计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.200900088

俄语多模态情感语料库的构建及应用

徐琳宏[1]　刘　鑫[1]　原　伟[2]　祁瑞华[1]
1 大连外国语大学语言智能研究中心　辽宁 大连 116044
2 信息工程大学　河南 洛阳 471003

摘　要　俄语的多模态情感分析技术是情感分析领域的研究热点,它可以通过文本、语音和图像等丰富信息自动分析和识别情感,有助于及时了解俄语区民众和国家的舆论热点。但目前俄语的多模态情感语料库还较少,因而制约了俄语情感分析技术的进一步发展。针对该问题,在分析多模态情感语料库的相关研究及情感分类方法的基础上,首先制定了一套科学完整的标注体系,标注内容包括话语、时空和情感3个部分的11项信息;然后在语料库的整个建设和质量监控过程中,遵循情感主体原则和情感连续性原则,拟订出操作性较强的标注规范,进而构建出规模较大的俄语多模态情感语料库;最后探讨了语料库在解析情感表达特点、分析人物性格特征和构造情感识别模型等多个方面的应用。

关键词:多模态;情感分析;语料库;俄语
中图法分类号　TP393

Construction and Application of Russian Multimodal Emotion Corpus

XU Lin-hong[1],LIU Xin[1],YUAN Wei[2] and QI Rui-hua[1]
1 Research Center for Language Intelligence of Dalian University of Foreign Languages,Dalian,Liaoning 116044,China
2 Information Engineering University,Luoyang,Henan 471003,China

Abstract　As a research hotspot in the field of emotion analysis,Russian multimodal sentiment analysis technology can automatically analyze and identify emotions through rich information such as text,voice and image,which is helpful to timely understand the public opinion hotspots in Russian speaking countries and areas. However,there are only a few multimodal emotion corpora in Russian,which limits the further development of Russian emotion analysis technology. Based on the analysis of the related research and emotion classification methods of multimodal emotion corpus,this paper develops a scientific and complete tagging system,which includes 11 items of information in utterance,space-time and emotion. In the whole process of corpus construction and quality control,this paper follows the principle of emotional subject and emotional continuity,formulates a strong operational annotation specification and constructs a large-scale Russian emotional corpus. Finally,it discusses the application of corpus in the analysis of emotional expression characteristics,the analysis of personality characteristics and the construction of emotion recognition model.

Keywords　Multimodal,Sentiment analysis,Corpus,Russian

1　引言

俄语情感分析技术能从海量的社交媒体中自动解读俄语区民众的情感倾向和情绪变化,这不仅有助于政府进行舆情监控和国际文化交流,同时也为了解各国民众对待跨国贸易政策的态度提供了大数据支撑。俄语作为东斯拉夫语支的重要语种,俄罗斯联邦、白俄罗斯、哈萨克斯坦和吉尔吉斯斯坦等国都把俄语作为官方语言。使用俄语的人口数量对中亚具有首屈一指的影响力,而这部分地区正是“一带一路”政策中亚部分的核心地带。因此及时和深入地了解俄语区民众和国家的舆情热点和动向,对于西北和东北的边界省份及周边地区,乃至“一带一路”倡议和国家的安全都具有积极的作用。海量的社交文本分析耗时费力,不可能完全依靠人工来完成,这时就需要借助人工智能和自然语言处理技术,而情感分析技术正是自动从文本中提取用户观点和情感倾向性的重要方法,其能高效地从大数据中挖掘更多有价值的用户信息。

随着互联网技术的不断发展,大量网民情感和观点的载体不仅局限于单纯的文字,更多的是依靠语音或视频的形式呈现。如VK(VKontakte)、twitter、微信和微博等社交媒体每天会产生大量的语音和视频信息,并逐渐成为了人们日常

到稿日期:2020-09-10　返修日期:2021-03-26
基金项目:教育部人文社科青年基金项目(18YJCZH208);国家自然科学基金(61806038,61772103)
This work was supported by the Ministry of Education Humanities and Social Science Project(18YJCZH208) and National Natural Science Foundation of China(61806038,61772103).
通信作者:徐琳宏(qingniao1203@163.com)

沟通和交流的方式。此外，疫情的出现导致很多企事业单位采用在线的互动软件办公，这也会形成海量的以音频和视频形式表达意见和评论的数据。多模态情感分析就是融合文本、语音和视频等多个通道的信息，完成情感分析和观点挖掘的技术。多模态的加入使情感的表达更加丰富和立体，也更加准确和生动。通过辨别语音中的语气、语调以及说话人的肢体动作和面部表情，可以进一步提高情感分析模型的准确率，这也是目前的一个研究热点。因此，多模态情感分析不仅推动了情感分析技术的理论发展，同时也具有巨大的社会效益。

现有的情感分析模型大多是以深度学习和机器学习为基础的，而这类方法和模型的效果通常与语料的标注规模和标注质量密切相关。目前国外的多模态语料库还是以英语为主，分为会话形式和独白形式两种。俄语的多模态情感语料库还较少，这在很大程度上制约了俄语情感分析技术的进一步发展。针对目前俄语情感分析资源匮乏的现状，本文构建了一个俄语的多模态情感分析语料库，并在此基础上探讨了该语料库的应用场景。本文第 2 节梳理了多模态情感语料库建设的相关工作；第 3 节介绍了多模态情感语料的标注体系；第 4 节展示了语料库构建过程和相关的统计结果；第 5 节探讨了语料库的多种应用；最后总结全文并展望未来。

