Multivariate, Multi-frequency and Multimodal: Rethinking Graph Neural Networks for Emotion Recognition in Conversation

Feiyu Chen[†‡] Jie Shao[†‡*] Shuyuan Zhu[†] Heng Tao Shen[†‡]
[†]University of Electronic Science and Technology of China, Chengdu, China
[‡]Sichuan Artificial Intelligence Research Institute, Yibin, China
{chenfeiyu, shaojie, eezsy, shenhengtao}@uestc.edu.cn

Code:https://github.com/feiyuchen7/M3NET

—— CVPR 2023

2023. 9. 17 • ChongQing

Leibniz-Institut für Sozialwissenschaften

Reported by JiaWei Cheng

Motivation

(1) complex multivariate relationships in ERC may not be sufficiently modelled by previous GNN-based methods.

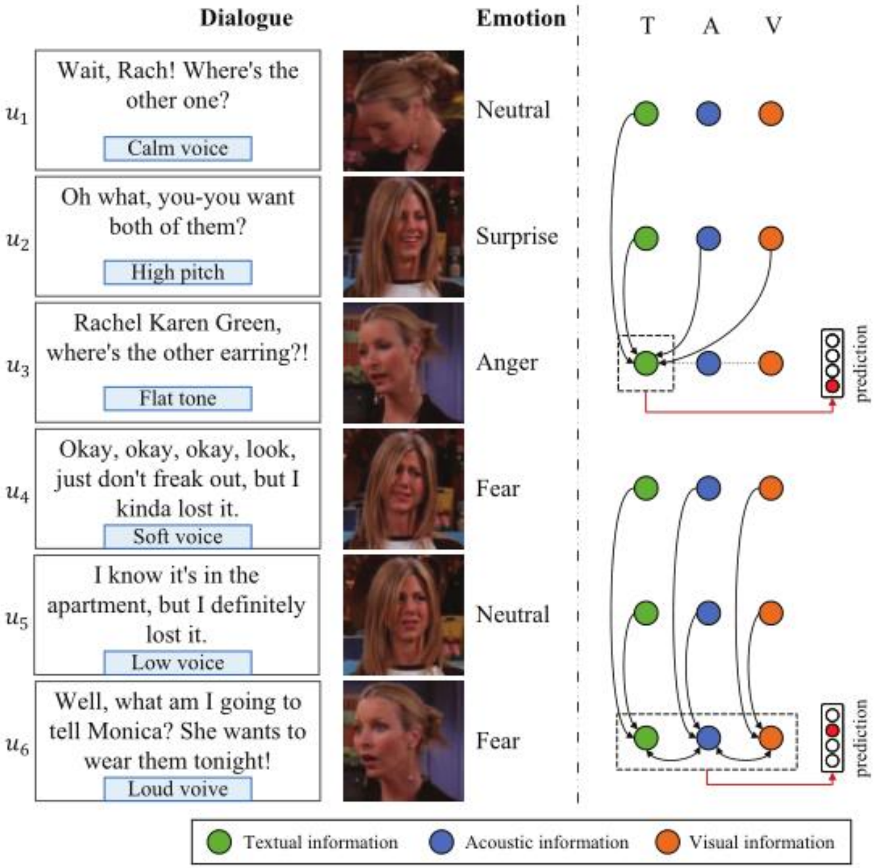


Figure 1. An example of multimodal dialogue (left) and the complex multivariate relationships of u_3 and u_6 (right).

Motivation

(2) It has been shown that the propagation rule of GNNs (i.e., aggregating and smoothing messages from neighbours) is an analogy to a fixed low-pass filter , and it is mainly low-frequency messages that flow in the graph while the effects of high-frequency ones are much weakened

