Relation-aware Graph Attention Networks with Relational Position Encodings for Emotion Recognition in Conversations

Softskills Seminar - M2 Data & AI - Institut Polytechnique de Paris

Thomas Wimmer, 11.01.2023

Ishiwatari, Taichi, et al. "Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations." *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020.

What will we do in the next 20 minutes?

- What is the problem solved in this work?
- Which work does this paper build on?
- What are the most important contributions?
- Which experiments were carried out? What are the results?
- Is it a valuable contribution to the field?
- What are the key findings?

Emotion Recognition in a nutshell

Anna

Bernard

It's such a lovely day today. How are you doing?

I'm devastated because France didn't win the World Cup.

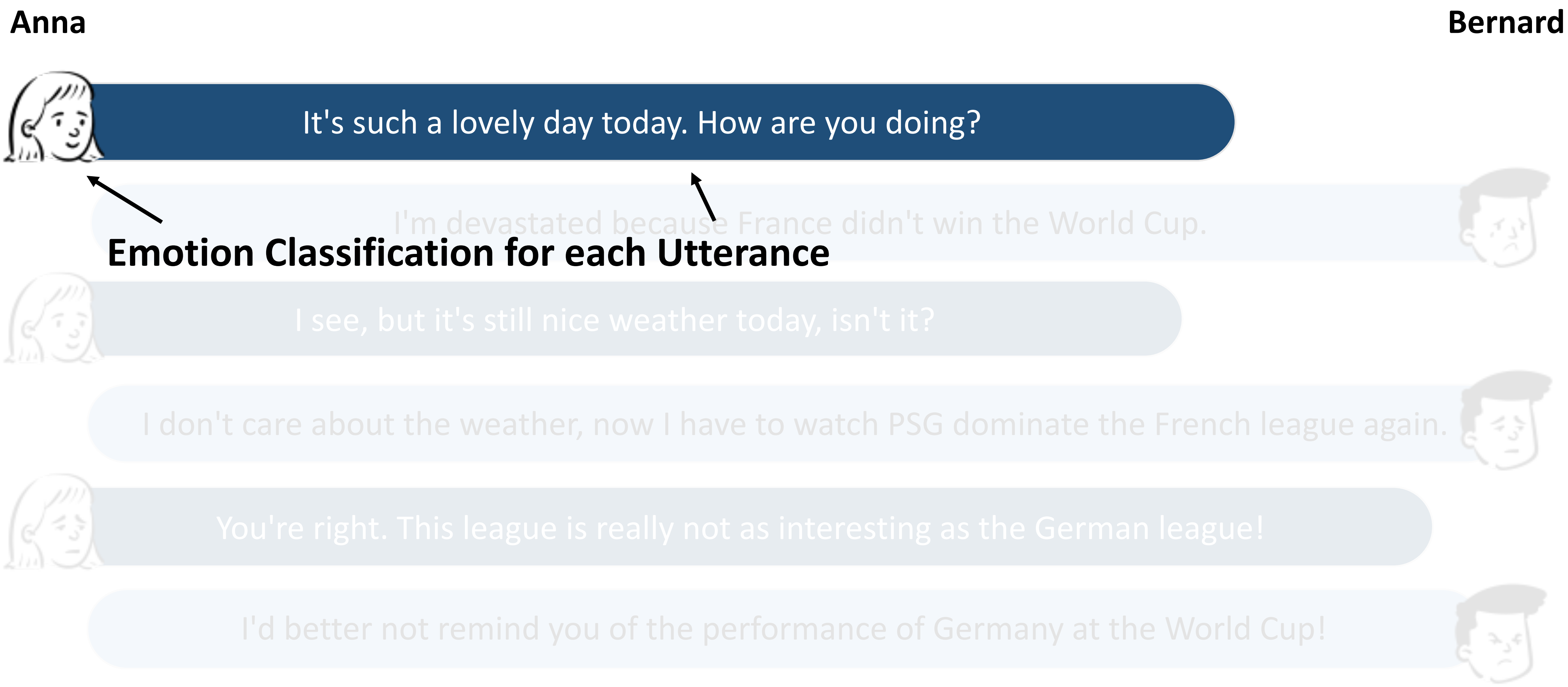
I see, but it's still nice weather today, isn't it?

I don't care about the weather, now I have to watch PSG dominate the French league again.

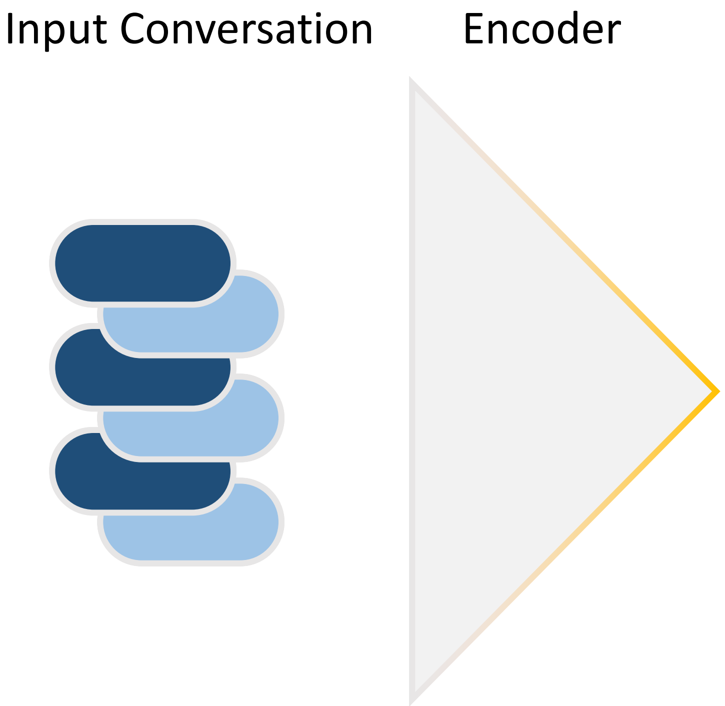
You're right. This league is really not as interesting as the German league!

I'd better not remind you of the performance of Germany at the World Cup!

