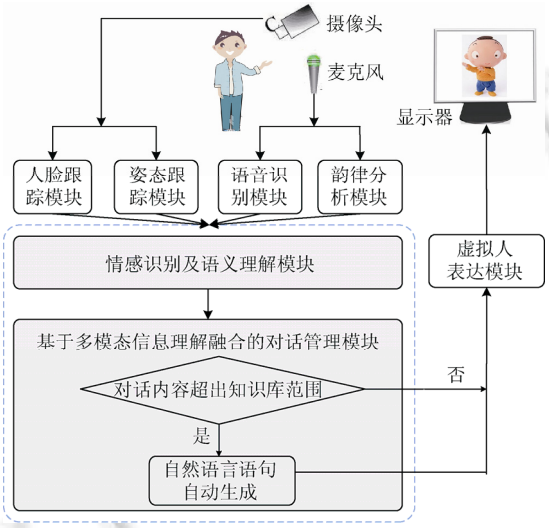


Fig.1 Framework of multi-modal human computer dialog

图 1 多通道人机对话系统框图

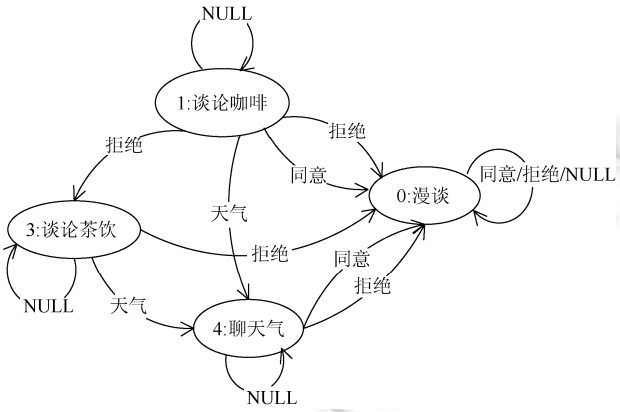


Fig.2 Dialog management for human computer dialog coffee shop

图 2 人机对话咖啡厅对话管理模型

2 实验分析

2.1 实验数据

实验中统计数据来源于自然交互的多通道人机对话系统[9,10]在如图 2 所示的对话管理模型下的对话日志.这些日志的时间跨度为 2014 年 12 月 23 日~2015 年 5 月 2 日,共 102 个对话片段,1 300 多条语句,所有对话都由

5 个标注者进行标注.表 1 是其中一个标注者的标注结果(其中,带删除线的黑体字表示语音识别错误;带下划线的黑体字表示模糊回答不当;加点的黑体字表示精确回答不当;其他字体表示其余对话过程).

Table 1 One of the labeled results for a piece of log file recorded in Sep.10 2014

表 1 一位标注者对 2014 年 9 月 10 日对话日志文件片段的标注

时间	角色	对话内容
[2014-9-10_14-4-19]	用户	我想喝咖啡,我想喝汽水
[2014-9-10_14-4-19]	系统	我们这目前只有速溶咖啡,您要来一杯吗?
[2014-9-10_14-4-39]	用户	在哪里?
[2014-9-10_14-4-36]	系统	就在不远的地方
[2014-9-24_14-4-46]	用户	你的领带是什么牌子的?
[2014-9-24_14-4-47]	系统	是雀巢的,您要来一杯吗?
[2014-9-24_14-4-56]	用户	马甲呢?
[2014-9-24_14-4-56]	系统	对不起,我无法回答这个问题.
[2014-9-24_14-5-6]	用户	不是很聪明吗?
[2014-9-24_14-5-6]	系统	谦虚使人进步, 骄傲使人落后
[2014-8-17_11-52-7]	用户	好的
[2014-8-17_11-52-32]	系统	切记咖啡不宜与茶同服.您还有什么想知道的吗?
[2014-8-17_11-52-32]	用户	嗯,不行
[2014-8-17_11-52-48]	系统	呵呵, 那就好.

