Framelet Based Dual Hypergraph Neural Networks for Student Engagement Prediction

@Workshop on **AI for Education - Bridging Innovation and Responsibility (AI4ED-AAAI)**

Ming Li[1], Jiandong Shi[1]

[1]Zhejiang Normal University, Jinhua, China

February 26-27, 2024

Work-in-Progress Short Paper

Highlights of Frame-HGNN

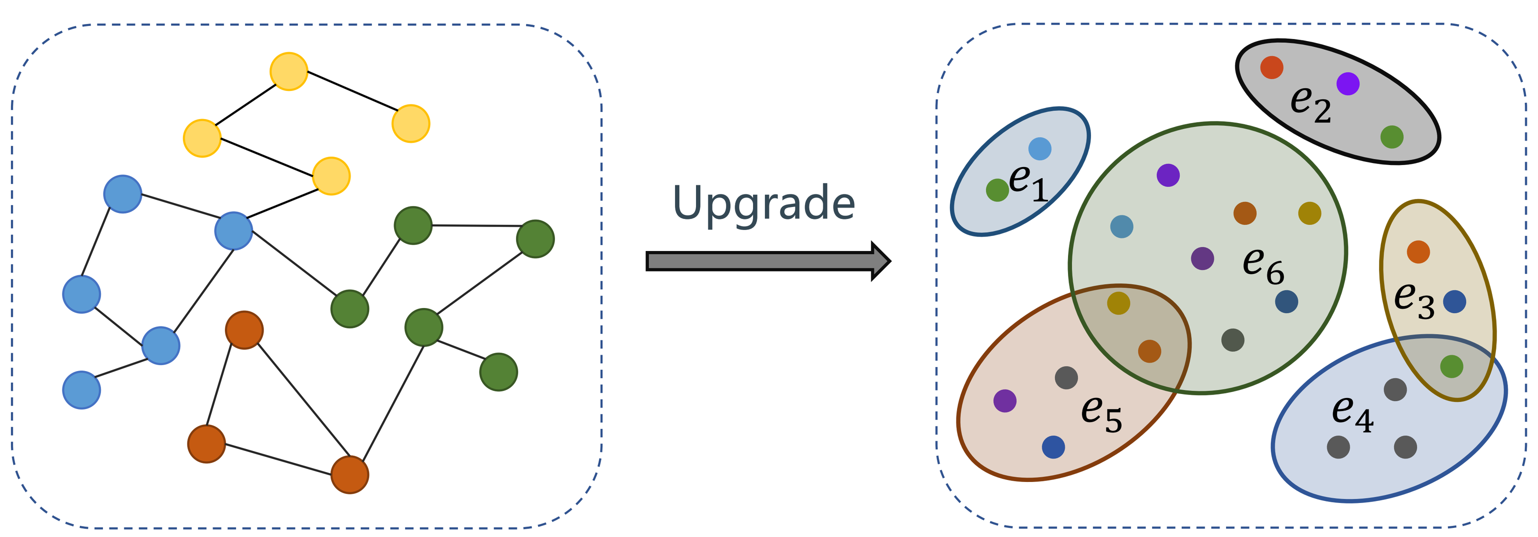
- **Harnessing the abundant educational data to predict student outcomes remains essential for proactive educational planning**
- **We introduce a framelet-based transform approach to capture high-order relationships among students from both low-pass and high-pass perspectives**
- **Instead of traditional GNN, we leverage a dual hypergraph neural network to learn and represent hypergraphs derived from both low-pass and high-pass components.**
- **Preliminary experimental findings on a real-world educational dataset highlight the promising potential of our framework**

Background and Motivation

- **Why framelet-based transform?**

Since the framelet-based transform identifies a frequency-domain operation, it refines and segregates the raw features into **a low-pass signal and six high-pass signals**, which distinctively highlights the similarities and differences among students

- **Why hypergraph neural network?**



✧ Beyond pair-wise relationship : Traditional graph structures can only reflect the relationships between linked nodes, while hypergraphs capture **higher-order information** between hyperedges

✧ Flexible hyperedge construction : Hypergraphs naturally provide a more flexible way to form hyperedges, e.g., prior knowledge, euclidean distance or different modalities.