Overview

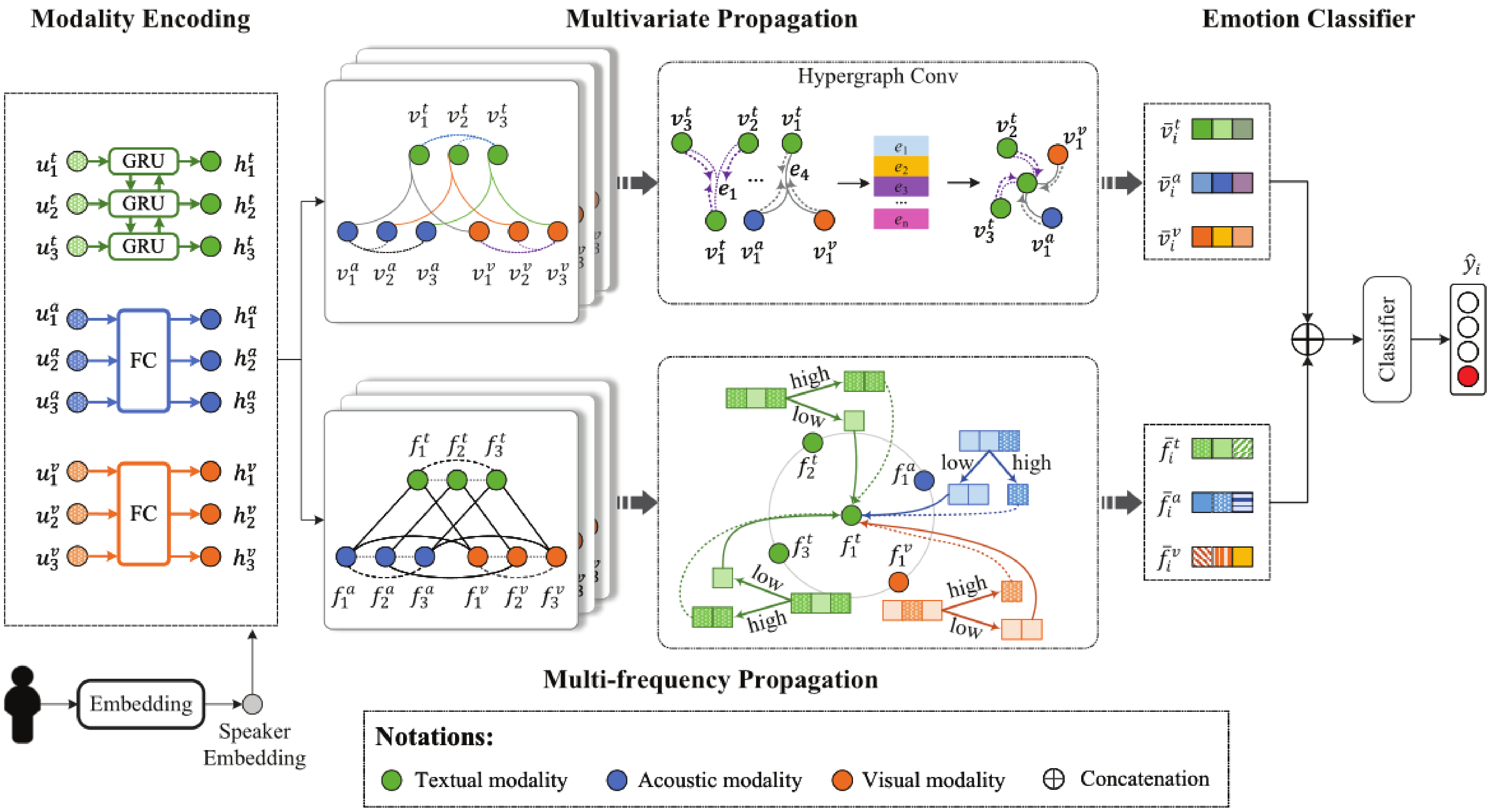
