Multimodal Dialogue Systems in Ping An Life

Dr. Haiqin Yang

Ping An Life

Who Are We?

great field to Explore and Exploit

- **Ping An Life Insurance**
  - Life insurance products: long period of protection, coverage comprehension, complex, …
  - Fin. + **Tech.**

Agents

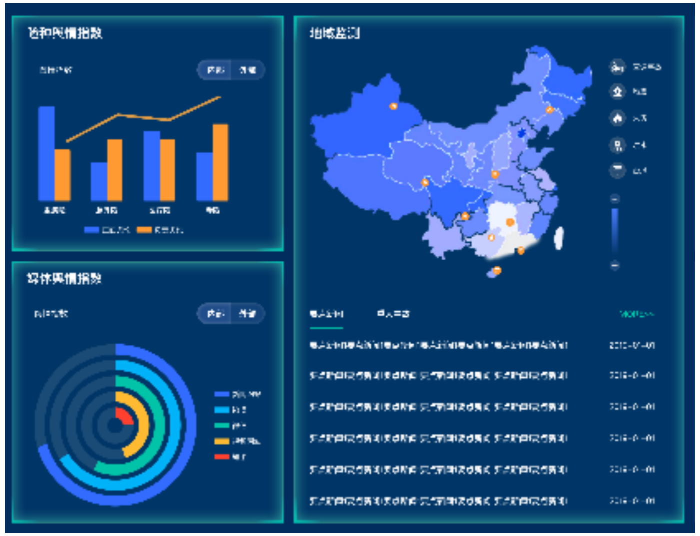
1.2M+

Customers

200M+

Our AI Technologies

Computational Intelligence



Deep Learning,
Recommendation

Perceptual Intelligence

Computer Vision

Cognitive Intelligence

NLP, Knowledge Graphs,
Chatbots

Core Applications

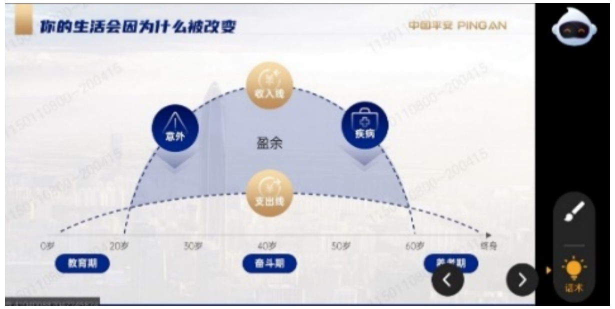
- **To A**gents
  - AskBob for agents
  - Training assistant
  - Visiting assistant
  - …

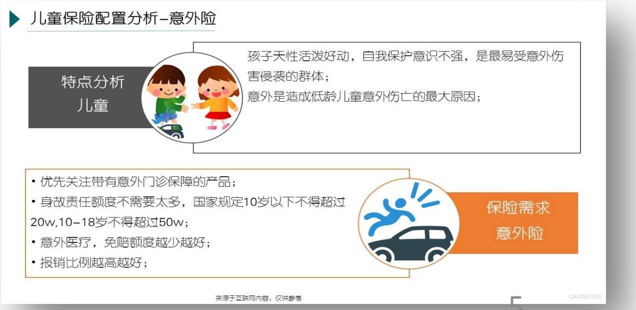
- **To C**ustomers
  - AskBob for financial customer services
  - Video follow-up chatbot
  - …

Professional Insurance Services: Visiting Assistant

- The first online Visiting Assistant
  - AI guest room
- Features
  - One-click explanation
  - Full accompany
  - Content creation

Dialog Q&A	Real-time Prediction
Real-time Conversation Assistance	Content Generation





Professional Insurance Services: Video Follow-up Chatbot

Prompt Initiate → Multimodal Dialogue → Realtime Services

- **Impersonate video interaction**
  - User-friendly
- **Flexible follow-up**
  - 7*24 online service
- **Efficient access to insurance policy**
  - Real-time Q&A

- The **first industry multi-modal** follow-up chatbot
  - Face-to-face interaction
  - Time reduces from 5 days to 2.8 minutes

Image Generation	Face Recognition
Lips Matching	Active Dialogue Guidance

Insurance Middle Platform (MP)

Statistical Artificial Intelligence & Learning

Applications

- Recruit-ment
- Train-ing
- Sales
- Services
- Manage ment
- Risk control
- Asset man.
- …

Personalized development

Standard deployment

Insurance MP

App. layer

- Chatbot MP
- NLP MP
- Knowledge MP
- Content MP
- Image MP

Comp. layer

- Deep Learning Platform
  - Training
  - Inference
  - Scheduling
- AI Algo.
  - Deep Learning
  - Reinforce. Learning
  - ……

Data layer

- Features Labels
- Dialogue
- Knowledge
- Corpus
- Multimedia
- ……

Chatbots

Text

Champions

7

DSTC8, SemEval2020

Speech

Papers

20+

IJCAI, EMNLP, …

Video

Patents

300 +

System, Technology, …

20+ Chatbots

10+ Platforms

100+ Services

A unified multi-modal chatbot platform

Minneapolis, MN
NAACL HLT 2019
June 2-7, 2019

Empathy Dialogue Systems

- HiGRU: Hierarchical Gated Recurrent Units for Utterance-level Emotion Recognition, NAACL'19, Joint work with CUHK
- DialogueTRM: Exploring the Intra- and Inter-Modal Emotional Behaviors in the Conversation (Under review)

Motivation: Emotion Recognition in Dialogues

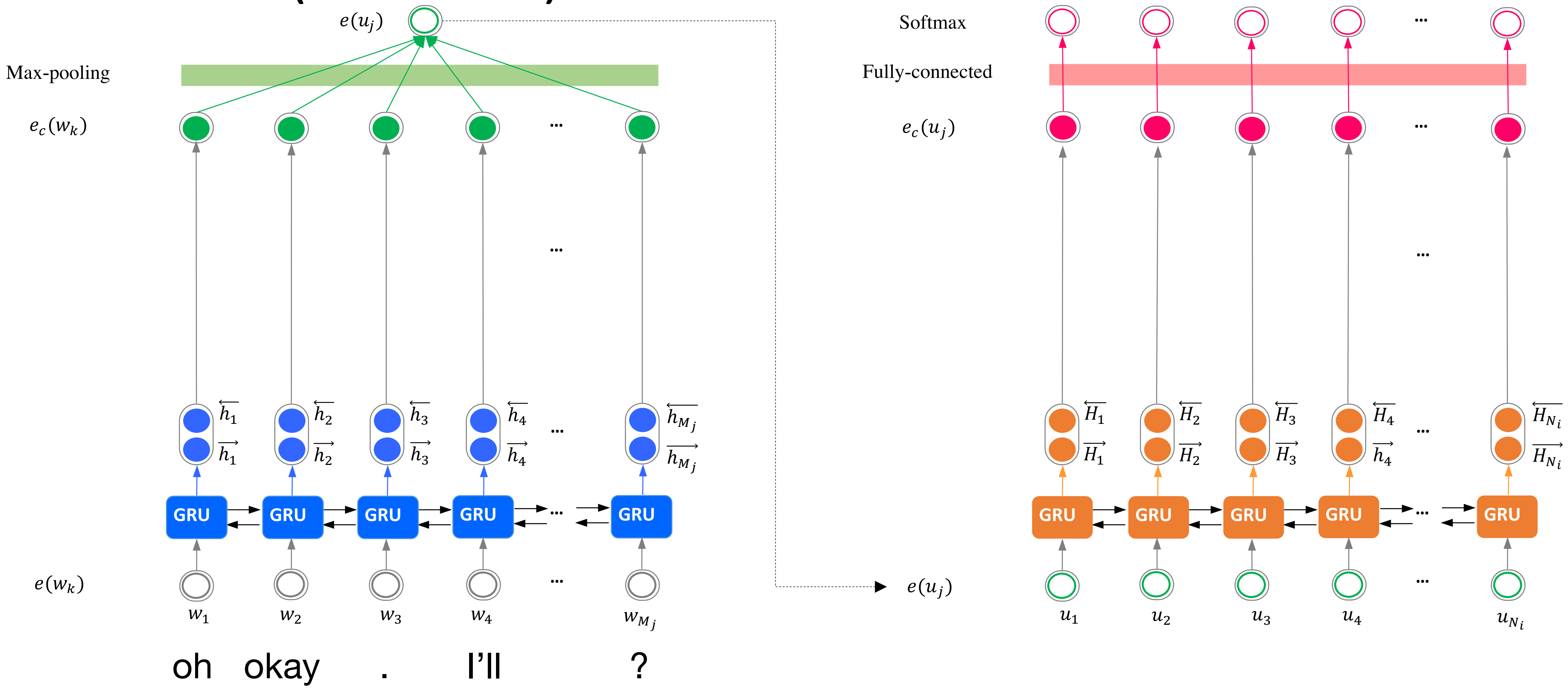
- The same word delivers different emotions
- Some emotions rarely appear
- Long-range contextual information is hard to captured

