软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2015,26(Suppl.(2)):177−188

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

面向自然交互的多通道人机对话系统中答句自动生成方法*

高廷丽, 陶建华, 杨明浩, 张大伟, 巢林林, 李 昊, 车 浩, 李 雅, 刘 斌

(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: 高廷丽, E-mail: tingli.gao@nlpr.ia.ac.cn

摘 要: 多通道自然人机对话系统要求计算机能够对用户的语句产生智能应答,传统的人机对话系统由于知识库的限制以及用户话语的随意性,当对话内容超出知识库范围时,系统将无法应答或产生与用户期望不符的回答,这在一定程度上影响了人机对话系统用户的体验感.为了解决该问题,提出了一种融合多模态历史交互信息和面向数据的句法分析(data-oriented parsing,简称 DOP)模型的最优答句生成方法:首先从大规模句法树库中提取上下文无关文法的语法规则,然后结合对话过程中用户呈现的表情、姿态等多模态历史交互信息,融合 DOP 模型对上下文无关文法生成的汉语句子进行过滤,最终生成一个符合语法规则且符合语义的答句返回给用户,让计算机在无法获得知识库支撑时,根据交互历史信息生成应对当前对话的语句,有效地提升了多通道自然人机交互系统用户的体验感.该方法应用于交通信息查询以及咖啡厅的多主题多模态人机自由对话系统.用户的体验表明,该方法能够有效提高用户交互的自然度和体验感.

关键词: 自然语言生成;上下文无关文法;面向数据的句法分析模型;多模态信息;对话管理

中文引用格式: 高廷丽,陶建华,杨明浩,张大伟,巢林林,李昊,车浩,李雅,刘斌.面向自然交互的多通道人机对话系统中答句自动生成方法.软件学报,2015,26(Suppl.(2)):177−188. http://www.jos.org.cn/1000-9825/15028.htm

英文引用格式: Gao TL, Tao JH, Yang MH, Zhang DW, Chao LL, Li H, Che H, Li Y, Liu B. Automatic generation of sentences for natural interaction of multi-channel interactive system. Ruan Jian Xue Bao/Journal of Software, 2015,26 (Suppl.(2)):177−188 (in Chinese). http://www.jos.org.cn/1000-9825/15028.htm

Automatic Generation of Sentences for Natural Interaction of Multi-Channel Interactive System

GAO Ting-Li, TAO Jian-Hua, YANG Ming-Hao, ZHANG Da-Wei, CHAO Lin-Lin, LI Hao, CHE Hao, LI Ya, LIU Bin

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

Abstract: Natural multimodal human computer interaction dialog requires computer be able to produce intelligent response to user's statement. Due to the limitations of knowledge base and randomness of user's discourse, a traditional human-computer dialogue system cannot answer or produce consistent answer with user's expectations when the conversation is beyond the scope of knowledge, thus affecting user's sense of experience to the natural machine dialogue system. To solve this problem, this paper presents a method of generating optimal sentence by integrating multi-modal interaction history information and data-oriented parsing model. First, rules of context-free grammar from large-scale syntax tree libraries are extracted. Then combining user's expressions, gestures and other multi-modal interaction history information in dialogue process, a data-oriented parsing (DOP) model is integrated to filter Chinese sentences which are generated by context-free grammars, ultimately generating a sentence which is grammatically and semantically sound. The method allows a computer to generate responses to the current dialogue according to the interaction history information when the system can't get the support of knowledge base, therefore enhancing user's experience to multi-channel natural-machine interaction

* 基金项目: 国家自然科学基金(61273288, 61233009, 61203258, 61305003, 61332017, 61375027); 国家社会科学基金(13&ZD189)

收稿时间: 2014-06-20; 定稿时间: 2014-08-20

system. The proposed method is applied to traffic information search and multi-modal multi-topic dialogue system, and the result shows it can effectively improve the naturalness and enhance user's experience.

Key words: natural language generation; context-free grammars; data-oriented parsing model; multi-modal information; dialogue management

随着语音识别、语音合成以及数字虚拟人表达技术的发展,人与计算机的自然对话已经获得很大的进步,如英国 BBC 电视台的网络女虚拟主播 Ananova[1]、日本名古屋工业大学的数字虚拟人[2]、美国南加州大学的数字智能生命体(creative agent)[3],这些虚拟人能够以逼真的语气朗读用户给定的文字,理解用户的查询需求,回答用户的购物问题和票务信息查询系统信息等等,甚至还可以以幽默的口气对语音识别不准确的问题进行反问.此外,国内外很多公司也纷纷推出了自己的面向自然人机对话的原型系统,例如,苹果公司的语音助手 Siri,百度公司的语音助手,科大讯飞的语音助手灵犀和雨点,小 i 机器人等.可以说,数字虚拟人与人的自然对话已经在实验室环境和市场应用中取得了长足的进步,并成为自然人机交互的重要发展方向.然而,目前的自然语音交互技术距离实用化以及进入人们的生活,还有很多问题需要解决,其中一个重要的方面就是这些系统一般都以大规模知识库作为支撑,然而在对话过程中,由于不同用户的个性化表达,用户意图超出知识库范围导致问答系统无法搜索答句的情况很常见,例如,在对话系统中输入"我想喝咖啡",输出"主人没有教我这个问题!".由于类似这种情况经常出现,在一定程度上降低了用户对系统的体验感,使用户不愿多次使用人机对话系统.另一方面,当一个人的语音或语气不足以反应具体表达的意思时,有时能从脸部表情或肢体动作上判断出说话者意图和情绪,甚至一个简单的表情,辅助伴随的手势动作快与慢、幅度变化也会蕴涵丰富的交互信息.因此,情感在提升用户体验感方面也是一个重要的指标.