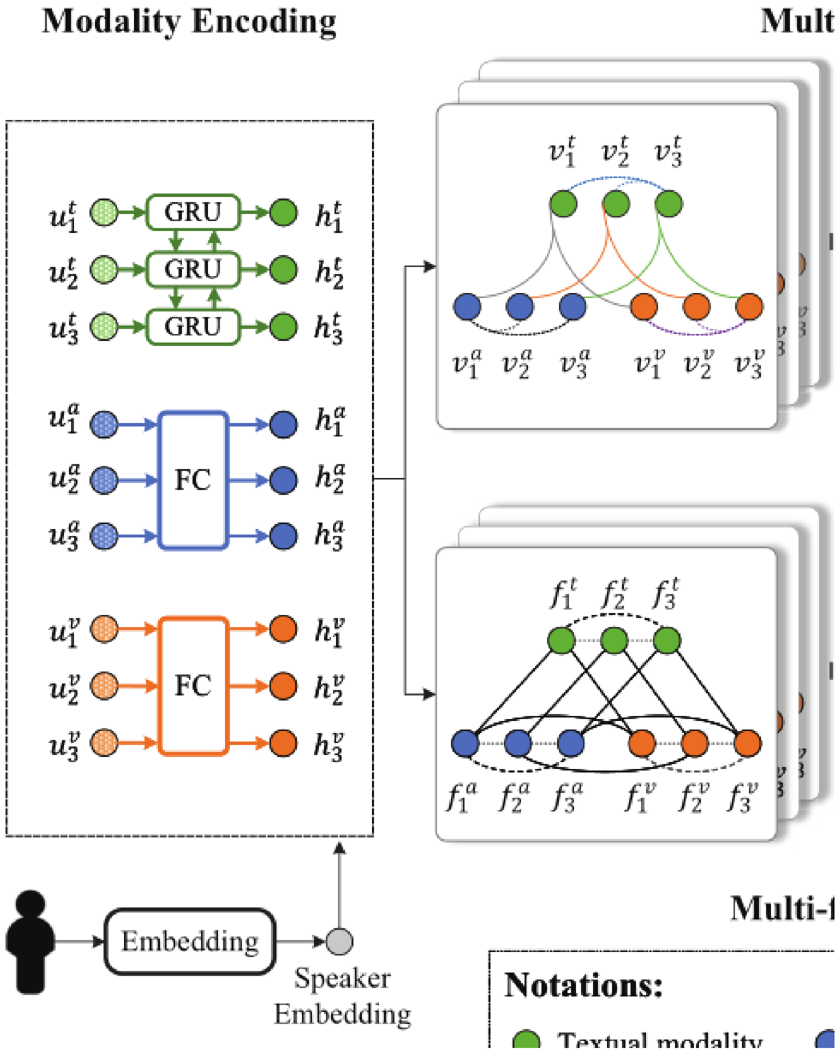