Role	Utterance	Emotion
Rachel	Oh okay, I'll fix that to. What's her e-mail address?	Neutral
Ross	Rachel!	Anger
Rachel	All right, I promise. I'll fix this. I swear. I'll-I'll- I'll-I'll talk to her.	Non-neutral
Ross	***Okay!***	***Anger***
Rachel	***Okay.***	***Neutral***
Nurse	This room's available.	Neutral
Rachel	***Okay!***	***Joy***
Rachel	Okay wait!	Non-neutral
Rachel	You listen to me!	Anger

Problem Definition: Utterance-level Emotion Recognition

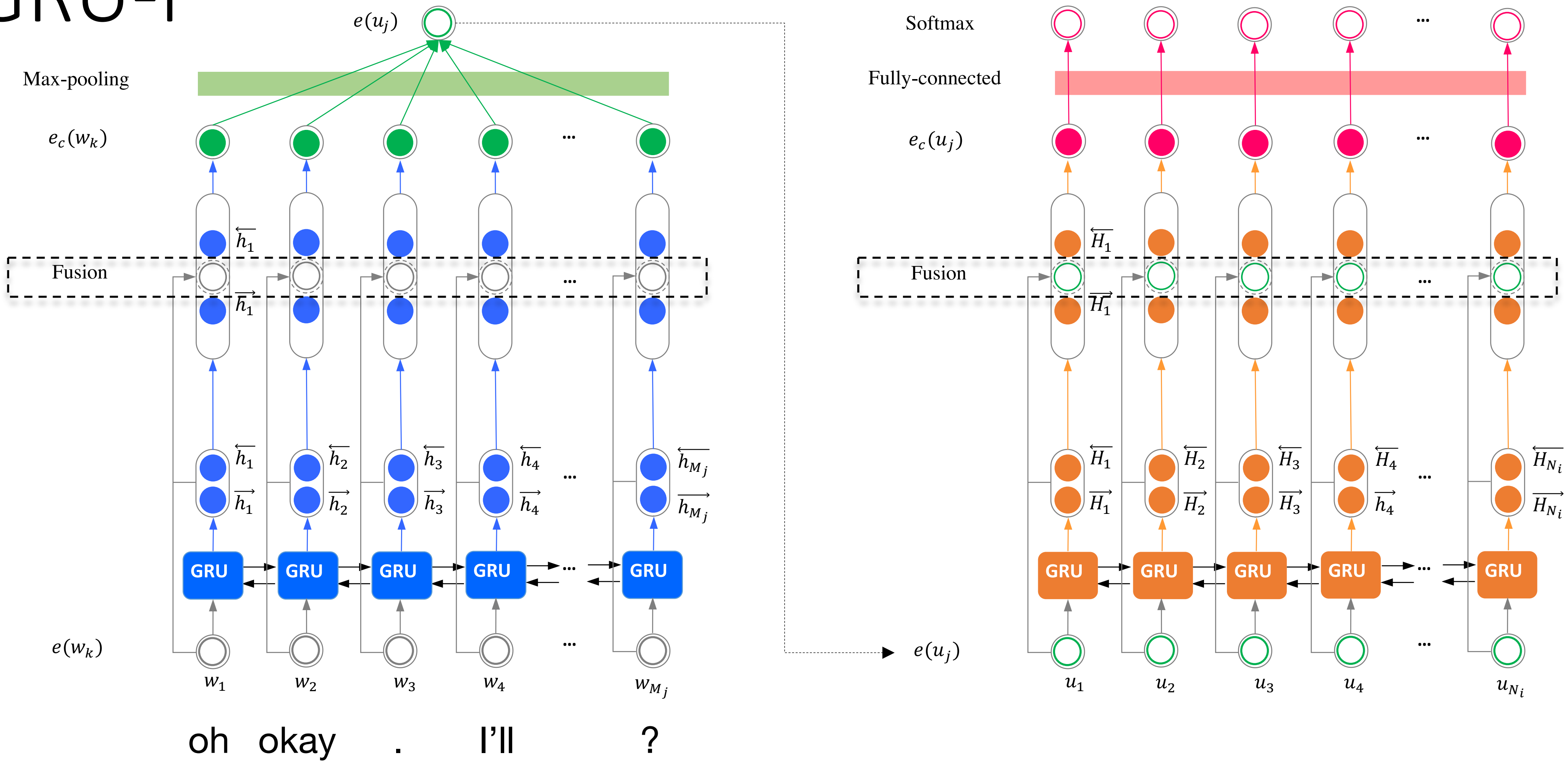
- **Input**: a set of dialogues, $\mathcal{D}= \{D_i\}_{i=1}^{L}$
  - L: the number of dialogues
  - $D_i = \{(u_j, s_j, c_j)\}_{j=1}^{N_i}$: a dialogue
    - u_j: utterance
    - s_j: speaker
    - c_j: emotion
- **Goal**: to train a model to tag each new utterance with an emotion label as accurately as possible

Our Proposal: Hierarchical Gated Recurrent Units (HiGRU)



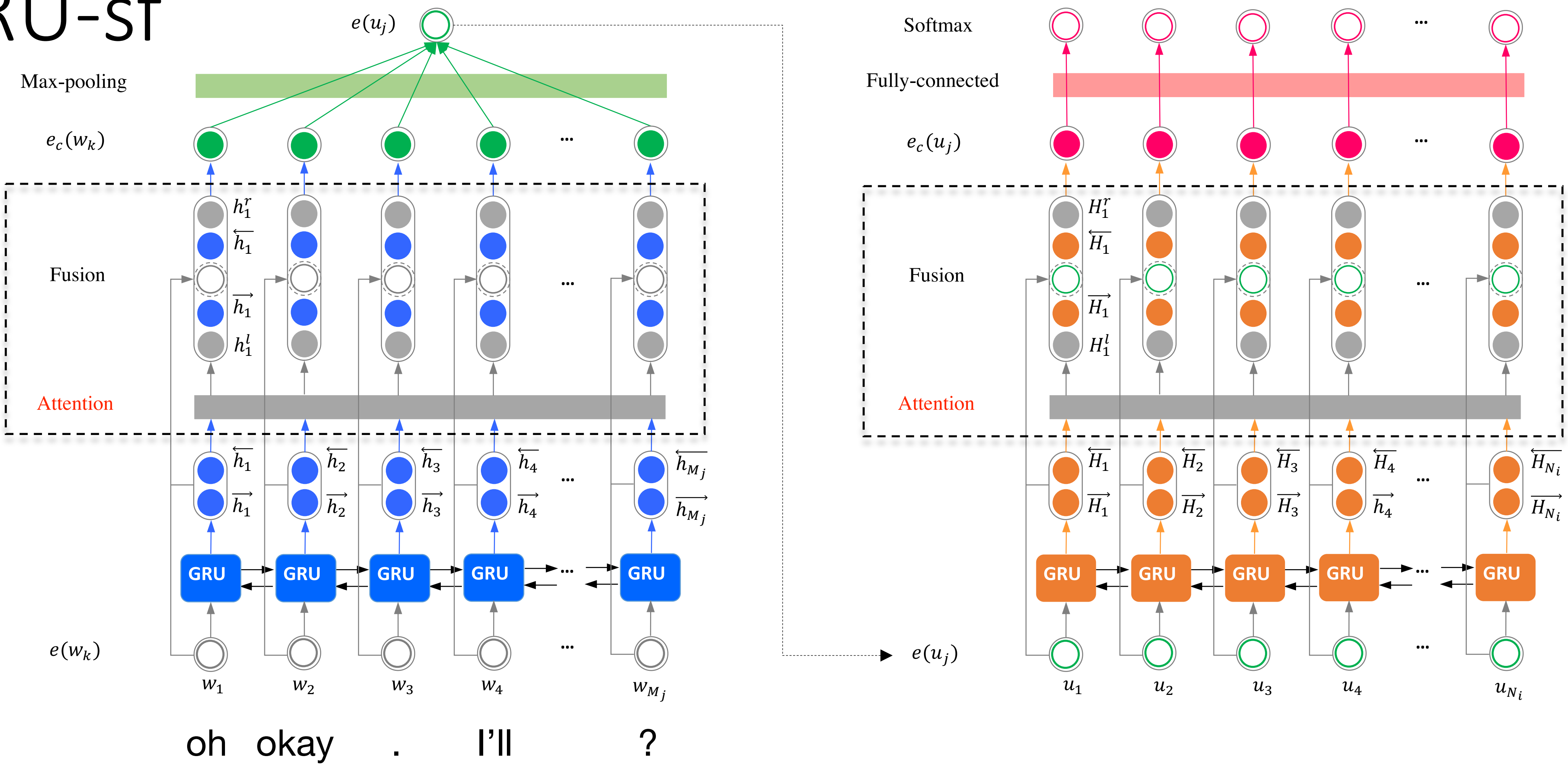
- **Lower-level bi-GRU**: individual utterance embedding
- **Upper-level bi-GRU**: contextual utterance embedding