为了解决用户意图超出知识库范围时系统无应答的情况,研究者们提出了很多解决方案,其中比较主流的方法有基于"模板"的生成(template-based generation)方法和基于语义片段的生成方法[4].基于模板的生成方法采用"罐装文本(canned text)"作为生成自然语言的基础.例如,要生成报告飞机航班信息的自然语言文本,就可以用一个简单的"模板":[航班号] [起飞时间] 由 [出发地]起飞,预计 [到达时间] 到达 [目的地],模板中,[ ]中的内容由数据库中的数据填充后,就可以生成一个自然语言的句子输出.基于语义片段的语言生成也是一种基于模板的方法,例如,语义模板"〈表示人的名词〉是〈表示职称的名词〉〈姓名|姓|名〉先生",通过语义单元填充后可生成"史密斯先生是工程师"这个句子.基于模板的方法准确率较高,但是大规模模板获取不易,且当用户意图超出模板库范围时,又将出现系统无法应答的情况.为了解决这个问题,本文提出了一种基于自然语言生成技术的系统答句生成算法.该算法在基于上下文无关文法的自然语言生成技术基础上进行.基于上下文无关文法的句子生成研究,在理论上和应用上都有意义[5].计算机上求解的许多问题的结构都可以用上下文无关语言表示,这些结构的实例生成都可以转化成文法的句子生成问题.但是,由上下文无关文法生成的是所有符合该文法语法规则定义的句子集,生成的句子存在以下问题:(1) 该集合中的句子符合汉语语法规则,但不一定符合句法和语义规则;(2) 符合句法和语义规则的句子不一定能够适用到当前的对话片段中,此时将出现答非所问的情况;(3) 虽然生成的句子能够与当前对话场景很好的匹配,但是无法针对用户不同的情感状态,生成不同的句子集,这也将在一定程度上降低用户对系统的体验感.为了解决上下文无关文法在答句生成过程中存在的以上问题,本文提出基于上下文无关文法的自然语言生成技术生成给定关键词下符合汉语语法规则的句子集;对生成出的候选句子集,经过 Data-Oriented Parsing 模型[6]进行过滤,并结合对话管理模型,最终从候选句子集中选出最优的符合用户当前意图的自然语言答句返回给用户.最终达到了提高虚拟人表现力,增强用户体验感的目的.

本研究针对目前自然人机语音交互系统中用户意图超出知识库范围时系统无法应答的现状,给出了一种面向实用的融合多模态自然人机语音交互信息的答句自动生成算法.相对于传统的人机对话模型,本研究工作的创新点主要在于:(1) 答句自动生成过程不依赖于知识库和模板库,有效解决系统无法应答的情况;(2) 有效地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)融合到答句自动生成中;(3) 对生成的候选句子集,通过融入到对话管理模型中,评测候选句子对对话管理模型的贡献程度挑选最优语句返回给用户,有效地解决了用户意图超出知识库范围时系统无应答的问题,提高了用户对人机交互系统的体验感.

本文第 1 节介绍本文提出的融合多模态信息和 Data-Oriented Parsing 模型的答句自动生成算法框架和其

中一些相关概念.第 2 节对实验进行介绍.第 3 节对全文进行总结.

1 融合多模态信息和 Data-Oriented Parsing 的自然语言语句自动生成

基于上下文无关文法的语句生成研究在理论上和应用上都有意义,计算机上求解的许多问题的结构都可以用上下文无关文法表示,这些结构的实例生成都可以转化成文法的句子生成问题.本文的答句自动生成算法基于上下文无关文法的自然语言生成技术,为了解决上下文无关文法生成的句子集不一定符合句法规则,且不一定都能适用到当前对话流程中的问题,本文在上下文无关文法的基础上融入了多模态交互信息,并使用 Data-Oriented Parsing 模型进行候选句子过滤,最终通过候选句子集中各个句子对当前对话管理模型的贡献程度选取最优答句返回给用户.具体来说,本文介绍的自然答句生成方法包括以下几个子模块:(1) 基于上下文无关文法的融入多模态交互信息的自然答句生成;(2) 基于 Data-Oriented Parsing 模型的句子过滤;(3) 融入对话管理模型的最优答句生成.具体处理流程如图 1 所示.

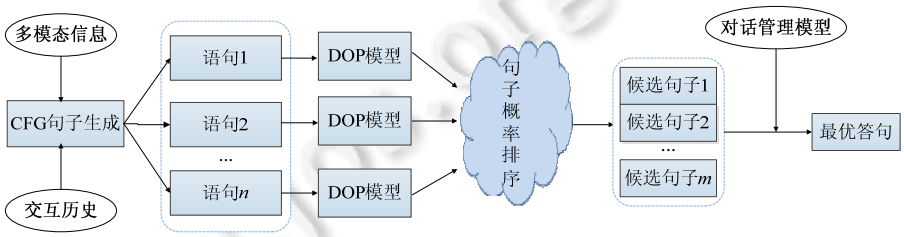


图 1 自然语言答句生成方法处理流程

该方法以对话过程中用户呈现的表情、姿态等多模态信息和系统历史交互信息作为关键词,采用上下文无关文法生成给定这些关键词下的符合汉语句法规则的句子集;针对上下文无关文法生成的句子集中每个句子,采用训练好的 DOP(data-oriented parsing)模型计算各个句子的语义得分,从而过滤候选句子集合中不符合汉语语义信息的句子,经过 DOP 模型过滤之后,将得到最终的符合汉语句法和语义规则的句子作为最终候选句子集合,系统最优答句将在最终候选集中产生.对最终候选句子集合中的每个句子,首先预测各个句子的意图,通过评估各个句子与本轮对话过程的匹配程度以及使得整个对话逻辑能够顺利进行的可能性,将得分最高的语句作为系统应答句返回给用户.

接下来将对上下文无关文法进行自然语言生成、Data-Oriented Parsing 模型、多模态交互历史信息在答句自动生成算法中的应用进行详细介绍.