Emotion Recognition in a nutshell

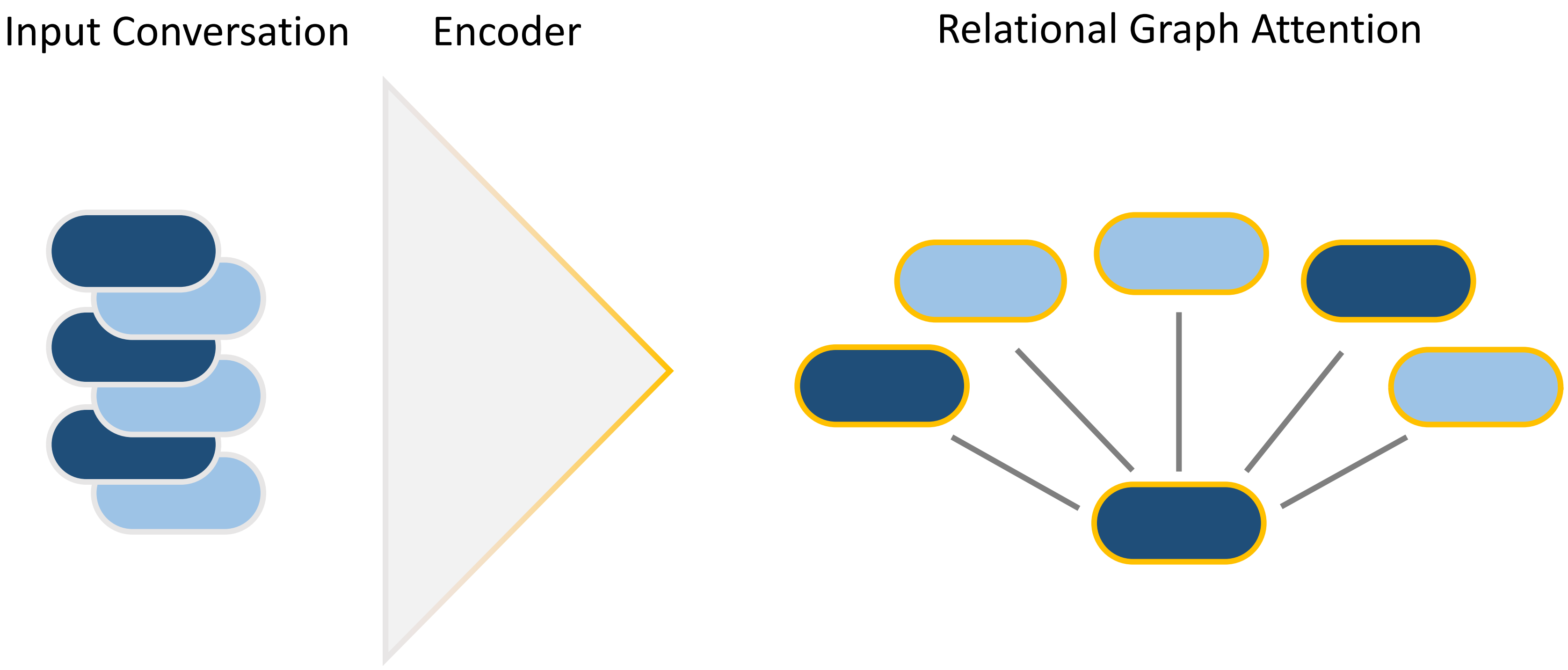


Basic structure of the architecture



Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. **DialogueGCN**: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

Basic structure of the architecture



Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. **DialogueGCN**: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

Emotion Recognition in a nutshell

Anna **Bernard**


It's such a lovely day today. How are you doing?

I'm devastated because France didn't win the World Cup.



I see, but it's still nice weather today, isn't it?

I don't care about the weather, now I have to watch PSG dominate the French league again.



You're right. This league is really not as interesting as the German league!

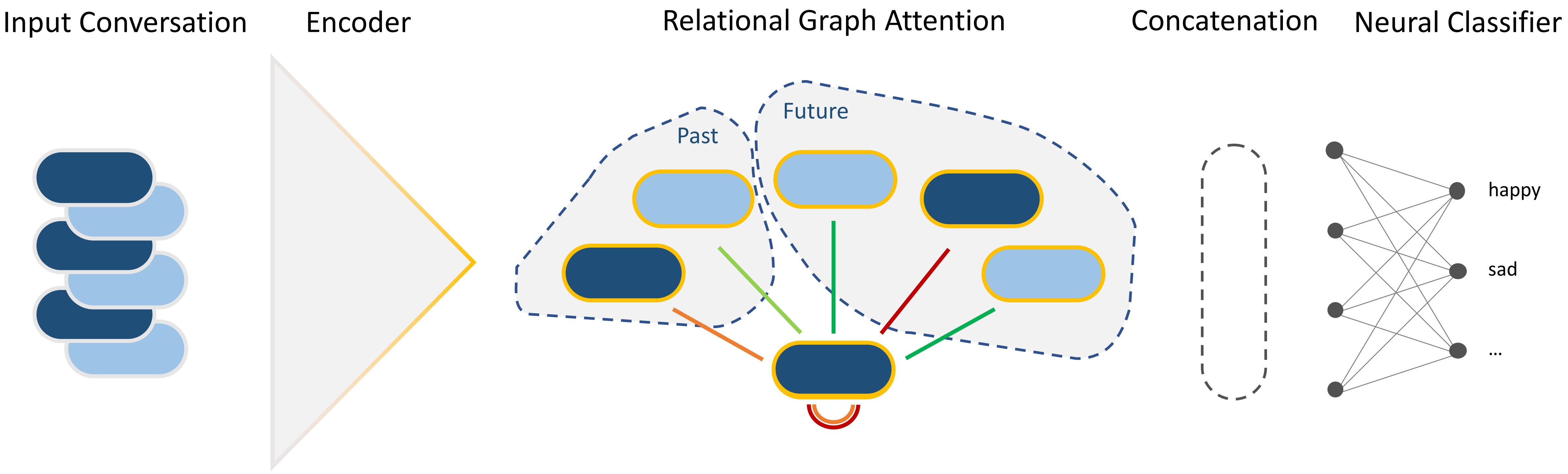
I'd better not remind you of the performance of Germany at the World Cup!


