



计算机科学

COMPUTER SCIENCE

基于三支决策的鲁棒增量模糊概念认知情感识别方法

徐伟华, 胡开平

引用本文

徐伟华, 胡开平. 基于三支决策的鲁棒增量模糊概念认知情感识别方法[J]. 计算机科学, 2026, 53(5): 257-267.

XU Weihua, HU Kaiping. [Robust Incremental Fuzzy Concept-cognitive Emotion Recognition Method Based on Three-way Decision](#) [J]. Computer Science, 2026, 53(5): 257-267.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[复杂网络下的概念认知学习与增量学习](#)

Concept-cognitive Learning and Incremental Learning in Complex Networks

计算机科学, 2026, 53(4): 208-214. <https://doi.org/10.11896/jsjcx.250600216>

[STWD-DLFRD: 基于序贯三支决策与深度学习的多粒度虚假评论检测方法](#)

STWD-DLFRD: Multi-granularity Fake Review Detection via Sequential Three-way Decisions and Deep Learning

计算机科学, 2026, 53(4): 188-196. <https://doi.org/10.11896/jsjcx.250500088>

[结合性别信息的多任务语音情感识别](#)

Multi-task Speech Emotion Recognition Incorporating Gender Information

计算机科学, 2026, 53(1): 180-186. <https://doi.org/10.11896/jsjcx.241200006>

[基于GRAM矩阵的粒感知机](#)

Granular Perception Machine Based on GRAM Matrix

计算机科学, 2025, 52(11A): 241200110-7. <https://doi.org/10.11896/jsjcx.241200110>

[部分不完备广义多尺度数据的最优尺度组合和属性约简](#)

Optimal Scale Combinations and Attribute Reduction for Partially Incomplete Generalized Multi-scale Data

计算机科学, 2025, 52(11): 49-61. <https://doi.org/10.11896/jsjcx.250700019>

基于三支决策的鲁棒增量模糊概念认知情感识别方法

徐伟华 胡开平

西南大学人工智能学院 重庆 400715

摘要 语音情感识别(Speech Emotion Recognition, SER)在人机交互系统中具有重要作用。为了解决现有深度学习模型在SER任务中决策过程不透明,以及传统概念认知学习(Concept-Cognitive Learning, CCL)在处理增量数据时易受噪声干扰而产生概念漂移的问题,构建了一种融合极端随机树权重机制的三支模糊概念认知分类框架(3WERT-WFCCL)。在特征处理上,模型采用Whisper提取高维语音特征,并经由多层感知机进行分层抽象表示;在认知学习阶段,引入极端随机树算法计算特征重要性以实现属性权重的自动量化分配,并在认知算子中嵌入三支决策的容错阈值参数,构建正负双向认知机制。面对增量数据时,模型依据特征辨识距离将新样本划分为正域、边界域和负域,并采用仅利用正域样本更新概念的鲁棒策略,有效抵御了噪声干扰。在特征边界较为复杂的SAVEE数据集上,鲁棒更新策略相比全局更新策略的准确率提升了0.16个百分点。在EmoDB和SAVEE两个公开数据集上进行相关实验,3WERT-WFCCL在多个关键评价指标上均优于现有基线方法。相比各数据集上表现最优的逻辑回归(Logistic Regression, LR)算法,所提出方法的准确率分别提升了1.53个百分点和0.62个百分点, F1分数分别提升了1.28个百分点和0.40个百分点。实验结果验证了引入三支决策机制的有效性,为构建兼顾高分类精度、强抗噪能力与逻辑可解释性的SER模型提供了新的方法。

关键词: 概念认知学习; 三支决策; 语音情感识别; 粒计算; 形式概念分析

中图分类号 TP183

Robust Incremental Fuzzy Concept-cognitive Emotion Recognition Method Based on Three-way Decision

XU Weihua and HU Kaiping

College of Artificial Intelligence, Southwest University, Chongqing 400715, China

Abstract Speech emotion recognition(SER) plays an important role in human-computer interaction systems. In order to solve the problems that the decision making process of existing deep learning models is opaque in the SER task, and the traditional concept-cognitive learning(CCL) is susceptible to noise interference and concept drift when processing incremental data, a three-way weighted fuzzy concept-cognitive classification framework that leverages extremely randomized trees(3WERT-WFCCL) is proposed. In the feature processing, Whisper is used to extract high dimensional speech features, and a multi-layer perceptron(MLP) is used for hierarchical abstract representation. In the cognitive learning stage, the extremely randomized trees(ERT) algorithm is introduced to calculate the importance of features to realize the automated quantitative allocation of attribute weights, and the three-way decision fault tolerance threshold parameter is embedded in the cognitive operator to construct a positive and negative two-way cognitive mechanism. In the face of incremental data, the model divides the new samples into a positive region, a boundary region and a negative region according to the feature identification distance, and adopts a robust strategy that only uses the positive region samples to update the concept, which effectively resists the noise interference. On the SAVEE dataset with more complex feature boundaries, the robust update strategy improves the accuracy by 0.16 percentage points compared with the global update strategy. Experiments on two public datasets EmoDB and SAVEE show that 3WERT-WFCCL is superior to the existing baseline methods in multiple key evaluation indicators. Compared with the baseline models Logistic Regression(LR) with the best performance on each data set, the accuracy of the proposed algorithm is increased by 1.53 percentage points and 0.62 percentage points respectively, and the F1 score is increased by 1.28 percentage points and 0.40 percentage points respectively. Experimental results verify the effectiveness of the three-way decision mechanism, which provides a new method for constructing SER models with high classification accuracy, strong noise robustness and logical interpretability.

到稿日期: 2026-02-12 返修日期:2026-04-04

基金项目:国家自然科学基金(62376229, 12371465);重庆市自然科学基金(CSTB2023NSCQ-LZX0027, CSTB2023NSCQ-MSX1063)

This work was supported by the National Natural Science Foundation of China (62376229, 12371465) and Natural Science Foundation of Chongqing(CSTB2023NSCQ-LZX0027, CSTB2023NSCQ-MSX1063).

通信作者:徐伟华(chxuwh@email.swu.edu.cn)

Keywords Concept-cognitive learning, Three-way decision, Speech emotion recognition, Granular computing, Formal concept analysis

1 引言

语言是人类文化中最核心的交流媒介,蕴含着丰富的语言信息与副语言信息。语音不仅传递字面意义,还承载着情绪、个性与心理状态等深层信息^[1]。人类大脑能够综合处理这些语言与副语言线索,从而实现对话语潜在含义的准确理解与有效沟通^[2]。随着人机交互技术的快速发展,如何使机器同样具备感知和理解语音情绪等副语言特征的能力,已成为语音情绪识别(SER)领域的重要研究课题。然而,如何从语音信号中提取有效的情绪特征,仍是当前SER研究所面临的核心挑战之一。

在SER研究中,特征提取是决定系统性能的关键环节。早期研究人员从语音信号中探索了多种声学特征,包括基于声源的激励特征、韵律参数、MFCC、对数梅尔频谱、过零率以及均方根能量在内的底层声学描述子(Low-Level Descriptors, LLDs)^[3-4]。这类特征计算简便,但在表达跨时间尺度的高层语义信息方面能力有限。近年来,以Whisper为代表的大规模预训练语音模型^[5]因能够提取兼具声学稳健性与语义丰富性的深层表征,逐渐成为高维特征提取的主流选择。例如,Pepino等^[6]基于wav2vec 2.0嵌入实现了语音情感识别的显著性能提升;Ma等^[7]提出的emotion2vec通过自监督预训练进一步增强了语音情感表征能力。然而,高维特征往往伴随着信息冗余,如何从中筛选出对情感识别真正有效的关键信息,是当前特征提取研究面临的重要问题^[8]。为还原人类认知中由底层感知向高层语义逐步凝练的过程,引入合适的特征变换手段十分必要。多层感知机(Multi-Layer Perceptron, MLP)^[9]以其非线性变换特性,实现了从初始语音特征到深层认知表征的过渡。

在分类阶段,研究人员通常将分类器分为线性与非线性两大类^[10]。线性分类器中,支持向量机(Support Vector Machine, SVM)、贝叶斯网络(Bayesian Network, BN)和线性判别分析(Linear Discriminant Analysis, LDA)^[11]是情感识别中较为常见的选择。由于语音信号本质上具有非平稳性,非线性分类器在SER任务中往往表现出更强的适应能力。高斯混合模型(Gaussian Mixture Model, GMM)和隐马尔可夫模型(Hidden Markov Model, HMM)是应用最为广泛的非线性分类器,能够对语音的时序动态特性进行有效建模,具备一定的可解释性。此外,K近邻(K-Nearest Neighbor, KNN)^[10]、决策树等方法也被应用于情感分类任务。后来,以卷积神经网络(Convolutional Neural Network, CNN)^[12]以及由多种神经网络构成的集成模型^[10,13]为代表的深度学习方法逐步被引入SER领域,并实现了识别准确率的跃升。但与此同时,模型的透明度却大幅降低,其内部多层非线性映射构成的复杂结构使决策路径几乎不可追踪,由此产生了典型的黑箱困境。针对这一局限,部分研究者开始探索将概念认知学习作为替代路径,以期在保持识别性能的同时,赋予模型更具可解释性的认知推理能力。

