

情感型对话机器人技术的研究综述

肖 鹏^{1,2}, 于 丹^{1,2}, 王建超^{1,2}, 来关军^{1,2}

(1.大连东软信息学院, 辽宁 大连 116023;
2.大连东软教育科技集团有限公司研究院, 辽宁 大连 116023)
✉xiaopeng@neuedu.com; yudan@neuedu.com;
wangjianchao@neuedu.com; laiguanjun@neuedu.com



摘 要: 对话机器人技术一直是人机交互领域的研究热点, 基于文本或者语音的对话机器人已经广泛应用于生活当中。然而, 构建能够与人类进行自然的、流畅的对话的机器人仍然充满挑战。情感作为拟人性的重要方面能够提高人机交互的自然性和流畅性。因此, 为了推进对话机器人技术的发展, 本文对情感型对话机器人的相关概念、发展历史、情感生成方式、设计思路 and 评价方式的相关研究展开了系统的梳理。情感型对话机器人主要分为指定类别情感回复和生成式情感回复两种, 其中生成式情感回复是未来发展的主要趋势。

关键词: 对话机器人; 情感; 设计; 评价

中图分类号: TP183 **文献标识码:** A

Overview of Emotional Chatbot Technology

XIAO Peng^{1,2}, YU Dan^{1,2}, WANG Jianchao^{1,2}, LAI Guanjun^{1,2}

(1. Dalian Neusoft University of Information, Dalian 116023, China;
2. Research Institute, Dalian Neusoft Education Technology Group Co. Limited, Dalian 116023, China)
✉xiaopeng@neuedu.com; yudan@neuedu.com;
wangjianchao@neuedu.com; laiguanjun@neuedu.com

Abstract: Chatbot technology has always been a research focus in the field of human-computer interaction. Chatbots based on text or voice have been widely used in practices. However, it is still challenging to build chatbots that can converse with human in a natural and fluent way. Emotion, an important aspect of anthropomorphism, can make human-computer interaction more natural and fluent. Therefore, in order to promote development of chatbot technology, this paper provides a systematic review of emotional chatbots, including related concepts, development history, emotion generation methods, design ideas, and evaluation methods. Emotion-enabled chatbots are divided into emotional responses of designated categories and generative emotional responses, of which generative emotional responses are the main trend.

Keywords: chatbot; emotion; design; evaluation

1 引言(Introduction)

对话机器人能够通过语音或者文本的方式使用自然语言与人类对话, 从而使人类能够轻松地与机器进行交流。对话机器人相关技术研究已经持续了很多年, 一部分研究将对话机器人的对话限制于特定的知识领域^[1], 例如预订助手、电商客服等; 而另一部分研究则要求对话机器人能够处理开放域内的任意输入并产生合理的回复, 例如Siri、Amazon Lex和

小冰^[2]等。近年来, 深度神经网络和大数据科学的发展加速了对话机器人技术的进步, 然而, 目前的对话机器人技术仍然处于初级阶段, 设计一个能够与人类进行连贯的、自然的且兼顾上下文语境的对话系统仍有很长的路要走。

作为人类智力的重要组成部分, 情感被定义为感知、整合、理解和调节情绪的能力^[3]。研究表明, 考虑了情感因素的对话机器人能够明显降低对话中断概率并提高用户的满意

度^[4]。因此,为了提高现有对话机器人对自然语言的理解,从而产生更符合对话情景的答复,一些研究将情感因素作为对话机器人的重要组成部分^[5]。引入了情感因素的对话系统会根据用户的情绪状态产生蕴含适当情绪的回复,能够有效提高用户参与度并创造更积极的对话环境,有效降低人机间的误解并保持人机对话上下文的情感一致性^[6],如表1所示。本文总结了现有工作在情感与对话机器人技术融合方面做出的努力,详细地介绍了情感型对话机器人技术的设计与实现及相关的评价方法。

表1 情感型对话机器人举例

Tab.1 Examples of emotional chatbot

问题	回答
大冷天喝杯热茶是真的好啊! [开心]	[正常] 今天真的冷。
	[悲伤] 我快要冻死了。
	[厌恶] 冬天是最糟的季节。
	[生气] 就你话多!
	[开心] 我也特别喜欢喝茶!
哎,宠物狗过得都比我好! [悲伤]	[厌恶] 那是你太坏了。
	[开心] 哈哈,我也是。

