赖河蒨,王丽丽, 李玲俐, 等 . 基于记忆迭代网络的多模态情感分析[J]. 智能计算机与应用,2024,14(11):200-205. DOI:10.20169/j. issn. 2095-2163. 241131

基于记忆迭代网络的多模态情感分析

赖河蒨[1], 王丽丽[2], 李玲俐[1], 许学添[1], 陈丽仪[1]
(1 广东司法警官职业学院 信息管理系, 广州 510520; 2 广东省外语艺术职业学院 基础教育学院, 广州 510640)

摘 要: 在开展多模态情感分析的过程中,需要考虑 2 方面的关键信息:一是每种模态的内部信息,二是各模态之间的交互信息。针对如何有效地捕获以上核心信息的问题,提出一种记忆迭代网络模型。将每种模态的序列数据分别输入到其各自的循环神经网络以及注意力机制模块,处理后将所有模态序列级联并送到记忆迭代模块进行迭代处理,迭代结束后将数据经过完全连接层,降低维度后进行分类。在公开的数据集 CMU-MOSI、ICT-MMMO 以及 YouTube 上进行的实验表明,记忆迭代网络模型的准确率和 *F*1 值均有所提升。

关键词: 记忆网络; 迭代; 注意力机制; 多模态; 情感分析
中图分类号: TP391 **文献标志码**: A **文章编号**: 2095-2163(2024)11-0200-06

Multimodal sentiment analysis based on memory iterative network

LAI Helang[1], WANG Lili[2], LI Lingli[1], XU Xuetian[1], CHEN Liyi[1]
(1 Department of Information Administration, Guangdong Justice Police Vocational College, Guangzhou 510520, China; 2 School of Basic Education, Guangdong Teachers College of Foreign Language and Arts, Guangzhou 510640 , China)

Abstract: In the process of multimodal sentiment analysis, two key aspects need to be considered: one is the internal information of each modality, and the other is the interaction information between each modality. Aiming at the problem of how to effectively capture the above core information, a memory iterative network model is proposed. Each modal sequence data is input into its own recurrent neural network and attention mechanism module respectively. After processing, all modal sequences are concatenated and sent to the memory iteration module for iterative processing. After the iteration, the data is passed through the fully connected layer and classified after reducing the dimension. Experiments on the public datasets CMU-MOSI, ICT-MMMO, and YouTube show that both the accuracy and *F*1 value of the memory iterative network model are improved.

Key words: memory network; iterative; attention mechanism; multimodal; sentiment analysis

0 引 言

随着互联网、社交媒体平台以及移动智能手机等应用的发展与普及,人们可以方便地通过上传视频的方式在网络上进行陈述、表达观点。由此,在 YouTube、Facebook 或 Twitter 等网络平台上就会涌现出大量带有个人情感色彩的数据,这些数据往往含有一些非常重要而且有价值的信息。例如,用户评论产品的信息,可以为商家提供产品反馈,使得商家可以发现用户对产品的倾向性,进一步改进服务质量,优化营销策略;又例如,人们对新闻、热点话题或时事政治等方面的内容发表主观观点或看法,可以为相关监管部门提供舆论民意,使其可以及时了解关注内容、并做出应对。为此,对网络评论进行深入的情感分析和挖掘,在经济和社会层面都具有重要意义。

情感分析也称为观点挖掘,是对带有情感色彩的主观性数据(尤其是用户在社交媒体或产品评论

基金项目: 广东省科技创新战略专项资金(大学生科技创新培育)项目(pdjh2024b593); 广东省教育科学规划课题(高等教育专项)(2024GXJK989); 广东省普通高校特色创新类项目(2023KTSCX250,2024WTSCX300,2023KTSCX295); 广东省第二批高职院校高水平专业群建设项目(GSPZYQ2021123); 广东司法警官职业学院第五届院级课题(2023YB08)。

作者简介: 赖河蒨(1985—),男,博士,讲师,主要研究方向:网络安全,情感分析; 王丽丽(1993—),女,博士,讲师,主要研究方向:数值计算; 李玲俐(1977—),女,教授,主要研究方向:模式识别; 陈丽仪(1980—),女,讲师,主要研究方向:数学建模。
通信作者: 许学添(1984—),男,副教授,主要研究方向:神经网络模型。Email:hmilyxxt@ 163. com。

