基于极端随机树加权模糊概念认知的语音情感识别

胡开平[1)] 徐伟华[1)] 钱宇华[2)]

[1)](西南大学人工智能学院 重庆 400715)

[2)](山西大学大数据科学与产业研究院 太原 030006)

摘 要 语音情感识别(Speech Emotion Recognition,SER)在提升人机交互体验中扮演着关键角色。当前主流深度学习模型虽在 SER 任务上取得了优异精度,但普遍存在可解释性不足的问题。为突破这一局限,本文借鉴概念认知学习(Concept-cognitive Learning,CCL)思想,将形式概念分析(Formal Concept Analysis,FCA)与粒计算(Granular Computing,GrC)理论相结合,提出一种基于极端随机树的加权模糊概念认知分类模型(Extremely Randomized Trees based Weighted Fuzzy Concept-cognitive Learning,ERT-WFCCL)。该模型采用大规模预训练语音模型 Whisper 作为特征提取器,从原始语音信号中获取高层语义表征,并通过多层感知机制(Multi-layer Perceptron,MLP)将高维特征解耦为具有明确语义的低维属性向量,以构建形式背景;在分类阶段,利用极端随机树(Extremely Randomized Trees,ERT)评估各属性对情感分类的贡献度,据此构造加权模糊概念空间,并在该概念空间中完成分类推理。该框架以感知到认知的路径为主线,实现了从连续声学信号到可解释概念空间的映射,为 SER 任务提供了一种具备内生可解释性的建模思路。本文在多伦多情感语音集(Toronto Emotional Speech Set,TESS)和众包情感多模态演员数据集(Crowd-sourced Emotional Multimodal Actors Dataset,CREMA-D)上进行了实验。结果表明,ERT-WFCCL 在 TESS 上取得 99.60%的分类准确率和 99.50%的宏 F1 值,在更具挑战性的 CREMA-D 上取得 85.71%的准确率和 85.81%的宏 F1 值,在多项指标上优于支持向量机(Support Vector Machine,SVM)、逻辑回归(Logistic Regression,LR)等对比方法。此外,增量学习实验验证了所提加权模糊概念认知更新策略能够在保持较低计算开销的前提下稳定更新概念空间,展现出较强的鲁棒性和可扩展能力。

关键词 概念认知学习;粒计算;形式概念分析;语音情感识别;认知学习

中图法分类号 TP18 **DOI 号** 10.11897/SP.J.1016.2026.02056

Speech Emotion Recognition Based on Weighted Fuzzy Concept Cognition Using Extremely Randomized Trees

HU Kai-Ping[1)] XU Wei-Hua[1)] QIAN Yu-Hua[2)]

[1)](College of Artificial Intelligence, Southwest University, Chongqing 400715)

[2)](Institute of Big Data Science and Industry, Shanxi University, Taiyuan 030006)

Abstract Speech Emotion Recognition (SER) plays a key role in improving the naturalness and effectiveness of human-computer interaction by enabling machines to perceive and respond to the emotional states conveyed in human speech. In recent years, deep learning models such as convolutional neural networks and recurrent architectures have achieved high recognition accuracy on various SER benchmarks. However, these models generally operate as black boxes and suffer from insufficient interpretability, making it difficult to provide transparent and traceable explanations for their classification decisions. This interpretability limitation restricts the deployment of

收稿日期:2025-10-07;在线发布日期:2026-04-07。本课题得到国家自然科学基金项目(62376229,12371465)、重庆市自然科学基金(CSTB2023NSCQ-LZX0027)资助。**胡开平**,硕士研究生,主要研究领域为概念认知学习、粒计算、语音情感识别。E-mail:hkp0922@email.swu.edu.cn。**徐伟华(通信作者)**,博士,教授,博士生导师,主要研究领域为粒计算、认知计算、数据挖掘、机器学习。E-mail:chxuwh@swu.edu.cn。**钱宇华**,博士,教授,博士生导师,主要研究领域为人工智能、大数据、机器学习。

SER systems in trust-sensitive and safety-critical applications such as mental health monitoring, intelligent customer service, and affective computing. To overcome this limitation, this paper draws on the theory of Concept-cognitive Learning (CCL), which integrates Formal Concept Analysis (FCA) and Granular Computing (GrC) to model human-like cognitive processes of concept formation and classification. Based on this theoretical foundation, we propose a novel classification framework termed Extremely Randomized Trees based Weighted Fuzzy Concept-cognitive Learning (ERT-WFCCL). The proposed framework comprises three main stages: feature extraction, feature decoupling, and concept-cognitive classification. In the feature extraction stage, the model employs Whisper, a large-scale pre-trained speech model, as the backbone feature extractor to obtain high-level semantic representations from raw audio waveforms. Whisper benefits from extensive multilingual and multitask supervised training, enabling it to capture rich acoustic and linguistic information relevant to emotional expression. In the feature decoupling stage, a Multi-layer Perceptron (MLP) based module is designed to transform the high-dimensional Whisper features into low-dimensional attribute vectors with explicit semantic meaning. Each dimension of the decoupled vector corresponds to a specific interpretable attribute, which enables the construction of a formal context as the standard input structure required by FCA. This decoupling step is critical for bridging the gap between deep feature representations and concept-based reasoning. In the classification stage, the Extremely Randomized Trees (ERT) algorithm is applied to evaluate the discriminative contribution of each decoupled attribute to emotion classification. Based on the resulting feature importance scores, a weighted fuzzy concept space is constructed, in which each attribute is assigned a weight reflecting its relevance to the target emotion categories. Classification inference is then performed by navigating this weighted fuzzy concept space according to the Concept-cognitive Learning strategy. This design endows the model with inherent interpretability, as each classification decision can be traced back through the concept lattice to specific attributes and their weighted contributions. The overall framework follows a perception-to-cognition pathway, achieving a mapping from continuous acoustic signals to an interpretable concept space and providing an inherently interpretable modeling approach for SER tasks. Comprehensive experiments are conducted on two widely used public SER datasets: the Toronto Emotional Speech Set (TESS), which contains 2800 utterances covering seven emotion categories, and the Crowd-sourced Emotional Multimodal Actors Dataset (CREMA-D), which comprises 7442 utterances from 91 actors across six emotion categories with greater speaker and recording variability. The proposed ERT-WFCCL model achieves a classification accuracy of 99.60% and a macro F1 score of 99.50% on TESS, and an accuracy of 85.71% with a macro F1 score of 85.81% on the more challenging CREMA-D, outperforming comparison methods including Support Vector Machine (SVM) and Logistic Regression (LR) across multiple evaluation metrics. Furthermore, incremental learning experiments verify that the proposed weighted fuzzy concept-cognitive update strategy can stably update the concept space with low computational overhead, demonstrating strong robustness and scalability.

Keywords　concept-cognitive learning; granular computing; formal concept analysis; speech emotion recognition; cognitive learning

1 引 言

近年来，人机交互（Human-Computer Interaction，HCI）领域飞速发展，其核心目标在于研究人类如何与计算机互动，并利用技术进步提升交互效率[1]。语音是传递情感与感受的有力媒介，而情感识别则是指识别、理解并响应情感的能力[2]。在这一背景下，语音识别通过语音直接驱动计算机完成特定任务，以及语音情感识别（SER）利用人类情感进行交互，这两个内容已成为促进人机交流的关键研究方向[3]。准确理解人类的情感表达，不仅能优化用户体验，更是实现个性化服务的基础。然而，现有 SER 方法在追求高性能的同时，往往以模型可解释性的显著下降为代价，使得情感判别过程难以从认知层面进行理解。

一个典型的 SER 系统流程通常包括预处理、特征提取和情感分类三个阶段。在特征提取阶段，研究者致力于从语音信号中发掘有效的声学特征，如连续语音特征、音质特征、谱特征、Teager 能量算子特征以及应用广泛的低级描述符（Low-Level Descriptors，LLDs）[4]。LLDs 能够捕捉语音信号的短时变化，涵盖梅尔频率倒谱系数、对数梅尔谱图、过零率和均方根等时域与频域信息，从而反映语音中的情感表现[5]。尽管 LLDs 在早期研究中发挥了重要作用，这类特征在捕捉短时声学模式方面具有明确物理意义，但难以建模情感表达中跨时间尺度的高层语义与上下文依赖。近年来，随着大规模自监督学习（Self-Supervised Learning，SSL）的兴起，基于预训练的语音表征模型展现出了卓越的性能。例如，WavLM[6]通过掩码预测任务学习全栈语音处理能力，emotion2vec[7]则针对性地提出了首个通用的情感表征模型，显著提升了跨任务的情感识别精度。此外，Whisper[8]凭借其在海量弱监督数据上的训练，不仅具备强大的语音识别能力，其编码器输出的深层特征也被证明富含高鲁棒性的声学与语义信息。受此启发，本研究采用 Whisper 作为特征提取器，以获取比传统特征更具判别力的语音表征。但是如何从这些高维特征中选择性地关注关键信息，以更贴近人类大脑高效整合上下文信息的认知模式[9]，仍然是一个开放性问题。为了模拟认知过程中对初始感知信息进行逐层抽象和提炼的过程，需要一个有效的特征变换模块来架起这座桥梁。深度神经网络，如多层感知机（MLP）[10]凭借其强大的非线性映射能力实现这种分层特征学习。

在情感分类层面，方法的演进历程清晰地揭示了当前面临的方法学瓶颈。早期研究多采用传统机器学习模型，如支持向量机（SVM）[11]、隐马尔可夫模型（Hidden Markov Model，HMM）[12]，朴素贝叶斯（Naive Bayes，NB）[13]。这些模型虽然具备一定的可解释性，但在处理复杂、非线性的情感数据时能力有限。随着技术发展，以卷积神经网络（Convolutional Neural Network，CNN）[14]、循环神经网络（Recurrent Neural Network，RNN）[15]为代表的深度学习模型因其强大的复杂情感模式学习能力，在 SER 任务中取得了显著进展。然而，这种性能的提升是以牺牲可解释性为代价的。这些模型内部复杂的、逐层抽象的变换过程使其决策依据难以被直观理解，形成了所谓的黑箱问题。这一本质缺陷极大地限制了它们在心理健康监测、司法鉴定等需要高度透明度和可信度的场景中的应用。为解决这一可解释性危机，本研究不再单纯追求端到端的黑箱拟合，而是转向探索一种更具认知合理性的新范式——概念认知学习（CCL）。CCL 模拟人类通过概念来认知世界的过程，它通过显式刻画对象-属性-决策之间的概念结构，使模型决策可被解释为一系列概念选择与关系推理过程，而非隐式权重叠加。在 SER 任务中，我们并非直接处理数值，而是将其抽象为“高音调”、“快语速”等概念，并基于概念间的逻辑关系进行推理，比如“高音调且快语速”通常意味着“愤怒”。

概念认知学习是人工智能与认知计算领域新兴的学习范式[16]，它是以概念为知识的基本载体，研究事物认知学习过程的科学[17]。与深度学习的端到端拟合不同，CCL 框架下的每一个决策都能追溯到其所依据的概念，从而天然地具备了良好的可解释性。为提升模型性能，研究者们发展了多种 CCL 模型，如记忆 CCL[18]、增量 CCL[19]、双向 CCL[20]、三支 CCL[21]、渐进式 CCL[22]、模糊 CCL[23]、加权 CCL[24]等。传统的模糊概念分析往往默认所有属性对决策的贡献是等权的，这与 SER 任务中不同声学特征具有不同判别力的事实不符。为此，Zhang 等人[25]提出了一种通过信息熵获取加权概念而无需依赖先验知识的方法；为缩减概念规模，Singh 等人[26]则研究了属性权重间的相关性以优化模糊概念的计算；Zhang 等人[27]则从决策信息粒度的一致性中提取条件属性的权重，进一步提升了模型的鲁棒性。在这些工作的基础上，为增强模型的泛化能

力，研究者们引入了集成学习。其中，Xu 等人[28]的研究表明，将随机森林（Random Forest，RF）与模糊 CCL 结合的集成方法能够优化预测结果。但是，RF 尽管在许多任务中表现出色，其训练时需寻找最优分裂点，导致训练过程缓慢，且在某些情况下泛化能力较弱。针对此问题，本研究引入极端随机树（ERT）算法来测量判别力。ERT 不仅通过完全随机分裂显著降低了计算开销，还能客观量化高维深度特征的基尼指数贡献，使概念权重不再是经验设定，而是由决策行为统计驱动。从而为模糊概念空间提供精确的属性权重，解决了 CCL 中的特征赋权难题。基于以上思考，本文的核心目的在于设计并深入阐述一个全新的、受认知科学启发的 SER 方法学框架。本文的主要贡献如下：

（1）首次将 CCL 方法应用于 SER 任务，为该领域提供了全新的研究视角和解决方案。该方法利用模糊概念的层级结构，有效解决了情感状态的模糊边界问题，在保证高性能的同时提供了内生的可解释性。此外，还通过特征可视化与重要性分析，验证了模型在保持高分类性能的同时，具备稳定且一致的认知解释能力。

（2）提出一种加权模糊 CCL 模型（Fuzzy-based concept-cognitive learning method with Extremely Randomized Trees，ERT-WFCCL）。该模型利用极端随机树计算特征重要性得分，实现了对模糊概念空间中属性权重的动态、自适应分配，有效提升了概念推理的判别能力。