概念认知学习(CCL)作为智能计算与认知科学交叉方向的一种新兴学习框架,其核心目标在于模拟人类借助概念体系实现知识发现与整合的思维机制^[14]。Shi等^[15]最早提出了面向增量概念学习的CCL模型。此后,Mi等^[16]将CCL与概念空间结合,提出半监督学习方法。Xu等^[17]引入模糊机制构建了双向CCL框架。Zhang等^[18]对增量CCL过程的稳定性进行了系统分析。Deng等^[19]与Xin等^[20]则进一步将粒计算与属性波动机制融入模糊CCL模型中。尽管上述研究提升了CCL的分类性能与可解释性,但传统模糊CCL采用二支决策逻辑,在处理边界模糊的情感样本或含噪增量数据时缺乏缓冲机制,易引发概念漂移问题。为此,三支决策已被广泛应用于各类概念认知模型的研究中,为处理不确定性信息提供了新视角。例如,Qi等^[21]从正反两视角研究三支概念分析;Li等^[22]探讨了多粒度三支概念学习;Yuan等^[23]提出渐进式模糊三支增量学习机制;Xu等^[24]构建了面向三支概念的新型CCL模型;Guo等^[25-26]则深入研究了模糊三支概念在动态数据分类中的应用,并指出其在知识描述与减少认知偏差方面的独特优势^[27]。受此启发,本文提出了一种基于极端随机树加权的三支模糊概念认知分类模型(3WERT-WF-CCL),通过设定置信度阈值将样本划分为正域、边界域与负域,并仅对正域样本实施增量更新,在保留高解释性的同时增强了模型对噪声数据的鲁棒性。本文的主要贡献如下:

1)提出了基于三支决策的鲁棒增量模糊概念认知情感识别方法3WERT-WFCCL,该框架融合了三支决策理论与模糊加权概念认知学习方法,解决了传统二支CCL模型在处理增量数据时面临的概念漂移问题。

2)设计了带有阈值调节机制的正负双向认知算子,该机制利用极端随机树算法客观量化特征权重,实现了对复杂语音特征的数据驱动加权与抗噪提取。

3)提出了一种基于正域更新的鲁棒增量学习策略,通过计算特征辨识距离,算法将边界模糊的情感样本划入边界域并延迟更新,在不丧失对新知识学习能力的前提下,有效防止低质量数据对概念空间的干扰。在EmoDB和SAVEE两个通用SER数据集上的对比实验表明,该模型的准确率分别达到97.42%和87.08%,相较于现有主流基线方法在分类性能与稳定性方面均表现出一定优势。

2 预备知识

基于经典形式背景的CCL强调对概念的确定性描述,难以有效处理连续型数据中普遍存在的模糊性问题。而引入隶属函数的CCL能够对样本与声学属性间的渐进映射关系进行细粒度建模,在处理连续型声学特征数据的概念学习场景中具有更强的适应性。为支撑后续模型的构建,本章重点阐述模糊形式背景的相关概念,认知算子与三支概念的理论细节将在下一章展开。

在实际分类任务中,模糊决策形式背景具有更高的实用价值。其中,声学特征以模糊隶属度的形式作为条件属性,情

感类别以离散标签作为决策属性。此外,本文所研究的问题聚焦于条件属性为模糊数据,决策属性为离散标签的模糊-经典决策形式背景,不涉及模糊-模糊或经典-模糊等其他形式,故不展开讨论。

设 (U, V, \tilde{E}) 与 (U, D, T) 为两个相异的形式背景,若五元组 (U, V, \tilde{E}, D, T) 满足以下条件,则称其为模糊决策形式背景:

- 1) $U = \{u_1, u_2, \dots, u_n\}$ 为非空有限语音样本集;
- 2) $V = \{v_1, v_2, \dots, v_m\}$ 为非空有限声学特征集;
- 3) $\tilde{E} = \{\langle (u_i, v_j), \tilde{E}(u_i, v_j) \rangle \mid (u_i, v_j) \in U \times V\}$ 为 $U \times V$

上的模糊关系,其中 $\tilde{E}: U \times V \rightarrow [0, 1]$, $\tilde{E}(u_i, v_j)$ 表示样本 u_i 关于特征 v_j 的隶属度;

4) $U/D = \{D_1, D_2, \dots, D_l\}$ 为基于情感标签 D 的决策划分,其中 $D = D_1 \cup D_2 \cup \dots \cup D_l$;

5) $T: U \times D \rightarrow \{0, 1\}$ 为 U 与 D 之间的二元关系,用于指示样本所属的情感决策类。

为直观呈现两类形式背景的区别,给出如下具体示例。对于模糊形式背景 (U, V, \tilde{E}) , 设对象集 $U = \{u_1, u_2, u_3, u_4, u_5\}$ 包含 5 个语音样本,声学特征集 $V = \{v_1, v_2, \dots, v_8\}$ 包含 8 个声学特征。模糊关系 $\tilde{E}: U \times V \rightarrow [0, 1]$ 量化了每个样本在各特征上的表现强度,例如 $\tilde{E}(u_1, v_1) = 0.3$ 表示样本 u_1 在特征 v_1 上的隶属度为 0.3, $\tilde{E}(u_4, v_1) = 0.8$ 则表明样本 u_4 在特征 v_1 上具有更强的表现。类似地, $\tilde{E}(u_1, v_7) = 0.9$ 说明 u_1 在特征 v_7 上具有很高的隶属度,而 $\tilde{E}(u_1, v_6) = 0.1$ 则表明 u_1 在特征 v_6 上具备该特征的程度较低。这种连续取值的方式使算法能够细粒度地刻画样本的声学特性。

对于模糊决策形式背景 (U, V, \tilde{E}, D, T) , 在保留上述 5 个样本与条件特征集 $V = \{v_1, v_2, \dots, v_5\}$ 的基础上,引入决策划分 $U/D = \{D_1, D_2, D_3\}$ 对应 3 种情感类别。二元关系 T 将每个样本唯一映射至其所属情感类,例如 $T(u_1, D_1) = 1$ 表示样本 u_1 属于情感类 D_1 , $T(u_2, D_2) = 1$ 表示样本 u_2 属于情感类 D_2 , $T(u_3, D_3) = 1$ 表示样本 u_3 属于情感类 D_3 , 而对应的其余决策属性值均为 0, 即对任意 $k \neq i$, 有 $T(u_i, D_k) = 0$ 。两者的核心区别在于,模糊决策形式背景在模糊条件属性的基础上附加了离散决策标签,使特征描述与分类目标得以统一表达,为后续概念认知学习提供了完整的建模基础。

3 基于极端随机树加权的三支模糊概念认知学习

当前,语音情感识别研究在提升性能与保持模型可解释性之间存在平衡问题。以卷积神经网络和循环神经网络为代表的深度学习模型具有较强的非线性拟合能力,能够提升识别精度,但其内部的层级映射结构复杂,决策过程不透明,难以满足人机交互中对逻辑可解释性的需求。另外,Whisper 等预训练模型为获取鲁棒的声学特征提供了方法,但如何从这些高维特征中提取关键信息,并建立清晰的情感推理过程,仍是当前研究需要解决的问题。针对上述问题,本章结合三支决策理论,提出了一种基于极端随机树权重的三支模糊概

念认知学习模型 3WERT-WFCCL。该方法利用 Whisper 预训练模型与多层感知机 (MLP) 提取特征,引入极端随机树 (Extremely Randomized Trees, ERT) 算法量化高维特征的基尼指数贡献,实现属性权重的自动量化分配。同时,本章在定义模糊概念认知算子时,加入了三支决策的阈值参数 (α 与 β), 利用正、负双向算子从两个角度描述情感认知学习过程中的特征范围,降低了高维特征中冗余和噪声带来的不确定性。此外,本章设计了基于三支概念空间的增量学习策略,以实现对新情感模式的分类与知识更新。

3.1 极端随机树与基尼指数

声学特征通常具有高维度和高冗余的特点。为了从复杂的语音信号中筛选出具有分类能力的特征,本节引入极端随机树算法。极端随机树通过引入随机性来扩展决策树,在节点分裂时,ERT 采用完全随机的方式选择分裂点,而不是寻找局部最优特征。该机制引入的额外随机因素能够有效抑制模型在跨说话人、跨语境条件下的预测波动,从而增强其泛化表现。与传统决策树相比,ERT 减少了寻找最优分裂阈值的计算过程,能够快速处理高维语音特征向量。