Emotion Recognition in a nutshell

Anna **Bernard**

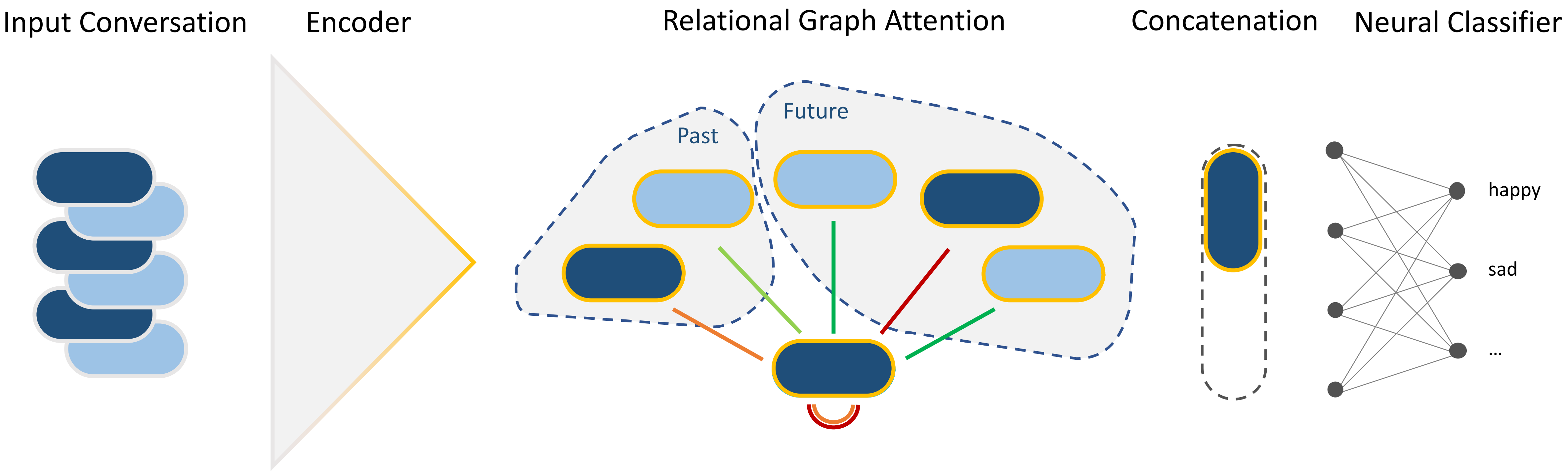
Emotions from other time steps (past or future), both of the current speaker and of other speakers, influence the emotion recognition result for a particular utterance.

Basic structure of the architecture



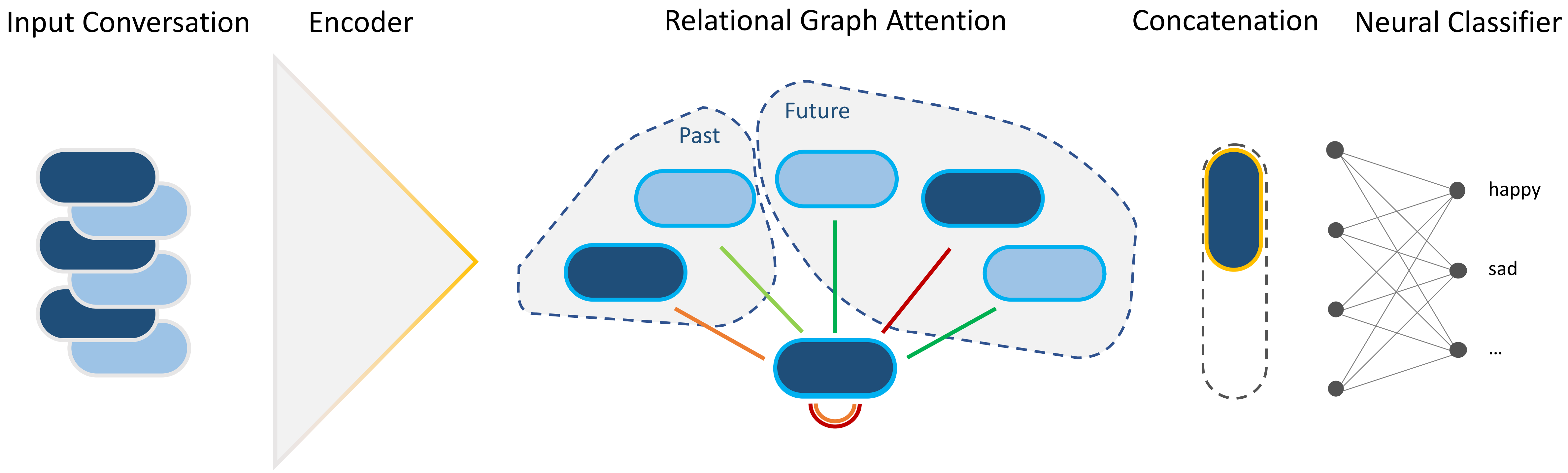
Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