（3）设计并集成了一个基于 Whisper 预训练特征与 MLP 的深度特征提炼网络，它与后续的概念认知模块共同构成了一个更完整的感知-抽象-认知处理流程，体现了对人类认知过程的深度模拟。

（4）在两个通用 SER 基准数据集 TESS、CREMA-D 上进行了全面的实验，结果表明该框架在处理跨语言复杂情感数据时具有优越的性能，且在增量学习模式下展现出极高的鲁棒性。

2　预备知识

本节简要介绍模糊形式概念分析与极端随机树的核心理论，为后续提出的模型奠定基础。

2.1　模糊形式概念分析

模糊形式概念分析作为一种数学工具，可同时用于数据分析与知识表示[23]。模糊形式背景侧重将模糊集中隶属度的观点引入对象集和属性集之间的关系描述中[29]，解决经典形式概念分析仅适用于离散型数据这一局限性，这对于后续的讨论至关重要。

令 B 表示论域，$\widetilde{B}$ 是定义在 B 上的一个模糊集，其特征由隶属度函数 $\widetilde{B}(v):B\to[0,1]$ 描述。对于任意 $v\in B$，值 $\widetilde{B}(x)$ 是元素 v 在模糊集 $\widetilde{B}$ 中的隶属度。一个模糊形式背景由一个三元组 $(U,V,\widetilde{E})$ 表示，其中 U 表示对象集，V 表示属性集，$\widetilde{E}:U\times V\to[0,1]$ 表示模糊二元关系，$\widetilde{E}(u_i,v_j)$ 表示对象 u_i 在属性 v_j 上的隶属程度。我们用 $\mathcal{P}(U)$ 和 $\mathcal{F}(V)$ 分别表示对象集 U 和属性集 V 的幂集。对于任意对象子集 $A\subseteq U$ 和模糊属性集 $\widetilde{B}\in\mathcal{F}(V)$，定义如下两个算子 $\mathcal{M}:\mathcal{P}(U)\to\mathcal{F}(V)$ 和 $\mathcal{N}:\mathcal{F}(V)\to\mathcal{P}(U)$：

$$\mathcal{M}(A)(v_j)=\bigcap_{u_i\in A}\widetilde{E}(u_i,v_j),\quad \forall v_j\in V,$$

$$\mathcal{N}(\widetilde{B})=\{u_i\in U\mid \forall v_j\in V,\widetilde{B}(v_j)\leqslant\widetilde{I}(u_i,v_j)\}。$$

算子 $\mathcal{M}(A)$ 提取了对象集 A 中所有对象共同拥有的属性特征，取隶属度的最小值；而算子 $\mathcal{N}(\widetilde{B})$ 则表示所有满足模糊属性集 $\widetilde{B}$ 描述的对象集合。

如果一个序对 $(A,\widetilde{B})$ 满足 $\mathcal{M}(A)=\widetilde{B}$ 和 $\mathcal{N}(\widetilde{B})=A$，则它构成一个模糊概念。其中，$A$ 被称为概念的外延，表示该概念覆盖的对象实例，即这个概念包含哪些对象。$\widetilde{B}$ 被称为概念的内涵，用来描述该概念的属性特征，表示这些对象共同具有哪些属性。例如，在 SER 任务中，若 $A=\{u_1,u_2,u_3\}$ 表示三条被标注为“愤怒”的语音样本，$\widetilde{B}$ 刻画了“高音调、高能量、语速较快”等属性在各维度上的隶属度，则 $(A,\widetilde{B})$ 就可以理解为“愤怒语音”这一模糊概念，外延 A 对应具体的语音实例，内涵 $\widetilde{B}$ 则抽象出这些实例在声学属性上的共性模式。

令 $(U,V,\widetilde{E})$ 和 (U,D,T) 表示两个不同的形式背景，其中 $\widetilde{E}:U\times V\to[0,1]$ 表示前一个背景中的模糊关系；$T:U\times D\to\{0,1\}$ 表示后一个背景中的经典关系。当 V 和 D 的交集为空时，结构体 $(U,V,\widetilde{E},D,T)$ 被称为模糊形式决策背景，V 是条件属性集，而 D 对应于决策属性集。给定两个加权模糊概念 $(A_1,\widetilde{B}_1)$ 和 $(A_2,\widetilde{B}_2)$，其层级序关系定义为

$$(A_1,\widetilde{B}_1)\leqslant(A_2,\widetilde{B}_2)\Leftrightarrow A_1\subseteq A_2\Leftrightarrow\widetilde{B}_2\subseteq\widetilde{B}_1 \tag{1}$$

该层级序关系意味着模糊概念间的包含关系取决于其外延的子集关系、内涵的隶属度关系。对于两个概念 $(A_1,\widetilde{B}_1)$ 和 $(A_2,\widetilde{B}_2)$，如果存在 $A_1 \subseteq A_2$（或 $\widetilde{B}_2 \subseteq \widetilde{B}_1$），我们称 $(A_1,\widetilde{B}_1)$ 是 $(A_2,\widetilde{B}_2)$ 的子概念或 $(A_2,\widetilde{B}_2)$ 是 $(A_1,\widetilde{B}_1)$ 的父概念。

例 1. 在一个如表 1 所示的模糊形式决策背景中，对象集 $U=\{u_1,u_2,u_3,u_4,u_5\}$ 代表五段不同的语音样本。条件属性集 $V=\{v_1,v_2,v_3\}$ 是用于情感分类的声学特征，具体来说，v_1 代表语速，v_2 代表音高，v_3 代表能量。决策属性 d 定义了情感类别，其中 $d=0$ 对应于悲伤，$d=1$ 对应于开心。模糊关系 $\widetilde{E}$ 如表 1 所示，$\widetilde{E}(u_1,v_1)=0.2$ 表示样本 u_1 在语速这一特征上的隶属度为 0.2。决策属性 d 将对象集 U 划分为两个类别：悲伤类 $\mathcal{D}_0=\{u_1,u_3\}$ 和开心类 $\mathcal{D}_1=\{u_2,u_4,u_5\}$。通过计算生成的模糊概念总结在表 2 中，为简洁起见，概念 $\left(\{u_2,u_4\},\left(\frac{0.8}{v_1},\frac{0.6}{v_2},\frac{0.7}{v_3}\right)\right)$ 被缩写为(24,(0.8,0.6,0.7))。

表 1 模糊形式决策背景

对象	v_1	v_2	v_3	d
u_1	0.2	0.3	0.4	0
u_2	0.8	0.6	0.7	1
u_3	0.4	0.3	0.5	0
u_4	0.9	0.8	0.8	1
u_5	0.7	0.6	0.6	1

2.2 极端随机树

极端随机树(ERT)是由 Geurts 等人[30]提出的一种集成学习方法，它通过引入额外的随机性层来扩展决策树的概念。虽然在概念上与随机森林相似，但 ERT 的区别在于其在树构建过程中，特别是在节点分裂时，会放大随机性程度。这种更强的随机性有助于降低模型的方差，从而提高其对未见数据的泛化能力。同时，由于分裂点是随机选择的，与传统决策树相比，其计算开销也大大减少。

特征重要性分析是衡量单个特征对模型预测性能贡献的关键工具。通过分析特征的重要性，可以辨别哪些特征对模型的预测影响最大，这不仅有助于选择最相关的特征，还能简化模型并增强其可解释性。在随机森林和极端随机树的背景下，特征重要性通常通过每个特征实现的基尼指数累计减少量和信息增益量来评估。基尼指数本身是衡量数据集“不纯度”的指标，其值越高，表示数据的分布越混乱、结构化程度越低[27]。节点的基尼指数源于该节点内对象类别标签的分布，计算基尼指数的公式如下所示：

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

此处，r 表示类别的总数，p_m 表示一个对象属于类别 m 的概率。在 RF 和 ERT 中，每棵独立的树都基于不同的特征进行数据分裂，每次分裂都有助于降低整体的基尼指数，从而提高模型区分不同类别的能力。一个特征降低基尼指数的程度越大，它对模型整体性能的贡献就越大，其重要性也越高。下式概述了在此背景下计算特征重要性的过程：

$$VI(v_i)=\frac{1}{T}\sum_{t=1}^{T}\sum_{n\in N_t,v_n=v_i}P(n)\cdot\Delta G(v_i,n) \tag{3}$$

其中，$VI(v_i)$ 表示特征 v_i 的重要性得分，T 表示森林中树的总数。每棵树 t 包含一组节点 N_t，其中节点 n 使用特征 v_n 进行分裂。$P(n)$ 代表节点 n 的权重，即该节点中的样本数占总样本数的比例，项 $\Delta G(v_i,n)$ 捕捉了由节点 n 处分裂引起的不纯度变化，其定义为

$$\Delta G(v_i,n)=G(n_{\text{parent}})-(G(n_{\text{left}})+G(n_{\text{right}})) \tag{4}$$

其中，$G(n_{\text{parent}})$，$G(n_{\text{left}})$ 和 $G(n_{\text{right}})$ 分别表示父节点、左子节点和右子节点的基尼指数值。为计算 $VI(v_i)$，需要找出所有 $v_n=v_i$ 的节点，并在所有树中对它们各自的 $\Delta G(v_i,n)$ 值进行求和。然后，将特征 v_i 的总贡献在所有树上取平均，得到最终的重要性得分。最终得到的重要性得分 $VI(v_i)$ 是一个非负值，值越高表明该特征对降低不纯度的贡献越大。模型中所有特征的重要性得分都经过归一化处理，使其总和为 1。该框架根据特征影响模型决策的能力对其进行有效排序，为特征选择和模型评估提供了宝贵的可解释性工具。

3 ERT-WFCCL

在本节中，我们提出一种名为 ERT-WFCCL 的认知学习方法。传统的模糊形式概念分析在构建概念格时，通常默认所有属性具有相同的权重。然而在 SER 任务中，不同声学特征对情感的判别能力存在显著差异，若不加区分地同等对待，噪声特征可能会掩盖关键情感线索。为此，本研究引入 ERT 来判别不同声学特征的重要程度。ERT 不仅能通过基尼指数减少量客观量化特征的重要性，解决 CCL

的特征等权缺陷，还能通过其随机分裂机制降低计算过拟合风险，二者的结合实现了判别性赋权与显式概念推理的优势互补。

3.1　基于 ERT 的概念加权

在一个模糊形式决策背景 $(U,V,\widetilde{E},D,T)$ 中，$U=\{u_1,u_2,\cdots,u_n\}$ 表示对象集，$V=\{v_1,v_2,\cdots,v_m\}$ 表示属性集。决策属性集为 $D=\{d_1,d_2,\cdots,d_r\}$，而 U 关于 D 的决策划分记为 $U/D=\{\mathcal{D}_1,\mathcal{D}_2,\cdots,\mathcal{D}_r\}$，它基于决策属性将 U 划分为互不相交的决策类。

定义 1. 令 $(U,V,\widetilde{E},D,T)$ 为一个模糊形式决策背景。每个属性 $v\in V$ 的特征重要性 $VI(v)$ 通过 ERT 模型计算得出。属性 v 被赋予一个权重，记为 $\boldsymbol{\omega}(v)$，其表达式为

$$\boldsymbol{\omega}(v)=VI(v) \tag{5}$$

其中，$\boldsymbol{\omega}(v)$ 满足 $\sum_{v\in V}\boldsymbol{\omega}(v)=1$。对于一个模糊概念对 $(A,\widetilde{B})$，其中 A 表示外延，$\widetilde{B}$ 表示内涵，我们定义加权模糊概念 $(A,\widetilde{B},\boldsymbol{\omega})$ 的权重表示如下：

$$\boldsymbol{\omega}=\sum_{v_i\in V}\widetilde{B}(v_i)\boldsymbol{\omega}(v_i) \tag{6}$$

权重 $\boldsymbol{\omega}$ 代表了外延 A 的平均信息重要性，反映了多属性内涵在模糊概念中的显著性。下述命题展示了利用 ERT 权重构造加权模糊概念的可行性。

命题 1. 给定一个模糊形式决策背景 $(U,V,\widetilde{E},D,T)$，其中 $\boldsymbol{\omega}$ 是通过 ERT 计算得到的与 V 中属性相关联的权重向量。对于任意 $u\subseteq U$，三元组 $(\mathcal{N}(\mathcal{M}(u)),\mathcal{M}(u),\boldsymbol{\omega})$ 是一个加权模糊粒概念。

例 2.（接例 1）首先，我们假设通过 ERT 模型分析了表 1 中的属性，并计算出它们对应的权重如下：语速 v_1 的重要性最高，音高 v_2 最低。具体权重值为 $\boldsymbol{\omega}(v_1)=0.45$，$\boldsymbol{\omega}(v_2)=0.2$，$\boldsymbol{\omega}(v_3)=0.35$。随后，我们利用公式 5 和上一节已生成的模糊概念，计算得到一系列加权模糊概念，如表 2 所示。

表 2　加权模糊概念（基于表 1）

加权模糊粒概念	概念简称
(13,(0.2,0.3,0.4), 0.29)	$\mathcal{WC}_1$
(24,(0.8,0.6,0.7), 0.725)	$\mathcal{WC}_2$
(3,(0.4,0.3,0.5), 0.415)	$\mathcal{WC}_3$
(4,(0.9,0.8,0.8), 0.845)	$\mathcal{WC}_4$
(245,(0.7,0.6,0.6), 0.645)	$\mathcal{WC}_5$