收稿日期: 2023-06-16

中产生的带有主观性的信息）进行分析、处理、提取、挖掘、归纳和推理的过程[1-3]。在情感分析过程中，基于单模态的研究会出现信息量不足的现象，而且情感识别率还容易受到外界各种因素的影响，而多模态信息丰富、形式多样，可以充分融合多个模态之间的互补信息，有助于提高情感识别的准确性[4]。当前，多模态情感分析已受到众多学者的广泛关注，研究热点已经从单模态转移到实际应用场合下的多模态情感分析[5]。

理解这种多模态交流对人类来说是很自然的，人类大脑潜意识中也是每天进行这种交流。然而，想要让人工智能（AI）像人一样通过整合各种相关的模态去理解这种交流形式，却是极大的挑战[6]。多模态情感分析在各模态信息融合时需要考虑如何最大化地保存每种模态内部变化的信息以及各模态之间交互的信息。即多模态情感分析面临的两大挑战，具体是：

（1）不同模态数据各自所包含的情感信息可能是不同的，需要有效地获取各模态数据内部的动态变化。

（2）不同模态数据在融合时需要获取不同模态之间的动态交互[6]。

针对上述 2 个关键的挑战，本文提出记忆迭代网络。首先，为每种模态序列数据分配独立的循环神经网络以及注意力机制模块，经过处理得到的输出序列数据在特征维度层面进行级联；然后，输送到记忆迭代模块进行迭代处理，按照时间发生的顺序，在每个时刻分别训练 2 个由神经网络控制的权重向量，将得到的权重向量作为当前时刻特征数据和历史记忆数据的比例因子，用来计算当前记忆数据；最后，迭代结束、即将记忆序列中的最后一个记忆特征数据输入到完全连接层，降低维度后进行预测输出。在多个数据集上开展了对比实验，结果显示，记忆迭代网络模型在多模态情感分析方面具有较好的性能。

1　相关工作

CMU-MOSI 是多模态情感分析中常用的一个数据集。图 1 是从该数据集中抽取出来的 2 个片段。第一个片段体现了同一种模态内部信息对情感类别的影响，第二个片段体现了不同模态之间交互信息对情感类别的影响。

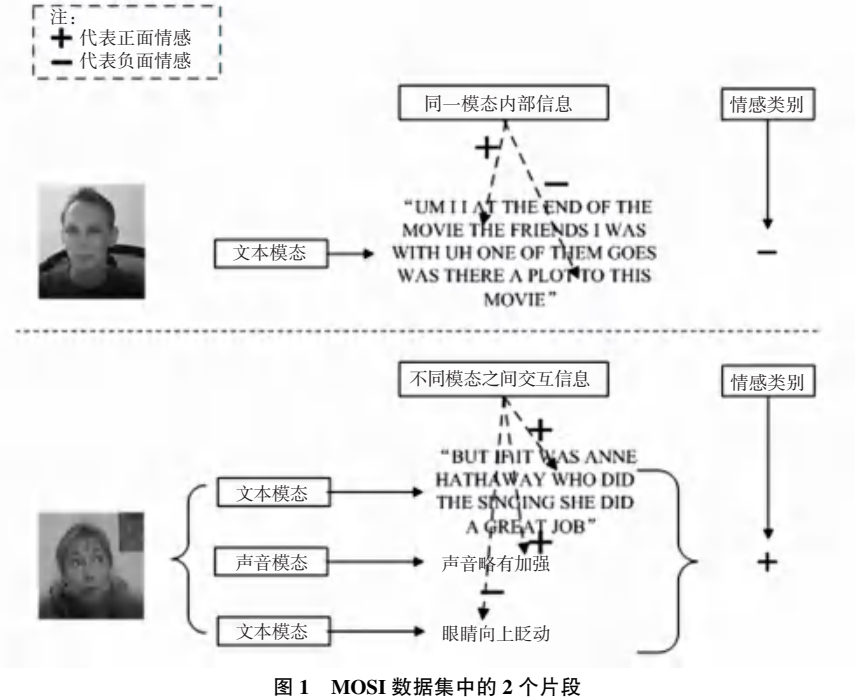

图 1　MOSI 数据集中的 2 个片段

Fig. 1　Two fragments in the MOSI dataset

随着人工神经网络的研究和应用上的更趋深入，现在已有日渐增多的研究人员在该数据集上开展多模态情感分析的研究。王旭阳等学者[7]提出了结合时域卷积网络和软注意力机制建立复合层次融合的多模态情感分析模型，并在数据集 CMU-MOSI 上的实验结果表明，所提模型能够有效提升多模态情感分析的准确率。胡新荣等学者[8]提出了一个基于多模态表示学习的多子空间情感分析框