Basic structure of the architecture



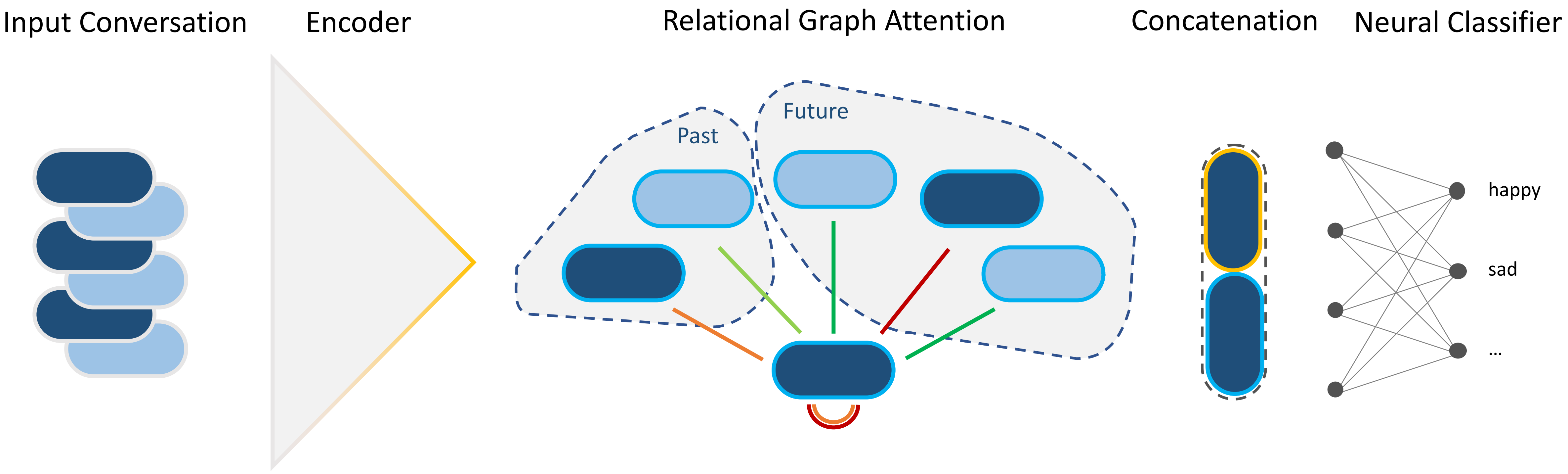
Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

Basic structure of the architecture



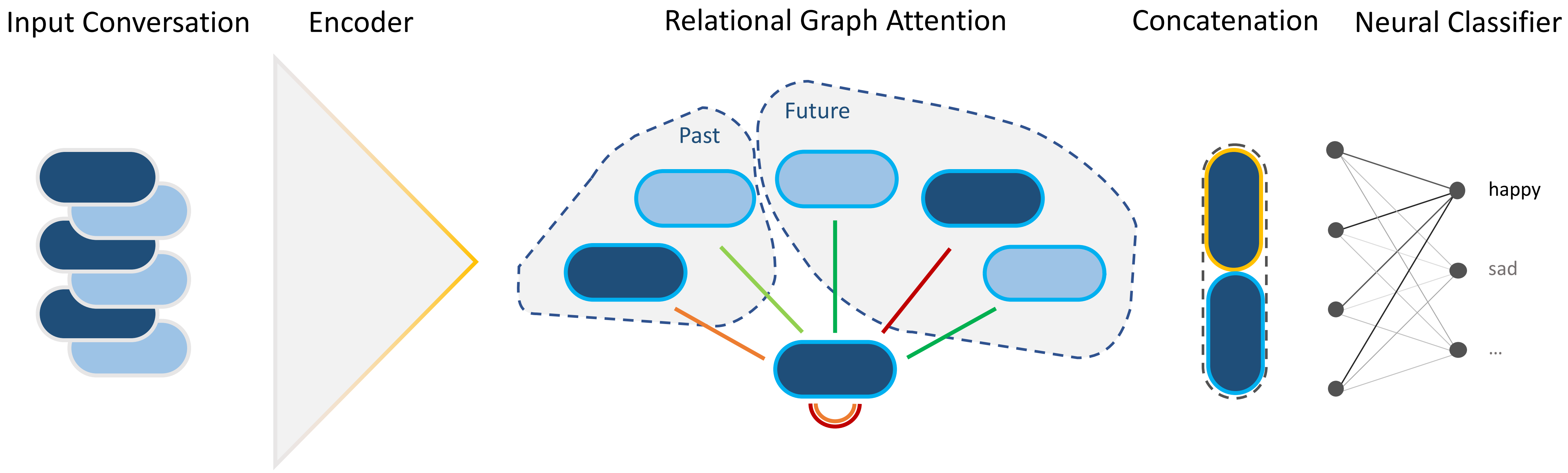
Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

Basic structure of the architecture



Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

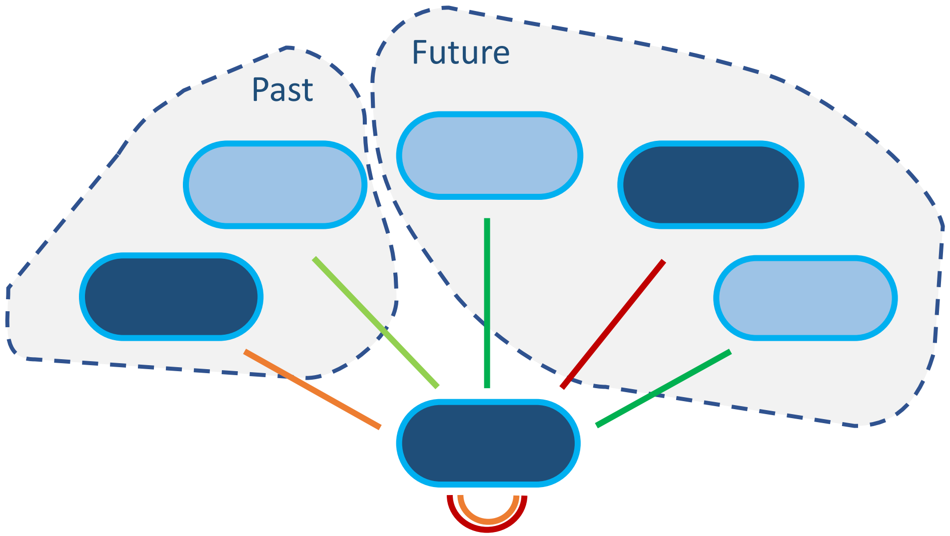
Basic structure of the architecture



Architecture is end-to-end trainable!

Deepanway Ghosal, Navonil Majumder, Soujanya Poria, Niyati Chhaya, and Alexander Gelbukh. 2019. DialogueGCN: A Graph Convolutional Neural Network for Emotion Recognition in Conversation. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 154–164, Hong Kong, China. Association for Computational Linguistics.