在该表中，每个加权模糊概念节点的内涵通过一组权重赋值来量化其重要性和相对价值。在认知过程中，人们通常会根据偏好和需求优先关注某些节点并选择性地提取相关信息。这里概念的权重是相对的，权重的大小针对同一概念子空间是有效的，对于不同子空间下的不同概念比较权重是没有意义的。这种方法展示了加权模糊概念的自适应性，为解决以人为中心的问题和增强认知学习提供了一种更高效的途径。通过快速筛选并汇集相关知识，该策略显著降低了不必要的存储开销。通过前述讨论，我们得到了所需的加权模糊概念。在例 1 中，随着对象维数的增加，穷举计算所有模糊概念的开销会呈指数级增长。信息粒作为粒计算理论的核心，在人类认知中扮演着至关重要的角色。为了在认知学习中降低计算复杂度，将信息粒融入学习过程至关重要。最终，如何利用这些加权模糊粒概念来构建一个完备的概念空间，仍然是认知学习领域重点关注的方向。

3.2　加权模糊伪概念学习

本小节我们提出一种基于加权模糊概念和粒计算的新型分类算法，旨在从加权模糊粒概念中构建和更新概念空间。该框架包含两个主要部分：(1) 建立一个初始概念空间以构成分类的基础；(2) 对概念空间进行持续更新，以适应数据变化并满足分类需求。

定义 2. 给定一个模糊形式决策背景 $(U,V,\widetilde{E},D,T)$，其中 $U/D=\{\mathcal{D}_1,\mathcal{D}_2,\cdots,\mathcal{D}_t\}$ 是对象集 U 基于决策属性 D 的划分，$\boldsymbol{\omega}$ 是赋予属性集 V 的权重向量。对于每一个决策类子集 $\mathcal{D}_i$，其关联的加权模糊粒概念子空间 $\mathcal{C}_i$ 定义为：

$$\mathcal{C}_i=\{(\mathcal{N}(\mathcal{M}(u)),\mathcal{M}(u),\boldsymbol{\omega})\mid \forall u\subseteq\mathcal{D}_i\},$$

所有子空间构成了总的加权模糊概念空间 $\mathcal{G}=\{\mathcal{C}_1,\mathcal{C}_2,\cdots,\mathcal{C}_t\}$，每个对象都可以被全面地学习，这有助于提高分类的准确性。算法 1 概述了构建该空间的过程。

算法 1. 构建加权模糊粒概念空间

输入：一个模糊形式决策背景 $(U,V,\widetilde{E},D,T)$ 及属性权重 $\boldsymbol{\omega}$.

输出：加权模糊概念空间 $\mathcal{G}_0$.

1. 初始化 $\mathcal{G}_0=\emptyset$；
2. FOR $\mathcal{D}_i\in U/D$
3.　初始化当前决策类的子空间 $\mathcal{C}_i=\emptyset$；
4.　根据定义 2 构建加权模糊粒概念子空间 $\mathcal{C}_i$；
5.　$\mathcal{G}\leftarrow\mathcal{C}_i$；
6. END FOR
7. RETURN $\mathcal{G}_0$

例 3.（接例 2）根据决策属性 d，对象被分为两个决策类 $\mathcal{D}_0=\{u_1,u_3\}$ 和 $\mathcal{D}_1=\{u_2,u_4,u_5\}$，基于此

划分构建相应的概念子空间。与决策类 $\mathcal{D}_0$ 相关的加权模糊概念子空间 $\mathcal{C}_0$ 包含所有外延是 $\mathcal{D}_0$ 子集的概念。例如，之前计算出的概念(13,(0.2,0.3,0.4),0.29)就属于这个子空间。根据算法 1 可以构建出 $\mathcal{C}_0=\{(13,(0.2,0.3,0.4),0.29),(3,(0.4,0.3,0.5),0.415)\}$。

同理，与决策类 $\mathcal{D}_1$ 对应的加权模糊概念子空间 $\mathcal{C}_1=\{(4,(0.9,0.8,0.8),0.845),(24,(0.8,0.6,0.7),0.725),(245,(0.7,0.6,0.6),0.645)\}$。

两个加权模糊粒概念之间的关系不仅受决策类影响，还可能被数据噪声干扰。噪声会削弱概念间的关联强度，因此必须过滤掉受其影响的加权模糊概念。为此，我们定义了如下的相似度度量。

定义 3. 给定一个加权模糊子空间 $\mathcal{C}_i$ 和一个阈值 ε。对于 $(A_p,\widetilde{B}_p,\boldsymbol{\omega}_p),(A_s,\widetilde{B}_s,\boldsymbol{\omega}_s)\in\mathcal{C}_i$，如果存在 $A_s\subseteq A_p$，则两个概念间的相似度定义为

$$\eta_{p,s}^{c_i}=\frac{|A_s|}{|A_s|+2\lambda(|A_p|-|A_s|)} \tag{7}$$

其中，$\lambda=|\boldsymbol{\omega}_p-\boldsymbol{\omega}_s|$，$|\cdot|$ 表示基数。$\eta_{p,s}^{c_i}$ 的值越大，表示相似度越高。当 $\eta_{p,s}^{c_i}>\varepsilon$ 时，我们认为这两个概念的关联是显著的；反之，则认为关联较弱，子概念 $(A_s,\widetilde{B}_s,\boldsymbol{\omega}_s)$ 可能由噪声引发，应当被剔除。本研究的阈值参数采取自适应设计，算法 2 描述了更新过程。

算法 2. 更新加权模糊粒概念空间

输入：初始加权模糊粒概念空间 $\mathcal{G}_0$ 和阈值 ε.

输出：更新后的加权模糊粒概念空间 $\mathcal{G}$.

1. 初始化更新空间 $\mathcal{G}=\emptyset$；
2. FOR $\mathcal{C}_i\in\mathcal{G}_0$
3. 　初始化更新子空间 $\mathcal{C}_\varepsilon^i=\mathcal{C}_i$；
4. 　FOR 概念 $\mathcal{WC}_i\in\mathcal{C}_i$
5. 　　FOR 其子概念 $\mathcal{WC}_j\in\mathcal{C}_i$
6. 　　根据定义 3 计算相似度 $\eta_{p,s}^{c_i}$；
7. 　　IF $\eta_{p,s}^{c_i}\leqslant\varepsilon$
8. 　　从 $\mathcal{C}_\varepsilon^i$ 移除子概念 $\mathcal{WC}_j$；
9. 　　END IF
10. 　　END FOR
11. 　END FOR
12. 　$\mathcal{G}\leftarrow\mathcal{C}_\varepsilon^i$；
13. END FOR
14. RETURN $\mathcal{G}$

例 4. (续例 3)假设定阈值 $\varepsilon=0.5$ 计算 $\mathcal{C}_1$ 中概念的相似度。令父概念为 $C_p=(245,(0.7,0.6,0.7),0.645)$，子概念为 $C_s=(24,(0.8,0.6,0.7),0.725)$。$|A_p|=3$，$|A_s|=2$，$\lambda=|0.645-0.725|=0.08$，$\eta_{p,s}^{c_1}=\frac{2}{2+2\times0.08\times1}=\frac{2}{2.16}\approx0.9259>0.5$。由于相似度高于阈值，子概念 C_s 被保留。经过对子空间所有概念对的计算，最终更新后的概念子空间为 $\mathcal{C}_1^\varepsilon=\{(4,(0.9,0.8,0.8),0.845),(24,(0.8,0.6,0.7),0.725),(245,(0.7,0.6,0.7),0.645)\}$。

定义 4. 令 $\mathcal{C}_\varepsilon^i$ 为一个更新后的加权模糊子空间。如果存在概念外延满足层级包含关系 $A_1\subseteq A_2\subseteq\cdots\subseteq A_t$ 的概念簇，记该簇为 $\{(A_h,\widetilde{B}_h,\boldsymbol{\omega}_h)\mid h=1,\cdots,t\}$。其中外延最大的概念 $(A_t,\widetilde{B}_t,\boldsymbol{\omega}_t)$ 定义为该概念簇的上确界概念。我们通过构建一个加权模糊伪概念 $p_{i,k}^\varepsilon=(A_{i,k}^\varepsilon,\widetilde{B}_{i,k}^\varepsilon,\boldsymbol{\omega}_{i,k}^\varepsilon)$ 来表示该簇，其计算公式如下：

$$A_{i,k}^\varepsilon=A_t \tag{8}$$

$$\widetilde{B}_{i,k}^\varepsilon=\frac{1}{2^{t-1}}(\widetilde{B}_1+\sum_{h=2}^{t}2^{h-2}\widetilde{B}_h) \tag{9}$$

$$\boldsymbol{\omega}_{i,k}^\varepsilon=\frac{1}{|B|}\sum_{v\in B}\widetilde{B}_{i,k}^\varepsilon(v)\boldsymbol{\omega}(v) \tag{10}$$

其中，$A_{i,k}^\varepsilon$ 表示第 i 个决策类中第 k 个概念簇生成的外延，$\widetilde{B}_{i,k}^\varepsilon$ 是融合了历史层级信息的内涵，通过指数加权的方式赋予了高层级概念更大的权重。在伪概念生成过程中，同一概念不会被学习两次，我们首先从外延最大的概念出发寻找一条连接最细粒度子概念的路径，该路径上的概念构成第一个概念簇，之后剔除该簇中的概念，重复该过程直到所有概念都被学习。

经过上述过程构成了最终的伪概念空间 $\mathcal{P}^\varepsilon=\{P_1^\varepsilon,P_2^\varepsilon,\cdots,P_r^\varepsilon\}$，其中每个 P_i^ε 表示第 i 个决策下所有伪概念构成的子概念空间。算法 3 详述了此过程。

算法 3. 构建加权模糊伪概念空间

输入：更新后的加权模糊概念空间 $\mathcal{G}$.

输出：渐进式加权模糊概念空间 $\mathcal{P}^\varepsilon$.

1. 初始化渐进式空间 $\mathcal{P}^\varepsilon=\emptyset$；
2. FOR $\mathcal{C}_\varepsilon^i\in\mathcal{G}$
3. 　初始化 $P_i^\varepsilon=\emptyset$；
4. 　在 $\mathcal{C}_\varepsilon^i$ 中根据外延的包含关系寻找概念簇；
5. 　根据定义 3 中的公式计算所有渐进加权模糊伪概念 $\{p_{i,1}^\varepsilon,p_{i,2}^\varepsilon,\cdots,p_{i,k}^\varepsilon\}$
6. 　将所有伪概念添加到对应的子空间 P_i^ε；
7. END FOR
8. $\mathcal{P}^\varepsilon\leftarrow P_i^\varepsilon$；
9. RETURN $\mathcal{P}^\varepsilon$

例 5.（续例 4）根据定义 4，我们对更新后的子空间 $\mathcal{C}_1$ 中的概念构造伪概念，由于子空间中的三个概念构成一个外延呈包含关系的概念簇，由 $\mathcal{WC}_2=(A_2,\widetilde{B}_2,\boldsymbol{\omega}_2)$，$\mathcal{WC}_4=(A_4,\widetilde{B}_4,\boldsymbol{\omega}_4)$，$\mathcal{WC}_5=(A_5,\widetilde{B}_5,\boldsymbol{\omega}_5)$ 构建的伪概念外延为 $A^{\varepsilon}_{1,1}=\{u_2,u_4,u_5\}$，内涵为 $\widetilde{B}^{\varepsilon}_{1,1}=0.25\times(\widetilde{B}_4+\widetilde{B}_2+2\widetilde{B}_5)=0.25\times((0.9,0.8,0.8)+(0.8,0.6,0.7)+2\times(0.7,0.6,0.6))=(0.775,0.65,0.675)$。每个决策类都可由一个或多个伪概念来表示，这些概念不仅保留了关键信息，还剔除了冗余，优化了认知学习的效率。

3.3　增量学习机制

对于一个已经构建好的渐进式概念空间，模型的核心挑战在于如何处理新加入的样本 Δu。一方面要准确判定 Δu 的类别标签，另一方面要高效地将 Δu 中蕴含的新知识融入现有空间，即实现增量学习。

3.3.1　新样本的类别预测

我们采用基于相似度的最小距离原则来预测新样本的类别，核心思想是一个新样本与某个情感类别的距离越近，它就越有可能属于该类别。通过使用定义 5 中提出的加权欧氏距离来度量这种相似性，算法 4 的流程确定最终的预测标签。这个过程在例 6 中展示，该方法能够为一个新的“开心”语音样本准确地判定其类别。

定义 5. 给定一个模糊形式决策背景 $(U,V,\widetilde{E},D,T)$。对于一个新加入的对象 Δu，我们将 Δu 及其属性构成的序对视为单粒度概念 $(\{\Delta u\},\widetilde{B})$，它到 P^{ε}_i 中第 k 个伪概念 $p^{\varepsilon}_{i,k}=(A^{\varepsilon}_{i,k},\widetilde{B}^{\varepsilon}_{i,k},\boldsymbol{\omega}^{\varepsilon}_{i,k})$ 的加权欧氏距离定义为

$$d(\Delta u,p^{\varepsilon}_{i,k})=\sqrt{\sum_{v\in B}(\boldsymbol{\omega}(v)(\widetilde{B}(v)-\widetilde{B}^{\varepsilon}_{i,k}(v)))^2} \tag{11}$$

