云南大学软件学院期中课程报告

Midterm Course Report

School of Software, Yunnan University

**个人成绩**

|  |  |  |  |
| --- | --- | --- | --- |
| **序号** | **学号** | **姓名** | **成绩** |
| **1** | **20211060245** | **陈俊宏** |  |
| **2** | **20211120171** | **孔令高** |  |
| **3** |  |  |  |
| **4** |  |  |  |

学　　期: 　2023秋季学期

课程名称: 　大数据分析及应用

任课教师: 　何婧

报告题目: 基于多模态数据的情感分析

组长姓名: 　陈俊宏

联系电话: 　18288116722

电子邮件: 　chenjunhong@stu.ynu.edu.cn

完成提交时间：2023年11月27日

作业截止时间：2023年11月27日

**小组成员及分工**

**（工作量总和为100%，真实体现成员工作量，不能直接用平均值）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **序号** | **学 号** | **姓 名** | **工作内容** | **工作量（%）** |
| 1 | 20211060245 | 陈俊宏 | 算法分析，实验结果分析与文档制作 | 50% |
| 2 | 20211120171 | 孔令高 | 算法分析与文档制作，PPT制作 | 50% |
| 3 |  |  |  |  |
| 4 |  |  |  |  |
| 5 |  |  |  |  |
| 6 |  |  |  |  |

评分标准：

1. 选题。 10%

2. 综述完整，结构组织合理，内容丰富，文字表达清楚。40%

3. 报告格式规范，图表清晰。20%

4. 有基础算法实验代码及分析。10%

5. 有自己的观点。10%

6. 课堂汇报逻辑清楚，PPT展示完整。10%

基于多模态数据的情感分析

Different Ways to Implement Multimodal Sentiment Analysis

陈俊宏1 孔令高1

1（云南大学软件学院 昆明 650000）

摘要 本文对不同算法在多模态情感分析任务中的应用进行了比较研究。我们探讨了文本、图像和音频等多种模态的情感分析方法，并对不同算法的性能进行了评估。实验结果表明，多模态情感分析能够提高情感分析的准确性，并且不同算法在不同模态下表现有所差异。最后，本文展望并指出了未来可能的研究方向。

关键词 多模态情感分析；模态融合；文本；图像；音频；维度情感模型；特征提取；注意力机制

中图法分类号 TP391

0 引言

情感分析是自然语言处理领域的一个重要研究方向，旨在自动识别和提取文本中的情感信息。传统的情感分析方法主要基于文本模态，利用词汇、语法和上下文等信息来推断情感极性。然而，这种方法往往忽略了其他模态的信息，如图像和音频等，这些模态同样包含丰富的情感信息。因此，如何综合利用多种模态的信息来进行情感分析，成为了一个值得研究的问题。

1 相关工作

在情感分析的发展过程中，许多研究者用一种模态来进行情感分析。由于用单模态来进行情感分析时只能在该模态获得情感信息，在某些情况下有很多局限性。如图 1 所示，在对人物进行情感分析时，若仅仅考虑文本信息，会得到一样的结果，只有结合面部表情后才能得到正确的情感极性。

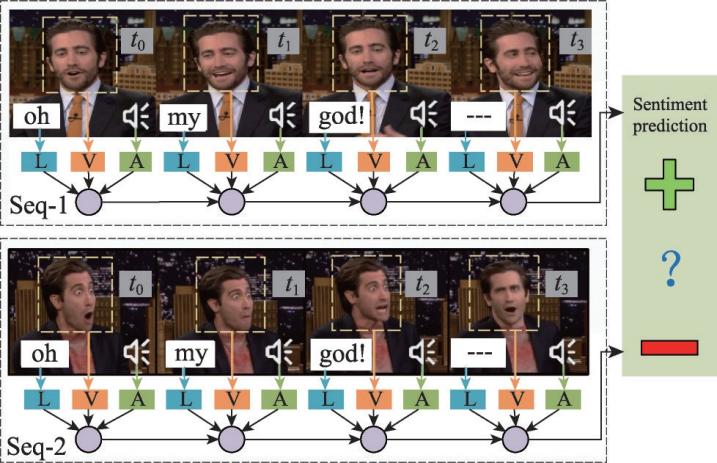


图 1 单一模态的局限性

在情感分析中，目前常用的信息有面部表情信息、文本信息和语音信息，也有一些研究者尝试用姿态、脑部信息来进行情感分析。多模态的情感分析是指由两种及两种以上的模态信息结合来进行情感分析。在特征提取阶段，多模态的情感分析与单模态的特征提取方法相同。利用多模态和单模态进行情感分析最大的区别就是在于多模态需要将单模态的信息进行融合，从而得到情感极性。结合现有文献，模态融合主要包括三种方法，分别是特征级融合、决策级融合以及混合融合。