2 相关工作

不同语种间的多模态语料库建设在标注体系和标注规范等多个方面是相通的，具有一定的借鉴价值。目前资源最为丰富的是英语语料库，因此本文从英语多模态语料库和俄语多模态语料库两个方面来介绍相关的研究工作。

前些年多模态情感分析受到数据获取和计算能力的制约而发展缓慢，相关语料库的建设工作也起步较晚，但近几年其逐步成为了一个研究热点。从内容的角度看，多模态语料可以分为会话形式和独白形式两类。会话形式一般指同一样本中包含两个或者多个说话者，他们之间具有一定程度的交互性。较早的会话式多模态情感语料库是 IEMOCAP 数据集，该数据库由南加州大学 SAIL 实验室收集。它包含视频、语音、面部动作捕捉和文本几个模态，总时长约 12 h。由 10 位演员即兴创作和通过剧本表达情感，共生成了 10 000 个句子[1]。除了通过专业演员表演的方式获取语料来源外，另一种常见的会话式多模态语料是从情景喜剧中获取的。例如，Bertero 等以两个英文情景剧为数据源，根据字幕分割视频，分别标注了《生活大爆炸》中 1 589 个场景的 43 672 条语句和《宋飞正传》中 2 267 个场景的 45 734 条语句。它以情景剧中的背景笑声为区分，将每个语句分为幽默和非幽默两类，并采用 CNN，LSTM 或 CRF 进行建模，抽取音量、语速、声调、Ngram、句子长度、形容词比例、说话人变化以及说话频率等特征进行识别[2-4]。此外，Poria 等选择情景剧《老友记》为数据源，将文本情感语料集 EmotionLines[5] 改进成多模态数据集 MELD(Multimodal EmotionLines Dataset)。标注的模态有文本、语音和视频，共包含 3 407 个场景、13 708 条语句。标注的属性包括发言人、情感、话语 ID、季、场景、开始时间和结束时间等。在此数据集上，Poria 等分别使用 CNN，LSTM 和 DialogRNN 等方法来建模，发现多模态融合有助于提高情绪识别的准确性[6]。

除了会话式多模态情感语料外，独白形式的多模态情感语料库是另外一大类比较常见的表达形式，该类型语料中只包含单个说话者的情感表达或对某事物的评论，该数据大多来源于社交媒体上的视频评论，如 YouTube 等。早在 2013 年，Wollmer 等从 YouTube 下载视频电影评论作为数据源，构建包括文字、语音和视频 3 个模态的英文语料库 ICT-MMMO(Multi-Modal Movie Opinion)，它包括 370 个电影评论视频，其中包含 228 条正面、23 条中立和 119 条负面评论，单个视频的时长为 1～2 min，以口语表达形式为主[7]。Poria 等也从 YouTube 上收集了 47 段产品评论的视频，涉及的产品包括香水、牙膏、化妆品和相机等多个种类，共 300 句。并在实验中使用 SVM，ELM 和 ANN 等模型进行情感极性分类，结果表明加入声音和视频后分类的准确率更高[8]。除了上述早期的、规模相对较小的多模态语料，目前 CMU-MOSEI (CMU Multimodal Opinion Sentiment and Emotion Intensity) 是规模最大的独白式多模态情感语料库，其前身为 CMU-MOSI 语料库[9]，共包含来自 1 000 个不同的演讲者和 250 个主题的23 453 个带注释的视频片段，每个片段都包括文字、语音和视频 3 种数据模态。采用亚马逊众包的方式进行标注，标注后发现各类情感的比例差别较大。文献[10]采用 Graph-MFN 等方法进行建模，发现加入语音和视频后，情感分类的 F1 值有一定幅度的提高。多模态语料库除了语音和视频模态，也有些眼球追踪等姿势和动作相关的语料库[11]。

与丰富的英语语料资源相比，俄语的多模态情感语料库相对匮乏。目前具有一定规模的是 RUSLANA(RUSsian LANguage Affective speech database)和 RAMAS(Russian Acted Multimodal Affective Set)。RUSLANA 是独白形式的多模态语料库，共包括 3 660 条语句，由 12 位男性说话者和 49 位女性说话者尽量以无情感的方式[12]录制完成，语料库的情感分类采用激活评估轮[13]。另一个语料 RAMAS 是一个会话式的语料库，它收集了时长大约 7 h 的特写视频，这些视频包括受试者面部、语音、运动以及生理信号等模态特征。由 10 位演员扮演互动式场景，共包含 581 个视频，视频长度从二十几秒到几分钟不等[14]。情感表达需要通过肢体、表情、声音和文字等多种形式输出，只有综合多个模态才能更加全面和生动地分析和理解情感。针对俄语多模态情感语料资源较少的现状，本文尝试构建了一个俄语的多模态情感语料库，可用于进一步探索俄语情感分析的特点，生成或训练情感自动分类模型。

