

# 人机对话中的情绪感知与表达

黄民烈 朱小燕  
清华大学

关键词：人机对话 情绪感知 情绪表达 情感分析

## 引言

人机对话与生俱来的自然性和便捷性，使其作为一种与计算机交流的新型方式，被认为是继鼠标、键盘、触控之后，未来的一大交互范式。随着人工智能的发展，人机对话的重要性越来越凸显。起源于1950年的图灵测试，衡量的是计算机程序智能水平的测试框架，其最初的设计方式就是以人机对话方式开展的。

人机对话也是最具挑战性的人工智能任务之一，得到了学术界的持续关注。工业界对人机对话可能产生的影响和变革也高度重视。微软首席执行官萨提亚·纳德拉(Satya Nadella)认为，聊天机器人会显著地变革现有的计算方式，因而，微软也在积极地推行“对话即平台”(Conversation as a Platform, CaaP)的概念。

简略地回顾一下人机对话技术发展的一些重要事件，或许可以让人们更好地了解人机对话的发展脉络。1966年麻省理工学院(MIT)约瑟夫·魏泽鲍姆(Joseph Weizenbaum)开发出ELIZA聊天机器人用于在临床治疗中模拟心理医生。1995年A.L.I.C.E(Alicebot)聊天机器人的推出，多次获得欧洲人工智能大奖。2011年IBM研发的深度问答系统Watson在“危险边缘”电视节目中击败了人类冠军选手。2014年，在阿兰·图灵逝世60周年之际，欧洲的知识问答系统Eugene Goostman号称通过了图灵测试，被测试者认为它是一个13岁的犹太小男孩。2014年微软推出了聊天机器人“小

冰”，将其定位为人工智能伴侣。2016年微软又推出了英语版小冰Tay(后因故下线)。许多世界著名公司都推出了各式助理类的虚拟机器人。不难看出，人机对话的技术得到了长足发展，应用范围涵盖了心理治疗、情感陪护、闲聊、特定任务完成等各种场景，并快速深入各行各业。

人机对话技术与实体机器人的结合，还催生了广泛的服务机器人市场。结合人机对话技术，服务机器人有了更加广泛的应用场景，例如“情感机器人”和“社交机器人”。其中代表性产品有日本软银公司的Pepper，法国蓝蛙公司的Buddy和美国MIT推出的Jibo。Pepper被定义为“第一款面向个人且具备情感认知能力的机器人”。它能够从语音语调、面部表情中感知用户的情绪，并进行对应的情绪响应与表达。Buddy和Jibo类似，被定义为“社交机器人”，即在“机器人”的基础上，是一个能够与人类沟通和互动的系统。Jibo创始人，MIT副教授辛西娅·布雷齐尔(Cynthia Breazeal)认为，这种机器人具有两个重要维度：使用价值和情感纽带。过分强调使用价值，机器人就会变成工具；只有情感纽带，就变成了玩具。社交机器人则兼具两个维度，具有一定的闲聊对话、任务完成、情感陪护的功能。

无论是虚拟机器人还是实体机器人，情感、情绪的感知和表达是这类机器人走进千家万户的一个重要的先决条件。它不仅需要具备基本的人机对话功能，更重要的是如何感知和理解人类的情感和情绪，并做出符合情境的情感响应。

## 情感和情绪

人类情感分两种,一种是低级情绪,通常认为就是高兴、喜欢、悲伤、愤怒、恐惧等基本情绪——它们是生物性的,理论上是可以通程序模拟的。另一种是高级情绪——它们是社会性的、精神性的,比如道德、爱情、自私、利他等。高级情绪是否可以通过程序模拟尚不可知,但如果人工智能拥有完善的低级情绪及学习能力,则有可能自行发展出高级情绪。目前人工智能对于情绪的研究还处在“低级情绪”研究的初级阶段。