2 研究背景(Research background)

近年来,基于深度学习算法的自然语言处理技术取得了巨大的进步,使用自然语言作为人机交互的媒介成为一种可靠又新颖的方式。因此,人机交互方式正从预先设定的界面交互方式向基于自然语言的交互方式转变。通过自然语言与机器交互时,使用者通常使用文本或者语音的方式向对话机器人提出问题,对话机器人则使用同样的方式回答用户的问题,这种基于自然语言的人机交互系统被称作对话机器人或者对话代理系统。目前,越来越多的对话机器人被应用到人们的日常生活中,耳熟能详的有苹果手机中的Siri、百度的小度、小米的小爱同学等。普华永道最近的调查报告显示,每1,000名参与调研的用户中,有700多名受访者经常使用对话机器人完成日常的工作需求。根据Gartner的最新预测,到2022年,对话机器人将会代替人工客服完成85%的客服服务项目^[7]。

关于对话机器人的研究已经持续了多年,新技术的不断出现使得对话机器人具有多种分类方式。通常,可以根据对话机器人的交互方式、对话内容的知识领域、使用方式以及对话的生成方式来划分对话机器人的类别^[1]。如图1所示,这些划分标准涵盖了对话机器人的核心设计理念、对话理解方式以及对会话上下文考虑的程度。根据交互方式的差异,对话机器人可以划分为基于文本交互的对话机器人与基于语音交互的机器人;根据对话内容的知识领域划分,对话机器人可以划分为开放域对话机器人与封闭域对话机器人,前者要

求对话机器人能够回答任何知识领域内的问题,而后者只要求对话机器人能够回答某个知识领域内的问题即可。当以使用方式或者应用任务领域划分对话机器人时,可以将对话机器人划分为任务型对话机器人和非任务型对话机器人,前者将对话机器人的设计目的限制于使其能够帮助人们完成指定的工作任务,如预订助手、客服服务等。最后,还可以基于对话机器人的对话生成方式对其进行分类,此时可以将其划分为检索式对话机器人与生成式对话机器人。前者是对话机器人最常用的对话生成技术,它以会话的上下文作为输入,并使用启发式的函数从数据库中获取最佳的匹配回答;后者则不依赖于数据库中预先定义好的问答对,使用特定的算法结合输入问题生成新的回答,但生成式对话机器人需要大量的训练数据,并且难以优化。

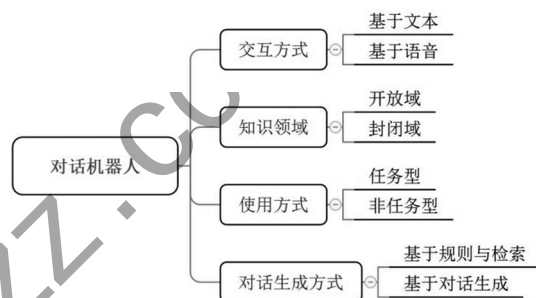


图1 对话机器人的分类标准

Fig.1 Classification criteria of chatbot

发展至今,基于对话机器人的人机交互场景中仍然存在着连贯性差、回答内容生硬以及上下文语境不符等问题。现有工作表明,考虑了情感因素的对话机器人能够有效地改善以上问题。构建具有情感属性的拟人性对话机器人能更好地理解对话中的情感因素,并在生成的回复中蕴含特定的情感,使对话机器人具有移情能力^[8]。这种移情能力能够改善用户的使用体验,提高用户的参与度并减少人机对话的中断次数,从一定程度上改善了对话机器人存在的连贯性差和自然性差等问题。因此,越来越多的对话机器人开始考虑情感因素的影响,使得构建情感因素相关的对话机器人成为对话机器人领域的研究热点。本文主要介绍情感因素在对话机器人系统中的应用。为了区别于普通对话机器人,本文将能够理解用户的情绪状态,并在生成的对话中蕴含特定情感的对话机器人称为情感型对话机器人。至此,在总结了前人的工作之余,本文将会进一步介绍三方面内容:一是情感型对话机器人的发展历史;二是情感型对话机器人技术的设计与实现;三是情感型对话机器人的评价方法。

3 情感型对话机器人的历史(History of emotional chatbot)