3 俄语多模态情感语料的标注体系

本文中的多模态情感语料库以俄语情景喜剧为数据源，通过分割每个发言人的话语片段，来完成从视频到文字、声音和图片的转换，并采用人工标注和自动标注相结合的方式，标注话语的情感、场景和发言人等信息。制定良好的标注体系是保障建设过程顺利进行的基础，更是保证语料库质量的关键，因此本文的标注体系在注重选择合理的标注粒度、确定恰当的标注框架和完善适宜的分类方法的同时，还尽量平衡标注效率和标注一致性。

3.1 标注单元及标注框架

语料的标注单元是某个发言人的一段话语。标注单元以俄文语句为基础，对同一发言人的连续相同情感的短句进行酌情合并，对较长的语句以语义完整为原则做适当的分割，例如，剧集开始和结束的独白部分大多是由 7～8 个短句组成的长句，这时会按短句分割后进行标注。总之，标注单元的合并和分割都会遵循不跨发言人和不跨场景的原则，即不会合并不同发言人或场景的话语，这样才能够保证在后续分析发言人性格特点和情感倾向时，数据的准确性和客观性。

标注框架包括内容信息(ContentInfo)、时空信息(SpaceTimeInfo)和情感信息(EmotionInfo)三大部分，其中内容信息包括俄语原文的文本信息、语音信息和图像信息以及中文译文，这部分主要从原始视频中自动获取；时空信息包括开始时间、结束时间、发言者和场景，此部分信息可以通过半自动的方式获取；情感信息包括情感类型、情感极性和内心情感，其中情感类型和极性是必选项，而内心情感为可选项，且第三部分信息需要大量人工标注才能完成。俄语多模态情感语料库的标注框架由下文的 11 元组构成：

(ContentInfo(RT, RV, RI, CT), SpaceTimeInfo (ST, ET, SP, SC), EmotionInfo (EM, SE, IF))

其中，各部分的具体含义如下。

内容信息部分：RT(Russian Text)是剧中的原始俄语文本；RV(Russian Voice)是剧中的俄语原声音频；RI(Russian Images)是从一段话语视频中截取若干图像构成的集合；CT(Chinese Text)则是针对原始俄语文本的中文译文。

时空信息部分：ST(Start Time)和 ET(End Time)分别是话语的开始时间和结束时间，它们均以当前剧集的起始点为时间参考点；SP(Speaker)是话语的发言人，大部分发言人为情景剧的主演，主要取值为阿列克斯、罗马、叶卡捷琳娜、阿尼娅、安纳托利 5 位，其他配角也会按名字标出，但所占的比例较小；SC(Scene)是场景信息，它不仅包含事件发生的地点，也可能描述事件发生时发言人之间的关系，没有固定的取值，具有一定的主观性。

情感信息部分：EM(Emotion)指情感分类，取值为喜、怒、悲、恐、愧和中性，每个大类还会划分成更细的情感类别，这部分会在 3.2 节中的情感分类中详细介绍；SE(Sentiment)指情感极性，分为积极、消极和中性 3 类；IF(Inner Feelings)代表内心的实际情感。情感表达过程中会出现发言人试图掩饰内心真实情感的情况，即字面语言表达的情感和人物真实的内心情感会出现偏差，遇到这种情况时 EM 部分主要标注字面情感，IF 部分标注发言者内心的真实情感。大部分话语中情感的表达都是所见即所得，从字面含义和语音语调就可以获悉，因此 IF 部分是可选的。

3.2 情感分类

情感信息部分是整个标注体系的重点，其中情感类别的划分关系到语料资源的质量，也是整个语料库建设的基石，分类标准是需要在保证各情感类别的互斥性的同时兼顾整体情感类别的覆盖面。为了合理规划情感的分类，本文参考了国外多个英文和俄文的多模态情感语料库，将它们的情感分类方法汇总，如表 1 所列。

表 1 多模态情感语料库的情感分类

Table 1 Emotion classification of other multimodal emotion corpus

类型	语料库名称	情感分类	极性分类	语种	来源
对话形式	RAMAS	幸福、悲伤、愤怒、恐惧、厌恶、惊奇和中性	无	俄语	演员表演
	MELD	欢乐、悲伤、愤怒、恐惧、厌恶、惊奇和中性	欢乐是积极情感，惊奇可能积极也可能消极，其余为消极	英语	老友记剧集
	IEMOCAP	幸福、兴奋、悲伤、愤怒、恐惧、沮丧和惊讶	无	英语	演员表演
非对话形式	RUSLANA	幸福、悲伤、愤怒、恐惧、惊奇和中性	无	俄语	演员表演
	CMU-MOSEI	幸福、悲伤、愤怒、恐惧、厌恶、惊奇和中性	高度否定、否定、弱否定、中性、弱肯定、肯定和高度肯定	英语	用户独白
	ICTMMO	无	强否定、弱否定、中性、弱褒义和强褒义	英语	电影评论

由表 1 可知，大部分多模态情感语料库的情感分类都包含幸福、悲伤、愤怒和恐惧几种情感，根据每个语料来源的不同，会有个别情感的增加。而极性分类大多是依据情感倾向分类，不同语料库在极性方面大致分为积极、消极和中性 3 种主要类别，根据情感强度的不同可能细化为 3 类、5 类和 7 类不等。本文根据情景喜剧自身的特点以及上述的分类方法，制定了语料库的情感标注类别，一方面力求情感的类别涵盖更多的标注样例，尽量使每个标注单元都有准确的类别，另一方面控制情感类别的数量以及各类之间的包含关系，保证情感类别的互斥性。