情感和情绪的理解是人类最重要的认知行为之一<sup>[1,2]</sup>,为了实现具有人类水平的行为和交流能力,机器人必须能够理解人类的认知行为。赋予计算机或机器人人类式的情感,使之具有识别、理解与表达喜怒哀乐的能力,这也是许多科学家的梦想。

情感机器人就是赋予了情感能力的机器人,具有识别、理解和表达情感与情绪的能力。相关研究表明,通过文字、面部表情、语音等具有情绪感知的人机交互系统能够很大程度提升用户满意度,避免陷入对话僵局,改善双方的对话行为。情感机器人的研究包括两个重要的问题:一是情感的感知与识别,即如何从字面信息、语音语调、表情、肢体语言中感知用户所表达的情感和情绪;二是情感的模拟表达,即如何通过声音、动作等体现出合适的情感响应。对应地,具有类人情感的机器人必须具备两个基本系统:情感计算系统和情感表达系统。

情感计算的概念是在1997年由MIT媒体实验室皮卡德(Picard)教授提出的<sup>[2]</sup>。她指出,情感计算是与情感相关,来源于情感或能够对情感施加影响的计算。情感计算的目的是通过赋予计算机识别、理解、表达和适应人的情感的能力来建立和谐人机环境,并使计算机具有更高的、全面的智能。情感机器人需要从用户的表达内容、语音语调、表情动作中分析其所表达的情感和情绪,并根据情境做出合适的响应。

情感表达使得机器人可通过多维度输出情绪,如文字、语音、表情及动作等。特定的情绪表达可

以通过内容体现,不同的内容可以表达不同的情感;也可以通过语音语调体现,同一句话可以用不同的语气和语调体现不同的情感表达;还可以通过表情和动作体现,例如面部表情(笑脸、哭脸、皱眉等)和肢体动作等。对于通过文字输入输出的虚拟机器人,内容表达成为最主要的表达方式,“说什么”变得更加重要。

## 情绪化内容生成

目前已有一些工作开始探索如何让计算机生成具有情感表达的文字内容。Radford等人<sup>[3]</sup>在大量的评论数据中训练了字节级别的循环神经网络模型,并发现了可以表达情感的神经元(Sentiment Unit),手工控制该神经元的响应输出可以得到不同极性的情感表达内容。Hu等人<sup>[4]</sup>则在变分自编码器的基础上,提出了混合随机噪声和可控编码的文本生成方法,实现了能控制情感极性、时态的可控文本生成。Ghosh等人<sup>[5]</sup>研究了利用神经网络语言模型表达情绪化的内容,在生成过程中不仅考虑已生成内容的上下文,还考虑其所携带的情感信息(利用外部情感资源计算)。

本文探索的是一种基于深度学习的虚拟机器人(即文字输入、文字输出形式)表达情感的方法<sup>[6]</sup>。基于深度学习的闲聊对话技术取得了长足的发展,一般采用编码-解码(Encoder-Decoder)的序列学习模型,在大规模闲聊对话语料上训练得到的模型能生成较为自然和流畅的对话。但已有的工作还聚焦在如何提高对话生成质量上<sup>[7-9]</sup>,在对话场景下几乎没有人同时考虑内容的相关性和情感表达的合适性。本文探索了当给定一个语句以及指定一个情绪类别时,计算机能否生成一个合适的答复。在未来工作中,我们将探索如何根据用户输入的内容以及当前的情境,预测计算机下一步最适合响应的情绪内容。

## 总体框架

模型框架如图1所示。首先,我们从社交媒体



图1 情绪化内容生成的总体框架

中得到几百万条互动数据，每条数据包含一个发帖 (Post) 和回复 (Response)，这些数据将作为我们的训练数据。之后，我们采用预先训练好的情绪分类器对每条数据的发帖和回复进行情绪类别的自动标注，每个发帖和回复都会被标记为“快乐、喜欢、悲伤、厌恶、愤怒、其他”中的一类。情绪分类器采用的是双向长短期记忆 (LSTM) 模型，即每篇文本内容用一个双向 LSTM 模型表示为向量，然后在 Softmax 层进行分类。训练数据来自 NLPCC 2013、2014 评测的情绪分类数据集，分类准确度 (accuracy) 可以达到 62%。最后，这些被标记好类别的数据，会送入情绪聊天机器 (ECM) 的情绪表达模块，以便生成指定情绪类别的回复内容。