发布于1966年的ELIZA是第一个能够通过图灵测试的对话机器人^[9],它使用简单的模式匹配与替换规则,就使得用

户误以为正在与人进行沟通。Parry是第一个考虑了情感因素并通过图灵测试的对话机器人。Parry的设计模式与ELIZA类似,但其在基于规则生成对话的同时还设置了一个能够产生情绪的情感模块,因此Parry被认为是第一个情感型对话机器人。Polzin和Waibel在2000年提出的人机交互系统使用分类器对用户的音调进行分类,并根据感知到的情感类别使用相应的话术,使得人机间的交互更加自然和流畅。2010年,Skowron提出一款能够监听情感的对话系统,该系统能够检测对话内容中包含的情感类别,旨在感知用户的情感状态,并在内容和情感相关性上产生对用户来说更有意义的回复^[10]。这些早期工作的灵感大多来自心理学,并在小规模数据集下使用规则生成回复,难以推广到更广泛的对话生成场景中。

2014年推出的社交对话机器人小冰在设计原理上同时考虑了智商与情商,并基于检索的方式生成对话。作为一款情感伴侣对话机器人,小冰能够满足用户对沟通、情感和社会归属感的需求^[2]。当前,使用深度神经网络算法构建对话机器人系统成为主流。2014年提出的序列到序列(Seq2Seq)^[11]模型能够提高对话生成的质量,促进多样性以及处理未知单词等,因此成为构建对话系统的主流模型。2018年,ZhouHao首次使用深度神经网络算法并结合情感因素构建大型情感对话机器人(Emotional Chatting Machine, ECM),其使用三个情感相关模块共同影响对话的产生,使得系统不仅在内容的相关性和语法正确性上产生适当的回复,还能够保证对话上下文情感的一致性^[12]。以ECM为代表的情感型对话机器人为了控制回复中的情感因素,需要手动指定生成的对话情感类别,因此,这些情感对话机器人更多的是关注生成的对话中所包含的情感与预先设定的情感类别是否一致。然而,在实际交流中,人类能够自然而然地从交谈中感知对方的情感类别。为了省略手动指定生成对话所包含的情感类别,使情感型对话机器人具有更自然的同理心,论文^[13]借助VA模型^[14]编辑情感词汇,进而构建了情感相关的损失函数,以鼓励在对话的生成中产生丰富的情感词汇,使得问题与回复在情感上有自然而然的承接性,增加了对话机器人的同理心。论文^[15]借助迁移学习的思想,对一个基于多任务训练的大规模语言模型进行微调,在对话情绪检测和移情对话生成方面都取得了最佳结果。

在情感型对话机器人的数据支持方面,伴随着研究的深入,带有情感标记的对话数据集也相继公布,更权威的金标准数据改善了对话系统的性能并为该领域的研究持续赋能。DAILYDIALOG^[16]是一个针对日常聊天场景的多轮对话数据集,并且数据集中的每一句对话都进行了手动情感类别标注。EMOTIONLINES^[17]将对话内容标注为七个基本情绪类

别:中性、高兴、惊讶、悲伤、愤怒、厌恶及恐惧,该数据集由两部分组成,分别采集于Facebook聊天记录和电视剧《老友记》,由29,245句短句构成2,000段日常对话。数据集STC^[18]是一个采集自新浪微博的中文单轮对话数据集,无情感标注,但ECM使用分类器将STC数据划分为六类情感类别,作为情感对话数据集,STC需要借助情感分类器预先划分类别,因此数据质量较差。最后,论文^[19]提出一个包含了150万条Twitter对话的数据集,可以用于训练通过音调感知用户情绪的对话机器人。

4 情感型对话机器人技术的设计与实现(Design and implementation of emotional chatbot technology)

情感型对话机器人的主要任务是能够感知用户的情绪,并在产生的回复中蕴含针对性的情感,提升人机交流的自然性、流畅性。早期的情感型对话机器人使用基于规则的方式生成带有情感信息的回复。目前,主流的情感型对话机器人都使用了深度神经网络算法,这些对话机器人使用编码器-解码器的结构训练一个Seq2Seq模型,编码器负责对输入信息进行编码,解码器负责对编码后的信息解码并产生适当的回复内容。Seq2Seq模型通常由两个递归神经网络构成(RNNs)^[20],一个充当编码器,一个充当解码器。作为RNNs最重要的改进,LSTM^[21]与GRU^[22]如今广泛地应用在对话机器人系统中。