HiGRU-f



- Fuse individual word/utterance embeddings to strengthen individual information

HiGRU-sf



- Self-Attention + Feature Fusion

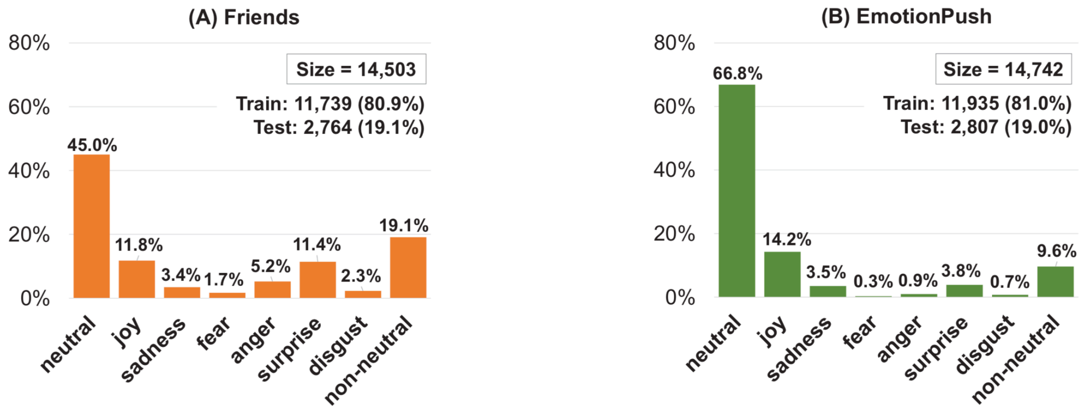
Model Training

- **Minimize weighted categorical cross-entropy**

$$loss = -\frac{1}{\sum_{i=1}^{L} N_i} \sum_{i=1}^{L} \sum_{j=1}^{N_i} \omega(c_j) \sum_{c=1}^{|\mathcal{C}|} y_j^c \log_2(\hat{y}_j^c)$$

  - predicted label: $\hat{y}_j = \mathrm{softmax}(W_{fc} \cdot e_c(u_j) + b_{fc})$
  - weight: $\frac{1}{\omega(c)} = \frac{I_c^\alpha}{\sum_{c'=1}^{|\mathcal{C}|} I_{c'}^\alpha}$

Experiments



- **Datasets**

Dataset	#Dialogue (#Utterance)			#Emotion				
	Train	Val	Test	Ang	Hap/Joy	Sad	Neu	Others
IEMOCAP	96 (3,569)	24 (721)	31 (1,208)	1,090	1,627	1,077	1,704	0
Friends	720 (10,561)	80 (1,178)	200 (2,764)	759	1,710	498	6,530	5,006
EmotionPush	720 (10,733)	80 (1,202)	200 (2,807)	140	2,100	514	9,855	2,133

- IEMOCAP: ~12 hours of audiovisual data, including video, speech, motion capture of face, text transcriptions
- Friends: Friends TV show transcripts
- EmotionPush: Facebook messenger logs

- **Metrics**
  - Weighted Accuracy (WA)
  - Unweighted Accuracy (UWA)

$$\mathrm{WA} = \sum_{c=1}^{|\mathcal{C}|} p_c \cdot a_c, \qquad \mathrm{UWA} = \frac{1}{|\mathcal{C}|} \sum_{c=1}^{|\mathcal{C}|} a_c,$$

Methods and Setup

- **Methods**
  - Existing methods: bcLSTM, CMN, SA-BiLSTM, CNN-DCNN
  - Our implementation: bcLSTM$_*$(weighted loss), bcGRU(weighted loss), HiGRU, HiGRU-f, HiGRU-sf
- **Parameters of** HiGRU, HiGRU-f, HiGRU-sf
  - # hidden states: 300
  - FC layer: two sub-layers with 100 neurons each
- **Training:** Adam, Anneal strategy, early stop, gradient clipping, dropout

Results on IEMOCAP

Model (Feat)	Ang	Hap	Sad	Neu	WA	UWA
bcLSTM[1] (T)	76.07	78.97	76.23	67.44	73.6	74.6
(T+V+A)	77.98	79.31	78.30	69.92	76.1	76.3
CMN[2] (T)	-	-	-	-	74.1	-
(T+V+A)	**89.88**	81.75	77.73	67.32	77.6	79.1
bcLSTM$_*$ (T)	75.29	79.40	78.07	76.53	77.7(1.1)	77.3(1.4)
bcGRU (T)	77.20	80.99	76.26	72.50	76.9(1.6)	76.7(1.3)
HiGRU (T)	75.41	**91.64**	79.79	70.74	80.6(0.5)	79.4(0.5)
HiGRU-f (T)	76.69	88.91	80.25	75.92	81.5(0.7)	80.4(0.5)
HiGRU-sf (T)	74.78	89.65	**80.50**	**77.58**	**82.1**(0.4)	**80.6**(0.2)

[1] by (Poria et al., 2017); [2] by (Hazarika et al., 2018).

- In WA: HiGRU achieves at least
  - 8.7% improvement over CMN (T) and
  - 3.8% improvement over CMN (T+V+A)

Results on Friends and EmotionPush

Model	Train	Friends (F)						EmotionPush (E)					
		Ang	Joy	Sad	Neu	WA	UWA	Ang	Joy	Sad	Neu	WA	UWA
SA-BiLSTM[1]	F+E	49.1	68.8	30.6	**90.1**	-	59.6	24.3	70.5	31.0	**94.2**	-	55.0
CNN-DCNN[2]	F+E	55.3	71.1	55.3	68.3	-	62.5	45.9	76.0	51.7	76.3	-	62.5
bcLSTM$_*$	F(E)	64.7	69.6	48.0	75.6	72.4(4.2)	64.4(1.6)	32.9	69.9	47.1	78.0	74.7(4.4)	57.0(2.1)
bcGRU	F(E)	69.5	65.4	52.9	74.7	71.7(4.7)	65.6(1.2)	33.7	71.1	57.2	76.1	73.9(2.9)	59.5(1.8)
bcLSTM$_*$	F+E	54.5	75.6	43.4	73.0	70.5(4.5)	61.6(1.6)	52.4	79.1	54.7	73.3	73.4(3.8)	64.9(2.1)
bcGRU	F+E	59.0	78.6	42.3	71.4	70.2(5.1)	62.8(1.4)	49.4	74.8	61.9	72.4	72.1(4.3)	64.6(1.8)
HiGRU	F(E)	66.9	73.0	51.8	77.2	**74.4**(1.7)	67.2(0.6)	55.6	78.1	57.4	73.8	73.8(2.0)	66.3(1.7)
HiGRU-f	F(E)	69.1	72.1	**60.4**	72.1	71.3(2.9)	68.4(1.0)	55.9	78.9	60.4	72.4	73.0(2.2)	66.9(1.2)
HiGRU-sf	F(E)	**70.7**	70.9	57.7	76.2	74.0(1.4)	**68.9**(1.5)	57.5	78.4	64.1	72.5	73.0(1.6)	68.1(1.2)
HiGRU	F+E	55.4	81.2	51.4	64.4	65.8(4.2)	63.1(1.5)	50.8	76.9	69.0	75.7	75.3(1.7)	68.1(1.2)
HiGRU-f	F+E	54.9	78.3	55.5	68.7	68.5(3.0)	64.3(1.2)	**58.3**	79.1	**69.6**	70.0	71.5(2.5)	69.2(0.9)
HiGRU-sf	F+E	56.8	**81.4**	52.2	68.7	69.0(2.0)	64.8(1.3)	57.8	**79.3**	66.3	77.4	**77.1**(1.0)	**70.2**(1.1)

[1] by (Luo et al., 2018); [2] by (Khosla, 2018).