值得注意的是，情绪分类器的性能还比较低，被自动标注情绪类别的数据是带有噪声的，但对于我们后续训练情绪生成模块来说已经够用了，基于这样的数据，我们的模型能够得到有意义的情绪表达内容。

## 情绪表达

情绪表达模块是基于深度学习的序列到序列生成算法，即 Encoder-Decoder 的编解码结构。这一模型的编码器部分将输入信息  $X$  (如“我今天好惨，考试没及格”) 用神经网络模型 (通常有门控循环网络 GRU、长短期记忆模型 LSTM) 编码为一个固定长度的向量。在解码器部分，采用注意力机制逐步解码生成一个个单词，当前解码生成的单词由输入编码、之前的解码内容共同决定。其本质是对概率公式  $P(y_i | X, y_{i-1}, y_{i-2}, \dots, y_1) = \text{softmax}(W * S_i)$  建模，其中每个  $y$  表示一个词， $\text{softmax}$  是一个指数函数形式的概率规

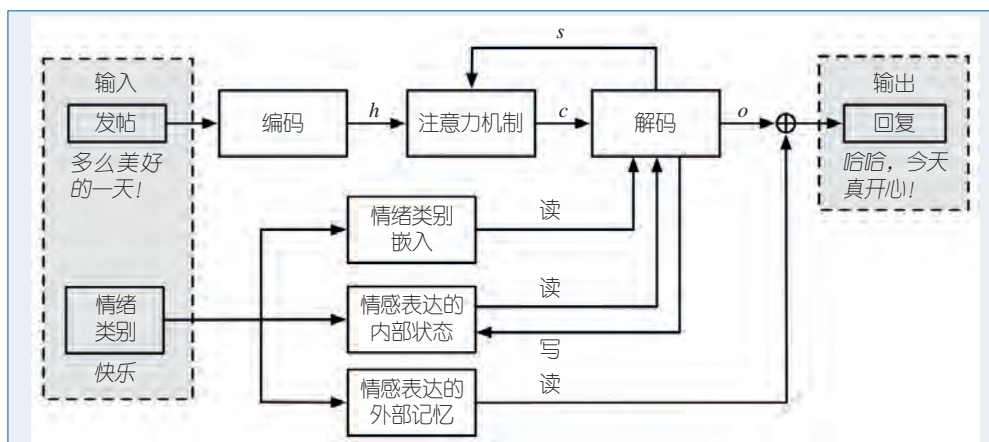
图2 情绪表达模块的模型结构图 ( $h$ : 编码之后的向量;  $c$ : 注意力机制产生的上下文向量;  $s$ : 解码器的状态变量;  $o$ : 输出的概率分布, 依据此概率进行解码)

表1 情绪化内容生成示例

指定情绪	帖子与回复
发帖	早上的天气不错啊！哈哈哈哈哈
喜欢	今天的天气不错啊！
悲伤	天气好，心情不好。
厌恶	这是哪里啊？求解释
愤怒	怎么了？什么情况？
快乐	今天的天气真好啊！
发帖	送一个这样的蛋糕，给你美丽的闺蜜吧！
喜欢	好漂亮的蛋糕啊！
悲伤	为什么我觉得很恶心
厌恶	这样的蛋糕你忍心吗？
愤怒	有没有想开网店的？
快乐	哈哈，好可爱的蛋糕！
发帖	承诺，就是一个骗子说给一个傻子听的。
喜欢	承诺是一个专一的傻子
悲伤	我就是一个悲伤的傻子
厌恶	一个骗子说给一个傻子听
愤怒	不管你信不信，反正我信了！
快乐	我就是一个开心的傻子
发帖	为了解决抄作业的坏习惯，高等数学老师这回要出绝招了！
喜欢	这老师太有才了！
悲伤	这样的老师伤不起啊
厌恶	这老师也太狠了点吧
愤怒	这老师是要逆天啊！
快乐	哈哈，这老师太有才了