综合考虑以上因素，本文将情感分为喜、怒、悲、恐和愧 5 大类别，同时每个类别按情感强度划分为 15 个小类，分别为得意、喜爱、快乐、惊喜、愧疚、抱怨、厌恶、生气、愤怒、沮丧、悲伤、担忧、害怕、惊慌和恐惧。每种情感都隶属于 5 大类别中的一种。需要说明的是，5 大类别中不包含“惊”，是因为该类别的情感包含惊喜和惊恐两种可能，极性可积极也可消极，因此本文把“惊”类情感根据极性的不同分别归入“喜”和“恐”两个类别中。不同等级情感的类别和关系如图 1 所示。

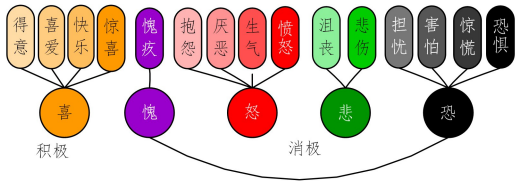


图 1 语料库的情感分类

Fig. 1 Emotion classification of corpus

内心情感的分类方法与情感的分类方法相同，但内心情

感指的是字面表达情感和人物真实的内心情感不一致时，标注发言人真实的内心的情感。这类情感一般不能通过文字直接表达，需要一定的背景知识和逻辑推理才能识别。例如，赌牌类的场景中，发言人话语的字面意思都是中性情感，但因为赌注较大，人物内心其实非常恐惧，这类内心情感的识别需要有赌牌和赌注价值等常识，而该标注项有助于情感知识图谱的构建和情感常识的挖掘，进而为深层次的情感识别服务。

4 语料库的标注规范及统计分析

根据第 3 节确立的标注体系，我们首先确定了合适的数据来源、提取和预处理 3 种模态的数据，然后制定了严谨和实用的标注规范，最后选择经验丰富的标注人员，来构建会话式的俄语多模态情感语料库。

4.1 数据来源及预处理

俄语情景剧通常是以矛盾和冲突的形式推进剧情，不仅情感信息丰富，用词比较口语化，而且一个场景下一般有 2～3 人发言，能在一定程度上展现当地人的真实生活，比较适合多模态情感语料库的构建。本文选择非常流行的俄语情景剧《我是如何成为俄罗斯人的》的 1～10 集作为数据源，该情景剧音频信号比较清晰、噪声较少，视频信号多为人物近景，方便捕获人物细节的肢体和表情信息，并且在情节设计上冲突和起伏明显，具有多种类型的情感信息，符合多模态情感语料库对数据的需求。

数据的预处理过程是按标注单元切割视频，以获取对应的文本、语音和图像 3 种模态的信息，为后续情感标注做准备。为此，我们首先通过 3ears[1)] 网站获取所有剧集的俄文脚本数据，即发言人会话的文本信息。其次，由于该情景剧没有对外发布各集对应的软字幕文件，无法获取时间信息，因此本文邀请熟悉俄语的相关领域专家为每集视频制作时间轴和字幕文件，并按标注单元进行划分，标记出每个单元的起始时间和结束时间。然后，利用 ffmpeg 软件分割视频，再进行视频和语音的转换。最后，借助 OpenSmile 工具包，处理切割后的每个视频文件，并从中提取图像信息。

4.2 标注规范

标注规范包括标注原则和标注流程两部分，用于控制整个语料库建设的全过程和标准化具体操作，它能在语料标注前、标注中和标注后 3 个阶段有效地规范相关人员的实际操作。首先，开始标注前需要根据标注体系制定相应的标注原则，帮助语料库建设人员更加准确地识别情感，理解各类情感间的差别。其次，语料标注过程中要合理分组，关注歧义样本的处理过程，通过一致性检测评估和衡量语料库的质量，修正出现偏差的部分。

4.2.1 标注原则

标注原则是相关人员在正式开始标注前需要认真学习的标注说明，它可以规范化多个标注者对各类情感的理解，提高标注结果的一致性。根据语料的特点和试标注的结果，本文采用的标注原则包含情感主体原则、连贯性原则和表达方式原则。

情感主体原则指标注的情感是话语发出人的情感，即发言人的情感，而不是话语收听者体会的情感，更不是观看剧集的观众的情感。

连贯性原则指一个场景下的对话具有时序性，前后互相关联，不能将它们割裂开，即发言人的情感通常具有前后的关联性，前一句是某类情感，后一句是该类情感的概率更大。

表 2 中的对话来源于第 5 集中阿尼娅和阿列克斯捉弄罗马的场景，罗马的情感从误认为阿列克斯自杀开始就进入持续的惊慌状态。这里的第 5 句是人名，根据字面含义是中性情感，但是根据情感的连贯性原则，发言人仍处于惊慌的情绪中，该句的语音和图像信息也从侧面证实了这一点，因此情感应该标注为“惊慌”。