- HiGRU gains at least 6.0% improvement over CNN-DCNN, the best performance
- HiGRU-f and HiGRU-sf perform better than HiGRU
- Training with mixed datasets can only help the imbalanced dataset, EmotionPush

Successful Cases

- **Scene-1:** both success
- **Scene-2:**
  - bcGRU: Joy → Ang
- **Scene-3:**
  - bcGRU:
    - Sad → Hap
    - Sad → Neu
  - HiGRU-sf:
    - Hap → Sad

Wrong labeled!

Role	Utterance	Truth	bcGRU	HiGRU-sf
Scene-1				
Phoebe	Okay. Oh but don't tell them Monica's pregnant because they frown on that.	Neu	Neu	Neu
Rachel	Okay.	Neu	Neu	Neu
Phoebe	Okay.	Neu	Neu	Neu
Scene-2				
Phoebe	Yeah! Sure! Yep! Oh, y'know what? If I heard a shot right now, I'd throw my body on you.	Joy	Ang	Joy
Gary	Oh yeah? Well maybe you and I should take a walk through a bad neighborhood.	Other	/	/
Phoebe	Okay!	Joy	Ang	Joy
Gary	All right.	Neu	Neu	Neu
Scene-3				
Female	Can I send you, like videos and stuff? What about when they start walking.	Other	/	/
Male	Yeah yeah yeah.	Sad	Hap	Sad
Male	You you record every second. You record every second because I want to see it all. Okay?	Hap	Hap	Sad
Male	If I don't get to see it now, I get to see it later at least, you know? You've got to keep it all for me; all right?	Other	/	/
Female	Okay.	Sad	Neu	Sad

Failed Cases

- **Scene-4:**
  - bcGRU: Joy → Sad
  - HiGRU: Joy → Neu
- **Scene-5:**
  - bcGRU:
    - Neu → Sad
    - Joy → Neu
  - HiGRU-sf
    - Neu → Sad
    - Joy → Sad

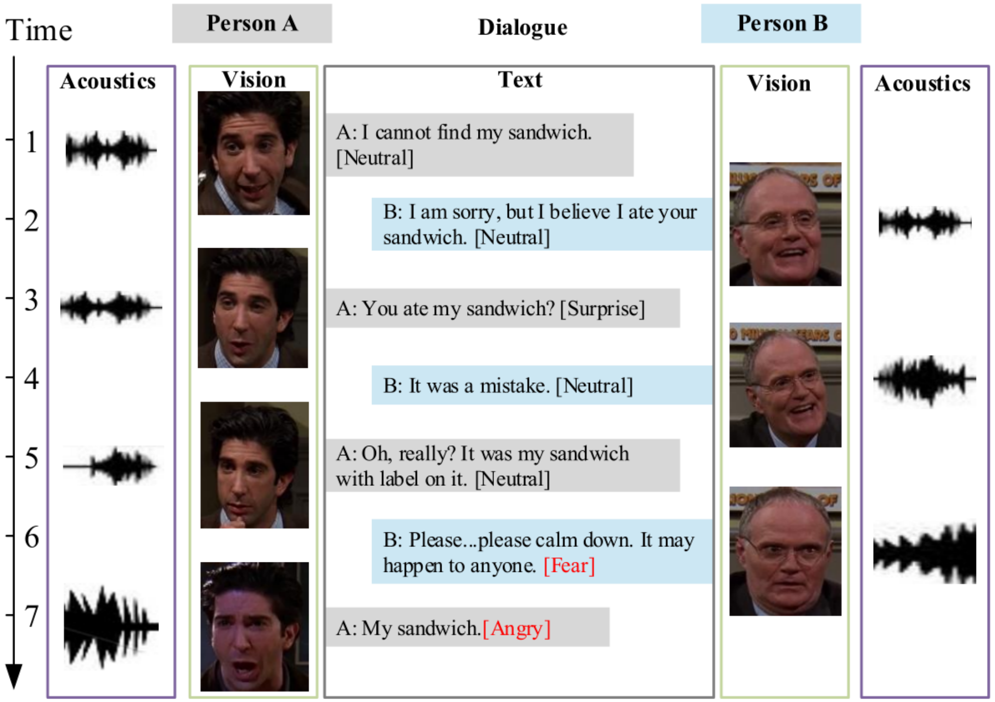
Role	Utterance	Truth	bcGRU	HiGRU-sf
Scene-4				
Ross	Hi.	Neu	Neu	Neu
Rachel	Hi.	Neu	Neu	Neu
Ross	Guess what?	Neu	Neu	Neu
Rachel	What?	Neu	Neu	Neu
Ross	They published my paper.	Joy	Sad	Neu
Rachel	Oh, really, let me see, let me see.	Joy	Neu	Neu
Phoebe	Rach, look! Oh, hi! Where is my strong Ross Sky-walker to come rescue me. There he is.	Other	/	/
Scene-5				
Speaker-1	Sorry for keeping you up	Sad	Sad	Sad
Speaker-2	Lol don't be	Joy	Joy	Joy
Speaker-2	I didn't have to get up today	Neu	Sad	Sad
Speaker-1	:p	Joy	Joy	Joy
Speaker-2	It's actually been a really lax day	Joy	Neu	Sad

Empathy Dialogue Systems

- HiGRU: Hierarchical Gated Recurrent Units for Utterance-level Emotion Recognition, NAACL’19, Joint work with CUHK
- DialogueTRM: Exploring the Intra- and Inter-Modal Emotional Behaviors in the Conversation (Under review)

Challenges of Emotion Recognition in Conversations (ERC)

- The emotional behaviors are not strictly synchronized either within or across modalities
  - Intra-modality: emotional expressions in different modalities exhibits different dependence on conversational context[1]
  - Inter-modality: emotional expressions in different modalities attain different contributions for emotional predictions.
- Existing methods either focus on **just** textual modality or assume all modalities have **the same** context preference.



Contributions

- Propose a DialogueTRM to explore the intra- and inter-modal emotional behaviors in ERC
  - a novel **Hierarchical Transformer (HT)** module for intra modal temporal modeling
  - a novel **Multi-Grained Interactive Fusion (MGIF)** module for multi-modal fusion through both neuron- and vector-grained interactive weighting **across** all modalities
- DialogueTRM achieves SOTA performance on three ERC benchmark datasets and outperforms SOTA fusion techniques in ERC settings