Figure 2. Detailed architecture of the proposed M^{3}Net.



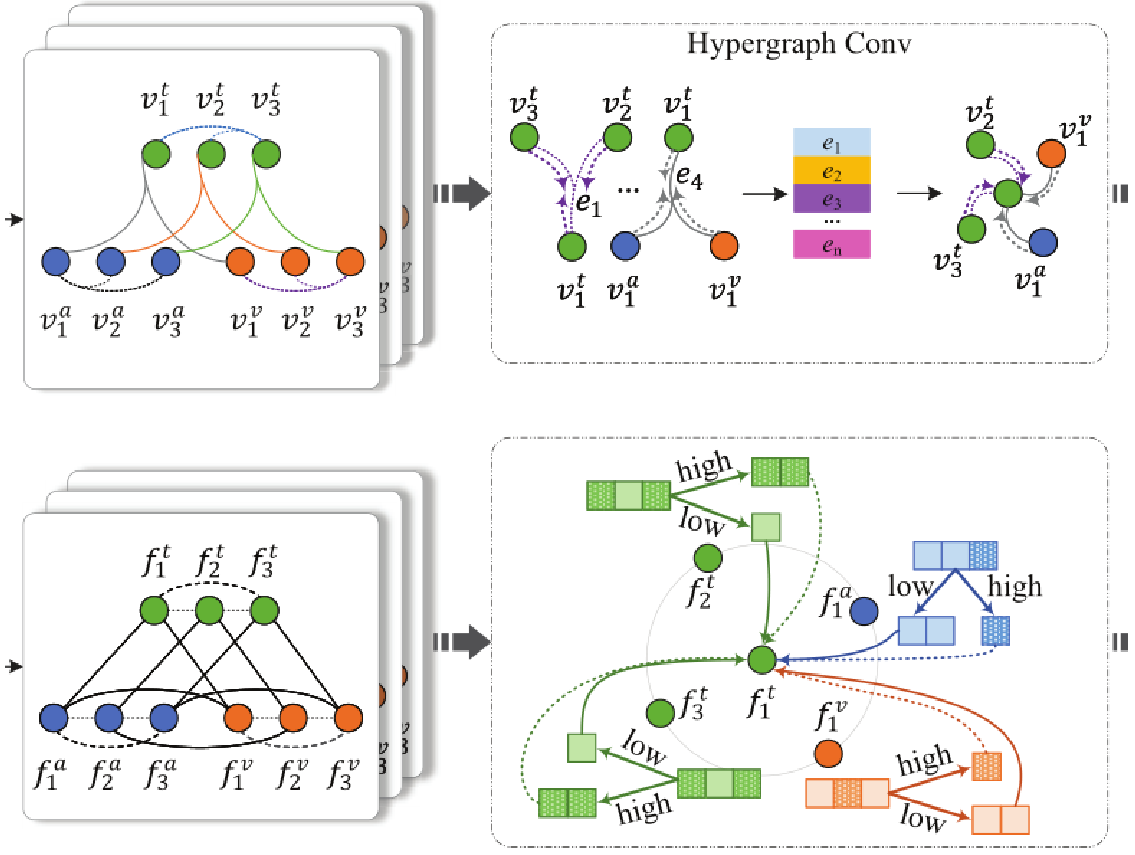
$$S_i = W_s s_i, \tag{1}$$

$$\begin{aligned} c_i^t &= \overleftrightarrow{GRU}(u_i^t, c_{i(+,-)1}^t), \\ c_i^a &= W_1 u_i^a + b_i^a, \\ c_i^v &= W_2 u_i^v + b_i^v, \end{aligned} \tag{2}$$

$$h_i^x = c_i^x + S_i, \quad x \in \{t, a, v\}. \tag{3}$$

Method

Multivariate Propagation



Multi-frequency Propagation

$$\hat{\mathbf{H}} = \begin{cases} \gamma_e(v), \text{ if hyperedge } e \text{ is incident with node } v; \\ 0, \text{ otherwise.} \end{cases} \quad (4)$$