借助特征重要性评估可以量化各声学维度对情感判别结果的贡献程度。利用 ERT 进行重要性度量,能够筛选出在情感类别区分中起关键作用的特征,有助于去除冗余的背景噪声特征,并为后续的三支模糊概念认知学习提供权重依据。在 ERT 中,特征重要性通过每个特征降低基尼指数的幅度进行度量。基尼系数反映了数据集中类别混杂的程度,该值越大,表示当前节点包含的情感类别混合程度越高。基尼指数公式如下所示:

$$G = 1 - \sum_{m=1}^c p_m^2$$

其中, c 为情感类别数目, p_m 为某语音样本归属于类别 m 的后验概率。在 ERT 中,每一次节点划分均有助于将不同情感的样本加以区分,进而减小总体基尼指数值。若某特征在降低该指数方面效果越显著,则认为其重要性越突出。特征重要性表示如下:

$$VI(v_i) = \frac{1}{T} \sum_{t=1}^T \sum_{n \in N_t, v_n = v_i} P(n) \cdot \Delta G(v_i, n)$$

此处, $VI(v_i)$ 量化了第 i 维声学特征 v_i 对分类决策的贡献程度,而 T 为集成模型中基学习器的数目。在单棵决策树 t 内部,记其节点集合为 N_t , 各内部节点 v_n 依据某一特征 v_i 执行二元划分。 $P(n)$ 代表节点 n 中的样本数在总样本中所占的比重, $\Delta G(v_i, n)$ 表示由节点 n 处分裂引起的类别不纯度变化, 定义为:

$$\Delta G(v_i, n) = G(n_{\text{parent}}) - (G(n_{\text{left}}) + G(n_{\text{right}}))$$

其中, $G(n_{\text{parent}})$, $G(n_{\text{left}})$ 与 $G(n_{\text{right}})$ 依次对应当前分裂节点及其左、右两个子节点处的基尼不纯度。具体来说,需要遍历整个森林中以 v_i 作为划分依据的全部节点,累加其不纯度降幅 $\Delta G(v_i, n)$ 后再除以树的数量以求均值。

经上述计算所得到的各维特征贡献度 $VI(v_i)$ 均为非负数,随后对全部维度的贡献度执行归一化,保证各分量之和等于 1。该评估体系按照各特征对决策过程影响的大小实施排序,从而确定了后续加权三支模糊 CCL 框架构造所需的属性权重参数 $\omega(v)$ 。

3.2 带有阈值调节机制的三支概念加权方法

人类对声音属性的感知是渐进的,并且包含否定的判断。模糊集理论与三支决策结合可以描述声学特征的不确定性。设定 B 为论域, \tilde{B} 为 B 上的一个模糊集,其通过隶属度函数 $\tilde{B}(v): B \rightarrow [0, 1]$ 加以刻画;相应地,其非隶属度可表示为 $\tilde{B}^-(v) = 1 - \tilde{B}(v)$ 。

定义 1 模糊形式背景可表示为三元组结构 (U, V, \tilde{E}) 。其中, U 为由语音样本构成的对象集合; V 为所提取的语音特征集合; $\tilde{E}: U \times V \rightarrow [0, 1]$ 为对象与属性之间的模糊关联,用以刻画对象 u_i 在属性 v_j 上的正向隶属度。此外,引入负向模糊关联 $\tilde{E}^-(u_i, v_j) = 1 - \tilde{E}(u_i, v_j)$, 描述对象 u_i 缺乏某属性的程度。分别以 $\mathcal{P}(U)$ 与 $\mathcal{F}(V)$ 作为对象集 U 与属性集 V 各自对应的幂集。

在概念构建阶段,引入参数 α 和 β , 在 3WERT-WFCCL 算法中, α 和 β 作为调节因子加在认知属性的计算中。作为调节因子,它们可以是正数也可以是负数,其正负决定了条件是放宽还是收紧。在实际应用中, α 通常取较小的负值,用于控制正域的准入严格程度; β 通常取较小的正值,用于界定边界域与负域的分界。两者的绝对值均远小于 1, 以确保阈值调节幅度在合理范围内。具体最优组合通过网格搜索确定。

定义 2 对于任意对象子集 $A \subseteq U$ 和模糊属性集 \tilde{B}^+ , $\tilde{B}^- \in \mathcal{F}(V)$, 基于调节因子 α 的正向认知算子及其逆算子 \mathcal{M}^+ 和 \mathcal{N}^+ 定义为:

$$\mathcal{M}^+(A)(v_j) = \bigcap_{u_i \in A} (\tilde{E}(u_i, v_j) - \alpha), v_j \in V$$

$$\mathcal{N}^+(\tilde{B}^+) = \{u_i \in U \mid \forall v_j \in V, \tilde{E}(u_i, v_j) \geq \tilde{B}^+(v_j) + \alpha\}$$

基于调节因子 β 的负向认知算子及其逆算子 \mathcal{M}^- 和 \mathcal{N}^- 定义为:

$$\mathcal{M}^-(A)(v_j) = \bigcap_{u_i \in A} (\tilde{E}^-(u_i, v_j) - \beta), v_j \in V$$

$$\mathcal{N}^-(\tilde{B}^-) = \{u_i \in U \mid \forall v_j \in V, \tilde{E}^-(u_i, v_j) \geq \tilde{B}^-(v_j) + \beta\}$$

结合上述算子定义带有调节因子的三支模糊概念认知算子及其逆算子 \mathcal{M}^∇ 和 \mathcal{N}^∇ 为:

$$\mathcal{M}^\nabla(A) = (\mathcal{M}^+(A), \mathcal{M}^-(A))$$

$$\mathcal{N}^\nabla(\tilde{B}^+, \tilde{B}^-) = \mathcal{N}^+(\tilde{B}^+) \cap \mathcal{N}^-(\tilde{B}^-)$$

算子 \mathcal{M}^∇ 从正、负两个角度学习对象集的属性特征。当调节因子 α 取正值时, $(\tilde{E}(u_i, v_j) - \alpha)$ 会降低对样本属性的隶属度要求; 当 α 取负值时, 对样本属性的隶属度数值要求更高。正向认知算子从样本在多大程度上拥有某一声学特征的角度出发, 刻画样本与情感概念之间的正向匹配程度, 隶属度越高, 说明样本越符合该情感类别的典型声学模式。负向认知算子则从相反角度出发, 关注样本在多大程度上缺乏某一特征, 描述样本与概念之间的排斥关系, 非隶属度越高, 说明样本越偏离该情感类别的特征分布。两个视角相互补充, 使模型能够同时利用具备什么特征和缺少什么特征两方面信息, 对情感样本的特征边界形成更完整的认知描述。

如果存在 $\mathcal{M}^\nabla(A) = (\tilde{B}^+, \tilde{B}^-)$ 且 $\mathcal{N}^\nabla(\tilde{B}^+, \tilde{B}^-) = A$, 则称序对 $(A, (\tilde{B}^+, \tilde{B}^-))$ 构成一个模糊三支概念。

定义 3 令 (U, V, \tilde{E}) 和 (U, D, J) 表示两个不同的形式背景, 当声学特征 V 和情感标签 D 的交集为空时, 构成模糊形式决策背景 (U, V, \tilde{E}, D, J) 。给定两个基于相同调节因子构建的加权模糊三支概念 $(A_1, (\tilde{B}_1^+, \tilde{B}_1^-))$ 和 $(A_2, (\tilde{B}_2^+, \tilde{B}_2^-))$, 其层级序关系定义为:

$$(A_1, (\tilde{B}_1^+, \tilde{B}_1^-)) \leq (A_2, (\tilde{B}_2^+, \tilde{B}_2^-)) \Leftrightarrow A_1 \subseteq A_2 \Leftrightarrow (\tilde{B}_2^+ \subseteq \tilde{B}_1^+, \tilde{B}_2^- \subseteq \tilde{B}_1^-)$$

定义 4 对于一个模糊三支概念 $(A, (\tilde{B}^+, \tilde{B}^-))$, 结合 ERT 获取的属性权重为 $\omega(v)$, 定义加权模糊三支概念为 $(A, (\tilde{B}^+, \tilde{B}^-), \omega)$, 其中综合权重表示为:

$$\omega = \sum_{v_i \in V} \frac{1}{2} (\tilde{B}^+(v_i) + \tilde{B}^-(v_i)) \omega(v_i)$$

综合权重 ω 结合了正向隶属与负向非隶属强度, 反映了该三支内涵在情感分类中的特征表示程度。

命题 1 给定一个模糊形式决策背景 (U, V, \tilde{E}, D, J) 与调节因子 (α, β) 。对于任意语音样本子集 $u \subseteq U$, 则三元组 $(\mathcal{N}^\nabla(\mathcal{M}^\nabla(u)), \mathcal{M}^\nabla(u), \omega)$ 构成一个加权模糊三支粒概念。