Problem Definition

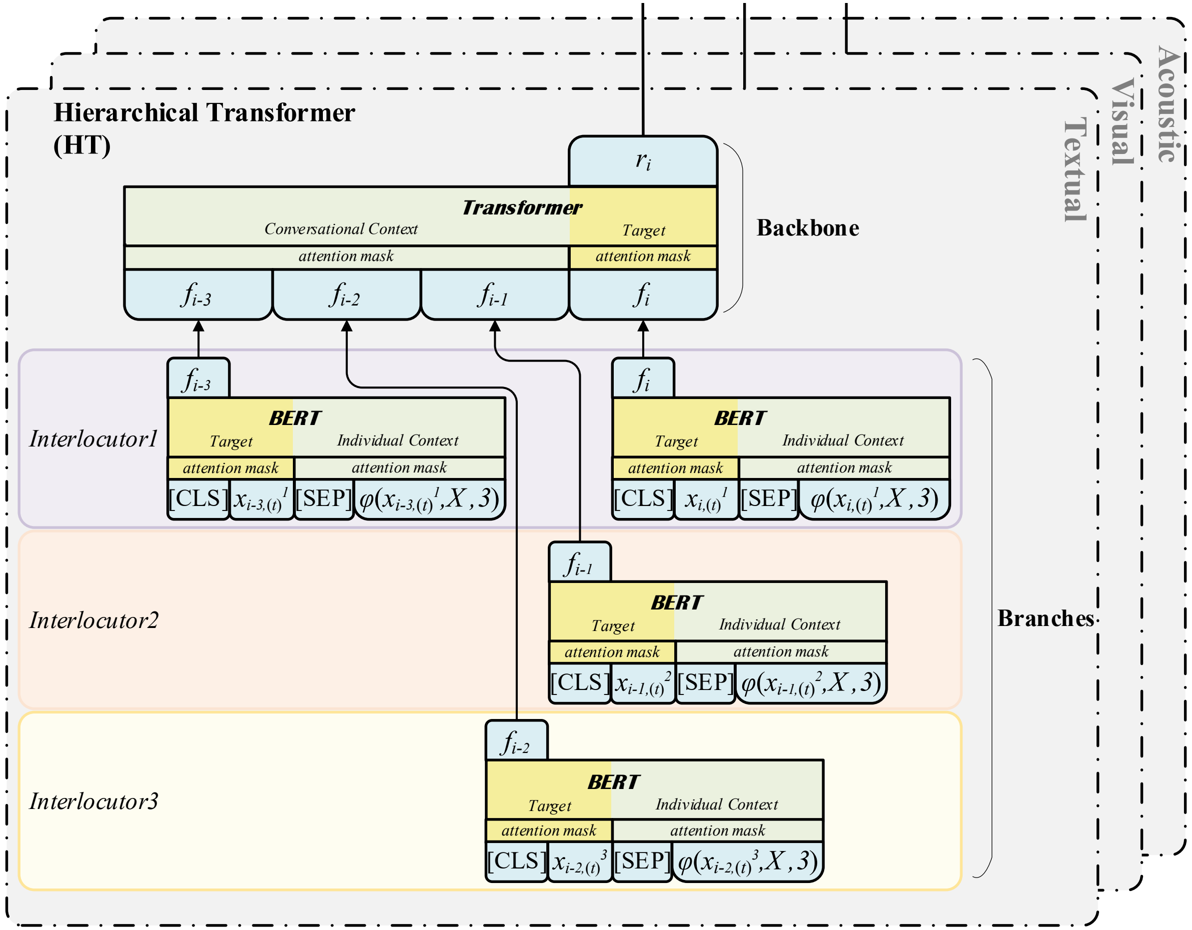
- Given a **conversation** of three modalities:
  - $X = \{x_\tau^\lambda = \langle x_{\tau,(t)}^\lambda, x_{\tau,(a)}^\lambda, x_{\tau,(v)}^\lambda \rangle | \tau \in [1, L], \lambda \in [1, N]\}$
    - x_τ^λ is the τ-th emotional expression produced by the λ-th interlocutor
    - t: textual, a: acoustic, v:visual
- x_i^j be the **target** emotional expression with two kinds of context,
  - **Individual context**:
    - $\varphi(x_i^j, X, K) = \{x_\tau^\lambda | \tau \in [i-K, i), \lambda = j\}$
  - **Conversational context**:
    - $\phi(x_i^j, X, K) = \{x_\tau^\lambda | \tau \in [i-K, i), \lambda \in [1, N]\}$

conversation	$X = \{x_1^1, x_2^1, x_3^2, x_4^1, x_5^3, x_6^2, x_7^1, x_8^2\}$
target	x_7^1
individual context	$\varphi(x_7^1, X, K) = \{x_4^1\}$
conversational context	$\phi(x_7^1, X, K) = \{x_4^1, x_5^3, x_6^2\}$

- Goal: train a model to predict the emotion of x_i^j given context information in multi-modal settings

Proposal

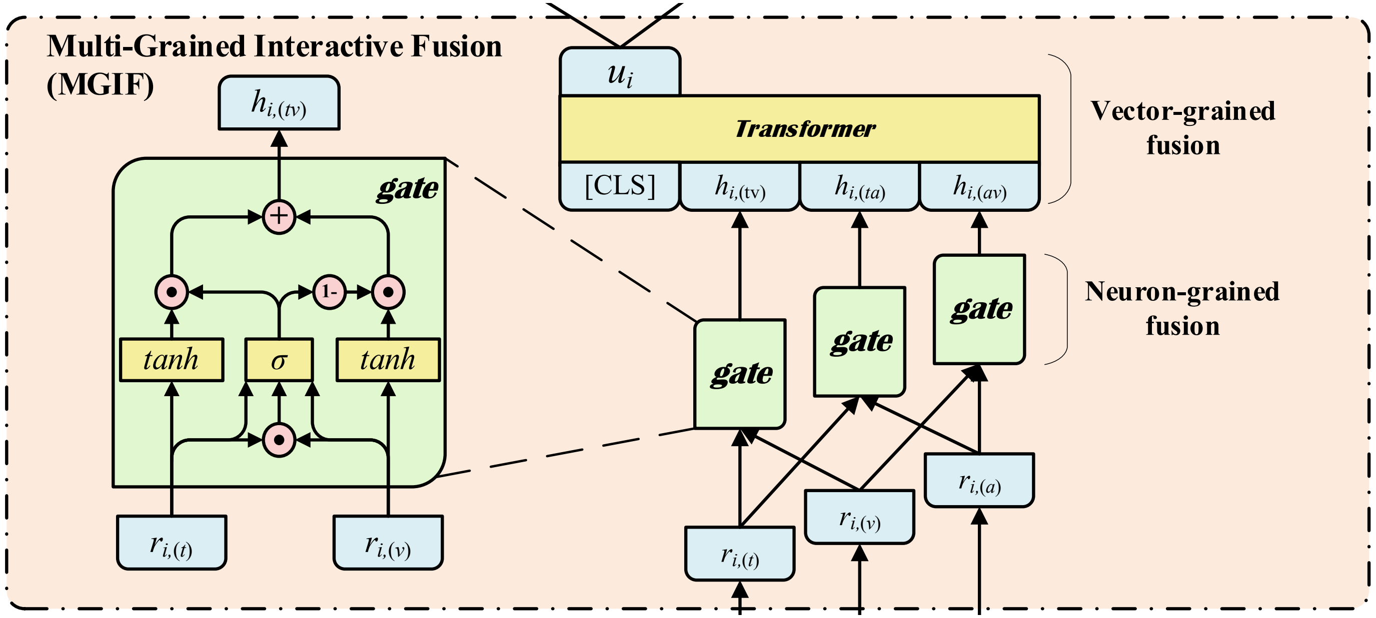
- Hierarchical Transformer:
  - BERT warps the individual context:
    - $f_i^j = \mathrm{BERT}\left([\mathrm{CLS}], x_i^j, [\mathrm{SEP}], \varphi(x_i^j, X, K)\right)$
  - Transformer wraps the conversational context:
    - $F = \{f_1, \dots, f_L\}$, index j is omitted
    - $r_i = \mathrm{TRM}(\phi(x_i^j, F, K), f_i)$



TRM : The encoder of TRansforMer

Proposal

- Multi-Grained Interactive Fusion:
  - Output from Hierarchical Transformer:
    - $r_i = \langle r_{i,(t)}, r_{i,(a)}, r_{i,(v)} \rangle$
  - Neuron-grained fusion:
    - Multi-modal gate (textual and visual):
      - $h_{i,(t)} = \tanh\left(W_t \cdot r_{i,(t)}\right)$
      - $h_{i,(v)} = \tanh\left(W_v \cdot r_{i,(v)}\right)$
      - $z = \sigma\left(W_z \cdot \left[r_{i,(t)}; r_{i,(v)}; r_{i,(t)} * r_{i,(v)}\right]\right)$
      - $h_{i,(tv)} = z * h_{i,(t)} + (1 - z) * h_{i,(v)}$
    - Simplified:
      - $h_{i,(tv)} = \mathrm{GATE}\left(r_{i,(t)}, r_{i,(v)}\right)$
      - $h_{i,(ta)} = \mathrm{GATE}\left(r_{i,(t)}, r_{i,(a)}\right)$
      - $h_{i,(av)} = \mathrm{GATE}\left(r_{i,(a)}, r_{i,(v)}\right)$
  - Vector-grained fusion:
    - $u_i = \mathrm{TRM}\left([\mathrm{CLS}], h_{i,(tv)}, h_{i,(ta)}, h_{i,(av)}\right)$