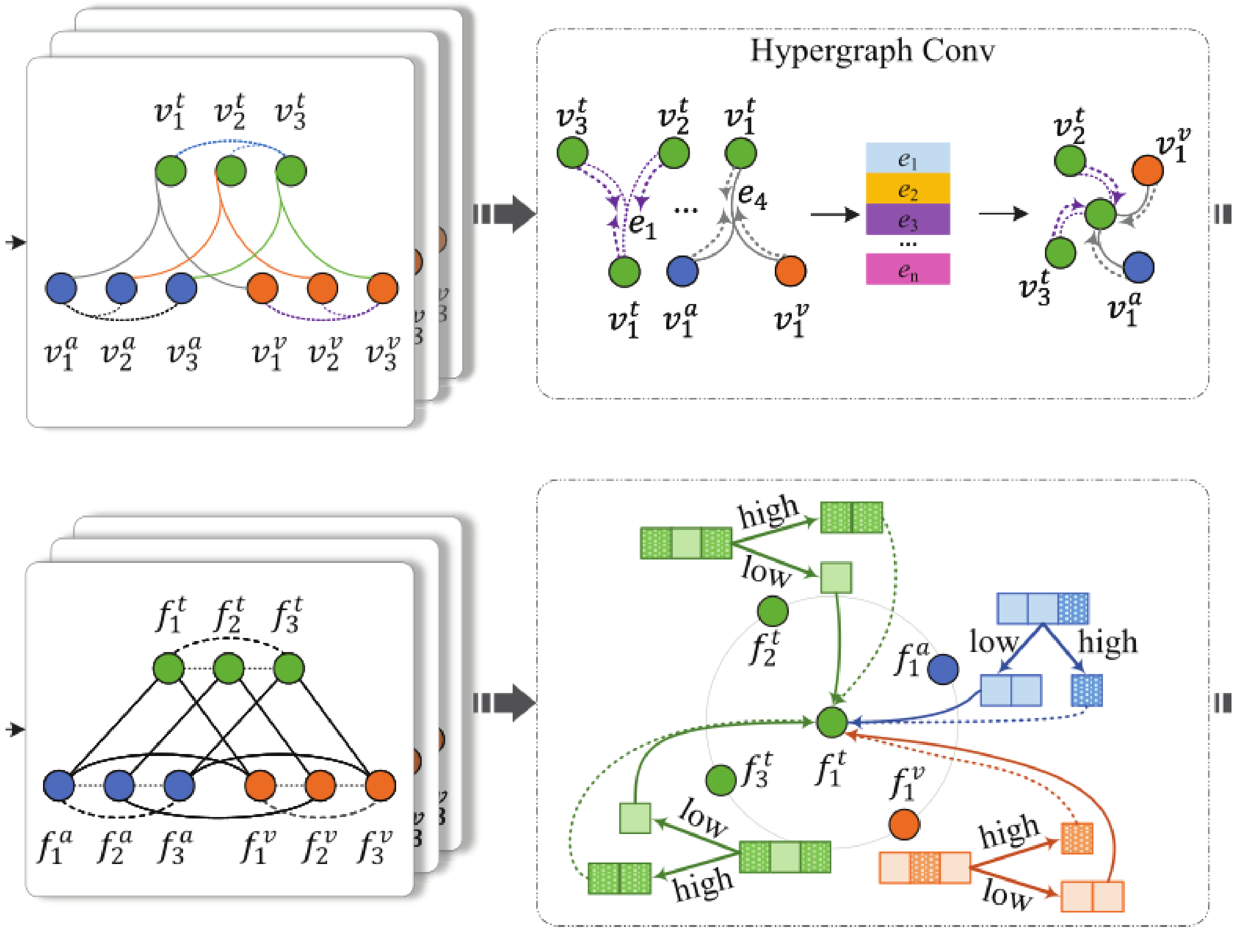
$$\mathbf{V}^{(l+1)} = \sigma(\mathbf{D}_{\mathcal{H}}^{-1}\mathbf{H}\mathbf{W}_e\mathbf{B}^{-1}\hat{\mathbf{H}}^{\top}\mathbf{V}^{(l)}), \quad (5)$$

Let $\mathbf{H} \in \mathbb{R}^{|\mathcal{V}_{\mathcal{H}}| \times |\mathcal{E}_{\mathcal{H}}|}$ represent the incidence matrix, in which a nonzero entry $H_{ve} = 1$ indicates that the hyperedge e is incident with the node v; otherwise $H_{ve} = 0$. in which $\mathbf{V}^{(l)} = \{v_{i,(l)}^{x} | i \in [1, N], x \in \{t, a, v\}\} \in \mathbb{R}^{|\mathcal{V}_{\mathcal{H}}| \times D_h}$ is the input at layer l. σ is a non-linear activation function. $\mathbf{W}_e = \text{diag}(w(e_1), ..., w(e_{|\mathcal{E}_{\mathcal{H}}|}))$ is the hyperedge weight matrix. $\mathbf{D}_{\mathcal{H}} \in \mathbb{R}^{|\mathcal{V}_{\mathcal{H}}| \times |\mathcal{V}_{\mathcal{H}}|}$ and $\mathbf{B} \in \mathbb{R}^{|\mathcal{E}_{\mathcal{H}}| \times |\mathcal{E}_{\mathcal{H}}|}$ are the node degree matrix and hyperedge degree matrix,

$$\overline{v_i^t} = v_{i,(L)}^t, \ \overline{v_i^a} = v_{i,(L)}^a, \ \overline{v_i^v} = v_{i,(L)}^v. \quad (6)$$