1.1 上下文无关文法(CFG)

1.1.1 上下文无关文法

一个上下文无关文法 G 由 4 个部分组成,可记作 $G=\{V_N,V_T,S,P\}$,其中,V_N 是非终结符号组成的有限集合;V_T 是终结符号组成的有限集合;$V_N\cap V_T=\varnothing$;S 是开始符号组成的有限集合;P 是一组重写规则组成的集合,每个重写规则具有如下的形式:$A\to\alpha$,其中,$A\in V_N,\alpha\in(V_N\cup V)^*$.

上下文无关文法构造句子的任务,就是从“句子”这个初始结点出发,不断调用规则,产生越来越复杂的句型框架,然后从词库中选择相应词性的单词,填进这个框架里.因此,上下文无关文法生成的句子集由给定的重写规则集来确定,提取出重写规则后,借助上下文无关文法可生成所有符合重写规则的自然语言句子[7].接下来介绍上下文无关文法规则提取方法.

1.1.2 上下文无关文法规则提取

汉语上下文无关文法重写规则从汉语句法树库中提取,句法树有多种格式,本文针对的是类似“(IP-HLN (NP-SBJ (NP-PN(NR 上海) (NR 浦东)) (NP (NN 开发) (CC 与) (NN 法制) (NN 建设))) (VP(VV 同步)))”这样的用括号标注的宾州树库[8]格式的句法树.一个句法树中左、右括号完全匹配;左括号后紧跟的是句法规则符号(可能是终结符号,也可能是非终结符号),随后是一个空格,再随后有两种可能性:一是下一层的左括号;二是

具体的词或标点符号.另外,每一个左括号都有一个与之相匹配的右括号.

从宾州树库中抽取上下文无关文法规则的流程如下:

(1) 确定每对括号在一个句法树中的具体位置及其包含的句法规则符号,顺便记录该括号中可能包含的具体词或标点符号;

(2) 在得到每对括号的具体数据后,针对这些数据构成的结构数组进行分析,得到整个句法树的全部上下文无关文法规则.

我们以句子"(IP-HLN (NP-SBJ (NP-PN (NR 上海) (NR 浦东)) (NP (NN 开发) (CC 与) (NN 法制) (NN 建设))) (VP (VV 同步)))"为例具体介绍自动提取上下文无关文法规则的方法.

(1) 确定每对括号的具体位置及其包含的句法规则符号,我们用一个结构来表示每对括号对应的值:

```
struct
{
int numStart;          //括号对开始位置
int numEnd;            //括号对结束位置
string strPos;         //括号对所对应的句法规则符号
string strCluster;     //括号对直接包含的具体词或标点符号
}
```

例如,例句中第 1 个括号对是{0,"IP-HLN","",89},第 1 个包含具体词的括号对是{23,"NR","上海",29}.一个句法树中全部的括号对可以用一个定长的结构数组表示,数组长度可以通过计算括号对的数量得到.这样,用两个 for 循环加上一个 switch 语句就可以确定每对括号对应的值.例如,例句中括号对的对应数据如下:

a) 0 IP-HLN 89;b) 8 NP-SBJ 75;c) 16 NP-PN 38;d) 23 NR 上海 29;e) 31 NR 浦东 37;f) 40 NP 74; g) 44 NN 开发 50;h) 52 CC 与 57;i) 59 NN 法制 65;j) 67 NN 建设 73;k) 77 VP 88;l) 81 VV 同步 87.

(2) 在得到每对括号的具体数据后,用 3 个嵌套的 for 循环得到整个句法树的全部上下文无关文法规则,每条上下文无关文法规则从其自身包含的规则符号开始向下一层找第 1 级子节点,但不再向下面第 2 层扩展.得到的上下文无关文法规则结果:1) IP-HLN->NP-SBJ VP;2) NP-SBJ->NP-PN NP;3) NP-PN->NR NR;4) NR->上海;5) NR->浦东;6) NP->NN CC NN NN;7) NN->开发;8) CC->与;9) NN->法制;10) NN->建设;11) VP->VV;12) VV->同步.

提取出汉语语法规则之后,通过对语言规则的不断替换和推导,最终只包含非终结符时就得到了符合该文法规则定义的句子集合[9].

1.1.3 基于上下文无关文法的语句生成

提取出上下文无关文法规则后,将上下文无关文法规则构建成网络形式,在网络中通过不断的用产生式进行替换,最终当网络中从初始节点到终止节点都替换成终结符之后(终结符集由交互过程中最近 5 轮历史交互信息关键词构成),该路径就构成上下文无关文法生成的一个句子.以产生式 $S \rightarrow NpVp|Np|Vp$ 为例,构建的网络图如图 2 所示.

当网络图中各个节点都被替换成文法中的非终结符(中文句子生成中对应于每个关键词)时,从 Sent_Start 到 Sent_End 的每条路径都构成符合该文法定义的一个句子.

通过上下文无关文法生成的句子符合文法定义的重写规则,但不一定符合汉语句法规则,句子集存在以下两个问题:

(1) 上下文无关文法是基于规则的方法.规则所能刻画的知识颗粒度太大,无法用有限的规则来刻画自然语言复杂多变的现象,很难处理自然语言的不确定性.

(2) 不能保证语言学规则之间相容.也就是说,在自然语言处理系统中,随着规则数量的增加,规则之间常常发生矛盾和冲突.

因此,需要对基于上下文无关文法生成的句子集进行过滤.