本文对多模态情感分析的相关技术和方法进行归纳总结，并且对文中提到的算法进行对比分析，最后重点介绍了多模态融合技术并对现有问题进行总结。

2 单模态情感分析

如图2所示，根据模态不同分别对文献进行叙述。在FER中，现有算法多用传统方法与深度学习相结合的方法来进行情感分析，在数据集方面用GAN、迁移学习等进行扩充。

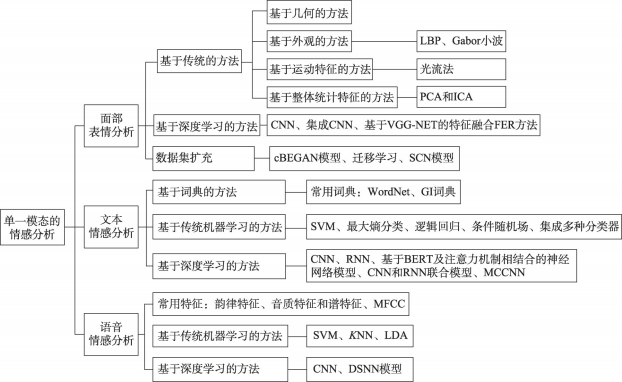


图2 单模态情感分析研究现状结构框图

在文本情感分析中，由于传统方法中情感词典受情感词数量和个数的限制，大多数研究者使用深度学习中的RNN、LSTM等模型来进行分析，同时加入注意力机制来提高分析效果；在语音情感分析中，多用深度学习的方法来进行分析，而难以采集到大量包含情感的语音数据是限制对其深入研究的主要因素之一。

由于从单模态中获得的信息量有限，想要进一步提高情感分析的准确度变得十分困难。因此尝试从多种模态中获取更多的信息进行情感分析来提高准确度。

3 多模态情感分析的框架

图3显示了一个多模态情感分析的框架。该框架包含两个基本步骤：分别处理单模态的数据和将处理后的数据进行融合。这两个步骤都很重要，如果单一模态的数据处理不好，会对多种模态的情感分析结果产生负面影响，而融合方式的性能不好会破坏多模态系统的稳定性。

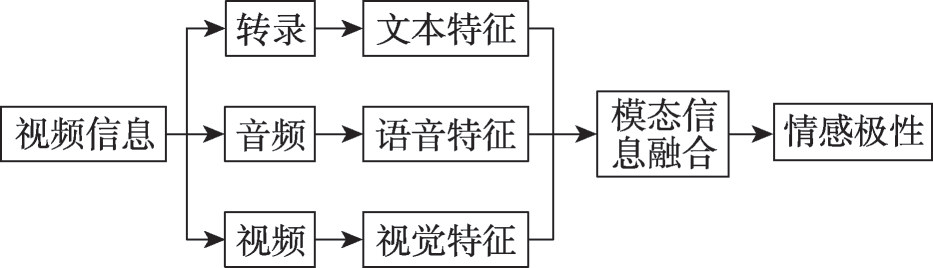


图3 多模态情感分析的框架

4 多模态情感分析研究

从特征融合的角度来看，多模态情感分析研究主要分为特征级融合、决策级融合和混合融合。

4.1 特征级融合

多模态情感分析中的特征级融合是一种策略，它通过拼接、相加等方式融合不同模态特征的向量表示，形成一维的向量。这种方法通常在多模态识别系统中使用，旨在从不同模态中提取特征并连接成单个高维特征向量。在特征级融合中，来自不同模态的特征首先被提取并处理，然后对特征信息进行综合分析和处理。这种方法有助于实现可观的信息压缩，有利于实时处理，且所提取的特征直接与决策分析有关，因此融合结果能最大限度地给出决策分析所需要的特征信息。例如，在情感分析中，可以通过多通道传感器获取交互者当前情感状态下的不同模态情感信号，如文本、音频和视觉特征。这些特征可以经过特征级融合，形成综合的情感特征向量，然后输入到分类器中，如多层感知器（MLP）和softmax层，以得出情感分类结果。

然而，特征级融合的局限性在于生成的高维特征向量，并且由于它以一种直接的方式融合了三种模态特征，不能对复杂的关系进行建模。此外，这种方法对通信带宽的要求较低，但可能会因数据丢失而使其准确性有所下降。