距离 $d(\Delta u,p^{\varepsilon}_{i,k})$ 的值衡量了 Δu 与伪概念的相似程度，值越小，相似度越高，Δu 的分类遵循最小距离原则。算法 4 阐述了为新加入对象确定类别标签的方法。

算法 4. 新样本的类别标签预测

输入：伪概念空间 $\mathcal{P}^{\varepsilon}=\{P^{\varepsilon}_1,P^{\varepsilon}_2,\cdots,P^{\varepsilon}_r\}$ 和新样本 Δu

输出：Δu 的预测类别标签 $\mathcal{L}$

1. FOR $p^{\varepsilon}\in\mathcal{P}^{\varepsilon}$
2. 计算 Δu 到所有伪概念 p^{ε} 的距离；
3. END FOR
4. 找出最小距离：$md=\min(d(\Delta u,p^{\varepsilon}_{i,k}))$；
5. 确定全局最小距离对应的类别索引 $i=\mathrm{argmin}(md_i)$；
6. $\mathcal{L}=i$；
7. RETURN 预测标签 $\mathcal{L}$

例 6.（续例 5）引入一个新的语音样本 Δu，假设其声学特征向量为 $\mathcal{M}(\Delta u)=(0.85,0.8,0.75)$，其真实情感标签为“开心”，属于决策类 $\mathcal{D}_1$。现在，我们根据定义 5 计算 Δu 到两个决策类伪概念的距离。$p^{\varepsilon}_{0,1}$ 的内涵为 $\widetilde{B}^{\varepsilon}_{0,1}=(0.3,0.3,0.45)$，$p^{\varepsilon}_{1,1}$ 的内涵为 $\widetilde{B}^{\varepsilon}_{1,1}=(0.775,0.65,0.675)$，计算距离如下：

$$d(\Delta u,p^{\varepsilon}_{0,1})=\sqrt{0.0613+0.01+0.011}\approx 0.2869$$

$$d(\Delta u,p^{\varepsilon}_{1,1})=\sqrt{0.0011.+0.0009+0.0007}\approx 0.0520$$

由于 Δu 到伪概念 $p^{\varepsilon}_{1,1}$ 的距离 0.052 远小于到 $p^{\varepsilon}_{0,1}$ 的距离 0.2869，因此模型将 Δu 预测为决策类 $\mathcal{D}_1$，与真实标签“开心”一致。

3.3.2　概念空间的动态更新与多样本的自适应更新

当新样本 Δu 被预测为类别标签 i 后，它将被加入对应的决策类 $\mathcal{D}_i$ 中，形成 $\mathcal{D}'_i=\mathcal{D}_i\cup\{\Delta u\}$，此时，如果完全重新计算整个概念空间，计算开销将非常巨大。为解决这个问题，我们提出一种高效的动态更新机制，其核心思想仅在于重构与新样本同类的概念子空间，同时微调异类子空间，从而避免全局性的重复计算。算法 5 展现了单个样本的动态更新过程。

算法 5. 渐进式加权模糊概念的动态增量更新

输入：原伪概念子空间 $\mathcal{P}^{\varepsilon}_i$、新样本 Δu

输出：更新后的伪概念子空间 $\hat{\mathcal{P}}^{\varepsilon}_i$

1. 根据算法 4 预测 Δu 的类别标签 i；
2. $\mathcal{D}'_i=\mathcal{D}_i\cup\{\Delta u\}$；
3. FOR $u_j\in\mathcal{D}'_i$
4.　IF $\widetilde{E}(\Delta u,v)\geqslant\widetilde{E}(u_j,v)$ 对于 $\forall v\in V$
5.　　将新对象 Δu 加入 $(\mathcal{NM}(u_j),\mathcal{M}(u_j),\boldsymbol{\omega})$ 的外延；
6.　END IF
7.　根据更新后的粒概念重新寻找概念簇，根据算法 3 更新得到新的伪概念子空间 $\hat{\mathcal{P}}^{\varepsilon}_i$；
8. END FOR
9. RETURN $\hat{\mathcal{P}}^{\varepsilon}_i$

当存在一批新样本需要处理时，我们采用一种序贯学习策略，即逐个处理新样本。每处理一个样本，都执行一次完整的分类与动态更新流程。这样可以确保模型在学习下一个样本时，已经充分吸收了前一个样本的信息，该过程如算法 6 所示。

算法 6. 多样本批量增量更新

输入：原伪概念空间 $\mathcal{P}^{\varepsilon}$、新样本集合 $\Delta U=\{\Delta u_1,\Delta u_2,\cdots,\Delta u_k\}$

输出：最终更新后的伪概念空间 $\hat{\mathcal{P}}^{\epsilon}$ 及新样本的类别标签

1. FOR 新样本 $\Delta u_j \in \Delta U$

2. 使用算法 4 预测 Δu_j 的类别标签；

3. 使用算法 5 将 Δu_j 的信息融入并更新伪概念子空间 $\hat{\mathcal{P}}^{\epsilon}_{*}$；

4. END FOR

5. $\hat{\mathcal{P}}^{\epsilon} \leftarrow \hat{\mathcal{P}}^{\epsilon}_{*}$

6. RETURN 更新后的 $\hat{\mathcal{P}}^{\epsilon}$ 和所有 Δu_j 的标签

3.4 时间复杂度分析

为分析 ERT-WFCCL 方法的计算代价，本文对模型中各关键算法的时间复杂度进行讨论。为统一记号，设对象集 U 的规模为 $|U|=n$，条件属性集 V 的维数为 $|V|=m$，决策类个数为 $|D|=r$。记第 i 个决策类对应的对象子集为 $\mathcal{D}_i$，其规模为 $|\mathcal{D}_i|=n_i$；$\mathcal{C}_i$ 表示第 i 个决策类对应的加权模糊概念子空间，概念数量为 $|\mathcal{C}_i|=c_i$；$\mathcal{P}^{\epsilon}_i$ 表示经相似度剪枝与融合后生成的加权模糊伪概念子空间，其伪概念数量为 $|\mathcal{P}^{\epsilon}_i|=p_i$。为便于分析，本文不区分更新前后概念子空间的记号，统一使用 $\mathcal{C}_i$ 表示各阶段的概念子空间规模。

算法 1 在每个决策类 $\mathcal{D}_i$ 内部，对其对象子集进行枚举，并据此构造加权模糊粒概念。对于单一决策类，其对象子集的数量为 2^{n_i}。对任意对象子集 $u \subseteq \mathcal{D}_i$，算子 $\mathcal{M}(u)$ 的计算需要在 m 个条件属性上取最小隶属度，其最坏情况下的时间复杂度为 $O(mn_i)$。由算子 $\mathcal{M}(u)$ 生成的模糊属性集作为输入，算子 $\mathcal{N}(\cdot)$ 需对全集 U 中的对象进行逐一校验，其时间复杂度为 $O(nm)$。因此，算法 1 的总体时间复杂度可表示为 $O(\sum_{i=1}^{r} 2^{n_i} \cdot m \cdot (n_i+n))$，其中，$n_i$ 表示算子 $\mathcal{M}(\cdot)$ 在决策类内部的计算规模，而 n 对应算子 $\mathcal{N}(\cdot)$ 在全集上的校验规模。在每个概念子空间 $\mathcal{C}_i$ 中，需要对其中的概念对进行两两比较。由于相似度计算仅涉及概念外延的基数及权重差值，其计算复杂度为常数时间。因此，对于单个子空间 $\mathcal{C}_i$，算法 2 的时间复杂度为 $O(c_i^2)$，其总体时间复杂度为 $O(\sum_{i=1}^{r} c_i^2)$。

算法 3 在更新后的概念子空间中，根据概念外延之间的层级包含关系构造概念簇，并进一步生成加权模糊伪概念。概念簇的识别需要进行成对的外延包含关系判断，其时间复杂度为 $O(c_i^2)$。在生成每一个伪概念时，需要对 m 维条件属性进行加权融合，其时间复杂度为 $O(m)$。设第 i 个子空间中生成的伪概念数量为 p_i，则算法 3 的时间复杂度可表示为 $O(\sum_{i=1}^{r}(c_i^2+mp_i))$。算法 4 通过计算新样本与伪概念空间中各伪概念之间的加权欧氏距离来确定其类别标签，单次距离计算涉及 m 个条件属性，其时间复杂度为 $O(m)$。因此，算法 4 的整体时间复杂度为 $O(m\sum_{i=1}^{r} p_i)$。在算法 5 中，新样本先通过算法 4 完成类别预测。该预测阶段需要遍历所有决策类下的伪概念，其时间复杂度为 $O(m\sum_{j=1}^{r} p_j)$。之后仅对预测得到的目标决策类对应的概念子空间进行更新。新样本需与该决策类中的 n_i 个对象进行属性层面的比较，其时间复杂度为 $O(n_i m)$。接着对更新后的局部概念子空间执行概念聚类及伪概念重构，其时间复杂度为 $O(c_i^2+mp_i)$。那么算法 5 的总体时间复杂度为 $O(m\sum_{j=1}^{r} p_j + n_i m + c_i^2 + mp_i)$。其中，$\sum_{j=1}^{r} p_j$ 对应预测阶段的全局伪概念遍历开销，而 p_i 仅对应目标决策类内的局部更新规模。

当需要处理多个新样本时，算法 6 采用序贯策略，依次对每个新样本执行算法 5。设新样本数量为 k，其总时间复杂度为 $O(\sum_{j=1}^{k}(m\sum_{i=1}^{r} p_i + n_{i_j} m + c_{i_j}^2 + mp_{i_j}))$，其中 i_j 表示第 j 个新样本预测得到的决策类。

最后，ERT-WFCCL 算法的整体工作流程如图 1 所示。模型以语音样本构成的模糊形式决策背景为输入，其中每个语音对象由多维声学特征表示，并以模糊隶属度刻画对象与属性之间的关系。模型先基于 ERT 模型对声学特征进行重要性评估，并将所得特征重要性作为属性权重引入模糊形式概念分析过程，从而得到加权的模糊形式背景。在此基础上，模型针对不同情感类别分别构建对应的加权模糊概念子空间，并生成加权模糊粒概念。随后通过引入相似度度量与权重约束，对概念子空间进行更新，过滤由噪声或弱相关特征引入的冗余概念，使概念结构更加紧凑且具有判别性。更新后的概念进一步经过层级融合，形成能够表征情感类别语义的加权模糊伪概念空间。当有新的语音样本加入时，模型首先在已构建的加权伪概念空间中计算该样本与各情感类别的相似性，并按照最小距离原则确定其情感标签，实现情感分类。与此同时，模型通过渐

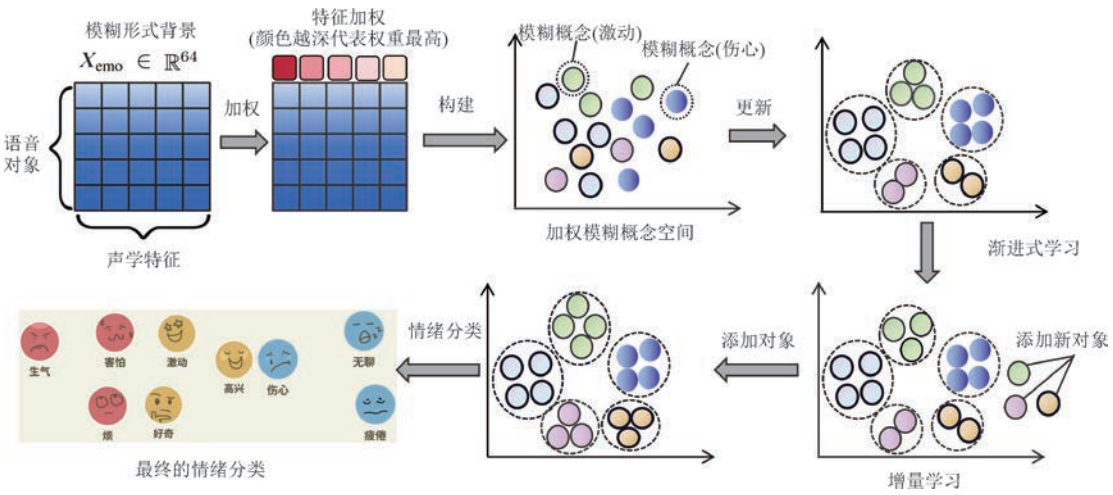

图 1　ERT-WFCCL 算法的工作流程

进式学习机制，将新样本的信息动态地融入对应的概念子空间，仅对相关局部结构进行更新，而无需对整个概念空间进行重构。

4　实验配置

为验证本文所提模型 ERT-WFCCL 的有效性，本节将该模型与多种经典的机器学习算法在两个公开语音情感数据集上进行性能对比。实验的核心评估指标为分类准确率。所有实验均在 macOS Ventura 13.4 操作系统、Apple M2 CPU 及 8GB RAM 的平台下，使用 Python 3.11.7 语言完成。为全面评估模型的性能与跨语言的泛化能力，本研究选用两个基准数据集：TESS、CREMA-D。实验流程统一遵循严格的验证标准，在 CREMA-D 数据集上进行严格评估，遵循严格的说话人独立（Speaker-Independent）原则，以保证结果的客观性和可信度。共进行了 10 次独立的重复实验，并取所有实验结果的平均值作为最终结果。模型的特征提取与模糊化处理流程严格遵循本节和第 2 节所述的方法。