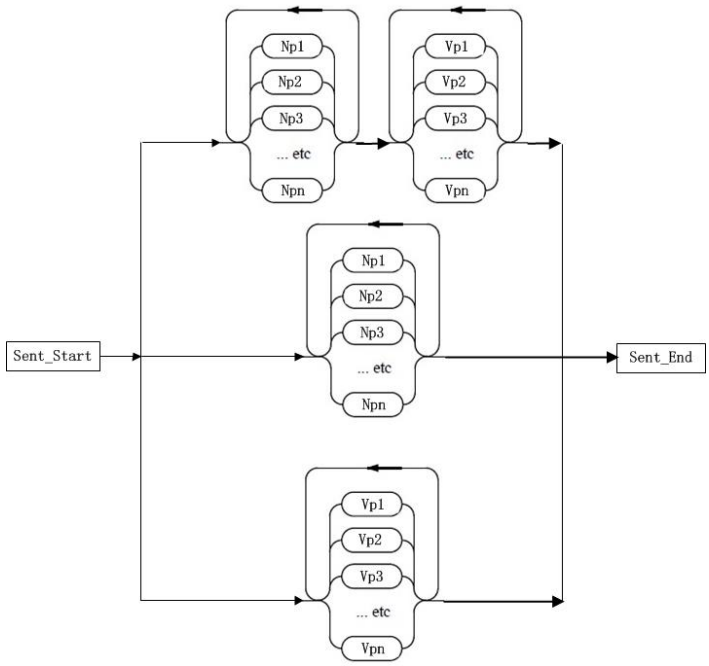


图 2 汉语语法规则网络

1.2 基于Data-Oriented Parsing模型的句子过滤

DOP 模型由 Scha 首先提出.该处理技术具体表达了这样的假设:人类对语言的领悟和创造依赖于以往具体的语言经验,而不是依赖于抽象的语法规则.Data-Oriented Parsing 技术框架可以分为:

(1) 建立包括以往成功分析的语言经验的标注语料库.

(2) 从语料库中抽取片段单元来构造新语言的分析过程.

(3) 计算分析过程的概率.Data-Oriented Parsing 模型建立在包含大量语言现象的语料库基础上,把经过标注的语料库看作一个语法(grammar)[10].

当输入一个新的现象时,系统通过对语料库中片段单元的组合操作来组合分析过程.根据所有片段单元的共现频率来评估最有可能性的分析结果.这个模型预设了一个具有带标短语结构树标注的语料库,然后从这个语料库中抽取所有任意大小规模和任意复杂结构的子树;通过对语料库中子树的组合操作来实现新输入的分析,然后考虑输入的所有派生结果的概率总和的大小来选择最有可能性的分析结果.

由上下文无关文法生成的自然语句经过 Data-Oriented Parsing 模型分析之后,可过滤不符合汉语语义的句子.在得到的候选语句集合基础上,将使用用户当前情绪的分析结果对生成的语句进行简单处理.如在天气查询的对话应用中,当用户说出“天气真糟糕”时,对话系统可以根据用户的这种负面情感,提供一些积极的信息,如第 2 天的天气会好转时,系统可以说“不用担心,明天天气就会好转”等模式,以达到更为自然的人机交互目的.因此,需要获取交互过程中用户的表情、姿态等多模态信息.

1.3 用户情感判断

一个真正智能的人机对话系统应该能够根据对话过程中用户呈现的表情、姿态等多模态信息判断用户的情感,并根据用户不同情感给出不同的应答.多模态融合,是对话管理模块根据用户语音、表情、姿态的识别结果对其意图进行理解的前提.对话过程中,系统同时接收并记录用户语音、头姿和手势等多个通道的信息输入;如,当用户说“请将这边的内容告诉我”时,对话系统会根据手势所指方向来判断需要解释的内容;如果在一段时间内没有语音信息输入,系统根据近一段时间的用户,脸部表情、头姿或者手势变化历史记录,根据这几个通道

的信息判断用户的语义表达.如,用户“点头”或者举出“OK”手势时,表示同意;而“摇头”或者“摆手”则表示不同意;在不同的上下文环境时,“摆手”又可以表示再见的意思.

情感作为人机对话的重要组成部分,对交互过程起推动和辅助作用.一个愉快的对话过程同时也是一个情感的交流过程.当用户产生负面情绪时,对话系统生成语句的过程中应该采取一些特殊的用语,以安抚用户的情绪;当用户产生明显的正面情绪(如高兴)时,系统生成的语句会偏向漫谈状态,以活跃现场气氛.在自然人机对话过程中,用户的情感识别是一个较为复杂的过程,可以从用户的语音内容、语气、脸部表情和部分手势等参数进行综合判断.本文采用两级结构的分类体系对用户情绪进行预测,分类器结构框图如图 3 所示.图中 *ft* 表示 *t* 时刻用户的表情、姿态等多模态信息特征,*dt* 表示第 1 级分类器预测的输出结果,*ot* 是最终用户情绪预测的输出结果.采用两级分类器的目的是将每一层的分类结果进行融合,常用的分类器模型有 SVM 分类器、贝叶斯分类器、神经网络等,本文使用支持向量机(support vector machine,简称 SVM)分类器和贝叶斯分类器[11].

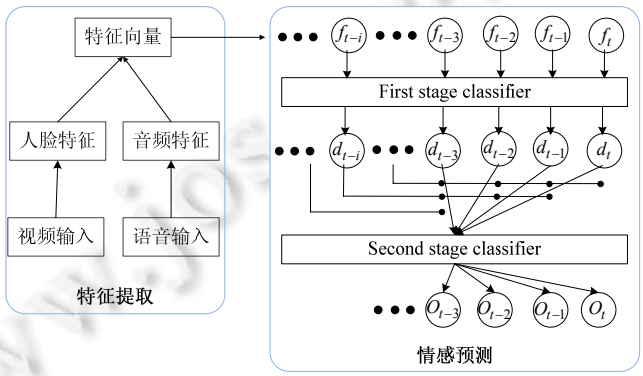


图 3 用户情感预测分类器框图

本文在 NLPR 情感数据库[12,13]上对情感预测模型进行测试,图 4 列出了 NLPR 情感数据库中的一些例子.