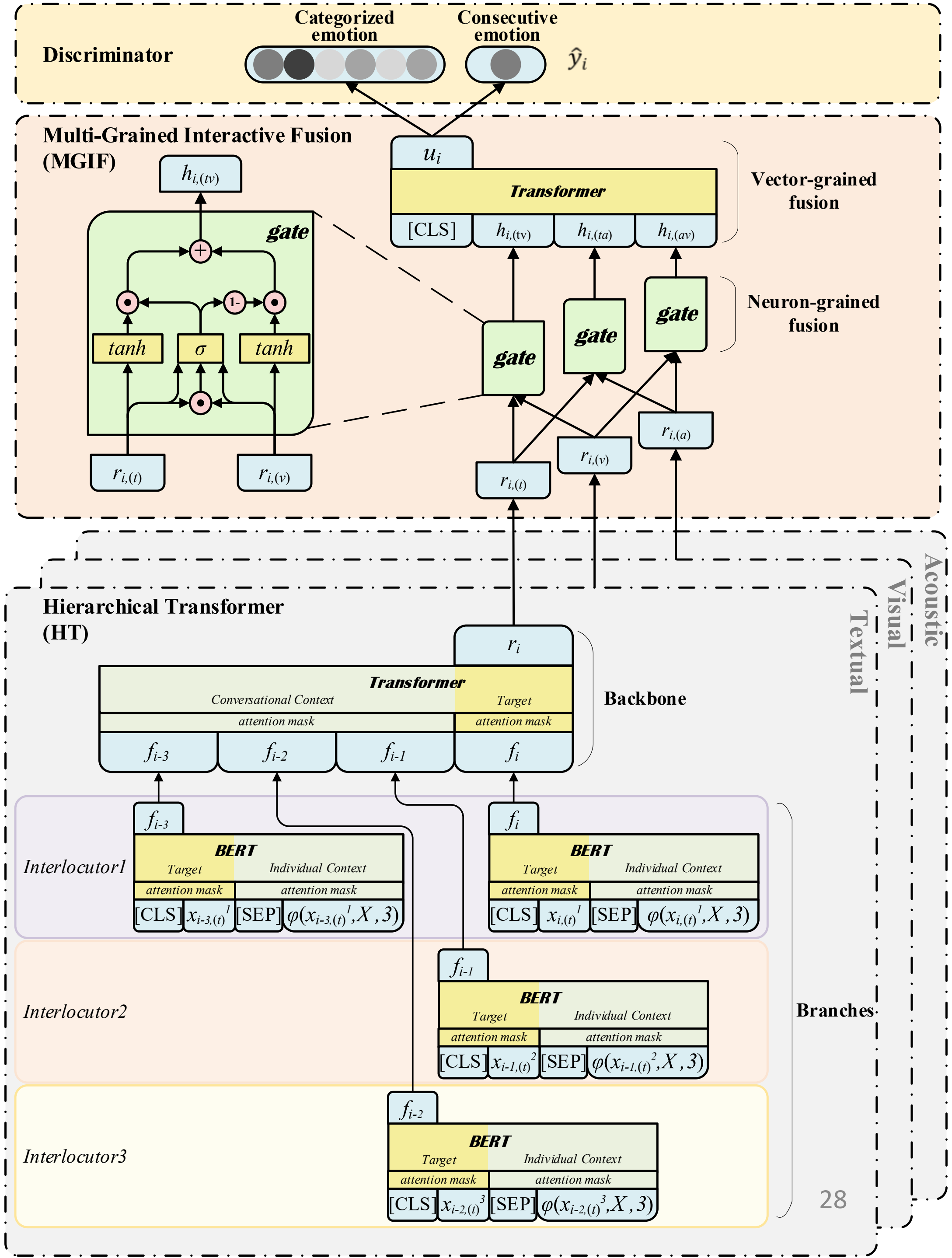


Proposal

- Discriminator:
  - $o_i = \tanh(W_l \cdot u_i)$
  - $\mathcal{P}_i = \begin{cases} softmax(W_{cat} \cdot o_i) & \text{categorized emotion} \\ \mathrm{W}_{con} \cdot o_i & \text{consecutive emotion} \end{cases}$
  - $\hat{y}_i = \begin{cases} \arg\max_k \mathcal{P}_i[k] & \text{categorized emotion} \\ \mathcal{P}_i & \text{consecutive emotion} \end{cases}$
- Loss function:
  - $Loss = \begin{cases} \frac{1}{L}\sum_i^L y_i \log(\hat{y}_i) & \text{categorized emotion} \\ \frac{1}{L}\sum_i^L (y_i - \hat{y}_i)^2 & \text{consecutive emotion} \end{cases}$

Consecutive emotion

- Valence: [-1,1]
- Arousal: [-1, 1]
- Expectancy: [-1, 1]
- Power: [0, ∞)



Dataset

- Categorized emotion:
  - **IEMOCAP**: six emotional types, including happy, sad, neutral, angry, excited, and frustrated.
  - **MELD**: seven emotional types, including anger, disgust, sadness, joy, neutral, surprise, and fear.
- Consecutive emotion:
  - **AVEC**: four emotional perceptions from visual and acoustic perspectives, which are Valence, Arousal, Expectancy, and Power.

dataset	conversations			utterances		
	train	val	test	train	val	test
IEMOCAP	120		31	5810		1632
AVEC	63		32	4368		1430
MELD	1039	114	280	9989	1109	2610

Main Results

Table 2: Categorized emotion results on IEMOCAP and MELD.

Model	IEMOCAP														MELD	
	happy		sad		neutral		angry		excited		frustrated		**Average**		**Average**	
	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1	ACC	F1
scLSTM	37.5	43.4	67.7	69.8	64.2	55.8	61.9	61.8	51.8	59.3	61.3	60.2	59.2	59.1	57.5	55.9
TL-ERC	-	-	-	-	-	-	-	-	-	-	-	-	-	58.8	-	-
DialogueRNN*	25.69	33.18	75.1	78.8	58.59	59.21	64.71	65.28	**80.27**	71.86	61.15	58.91	63.4	62.75	56.1	55.9
DialogueGCN*	40.62	42.75	**89.14**	**84.54**	61.92	63.54	67.53	64.19	65.46	63.08	64.18	66.99	65.25	64.18	-	58.1
AGHMN	48.3	**52.1**	68.3	73.3	61.6	58.4	57.5	61.9	68.1	69.7	67.1	62.3	63.5	63.5	59.5	57.5
BiERU*	**54.24**	31.53	80.6	84.21	64.67	60.17	**67.92**	65.65	62.79	**74.07**	61.93	61.27	66.11	64.65	-	60.84
DialogueTRM	54.19	48.7	71.04	77.52	**77.1**	**74.12**	64.74	**66.27**	63.91	70.24	**70.71**	**67.23**	**68.92**	**69.23**	**65.66**	**63.55**

Symbol * in both Table 2 and 3 indicate that models are fed with extra succeeding context.

Table 3: Consecutive emotion results on AVEC.

model	Valence	Arousal	Expectancy	Power
scLSTM	0.14	0.23	0.25	-0.04
TL-ERC	0.65	0.42	0.35	-0.03
DialogueRNN*	0.35	0.59	0.37	0.37
BiERU*	0.36	**0.64**	0.38	0.37
DialogueTRM	**0.756**	0.52	**0.4**	**0.4**

- DialogueTRM outperforms TL-ERC (BERT-based)
- DialogueTRM outperforms models using succeeding context, DialogueRNN, DialogueGCN, and BiERU
- DialogueTRM attains significant better R score in ``Valence'' and marginally better in “Expectancy” and “Power” than all baselines

Summary

- Two research work for empathy dialogue systems
  - a hierarchical Gated Recurrent Unit (HiGRU) framework for better representation learning
  - a DialogueTRM to explore the intra- and inter-modal emotional behaviors in ERC
- Experiments demonstrate the superior performance on the proposed models