4.1　实验数据集

为保证实验结果的纯粹性和可靠性，避免类别不均衡等外部因素的干扰，本研究选用了两个在学术界被广泛使用的多伦多情感语音集（TESS）以及众包情感多模态演员数据集（CREMA-D）。数据集的关键特征如下：

多伦多情感语音集（TESS）[31]是一个纯音频的英语数据集，包含由 2 位女演员录制的 2800 个音频刺激，涵盖 7 种不同的情感状态，分别为：恐惧（fear）、惊喜（pleasant_surprise）、悲伤（sad）、愤怒（angry）、厌恶（disgust）、开心（happy）和中性（neutral）。该数据集为完全平衡设计，每个情感类别下均包含 400 个样本，这为公平比较各模型内在的分类能力提供了理想的实验平台。我们在 TESS 数据集上使用 80％训练集和 20％测试集来划分，由于该数据集仅包含 2 位说话人，若采用说话人独立划分，训练集将仅含 1 位说话人的数据，样本量不足且易引入说话人偏差，因此本文对 TESS 采用随机划分方式。

众包情感多模态演员数据集（CREMA-D）[32]是一个更具挑战性的多模态情感数据集，本研究中使用其音频部分，其核心特点在于数据的多样性与更接近真实场景的自然度。该数据集由 91 位不同年龄（20～74 岁）、不同种族和民族背景的专业演员录制。演员们通过说出 12 句内容不同的句子来表达 6 种基本情感：愤怒（angry）、厌恶（disgust）、恐惧（fear）、开心（happy）、悲伤（sad）和中性（neutral）。选用 CREMA-D 的主要目的在于检验模型在面对更复杂、更多样化、更接近真实世界声学变化时的泛化能力与稳健性。对于 CREMA-D 数据集，实验采用说话人独立划分，将数据集按说话人 ID 分割为 70％训练集、15％验证集和 15％测试集。为确保结果的统计稳健性，均进行 10 次独立的重复实验，并取平均值作为最终结果。两个数据集的情绪类别样本数量分布由图 2 的饼状图表示。

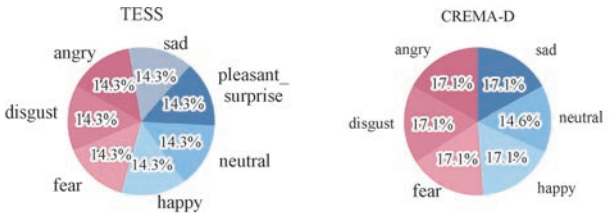

图 2　TESS 和 CREMA-D 语音数据集的类别样本数量分布

4.2 特征提取

SER 的特征处理流程旨在将原始的声学信号转化为一个紧凑且具有高判别力的特征表示[33]。传统的语音情感识别 SER 系统主要依赖于手工提取的声学特征，例如梅尔频率倒谱系数(MFCC)、能量和过零率等。然而，这些低级描述符往往难以捕捉语音信号中深层的语义信息以及长时依赖关系。为了克服这一局限性，本研究采用基于大规模弱监督学习的预训练模型 Whisper[8] 作为特征提取器，并引入深度感知聚焦层 MLP 以提炼情感特征。

4.2.1 基于 Whisper 的语音表征

本研究选用 OpenAI 发布的 Whisper-Base 模型。相比于传统的声学特征，Whisper 特征不仅编码了丰富的声学细节，还隐式地包含了语言学语义信息。这种深层语义对于区分依赖语境的情感至关重要。该模型在约 680000 小时的多语言、多任务监督数据上进行了训练，展现出极强的鲁棒性和泛化能力。特征提取的具体流程如下：先对原始音频信号进行预处理，将其重采样至 16 kHz，并通过补零或截断操作统一调整为 30 秒的固定长度。随后，计算音频的对数梅尔声谱图作为模型输入。该谱图被送入 Whisper 的 Transformer 编码器中。我们提取编码器最后一层的输出作为语音的深层表征，并对时间步维度进行平均池化，最终获得一个全局特征向量 $\boldsymbol{X}_{\text{whisper}} \in \mathbb{R}^{512}$。所有 512 维特征均经过 StandardScaler 标准化处理，确保数据分布的均一性。

4.2.2 深度特征提炼网络 MLP

尽管 Whisper 提供了具有丰富语义信息的 512 维深层声学表征，但该高维特征中仍可能包含与情感判别弱相关的冗余成分，例如说话人身份和背景噪声信息。更重要的是，后续 ERT-WFCCL 模型中模糊概念格的构建复杂度随特征维度呈非线性增长，高维输入将显著增加概念节点数量，从而影响认知推理效率。因此，本文在 Whisper 特征之后引入一个受监督的多层感知机作为深度特征提炼模块，其目标并非直接完成分类任务，而是通过监督学习引导特征空间向情感判别方向聚焦，生成维度紧凑且冗余较低的中间表示，用以支撑后续的概念认知建模。该 MLP 网络架构包含以下层次变换，具体架构设计如图 3 所示：

(1)输入层：接收由 Whisper 提取的特征向量 $\boldsymbol{X}_{\text{whisper}}$，输入维度为 512。

(2)隐藏层 1：将特征维度从 512 映射至 256。该层包含全连接操作、批归一化(BN)、ReLU 激活函数以及 Dropout ($p=0.3$)，以防止模型过拟合。

(3)隐藏层 2：进一步将维度从 256 压缩至 128。结构同样包含 BN、ReLU 和 Dropout ($p=0.2$)。

(4)输出层：最终将特征压缩至 64 维，作为后续模糊概念认知模型的输入。

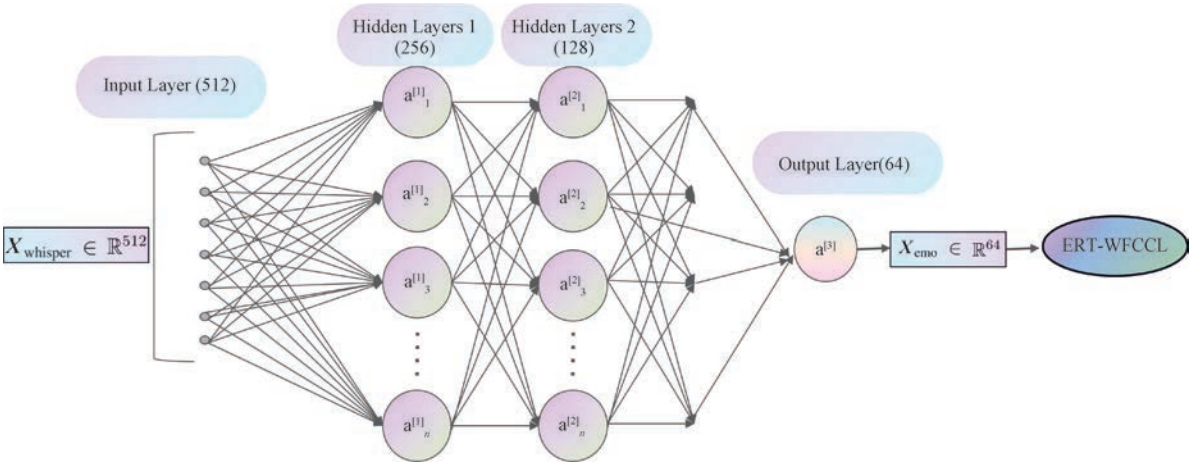

图 3 深度特征提炼网络 MLP 架构

为了在特征提炼阶段显式引入情感判别监督，本文在训练阶段于 64 维特征输出端附加一个线性分类头 Linear(64,6)，构建 MLPWithClassifier 网络结构。整个网络的优化目标是最小化情感分类的误差。损失函数采用标准的交叉熵损失 nn.CrossEntropyLoss，通过计算预测 logit 与真实情感类别标签之间的差异，产生监督信号。该梯度信号反向传播回 MLP，有效指导 64 维特征的学习，使其最大化保留情感相关的判别信息。优化器选用 Adam，初始学习率设置为 0.0005。为增强模型的泛化能力和鲁棒性，我们引入了学习率调度和严格的数据划分策略。学习率调度采用 CosineAn-

nealingWarmRestarts，初始周期 $T=50$ 个 epoch。数据划分严格遵循说话人独立 Speaker-Independent 原则，利用 GroupShuffleSplit 确保训练集和验证集的说话人集合完全不重叠，从而有效避免模型学习特定说话人的声音特征而导致的过拟合。此外，模型采用基于验证集准确率的早停机制，若连续 50 个 epoch 验证准确率未提升，则停止训练并保存最佳 MLP 权重，以确保最终用于 ERT-WFCCL 的 64 维特征是最优且具有泛化能力的。

本文采用 t-SNE 算法将高维特征映射到二维空间进行可视化，来直观评估 MLP 深度特征提炼网络对声学特征的优化效果。图 4 和图 5 展示了 TESS 数据集在特征提炼前后的分布对比。如图 4 所示，原始的 512 维 Whisper 特征虽然能够大体区分不同的情感类别，但各类别簇的分布较为松散，且在决策边界处存在明显的混叠现象，例如“Neutral”与“Sad”之间以及“Fear”与“Disgust”之间。这表明原始的预训练特征中仍包含一定的背景噪声或非情感相关的冗余信息。相比之下，如图 5 所示，经过 MLP 网络提炼后的深度特征展现出了极佳的判别性，类内聚合度显著提升，同一情感类别的样本点聚集得更加紧密，说明模型成功滤除了特征中的噪声干扰。而且类间分离度显著增大，不同情感簇之间的间隔明显拉大，即使是原始空间中容易混淆的类别，在经过非线性映射后也实现了有效的解耦。

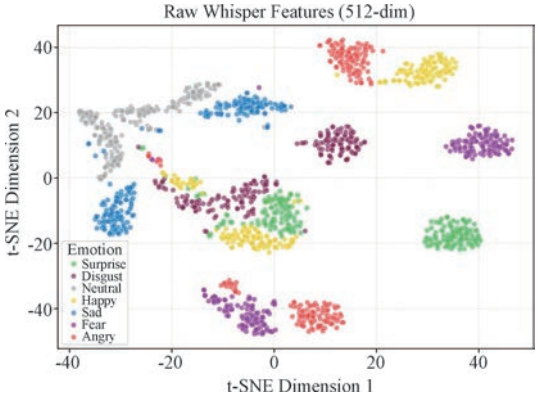

图 4　原始 Whisper 编码器输出特征分布 t-SNE 可视化

并且为了进一步探究经 MLP 提炼后 64 维特征的内在质量，我们计算了特征间的皮尔逊相关系数矩阵，并将其可视化为热力图，如图 6 所示。图中采用层次聚类对特征维度进行了重新排序，以更清晰地呈现特征间的关联结构。通过分析非对角线区域可知，热力图中虽然在如左上角、中间及右下角的局部区域出现了若干高相关性的区块，表明部分特征之间存在一定的聚类关系，但这些区块的规模有

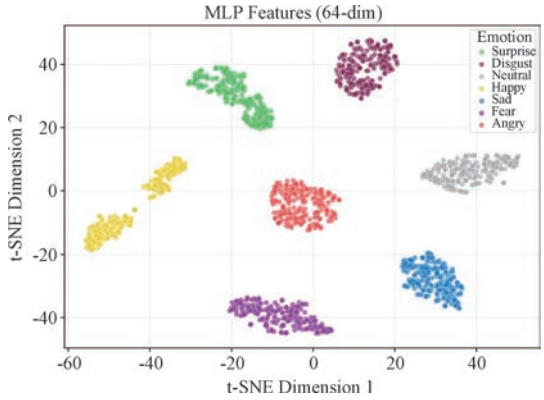

图 5　经 MLP 提炼后的深度特征分布 t-SNE 可视化

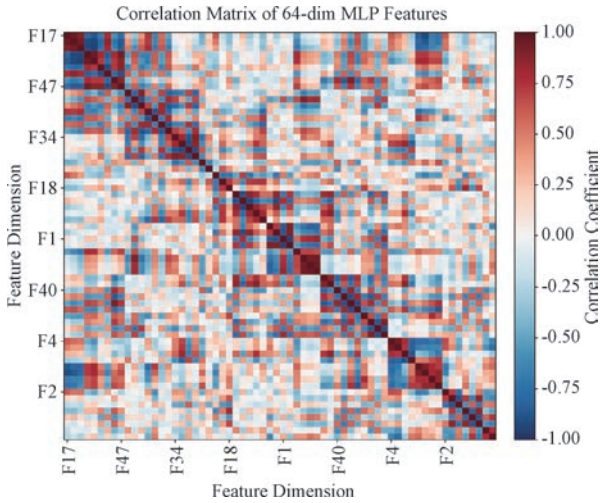

图 6　64 维声学特征的相关性热力图

限且彼此之间以浅色区域分隔，整体上并未形成大范围的强相关带状结构。经计算，特征矩阵的平均绝对相关系数仅为 0.3255。该结果从统计角度表明，经 MLP 非线性映射后，特征间的线性相关性得到有效缓解。高独立性的特征对于后续的 ERT-WFCCL 模块至关重要，它避免了因特征高度相关导致的权重分散问题，确保了极端随机树能够准确评估每个特征维度的独特贡献，从而为构建精简且高效的模糊概念空间提供了高质量的信息输入。