4.2 决策级融合

决策级融合对每个模态的特征进行独立建模与分析，将分析的情感类型概率结果融合为最终预测的情感结果。多模态情感分析中的决策级融合是一种高级别的融合方法，它涉及在融合之前，每个局部传感器相应的处理部件已独立完成了决策或分类任务。实质是按一定的准则和每个传感器的可信度进行协调，做出全局最优决策。与特征级融合不同，决策级融合是直接针对具体决策目标的，融合结果直接影响决策水平。这种方法必须从具体决策问题的需求出发，充分利用特征级融合所提取的测量对象的各种特征信息，采用适当的融合技术来实现。在多模态情感分析中，决策级融合可以按一定的准则将多个单传感器决策结果进行组合和协调，以获得更精确、更明确的决策结果。这种融合方法通常采用投票法、加权平均法等策略，根据不同传感器之间的可信度进行权重的分配，从而得到更加准确和可靠的决策结果。

总之，决策级融合是多模态情感分析中一种高层次的融合方法，它可以实现对多个单传感器决策结果的优化组合，提高决策的准确性和可靠性。但是虽然决策级融合方法可以针对不同模态数据选择最优的分类器模型，分类器的增加也会使得整体模型非常耗时和繁琐，且没有考虑到模态间的交互信息。

4.3 混合融合

多模态情感分析中的混合融合是一种综合方法，它结合了特征级融合和决策级融合的优点，以实现更准确和可靠的情感分析结果。混合融合首先对不同模态的特征进行特征级融合，然后使用这些特征进行独立的分类或回归任务，最后将各个模态的分类或回归结果进行决策级融合，以得到最终的决策结果。混合融合方法结合了特征级融合和决策级融合的优点，既能够提取到不同模态的特征信息，又能够实现对多个单传感器决策结果的优化组合，从而提高情感分析的准确性和可靠性。此外，混合融合还具有对不同模态特征的同步要求较低的优点，因为它在特征级融合后，各个模态的特征已经具有了一定的同步性，从而减少了不同模态特征之间的时间差异对情感分析结果的影响。

总之，混合融合是多模态情感分析中一种综合性的方法，它能够实现更准确、可靠的情感分析结果，并且在不同模态特征之间的同步要求较低。尽管基于混合融合的方法取得了较好的效果，但对于模态间的层次关联信息并没有进行充分的考虑。

5 多模态情感分析数据集

目前国内外多模态情感数据库大多来源于网络视频评论或人为制作，对于科研领域仍是半公开或者不公开的状态。由于模态选择的不同以及数据集的局限性，一些研究者会根据自己的需求来建立所需要的情感数据集。用于多模态情感分析的可用数据集大多是从不同在线视频共享平台上的产品评论收集的。表1 总结了常用的多模态情感分析数据集。

表 1 常用的多模态情感分析数据集汇总

|  |  |  |  |
| --- | --- | --- | --- |
| 数据集名称 | 语言 | 所含模态 | 情感标签 |
| SEED 数据集 | 中文 | 脑电 | 积极、中性、消极三分类 |
| 新浪微博数据集 | 中文 | 文本、图像 | 积极、消极、中性三分类 |
| Yelp 数据集 | 英文 | 文本、图像 | 1~5 的5 个情感分数 |
| Multi-ZOL 数据集 | 英文 | 文本、图像 | 1~10 的10 个分数 |
| DEAP 数据集 | 英文 | 脑电、视觉 | 消极到积极1~9 的9 个分数 |
| CH-SIMS 数据集 | 中文 | 文本、图像、音频 | -1（负）、0（中性）、1（正） |
| YouTube 数据集 | 英文 | 文本、图像、音频 | 积极、消极、中性三分类 |
| ICT-MMMO 数据集 | 英文 | 文本、图像、音频 | 积极、消极、中性三分类 |
| MOSI 数据集 | 英文 | 文本、图像、音频 | 从-3 到+3 的7 类情感倾向 |
| News Rover Sentiment 数据集 | 英文 | 文本、图像、音频 | 积极、消极、中性三分类 |
| IEMOCAP 数据集 | 英文 | 文本、图像、音频、姿态等 | 快乐、愤怒、悲伤等10 个标签 |

目前国内外多模态情感数据库大多来源于网络视频评论或人为制作，对于科研领域仍是半公开或者不公开的状态。由于模态选择的不同以及数据集的局限性，一些研究者会根据自己的需求来建立所需要的情感数据集。用于多模态情感分析的可用数据集大多是从不同在线视频共享平台上的产品评论收集的。表1 总结了常用的多模态情感分析数据集。

6 不同算法对比

为了得到影响情感分析准确率的因素，本节将对不同的多模态情感分析算法进行对比研究，对比结果如表2~表5 所示。以下表中的评价指标都为Accuracy，并且表中的模态信息 A、V、T分别代表Audio、Video、Text。