Learning with Knowledge Bases

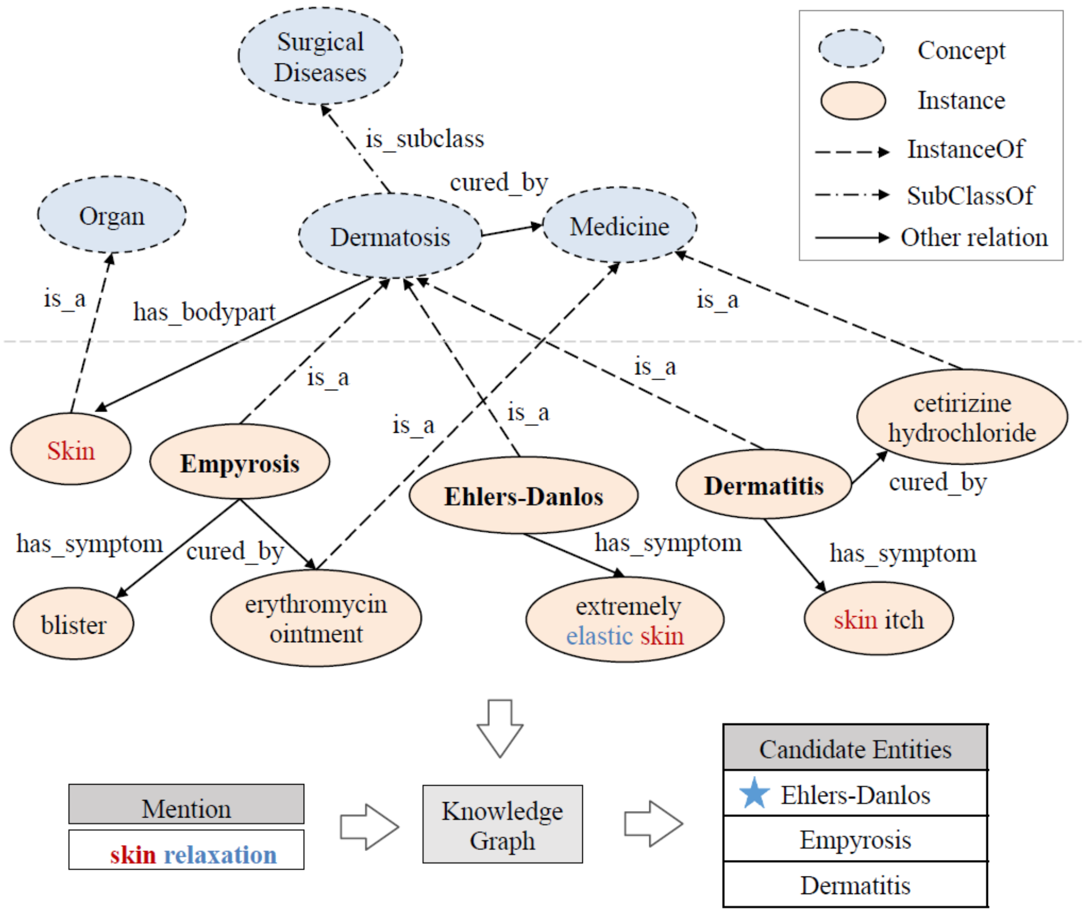
MedSynNet: A Novel Medical Entity Synonyms Discovery Framework

Motivation

- Medical entity synonyms discovery aims to accurately map the mention to synonymous medical entity in knowledge graphs

- Challenges
  - medical terms are often out-of-vocabulary (OOV)
  - different semantic spaces: oral vs. professional
  - some entities rarely appear in existing mention-entity pairs

- Existing methods via *matching of syntactic strings or lexical embeddings* **cannot** capture external knowledge well



Contribution

- A novel framework, MedSynNet, to
  - jointly learn both semantic features and knowledge representation of entities from knowledge graphs
  - craftily design fusion gate to enhance information interaction
  - demonstrate the effectiveness through experiments on both offline and online test

- The **first health insurance benchmark** consists of
  - a large-scale Chinese health insurance related knowledge graph: occupations, diseases, and insurance products
  - a dataset of annotated mention-to-entity pairs of diseases

Problem Definition: Medical entity synonyms discovery

- **Inputs**
  - mention-to-entity pairs $\mathcal{D} = \{(q_i, t_i)\}_{i=1}^{N}$
    - e.g., (bow legs, knee varus), (high blood pressure, hypertension), …
    - N is the number of annotated pairs
  - a cross-domain knowledge graph $\mathcal{KG} = \{\mathcal{C}, \mathcal{I}, \mathcal{R}, \mathcal{S}\}$
    - $\mathcal{C}$: Concept; $\mathcal{I}$: Instance
    - $\mathcal{R}$: relation set; $\mathcal{S}$: triple set
- **Output**
  - determine a list of synonymous entities for the disease mention
    - e.g., skin relaxation -> Ehlers-Danlos

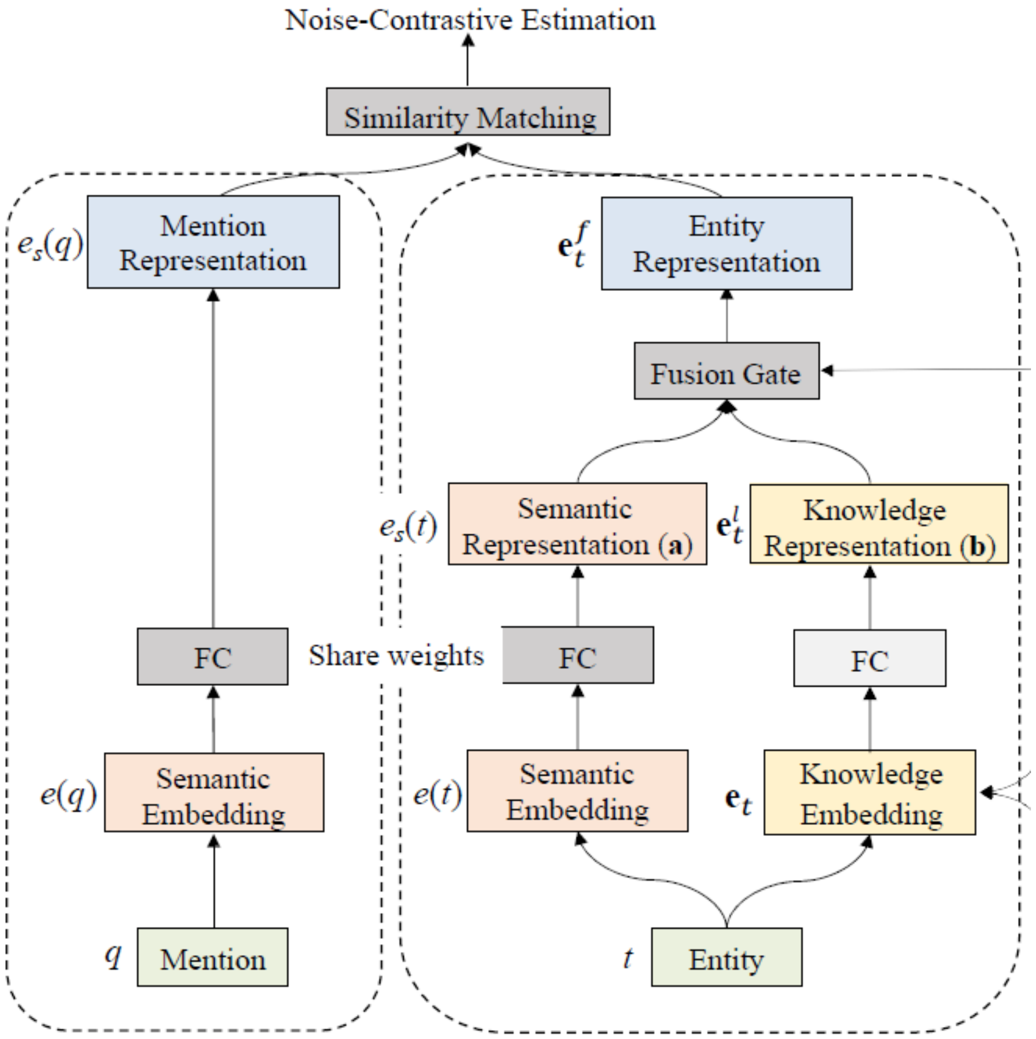
Our Proposal

1. Input representations
   - Char embeddings (semantic information)

$$e(q) = \frac{1}{|q|}\sum_{k=1}^{|q|} e(q_k), \quad e(t) = \frac{1}{|t|}\sum_{k=1}^{|t|} e(t_k).$$