经过 MLP 网络的非线性映射，原始 512 维表征被提炼为 64 维的情感判别性特征向量 $\boldsymbol{X}_{\text{emo}} \in \mathbb{R}^{64}$。随后，该特征将被输入 ERT-WFCCL 认知分类器中进行模糊概念推理。

4.2.3　模型分类与评估

在特征提取之后，我们对所提出的 ERT-WFCCL 模型的分类性能进行评估。由于原始特征不适用于模糊背景实验，我们首先需要对数据集的条件属性进行模糊化处理。其模糊化公式如下：

$$\beta(\mu,v)=\frac{\rho(\mu,v)-\rho_{\min}(v)}{\rho_{\max}(v)-\rho_{\min}(v)} \tag{12}$$

此处，$\beta(\mu,v)$ 是对象 μ 在属性 v 上的模糊化值，而 $\rho_{\max}(v)$ 和 $\rho_{\min}(v)$ 分别是属性 v 的最大值和最小值。模糊化后的数据被送入 ERT-WFCCL 概念认知分类模型中执行分类任务。

5 实验结果与分析

本章旨在通过一系列严格控制的量化实验，对本文所提出的 Whisper＋MLP＋ERT-WFCCL 认知框架进行全面验证。实验的核心目的是评估在不同难度数据集上的泛化能力，客观深入地分析该框架在处理多类别情感数据时的有效性、与主流方法的性能差异，并对框架内部组件的贡献进行剖析。

5.1 评价指标

为科学评估模型性能，本研究采用四种关键评价指标：准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 $F1$ 分数。这些指标均基于混淆矩阵中的四个基本术语计算得出：真正例（TP_i）、真负例（TN_i）、假正例（FP_i）和假负例（FN_i）。

准确率是衡量模型正确预测的样本（包括正确预测的正例和负例）占总样本的比例，它反映了分类器的整体效能。其计算公式为

$$\text{Accuracy}=\frac{\sum_i(TP_i+TN_i)}{\sum_i(TP_i+TN_i+FP_i+FN_i)} \tag{13}$$

然而，在类别不均衡的情况下，准确率可能会产生误导。因此，我们引入了精确率和召回率。精确率衡量的是所有被模型预测为正例的样本中，实际也为正例的比例，反映了模型预测结果的可靠性。计算公式为

$$\text{Precision}=\frac{\sum_i TP_i}{\sum_i(TP_i+FP_i)} \tag{14}$$

召回率则衡量所有实际为正例的样本中，被模型成功预测为正例的比例，这反映了模型发现所有正例的能力。其计算公式为

$$\text{Recall}=\frac{\sum_i TP_i}{\sum_i(TP_i+FN_i)} \tag{15}$$

在本研究中，为避免样本量大的类别主导评估结果，我们采用 Macro F1 分数作为最终的评估指标。这确保了每个情感类别对模型的总评估结果贡献相等，从而更客观地反映模型在所有情感类别上的综合性能和泛化能力。其计算公式为

$$\text{Macro } F1=2\cdot\frac{\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}} \tag{16}$$

5.2 总体性能对比分析

为了客观评估本文所提认知分类器的有效性，我们将本文提出的 ERT-WFCCL 模型以及多种经典机器学习[34]、神经网络模型在内的分类算法，包括 MLP、SVM、NB、LR、RF、K-NN（$K=3$），在 TESS 和 CREMA-D 两个不同的数据集上进行了全面对比。为确保对比的公平性，本研究并未直接引用数据，而是严格复现了以上所有对比算法。所有模型均接收经由 Whisper 编码并通过 MLP 深度提炼后的 64 维特征向量。详细的性能指标如表 3 和表 4 所示（Accuracy，Precision，Recall，Macro $F1$）。

表 3 不同模型在 TESS 数据集上的详细性能指标对比（%）

模型	准确率	精确率	召回率	$F1$ 分数
MLP＋ERT-WFCCL	**99.60**	**99.50**	**99.40**	**99.50**
Whisper(512)＋ERT-WFCCL	92.81	91.93	92.81	92.37
MLP	99.40	99.30	99.20	99.30
SVM[35]	99.50	99.40	99.30	99.40
NB[13]	99.30	99.20	99.10	99.20
LR[36]	99.00	98.90	98.80	98.90
RF[13]	98.70	98.60	98.50	98.60
K-NN[13]	97.70	97.60	97.40	97.50

表 4 不同模型在 CREMA-D 数据集上的详细性能指标对比（%）

模型	准确率	精确率	召回率	$F1$ 分数
MLP＋ERT-WFCCL	**85.71**	**86.01**	**85.74**	**85.81**
Whisper(512)＋ERT-WFCCL	78.39	78.49	78.45	78.47
MLP	84.55	84.95	84.57	84.65
SVM[35]	85.42	85.67	85.47	85.48
NB[13]	84.93	85.22	84.92	85.01
LR[36]	85.54	85.73	85.58	85.60
RF[13]	85.28	85.55	85.29	85.35
K-NN[13]	85.05	85.26	85.09	85.10

由表 3 可知，所有模型在 TESS 数据集上均展现出了卓越的分类性能。一方面 TESS 数据集本身录制于受控实验室环境，信噪比高且情感表达清晰。其次，更关键的是前端特征提取模块的有效性。经过 Whisper 预训练模型编码并经由 MLP 深度提炼后的 64 维特征，保留了极高纯度的情感判别信息。LR 能达到 99.00％的准确率，充分证明了该特征空间具有极佳的线性可分性。在如此高的基准性能之上，本文提出的 ERTWFCCL 模型依然实现了性能

的进一步突破，准确率和综合得分最高。这表明在特征极其纯净的情况下，基于加权模糊概念的推理机制比最大间隔超平面模型更能捕捉到数据中极细微的非线性结构。基于集成学习的 RF 在 TESS 上的表现略逊于基于几何边界的模型 SVM。这可能是因为在极高维且特征高度抽象的低维空间中，连续的几何边界比轴平行的树分裂规则更适合描述数据分布。而 K-NN 的表现相对最弱，为 97.70%，说明在高密度特征空间中，简单的欧氏距离度量不如基于概念结构 ERT-WFCCL 的度量有效。

分析表 4 可知，在 CREMA-D 数据集上，本文提出的 ERT-WFCCL 同样在四个评价指标上均位居第一，其 Macro $F1$ 分数达到 85.81%。其他对比算法都受益于高质量的 Whisper＋MLP 深度特征，但 ERT-WFCCL 依然能够从高度抽象的特征中挖掘出额外的判别信息，超越了 LR 的 85.60%和 SVM 的 85.48%。这证明了在特征空间受限的情况下，基于加权模糊概念的非线性推理比传统的几何划分更具优势。图 7 为不同模型在两个数据集上性能对比的柱状图。

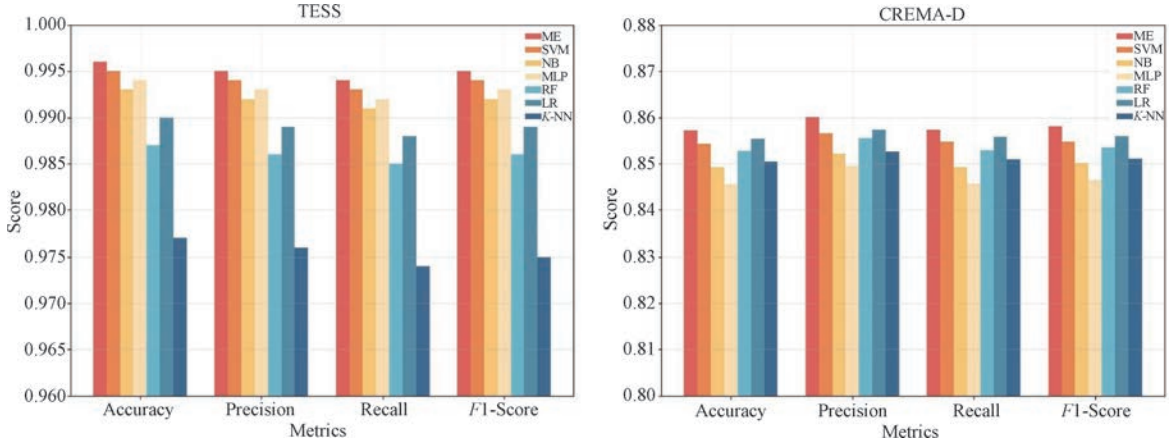

图 7　不同模型分别在 TESS 和 CREMA-D 数据集上性能对比图

5.3　混淆矩阵分析

为进一步深入探究本文所提框架在具体情感类别上的分类行为，并客观评估其在不同难度数据集上的泛化能力，本小节对其在 CREMA-D 上的分类结果进行了混淆矩阵分析。混淆矩阵不仅能够展示模型在每个类别上的分类准确性，更能揭示其内在的决策模式与稳健性。

不同模型在 CREMA-D 数据集上的混淆矩阵如图 8 所示。在所有对比模型中，ERT-WFCCL 取得了最优表现，其准确率和 F1-Macro 都达到了最高。然而所有模型存在负向和中性情感的混淆，高唤醒度的负向情感如“愤怒”、“厌恶”之间存在显著的相互混淆，这是由于这些情感在声学特征例如语速、能量上具有高度的相似性，符合人类情感认知的生理学规律。同样，由于缺乏鲜明的声学标记，Neutral 中性情感的样本也显示出较高的分散性，容易被模型误判为低唤醒度的悲伤或厌恶情感。这表明模型的鲁棒性测试焦点不应仅仅放在整体准确率上，而应放在对这些模糊和重叠情感边界的处理能力上。

ERT-WFCCL 模型的混淆矩阵在处理上述挑战时展现了独特的优势。通过与 LR 和 SVM 的矩阵进行对比，可以发现 ERT-WFCCL 的错误分布更加分散和均匀，而沿主对角线的数值集中度最高。这表明 ERT-WFCCL 能够更有效地抑制系统性的错误偏差。当样本位于特征空间的模糊边界时，ERT-WFCCL 不会像传统几何模型那样进行硬性超平面划分，而是基于其在不同加权概念内涵中的隶属度进行柔性推理。这种机制有效地减少了跨越不同情感效价和唤醒度的错误，确保了在高度相似的负向情感之间，模型依然能做出更精确的判断，最终使其 F1-Macro 达到 85.81%的最优水平，从而有力地验证了认知决策模块的精细化优势。

5.4　消融实验

为了深入探究框架内部各组件的贡献，本节进行了消融实验，旨在回答三个核心问题：选择 Whisper 预训练模型相比于其他主流预训练模型有何优势，深度特征提炼阶段是否必要，以及认知分类器 ERT-WFCCL 相比于传统的神经网络分类层有何优势。

5.4.1　不同预训练语音特征的对比分析

本小节在 TESS 和 CREMA-D 数据集上将 Whisper 预训练模型与目前先进的语音编码器 WavLM 还有专门针对语音情感的 emotion2vec 模型进行了