架,该框架在数据集 CMU-MOSI 上进行仿真。得到的实验结果表明,大多数评价指标都优于基线模型。丁健等学者[9]提出了一种异质的动态融合方法,通过层次化的异质动态融合方式更完备地进行模态融合,能动态地捕捉到模态间的相互作用,在数据集 CMU-MOSI 上的实验表明所提模型相比于主流模型具有优势。冯广等学者[10]提出了使用双向门控循环网络结合模态内和跨模态的上下文注意力机制进行情绪分析,在 MOSI 数据集上验证了从多模态特征和时序特征的角度进行情绪分析,可以有效提高情绪分析研究的分类准确率。

以上方法都取得了不错的效果,但这些方法都没有考虑将记忆特征沿着时序推进的方向进行迭代。相比之下,本文提出的记忆迭代网络考虑了这方面的因素,在计算下一个记忆数据的时候,迭代之前的历史记忆数据,使得模型捕获到了更丰富的特征数据,并在实验中获得了更好的性能。

2 记忆迭代网络模型

在多模态情感分析过程中,为了捕获同一种模态在不同时刻的内部信息以及不同模态之间的交互信息,本文提出了记忆迭代网络,模型结构如图 2 所示。首先,每个“话语”(一个“话语”是以说话者在说话过程中的呼吸或停顿为界限的视频片段)的特征都由多模态序列数据表示(多模态序列数据是指每一种模态均有其对应的序列特征数据),然后将每种模态下的序列数据送到各自的循环神经网络和注意力机制模块进行处理。经过注意力模块处理后,将多种模态序列数据在特征维度进行级联,合成一个序列数据。此时,用零向量初始化记忆序列的第一个记忆(零号记忆),与合成序列进行记忆迭代。迭代结束后,取出记忆序列中的最后一个记忆,将其输入到完全连接层进行降维,最终输出预测分类结果。文中模型的详细处理过程如下。