表 2 连贯性原则示例

Table 2 Examples of coherence principles

序号	发言人	脚本	情感
1	罗马	Алекс! Алекс! Алекс! Ты слышишь меня, Алекс?（阿列克斯，阿列克斯，阿列克斯，听得到我说话吗？阿列克斯？）	惊慌
2	阿尼娅	Пульс нитевидный!（他的脉象很微弱！）	惊慌
3	罗马	Алекс，слышишь меня? Это шутка была，паспорт твой у меня，Алекс，паспорт твой，это шутка!（阿列克斯，你能听到吗？你的护照在我这儿，玩笑而已！）	惊慌
4	阿列克斯	Будь ты проклят.（我会诅咒你的）	愤怒
5	罗马	Алекс!（阿列克斯）	惊慌

表达方式原则指每种情感都有一些特定的表达方式，如赞扬多与“喜”类相关，批评、谩骂和讽刺则一般与“怒”类相关，恳求和催促经常与“恐”类相关。这些表达方式通常比情感更容易识别，因此有些情况下表达方式能够辅助情感的识别，提高标注的准确率和一致性。但需要说明的是，并非每种情感都借助某种表达方式来体现，而且情感与表达方式之间也不存在一一对应的关系，它只用于提示标注者在话语中可能存在的某种情感。例如，恐吓的表达方式可能是愤怒引发的，也可能是恐惧引发的。

4.2.2 标注流程及质量监控

在语料库建设过程中采用两人一组标注、五人合作互助的方法处理歧义。前期，一个话语由两个人同时标注，如果标注结果不一致，则由第三个标注者裁定，最终对于还是不能达成一致的疑难语料，由五人小组开会讨论来决定其最终的情感。少量争议较大的话语可能是因为目前的情感分类没有覆盖到，这种话语标注成“未知”。标注人员经过标注体系和标注规范的培训后，先进行试标注，在一致性达到一定水平后便可以进组进行正式标注。

标注的一致性是评价语料库建设质量的关键指标，本文采用 Kappa[15] 一致性检测，表达式如下：

$$Kappa=\frac{P_o-P_e}{1-P_e} \tag{1}$$

其中，P_o 是正确分类的话语数量之和除以总话语数，P_e 为按类别累加标准话语的数量与预测话语的数量的乘积，然后除以话语总数量的平方。随着标注人员对标注规范的不断深入理解，在语料库建设过程中，一致性会逐步提高。俄语多模态

1) https://3ears.com

情感语料库在建设初期，其 Kappa 值就已达到 78.51%，可见本文构建的语料库在标注一致性上是可靠的。

4.3 统计结果

本文构建的俄语多模态语料库共包含 181 个场景和 3278 条话语，涉及的发言人有 82 名，其中主要发言人有 5 个，分别为 Алекс（阿列克斯）、Анатолий（安纳托利）、Екатерина（叶卡捷琳娜）、Ром（罗马）和 Аня（阿尼娅）。各模态的一般统计信息如表 3 所列。

表 3 语料库的统计信息

Table 3 Statistical information of corpus

模态	指标	最大	最小	平均
文本	语句长度	26	1	5.4
	词汇频率	559	1	3.8
	每集话语数	387	258	328
音频	时长	28.927	0.042	2.839
	基频	490	0	52
	能量	$5.98*10^{-2}$	$1.68*10^{-5}$	$2.225*10^{-5}$
图像	饱和度(0～360)	236	9	98
	色调(0～255)	147	8	74
	亮度(0～255)	253	3	101

表 3 列出了语料库中各模态的信息，其中语句的平均长度为 5～6 个单词，句长都比较短，符合情景剧口语化的特点。语料共有词汇 4609 个，词汇信息比较丰富，涵盖了大部分常用的俄文词汇，使语料库具有良好的泛化性。个别词汇的使用频率较高，其中包括人名和代词等，每集的话语数基本集中在 300 句左右。语料中多是口语化的短句，因此音频的平均时长大约为 3 s。尽管其中个别音频时长有 20 s 左右，但大多是因为发言者在叙述过程中存在较长时间的思索和停顿，所以音频语句中包含的单词数量并不多。

除了上述的基本信息外，本文还分别统计了该语料库中每个情感大类和小类的占比，如图 2 所示。图 2 中，“中性”类别的话语共计 1495 个，占总话语数量的 46%，而“悲”和“愧”类别的话语最少，均为 48 条，各自仅占总数量的 1%左右。大约 25%的语句为“怒”类，是情感类别中占比最高的，此类情感按强度可大致分为“抱怨”“厌恶”“生气”和“愤怒”，其中“生气”类别的话语最多，其次是“愤怒”类别。“喜”和“恐”两类数量大致相当，占比分别为 12%和 14%，“喜”类中“快乐”类的话语最多，“恐”类中“惊慌”类占比最大，而强度较大的“恐惧”类别占比最小。