图 4 NLPR 情感数据库示例

该数据库包含 30 人(15 个男生、15 个女生)的数据,每个人在降噪环境中用夸张的表情录制了 2 个小时的视频数据,3 个标注者分别将视频片段数据标注为“高兴”、“悲伤”、“生气”、“害怕”、“自然”[14,15],在本文的实验中,将“高兴”、“自然”归入正向情感类;将“悲伤”、“生气”、“害怕”归入负向情感类.随机选取 500 个正向情感片段,500 个负向情感片段作为测试集,经过测试,该模型在正向情感测试数据上准确率为 77.02%,负向情感测试数据上准确率为 71.69%.判断出用户情感状态之后,将经过 Data-Oriented Parsing 模型过滤后选出的最优答句填入预先定义的情感子句模板中,作为最终的系统应答句返回给用户.例如,当检测出用户表情或语句中包含负面情感时,系统在生成的答句中会加入类似“不用担心”这样的子句,以安抚用户的情绪.

在生成的所有自然语言句子集合中,需要挑选出最符合当前对话意图的句子返回给用户,本文采用结合对话管理模型的方法,通过评估候选句子集中各个句子对对话逻辑跳转的影响,从候选句子集合中挑选最优语句.该部分内容在实验部分具体介绍.

本文提出的系统无应答时答句自动生成方法就是将上下文无关文法、Data-Oriented Parsing 模型和多模态信息获取等技术通过图 1 进行有机融合之后实现的,用户体验表明,本文方法能够有效提高用户交互的自然度和体验感.

2 实 验

2.1 对话管理框架

本文实验是自然交互的多通道人机对话系统中的一个子模块.该对话系统采用有限状态转换图模型作为对话管理模型.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,有效地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;针对人机对话中较为常用的数字虚拟人的行为控制,采用一种简化的多模态协同置标语言[16],实现了虚拟人的多通道情感动作表达和语音协调控制,提高了虚拟人的表现力[17].系统框图如图 5 所示.

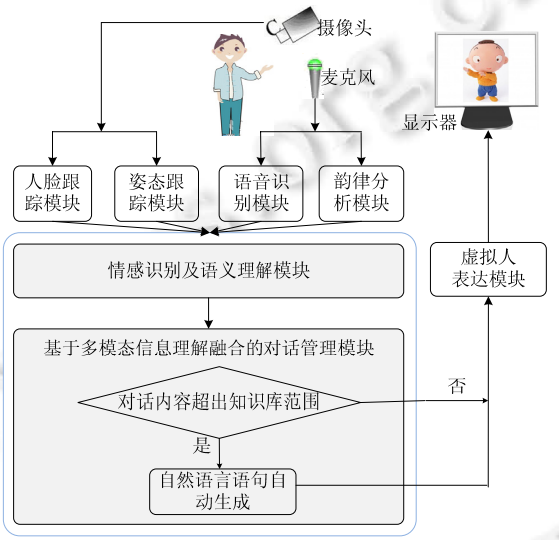


图 5 自然交互的多通道人机对话系统框图

在图 5 中,对话管理模块采用有限状态转换图来实现.文献[17]介绍了基于以上框架设计的交通路况信息查询多模态人机对话系统,本文在此框架基础上,通过修改有限状态转换图,实现了人工智能咖啡厅多模态对话系统,由于交通路况信息查询系统属于精准信息问答,而用户问句超出知识库范围的状况在咖啡等主题的漫谈对话过程中大量出现,因此本文提出的方法将在人工智能咖啡厅多模态对话系统中进行实验和测试.整个对话中的状态对应于有限自动机的 4 个节点,分别是:谈论咖啡、谈论茶饮、聊天气、漫谈,对话过程就在这 4 个状态之间跳转.图 6 给出了人工智能咖啡厅多模态对话系统的状态转换图.

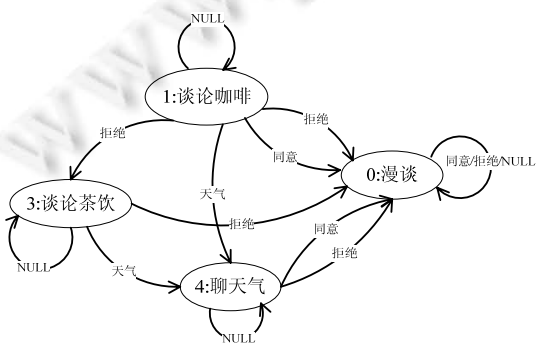


图 6 人工智能咖啡厅多模态对话系统对话管理模型

2.2 自然答句生成过程

例如,当交互历史关键词包含“不喝”“有”“咖啡”“饮料”“其他”“好喝”“咖啡机”这些关键词时,通过上下文无关文法生成的部分句子:

SENT-START 不喝 饮料 不喝 咖啡机 SENT-END;

SENT-START 不喝 饮料 不喝 咖啡 SENT-END;

SENT-START 有 咖啡 不喝 饮料 SENT-END;

SENT-START 不喝 咖啡 有 咖啡 SENT-END;

SENT-START 有 饮料 不喝 咖啡 SENT-END;

SENT-START 有 饮料 不喝 咖啡 不喝 饮料 有 饮料 SENT-END;

SENT-START 不喝 饮料 不喝 咖啡 不喝 饮料 有 饮料 SENT-END.

这些句子中有些是前后矛盾的,有些是符合语法但不符合汉语语义的,例如,子句“不喝咖啡机”.将上下文无关文法生成的所有句子分别作为 Data-Oriented Parsing 模型的输入,由于 Data-Oriented Parsing 模型是根据所有片段单元的共现频率来评估最有可能性的分析结果[18],经过 Data-Oriented Parsing 模型分析后,这些不符合语义的句子和前后矛盾的句子得到的句法分数将被过滤,以上几个关键词经过上下文无关文法后可生成 100 个句子,再经过 Data-Oriented Parsing 模型输出后选出句法分数最高的前 20 个句子作为最终的候选句子集.例如,以上 6 个句子经过 Data-Oriented Parsing 模型过滤之后剩下以下 3 个符合汉语句法和语义的句子:

SENT-START 不喝 饮料 不喝 咖啡 SENT-END;

SENT-START 有 咖啡 不喝 饮料 SENT-END;

SENT-START 有 饮料 不喝 咖啡 SENT-END.