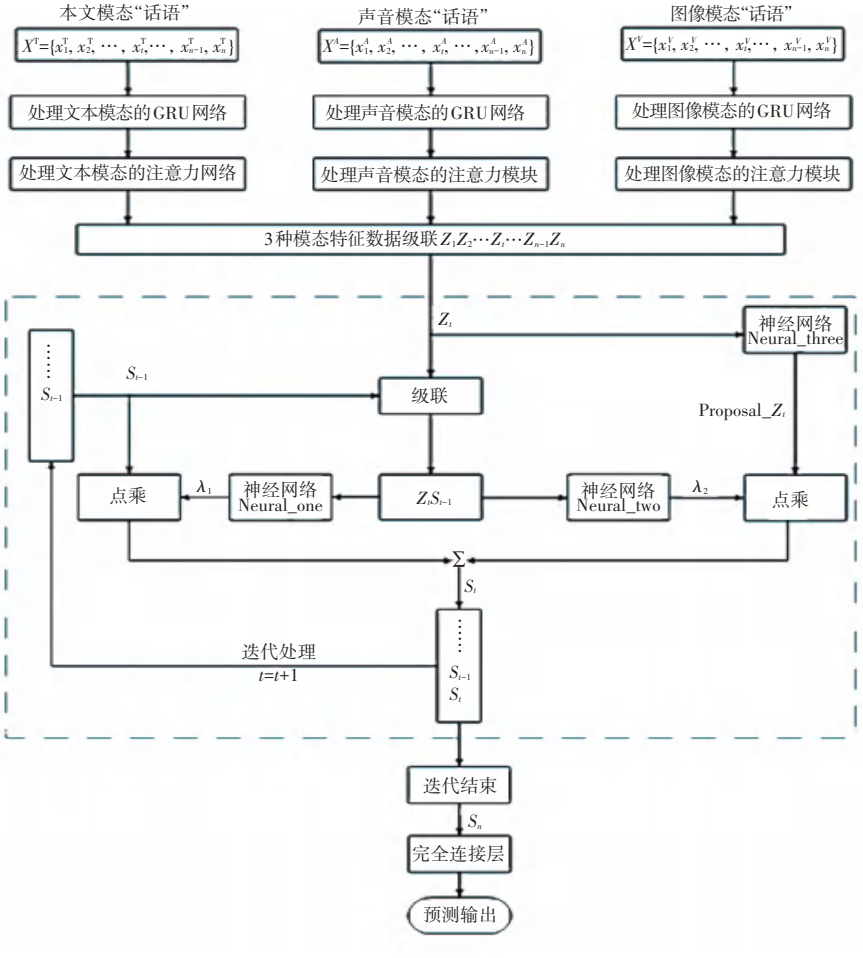

图 2 记忆迭代网络模型图

Fig. 2 Model of memory iterative network

2.1 模型的输入处理

文中模型采用了文本 (T)、声音(A) 以及图像 (V) 共 3 种模态来表示一个“话语” 的情感数据,一种模态就有一个序列,具体见图 2。令 M 为多模态

的集合，则 $M=\{T,A,V\}$，那么一个“话语”在第 $m(m\in M)$ 种模态下的数据形式可以表示为 $X^m=\{x_1^m,x_2^m,\cdots,x_t^m,\cdots,x_{n-1}^m,x_n^m\}$。其中，$x_t^m$ 是“话语”在第 m 种模态下 t 时刻的特征数据，n 是序列的长度。

2.2　循环神经网络

为了捕获同一种模态的内部信息，文中模型先将输入数据送到循环神经网络进行处理。文中使用了基于门控循环单元（Gated Recurrent Unit，GRU）的循环神经网络。GRU 是由 Cho 等学者[11]提出的一种网络结构，是长短期记忆网络 LSTM 的变体。GRU 将遗忘门和输入门合成到一个单独的“更新门”中，同时也合并了单元状态和隐藏状态，并且做了一些其它改动，使得其模型比标准 LSTM 模型更简单。门控循环单元不会随时间变化就清除以前的信息，而是会保留相关的信息并传递到下一个单元，因此就利用全部信息而避免了梯度消失问题。本文的 3 种不同模态数据序列，分别对应输入到 3 个不同的 GRU 神经网络。3 组序列数据在经过 GRU 网络后，可得到各个序列自身内部信息的输出。

GRU 使用了更新门 z_t 和复位门 r_t 来控制输入和记忆等信息，其中复位门决定了如何将新的输入信息与前面的记忆相结合，更新门定义了前面记忆保存到当前时间步的权重。假设当前时间步的输入为 x_t，那么，当前时间步的输出 y_t 的计算方法如下：

$$z_t=\sigma(x_t\cdot\boldsymbol{U}^z+y_{t-1}\cdot\boldsymbol{W}^z+\boldsymbol{b}^z) \tag{1}$$

$$r_t=\sigma(x_t\cdot\boldsymbol{U}^r+y_{t-1}\cdot\boldsymbol{W}^r+\boldsymbol{b}^r) \tag{2}$$

$$\hat{y}_t=\tanh(x_t\cdot\boldsymbol{U}^h+(y_{t-1}\otimes r_t)\cdot\boldsymbol{W}^h+\boldsymbol{b}^h) \tag{3}$$

$$y_t=(1-z_t)\otimes\hat{y}_t+z_t\otimes y_{t-1} \tag{4}$$

其中，$\boldsymbol{U}^z$、$\boldsymbol{U}^r$、$\boldsymbol{U}^h$、$\boldsymbol{W}^z$、$\boldsymbol{W}^r$ 和 $\boldsymbol{W}^h$ 是权重矩阵，$\boldsymbol{b}^z$、$\boldsymbol{b}^r$ 和 $\boldsymbol{b}^h$ 是偏置量。

2.3　注意力机制

注意力是一种人类不可或缺的复杂认知功能，指人可以在关注一些信息的同时忽略另一些信息的选择能力。同理，神经网络中的注意力机制（Attention Mechanism，AM）[11]可以模仿人类的注意力原理，对数据进行聚焦，进一步地关注到某些重要信息以及忽略掉不重要信息。