3.3 加权模糊三支概念学习情感分类算法

基于上述算子生成的概念空间具有三支决策的划分属性。本节通过这些概念构建情感空间。

定义 5 给定模糊形式决策背景及决策类划分 $U/D = \{\mathcal{D}_1, \mathcal{D}_2, \dots, \mathcal{D}_c\}$, 其中 c 为情感类别总数。对于情感决策类 \mathcal{D}_i , 与之对应的加权模糊三支粒概念子空间 C_i 可形式化表示为:

$$C_i = \{(\mathcal{N}^\nabla(\mathcal{M}^\nabla(u)), \mathcal{M}^\nabla(u), \omega) \mid \forall u \subseteq \mathcal{D}_i\}$$

将所有这类子空间汇聚, 得到面向全部情感类别的加权模糊三支概念全局空间 $\mathcal{G} = \{C_1, C_2, \dots, C_c\}$ 。具体构建过程如算法 1 所示。为了减少特殊样本的干扰, 本文将相似的概念进行概念融合。

算法 1 构建基于 ERT 加权的三支模糊粒概念空间

输入: 模糊形式决策背景 (U, V, \tilde{E}, D, J) , ERT 属性权重 $\omega(v)$, 调节因子 (α, β)

输出: 全局加权模糊三支情感概念空间 \mathcal{G}

1. 初始化 $\mathcal{G} \leftarrow \emptyset$;
2. for each 情感决策类 $\mathcal{D}_i \in U/D$ do
3. 初始化当前决策类的子空间 $C_i \leftarrow \emptyset$;
4. for each 样本子集 $u \subseteq \mathcal{D}_i$ do
5. 根据调节因子 α 和 β 计算双向认知算子 $\mathcal{M}^\nabla(u)$ 与 $\mathcal{N}^\nabla(\mathcal{M}^\nabla(u))$;
6. 根据定义 4 计算该三支概念的综合权重 ω ;
7. 生成加权模糊三支粒概念 $(\mathcal{N}^\nabla(\mathcal{M}^\nabla(u)), \mathcal{M}^\nabla(u), \omega)$ 并加入 C_i ;
8. end
9. $\mathcal{G} = \mathcal{G} \cup \{C_i\}$;
10. end
11. return \mathcal{G}

定义 6 在概念子空间 C_i 中, 若存在外延满足层级包含关系 $A_1 \subseteq A_2 \subseteq \dots \subseteq A_q$ 的概念集合 (q 为层级深度), 则通过

指数加权的方式构建一个具有相似概念簇信息的加权模糊三支伪概念。

$$p_{i,k} = (A_{i,k}, (\tilde{B}_{i,k}^+, \tilde{B}_{i,k}^-), \omega_{i,k});$$

$$A_{i,k} = A_q$$

$$\tilde{B}_{i,k}^+ = \frac{1}{2^{q-1}} (\tilde{B}_1^+ + \sum_{h=2}^q 2^{h-2} \tilde{B}_h^+)$$

$$\tilde{B}_{i,k}^- = \frac{1}{2^{q-1}} (\tilde{B}_1^- + \sum_{h=2}^q 2^{h-2} \tilde{B}_h^-)$$

$$\omega_{i,k} = \frac{1}{|V|} \sum_{v \in V} \frac{1}{2} (\tilde{B}_{i,k}^+(v) + \tilde{B}_{i,k}^-(v)) \omega(v)$$

该过程赋予了外延较大的概念更大的数值权重,形成具有正负特征表示的情感伪概念,由此构成当前类别的伪概念集合 $P_i = \{p_{i,1}, p_{i,2}, \dots\}$ 。所有类别的伪概念集合共同构成渐进式三支伪概念空间 $= \{P_1, P_2, \dots, P_c\}$,具体过程如算法 2 所示。

算法 2 构造渐进式加权模糊三支伪概念空间

输入:全局加权模糊三支概念空间 $\mathcal{G} = \{C_1, \dots, C_c\}$

输出:渐进式三支伪概念空间 \mathcal{P}

1. 初始化渐进式空间 $\mathcal{P} \leftarrow \emptyset$;
2. for each 子空间 $C_i \in \mathcal{G}$ do
3. 初始化伪概念子空间 $P_i \leftarrow \emptyset$;
4. 在 C_i 中寻找满足外延层级包含关系 $A_1 \subseteq A_2 \subseteq \dots \subseteq A_q$ 的概念集合簇;
5. for each 概念集合簇 do
6. 根据定义 6,利用指数加权计算融合后的外延 $A_{i,k}$ 、双向内涵 $(\tilde{B}_{i,k}^+, \tilde{B}_{i,k}^-)$ 及综合权重 $\omega_{i,k}$;
7. 生成加权模糊三支伪概念 $p_{i,k} = (A_{i,k}, (\tilde{B}_{i,k}^+, \tilde{B}_{i,k}^-), \omega_{i,k})$;
8. 将 $p_{i,k}$ 添加到 P_i ;
9. end
10. $\mathcal{P} = \mathcal{P} \cup \{P_i\}$;
11. end
12. return

3.4 \mathcal{P} 基于三支边界的增量学习与更新策略

在处理增量数据时需要对新样本进行分类并更新。在处理新样本 Δu 时,根据其与其与伪概念空间的匹配关系,实施增量学习。

3.4.1 新样本的加权辨识距离与三支划分

定义 7 对于新样本 Δu ,其在特征空间中的表现直接由正向模糊关系 $\tilde{E}(\Delta u, v)$ 和负向模糊关系 $\tilde{E}^-(\Delta u, v)$ 表示。样本 Δu 到类别 P_i 中情感三支伪概念 $p_{i,k}$ 的加权辨识距离为:

$$d(\Delta u, p_{i,k}) = \left(\sum_{v \in V} \omega(v) ((\tilde{E}(\Delta u, v) - \tilde{B}_{i,k}^+(v))^2 + (\tilde{E}^-(\Delta u, v) - \tilde{B}_{i,k}^-(v))^2) \right)^{\frac{1}{2}}$$

由于伪概念空间 P_i 中可能包含多个伪概念,因此样本 Δu 到情感类别 i 的距离取其到该类下所有伪概念的最小距离:

$$d(\Delta u, P_i) = \min_{p \in P_i} d(\Delta u, p)$$

为了确定三支区域的划分标准,需要事先定义一种基础距离阈值 D_{base} 。该阈值反映了各个情感类别在特征空间中的平均内聚程度,通过计算训练集中所有已知语音样本到其所属类别伪概念的距离均值来动态确定。 D_{base} 的计算式为:

$$D_{\text{base}} = \frac{1}{|U|} \sum_{i=1}^c \sum_{u \in D_i} d(u, P_i)$$

基于数据驱动确定的 D_{base} 将调节因子 (α, β) 添加至基础阈值上,设定正域划分标准 $T_{\text{pos}} = D_{\text{base}} + \alpha$ 以及负域划分标准 $T_{\text{neg}} = D_{\text{base}} + \beta$ (其中 $\alpha \leq \beta$)。基于新样本到各情感类的最小距离 $\min d = \min_i d(\Delta u, P_i)$ 及对应的类别 i^* ,将新样本 Δu 划入三支决策区域。

1) 正域 (POS_i^*): 接受预测。若 $\min d \leq D_{\text{base}} + \alpha$,则说明该样本符合情感类别 i^* 的特征。

2) 边界域 (BND_i^*): 延迟预测。若 $D_{\text{base}} + \alpha < \min d \leq D_{\text{base}} + \beta$,则说明该样本包含其他情感特征或部分噪声。

3) 负域 (NEG_i^*): 拒绝预测。若 $\min d > D_{\text{base}} + \beta$,则说明该样本偏离已知类别。

该基础阈值基于训练集的距离分布计算得出,当增量数据的分布与训练集存在较大偏移时,阈值的区分效力可能有所下降。然而,三支决策的边界域机制为此提供了一定程度的缓冲。分布偏移导致的不确定样本会被划入边界域而非直接接受或拒绝,从而避免了对概念空间的错误更新。在极端偏移场景下,可考虑周期性地利用累积的正域样本重新估计阈值,以维持划分的合理性,具体过程如算法 3 所示。

算法 3 基于三支边界的新样本情感状态辨识

输入:伪概念空间 $= \{P_1, \dots, P_c\}$,新语音样本 Δu ,特征权重 $\omega(v)$,基础距离阈值 D_{base} ,调节因子 (α, β)