Relational Graph Attention

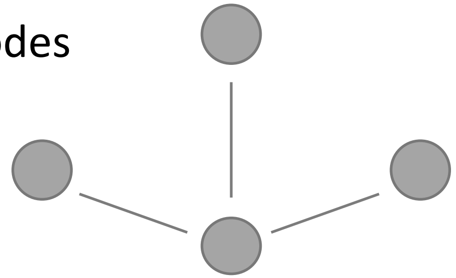


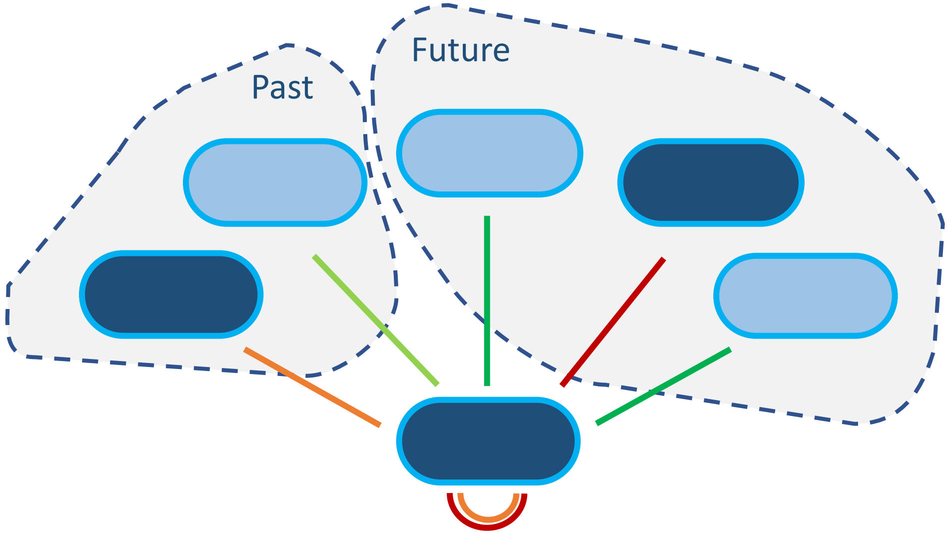
Relational Graph Attention

Graph Neural Network Basics

Message Passing Network

1. Process nodes
2. Aggregate neighbourhood nodes
3. Update nodes



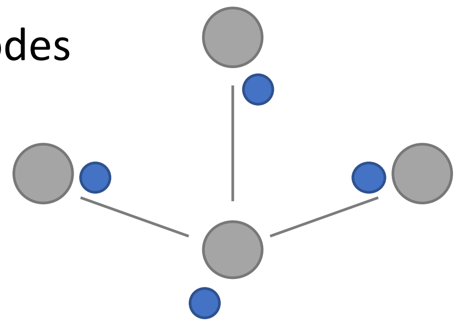


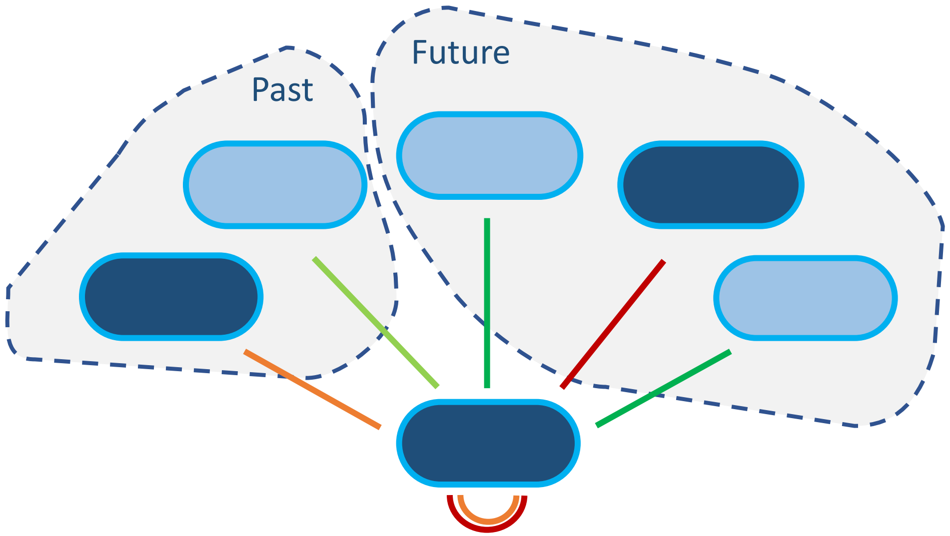
Relational Graph Attention

Graph Neural Network Basics

Message Passing Network

1. **Process nodes**
2. Aggregate neighbourhood nodes
3. Update nodes



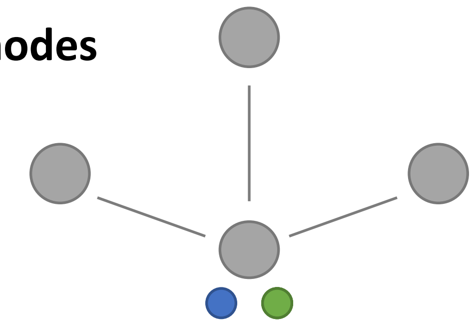


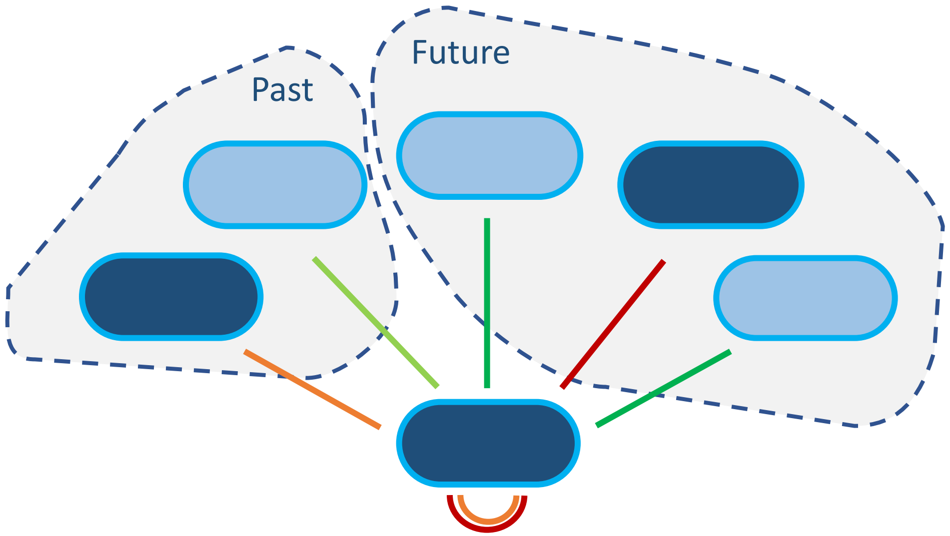
Relational Graph Attention

Graph Neural Network Basics

Message Passing Network

1. Process nodes
2. **Aggregate neighbourhood nodes**
3. Update nodes

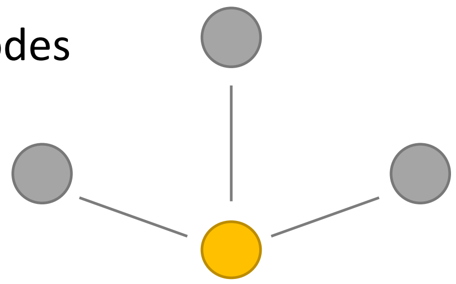