Method

Multivariate Propagation



Multi-frequency Propagation

$$\mathcal{F}_l = \mathbf{I} + \mathbf{D}_{\mathcal{G}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathcal{G}}^{-1/2} = 2\mathbf{I} - \mathbf{L},$$
$$\mathcal{F}_h = \mathbf{I} - \mathbf{D}_{\mathcal{G}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathcal{G}}^{-1/2} = \mathbf{L}. \quad (7)$$

$$\mathcal{F}_l *_C \varphi = \mathcal{F}_l \cdot \varphi,\ \mathcal{F}_h *_C \varphi = \mathcal{F}_h \cdot \varphi. \quad (8)$$

$$\begin{aligned}\mathbf{F}^{(k+1)} &= \mathbf{R}^l(\mathcal{F}_l \cdot \mathbf{F}^{(k)}) + \mathbf{R}^h(\mathcal{F}_h \cdot \mathbf{F}^{(k)}) \\ &= \mathbf{F}^{(k)} + (\mathbf{R}^l - \mathbf{R}^h)\mathbf{D}_{\mathcal{G}}^{-1/2}\mathbf{A}\mathbf{D}_{\mathcal{G}}^{-1/2}\mathbf{F}^{(k)},\end{aligned} \quad (9)$$

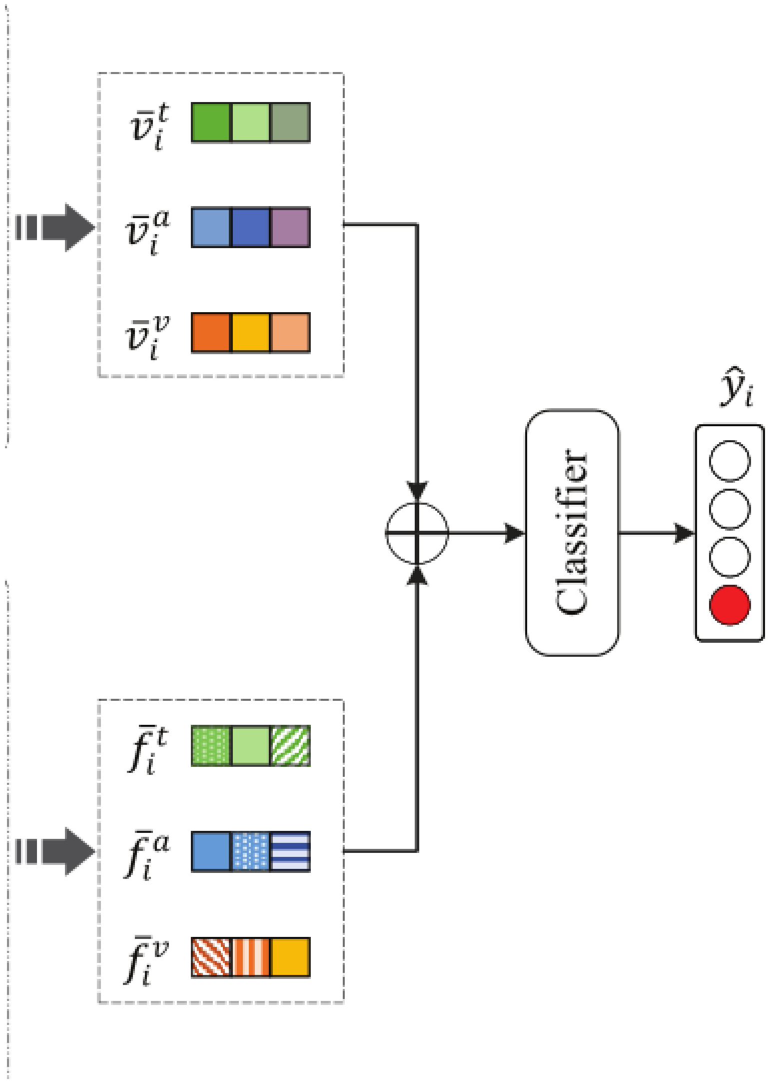
$$f_{i,(k+1)} = f_{i,(k)} + \sum_{j \in \mathcal{N}_i} \frac{r_{ij}^l - r_{ij}^h}{\sqrt{|\mathcal{N}_j|}\sqrt{|\mathcal{N}_i|}} f_{j,(k)}, \quad (10)$$

$$r_{ij}^l - r_{ij}^h = \tanh(W_3(f_{i,(k)} \oplus f_{j,(k)})). \quad (11)$$

$$\overline{f_i^t} = f_{i,(K)}^t,\ \overline{f_i^a} = f_{i,(K)}^a,\ \overline{f_i^v} = f_{i,(K)}^v. \quad (12)$$