输出:预测标签 \mathcal{L} ,样本所属三支划分区域 Region

1. 提取新样本特征并获取模糊关系矩阵表现 $\tilde{E}(\Delta u, v)$ 与 $\tilde{E}^-(\Delta u, v)$;
2. for each 类别伪概念集合 $P_i \in \mathcal{P}$ do
3. 根据定义 7,计算 Δu 到该类别集合 P_i 的加权辨识距离 $d(\Delta u, P_i)$;
4. end
5. 找出全局最小距离: $\min d = \min_i d(\Delta u, P_i)$;
6. 确定对应的最佳匹配情感类别: $i^* = \arg \min_i d(\Delta u, P_i)$;
7. if $\min d \leq D_{\text{base}} + \alpha$ then
8. $\mathcal{L} = i^*$; Region = POS_i^* , 接受预测;
9. else
10. if $D_{\text{base}} + \alpha < \min d \leq D_{\text{base}} + \beta$ then
11. $\mathcal{L} = i^*$; Region = BND_i^* , 延迟预测;
12. else
13. $\mathcal{L} = \text{Unknown}$; Region = NEG_i^* , 拒绝预测;
14. end
15. end
16. return \mathcal{L} , Region

3.4.2 调节因子在增量学习中的应用与概念空间更新

在增量学习中,符合 D_{base} 的众多新样本可能会将带有噪声的数据加入模型中,导致模型性能下降。因此,本算法采用仅更新正域的策略。当设定 α 为负数时,正域的划分标准变小(即 $T_{\text{pos}} < D_{\text{base}}$)。这意味着模型要求新样本必须具有较小的特征距离,才能进入 POS_i^* 区域。进入该区域的样本会被加入到对象集 $\mathcal{D}_i' = \mathcal{D}_i \cup \{\Delta u\}$,并触发底层概念子空间的重新计算。

对于距离在 $D_{\text{base}} + \alpha$ 与 $D_{\text{base}} + \beta$ 之间的样本,它们会被划分到 BND_i^* 区域。模型会输出这些样本的预测结果,但不

将它们加入到对象集进行概念空间的更新。这种负值调节的方式降低了噪声数据进入模型更新过程的概率,保持了模型在处理增量数据时的稳定性,具体逻辑如算法 4 所示。

算法 4 基于三支正域的概念空间动态增量更新

输入:原决策类对象集 \mathcal{D} ,原概念空间 \mathcal{G} ,伪概念空间 \mathcal{P} ,新样本 Δu ,

预测标签 i^* ,划分区域 Region

输出:更新后的空间 \mathcal{G} 和 \mathcal{P} ;

1. if Region == POS $_i^*$ then
2. 将新样本并入对应的情感决策类: $\mathcal{D}_i^* = \mathcal{D}_i \cup \{\Delta u\}$;
3. 调用算法 1 中的逻辑,根据更新后的 \mathcal{D}_i^* 重新计算并更新底层概念子空间 \mathcal{C}_i^* ;
4. 调用算法 2 中的融合机制,根据新的 \mathcal{C}_i^* 重新寻找包含关系并更新渐进式伪概念子空间 \mathcal{P}_i^* ;
5. 用更新后的 \mathcal{C}_i^* 和 \mathcal{P}_i^* 替换 \mathcal{G} 和 \mathcal{P} 中的对应项;
6. else
7. 样本落入 BND $_i^*$ 或 NEG $_i^*$ 区域,不触发空间更新;
8. end
9. return 更新后的 \mathcal{G} 和 \mathcal{P}

为进一步展示算法的可解释性优势,给出一个简要的推理示例。假设某待分类语音样本经特征提取与模糊化后,其在声学属性上的隶属度向量为(0.85,0.72,0.61)。3WERT-

WFCCCL 首先通过正向认知算子计算该样本与各情感类别伪概念的隶属匹配程度,同时通过负向认知算子评估其在各属性上的非隶属排斥程度,两者结合 ERT 属性权重生成加权辨识距离。若该样本到愤怒类伪概念的距离最小且落入正域范围,则分类算法输出愤怒的分类结果。整个推理过程中,每一步决策均可追溯至具体的属性隶属度与权重数值。相比深度学习模型的黑箱输出,本文方法能够清晰地揭示分类判据的来源。

图 1 展示了本文提出的基于极端随机树加权的三支模糊概念认知学习 3WERT-WFCCCL 算法的整体框架,该框架由 4 个核心模块构成:1)原始语音信号经 Whisper 预训练模型与 MLP 依次处理,提取高维声学特征矩阵并转换成模糊形式背景;2)随后 ERT 算法通过量化各特征的基尼指数下降量计算得到属性权重 $\omega(v)$;3)在认知学习阶段,模型结合属性权重与容错阈值 (α, β) 构建正负双向认知算子 $(\mathcal{M}^\nabla, \mathcal{N}^\nabla)$,生成加权模糊粒概念空间 \mathcal{G} ,并进一步融合为渐进式三支伪概念空间 \mathcal{P} ;4)面对增量样本 Δu ,计算其与 \mathcal{P} 的辨识距离,依据三支决策将样本划入正域、边界域或负域,分别对应接受预测并触发增量更新、延迟预测与拒绝预测,最终输出情感分类标签 \mathcal{L} 。

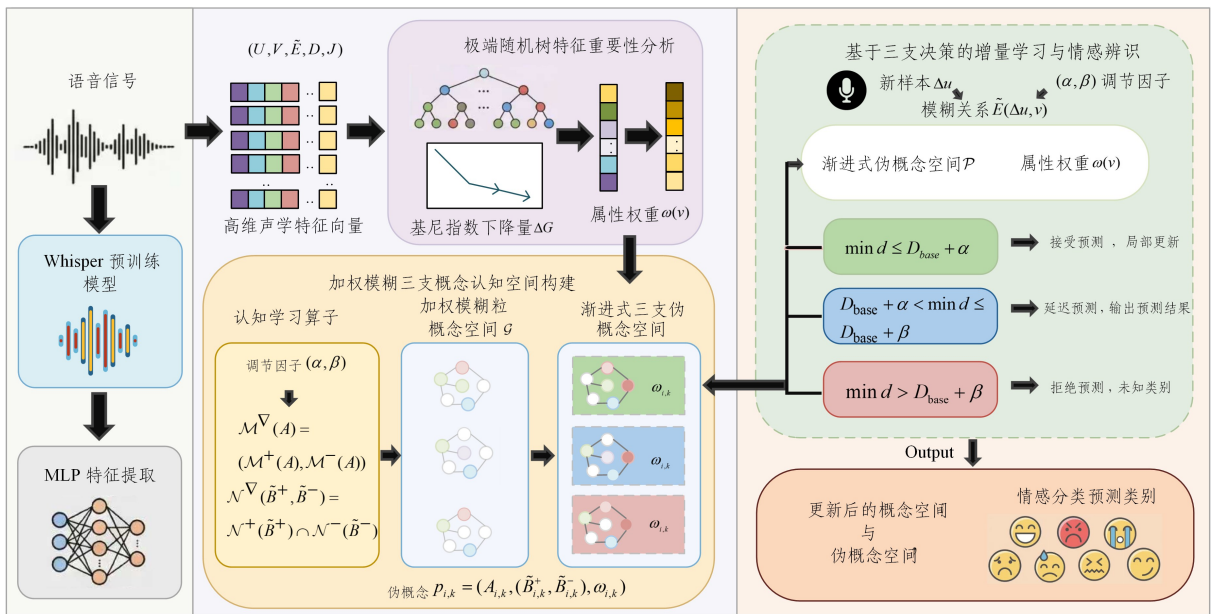


图 1 基于极端随机树加权的三支模糊概念认知学习 3WERT-WFCCCL 算法的整体框架

Fig. 1 Overall framework of the 3WERT-WFCCCL algorithm leveraging ERT weighting for three-way fuzzy concept-cognitive classification

4 实验设置

为评估所提 3WERT-WFCCCL 的实际效果,选取了多个代表性机器学习方法作为基线,在两个广泛使用的语音情感基准数据集上开展对比实验。实验在配备 Apple M2 芯片与 8GB 运行内存、运行 macOS Ventura 13.4 的计算平台上完成,所有代码均以 Python 3.11.7 实现。为降低随机因素的影响,每组评估实验均独立执行 10 轮随机数据划分,并以 10 轮实验所得各项指标的平均值作为最终报告结果。

4.1 数据集介绍

本文选用了两个在学术界广泛使用的基准数据集:柏林情感语音数据库 (EmoDB) 和萨里视听表达情感数据库 (SAVEE),以此保证实验结果的可靠性,并验证模型在不同语言和声学条件下的鲁棒性。

4.1.1 柏林情感语音数据库 (EmoDB)

柏林情感语音数据库^[28]是一个经典的德语纯音频数据集,它包含由 5 男 5 女组成的 10 位专业演员在专业录音室环境下录制的 535 条音频语句。该数据集涵盖了 7 类情感标签,分别为 angry(愤怒)、bored(无聊)、disgust(厌恶)、fear(恐惧)、

happy(开心)、sad(悲伤)与 neutral(中性)。EmoDB 的语音质量极高且背景纯净,非常适合用于评估模型在理想声学环境下的基础情感特征提取和认知分类能力。

4.1.2 萨里视听表达情感数据库(SAVEE)