对候选句子集中的 3 个句子,最终将结合对话管理模型,选择一个最优语句作为系统答句返回给用户.选择过程如下:

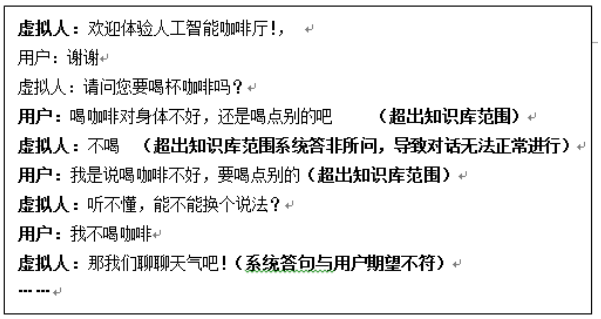
① SENT-START 不喝 饮料 不喝 咖啡 SENT-END.该句虚拟人意图是从“谈论茶饮”状态转移到“漫谈”状态,所以当系统上一个状态是“谈论咖啡”,当前状态是“谈论茶饮”且用户上一轮意图为“拒绝”时,该句子作为最优答句.

② SENT-START 有 咖啡 不喝 饮料 SENT-END.该句虚拟人意图是从“谈论茶饮”状态转移到“谈论咖啡”状态,该语句将不会被选中,因为从状态转换图中看出并没有从“谈论茶饮”到“谈论咖啡”状态的转移.

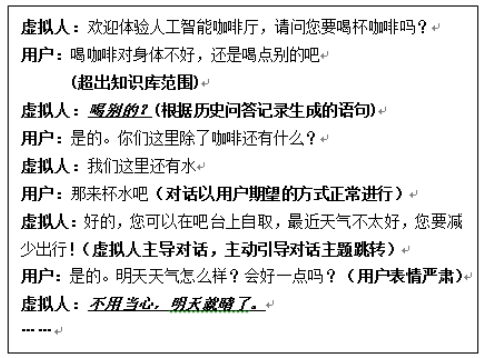
③ SENT-START 有 饮料 不喝 咖啡 SENT-END.该句虚拟人意图是从“谈论咖啡”状态转移到“谈论茶饮”状态,所以当系统当前状态是“谈论咖啡”,用户上一轮意图为“拒绝”时,该句子作为最优答句.

2.3 主观评测

本文实验主要测试在多模态人机对话系统中,当用户输入语句超出知识库范围时,系统根据对话历史和用户当前情感状态信息下答句自动生成模块的性能.图 7 给出了对同一个对话片段采用自然答句生成模块和不采用自然答句生成模块时系统的对话片段,其中图 7(a)是不使用自然答句生成模块时系统对话情况,图 7(b)是使用自然答句生成模块时对话情况(下划线句子是原系统中超出知识库范围后采用本文方法生成的句子).



(a) 不使用自然答句生成模块时系统对话情况

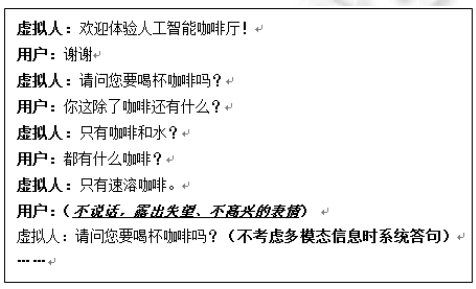


(b) 使用自然答句生成模块时系统对话情况

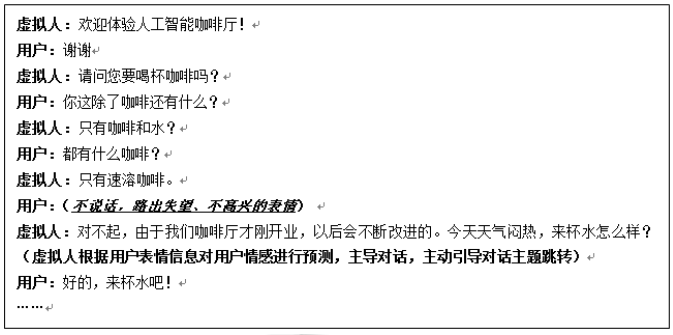
图 7

从图 7(a)和图 7(b)两个图的对话片段中可以看出,使用本文提出的答句自动生成方法能够有效应对用户意图超出知识库范围时系统无法应答或答句不符合用户期望的问题,使对话过程能够正确地进行下去.

图 8 给出了对同一个对话片段考虑多模态信息时自然答句生成和不考虑多模态信息时的对话片段,其中,图 8(a)是不考虑多模态信息时系统的对话情况,图 8(b)是考虑多模态信息时系统的对话情况(下划线句子是对用户呈现的多模态信息的分析).



(a) 不考虑多模态信息时系统对话情况

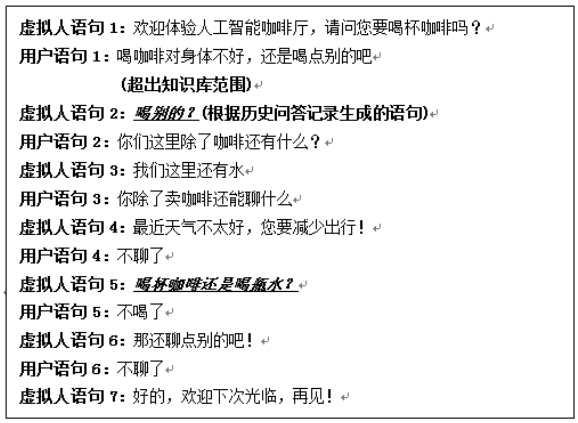


(b) 考虑多模态信息时系统对话情况

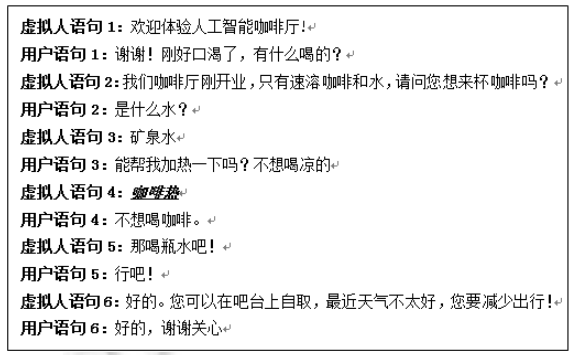
图 8

从图 8 对话片段可看出,多模态信息在对话过程中起到很关键的作用,在对话过程中,当用户的语音或语气不足以反映具体表达的意思时,有时能从脸部表情或肢体动作上判断出说话者意图,甚至一个简单的表情,辅助伴随的手势动作快与慢、幅度变化也会蕴涵丰富的交互信息,可见,多模态的人机交互方式在表达效率和完整性上都要优于传统的单一模式.