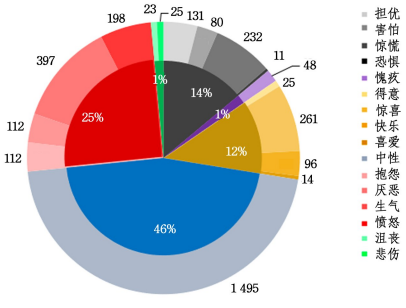


图 2 各类情感语料的占比

Fig. 2 Proportion of various emotional corpus

由图 2 可以看出，虽然该情景喜剧的情感元素比较丰富，但大部分的话语仍为“中性”，这正符合人们日常生活的状态，即多数时间处于平静状态，处于情感状态的比例相对较小，尤其是比较强烈的情感状态则更不常见。如“恐惧”类别的数量很少，它是一个强度较大的情感状态。“怒”类情感约占情感类别话语数量的一半，这主要是因为情景喜剧通常采用制造夸张冲突的方法来达到一定的喜剧效果。从情感极性的角度看，“喜”属于积极情感，其他均为“消极”情感，因此积极情感大约占 12%，中性为 46%，消极情感为 42%。情景喜剧中通过人物的冲突和误会等使观众感受喜悦，但并不是将人物本身的喜悦共情给观众，因此积极情感偏少，消极情感更多。

5 语料库的应用

俄语多模态情感语料库在情感分析的许多方面都具有广泛应用，除了可以构建更加精准的情感识别模型外，还可以从语料库中统计和分析俄语情感表达的特点以及对发言人的性格建模等。

5.1 解析情感表达的特点

每种情感的表达都有其自身的特点，掌握各类词汇使用上的特点和语音特征不仅有助于情感识别，同时也能大幅提升文本和语音生成的效果，例如，在构建智能聊天系统时，针对特定情感类别加载相应的音频合成模型，可以使机器的输出更加贴近人类的表达。图 3 给出了怒、喜、恐和悲 4 种主要情绪的词云分布图。

(a)怒 (b)喜

(c)恐 (d)悲

图 3 4 种情感的词云分布

Fig. 3 Distribution of words cloud by four emotions

图 3 给出了每种情感的词云分布，每个词云给出了该类情感下的常用词汇，图中词汇的字号越大，说明词汇的使用频率越高。从图 3 可以看出，各类情感的词汇分布差异较大，其中包含了很多带有明显情感倾向的词汇。例如，“怒”类中包含 урод（怪物）、надоело（厌倦）、аферист（骗子）和 Заткнись（闭嘴）；“喜”类包含 люблю（爱）、Молодец（做得好）和 нравится（喜欢）；“恐”类包含 позвонить（呼叫）、Помогите（救命）、убьет（杀死）和 скорая（救护车）。但不是所有的情感类别都有明显的词汇特征，如“悲”类中很少包含具有情感倾向的词汇，更多的是通过表情和声音来表达沮丧和悲伤。

此外，从语音的角度分析，各类情感的特征也比较明显。“恐”和“怒”类情感的能量值分别为 0.0035 和 0.0030，相比其他情感能量值，该类情感的能量值都比较高，因为这两类都属于唤醒度比较高的情感，一般需要较高的能量值表达。而“悲”类的能量值较低，仅为 0.0024，这可能与悲伤类情感多通过低沉幽怨的语调表达有关。同时愧疚类的基频是最低

的，只有 41.46，而其他类别的基频都在 60 左右，差异较大。了解上述语音特征有助于更高效地识别情感类别和生成质量更好的情感语句。

5.2 分析人物性格特征

情感的表达与情感主体的性格之间存在密切的关系，面对同一事件不同人的情感体验也是千差万别的，仅仅对文本、音频和视频进行建模，难以全面和准确地把握和识别情感，因此还需要考虑情感主体的性格特征。作为情感的发出者，不同的情感主体因为生活的环境和知识结构等方面存在差异，表达情感的方式也存在较大区别，因此对每个情感主体建模是情感分析的重要环节。利用带标注的多模态情感语料库可以对情感主体建模，例如，以情感为基础分析人物的性格特点或者在提取音频和视频特征的基础上构建人物的语音和语调特点等。图 4 给出了从语料库中统计 5 个主要发言人的情感比例。

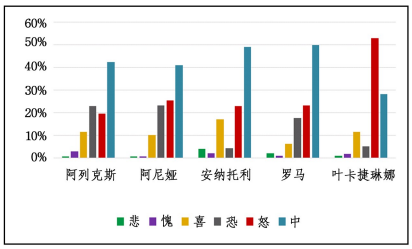


图 4 发言人话语的情感比例

Fig. 4 Emotional proportion of speaker's utterances

图 4 中，横轴为语料库中的 5 个主要发言人，纵轴为各类情感所占的比例。由图可以看出，前面 4 个发言人的话语均以中性表达为主，而发言人叶卡捷琳娜的话语则以愤怒为主要情感类别，且愤怒类别所占比例几乎是中性的两倍，其他情感类别所占的比例较少，可见该人物的性格属于易怒的类型。相比而言，发言人阿列克斯的话语中，各类情感的比例相对均衡，人物属于性格稳定型。5 个人物中，“喜”类情感占比最少的是罗马，由于剧情和人物塑造的需要，该发言人的情感以恐惧和愤怒为主。恐惧类情感占比最少的人物是安纳托利，同时他也是“喜”类占比最高的人物，因为该发言人是个一言不合就举猎枪的富商，张扬的性格特点本身就比较鲜明。上述对发言人的分析表明，依托质量较好的多模态情感语料库，可以准确地对情感主体的性格特征进行建模，进而构建和训练更加精确的情感识别和生成模型。