萨里视听表达情感数据库^[29]是一个由4位男性英语母语者录制的多模态数据集,本研究中仅使用其音频模态。该数据集共包含480个发音样本,涵盖了7类情感标签,即 angry(愤怒)、disgust(厌恶)、fear(恐惧)、happy(开心)、sad(悲伤)、surprise(惊讶)及 neutral(中性)。相比于 EmoDB,SAVEE 数据集的样本量较小,但具有不同的语言和发音特点。

为验证所提出的算法在增量学习场景下的抗概念漂移能力,本文对两个语音数据集进行了特定的划分,划分方式为保留约20%的数据作为固定的测试集,剩余80%的数据作为训练集。上述划分均采用分层抽样策略,确保训练集与测试集中各情感类别的样本比例与原始数据集保持一致,以避免类别不平衡对实验结果的影响。进一步将训练集划分为两个部分,随机抽取其中70%用于构建基础的概念认知空间,剩余30%的训练样本被依次输入模型进行增量更新,每次更新完成后均在固定的测试集上评估模型的综合性能。

4.2 特征提取

在SER系统中,从语音信号中构建有效的情感表征是分类任务成功的前提。早期方法通常依赖人工设计的声学参数,如韵律特征、MFCC及过零率等低层描述子,这类特征虽计算高效,但受限于固定的变换规则,在刻画情感语音中跨时间尺度的高阶语义关联时表现不佳。为解决上述问题,本文引入以Whisper^[5]为代表的大规模预训练语音模型来承担底层特征获取任务,并通过有监督的MLP网络对原始表征进行语义层面的压缩与聚焦,以适配后续模糊概念认知建模的需求。

4.2.1 Whisper 特征

Whisper由OpenAI基于约68万小时的多语言语料进行弱监督预训练,其Transformer编码器在完成语音识别的过程中同步习得了丰富的跨语言声学规律与隐式语义模式,使所提取的表征兼顾声学细粒度与语言层语义,在SER任务中展现出优于传统手工特征的判别能力。本文采用Whisper-Base变体进行特征提取,流程如下:将原始音频重采样至16kHz并统一补零或截断至固定长度,再计算对数梅尔声谱图送入Transformer编码器,对末层隐状态沿时间轴进行均值池化,得到512维全局表征向量 $\mathbf{x}_{\text{whisper}} \in \mathbb{R}^{512}$,最终通过StandardScaler对全部维度的特征执行零均值单位方差标准化。

4.2.2 MLP 特征精炼网络

Whisper编码器输出的512维表征携带了充分的声学与语义线索,但其中不可避免地混入了与情绪状态无关的信道背景等干扰因素。此外,直接将512维向量输入模糊概念格会使格节点数目随维度呈指数级膨胀,导致认知学习阶段的计算资源消耗大幅增加。因此,本文设计了一个以情感类标签为监督信号的MLP特征精炼网络,将高维通用表征映射至面向情感分类的低维紧凑空间,同时降低后续概念构建的复杂度。该网络由以下三级变换组成。

第一层是压缩层(512 → 256维):全连接线性映射将输入维度减半,后接批归一化稳定激活分布,配合ReLU引入非线性,并以Dropout($p=0.3$)抑制高维空间下的协同适应。

第二层为聚焦层(256 → 128维):在上一层基础上进一步压缩特征维度,保留对情感区贡献最显著的方向,结构设计沿用批归一化+ReLU+Dropout($p=0.2$)的组合方式。

最一层是输出层(128 → 64维):经末级全连接得到64维情感判别特征向量,该向量将作为三支模糊概念认知模型的条件输入。

在提炼特征的训练阶段于64维输出端接入线性分类头,以交叉熵nn.CrossEntropyLoss为优化目标驱动整个网络进行端到端学习;同时优化器采用Adam,初始学习率设为0.0005,并配合CosineAnnealingWarmRestarts调度策略以防止陷入局部极值。

4.3 模型分类与评估

完成特征提取后,本节就3WERT-WFCCL的分类效果展开系统性验证。在将声学特征输入模型前,有必要对其进行预处理:Whisper与MLP所输出的特征向量以连续实数形式存在,而模糊概念认知框架要求条件属性以隶属度值表达。若将实数特征直接嵌入模糊概念空间,一方面会导致概念节点的粒度分布极不均匀,另一方面,大量稀疏节点会显著增加概念结构的构建开销。为此,本文对全部声学特征执行基于极差归一化的模糊化变换,使各属性值落在 $[0,1]$ 之间,具体公式如下:

$$\mu(u_i, v_j) = \frac{x(u_i, v_j) - x_{\min}(v_j)}{x_{\max}(v_j) - x_{\min}(v_j)}$$

其中, $\mu(u_i, v_j)$ 表示样本 u_i 在声学属性 v_j 上经模糊化后的隶属度, $x(u_i, v_j)$ 为原始特征观测值, $x_{\max}(v_j)$ 与 $x_{\min}(v_j)$ 分别为该属性在训练集上的全局上界与下界。隶属度值越高,意味着该样本在对应声学维度上的表现越好,由此生成的模糊概念可被解释为一类具有特定声学强度的情感感知单元。经模糊化处理后的特征数据随后作为3WERT-WFCCL的形式背景输入,进入概念认知分类阶段。

4.4 评价指标

本文依据混淆矩阵中真正例 TP_i 、真负例 TN_i 、假正例 FP_i 和假负例 FN_i 4类统计量,选取准确率(Accuracy)、精确率(Precision)、召回率(Recall)和Macro F1分数4项关键指标来科学评估模型性能。

准确率表示正确预测的样本占总样本的比例,反映模型的整体分类效能。其计算式为:

$$Accuracy = \frac{\sum_i (TP_i + TN_i)}{\sum_i (TP_i + TN_i + FP_i + FN_i)}$$

精确率度量的是被判为正类的样本中真正属于该类的占比,体现了预测的精准程度;召回率则度量真正属于某类的样本中被成功识别的占比,体现了模型对该类样本的覆盖程度。两者的计算式分别为:

$$Precision = \frac{\sum_i TP_i}{\sum_i (TP_i + FP_i)}$$

$$Recall = \frac{\sum_i TP_i}{\sum_i (TP_i + FN_i)}$$

考虑到各情感类别样本量存在差异,直接以整体指标进行评价可能导致偏倚,本文选用 Macro F1 值作为核心评价指标。该指标对每种情感类别赋予同等评估权重,因此能够更全面地衡量模型在多类别场景下的整体判别水平。为便于表述,后续实验中统一将 Macro F1 简记为 F1 分数。其计算式为:

$$\text{Macro F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5 实验结果与分析

为了全面评估本文提出的 3WERT-WFCCCL 鲁棒增量学习算法的有效性,本章的实验设计旨在回答以下 3 个关键研究问题:1) 阈值参数的选择如何影响模型的性能与稳定性? 2) 基于三支决策的 POS 区域更新策略是否能有效提升增量学习的抗噪能力? 3) 与现有的经典机器学习分类器以及基础二支 CCL 算法相比,3WERT-WFCCCL 的总体表现如何? 针对上述问题,本章将依次进行详细的验证与分析。

5.1 参数敏感性分析

在 3WERT-WFCCCL 算法中, α 和 β 是构建三支决策概念空间的两个核心参数,它们分别决定了高置信度 POS 区域的准入门槛和低置信度 NEG 区域的拒绝边界,其取值变化会直接改变增量学习过程中的样本划分结果,进而对最终的性能产生显著影响。在 EmoDB 数据集上的初步网格搜索实验表明,最优参数区间与 SAVEE 上的结果一致,均落在 $\alpha \in [-0.01, 0.005]$, $\beta \in [0.005, 0.02]$ 的范围内。考虑到两个数据集的响应趋势具有较高的一致性,且 SAVEE 的特征边界更为复杂且更具分析价值,本节以 SAVEE 数据集为代表性案例进行阈值参数的敏感性分析。

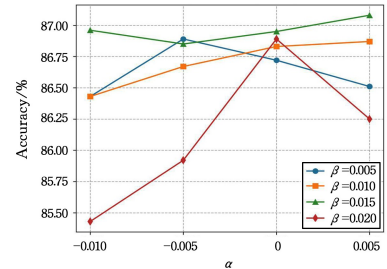
设定高置信度阈值参数 α 的区间为 $[-0.01, -0.005]$, $[0.0, 0.005]$, 低置信度阈值参数 β 的区间为 $[0.005, 0.01]$, $[0.015, 0.02]$ 。这两种参数两两组合共构成 16 组网格搜索实验。表 1 详细记录了 3WERT-WFCCCL 模型在不同参数配置下的分类准确率与 F1 分数,同时图 2 直观展示了模型性能的动态演变趋势。