Main Framework and Key Modules

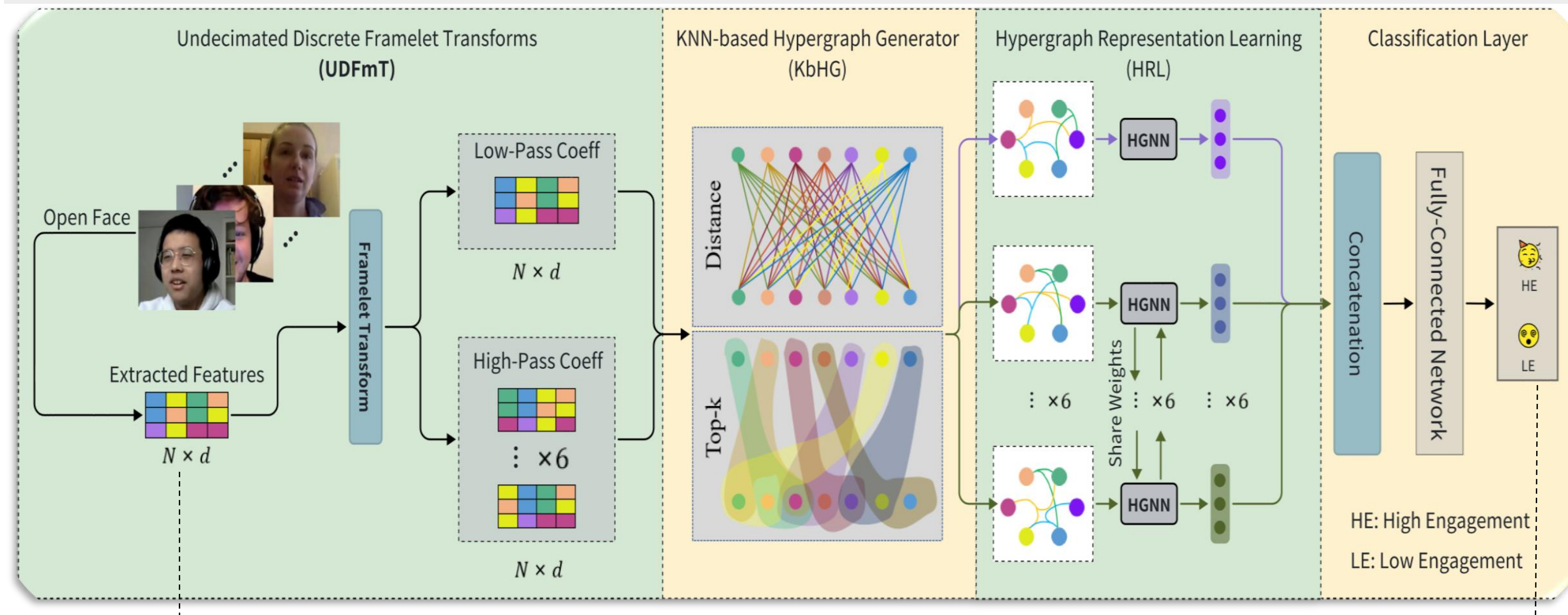


Figure 1: Overall framework of Frame-HGNN

Input: Pre-processed Student Feature Matrix X

Output: Student Engagement Label Y

$$\mathrm{DHF}_2 = \{a^H, b_1, \ldots, b_6\}$$

S1: Transform the features X

Construct a 2D directional Haar filter bank DHF_2

Decompose the features through DHF_2

Compute Euclidean distance $dis[i, j]$:

$$dis[i,j] = \sqrt{\sum_{k=1}^{d} (\mathbf{X}[i,k] - \mathbf{X}[j,k])^2}$$

S2: Generate a set of hypergraphs $\mathcal{G} = \{\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_7\}$

Treat the vectors in 7 coefficients matrices as nodes

Calculate the distance between nodes, then the nearest k-neighbors form a hyperedge.