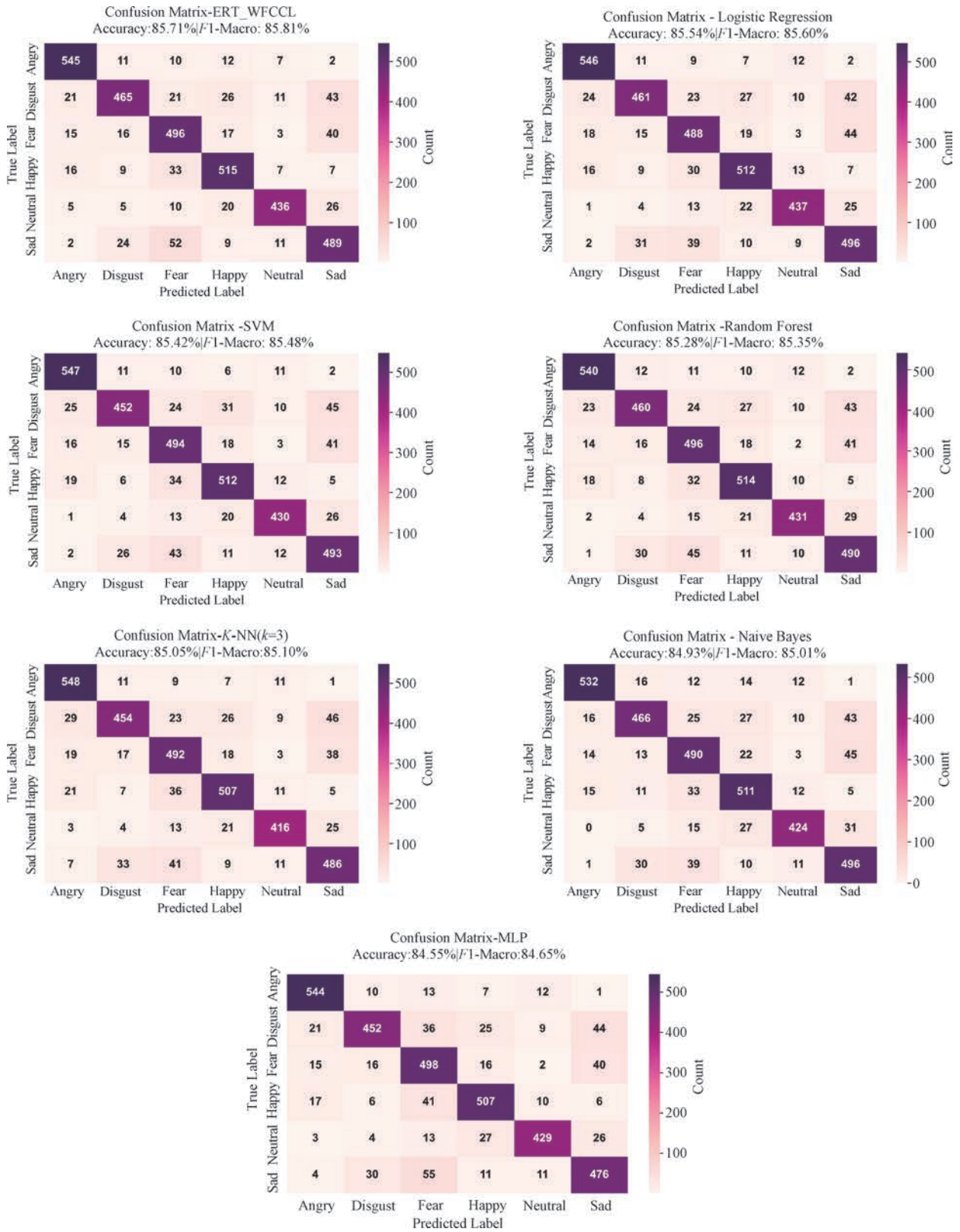

图 8 不同模型在 CREMA-D 数据集上的混淆矩阵

对比。实验均经过相同的 MLP 深度特征提炼降维至 64 维，再接入 ERT-WFCCL 进行分类，实验结果如下表所示，表中名称后括号的数字代表提取特征的维度。

如表 5 所示，在受控环境下录制的 TESS 数据集上，三种预训练模型均展现出了极高的性能，各项指标接近或超过 99.2%。其中，emotion2vec 的 $F1$ 分数为 99.53%，略微领先于 Whisper 的 99.50%。

这表明对于情感特征较为清晰、噪声干扰较小的数据集，主流的预训练模型都能提取出具有高度判别性的声学表征，性能已接近该数据集的理论上限。

表 5　TESS 数据集上不同预训练特征的性能对比（%）

方法	准确率	精确率	召回率	$F1$ 分数
Whisper(512)	**99.60**	99.50	99.40	99.50
emotion2vec(768)	99.52	**99.54**	**99.52**	**99.53**
WavLM(768)	99.28	99.19	99.28	99.23

然而，Whisper 模型在面对发音人多样、情感表达更复杂且更贴近真实场景的 CREMA-D 数据集时，可以体现很大优势。我们观察表 6 可知，Whisper 取得 85.81%的最高 $F1$ 分数，分别比 emotion2vec 和 WavLM 提取特征的 $F1$ 高出 0.66%和 0.85%。这得益于 Whisper 在极大规模、多语种且包含真实噪声的语料库上进行的弱监督训练，使其提取的深度特征具有更强的泛化能力和抗干扰鲁棒性，从而在处理高难度的复杂情感分类任务时表现更优。

表 6　CREMA-D 数据集上不同预训练特征的性能对比（%）

方法	准确率	精确率	召回率	$F1$ 分数
Whisper(512)	**85.71**	**86.01**	**85.74**	**85.81**
emotion2vec(768)	85.08	85.20	85.10	85.15
WavLM(768)	84.88	85.01	84.92	84.96

5.4.2　MLP 深度特征提炼模块的必要性分析

在确定 Whisper 作为前端特征提取器之后，本节进一步验证 MLP 深度特征提炼模块的必要性。该实验将 Whisper 编码器输出的原始 512 维特征直接输入 ERT-WFCCL 进行分类，不经过 MLP 提炼，我们将此方法命名为 Whisper＋ERT-WFCCL。

由表 3 和表 4 可知，移除 MLP 模块后 Whisper＋ERT-WFCC 的性能出现大幅下降，在 TESS 上 $F1$ 分数从 99.50%降至 92.37%，下降了 7.13 个百分点；在 CREMA-D 上从 85.81%降至 78.47%，下降了 7.34 个百分点。该性能差距可从以下几个方面解释：一方面，原始 512 维 Whisper 特征中包含大量与情感判别弱相关的冗余信息，这些冗余维度在模糊概念空间构建中会引入大量无效概念节点，稀释了关键情感概念的判别力。另一方面，模糊概念空间的构建复杂度随特征维度呈非线性增长，512 维输入导致概念空间过于庞大，概念间的欧氏距离度量在高维空间中变得不可靠，产生维度灾难问题。以上结果表明，MLP 模块的作用不仅在于降维，更在于通过有监督的特征重塑，为概念空间提供紧凑且具有强判别力的输入表征。

5.4.3　认知分类器 ERT-WFCCL 的贡献分析

在 Whisper 特征提取和 MLP 特征提炼均已确定的基础上，本节分析认知分类器 ERT-WFCCL 相比传统神经网络分类层的优势。在本实验中，MLP 基准模型代表端到端的神经网络方法，直接使用特征提炼网络末端的线性分类层配合 Softmax 函数进行预测。而 ERT-WFCCL 则是移除该线性层，将 64 维特征向量直接输入认知概念模型中进行推理。

如表 4 所示，ERT-WFCCL 的 Macro $F1$ 显著优于标准 MLP 分类头，提升了 1.16 个百分点。由于两者的输入特征完全相同，这一差距直接反映了决策机制的差异。MLP 分类头本质上是一种概率映射，容易受到训练数据分布偏差的影响，且缺乏对特征组合逻辑的显式建模。相比之下，ERT-WFCCL 通过 ERT 赋予的特征权重构建加权概念格，实现了从数值特征到语义概念的跃迁。这一结果证实，在深度特征提取之后，引入一个基于规则和概念的认知决策模块，比继续堆叠全连接层能获得更鲁棒的分类效果。

5.5　增量学习机制的有效性与鲁棒性分析

为了验证本文提出的自适应更新算法在处理流式数据时的有效性，我们在 CREMA-D 数据集上对比了批处理模式与自适应更新模式的性能表现以及样本输入顺序对模型性能的影响。批处理模式一次性构建概念空间，不对测试集样本进行模型更新。自适应更新模式模拟在线学习场景，逐个输入测试样本，并利用算法 5 动态更新概念空间，实验结果如表 7 所示。

由表 7 可知，本文提出的增量更新算法在各项指标上均略优于批量处理模式，其中 Macro $F1$ 提升了 0.31%。在计算效率方面，逐个处理相比批量处理节省了约 45.82 秒，约 5%的时间开销。这表明 ERT-WFCCL 的动态更新机制能够高效地将新样本知识融合进概念格，且这种细粒度的渐进式修正比粗粒度的批量重构更能捕捉局部特征的微小变化。图 9 展现了两种模式的性能对比。

表 7　批处理与自适应更新模式性能对比

模式	准确率（%）	精确率（%）	召回率（%）	$F1$ 分数（%）	平均耗时（s）
自适应更新	**85.71**	**86.01**	**85.74**	**85.81**	**840.25**
批处理	85.42	85.67	85.48	85.50	886.07

在增量学习中，样本的输入顺序往往会影响最终的概念格结构。为验证模型对输入顺序的敏感

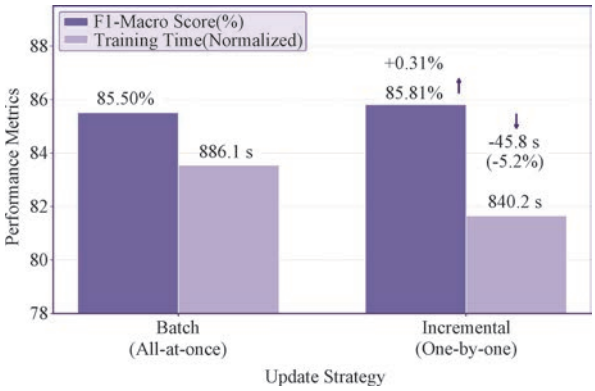

图 9 批处理模式与自适应更新模式的性能表现

度，我们对比了原始顺序与三种随机打乱顺序(Random 1-3)下的模型性能。结果如表 8 和图 10 所示。分析显示，不同处理顺序导致的性能波动极小。四组实验中，Macro $F1$ 的最大差值仅为 0.25%，且所有结果均维持在 85.6%以上。实验证明了本研究提出的算法对样本添加顺序具有良好的鲁棒性，模型最终收敛的性能对输入序列不敏感，这为实际应用中处理流式数据提供了可靠性保障。

表 8 不同样本处理顺序对模型性能的影响

样本顺序	准确率(%)	精确率(%)	召回率(%)	$F1$ 分数(%)	平均耗时(s)
Original	85.71	86.01	85.74	85.81	884.97
Random 1	**85.78**	**86.22**	**85.85**	**85.88**	**823.76**
Random 2	85.74	86.15	85.82	85.86	886.41
Random 3	85.56	85.99	85.63	85.63	886.84

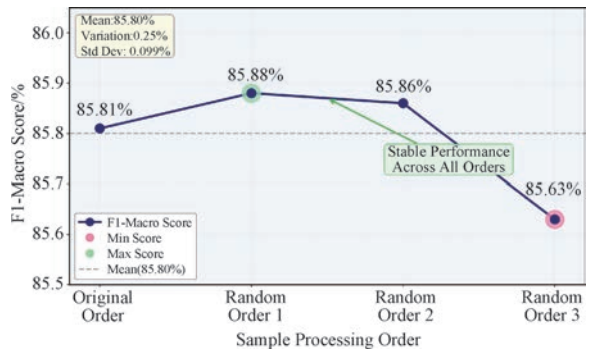

图 10 不同样本处理顺序对模型性能的影响比较

5.6 模型可解释性与认知一致性分析

为验证所提出的 ERT-WFCCL 模型在实际决策过程中是否体现认知一致性与结构可解释性，本节从特征层贡献分布与认知决策机制两个层面对模型的内部行为进行分析，并且最后将本文提出的可解释性方法与现有可解释性方法进行了比较。

5.6.1 特征贡献分布分析

在本文提出的 ERT-WFCCL 模型中，情感分类并非直接基于深度嵌入向量完成，而是通过加权模糊概念空间中的认知推理过程实现。因此，不同特征维度在决策过程中所承担的贡献程度成为衡量模型可解释性的关键因素之一。本文基于 ERT 在训练过程中产生的特征重要性统计结果，对经 MLP 提炼后的 64 维情感特征进行分析，如图 11 所示。

图 11(a)给出了重要性得分排名前 20 的特征维度，可以观察到，排名最高的特征 F59 的重要性得分约为 0.0328，排名第 20 的特征约为 0.017，两者仅相差约一倍，整体呈现平缓递减的分布趋势，并未出现单一特征占据绝对主导的极端情况。这表明 ERT-WFCCL 在决策过程中并非依赖于某个孤立的声学线索，而是综合利用多个互补特征共同完成情感判别。进一步地，图 11(b)展示了全部 64 维特征的重要性分布情况及其平滑趋势线。我们可以看到，特征重要性整体呈现连续且平滑的长尾分布，重要性随特征排名逐步衰减，即使排名最末的特征仍保持约 0.007 的贡献值，而非出现明显断崖或坍缩现象。这一结果说明，经 MLP 非线性映射后得到的特征整体质量较高，大多数特征在决策过程中均发挥了一定作用，ERT 仅对其贡献程度进行细粒度区分，而非执行硬性的特征选择。该分布特性与前文对 MLP 输出特征进行的相关性分析结果保持一致，从决策层面进一步验证了特征提炼阶段对冗余信息的有效抑制。

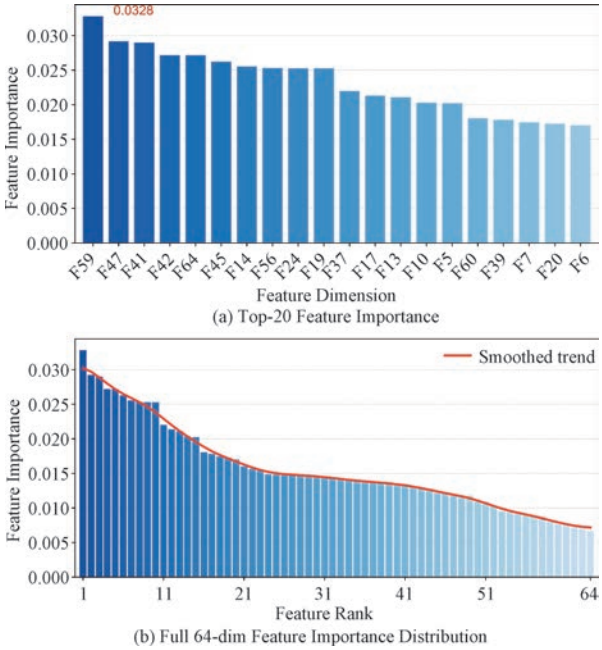

图 11 特征重要性分布情况

5.6.2 认知一致性分析

从认知建模角度来看，情感判断通常并非由单一感知线索触发，而是由多个声学与语义因素的综

合作用逐步形成。ERT-WFCCL 所呈现的特征重要性分布恰好体现了这一认知特征。少数关键维度在情感判别中起到主导作用，其余特征则作为辅助信息共同参与推理过程。这种决策模式与人类在语音情感感知中的判断机制具有良好的认知一致性。此外，由于特征权重直接来源于极端随机树在模糊概念空间中的分裂统计结果，特征贡献的估计过程与具体决策路径具有明确对应关系，使得模型输出能够被追溯至具体的特征维度及其在概念推理过程中的作用位置。这种结构内嵌的解释机制避免了事后分析带来的不稳定性问题，为模型决策提供了更具可信度的解释依据。