表 1 SAVEE 上不同阈值参数组合的分类性能

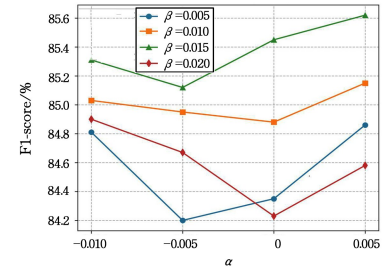
Table 1 Classification performance of different threshold parameter combinations on the SAVEE dataset

$\alpha(\text{POS})$	$\beta(\text{BND})$	准确率(Acc)	Macro-F1
-0.010	0.005	86.43	84.81
-0.010	0.010	86.43	85.03
-0.010	0.015	86.96	85.31
-0.010	0.020	85.43	84.90
-0.005	0.005	86.89	84.20
-0.005	0.010	86.67	84.95
-0.005	0.015	86.85	85.12
-0.005	0.020	85.92	84.67
0.000	0.005	86.72	84.35
0.000	0.010	86.83	84.88
0.000	0.015	86.95	85.45
0.000	0.020	86.89	84.23
0.005	0.005	86.51	84.86
0.005	0.010	86.87	85.15
0.005	0.015	87.08	85.62
0.005	0.020	86.25	84.58

结合图 2 能够发现,模型分类性能随 α 和 β 的调整呈现出明显波动。 α 决定了模型更新的严苛程度。 α 取值过小(如 -0.01),会让高置信度区域的判定条件变得极其严格,这会导致漏掉增量数据中大量蕴含一定信息量但特征略微偏移的有效样本,进而导致模型无法充分学习新知识。此时模型虽然保留了一定程度的鲁棒性,但整体准确率基本停留在 $86.43\% \sim 86.96\%$,难以实现性能突破。参数 β 直接关系到模型的容错空间。 β 值增大至 0.02 时,会大幅拓宽边界区域并过度压缩拒绝区域,这种过于宽松的准入机制极易导致特征模糊或含有严重噪声的样本混入增量更新环节,进而污染概念空间。在绝大多数 α 下, β 增至 0.02 均引发了分类性能不同程度的回落,在 $\alpha = -0.01$ 且 $\beta = 0.02$ 时,准确率更是明显下滑至 85.43% 。调整参数配置为 $\alpha = 0.005$ 与 $\beta = 0.015$ 时,模型达到全局最优状态,分别取得 87.08% 的最高准确率和 85.62% 的最优 F1 分数。这些数据表现证明了,赋予模型适当宽松的高置信度准入阈值并结合适度的边界容错空间,能够最大程度地平衡吸收新知识与抵御噪声干扰之间的矛盾。



(a) Accuracy sensitivity analysis



(b) F1-score sensitivity analysis

图 2 SAVEE 数据集上模型分类性能对不同阈值参数的敏感性变化趋势

Fig. 2 Trend of sensitivity of model classification performance to different threshold parameters on the SAVEE dataset

5.2 鲁棒增量更新策略有效性分析

第 5.1 节确定了构建三支决策概念空间的阈值参数组合。划分正域边界域和负域的目的旨在优化增量学习阶段的模型更新过程,为了验证仅利用高置信度样本进行更新这一策略的有效性,本节设计了更新机制的消融实验。实验在 EmoDB 和 SAVEE 两个数据集上测试了 3WERT-WFCCCL 算法在两种更新策略下的分类结果:第一种是全局更新策略,这是一种基础的增量学习逻辑,当新样本被模型预测正确时,会被加入到现有概念空间进行更新;第二种是鲁棒更新策略,即本节采用的三支决策逻辑,模型只接受预测正确且特征距离满足高置信度门槛的样本来更新概念空间,同时把落入边

界域的样本删除。

表 2 详细记录了这两种更新策略在两个数据集上的各项性能指标,从具体数值可以看出,两种策略在不同复杂度的数据集上呈现出差异化结果。在背景噪声极小且情感特征高度纯粹的 EmoDB 数据集中,全局更新策略的准确率为 97.69%,略优于鲁棒更新策略。图 3 则直观展示了算法性能在不同场景下的对比演变。如柱状图所呈现,在特征边界更为复杂的情感数据集 SAVEE 上,全局更新策略的准确率仅为 86.92%,而采用鲁棒更新策略后模型准确率升到了 87.08%。

表 2 两种增量更新策略在不同数据集上的分类性能对比

Table 2 Classification performance comparison of two incremental update strategies on different datasets

数据集	更新策略	准确率 (Acc)	精确率 (Pre)	召回率 (Rec)	Macro-F1
EmoDB	全局更新	97.69	97.32	97.50	97.38
	鲁棒更新	97.42	97.06	97.27	97.12
SAVEE	全局更新	86.92	85.68	85.79	85.44
	鲁棒更新	87.08	85.83	85.71	85.62

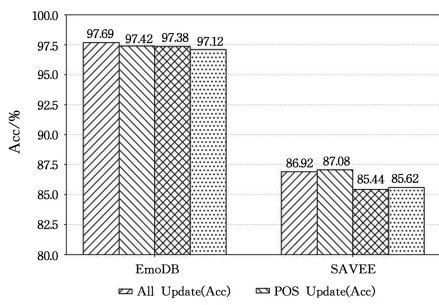


图 3 两种增量更新策略在不同数据集上的性能对比

Fig. 3 Performance comparison of two incremental update strategies on different datasets

EmoDB 数据集属于高质量实验室录音数据,绝大多数样本特征极其清晰。在这种极度纯净的环境下,全局更新策略能够最大程度地吸收所有正确样本的细微特征变化,从而触及极高的分类上限,而鲁棒更新的严格准入机制反而可能滤除掉一部分有效的概念外延信息。相反,在处理 SAVEE 这类复杂边界相对模糊的数据时,部分样本虽然在当前阶段被勉强预测正确,但其声学特征可能处于两种不同情感的过渡区域。将这些低质量的边缘样本直接用于更新会导致原本清晰的概念中心发生偏移。随着增量学习的不断推进,概念空间会变得逐渐模糊,从而引起整体分类性能的下降。在实际应用场景中,数据质量往往无法保证如此纯净,鲁棒策略在特征边界更复杂的 SAVEE 数据集上展现出的优势更具现实参考意义。因此,本文选择以牺牲极少量纯净数据上的性能为代价,换取模型在复杂环境下更强的泛化能力和整体稳定性。三支决策机制通过正域的准入条件对样本进行了筛选,有效过滤了边界模糊样本对已有概念空间的干扰。这表明,鲁棒更新策略在面对理想纯净数据时略显保守,但在复杂的增量学习环境中能够保留模型对各情感类别本质特征的刻画,保持了更优抗噪能力以及稳定性。

5.3 总体性能对比分析

本节通过比较 3WERT-WFCCL 算法与多种分类算法的评价指标,进一步评估其在语音情感识别任务中的整体表现。表 3 和表 4 列出了各算法在 EmoDB 和 SAVEE 数据集上的准确率、精确率、召回率以及 F1 分数。3WERT-WFCCL 引入三支决策机制后有效克服了二支决策模型处理模糊特征时的缺陷。如表 3 和表 4 所列,3WERT-WFCCL 算法在两个数据集上的表现均超越二支决策 ERT-WFCCL。EmoDB 数据集上的准确率由 97.01% 提高至 97.42%,SAVEE 数据集上的准确率由 85.42% 提高至 87.08%,证明了高置信度筛选机制对模糊概念认知模型性能提升的作用。

表 3 不同算法在 EmoDB 数据集上的总体性能对比

Table 3 Overall performance comparison of different algorithms on

EmoDB dataset				
方法	准确率 (Acc)	精确率 (Pre)	召回率 (Rec)	Macro-F1
3WERT-WFCCL	97.42	97.06	97.27	97.12
ERT-WFCCL	97.01	96.59	96.81	96.63
Logistic Regression	95.89	95.90	95.79	95.84
Random Forest	95.51	95.47	95.58	95.48
Neural Network	95.51	95.59	95.42	95.49
Naive Bayes	94.95	94.81	95.10	94.90
KNN	92.15	94.25	91.74	92.67
Decision Tree	90.65	90.31	89.85	90.00
AdaBoost	84.67	87.74	83.17	83.84

表 4 不同算法在 SAVEE 数据集上的总体性能对比

Table 4 Overall performance comparison of different algorithms on

SAVEE dataset				
方法	准确率 (Acc)	精确率 (Pre)	召回率 (Rec)	Macro-F1
3WERT-WFCCL	87.08	85.83	85.71	85.62
Logistic Regression	86.46	86.09	85.48	85.22
ERT-WFCCL	85.42	85.72	84.29	84.20
Neural Network	85.21	85.28	84.52	84.10
Random Forest	83.33	83.14	82.74	82.53
KNN	81.46	80.77	79.64	80.12
Naive Bayes	80.21	80.33	80.12	79.74
AdaBoost	77.08	76.71	76.31	76.19
Decision Tree	75.83	74.76	74.76	74.41