2.2 实验结果分析

本文从 3 个方面对 5 位标注者的数据进行统计分析:话题反馈不当引起中断的比例、话题正确情况下的模糊反馈不当引起话题中断的比例、精确反馈不当情况下引起话题中断的比例.同时也统计语音识别错误的情况下同话题回答不当和回答不在话题内等情况.其中,

(1) 精确回答不当定义为:用户问题涉及某个具体的事物,系统回答与用户问题不相关或没有解决用户的问题.

(2) 模糊回答不当定义为:用户问题比较宽泛,不是具体的某个事物,系统回答与用户问题不相关或没有解决用户的问题.

(3) 话题反馈不当(回答不在同一话题)定义为:用户问题与系统回答属于不同的对话话题.

(4) 在语音识别错误的情况下,同话题回答不当定义为:用户的问句识别结果中存在语音识别错误,系统的反馈与用户问句在同一个话题内,但答案没有解决用户的问题.

(5) 在语音识别错误的情况下回答不在同一话题内定义为:用户的问句识别结果中存在语音识别错误,系统的反馈与用户问句不属于同一个话题的范畴.

表 2 给出了标注者标注类别的一些示例(其中,带删除线的黑体字表示语音识别错误;带下划线的黑斜体字表示当前语句所属的意图).

Table 2 Sampels for label categories

表 2 标注类别示例

类型	对话片段
精确回答不当示例	用户:请问有什么品牌的咖啡? (*意图: 咖啡*) 系统:好的,那我们一会儿接着聊! (*意图: 闲聊*)
非精确回答不当示例	用户:这一带有什么? (*意图: 闲聊周边信息*) 系统:好的,您准备好了我们继续玩游戏! (*意图: 闲聊游戏*)
话题反馈不当示例	用户:你的领带是什么牌子的? (*意图: 闲聊中精确问领带品牌*) 系统:是雀巢的,您要来一杯吗? (*意图: 咖啡*)
语音识别错误的情况下同话题回答不当示例	用户:不仅好喝(正确:不经常喝) (*意图: 咖啡*) 系统:其实喝咖啡还是有好处的,你知道它的好处吗? (*意图: 咖啡*)
语音识别错误的情况下回答不在同一话题内示例	用户:那些看着比较好(正确:哪些咖啡比较好) (*意图: 咖啡*) 系统:我们这里还有矿泉水,您要来一杯吗? (*意图: 谈论茶饮*)

根据对以上标注的定义,对所有 5 个标注者的日志文件进行了统计,图 3(a)、图 3(c)、图 3(e)、图 3(g)、图 3(i)分别列出在语音识别正确情况下,第 1 位~第 5 位标注者针对精确回答不当、模糊回答不当以及话题回答不当(回答不在同一话题内)的标注结果;图 3(b)、图 3(d)、图 3(f)、图 3(h)分别列出在语音识别错误情况下,第 1 位~第 5 位标注者针对同一话题内回答不当及回答不在话题内的标注结果,其中,“中断概率/次=中断百分比×是/Sum(是)”,表示当一次某类别回答不当发生时,所引起的对话中断的概率.

话题中断	精确回答不当	非精确回答不当	回答不在同一话题
总共	135	43	123
是	63	25	98
否	72	18	25
中断百分比	46.67%	58.14%	79.67%
中断概率/次	33.41%	7.81	41.98%

(a)

	同话题回答不当	回答不在话题内
总共	108	134
是	23	119
否	85	15
中断百分比	21.30%	88.81
总比	3.19%	74.42%

(b)

话题中断	精确回答不当	非精确回答不当	回答不在同一话题
总共	94	67	156
是	35	34	135
否	59	33	21
中断百分比	37.23%	50.75%	86.54%
中断概率/次	18.88%	10.21%	73.94%

(c)