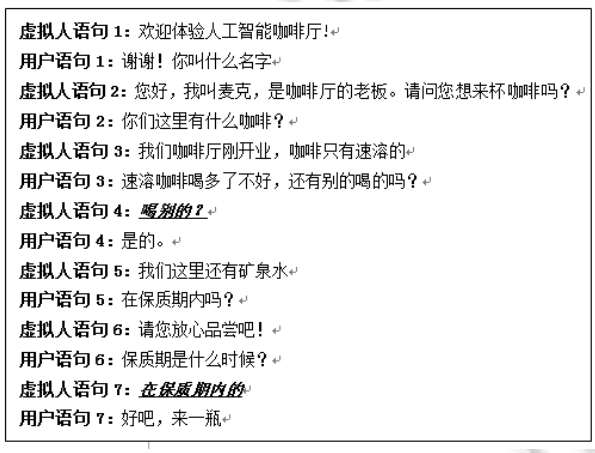
目前,国内外虽然有很多自然语言生成方面的研究和评测,但主要是从词法、句法角度去评测生成的自然语言的好坏,针对多模态人机对话系统中自然语言的生成,国内外尚缺乏有效的评价标准.这里采用了主观评测的方法对本研究的面向对话系统中自然语句生成评测[19],将用户在体验多模态信息查询系统中的对话历史记录下来,截取其中 3 段包含机器自动生成语句的片段作为评测文本,5 个评测者对这 3 段文本进行了评测,并判断文本中是否包含机器自动生成的语句,如果包含,指出机器生成的语句是哪一句或哪几句.3 段评测文本如图 9(a)~图 9(c)所示(下划线句子是原系统出超出知识库范围后采用本文方法生成的句子).


虚拟人语句 1：欢迎体验人工智能咖啡厅，请问您要喝杯咖啡吗？
用户语句 1：喝咖啡对身体不好，还是喝点别的吧
（超出知识库范围）
虚拟人语句 2：喝别的？（根据历史问答记录生成的语句）
用户语句 2：你们这里除了咖啡还有什么？
虚拟人语句 3：我们这里还有水
用户语句 3：你除了卖咖啡还能聊什么
虚拟人语句 4：最近天气不太好，您要减少出行！
用户语句 4：不聊了
虚拟人语句 5：喝杯咖啡还是喝瓶水？
用户语句 5：不喝了
虚拟人语句 6：那还聊点别的吧！
用户语句 6：不聊了
虚拟人语句 7：好的，欢迎下次光临，再见！

(a) 对话片段 1


虚拟人语句 1：欢迎体验人工智能咖啡厅!
用户语句 1：谢谢！刚好口渴了，有什么喝的？
虚拟人语句 2：我们咖啡厅刚开业，只有速溶咖啡和水，请问您想来杯咖啡吗？
用户语句 2：是什么水？
虚拟人语句 3：矿泉水
用户语句 3：能帮我加热一下吗？不想喝凉的
虚拟人语句 4：咖啡热
用户语句 4：不想喝咖啡。
虚拟人语句 5：那喝瓶水吧！
用户语句 5：行吧！
虚拟人语句 6：好的。您可以在吧台上自取，最近天气不太好，您要减少出行!
用户语句 6：好的，谢谢关心

(b) 对话片段 2


虚拟人语句 1：欢迎体验人工智能咖啡厅!
用户语句 1：谢谢！你叫什么名字
虚拟人语句 2：您好，我叫麦克，是咖啡厅的老板。请问您想来杯咖啡吗？
用户语句 2：你们这里有什么咖啡？
虚拟人语句 3：我们咖啡厅刚开业，咖啡只有速溶的
用户语句 3：速溶咖啡喝多了不好，还有别的喝的吗？
虚拟人语句 4：喝别的？
用户语句 4：是的。
虚拟人语句 5：我们这里还有矿泉水
用户语句 5：在保质期内吗？
虚拟人语句 6：请您放心品尝吧！
用户语句 6：保质期是什么时候？
虚拟人语句 7：在保质期内的
用户语句 7：好吧，来一瓶

(c) 对话片段 3

图 9

图中下划线部分语句是采用本文方法自动生成的句子,其他虚拟人的语句可以从数据库中查找到.表 1 给出了评测者对图 9(a)的评测结果.

表1 评测者对对话片段1的评测结果(判断正确的比例越高,说明使用本文方法生成的语句自然度越差)

评测者编号	评测者认为是系统生成的语句	判断正确的语句个数
1	虚拟人语句 2、5	2
2	虚拟人语句 2、4、6	1
3	虚拟人语句 4	0
4	虚拟人语句 2、4	1
5	虚拟人语句 5	1

采用准确率来衡量机器自动生成的句子的自然度[20]:

$$准确率=\frac{R}{N},$$

其中,R 表示评测者找出的由机器自动生成的正确语句总数,N 表示评测文本包含的由机器自动生成的语句总数.该对话片段中共包含 2 句机器生成的句子,因此 5 个人评测者进行评测时,评测文本中包含的由机器自动生成的语句共 10 句,最终计算得到的准确率是 50%.采用同样的方法,5 个评测者对图 9(b)和图 9(c)分别进行了评测,准确率如图 10 所示.

