

## 《人机对话中的情感处理》评介

Johannes Pittermann, Angela Pittermann, Wolfgang Minker 著

吴国良 冯春灿

《人机对话中的情感处理》(*Handling Emotions in Human-Computer Dialogues*) 2009年由Springer-Verlag出版社出版,作者是Johannes Pittermann、Angela Pittermann和Wolfgang Minker。

近年来,语言中的情感(emotion)已经成为语言学界一个热门的研究主题,在文本分析、人机对话等领域正日益得到研究者的关注。该书共分七章,以人们在自然语言交际过程中所流露出来的情感为研究对象,以计算机处理情感为研究主题,对人类情感、人机口语对话系统、对话系统评价等方面展开了深入讨论,并提供了翔实的研究实例样本。本书主要关注如何在人机口语对话系统中融入用户情感状态管理的问题,描写了情感的识别与分类,并提出了融情感于适应性对话管理的处理模型。

### 内容介绍

作者在其序言中对情感研究的研究背景、人机对话的应用领域和全书的章节安排作了简要的介绍。本书第一章是引言,按六个核心问题对人机口语对话系统(SLDS,即Spoken Language Dialogue System)的功能模块及其在当前人机对话系统中的应用作了详细介绍:(1)口语对话系统;(2)增强型口语对话系统;(3)对话管理中的挑战;(4)用户模型中的问题;(5)对话系统的评估;(6)作者对自身研究的评价。该书在人机口语对话系统上有两个关注点,即界面友好度和自然度。这涉及语音信号中的情感识别以及对话与用户情感状态的相适应问题。一个人机口语对话典型系统由自动语音识别、自然语言处理、对话管理器、文本生成、文本转换为语音等五个基本功能组件构成,如图1所示:

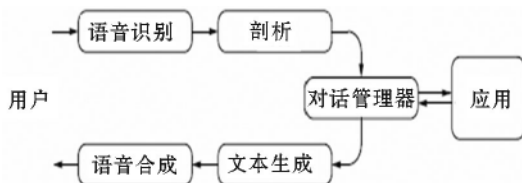


图1 人机口语对话典型系统

本章末概述了作者在人机对话情感研究方面的六项贡献: (1) 简单情感识别器的运用。采用隐马尔科夫模型( Hidden Markov Model) 对从语音信号分离出来的韵律与声学特征进行情感状态分类; (2) 语音情感识别器的开发。通过区分情感音素( emophoneme) 将语音识别与情感识别的两个过程合二为一。所谓情感音素, 就是附有情感状态的音素; (3) 语音情感识别器的优化。通过扩展成二次处理的识别器实现; (4) 改进识别器输出投票减错法( ROVER, 即 Recognizer output voting error reduction) 算法。使之结合多个语音识别器以实现较好的单词准确性。改进算法能够用来处理来自语音情感多处理器的输出数据。(5) 半随机对话模型的开发。无需借助于复杂的规则集, 即可使对话过程灵活地适应用户情感状态; (6) 适应性对话管理器的运用。将其对话模型的灵活性与“语音可扩展标记语言”( VoiceXML) 框架的方便编程界面相结合。

人类情感是作者的关注焦点, 第二章对此作了细致的描写, 对情感的种种定义、多种情感理论、情感标注、情感语音语料库等进行了全面的梳理。这四个方面影响到自动情感识别器能否有效工作。其中, 语音样本的质量及其标注对识别器模型的质量有重大影响。最后作者对语音中的情感分类、测量、标注等困难作了简要说明。一方面情感在人们日常生活中的有重要作用, 有研究价值; 另一方面, 情感无法客观地加以测量、标注, 情感标注只能采取主观的方法, 因此对现有的种种情感识别器的评估在很大程度上难以有效展开。在本书中, 作者出于实际考虑, 将所研究的情感样本数定为七种, 其中六种情感, 分别是愤怒( anger)、无趣( boredom)、厌恶( disgust)、畏惧( fear)、高兴( happiness)、悲伤( sadness), 再外加一个中性情感( neutral)。为便于情感状态的使用, 对上述情感状态采用数字标注法, 该数值根据情感在激活-评价空间( activation-evaluation space) 中的位置确定。本书中实例使用的语料库是柏林情感语音数据库( Berlin Database of Emotional Speech, Burkhardt, et al. 2006)。