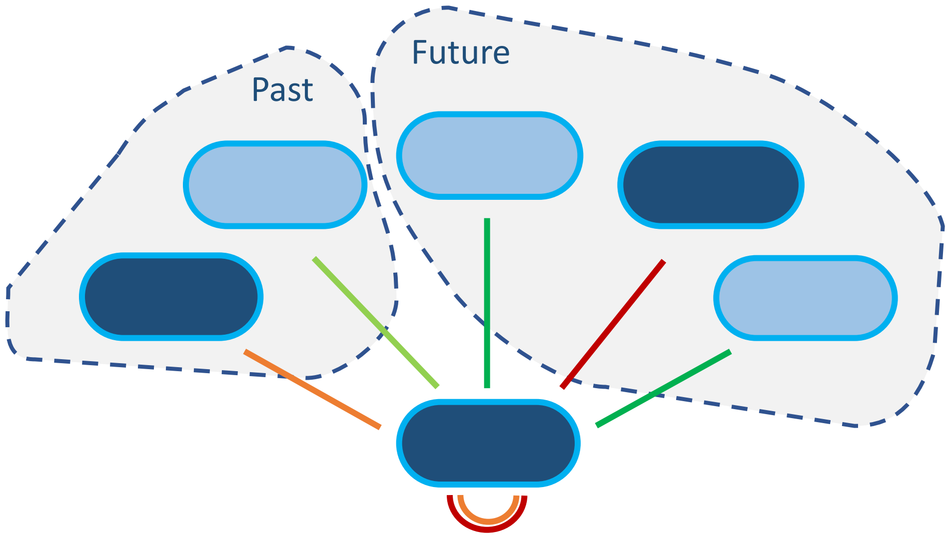


Relational Graph Attention

Graph Neural Network Basics

Message Passing Network
1. Process nodes
2. Aggregate neighbourhood nodes
3. **Update nodes**



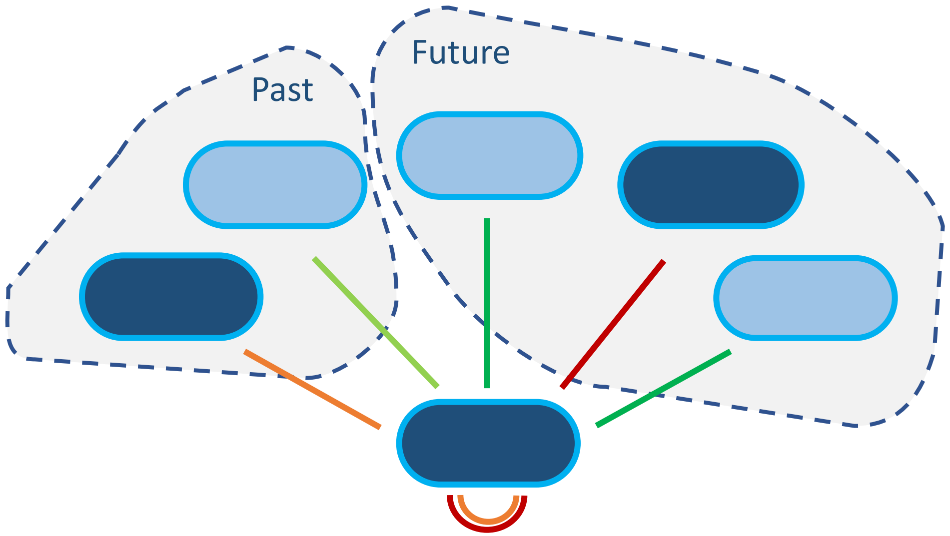


Active research field:
Many more variants, like e.g., **attention**, convolutions, gated units, …

Relational Graph Attention

Relation types:

	Speakers dependency	
Temporal dependency	self - past	inter - past
	self - future	inter - future



Attention independent for each relation type r:

$$\alpha_{ijr} = softmax_i\left(LeakyReLU\left(a_r^T[W_r h_i || W_r h_j]\right)\right)$$

Propagation of information through the graph:

$$h_{ir}^{(l-1)} = \sum_{j \in \mathcal{N}^r(i)} \alpha_{ijr}^{(l-1)} W_r^{(l-1)} h_j^{(l-1)}$$