表 2 MOSI 数据集上不同算法Accuracy 比较 %

|  |  |  |  |
| --- | --- | --- | --- |
| 算法 | T+V | T+A | V+A |
| LSTM-based model | 80.22 | 79.33 | 62.17 |
| Utterance-level | 72.39 | 72.88 | 68.86 |
| C-MKL | 85.46 | 84.12 | 83.69 |
| CRMKL | 96.21 | 84.12 | 95.68 |
| Multilogue-Net | 80.06 | 80.18 | 75.16 |
| SAL-CNN | 73.00 | 72.50 | 62.10 |

表 3 MOSI 数据集上不同算法Accuracy 比较 %

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | T | V | A | A+V+T |
| LSTM-based model | 78.12 | 55.80 | 60.31 | 80.30 |
| GME-LEST | — | — | — | 76.50 |
| DFF-TMF | — | — | — | 80.98 |
| Gated mechanism for attention | — | — | — | 83.91 |
| TSAM | 74.50 | 61.80 | 60.90 | 75.10 |
| Multilogue-Net | — | — | — | 81.19 |

表 4 MOUD 数据集上不同算法Accuracy 比较 %

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | T | V | A | A+V+T |
| LSTM-based model | 52.17 | 48.58 | 59.99 | 68.11 |
| Utterance-level | 70.94 | 67.31 | 64.85 | 74.09 |
| CRMKL | 79.77 | 94.50 | 74.22 | 96.55 |
| SAL-CNN | 73.20 | 63.60 | 61.80 | 73.00 |

表 5 相同算法在不同数据集上Accuracy 比较 %

|  |  |  |
| --- | --- | --- |
| 算法 | 数据集 | A+V+T |
| LSTM-based model | MOSI | 78.12 |
| MOUD | 52.17 |
| MultiSentiNet-Att | MVSA-Single | 69.84 |
| MVSA-Multi | 68.86 |
| Multilogue-Net | MOSI | 81.19 |
| MOSEI | 82.10 |
| DFF-TMF | MOSI | 80.98 |
| MOSEI | 77.15 |
| IEMOCAP | 81.37 |
| Gated mechanism for attention | MOSI | 83.91 |
| MOSEI | 81.14 |

通过表2可以看出，大多数算法用T+V和T+A进行情感分析时的准确率都要高于V+A，说明在基于多模态的情感分析中，文本信息仍然是重要的情感线索。

通过表3可得，在MOSI数据集上用三种模态进行情感分析时，Gated mechanism for attention算法的准确率最高，说明门控单元在模态选择时的重要性。此外，对每种模态的特征进行降噪也可以提高准确率。其次，MultiSentiNet-Att和DFF-TMF这两种算法的准确率也较高，可以看出注意力机制以及模态间的相关性在提高准确率方面也有重要的价值。

通过表 4 可得，在 MOUD 数据集上用三种模态进行情感分析时，CRMKL 算法的准确率最高，说明视频中上下文的信息、文本信息的预处理以及模态融合的选择对提高准确率很有帮助。

通过表 5 可得，大多数算法在不同数据集上的鲁棒性较好。LSTM-based model 方法在不同数据集上的准确度相差较大，产生这种效果的原因是模型用 MOSI 训练，MOSI 是英语，而 MOUD 是西班牙语，语言不同，因而情感表达方式不同，分析方式也不同。

通过表2~表4 中相同的算法进行对比可以看出，用三种模态进行情感分析的准确率高于用两种和一种模态方法的准确率，说明结合多种模态信息进行情感分析的必要性。

7 总结

本文主要对当前多模态情感分析技术及模态融合方法进行了总结和对比。引入了RNN、LSTM、GRU等网络以提取上下文信息，采用多层级GRU编码上下文信息来解决长时间依赖问题，实现更全面的情感信息获取。通过注意力机制在模态融合中寻找最优权值时发挥重要作用，通过张量将所有模态特征投影到同一空间，形成联合表征空间，便于计算模态间的交互作用。并对比了主流算法的性能。

参 考 文 献

1. LIU Jiming,ZHANG Peixiang,LIU Ying,ZHANG Weidong,FANG Jie.Summary of Multi-modal Sentiment Analysis Technology[J]. Journal of Frontiers of Computer Science and Technology,2021,15(07):1165-1182
2. WANGJinghao,LIU Zhen,LIU Tingting,WANG Yuanyi,CHAIYanjie.Multimodal Sentiment Analysis Based on Multilevel Feature Fusion Atention Network[J].JOURNAL OFCHINESEINFORMATION PROCESSING,2022,36(10):145-154
3. LI Xia,LU Guan-Ming,YAN Jing-Jie,ZHANG Zheng-Yan.A Survey of Dimensional Emotion Prediction by Multimodal Cues[J].ACTA AUTOMATICA SINCIA,2018,44(12):2142-2159