使用深度神经网络算法构建对话系统时,通常使用编码器-解码器的结构创建一个Seq2Seq模型,编码器和解码器分别由LSTM或GRU构成。以LSTM为例,解码器接收提问序列输入 $P=(p_1, p_2, \dots, p_n)$,并将其转化为隐藏状态序列 $H=(h_1, h_2, \dots, h_n)$,编码器的隐藏状态会被编码成上下文序列 $c_i=f(h_0, h_1, \dots, h_i)$,在生成状态 h_i 时,LSTM会兼顾历史隐藏状态 h_{i-1} 与上下文信息 c_{i-1} :

$$h_i = LSTM(x_i, h_{i-1}, c_{i-1})$$

同时,解码器在解码状态 s_i 时,通常会将前一时刻预测的单词的词向量编码 $e(y_{i-1})$ 与上下文序列 c 拼接作为一个整体,并同时考虑历史状态 s_{i-1} :

$$s_i = LSTM([c, e(y_{i-1})], s_{i-1})$$

对于生成的状态 s_i ,解码器使用以下公式生成对应的回复单元:

$$Word_i = \arg \max(\text{soft max}(W \cdot s_i))$$

本文主要介绍基于深度神经网络算法的对话机器人系统。如上所述,基于深度神经网络算法实现的情感型对话机器人可以分为两类:一类是生成蕴含情感因素的回复之前需要手动指定回复中带有情感类别,因此,这种方法在输入对话上文的同时还要指定生成的对话中需要包含的情感类别;另一类方法则认为生成的回复中所应包含的情感已经蕴含在上下文对话当中,不需要手动指定。前者能够按照需求

灵活产生包含指定情感的回复，但是需要大量的带有情感标记的训练数据；而后者能够基于现有的大量语料库训练而不受限于情感标注，但是其生成的对话所包含的情感类别是不受控制的。

ECM^[12]是第一个考虑了情感因素的大规模情感型对话系统，如图2所示。ECM使用GRU构建编码器-解码器结构，为了达到能够在回复中包含指定类型情感的目的，ECM在解码阶段设置了三个辅助机制：情感类别嵌入机制、内部记忆单元以及外部记忆单元。情感类别嵌入机制使用固定维度向量表示情感的类别，并将情感类别的向量表示与上下文 c 和 y_{t-1} 的词编码拼接后输入解码器中。内部记忆单元是一个辅助回复中情感表达的机制，根据论文^[23]提出的理论，有关情绪的回复具有更短的生命周期并具有较强的波动性。受此启发，作者将情绪相关的状态存储在内部记忆单元，通过读门和写门更新内部记忆单元中的情绪状态，并且随着解码过程的进行，内部记忆单元中的情绪相关状态会不断地衰减，当解码完成时，内部单元中的状态应该衰减为0。外部记忆单元的设计目的是显式地增强生成的回复中指定的情感类型的表达强度，它会同时计算情感字典和通用字典中的候选词概率，并根据状态 s_t 计算情感因子 α ，并使用 α 和 $(1-\alpha)$ 对不同词典的候选词加权，进而生成最终的候选词。受ECM启发，EmoDS^[24]认为回复中的情感表达既可以通过强烈的情感词汇显式地表现，也可以在不包含情感词汇的情况下隐晦地表达。因此，作者设计了基于词典的Attention机制，鼓励将回复中的情感词汇替换为在情感词典中的同义词，显式地增加回复中情感的表达强度。此外，作者还设计了一个句子级别的情感分类器，在整体层面上指导回复中情感的生成，以一种隐晦的方式提高回复中情感的表达强度。EmoDS与ECM类似，都使用编码器-解码器结构构造了一个端到端的Seq2Seq结构来实现对话系统。由于指定的情感类型的限定，以ECM为代表的情感型对话机器人的合理回复空间被进一步限制，因此情感型对话机器人面临严峻的安全问题。CDL^[25]为了解决这个问题，提出了课程对偶学习框架，在保证回复中情感表达强度的同时，还能够保持上下文语法及内容的合理性。CDL通过训练前向传播模型与反向传播模型构建对偶学习，并使用强化学习交替训练两个模型，前向传播模型是给出提问与指定的情绪类别生成对应的回复，反向传播模型是给出回复内容及相应的情绪类别反推提问，训练中的奖励由情感表达一致性奖励与内容一致性奖励的加权和表示。为了提高模型训练的稳定性，CDL引入了课程学习辅助训练，根据训练数据情感分类的难易程度选取了前1%最简单数据参与最初的训练，并随着训练的进行逐渐增加训练样本的难度，一