Method

Emotion Classifier



$$e_i = \overline{v_i^t} \oplus \overline{f_i^t} \oplus \overline{v_i^a} \oplus \overline{f_i^a} \oplus \overline{v_i^v} \oplus \overline{f_i^v}, \qquad (13)$$

$$\begin{aligned} \tilde{e}_i &= \mathrm{ReLU}(e_i), \\ \mathcal{P}_i &= \mathrm{softmax}(W_4 \tilde{e}_i + b_4), \\ \hat{y}_i &= \underset{\tau}{\mathrm{argmax}}(\mathcal{P}_i[\tau]), \end{aligned} \qquad (14)$$

$$L = -\frac{1}{\sum_{s=1}^{Num} c(s)} \sum_{i=1}^{Num} \sum_{j=1}^{c(i)} log \mathcal{P}_{i,j}[y_{i,j}] + \lambda \left\| \theta \right\|_2, \quad (15)$$

Experiments

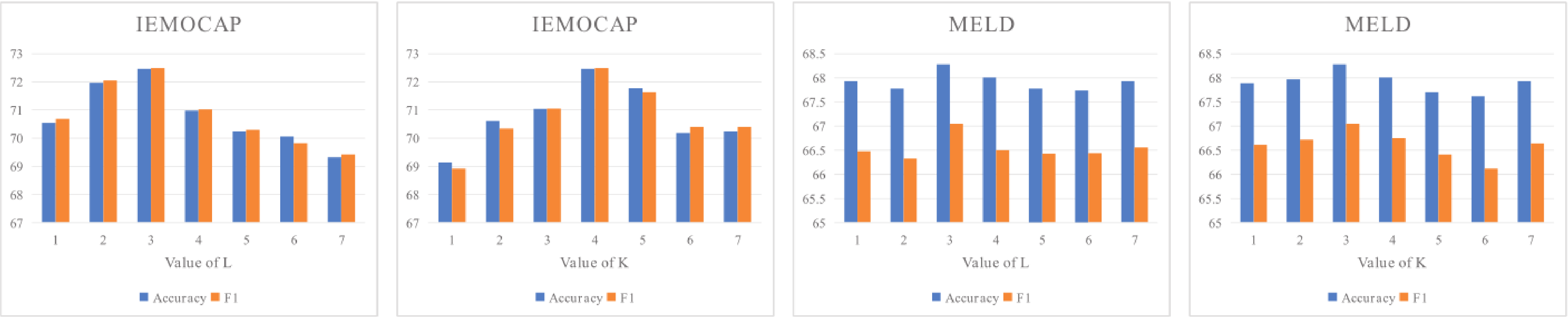
	Methods	Network	IEMOCAP Average (w)		MELD Average (w)	
			Accuracy	F1	Accuracy	F1
GloVe	CMN$^{\triangleleft}$ [14]	Non-GNN	-	58.50	-	-
	ICON* [13]	Non-GNN	64.00	63.50	-	-
	DialogueRNN† [24]	Non-GNN	63.51	62.90	59.92	57.60
	MetaDrop$^{\diamond}$ [5]	Non-GNN	65.76	65.58	-	58.30
	DialogueGCN† [12]	GNN-based	66.17	66.24	57.01	55.59
	MMGCN† [18]	GNN-based	65.80	65.41	60.42	58.31
	MM-DFN† [17]	GNN-based	68.21	68.18	59.81	58.42
	M^{3}Net (ours)	GNN-based	**69.50**	**69.08**	**61.65**	**59.22**
RoBERTa	DialogueGCN† [12]	GNN-based	63.96	64.44	63.49	62.78
	MMGCN† [18]	GNN-based	66.79	66.99	66.63	65.13
	DialogueRNN$^{\diamond}$ [24]	Non-GNN	68.64	68.72	65.94	65.31
	MetaDrop$^{\diamond}$ [5]	Non-GNN	69.38	69.59	66.63	66.30
	MM-DFN† [17]	GNN-based	69.87	69.48	67.01	66.17
	M^{3}Net (ours)	GNN-based	**72.46**	**72.49**	**68.28**	**67.05**