	同话题回答不当	回答不在话题内
总共	128	101
是	23	90
否	105	11
中断百分比	17.97%	89.11%
总比	3.66%	70.97%

(d)

话题中断	精确回答不当	非精确回答不当	回答不在同一话题
总共	163	51	151
是	88	36	125
否	75	15	26
中断百分比	53.99%	70.58%	82.78%
中断概率/次	19.08%	82.78%	41.55%

(e)

	同话题回答不当	回答不在话题内
总共	142	119
是	25	113
否	117	6
中断百分比	17.61%	94.96%
总比	3.19%	77.76%

(f)

话题中断	精确回答不当	非精确回答不当	回答不在同一话题
总共	80	68	473
是	68	64	344
否	11	4	129
中断百分比	85.00%	94.12%	72.73%
中断概率/次	12.01%	12.60%	52.03%

(g)

	同话题回答不当	回答不在话题内
总共	307	8
是	195	7
否	112	1
中断百分比	60.26%	87.50%
总比	58.17%	3.03%

(h)

话题中断	精确回答不当	非精确回答不当	回答不在同一话题
总共	32	34	8
是	12	6	6
否	20	28	2
中断百分比	37.50%	17.60%	75.00%
中断次概率/	19.75%	4.40%	18.75%

(i)

Fig.3

图 3

图 4 给出了图 3(a)、图 3(c)、图 3(e)、图 3(g)和图 3(i)的综合分析结果,即通语音识别正确时精确回答不当、模糊回答不当以及话题回答不当(回答不在同一话题内)引起的对话中断概率,5 位标注者的概率均值.从图 3 和图 4 的统计结果可看出,在语音识别正确的情况下,当系统对用户意图理解正确时,精确回答不当引起话题断裂的平均比例为 21.6%,模糊回答不当引起话题断裂的平均比例为 22.2%,意图理解不对(回答不在同一话题或话题反馈错误)引起话题断裂的平均比例为 60.1%.可见,当语音识别正确时,无论是精确回答不当还是模糊回答不当,引起对话中断的概率都远远比用户意图理解错误时引发系统中断的概率要低.

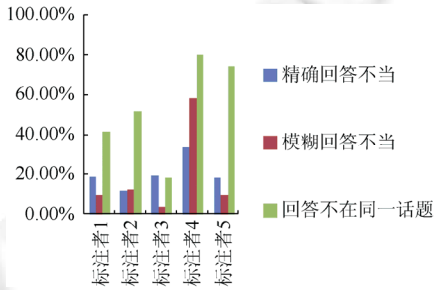


Fig.4 Dialog interrupt analysis without speechrecongnition error

图 4 语音识别正确时对话中断情况分析

图 5 给出了图 3(b)、图 3(d)、图 3(f)、图 3(h)的综合分析结果,即语音识别错误时,计算机同话题回答错误与计算机回答不在话题内引起话题中断概率,5 位标注者的概率均值.从图 5 统计结果可看出,在语音识别错误时,如果系统能够正确判断用户的意图(即,同话题回答不当),则语音识别出错引起话题断裂的平均比例为 20%,如果系统意图理解错误(回答不在用户提问的同一话题内),则语音识别错误引起话题断裂的平均比例为 55%.

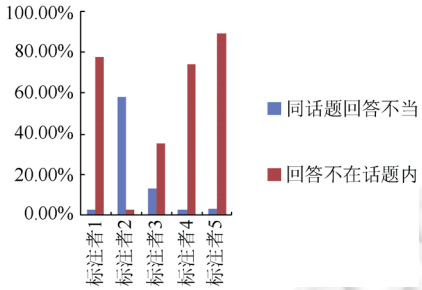


Fig.5 Dialog interrupt analysis on the cases of speech recongnition error

图 5 语音识别错误时对话中断情况分析

结合图 4 和图 5 的统计分析信息可知,意图或者话题反馈错误对系统中断的影响最大,比语音识别错误回答和精确回答错误的影响大得多.进一步来说,在语音识别错误的情况下,话题反馈正确会在很大程度上降低系统中断概率.因此,在对话系统中,如何正确地把不确定的话题转成正确话题并正确识别用户的意图,是提高自然人机对话体验的一个重要因素.