此外，语料库中除了标注话语的字面情感，还标注了发言人的内心情感，这也可以用于情感主体的深层建模。例如，发言人罗马经常会出现字面情感和内心情感不一致的情况，他通过快乐、生气和厌恶等情感来掩饰内心的恐惧，导致出现表里不一的情况，而其他人物则很少出现这类现象。上述统计信息有助于更加全面地完成情感主体的性格建模，体现情感表达的多面性。

5.3 构建情感分类模型

生成或训练准确率高的情感分类模型是目前情感分析的重要研究方向，目前在英文语料上构建的分类模型相对较多，但近几年针对其他语种进行情感分析的相关研究也越来越多，中文方面既有探索跨领域[16]情感分析的研究，也有专门探讨话题类情感分析的工作[17-18]，还有部分研究者依赖常识完成情感分析[19]。俄文方面有在新闻语料中使用 LSTM[20]和俄文评论数据上利用卷积神经网络进行建模的研究[21]。上述工作大多是依托文本信息数据为基础搭建模型，随着情感分析研究的深入开展，单纯依靠文本建模的方法因为信息量有限而遇到了瓶颈，综合文本、音频和图像的多模态情感分析是目前情感分析研究的新途径，也是对进一步提高模型识别的准确率、泛化性和鲁棒性的新探索。Zadeh 等合作构造了 MFN[22]和 DialogRNN[23]等多模态情感分析模型，并分别在 CMU-MOSI，MOUD，YouTube，ICT-MMMO 和 IEMOCAP 等多个数据集上进行了三分类和多分类的情感识别实验。结果发现，加入语音和视频特征后的分类结果均比单纯依靠文本的分类效果有不同程度的提高，这也说明包含音频和视频信息的多模态的情感语料能有效提升情感分类模型的效果。

情感表达的方式本来就多种多样，语言文字只是其中的一种形式，很多情绪是借助音量、音调、语速、表情甚至手势来表达的，单纯的文本形式容易错失很多情感信号。图 5 给出了俄语情感语料库中一段父女吵架的对话，如果仅仅从文本角度进行情感分析，这段对话也可以理解为父亲对女儿的交友状况的关心和询问。但通过急促和高分贝的语音信息和人物面部表情的图像特征，可以分析出这段对话是一场争吵，两个人物表达的都是明显的愤怒情绪。由此可见，本文构建的俄语多模态语料库能规避单纯依靠文本信息建模的缺陷，并有效提升俄语情感分析模型的准确率，丰富情感特征识别的深度和广度。

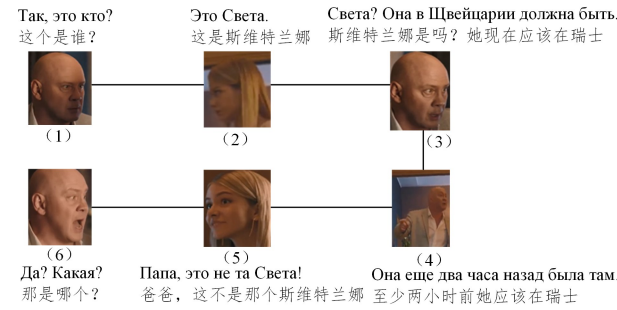


图 5 音频和视频模态的作用示例

Fig. 5 Examples of audio and video modes

结束语 多模态情感识别是目前情感分析领域的研究热点，而俄语的多模态情感分析工作目前还处于起步阶段，急需各种类型语料资源的支撑。本文对当今多模态情感分析语料的分类方法和构建过程做了详细的论述，在此基础上制定了一套详实全面和可操作性高的标注体系和标注规范，以俄文情景剧为原始材料，标注内容、时空和情感 3 部分的 11 项信息，构建了包括文本、语音和图像 3 个模态的俄语情感语料库，并进一步探讨了语料库在解析情感表达特点、分析人物性格特征和构造情感识别模型等多个方面的应用。本文构建的语料库是目前为止第一个以情景剧为原始数据的会话式俄语多模态情感语料资源。从数据来源看，情景剧具有场景模拟真实和剧情发展连贯的特点。不同于现有的俄语多模态语料库 RAMAS 和 RUSLANA，它们的数据来源均来自于雇佣演员的模拟表演。情景剧在语音模态上包含很多真实的噪声，

更好地还原了人们的原始生活场景。此外，前后剧情关系紧密，有许多铺垫和伏笔，能更多地体现出情感的迁移和变化规律。从模态和标注粒度上看，我们的语料库包含了文本、语音和视频3个模态，情感标注的粒度基本细化到每一句话，与会话式语料库RAMAS比，其增加了文本模态，且语料分割的粒度也更细。

虽然俄文情感语料的应用广泛，但本文构建的语料库也存在一些不足，例如，标注的范围和语料数量都有待提高，在标注的过程中也需要进一步完善质量监控等。此外，语料库中的统计结果是从单一喜剧语料中总结得出的，不一定覆盖真实生活的所有情感，如果另选一个基调比较低沉忧郁的影视剧，可能标注出的悲伤类别会更多，这也进一步提示我们应该拓宽标注数据来源，在后期的俄语多模态语料的建设过程中应考虑更多类型的影视剧，或者社交媒体上的用户视频评论等，进而丰富语料库的情感类别，平衡情感数量。在今后的研究工作中，除了完善语料库，我们还需要进一步挖掘语料库的应用场景，计划搭建俄语的情感模型，加入更多的针对发言人和话语时序的特征，加入俄语发音的独有特征，融合多种模态的信息，提高模型分类的准确性。

参 考 文 献

[1] BUSSO C, BULUT M, LEE C, et al. IEMOCAP: Interactive emotional dyadic motion capture database[J]. Journal of Language Resources and Evaluation, 2008, 42(4): 335-359.