$$h_i^{(l)} = \sum_{r=1}^{R} h_{ir}^{(l-1)}$$

Relational Graph Attention

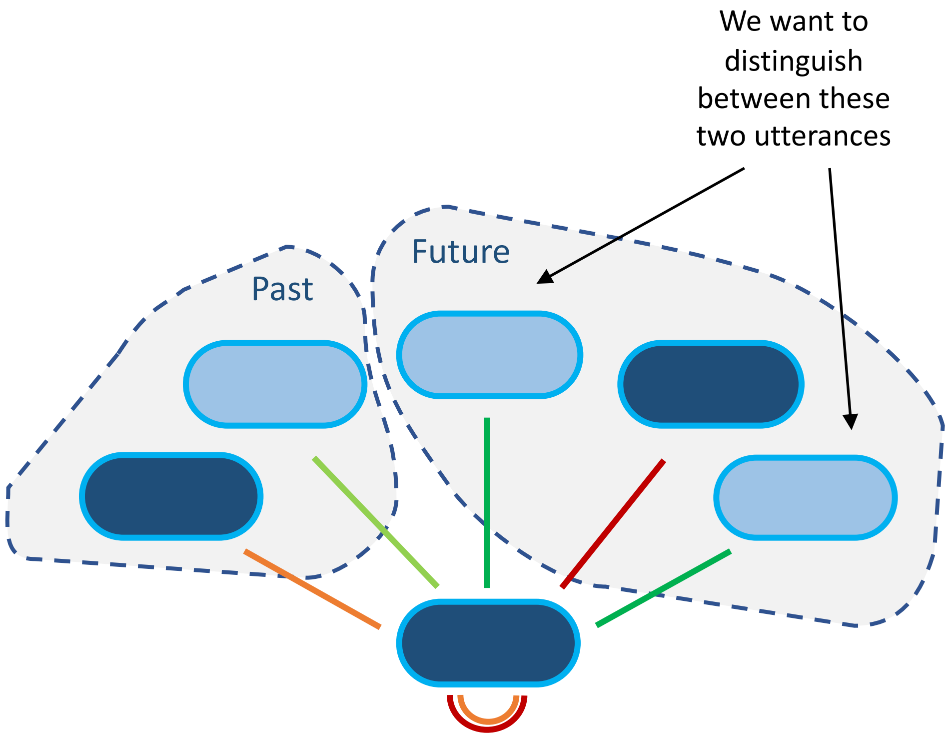
Relational position encodings:

Absolute Position	1	2	3	4	5	6
Relative Position	-2	-1	0	1	2	3
Relational Position	-1	-1	0	-1	-1	-2

+ learned positional embedding with relational position as input

Attention with added positional encoding PE_{ijr}:

$$\alpha_{ijr} = softmax_i\left(LeakyReLU\left(a_r^T\left[W_r h_i || W_r h_j\right]\right) + PE_{ijr}\right)$$



Experiments

Evaluation:

F1 score (micro-averaged / weighted-averaged as used in literature)

Datasets:

IEMOCAP, MELD, EmoryNLP, DailyDialog

https://giphy.com/friends/reactions/excited

Experiments

Results

Proposed model achieves SOTA performance on all benchmark datasets

Models	IEMOCAP	MELD	EmoryNLP	DailyDialog
CNN	48.18	55.86	32.59	49.34
CNN+cLSTM	54.95	56.87	32.89	50.24
BERT_BASE	53.31	56.21	33.15	53.12
KET	59.56	58.18	34.39	53.37
DialogueRNN	62.75	57.03	31.70	50.65
DialogueGCN	64.18	58.10	-	-
Ours	**65.22**	**60.91**	**34.42**	**54.31**

Ishiwatari, Taichi, et al. "Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.

Experiments

Influence of Positional Encoding

- BERT encoder contributes to better performance
- Proposed method is well-balanced over all categories

♯	Models	Background Components: Contextual Utterance Embedding	Background Components: Speaker Dependency Modeling	Happy	Sad	Neutral	Angry	Excited	Frustrated	Average
0	BERT_BASE	BERT	×	37.09	59.53	51.73	54.33	54.26	55.83	53.31
1	DialogueRNN	CNN, GRU		33.18	78.80	59.21	**65.28**	**71.86**	58.91	62.75
2	DialogueGCN	CNN, GRU	RGAT	42.75	**84.54**	63.54	64.19	63.08	**66.99**	64.18
3	Ours(without PE)	BERT	RGAT	50.69	76.78	**65.85**	59.66	64.04	62.37	64.36
4	Ours	BERT	RGAT with PE	**51.62**	77.32	65.42	63.01	67.95	61.23	**65.22**

Ishiwatari, Taichi, et al. "Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.

Assessment of the paper

- **State-of-the-art results** ☺
- Thought-provoking paper, generally well-written

- Inconsistencies and missing formulas / explanations
- **No published implementation**
- Best performing configurations (even architectural choices) not shared
- Work is mainly based on DialogueGCN
- Experiments could go even further

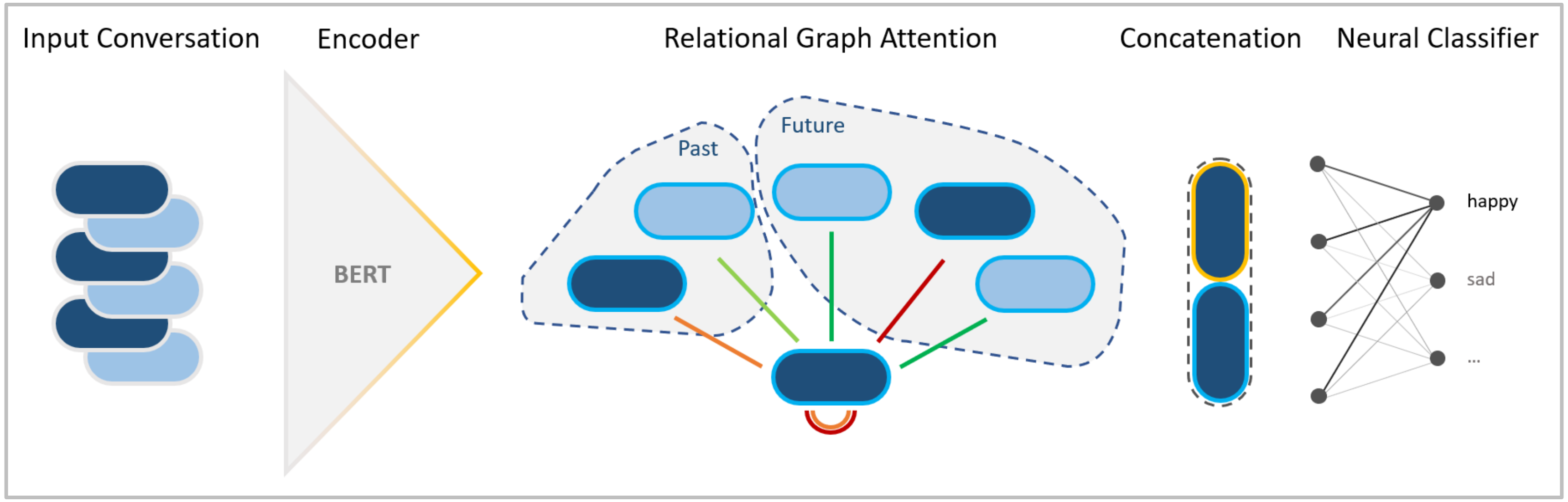
Ideas for Future Work

- Analyze the positional encodings by visualizing the learned 1D (!) function
- Try positional encodings that are multi-dimensional (not just a scalar value)
- Include positional encodings not just by adding it to the attention weights