注意力模块在接收上一模块（循环神经网络）的输出之后，将进行如下的处理。首先，采用了非对等窗口的技术，将时刻 t 的历史窗口和将来窗口设置成不同的数值。例如，令历史窗口为 3，将来窗口为 2，第 $m(m\in M)$ 模态下的序列数据为 $Y^m=\{y_1^m,y_2^m,\cdots,y_t^m,\cdots,y_{n-1}^m,y_n^m\}$，那么，序列数据中的 t 时刻数据 y_t^m 的上下文信息可表示为 $Context\_y_t^m=\{y_{t-3}^m,y_{t-2}^m,y_{t-1}^m,y_{t+1}^m,y_{t+2}^m\}$。然后，利用 Softmax 函数，计算出 $Context\_y_t^m$ 中每个单元与 y_t^m 之间的权重值，再将各个权重值与每个单元相乘，加权上下文信息，得到汇总表示 $Att\_Context\_y_t^m$。具体计算如下：

$$Att\_Score=Softmax((Context\_y_t^m)^{\mathrm{T}}\cdot y_t^m) \tag{5}$$

$$Att\_Context\_y_t^m=Context\_y_t^m\cdot Att\_Score \tag{6}$$

最后，$Att\_Context\_y_t^m$ 对 y_t^m 的上下文信息进行了加权汇总，y_t^m 合并该汇总信息后进行自身更新：

$$y_t^m=y_t^m+Att\_Context\_y_t^m \tag{7}$$

2.4　记忆迭代过程

3 种不同模态的序列数据在分别经过注意力机制之后，将按照序列发生的时间顺序，在每个时刻进行级联。级联后的序列数据表示为 $Z=\{\boldsymbol{Z}_1,\boldsymbol{Z}_2,\cdots,\boldsymbol{Z}_t,\cdots,\boldsymbol{Z}_{n-1},\boldsymbol{Z}_n\}$，该序列将输入到记忆迭代网络模块进行迭代处理，详见图 2 中的虚线框部分。

本模块的迭代过程如下。首先，令记忆序列表示为 $\boldsymbol{S}=\{\boldsymbol{S}_0,\boldsymbol{S}_1,\cdots,\boldsymbol{S}_{t-1},\boldsymbol{S}_t,\cdots,\boldsymbol{S}_{n-1},\boldsymbol{S}_n\}$，其中 $\boldsymbol{S}_0$ 为零向量。在 t 时刻，将 $\boldsymbol{Z}_t$ 和 $\boldsymbol{S}_{t-1}$ 进行级联，级联后的数据送到 2 个不同的神经网络 Neural_one 和 Neural_two，通过训练这 2 个神经网络，可以得到 2 个权重向量 λ_1 和 λ_2。此外，$\boldsymbol{Z}_t$ 还送到第 3 个神经网络 Neural_three，得到新的表示 $Proposal\_\boldsymbol{Z}_t$。此时，前一时刻的记忆 $\boldsymbol{S}_{t-1}$ 与权重 λ_1 相乘，新的表示 $Proposal\_\boldsymbol{Z}_t$ 与权重 λ_2 相乘。最后，将 2 个相乘后的结果相加，得到 t 时刻的记忆 $\boldsymbol{S}_t$。具体的计算可用如下公式表示：

$$\lambda_1=Neural\_one(\boldsymbol{Z}_t\boldsymbol{S}_{t-1}) \tag{8}$$

$$\lambda_2=Neural\_two(\boldsymbol{Z}_t\boldsymbol{S}_{t-1}) \tag{9}$$

$$Proposal\_\boldsymbol{Z}_t=Neural\_three(\boldsymbol{Z}_t) \tag{10}$$

$$\boldsymbol{S}_t=\lambda_1\cdot\boldsymbol{S}_{t-1}+\lambda_2\cdot Proposal\_\boldsymbol{Z}_t \tag{11}$$

2.5　完全连接层及预测输出

记忆迭代结束后，取出记忆序列中的最后一个数据 S_n。将 S_n 送到完全连接层，经过降低维度的处理后进行预测分类。

在训练过程中，对于多分类问题，使用带 *L2* 正则化的交叉熵损失函数（如 Pytorch 损失函数中的 *nn. CrossEntropyLoss*）来计算损失值；对于二分类问题，使用带 *L2* 正则化的平均绝对误差函数（如 Pytorch 损失函数中的 *nn. L1Loss*）来计算损失值。