在与其他机器学习算法的对比中,逻辑回归 LR 等统计学习模型依赖在特征空间中寻找全局决策边界。如表 3 所列,面对 EmoDB 这类样本特征分布相对独立且边界清晰的数据集,全局优化方法能够达到较高的拟合精度,其中 LR 算法在 EmoDB 数据集上的准确率为 95.89%,在 SAVEE 数据集上的准确率为次优。3WERT-WFCCL 采用模糊粒化机制模拟人类认知过程,模型通过计算声学特征与概念内涵之间的隶属度,将数据的不确定性转换为模糊概念表达,不同情感之间的特征交叠问题得到有效缓解。凭借该机制,3WERT-WFCCL 在 EmoDB 上达到 97.42% 的最高准确率。由表 4 可知,在处理特征边界更为复杂且存在交叠的 SAVEE 数据集时,传统决策树和随机森林等硬划分模型容易产生误判,它们

的准确率分别降至 75.83% 和 83.33%。3WERT-WFCCl 在该数据集上的准确率为 87.08%，优于 LR 与 KNN。该现象表明，基于模糊概念学习的模型在处理特征交叠的小样本语音数据时具备突出的泛化能力，同时规避了神经网络模型在小样本集上容易引发的过拟合风险。

结束语 本文提出了一种基于极端随机树加权的三支模糊概念认知分类模型，通过将三支决策理论与模糊加权概念认知学习相融合，有效解决了传统 CCL 模型在处理增量语音情感数据时易受噪声干扰而引发概念漂移的问题。模型采用 Whisper 提取高维语音特征，并经 MLP 进行分层抽象表示；之后在分类阶段中引入 ERT 算法实现特征权重的自动化分配，在认知算子中嵌入容错阈值构建正负双向认知机制；面对增量数据时，依据特征辨识距离将样本划分为正域、边界域与负域，并仅利用正域样本更新概念空间，增强了模型的鲁棒性与可解释性。在 EmoDB 和 SAVEE 两个公开数据集上的实验结果表明，3WERT-WFCCl 在准确率与 F1 分数等关键指标上均优于现有主流基线方法，验证了三支决策机制引入的有效性。

然而，本文方法仍存在一定局限性。一方面，模型中容错阈值参数 α 与 β 的设定目前依赖经验调节，缺乏自适应确定机制；另一方面，Whisper 与 MLP 的组合特征提取流程引入了较高的计算开销，在资源受限的实时场景中存在部署挑战。未来工作将探索阈值参数的自动化学习策略，例如引入贝叶斯优化或基于交叉验证的自动调参机制，根据增量数据的分布特征动态调整阈值，以替代当前依赖网格搜索的方式，进一步提升模型在不同数据分布下的泛化能力；还需研究模型的轻量化设计方案，降低计算复杂度，推动该方法在实时人机交互系统中的实际应用。此外，本文实验尚未严格遵循说话人独立原则进行数据划分，这可能在一定程度上高估模型的泛化性能。未来工作还将采用说话人独立的实验设置，确保训练集与测试集之间无说话人重叠，以更严格地验证模型在未知说话人条件下的鲁棒性与实际部署能力。

参 考 文 献

- [1] SCHULLER B. Speech emotion recognition: two decades in a nutshell, benchmarks, and ongoing trends [J]. *Communications of the ACM*, 2018, 61(5): 90-99.
- [2] HOOK J, NOROOZI F, TOYGAR O, et al. Automatic speech based emotion recognition using paralinguistics features [J]. *Bulletin of the Polish Academy of Sciences: Technical Sciences*, 2019, 67(3): 479-488.
- [3] INGALE A, CHAUDHARI D. Speech emotion recognition [J]. *International Journal of Soft Computing and Engineering*, 2012, 2(1): 235-238.
- [4] EYBEN F, SCHERER K R, SCHULLER B W, et al. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing [J]. *IEEE Transactions on Affective Computing*, 2016, 7(2): 190-202.
- [5] RADFORD A, KIM J W, XU T, et al. Robust speech recognition via large-scale weak supervision [C]// *International Conference on Machine Learning*. PMLR, 2023: 28492-28518.
- [6] PEPINO L, RIERA P, FERRER L. Emotion recognition from speech using wav2vec 2.0 embeddings [C]// *Interspeech*. 2021: 3400-3404.
- [7] MA W, RIVERA H, VALERI J, et al. emotion2vec: Self-supervised pre-training for speech emotion representation [J]. *arXiv*: 2312.15185, 2023.
- [8] LIU Z T, XU J P, WU M, et al. Review of emotional feature extraction and dimension reduction for speech emotion recognition [J]. *Chinese Journal of Computers*, 2018, 41(12): 2833-2851.
- [9] JOY J, KANNAN A, RAM S, et al. Speech emotion recognition using neural network and MLP classifier [J]. *International Journal of Engineering Science and Computing*, 2020, 10(4): 25170-25173.
- [10] KAUR J, KUMAR A. Speech emotion recognition using CNN, k-NN, MLP and random forest [C]// *Computer Networks and Inventive Communication Technologies: Proceedings of Third ICCNCT 2020*. Singapore: Springer, 2021: 499-509.
- [11] CHEN L, MAO X, XUE Y, et al. Speech emotion recognition: features and classification models [J]. *Digital Signal Processing*, 2012, 22(6): 1154-1160.
- [12] ISSA D, DEMIRCI M, YAZICI A. Speech emotion recognition with deep convolutional neural networks [J]. *Biomedical Signal Processing and Control*, 2020, 59: 101894.
- [13] AHMED M, ISLAM S, ISLAM A, et al. An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition [J]. *Expert Systems with Applications*, 2023, 218: 119633.
- [14] GUO D, XU W, DING W, et al. Concept-cognitive learning survey: mining and fusing knowledge from data [J]. *Information Fusion*, 2024, 109: 102426.
- [15] SHI Y, MI Y, LI J, et al. Concept-cognitive learning model for incremental concept learning [J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2018, 51(2): 809-821.
- [16] MI Y, LIU W, SHI Y, et al. Semi-supervised concept learning by concept-cognitive learning and concept space [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2020, 34(5): 2429-2442.
- [17] XU W, GUO D, QIAN Y, et al. Two-way concept-cognitive learning method: a fuzzy-based progressive learning [J]. *IEEE Transactions on Fuzzy Systems*, 2022, 31(6): 1885-1899.
- [18] ZHANG T, RONG M, SHAN H, et al. Stability analysis of incremental concept tree for concept cognitive learning [J]. *International Journal of Machine Learning and Cybernetics*, 2022, 13(1): 11-28.
- [19] DENG X, LI J, QIAN Y, et al. An emerging incremental fuzzy concept-cognitive learning model based on granular computing and conceptual knowledge clustering [J]. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2024, 8(3): 2417-2432.
- [20] XIN X W, YU H Y, XUE Z, et al. A novel fuzzy concept-cogni-

- itive learning model with attribute fluctuation and concept clustering [J]. *IEEE Transactions on Fuzzy Systems*, 2025, 33(10): 3570-3581.
- [21] QI J, WEI L, REN R. 3-way concept analysis based on 3-valued formal contexts [J]. *Cognitive Computation*, 2022, 14(1): 1900-1912.
- [22] LI J, HUANG C, QI J, et al. Three-way cognitive concept learning via multi-granularity [J]. *Information Sciences*, 2017, 378: 244-263.
- [23] YUAN K, XU W, LI W, et al. An incremental learning mechanism for object classification based on progressive fuzzy three-way concept [J]. *Information Sciences*, 2022, 584(1): 127-147.
- [24] XU W H, JIANG D. A novel concept-cognitive learning model oriented to three-way concept for knowledge acquisition [J]. *IEEE Transactions on Big Data*, 2025, 11(5): 2779-2791.
- [25] GUO D, XU W. Fuzzy-based concept-cognitive learning: an investigation of novel approach to tumor diagnosis analysis [J]. *Information Sciences*, 2023, 639: 118998.
- [26] GUO D, XU W, QIAN Y, et al. M-FCCL: memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion [J]. *Information Fusion*, 2023, 100: 101962.
- [27] GUO D, XU W, QIAN Y, et al. Fuzzy-granular concept-cognitive learning via three-way decision: performance evaluation on dynamic knowledge discovery [J]. *IEEE Transactions on Fuzzy Systems*, 2024, 32(1): 1409-1423.
- [28] BURKHARDT F, PAESCHKE A, ROLFES M, et al. A database of German emotional speech [C] // *InterSpeech*. 2005: 1517-1520.
- [29] HAQ S, JACKSON P J B. Multimodal emotion recognition [C] // *Machine Audition: Principles, Algorithms and Systems*. IGI Global, 2010: 398-423.



XU Weihua, born in 1979, Ph.D, professor, Ph.D supervisor. His main research interests include granular computing, cognitive computing, data mining and machine learning.

(责任编辑:柯颖)