格化函数，这个概率可以由解码器神经网络的输出状态  $S_t$  决定。

为了生成内容相关、情感表达合适的回复内容，我们尝试了三种策略：情绪类别嵌入、情感表达的  
内部状态、情感表达的外部记忆。模型表达的总体  
框架如图 2 所示。

1. 情绪类别嵌入 (Emotion Embedding) : 我们把  
每个情绪类别表示为一个向量 ( 模型参数的一部分，  
需要学习得到 )，然后把指定要生成的情绪类别作  
为一个额外信息输入到解码器神经网络 GRU 中。

2. 情感表达的  
内部状态 (Internal Memory) : 我们  
认为每个待生成的情绪类别对应一个情绪内部状  
态向量，随着解码器每次解码出来一个词，这个向  
量会衰减一部分信息，解码完成时该向量衰减到零，  
认为其对应的情感信息得到了完全表达。在实现上，  
我们采用了记忆网络中的读、写操作。每次解码前，  
利用读取门 (Read Gate) 从该向量中读取一部分信  
息；解码后，利用写入门 (Write Gate) 对该状态向  
量进行更新。同样的，每次读取的信息作为一个额  
外信息输入到解码器神经网络 GRU 中。

3. 情感表达的  
外部记忆 (External Memory) : 为了  
区分解码过程中的一般词汇 ( 如天气、你、我等 )  
和情感表达强烈的词汇 ( 如高兴、悲伤、惨等 )，我  
们引入一个外部记忆单元区分。在解码过程中，通  
过控制输出词表上的概率分布来引导当前解码器选  
择合适的词语表达情感。

通过以上三种策略，机器可以生成内容相关、  
情感表达合适的回复内容。表 1 展示了我们的模型  
生成的一些结果。量化评估的结果表明我们的模型  
比传统的 Seq2Seq 模型以及其他的一些配置更好。

情感交互模式分析

在人类对话行为中，大多数的情感交互模式为  
共情模式。我们通过对数据中的情感交互模式进行  
分析和统计，也发现了这个现象 ( 见图 3 )。一个情  
感交互模式定义为二元组  $\langle E_{post}, E_{response} \rangle$ ，其中  $E_{post}$   
和  $E_{response}$  分别表示发帖和回复的情绪类别。共情，  
即当发帖表达某种情感时，回应也表达同样的情感。  
这些例子包括  $\langle \text{快乐}, \text{快乐} \rangle$ ,  $\langle \text{喜欢}, \text{喜欢} \rangle$ ,  $\langle \text{悲伤}, \text{悲伤} \rangle$  等。换言之，当一方表达情绪 A 时，另一  
方表达情绪 A 的概率要显著高于表达其他情绪的概  
率。例外的情况也有，例如  $\langle \text{悲伤}, \text{喜欢} \rangle$ ，这通  
常是因为当一方表达悲伤情绪时，另一方往往表达  
安慰，也符合情感交互的常理。