Experiments

	Methods	IEMOCAP		MELD	
		Acc.	F1	Acc.	F1
	M^3Net	72.46	72.49	68.28	67.05
1	w/o multivariate info.	70.06	70.05	67.74	66.36
2	w/o multi-frequency info.	69.87	69.74	67.36	66.03
3	w/o hyperedge weight $\omega(e)$	70.30	70.45	68.11	66.99
4	w/o node weight $\gamma_e(v)$	70.98	71.02	68.05	66.92
5	w/o both weights	70.12	70.09	67.89	66.75
6	$\mathcal{H} \to \mathcal{G}$ in series	68.39	68.44	68.20	66.84
7	$\mathcal{G} \to \mathcal{H}$ in series	69.50	69.70	68.05	66.85

Table 3. Ablation studies of M^3Net.

Experiments



(a) Effect of L on IEMOCAP (b) Effect of K on IEMOCAP (c) Effect of L on MELD (d) Effect of K on MELD

Figure 3. Results of M^{3}Net at different graph layers. In (a) and (c), effects of L are tested by fixing K as in the best-performing models. In (b) and (d), effects of K are tested by fixing L as in the best-performing models.

Experiments

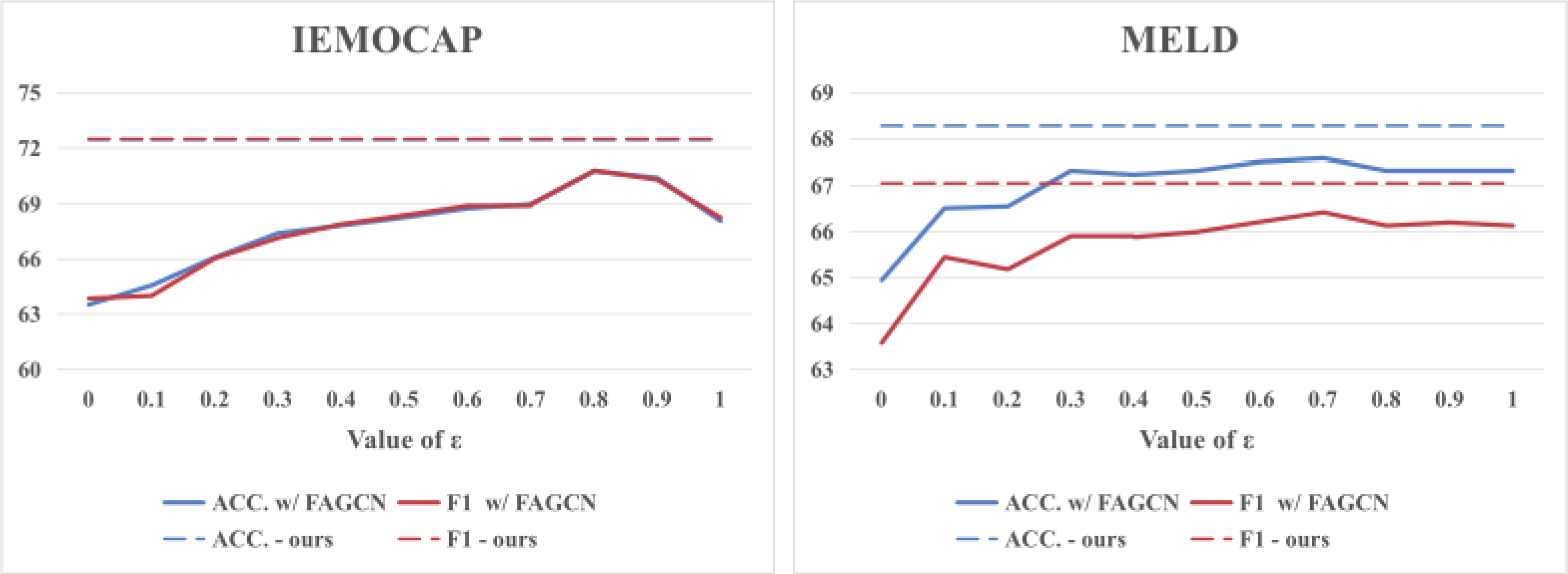


Figure 4. Performance comparison with FAGCN.

Thanks!