第三至第五章描写了作者将情感融入适应性对话管理的策略、基于语音的情感识别方法及其复合语音情感识别与优化方法。第三章描写了对话策略概念和情感对口语对话的影响, 对适应性人机对话进行了深入描写。在介绍了适应性对话管理的研究现状之后, 深入讨论了基于规则的方法和半随机对话与情感模型。作者的研究把这些模型综合成一个情感对话模型, 并详细阐述了该模型的扩充方式。

第四章的中心是语音信号的自动情感识别, 尤其是作者的语音情感识别器。主要讨论了包括信号处理与分类在内的各种技术细节。在简要介绍现有相关研究的基础上, 提出了作者根据一个简单语音识别器设计出来的一个简单情感识别器。情感的识别, 依赖从语音信号中分离出来的种种特征。本章第一部

分介绍了信号处理与情感特征的分离,涉及预处理、线性预测、梅尔频倒频谱系数、韵律与声学特征等。第二、三部分介绍了情感识别分离器和现有情感识别方法,描写了分类方法与实际识别过程的各种细节。作者的情感识别器与语音识别器关系密切,尤其是在特征的分离与分类方面。因此作者详细描写了其识别器的算法与处理方法,并设计出了采用其综合语音情感识别器的人机口语对话系统工作流程图。第四、五部分首先描写了独立的语音、情感及语音情感的识别器,然后详细介绍作者的综合语音情感处理方法。语音识别器与情感识别器均可从该方法获得启发。因为该方法从信号处理与分类上语音与情感处理的相似性得到启发,在深入描写其情感识别器之前,还选择部分简单的语音识别进行了必要的介绍。情感状态、情绪与倾向性不依赖于各种语音信号分析与分类方法也可以从文本的语言内容中加以确定,因此第七部分介绍了借助于各种语法识别情感的方法。第五、六部分也将这种识别方法融入了作者的语音情感识别器。

第五章主要对作者的情感识别器与语音情感识别器提供了一些改进方法,并示范其适应性对话管理器的实现与应用。第三、四章有关适应性对话管理与基于语音的情感识别,为适应性人机口语对话系统情感融入的理论层面打下了坚实的基础,但这个基础仍有很大的改进潜力和灵活性。因此作者对综合多个识别器结果的方法作了改进,以获得更好的识别效果。本章第一部分首先识别了其情感与语音情感识别器的改进方法,描写了其适应性对话管理器的实现,具体涉及其简单情感识别器与综合语音情感识别器的优化。识别的优化不总是需要改变识别器本身。有时,一个后处理算法就可以减少识别错误。这里还提出了一个语音情感识别的二次处理法:第一步分离出语音情感的各种特征,第二步语音情感识别器采用经优化的基于句子的语言模型。第二部分探讨了其运用多个简单情感识别器或综合语音情感识别器来降低总出错率的办法。这里作者把识别器结果投票减错法的原理运用到了其语音情感识别器。第三部分介绍了基于语音可扩展标记语言的适应性对话管理器,它融合了第三章第八部分所介绍的半随机情感对话模型,已实现对话过程与用户情感状态的相适应。图2是嵌入标准对话系统环境的适应性对话管理器的简化结构图:

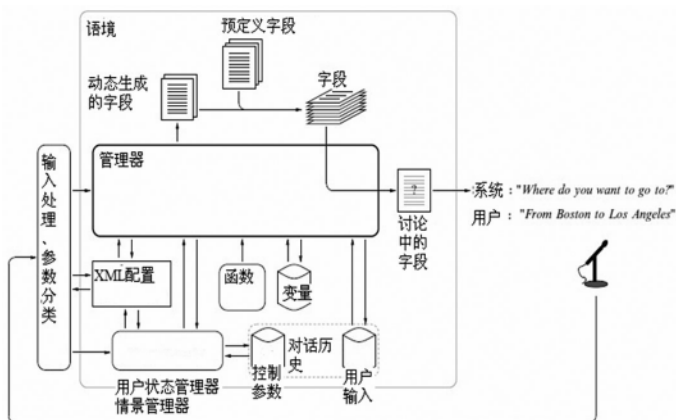


图2 嵌入标准对话系统环境的适应性对话管理器的简化结构图

第六章专门介绍人机口语对话系统的评估，以使不同的系统与概念可以相互比较。识别一个系统的弱点、制定测试的标准与手段，是人机界面评估的主要挑战之一。本章第一部分介绍了评估的各项参数。第二部分介绍了用于评估的语料库资源。尤其对作者在其研究中采用的柏林情感语音数据库进行了详细描述。第三部分对第四、五章所描述的情感识别器进行了深度评估，包括简单情感识别、语音情感识别、多语音情感识别器、情感识别的语言分析法等。第四部分对第五章第三部分所描写的对话管理器的可用性提出了自己的测量方法。

第七章首先重申了全书的关注焦点，即为了使人机口语界面更加有效、自然、友好，一个人机口语对话系统能如何、应如何对用户情感状态做出合适反应？指出作者的研究重心在于结合语音识别器从语音信号中识别情感和借助于扩展的用户状态与对话管理器将情感融入对话过程。接下来回顾总结了第三至第六章的主要内容，尤其是第三、四章中的适应性人机对话与综合语音情感识别，以及第五章的情感识别改进方法。全书提出了基于语音的情感识别的四种方法和改进措施。最后简要描述了存在的问题及作者今后的研究方向。语音与情感识别器总体效果受到语料库的质量与规模制约。作者寄希望于获得较大规模的语料库，以调试与提升其识别器的效率，并计划改进识别器结果投票减错法工作机制，以使之更好地融入其语音情感识别。作者还计划进一步关注语言与副语言信息，以获得更好更稳定的综合性识别器。要改进情感性的语言分析，尚需改进分类方法，需要脱离基于规则的语法方法，倾向于随机的语言模型法，或者两者的结合。本书第三章中的半随机对话模型遭遇语音情感识别器的相同困境，即没有足够的数据库。作者计划改进状态转换的数量、减少状态的数量来增进适应性对话系统的效果。在如何将智力与情感意识融入人机口语对

话系统的问题上,作者目前正在致力于能使用户与一个或多个应用进行自然友好交互的系统。

### 简评

本书存在的主要问题如下:

第一,诚如作者指出的那样,本书示例的人机口语对话系统的语音语料库偏小,影响到情感识别的效度。自然语言有语言、语言使用者、地域、环境、性别、年龄等多方面的变化因素,自然语言的语音处理难度因此增加许多。

第二,本书作为示例的情感集合只有七种情感,数量较少。人类的情感细腻而且丰富,情感在语音与文本中的切分与分离,其操作性仍无法获得较统一有效的方法。如何在较大情感集合上获得较好的情感识别以更进一步贴近自然语言特征,仍有许多研究空间。

有的情感是与生俱来的,在某种意义上是人类普遍具有的;有的情感是后天习得的,这些情感的分类随文化的不同而不同( Gu in press; Burkhardt, et al. 2006)。Gu 把情感定性为言语施事行为的内在要素( emotion is an intrinsic element of illocution),情感变化跟言语行为的实施是紧密联系在一起的。这些对于本书的研究架构都是有难度的挑战。

本书对人机对话中的情感识别提供了最新的研究示例,连同其书后附录所收录的近百个涉及 20 余种语言的情感语音数据库目录,对于任何想在情感性的人机口语对话系统领域进行研究的读者都不失为一部好的实用参考书。

### 引用文献

- Burkhardt, F., N. Audibert, L. Malatesta, O. Turk, L. Arslan, and V. Auberger. 2006. Emotional prosody – Does culture make a difference? In 3rd International Conference on Speech Prosody, Dresden, Germany. Pp. 245 – 8.
- Gu, Y. ( 顾曰国). In press. A conceptual model of Chinese illocution, emotion and prosody. To appear in 《第四届国际汉学大会论文选集》。台北: 中研院。

作者通讯地址: 312000 浙江省绍兴市越城区会稽路 428 号浙江越秀外国语学院  
E-mail: wgl11203@yahoo.cn ( 吴国良); jhfcc@163.com ( 冯春灿)