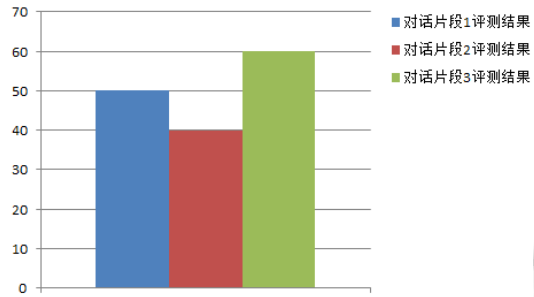


图 10 3 个对话片段的评测结果

从实验结果可以看出,评测者对系统生成的语句准确率平均在 50%左右,说明由本文所提出的包含用户情感、姿态变化的多模态自然答句生成模块生成的语句,其自然度与真实说话人在表达、情感、连贯性等方面较为匹配,有接近一半的语句评测者无法区分是由机器自动生成的语句还是说话人说出的语句.该模块能使人机交互的体验更为流畅,使得虚拟人与用户的交互显得比较自然,能够明显地提高人机对话的自然度和虚拟人的表现力,从而提高用户的交互体验.

3 总 结

本文介绍了一种融合多模态交互信息和 Data-Oriented Parsing 模型的最优答句选取方法.其思路是以上下文无关文法自然语言生成为基础,结合对话过程中用户呈现的表情、姿态等多模态信息以及历史交互信息采用上下文无关文法生成自然语言语句,并对生成的句子集采用 Data-Oriented Parsing 模型进行过滤,最后将候选句子融入到对话管理模型中,选取最优语句返回给用户.上下文无关文法保证了生成的答句符合汉语语法,同时,Data-Oriented Parsing 模型保证了生成的句子符合汉语句法和语义规则,最终对话管理模型保证了生成的句子符合当前系统的对话逻辑.通过该方法,有效地解决了在以大规模知识库作为支撑的对话系统中,当用户意图超出知识库范围时系统无应答导致对话提前终止的问题.实验结果表明,相对于简单问答的对话系统,本文提出的策略提高了虚拟人的表现力,并有助于提高人机对话的自然性,从而使用户在整个对话过程获得更为自然的体验.

References:

[1] 2014. http://en.wikipedia.org/wiki/Ananova

[2] 2014. http://www.mmdagent.jp/

[3] Morbini F, DeVault D, Sagae K, Gerten J, Nazarian A, Traum D. FLoReS: A forward looking, reward seeking, dialogue manager. In: Proc. of the 4th Int'l Workshop on Spoken Dialog Systems. 2012.

[4] Glass JR. Challenges for spoken dialogue systems. In: Proc. of the IEEE ASRU Workshop. 1999.

[5] Reiter E, Dale R. Building Natural Language Generation System. Cambridge University Press, 2000.

[6] Grune D, Jacobs C. Parsing Techniques: A Practical Guide. Ellis Horwood Limited, 1990.

[7] Dale R, Eugenio D, Scott D. Introduction to the special issue on natural language generation. Computational Linguistics, 1998,24 (3):345−353.

[8] 宾州树库. In: Proc. of the 9th Int'l Workshop on Natural Language Generation. http://www.cis.upenn.edu/-chinese/ctb.html.0

[9] 朱靖波,张玥杰,姚天顺.面向数据的句法分析技术.中文信息学报,1998,(1):1−8.

[10] Chao LL, Tao JH, Yang MH. Combining emotional history through multimodal fusion methods. In: Proc. of the Asia Pacific Signal and Information Processing Association (APSIPA 2013). 2013.

[11] 2014. http://www.cripac.ia.ac.cn/Databases/databases.html

[12] 2014. http://www.speech.kth.se/multimodal/

[13] Hall D, Llinas J. An introduction to multisensor data fusion. Proc. of the IEEE, 1997,85(1):6−23.

[14] 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.

[15] Raux A, Eskenazi M. A finite-state turn-taking model for spoken dialog systems. In: Proc. of the North American Chapter of the Association for Computational Linguistics (NAACL). 2009. 629−637.

[16] Tur G, Celikyilmaz A, Hakkani-Tur D. Latent semantic modeling for slot filling in conversational understanding. In: Proc. of the IEEE Int'l Conf. on Acoustics, Speech, and Signal Processing. 2013.

[17] 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.

[18] 18Jurafsky D, Martin JH. Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. Prentice-Hall, Inc., 2000.

[19] Lee C, Jung S, Kim K, Lee D, Lee GG. Recent approaches to dialog management for spoken dialog systems. Journal of Computing Science and Engineering, 2010,4(1):1−22.

[20] Pietzuch PR, Shand B, Bacon J. A framework for event composition in distributed systems. In: Proc. of the 4th ACM/IFIP USENIX Int'l Conf. on Middleware (Middleware 2003). LNCS 2672, 2003. 62−82.

高廷丽(1988－),女,云南双柏人,助理工程师,主要研究领域为人机对话系统.

陶建华(1972－),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语言技术,人机交互技术,虚拟现实技术.

杨明浩(1977－),男,博士,助理研究员,CCF 专业会员,主要研究领域为人机交互,发音可视化,多模态信息处理,虚拟现实.

张大伟(1989－),男,博士生,主要研究领域为面向智能人机对话的多模态信息融合技术.

巢林林(1988－),男,博士生,主要研究领域为多模态情感识别.

李昊(1989－),男,博士,工程师,主要研究领域为可视语音合成,多模态人机交互.

车浩(1983－),男,博士,工程师,主要研究领域为韵律分析.

李雅(1984－),女,博士,助理研究员,CCF 专业会员,主要研究领域为韵律模型,情感计算,人机对话系统.

刘斌(1984－),男,博士,助理研究员,主要研究领域为低速率语音编码,单通道语音增强.