S3: Hypergraphs representation learning

$$\mathbf{X}_{b_i}^{(\ell+1)} = \mathrm{HGNN}(\mathbf{X}_{b_i}^{(\ell)}, \mathbf{H}_{b_i}; \Theta_{high})$$

For low-pass filter coefficient matrix, employ an identical HGNN

For 6 high-pass filter coefficient matrices, leverage 6 share weights HGNN

Start hypergraph convolutional operation in 7 channels, respectively

Stack L layers and loop

S4: Predict the student engagement P

Concat the representation to get final prediction

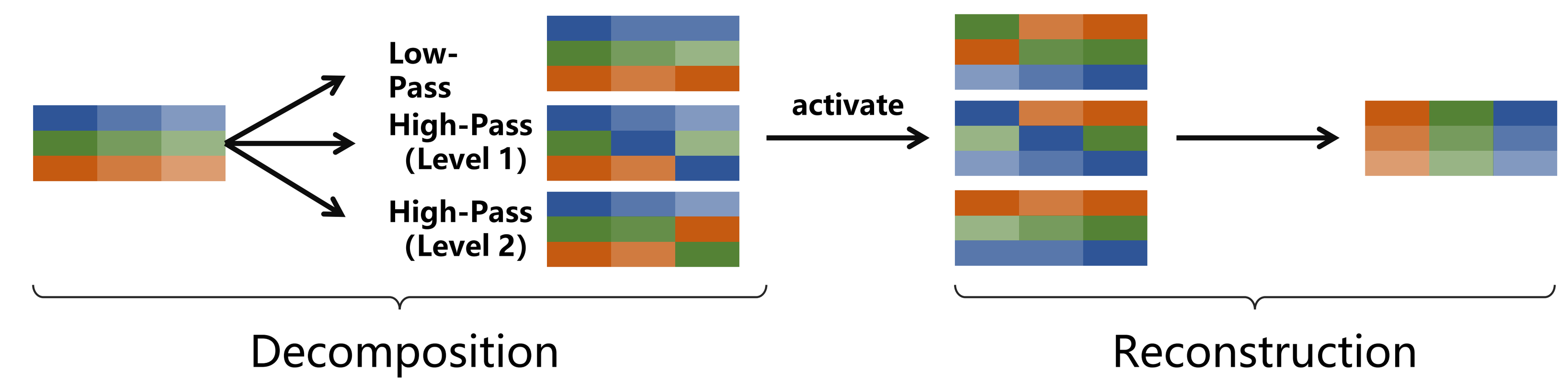
$$\mathbf{X}^{(\ell+1)} = \mathbf{D}_v^{-\frac{1}{2}} \mathbf{H} \mathbf{D}_e^{-1} \mathbf{H}^{\mathrm{T}} \mathbf{D}_v^{-\frac{1}{2}} \mathbf{X}^{(\ell)} \Theta$$

- **Through this, we can:**

✓ Refine the raw features

✓ Capture the higher-order relationships between students and peers from various views

Undecimated Discrete Framelet Transform



To decomposition the features, we initially define a 2D directional Haar filter bank $\mathrm{DHF}_2 = \{a, b_1, b_2, \ldots, b_6\}$, then, for a 2D filter h, we denote X_h the (circular) convolution of X with the 2D filter h, i.e., $X_h := X \star h$ with:

$$\mathbf{X}_h(k) := \sum_{k' \in \mathbb{Z}^2} \widetilde{\mathbf{X}}(k' - k) \cdot h(k'), \quad k = (k_1, k_2), k' = (k'_1, k'_2) \in \mathbb{Z}^2$$

Then, we can decompose X as follows:

$$\mathbf{X} = \mathbf{X}_a \star \bar{a} + \sum_{i=1}^{6} \mathbf{X}_{b_i} \star \bar{b}_i$$

Initial Experiments

- We conduct experiments on RoomReader dataset which includes over 8 hours of video and audio recordings, capturing the interactions of 118 participants across 30 sessions that take place in the online environment of Zoom.
- We compare our Frame-HGNN to four triditional methods including ConvLSTM, TEMMA, EnsModel and Bootstrap with the same data preprocess.

Table 1: The performance comparison of student engagement prediction accuracy

Method	ACC. (%)
ConvLSTM	76.50 ± 1.85
TEMMA	80.90 ± 2.47
EnsModel	75.30 ± 3.50
Bootstrap	73.80 ± 3.35
Ours	$\mathbf{85.38 \pm 1.41}$

Overall, the proposed Frame-HGNN achieves:

- Excellent performance on the student engagement prediction task with approximately 1.41 improvement on accuracy.
- Expressive feature repesentation ability across GNN achitectures.

Remark: Ongoing experiments are currently in progress, which will serve to extend and enhance this initial work further.