- Try differentiating between different speakers (instead of throwing everyone in the "inter-speaker dependency" bucket)
- Use attention (or simple learned weights) on the different propagation states of the GNN when using it as input to the classification network

Summary

A new state-of-the-art model for **emotion recognition in conversations** introducing relational position encodings



Relation types:

	Speakers dependency	
Temporal dependency	self - past	inter - past
	self - future	inter - future

Relational position encodings:

Absolute Position	1	2	3	4	5	6
Relative Position	-2	-1	0	1	2	3
Relational Position	-1	-1	0	-1	-1	-2

+ learned positional embedding with relational position as input

Backup Slides

Experiments

Study on best positional encoding

- Absolute position (node-based)
- Relative position (edge-based)
- Relational (proposed method)

Learned encoding always performs better than fixed

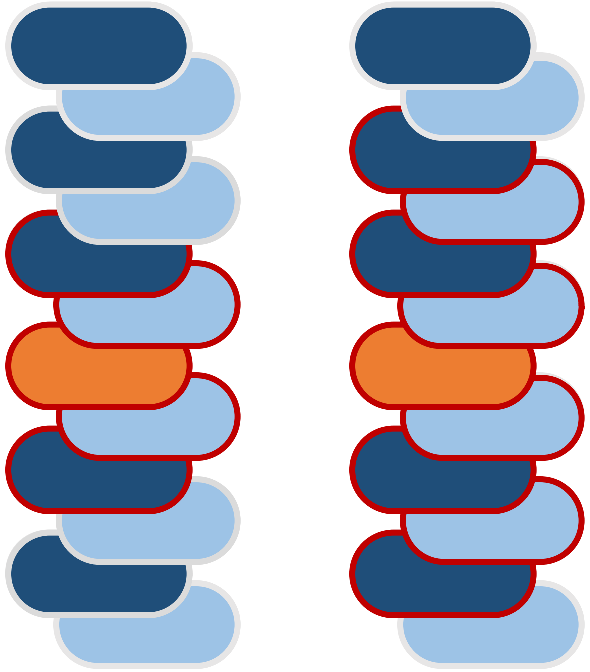
♯	Position Encodings (PE)	Type	Average
0	-	-	64.36
1	Node-based PE	*fixed*	63.95
2		*learn*	64.95
3	Edge-based PE	*fixed*	63.97
4		*learn*	64.59
5	Relational PE	*fixed*	63.99
6		*learn*	**65.22**

Table 5: Impact of various position encodings components on the IEMOCAP dataset. The base model using BERT and RGAT without position encodings is shown in ♯0. "*fixed*" and "*learn*" denote a fixed function and a learned representation respectively.

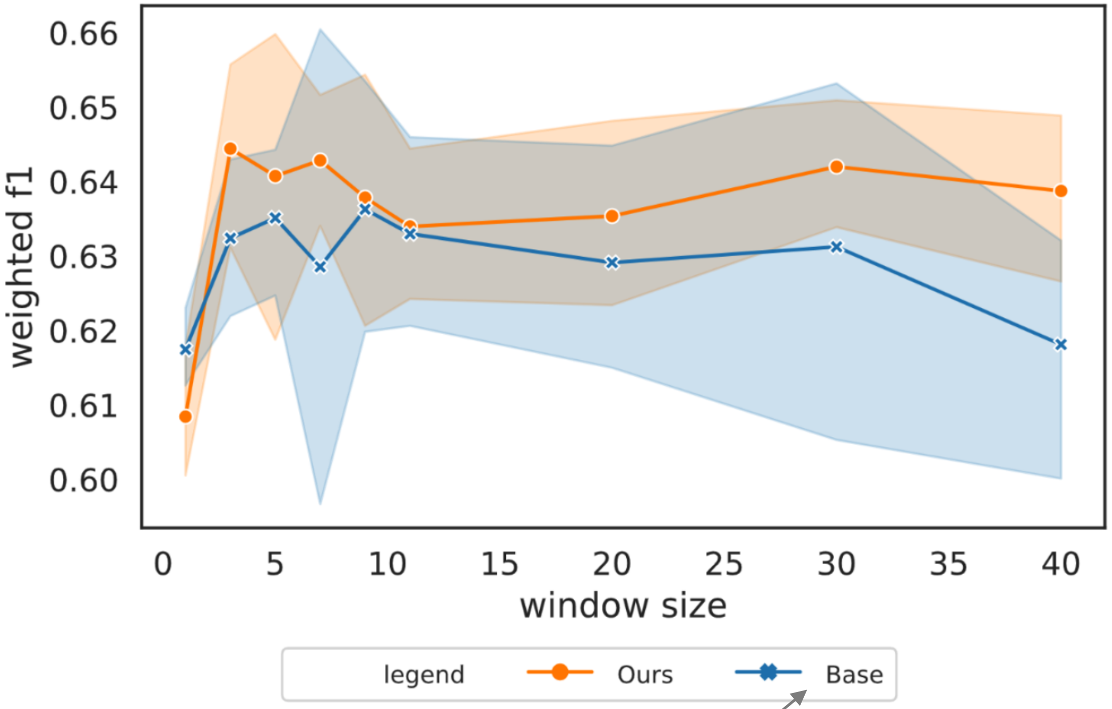
Ishiwatari, Taichi, et al. "Relation-aware graph attention networks with relational position encodings for emotion recognition in conversations." Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020.

Experiments

Influence of different context window sizes



K = 1 K = 2



Same architecture but without positional encoding