3　实验

3.1　数据集与实验设置

实验在 CMU－MOSI[12]、ICT－MMMO[13] 以及

YouTube[14]数据集上进行。其中,CMU-MOSI 数据集中包含了 2 199 个视频片段,ICT-MMMO 数据集包含了 340 个视频。这些视频是在线评论视频,里面含有人们发表的各种各样的意见观点。YouTube 数据集包含了 269 个视频片段,这些视频来自社交媒体网站 YouTube,涵盖了各种产品评论和意见。以上数据集中的每个片段都带有观点立场或标注了情感。Zadeh 等学者[6]已对上述各个数据集进行了数据预处理,并公开了相关特征数据(https://github.com/pliang279)。

在实验设置方面,本文模型用到的优化器是 Adam。为了防止过拟合,模型采用了 Dropout[15]机制。在训练阶段设置了早停措施,如果验证集的损失值连续 10 轮不再下降时,则停止训练。实验采用的评估方法是准确率(记为 *ACC*)和 *F*1 值(记为 *F*1)。

3.2 基线模型

为了验证本文提出的记忆迭代网络的性能,实验对比了当前一些经典的模型。以下是这些模型的简单描述。

(1)TFN(Tensor Fusion Network):是 Zadeh 等学者[16]提出的一种基于张量融合的神经网络模型。该模型通过创建一个多维度的张量来捕获 3 种模态中的单模态、双模态以及三模态之间的交互。

(2)BC-LSTM(Bidirectional Contextual LSTM):是 Poria 等学者[17]提出的一种依赖于上下文的双向循环神经网络模型。该模型通过捕获“话语”相邻的上下文信息,能较好地识别出多模态情感类别。

(3)MARN(Multi-attention Recurrent Network):是 Zadeh 等学者[6]提出的一种多级注意力循环神经网络模型。该模型首先使用一个 MAB 模块来分析不同模态之间的交互,然后使用一个 LSTMH 模块来存储这些交互信息。

(4)LMF(Low-rank Multimodal Fusion):是 Liu 等学者[18]提出的一种低秩多模态融合方法。该方法使用了低秩张量,改善了多模态在融合时的效率,降低了计算复杂度。

(5)MFN(Memory Fusion Network):是 Zadeh 等学者[19]提出的一种记忆融合网络模型。该模型首先使用一个 LSTMs 模块(LSTMs 为每种视图分配一个 LSTM)独立地捕获每种特定视图的动态交互,然后再使用一个 DMAN 模块(特定的注意力机制)来捕获不同视图之间的动态交互。

(6)MFM(Multimodal Factorization Model):是 Tsai 等学者[20]提出的一种多模态分解模型。该模型将多模态表示分解为多模态判别因子和模态特定生成因子。多模态判别因子在所有模态中都是共享的,并且包含判别任务所需的联合多模态特征。模态特定生成因子对于每个模态都是唯一的,并且包含生成每个模态所需的信息。

3.3 实验结果与分析

各种不同模型在 CMU-MOSI、ICT-MMMO 以及 YouTube 数据集上的实验结果见表 1~表 3。表 1~表 3 中,基线模型的数据来自于文献[20]。从表 1 可以看到,相比当前效果较优的模型 MFM,本文提出的模型将准确率和 *F*1 均提升了 1.3%;从表 2 可以看到,相比模型 MFM,本文提出的模型将准确率和 *F*1 分别提升 1.2%和 3.3%;从表 3 可以看到,本文提出的模型与模型 MFM 相比性能上较为接近,但也比 MFN 性能略好。由此可见,文中提出的记忆迭代网络的方法能有效捕获到多模态情感的动态变化信息。

表 1 不同模型在 CMU-MOSI 数据集上的结果

Table 1 Results of different models on CMU-MOSI

Models	*ACC*	*F*1
TFN	73.9	73.4
BC-LSTM	73.9	73.9
MARN	77.1	77.0
LMF	76.4	75.7
MFN	77.4	77.3
MFM	78.1	78.1
Our Model	**79.4**	**79.4**

表 2 不同模型在 ICT-MMMO 数据集上的结果

Table 2 Results of different models on ICT-MMMO

Models	*ACC*	*F*1
TFN	72.5	72.6
MFN	73.8	73.1
MFM	81.3	79.2
Our Model	**82.5**	**82.5**

表 3 不同模型在 YouTube 数据集上的结果

Table 3 Results of different models on YouTube

Models	*ACC*	*F*1
MFN	51.7	51.6
MFM	53.3	52.4
Our Model	**53.3**	**51.9**

4 结束语

本文提出了一种基于记忆迭代网络的多模态情

感分析方法。记忆迭代过程既考虑了每一时刻当前的特征数据，又融合了历史的记忆数据，可以更有效地捕获到每种模态的内部信息以及各种模态之间的交互信息。在多个数据集上进行了对比实验，结果表明，相比当前的一些主流神经网络模型，文中所提方法能有效地提高情感分类的准确率。下一步将研究如何降低模型复杂度以及优化神经网络模型结构。

参考文献

[1] LIU Bing. Sentiment analysis and opinion mining[J]. Synthesis Lectures on Human Language Technologies, 2012, 5(1): 1-167.