段时间后恢复为正常的训练模式。通过个性化的训练方式，CDL能够有效地降低对话机器人出现安全问题的概率，并提升对话机器人的对话品质。

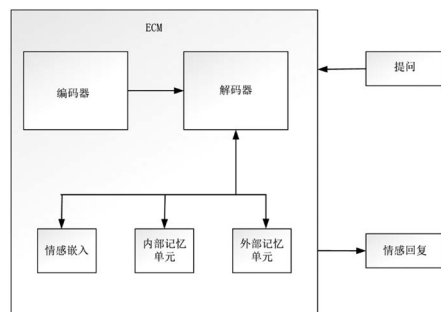


图2 ECM对话系统

Fig.2 ECM chatbot system

与ECM和CDL不同，另一部分研究者认为回复中所包含的情感类型应该由对话系统自己生成而非人为指定。ANRG^[15]认为传统的对话系统在情感语义的捕获上缺乏效率，因此在使用传统的词编码之外，还使用了VAD情绪模型构建词的情绪编码W2AV，然后将两种编码向量拼接作为输入，如图3所示。在训练阶段，作者通过对损失函数添加情绪相关正则项，鼓励网络在生成的回复中尽可能包含更多的情绪表达，从而使对话系统具有同理心。在解码阶段，作者提出一种改进的Beam Search算法，增加回复生成语义的多样性，从而提高对话系统的情感多样性。与ANRG显式地获取输入的情感词编码不同，MoEL^[26]使用一个情感追踪模块去探测输入中的情绪状态分布，并根据相关分布使用对应的监听器生成移情回复。除了充分利用对话上下文的信息外，研究者也使用多任务学习训练对话机器人辅助移情对话的生成。而Know-EDG^[9]认为人类的情感回复很大程度上依赖于外部知识与社会经验，因此作者为对话机器人引入了两个外部知识库：ConceptNet与NRC-VAD，前者是一个通用的常识图谱，后者则是一个被广泛接受的带有VAD情绪强度描述的情感词汇库。作者使用NRC-VAD度量ConceptNet中各元组尾实体的VAD值，并与输入序列中的各元素构成一个情感相关的对话上下文关系图谱，基于构建的上下文关系图谱对输入序列中的各元素进行编码，并在编码过程中同时考虑了局部信息与全局信息，使用self-attention编码当前元素与相连元素的关系，使用transformer结合LayerNorm编码当前元素与全局信息的关系。同时，作者还使用输入各元素情感值的softmax输出加权对应的各元素的编码向量，然后基于这些加权后的编码向量使用线性分类器预测输入中携带的情感类别，辅助解码器生成蕴含合理情感信息的回复。在解码阶段，作者使用改进的transformer进一步增强生成回复中的情感表达，使得解码阶段更加关注情感相关单元。

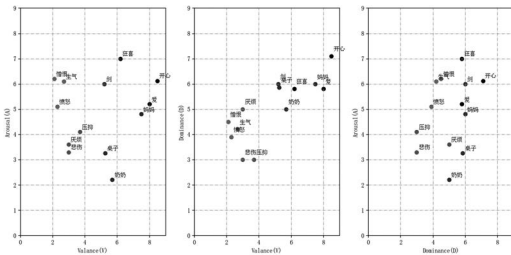


图3 VAD模型描述单词举例

Fig.3 Words described by VAD model

5 情感型对话机器人的评价(Evaluation of emotional chatbot)

