Cross-modal Affinity-aligned Multimodal Learning Analytics for Predicting Student Collaboration Satisfaction in Game-Based Learning

Wen-Hsin Tsai, Chia-Ming Lee, Yuk-Ying Tung

Insititute of Education, National Cheng Kung University

yytung@mail.ncku.edu.tw

CVPR JUNE 3-7, 2026 DENVER COLORADO

CV4Edu Workshop Archival paper

Abstract

Motivation: In educational deployments, modalities such as eye gaze exhibit ***inconsistent informativeness*** across cohorts or may be ***entirely absent*** due to expensive data collection, causing implicit fusion to degrade satisfaction prediction.

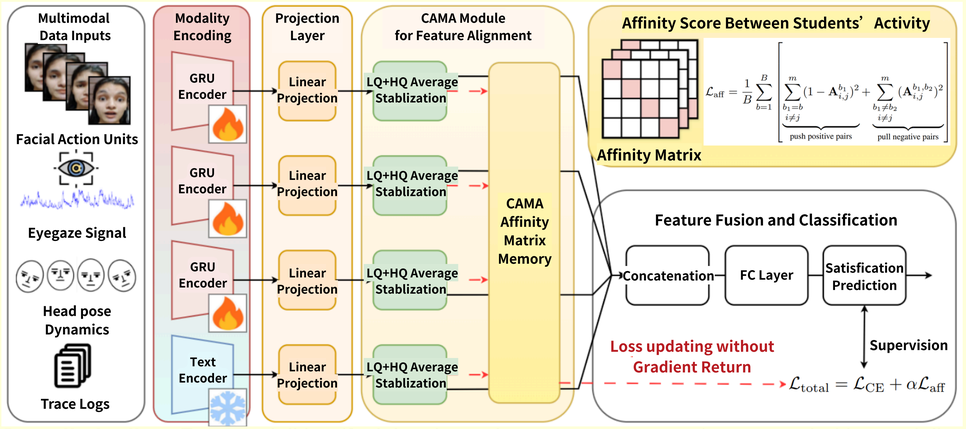
Abstract: We propose AAMLA, which explicitly aligns heterogeneous behavioral signals via affinity matrices to predict student collaboration satisfaction *robustly under real-world modality degradation*.

Contribution: CAMA suppresses uninformative modalities via affinity-guided contrastive learning, ensuring robust cross-modal representations without discarding any modality.

▲ *3D exploration view of the EcoJourneys collaborative learning environment.*

Dataset and Setting: 50 middle school students across 164 game sessions in **EcoJourneys**, capturing facial action units, head pose, eye gaze, and interaction trace logs. Inconsistent gaze signals across cohorts cause traditional fusion strategy to *over-weight noisy modalities*, cascading into prediction failure under real-world deployment conditions, such as missing modality.

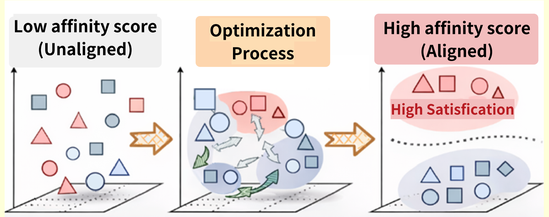
Affinity-Aligned Multimodal Learning Analytics (AAMLA)



Overall pipeline: Four modalities (facial action units, pose, gaze, trace) are encoded by modality-specific GRU encoders and projected into a *unified semantic space*. CAMA explicitly aligns cross-modal features via **affinity matrices and contrastive learning**, adaptively suppressing uninformative modalities such as inconsistent gaze signals. Aligned embeddings are jointly optimized with classification and affinity alignment losses before a shared FC classifier produces four-class satisfaction predictions, preventing cascading fusion errors under real-world degradation.

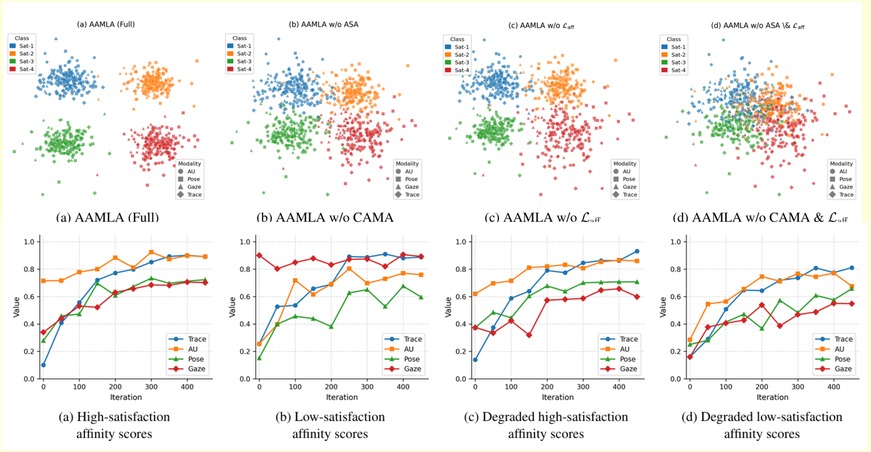
Key Insight - LQ-HQ Average Stablization: For each modality, original and perturbed features are averaged within each batch, mitigating within-batch discrepancies caused by student behavioral variation and providing stable representations regardless of modality quality fluctuations.

Cross-modal Affinity-guided Modality Alignment (CAMA)



CAMA Module: Affinity matrices explicitly model *inter-modal relationships* via contrastive learning, **pulling same-class embeddings together and pushing apart cross-activity pairs**, with average stabilization further ensuring robust alignment without additional computational overhead.

t-SNE Results and Affinity Score Analysis



(a) High-satisfaction affinity scores | (b) Low-satisfaction affinity scores | (c) Degraded high-satisfaction affinity scores | (d) Degraded low-satisfaction affinity scores

t-SNE Feature Distribution: Full AAMLA yields well-separated clusters; removing CAMA or contrastive loss progressively blurs class boundaries across modalities.

Affinity Score Evolution: Trace embeddings converge stably while gaze remains most variable; high-satisfaction activities reach alignment earlier.

Quantitative Results and Comparison

Model	F1-Score	Accuracy
AU (Unimodal)	0.66 ±0.03	0.66 ±0.03
Pose (Unimodal)	0.66 ±0.04	0.66 ±0.04
Gaze (Unimodal)	0.65 ±0.05	0.65 ±0.05
Trace (Unimodal)	0.65 ±0.04	0.65 ±0.04
Cross-Attention [1]	0.72 ±0.03	0.72 ±0.03
AAMLA (Ours)	**0.79** ±0.02	**0.77** ±0.02

Modality	Degradation	Cross-Attn [1]	AAMLA (Ours)
Gaze	Dropout 30%	0.68	0.77
	Dropout 50%	0.61	0.75
	Dropout 70%	0.52	0.72
AU + Pose	$\mathcal{N}(0, 0.01)$	0.65	0.78
	$\mathcal{N}(0, 0.05)$	0.54	0.72
Full Absence	w/o AU	0.54	0.73
	w/o Pose	0.53	0.72
	w/o Gaze	0.52	0.68
	w/o Trace	0.56	0.70

Unimodal vs. Multimodal Performance: Each modality contributes comparably; AAMLA consistently outperforms the cross-attention baseline.

Robustness under Modality Degradation and Missing: Cross-attention degrades sharply under modality degradation; AAMLA remains stable across all conditions.