[2] 牛利月, 郑秋生, 张龙, 等. 结合句法增强的多通道方面级情感分析模型[J]. 智能计算机与应用, 2022, 12(8): 48-53.

[3] 周萍. 基于多模态深度学习的音乐情感分类算法[J]. 智能计算机与应用, 2022, 12(9): 110-114.

[4] 严驰腾, 何利力. 基于 BERT 的双通道神经网络模型文本情感分析研究[J]. 智能计算机与应用, 2022, 12(5): 16-22.

[5] PORIA S, CAMBRIA E, BAJPAI R, et al. A review of affective computing: from unimodal analysis to multimodal fusion[J]. Information Fusion, 2017, 37:98-125.

[6] ZADEH A, LIANG P P, PORIA S, et al. Multi-attention recurrent network for human communication comprehension[C]// Proceedings of the 32nd AAAI Conference on Artificial Intelligence. New Orleans, USA:AAAI, 2018: 5642-5649.

[7] 王旭阳, 董帅, 石杰. 复合层次融合的多模态情感分析[J]. 计算机科学与探索, 2023, 17(1): 198-208.

[8] 胡新荣, 陈志恒, 刘军平, 等. 基于多模态表示学习的情感分析框架[J]. 计算机科学, 2022, 49(S2): 631-636.

[9] 丁健, 杨亮, 林鸿飞, 等. 基于多模态异质动态融合的情绪分析研究[J]. 中文信息学报, 2022, 36(5): 112-124.

[10] 冯广, 江家懿, 罗时强, 等. 基于话语间时序多模态数据的情绪分析方法[J]. 计算机系统应用, 2022, 31(5): 195-202.

[11] CHO K, MERRIëNBOER V B, GULCEHRE C, et al. Learning phrase representations using RNN encoder-decoder for statistical machine translation [C]//Proceedings of the Conference on Empirical Methods in Natural Language Processing. Doha, Qatar: ACL, 2014: 1724-1734.

[12] ZADEH A, ZELLERS R, PINCUS E, et al. Multimodal sentiment intensity analysis in videos: Facial gestures and verbal messages[J]. IEEE Intelligent Systems, 2016, 31(6): 82-88.

[13] 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.

[14] MORENCY L P, MIHALCEA R, DOSHI P. Towards multimodal sentiment analysis: Harvesting opinions from the Web [C]//Proceedings of the 13th International Conference on Multimodal Interfaces. Alicante, Spain: ACM, 2011: 169-176.

[15] SRIVASTAVA N, HINTION G, KRIZHEVSKY A, et al. Dropout: A simple way to prevent neural networks from overfitting [J]. The Journal of Machine Learning Research, 2014, 15(1): 1929-1958.

[16] ZADEH A, CHEN Minghai, PORIA S, et al. Tensor fusion network for multimodal sentiment analysis[C]//Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing. Denmark: dblp, 2017: 1103-1114.

[17] PORIA S, CAMBRIA E, HAZARIKA D, et al. Context-dependent sentiment analysis in user-generated videos[C]// Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada: ACL, 2017: 873-883.

[18] LIU Zhun, SHEN Ying, LAKSHMINARASIMHAN V B, et al. Efficient low-rank multimodal fusion with modality-specific factors[C]//Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics. Melbourne, Australia: ACL, 2018: 2247-2256.

[19] ZADEH A, LIANG P P, MAZUMDER N, et al. Memory fusion network for multi-view sequential learning[C]//Proceedings of the Thirty Second AAAI Conference on Artificial Intelligence. New Orleans: AAAI, 2018: 5634-5641.

[20] TSAI Y H H, LIANG P P, ZADEH A, et al. Learning factorized multimodal representations [J]. arXiv preprint arXiv, 1806.06176v3, 2019.