情感型对话机器人的评价分为自动评价与人工评价。常用的自动评价指标有BLEU评分、Dist-1与Dist-2、情感准确性、情感词评价、困惑度评价等。BLEU用于评价回复与提问在内容上的相关性,但是,文献[26]认为BLEU与人类的判断相关性较低,因此目前并不常用。Dist-1与Dist-2用于评价生成回复的多样性。情感准确性用于评价生成回复中的情感类别与手动指定的情感类别的一致性。情感词评价用于评价生成回复中包含指定情感词汇的百分比。困惑度评价则用于评价生成的回复在语法上的正确性。因此,常用的自动评价方法是从多样性、情感一致性、回复内容的语法正确性三方面评价对话机器人的性能。而对于人工评价,常见的做法是招募多名专业的注释人员在移情性、相关性和流利度三个方面评价生成的回复。移情性表示生成的回复是否理解了提问者的情感,相关性表示生成的回复是否符合对话上下文语境,流利度则表示生成的回复语法是否正确可读。每个方面共有五级标准,级别越高评价越积极。此外,也常常会对不同人员的评价进行一致性分析,作为对话机器人性能的评价。

6 结论(Conclusion)

本文介绍了对话机器人的研究背景及大致分类,并讨论了情感型对话机器人的发展历史与技术的设计和实现。对话机器人的研究具有很长的历史,情感作为拟人性的重要标志,成为构建更高级对话机器人的必要因素。因此,情感型对话机器人引起了越来越多的关注。目前,情感型对话机器人主要分为两类,一类是需要手动指定生成回复中需要蕴含的情感类别;另一类则根据上下文自动生成蕴含特定情感类别的回复。前者依赖于大量有情感标记的数据,但生成的回复中蕴含的情感类别灵活可控。后者对数据集的要求更宽松,但是生成的回复中蕴含的情感类别并不稳定。对情感型对话机器人的评估,研究者们通常从回复内容的正确性、多样性以及情感一致性三方面考虑,可以使用自动的评价方法,也可以基于人工进行评估。

参考文献(References)

- [1] 贾熹滨,李让,胡长建,等.智能对话系统研究综述[J].北京工业大学学报,2017,43(09):1344-1356.

- [2] Zhou L, Gao J, Li D, et al. The design and implementation of Xiaoice, an empathetic social chatbot[J]. Computational Linguistics, 2020, 46(1):53-93.
- [3] Sluyter, Peter S, David J. Emotional Development and Emotional Intelligence: Educational Implications[J]. Gifted Child Quarterly, 1997, 43(2):108-110.
- [4] Prendinger H, Mori J, Ishizuka M. Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game[J]. International journal of human-computer studies, 2005, 62(2):231-45.
- [5] Ghosh S, Chollet M, Laksana E, et al. Affect-lm: A neural language model for customizable affective text generation[DB/OL]. [2017-04-22]. <https://arxiv.org/pdf/1704.06851.pdf>.
- [6] Chaves AP, Gerosa MA. How should my chatbot interact? A survey on human-chatbot interaction design[DB/OL]. [2019-04-04]. <https://arxiv.org/pdf/1904.02743.pdf>.
- [7] Svikhnushina E, Pu P. Should Machines Feel or Flee Emotions? User Expectations and Concerns about Emotionally Aware Chatbots[DB/OL]. [2020-06-24]. <https://arxiv.org/pdf/2006.13883.pdf>.
- [8] Li Q, Li P, Chen Z, et al. Empathetic dialogue generation via knowledge enhancing and emotion dependency modeling[DB/OL]. [2020-09-21]. <https://arxiv.org/pdf/2009.09708.pdf>.
- [9] Weizenbaum J. ELIZA—a computer program for the study of natural language communication between man and machine[J]. Communications of the ACM, 1966, 9(1):36-45.
- [10] Skowron M. Affect listeners: Acquisition of affective states by means of conversational systems[C]. In Development of Multimodal Interfaces: Active Listening and Synchrony, 2010:169-181.
- [11] Shang L, Lu Z, Li H. Neural responding machine for short-text conversation[DB/OL]. [2015-03-09]. <https://arxiv.org/pdf/1503.02364.pdf>.
- [12] Zhou H, Huang M, Zhang T, et al. Emotional chatting machine: Emotional conversation generation with internal and external memory[DB/OL]. [2017-04-25]. <https://arxiv.org/pdf/1704.01074.pdf>.
- [13] Asghar N, Poupart P, Hoey J, et al. Affective neural response generation[C]. In European Conference on Information Retrieval, 2018:154-166.
- [14] Warriner AB, Kuperman V, Brysbaert M. Norms of valence, arousal, and dominance for 13,915 English lemmas[J]. Behavior research methods, 2013,45(4):1191-1207.