3 总 结

对自然交互的多通道人机对话系统中对话日志文件进行分析,实验结果表明,对话系统中用户意图理解正确在很大程度上决定了对话系统能否正常进行下去.即使系统中出现语音识别错误或其他错误,如果能够正确识别用户的意图,即话题反馈正确,那么绝大多数情况下对话都不会陷入中断.而当用户意图理解出错时,无论系统有无语音识别等其他错误,对话在很多情况下都会中断,这样会导致用户对话体验的下降.因此在对话系统中,正确识别用户意图,对用户的话题进行反馈,是提升用户人机对话体验的重要环节.

References:

[1] Zue V, Seneff S, Polifroni J, Phillips M, Pao C, Goodine D, Goddeau D, Glass J. PEGASUS: A spoken dialogue interface for on-line air travel planning. Speech Communication, 1994,15(3-4):331−340.

[2] http://www.zabaware.com/

[3] http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1

[4] Schwärzlery S, Maiery S, Schenk J, Wallhoff F, Rigoll G. Using graphical models for mixed-initiative dialog management systems with realtime policies. In: Proc. of the Annual Conf. of the Int'l Speech Communication Association—INTERSPEECH. 2009. 260−263.

[5] Pietquin O, Dutoit T. A probabilistic framework for dialog simulation and optimal strategy learning. IEEE Trans. on Audio, Speech, and Language Processing, 2006,14(2):589−599.

[6] Schatzmann J, Weilhammer K, Stuttle M, Young S. A survey of statistical user simulation techniques for reinforcement-learning of dialogue management strategies. Knowledge Engineering Review, 2006,21(2):97−126.

[7] Williams JD, Poupart P, Young P. Partially observable Markov decision processes with continuous observations for dialogue management. In: Proc. of the 6th SigDial Workshop on Discourse and Dialogue Lisbon. 2005.

[8] Young S. Using POMDPs for dialog management. In: Proc. of the IEEE Workshop on Spoken Language Technology. 2006.

[9] Yang MH, Tao JH, Li H, Chao LL. Nature multimodal human-computer-interaction dialog system. Computer Science, 2014,41(10): 12–18,35 (in Chinese with English abstract).

[10] Yang MH, Tao JH, Mu KH, Li Y, Che JF. A multimodal approach of generating 3D human-like talking agent. Journal on Multimodal User Interfaces, 2012,5(1-2):61–68.

[11] http://yuyin.baidu.com/asr

附中文参考文献:

[9] 杨明浩,陶建华,李昊,巢林林.面向自然交互的多通道人机对话系统.计算机科学,2014,41(10):12–18,35.

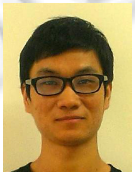
杨明浩(1977－),男,四川达州人,博士,副研究员,CCF 专业会员,主要研究领域为多模态人机自然口语对话与言语生成,病理语音康复及语音产生机理,情感计算.

高廷丽(1989－),女,工程师,主要研究领域为多模态人机口语对话.

陶建华(1971－),男,博士,研究员,博士生导师,CCF 杰出会员,主要研究领域为语音合成,虚拟现实,人机交互,情感计算.

张大伟(1989－),男,博士生,主要研究领域为人机自然口语对话,病理语音康复及语音产生机理.

孙梦伊(1987－),女,工程师,主要研究领域为人机口语对话,对话语料分析.



李昊(1989－),男,博士,主要研究领域为人机自然口语对话,病理语音康复及语音产生机理.

巢林林(1989－),男,博士,主要研究领域为人机自然口语对话,情感计算.