5.6.3　与现有可解释性方法的比较讨论

现有深度学习模型在可解释性方面的研究多集中于注意力权重可视化或事后归因方法，如 SHAP 和 LIME 等。注意力机制通常通过可视化不同时间步或特征通道的权重分布来刻画模型关注区域，但其权重大小并不一定与最终决策贡献严格对应，且缺乏明确的概念层语义对齐。SHAP 和 LIME 等事后解释方法则通过局部扰动样本来估计特征贡献，其解释结果对采样策略和扰动范围较为敏感，且解释模块与原始模型结构相互独立，稳定性和一致性难以保证。相比之下，本文提出的 ERT-WFCCL 方法将可解释性直接嵌入模型结构之中。特征重要性的评估过程源于极端随机树在加权模糊概念空间中的分裂统计，与实际分类决策路径耦合，使得解释结果不依赖于额外的事后分析或近似假设。同时，模糊概念认知框架为特征贡献提供了概念层语义中介，使模型输出能够被理解为概念推理的结果，而非仅仅是数值加权的产物。这种结构内嵌式解释方式为语音情感识别任务提供了一种更加稳定且符合认知直觉的解释路径。

6　结论与未来工作

本文面向语音情感识别中深度模型高性能但难以解释的核心矛盾，借鉴 CCL 思想，构建了一个兼顾表征能力与可解释性的两阶段框架，前端采用 Whisper 预训练模型与 MLP 进行深度特征提炼，后端通过基于 ERT 的加权模糊概念认知分类模型 ERT-WFCCL 完成情感判别。该框架以感知到抽象再到认知的认知路径为主线，实现了从连续声学信号到可解释概念空间的映射，为 SER 任务提供了一种具备内生可解释性的建模思路。实验结果表明，所提方法在两个公开情感语音数据集上均取得了优异性能。因此本文方法尤其适用于既需要较高识别性能，又必须给出决策依据的场景，例如智能客服、心理健康辅助评估、教育场景中的学习状态监测等。同时，增量学习实验验证了所提出的加权模糊概念认知更新策略在多批次新样本到来时，能够在保持较低计算开销的前提下稳定更新概念空间，展现出较强的鲁棒性和可拓展能力。这种设计为真实部署中的在线语音情感识别系统提供了实用支撑。一方面避免频繁重训带来的计算开销，另一方面保证模型在用户群体和交互环境持续变化时，仍能保持稳定的性能与一致的认知解释。

尽管如此，本文仍存在若干值得进一步研究的方向。实验数据主要来自受控条件下的公开数据集，未来我们将考虑引入更贴近实际应用场景的 in-the-wild 语音语料，系统评估该框架在复杂噪声环境和多说话人混合场景下的表现。再者，随着特征维度和样本规模的进一步增大，ERT-WFCCL 在大规模概念空间构建与多批次增量更新中的计算复杂度仍有优化空间，可结合更高效的树模型、近似搜索与并行计算策略，构建轻量化的认知分类器。另外，后续研究将视觉、文本等多模态信息融入当前的认知框架，有望构建一个更全面的情感理解模型。

参　考　文　献

[1] Mishra S, Bhatnagar N, P P, et al. Speech emotion recognition and classification using hybrid deep CNN and BiLSTM model. Multimedia Tools and Applications, 2024,83:37603-37620

[2] Liu M, Raj A N J, Rajangam V, et al. Multiscale-multichannel feature extraction and classification through one-dimensional convolutional neural network for speech emotion recognition. Speech Communication, 2024,156:103010

[3] Hashem A, Arif M, Alghamdi M. Speech emotion recognition approaches: A systematic review. Speech Communication, 2023,154:102974

[4] Ahmed M R, Islam S, Islam A K M M, et al. An ensemble 1D-CNN-LSTM-GRU model with data augmentation for speech emotion recognition. Expert Systems with Applications, 2023,218:119633

[5] Zou Yong-Pan, Wang Dan-Yang, Wang Dan, et al. Research on wearable emotion recognition based on multi-source domain adversarial transfer learning. Chinese Journal of Computers, 2024,47(2):266-286(in Chinese)
(邹永攀，王丹阳，王丹，等. 基于多源域对抗迁移学习的可穿戴情绪识别技术. 计算机学报，2024,47(2):266-286)

[6] Chen S, Wang C, Chen Z, et al. WavLM: Large-scale self-su-

pervised pre-training for full stack speech processing. IEEE Journal of Selected Topics in Signal Processing, 2022,16(6): 1505-1518

[7] Ma Z, Zheng Z, Ye J, et al. emotion2vec: Self-supervised pre-training for speech emotion representation//Findings of the Association for Computational Linguistics: ACL 2024. Bangkok, Thailand, 2024:15747-15760

[8] Radford A, Kim J W, Xu T, et al. Robust speech recognition via large-scale weak supervision//Proceedings of the 40th International Conference on Machine Learning. Honolulu, USA. 2023:28492-28518

[9] Schirmer A, Adolphs R. Emotion perception from face, voice, and touch: Comparisons and convergence. Trends in Cognitive Sciences, 2017,21(3):216-228

[10] Middya A I, Nag B, Roy S. Effective MLP and CNN based ensemble learning for speech emotion recognition. Multimedia Tools and Applications, 2024,83:83963-83990

[11] Luengo I, Navas E, Hernáez I, et al. Automatic emotion recognition using prosodic parameters//Proceedings of Interspeech. Lisbon, Portugal, 2005:493-496

[12] Nwe T L, Foo S W, De Silva L C. Speech emotion recognition using hidden Markov models. Speech Communication, 2003,41(4):603-623

[13] Chen H, Wu S, Wu X, et al. Speech emotion recognition using multiple classification models based on MFCC feature values. Applied and Computational Engineering, 2023, 6: 1037-1047(in Chinese)
陈翰文,吴圣平,吴昕睿,等.基于 MFCC 特征值的多分类模型语音情感识别.应用与计算工程,2023,6:1037-1047

[14] Zhao J, Mao X, Chen L. Speech emotion recognition using deep 1D & 2D CNN LSTM networks. Biomedical Signal Processing and Control, 2019,47:312-323

[15] Li D, Liu J, Yang Z, et al. Speech emotion recognition using recurrent neural networks with directional self-attention. Expert Systems with Applications, 2021,173:114683

[16] Guo D, Xu W, Ding W, et al. Concept-cognitive learning survey: Mining and fusing knowledge from data. Information Fusion, 2024,109:102426

[17] Guo D, Xu W. Fuzzy-based concept-cognitive learning: An investigation of novel approach to tumor diagnosis analysis. Information Sciences, 2023,639:118998

[18] Guo D, Xu W, Qian Y, et al. M-FCCL: memory-based concept-cognitive learning for dynamic fuzzy data classification and knowledge fusion. Information Fusion, 2023,100:101962

[19] Yuan K H, Xu W H, Li W T, et al. An incremental learning mechanism for object classification based on progressive fuzzy three-way concept. Information Sciences, 2022,584:127-147

[20] Xu W, Guo D, Mi J, et al. Two-way concept-cognitive learning via concept movement viewpoint. IEEE Transactions on Neural Networks and Learning Systems, 2023,34(10):6798-6812

[21] Guo D, Xu W, Qian Y, et al. Fuzzy-granular concept-cognitive learning via three-way decision: Performance evaluation on dynamic knowledge discovery. IEEE Transactions on Fuzzy Systems, 2023,32(3):1409-1423

[22] Li Jin-Hai, Mi Yun-Long, Liu Wen-Qi. Incremental cognitive of concept: Theories and methods. Chinese Journal of Computers, 2019,42(10):2233-2250(in Chinese)
(李金海,米允龙,刘文奇.概念的渐进式认知理论与方法.计算机学报,2019,42(10):2233-2250)

[23] Mi Y, Shi Y, Li J, et al. Fuzzy-based concept learning method: Exploiting data with fuzzy conceptual clustering. IEEE Transactions on Cybernetics, 2020,52(1):582-593

[24] Li J, He Z, Zhu Q. An entropy-based weighted concept lattice for merging multi-source geo-ontologies. Entropy, 2013, 15(6):2303-2318

[25] Zhang S, Guo P, Zhang J, et al. A completeness analysis of frequent weighted concept lattices and their algebraic properties. Data & Knowledge Engineering, 2012,81:104-117

[26] Singh P K, Cherukuri A K, Li J. Concepts reduction in formal concept analysis with fuzzy setting using Shannon entropy. International Journal of Machine Learning and Cybernetics, 2017,8(1):179-189

[27] Zhang C, Tsang E C C, Xu W, et al. Incremental concept-cognitive learning approach for concept classification oriented to weighted fuzzy concepts. Knowledge-Based Systems, 2023,260:110093

[28] Xu W, Zhang C. RF-WFCCL: a random forest-driven weighted fuzzy concept cognitive learning. Fuzzy Sets and Systems, 2025:109724

[29] Guo Dou-Dou, Xu Wei-Hua. R-FCCL: an approach of fuzzy-based concept-cognitive learning with robustness for high-dimensional data. Journal of Computer Research and Development, 2025,62(2):383-396(in Chinese)
(郭豆豆,徐伟华.R-FCCL:一种面向高维数据的稳健模糊概念认知学习方法.计算机研究与发展,2025,62(2):383-396)

[30] Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Machine Learning, 2006,63(1):3-42

[31] Dupuis K, Pichora-Fuller M K. Toronto emotional speech set (TESS). Toronto, Canada: University of Toronto, Psychology Department, 2010

[32] Cao H, Cooper D G, Keutmann M K, et al. CREMA-D: Crowd-sourced emotional multimodal actors dataset. IEEE Transactions on Affective Computing, 2014,5(4):377-390

[33] Liu Zhen-Tao, Xu Jian-Ping, Wu Min, et al. Review of emotional feature extraction and dimension reduction for speech emotion recognition. Chinese Journal of Computers, 2018,41(12):2833-2851(in Chinese)
(刘振焘,徐建平,吴敏,等.语音情感特征提取及其降维方法综述.计算机学报,2018,41(12):2833-2851)

[34] Agarwal B, Mittal N. Text classification using machine learning methods: A survey//Proceedings of the Second International Conference on Soft Computing for Problem Solving (SocProS 2012). Jaipur, India, 2012:701-709

[35] Zhou Zhi-Hua. Machine Learning. Beijing: Tsinghua University Press, 2016(in Chinese)
(周志华.机器学习.北京:清华大学出版社,2016)

[36] Zhu-Zhou F, Gil-Pita R, García-Gómez J, et al. Robust multi-scenario speech-based emotion recognition system. Sensors, 2022,22(6):2343

HU Kai-Ping, M. S. candidate. Her research interests include Concept-cognitive Learning, granular computing, and speech emotion recognition.

XU Wei-Hua, Ph. D., professor. His research interests include granular computing, cognitive computing, data mining, and machine learning.

QIAN Yu-Hua, Ph. D., professor, Ph. D. supervisor. His research interests include artificial intelligence, big data, and machine learning.

Background

With the continuous development of human-computer interaction (HCI) technology, computers are expected not only to understand the linguistic content of speech but also to accurately perceive the emotional states conveyed through it. Speech Emotion Recognition (SER) is gradually becoming a key link connecting users' emotional states with the responses of intelligent systems in scenarios such as intelligent customer service, vehicle interaction, and online education. Over the past decade, deep learning has greatly improved the recognition accuracy of SER, but it has also made the decision-making process of models increasingly opaque, forming a typical black box problem. This has significantly restricted the deployment of models in application scenarios with high requirements for safety and credibility. To alleviate this contradiction between performance and interpretability, this paper attempts to break away from the existing end-to-end deep network paradigm and construct a cognitive framework that combines deep feature learning and Concept-cognitive Learning (CCL). The original acoustic signal is deeply represented and features are extracted using a pre-trained encoder based on Whisper and a Multi-layer Perceptron (MLP), obtaining a compact 64-dimensional emotional feature representation. These features are then input into a weighted fuzzy concept-cognitive classification model based on Extremely Randomized Trees (ERT-WFCCL), completing the emotional discrimination in an explicitly interpretable concept space. Extremely Randomized Trees are used to automatically evaluate attribute importance from high-dimensional features, and the weighted fuzzy concept-cognitive component characterizes the reasoning basis of the model through concept lattice structures, thereby maintaining high recognition performance while providing more stable, traceable, and intuitive emotional recognition decisions for the aforementioned key application scenarios. This work was supported by the National Natural Science Foundation of China (62376229, 12371465) and the Natural Science Foundation of Chongqing (CSTB2023NSCQ-LZX0027).