   - Entity knowledge embedding
     - Joint TransC-TransE learning
2. Semantic space alignment
   - Shared weights in FC
3. Fusion of Entity's Semantic and Knowledge Representations
4. Similarity Matching:

   noise-contrastive estimation



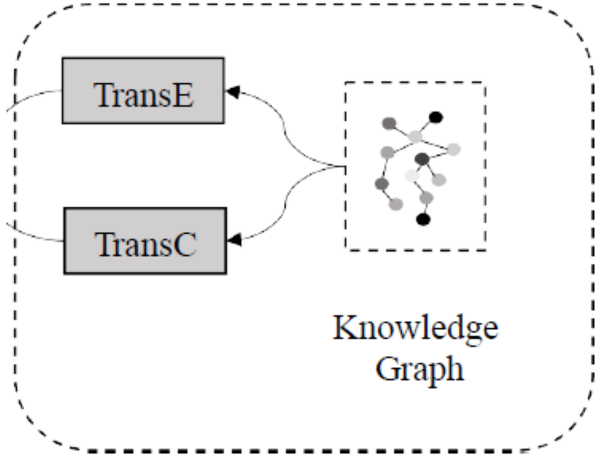
Properties: Adding External Knowledge

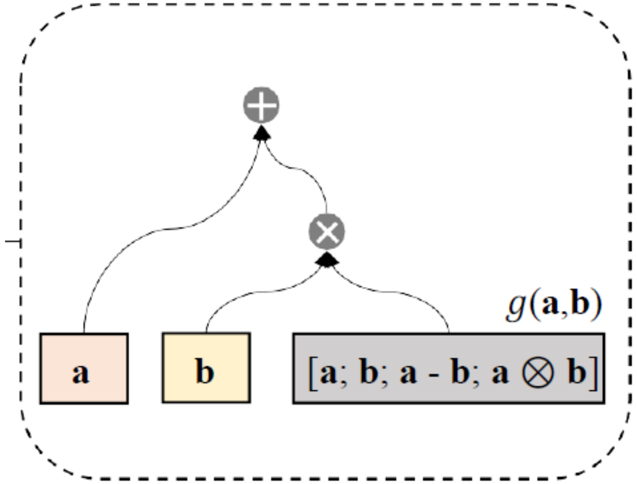
- Hybrid knowledge embedding
  - represent learning via joint TransC-TransE learning

$$\mathcal{L}_k = \sum_{(i,r_e,c)\in S_e} f_e(i,c) + \sum_{(c_i,r_c,c_j)\in S_c} f_c(c_i,c_j) + \sum_{(i,r_{ij},j)\in S_l} f_l(i,r_{ij},j)$$
$$+ \sum_{(i,r_{ic},c)\in S_{IC}} f_{IC}(i,r_{ic},c) + \sum_{(c_i,r_{c_ic_j},c_j)\in S_{CC}} f_{CC}(c_i,r_{c_ic_j},c_j). \quad (7)$$

- Fusion gate to adaptively incorporate knowledge with learned semantic features

$$\mathrm{e}_t^f = e_s(t) + \mathrm{e}_t^l \otimes g(e_s(t), \mathrm{e}_t^l)$$





Knowledge Graph (KG) and Dataset

- A heterogeneous KG with entities collected from three categories: **insurance products, occupation, and medicine**

Knowledge Graph	All	Insurance	Occupation	Medicine	Cross Domain
# Entities	75,153	1,409	2,587	71,157	0
# E_type	17	2	2	13	0
# Relations	1,120,792	2,827	2,580	1,098,280	17,105
# R_type	20	2	2	13	4
# Mention-entity pairs in Train/Dev/Test				45,500/5,896/5,743	
# Regular cases/# Difficult cases					5,303/440

- Difficult cases: no char overlap between mentions and entities
- Regular cases: at least one char overlap

Experimental Knowledge Graph (KG) and Dataset

- A heterogeneous KG with entities collected from three categories: **insurance products, occupation, and medicine**.

Knowledge Graph	All	Insurance	Occupation	Medicine	Cross Domain
# Entities	75,153	1,409	2,587	71,157	0
# E_type	17	2	2	13	0
# Relations	1,120,792	2,827	2,580	1,098,280	17,105
# R_type	20	2	2	13	4
# Mention-entity pairs in Train/Dev/Test					45,500/5,896/5,743
# Regular cases/# Difficult cases					5,303/440

- Difficult cases: no char overlap between mentions and entities
- Regular cases: at least one char overlap

Results

- Difficult: no char overlap between mentions and entities
- Regular: at least one char overlap
- All: Difficult + Regular

Methods	hits@3			hits@5			hits@10		
	All	Regular	Difficult	All	Regular	Difficult	All	Regular	Difficult
JACCARD [26]	52.28%	56.61%	0.00%	58.03%	62.83%	0.00%	63.76%	69.04%	0.00%
Word2Vec [4]	47.00%	50.88%	0.00%	52.28%	56.59%	2.30%	58.31%	63.10%	4.60%
CNN [24]	51.76%	55.69%	4.33%	57.75%	61.98%	6.38%	65.13%	69.72%	9.34%
BERT [6]	54.60%	58.87%	2.96%	60.41%	65.02%	4.78%	66.50%	71.39%	7.52%
DNorm [17]	56.23%	59.78%	12.76%	63.79%	67.58%	17.77%	71.89%	75.64%	26.42%
SurfCon [36]	58.29%	62.02%	12.98%	66.27%	70.11%	19.59%	75.20%	79.03%	28.93%
MedSynNet	**66.84%**	**70.81%**	18.91%	**73.09%**	**77.13%**	24.37%	**79.41%**	**83.35%**	31.89%
−Knowledge Embedding	64.91%	69.07%	14.58%	71.56%	75.77%	20.73%	79.12%	83.14%	30.52%
−TransC (TransE only)	65.80%	69.92%	15.95%	71.44%	75.79%	18.91%	78.94%	83.18%	27.80%
→Direct Addition	63.51%	67.19%	**19.13%**	70.85%	74.47%	**27.10%**	78.13%	81.77%	**34.17%**
→Ernie Fusion	61.98%	65.85%	15.26%	68.63%	72.54%	21.41%	76.28%	80.29%	27.79%

- MedSynNet beats all baselines
- Knowledge embedding plays a significant role in improving performance

Some Examples

Alias	Top 5 Entities
弹力过度性皮肤	埃莱尔-当洛综合症, 埃勒斯-当洛斯综合症, 皮肤松垂, 埃莱尔-当洛, 皮肤松弛，
cutis hyperelastica	Ehlers-Danglos syndrome, Ehlers-Danglos syndrome, dermatolysis, Ehlers-Danglos, cutis laxa
肚子疼	腹痛, 疼痛, 下腹痛, 全身疼痛, 疼痛性脂肪过多症
collywobbles	stomachache, pain, hypogastralgia, generalized pain, lipomatosis dolorosa
歪嘴风	面瘫, 面神经麻痹, 新生儿面部神经麻痹, 周围性面瘫, 特发性面神经瘫痪
facioplegia	prosopoplegia, facial paralysis, neonatal facial paralysis, peripheral facial paralysis, idiopathic facial paralysis
倦怠	虚弱, 乏力, 张力失常, 失眠症, 弱精
languor	asthenia,asthenia, dystonia, insomnia, asthenozoospermia

- facioplegia -> prosopolegia, facial paralysis
- “neonatal facial paralysis", “peripheral facial paralysis", and “idiopathic facial paralysis" are all hyponyms of the mention with specific clinical manifestations

Summary

- MedSynNet to tackle the task of medical entity synonyms discovery in
  - resolving OOV by newly learned character embeddings and map them into the same space
  - learning knowledge information by TransC-TransE model
  - designing a fusion gate to adaptively include knowledge information into the semantic features
- The **first health insurance benchmark** consists of
  - a large-scale Chinese health insurance related knowledge graph in multiple domains
  - a dataset of annotated mention-to-entity pairs of diseases
- Future work in
  - extending to other domains, e.g., education or justice
  - enriching the knowledge graph to improve the discovering ability
  - tackling the task of entity hypernyms/hyponyms discovery

Conclusion

- Briefly review AI technologies development and applications in Ping An Life

- Present some research work on
  - emotion recognition in dialogues
  - medical entity synonyms discovery in knowledge graphs

- Many potential applications and research problems exist …

https://iconip2020.apnns.org/

AI Department

Ping An Life

AI Empowers Insurance Services

hqyang.github.io

hqyang@ieee.org