[2] DARIO B, PASCALE F. Deep Learning of Audio and Language Features for Humor Prediction[C]// Proceedings of the 10th International Conference on Language Resources and Evaluation. LREC, 2016: 496-501.

[3] BERTERO D, PASCALE F. Predicting humor response in dialogues from TV sitcoms[C]// IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). 2016: 5780-5784.

[4] BERTERO D, FUNG P. A Long Short-Term Memory Framework for Predicting Humor in Dialogues[C]// Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2016: 130-135.

[5] HSU C C, KUO C C, CHEN S, et al. Emotionlines: An emotion corpus of multi-party conversations[C]// Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC). 2018: 1597-1601.

[6] PORIA S, HAZARIKA D, MAJUMDER N, et al. MELD: A Multimodal Multi Party-Dataset for Emotion Recognitionin Conversations[C]// Annual Meeting of the Association for Computational Linguistics. 2019: 527-536.

[7] WOLLMER M, WENINGER F, KNAUP T, et al. YouTube Movie Reviews: Sentiment Analysis in an Audio-Visual Context [J]. IEEE Intelligent Systems, 2013, 28(3): 46-53.

[8] PORIA S, CAMBRIA E, HOWARD N, et al. Fusing audio, visual and textual clues for sentiment analysis from multimodal content[J]. Neurocomputing, 2016, 174(JAN. 22PT. A): 50-59.

[9] ZADEH A, ZELLERS R, PINCUS E, et al. Mosi: Multimodal corpus of sentiment intensity and subjectivity analysis in online opinion videos[J]. arXiv: 1606. 06259, 2016.

[10] ZADEH A, LIANG P P, POIRA S, et al. Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph[C]// Annual Meeting of the Association for Computational Linguistics. 2018: 2236-2246.

[11] WANG X M, ZHAO X B. Survey of Construction and Application of Reading Eye-tracking Corpus[J]. Computer Science, 2020, 47(3): 174-181.

[12] MAKAROVA V, PERTRUSHIN V A. RUSLANA: a Database of Russian Emotional Utterances[C]// International Conference on Spoken Language Processing. 2002: 1-4.

[13] COWIE R, DOUGLAS E, TSAPATSOULIS N, et al. Emotion recognition in human-computer interaction[J]. IEEE Signal Processing Magazine, 2002, 18(1): 32-80.

[14] PEREPELKINA O, KAZIMIROVA E, KONSTANTINOVA M. RAMAS: Russian Multimodal Corpus of Dyadic Interaction for Affective Computing[C]// SPECOM. 2018: 1-6.

[15] CARLETTA J. Assessing agreement on classification tasks: the kappa statistic[J]. Computational Linguistics, 1996, 22(2): 249-254.

[16] LI X, LI Y, WANG S G. Text Similarity Calculation for Text Sentiment Clustering[J]. Journal of Chinese Information Processing, 2018, 32(5): 97-104.

[17] WANG J C, XU Y, LIU Q Y, et al. Dialog Sentiment Analysis with Neural Topic Model[J]. Journal of Chinese Information Processing, 2020, 34(1): 106-112.

[18] WANG K, PAN W, YANG B H. OTSRM-Based Approach for Sentiment Evolution and Topic Analysis[J]. Journal of the China Society for Scientific and Technical Information, 2019, 38(5): 534-542.

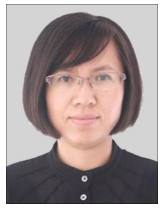
[19] YANG L, ZHOU F Q, LIN H F, et al. Sentiment Analysis Based on Emotion Commonsense Knowledge[J]. Journal of Chinese Information Processing, 2019, 33(6): 94-99.

[20] SAKENOVICH N S, ZHARMAGAMBETOV A S. On one approach of solving sentiment analysis task for Kazakh and Russian languages using deep learning[C]// International Conference on Computational Collective Intelligence. 2016: 537-545.

[21] GALINSKY R, ALEKSEEV A, NIKOLENKO S I. Improving neural network models for natural language processing in Russian with synonyms[C]// IEEE Artificial Intelligence and Natural Language Conference. 2016: 1-7.

[22] ZADEH A, LIANG P P, MAZUMDER N, et al. Memory Fusion Network for Multi-view Sequential Learning[J/OL]. Association for the Advancement of Artificial Intelligence, 2018: 5634-5641. https://arxiv.org/abs/1802.00927.

[23] MAJUMDER N, PORIA S, HAZARIKA D, et al. DialogueRNN: An Attentive RNN for Emotion Detection in Conversations[C]// Association for the Advancement of Artificial Intelligence. 2019: 681-682.



XU Lin-hong, born in 1979, associate professor. Her main research interests include nature language processing and sentiment analysis.