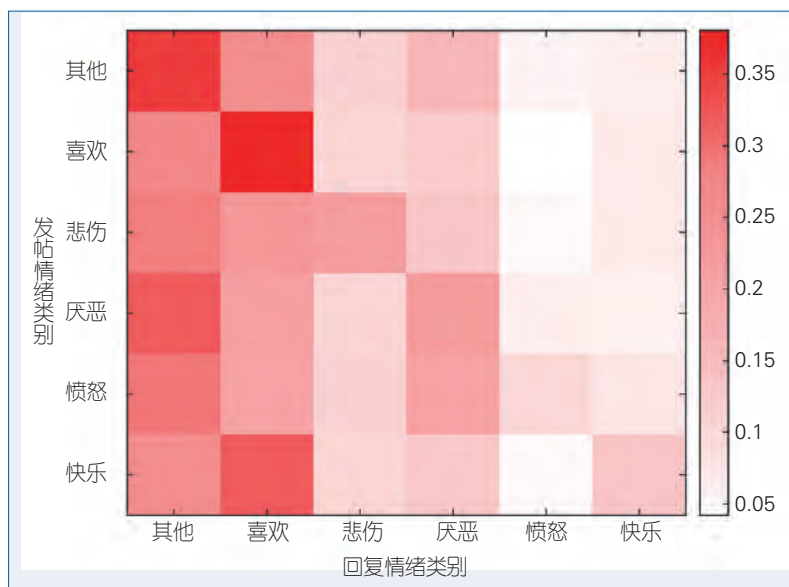


图3 情感交互模式分析。颜色深浅表示这种模式的出现次数，越深表示出现次数越多

上述情感交互模式分析并没有考虑发帖的内容。实际上，我们发现情感响应与发帖的具体内容密切相关。例如，当发帖的内容是时事政治类话题时，情绪响应多为厌恶、愤怒；而当话题为一般的情感、心情表述时，情绪响应多为快乐、喜欢、悲伤。但仅仅依据发帖的内容，有时候很难判断出合适的情绪响应，因为情绪响应往往还取决于用户当时的背景信息、情境以及其他更多的上下文。

## 总结与展望

本文探索了一种基于深度学习模型、面向文字的情绪表达方法，表明计算机可以具有表达初步情绪的能力。通过大量的对话语料训练模型，计算机可以生成表达不同情绪类别的内容。为了实现类人水平的人机对话技术，机器人的情感感知与表达能力是一个必然的要求，而这正是对话机器人能够获得更加广泛应用的重要前提，使之可以在情感陪护、心理咨询等更复杂的任务中得到应用。在情绪感知上，机器人要能够从多模态信息，如文字、语音语调和表情等识别与理解人类所表达的情感与情绪；在情感表达上，机器人要能够通过文字、语音、表情、

动作等输出多模态的情感情绪信息。这需要自然语言理解、语音识别与合成、视觉信息处理、机器人规划执行等多种技术的综合运用。

但是，目前计算机还只能处理单一低级情绪。在相对复杂的动态环境中，人类的情绪是动态变化的，而且是复杂多样的。计算机要发展到能够维护、发展自我情绪，尤其是进化到道德、精神层面的高级情感，还有相当遥远的路程。无论是低级情绪还是高级情感，都是影响人类决策与行为的重要因素，对其进行深入探索意义深远。



黄民烈

CCF专业会员。清华大学副教授。IJCAI 2017高级程序委员，IJCNLP 2017、ACL 2016、EMNLP 2014/2011等国际会议领域主席。主要研究方向为人工智能、深度学习、机器学习方法与应用，自然语言处理等。aihuang@tsinghua.edu.cn



朱小燕

CCF高级会员。清华大学智能技术与系统国家重点实验室主任、计算机系教授。加拿大国际发展研究中心(IDRC)特聘教授(IRCI)。主要研究方向为智能信息处理、新一代智能对话/交互系统。zxy-dcs@tsinghua.edu.cn

## 参考文献

- [1] Salovey P, Mayer J D. Emotional intelligence[J]. *Imagination Cognition and Personality*, 1990, 9(3):185-211.
- [2] Picard R W, Picard R. *Affective computing*[M]. Cambridge: MIT press, 1997.
- [3] Radford A, Jozefowicz R, Sutskever I. Learning to generate reviews and discovering sentiment[OL]. (2017). arXiv preprint arXiv:1704.01444.

更多参考文献: [www.ccf.org.cn/sztsg/cbw/zgjsjxhtx/](http://www.ccf.org.cn/sztsg/cbw/zgjsjxhtx/)