Knowledge Aware Dialogue Generation

Hao Zhou

Tsinghua University, Beijing, China

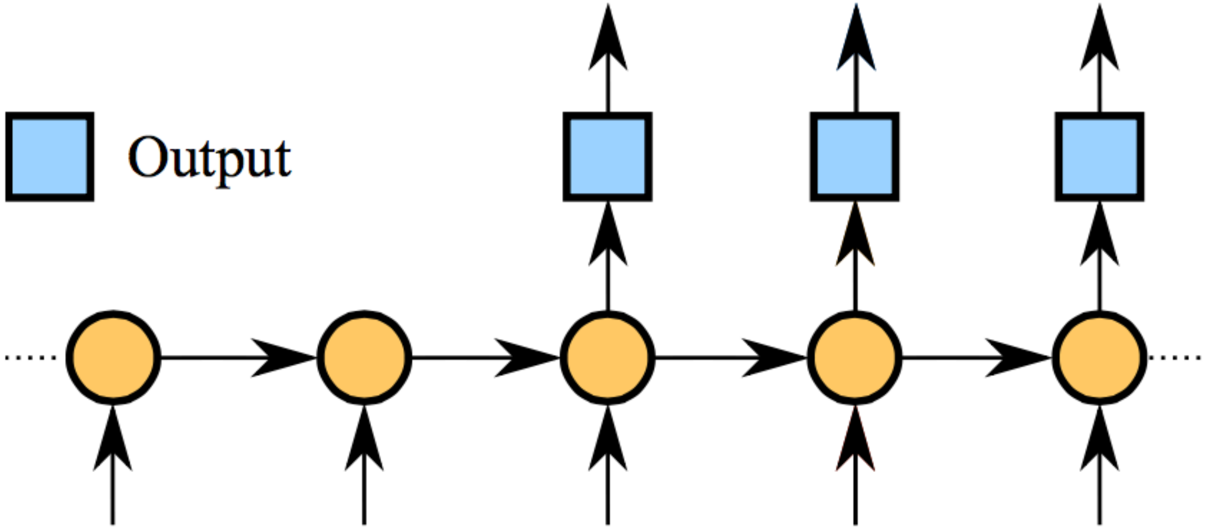
Outline

- Introduction
- Commonsense Knowledge Aware Conversation Generation with Graph Attention
- Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory
- Summary

Seq2Seq

RNN Unit:

$$
\begin{aligned}
h_t &= \operatorname{sigm}\left(W^{\text{hx}} x_t + W^{\text{hh}} h_{t-1}\right) \\
y_t &= W^{\text{yh}} h_t
\end{aligned}
$$



Goal:

$$p(y_1, \ldots, y_{T'} | x_1, \ldots, x_T) = \prod_{t=1}^{T'} p(y_t | v, y_1, \ldots, y_{t-1})$$

Seq2Seq

- Machine Translation
  - Source language sequence to target language sequence
  - Attention strengthen the key information because the corresponding words and phrase between two languages
  - Statistical translated sentence
- Dialogue System
  - Post sequence to response sequence in the same language
  - Attention strengthen the context information for generating response
  - Relevant and grammatical sentence

Difference

- Machine Translation
  - Certainty with the same meaning
  - Context-free
  - Statistical without understanding
  - No knowledge
  - Easy to evaluate

- Dialogue System
  - Uncertainty with diverse meanings
  - Context-dependent
  - Understanding is critical in some cases
  - Need knowledge
  - Hard to evaluate

Commonsense Knowledge Aware Conversation Generation with Graph Attention

Hao Zhou[1], Tom Young[2], Minlie Huang[1*], Haizhou Zhao[3], Jingfang Xu[3], Xiaoyan Zhu[1]

[1]Tsinghua University, Beijing, China

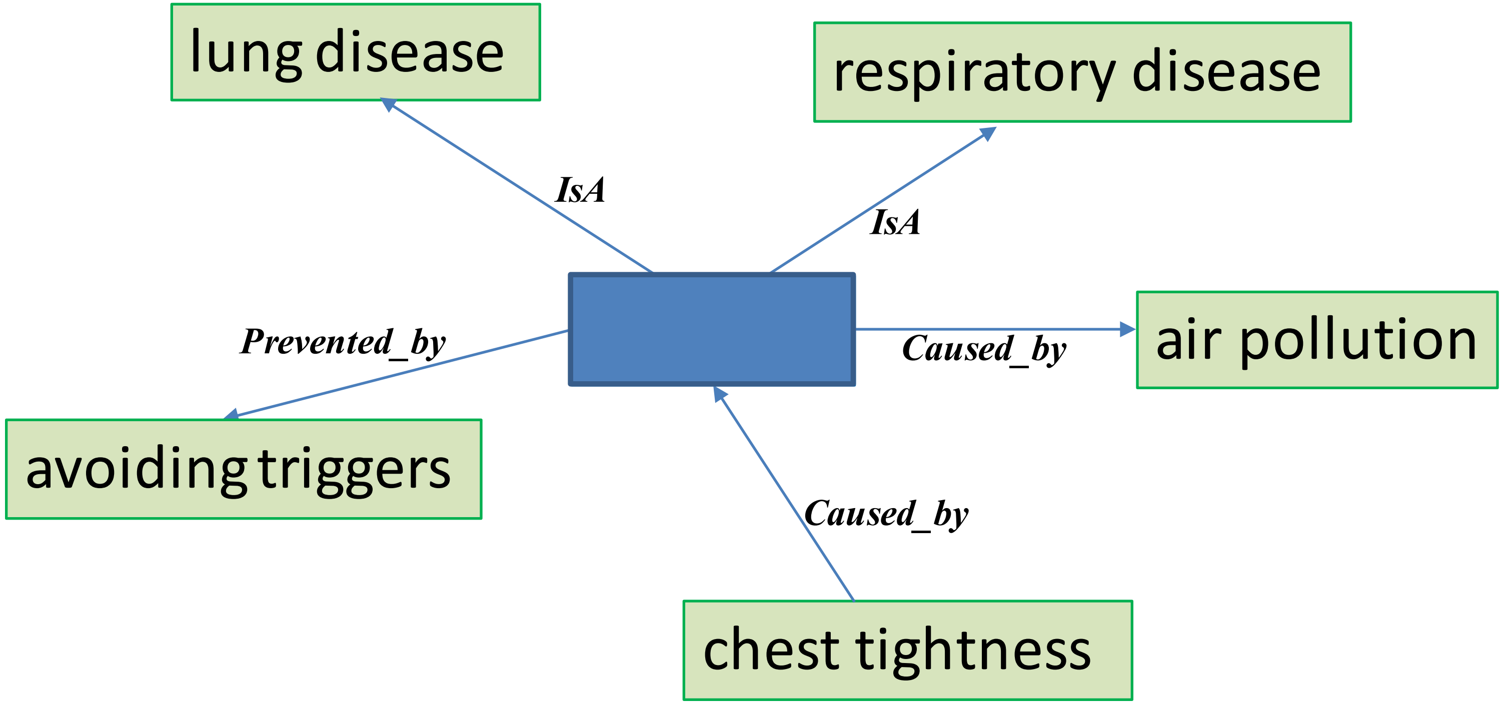
[2]Beijing Institute of Technology, China

[3]Sogou Inc., Beijing, China

Motivation

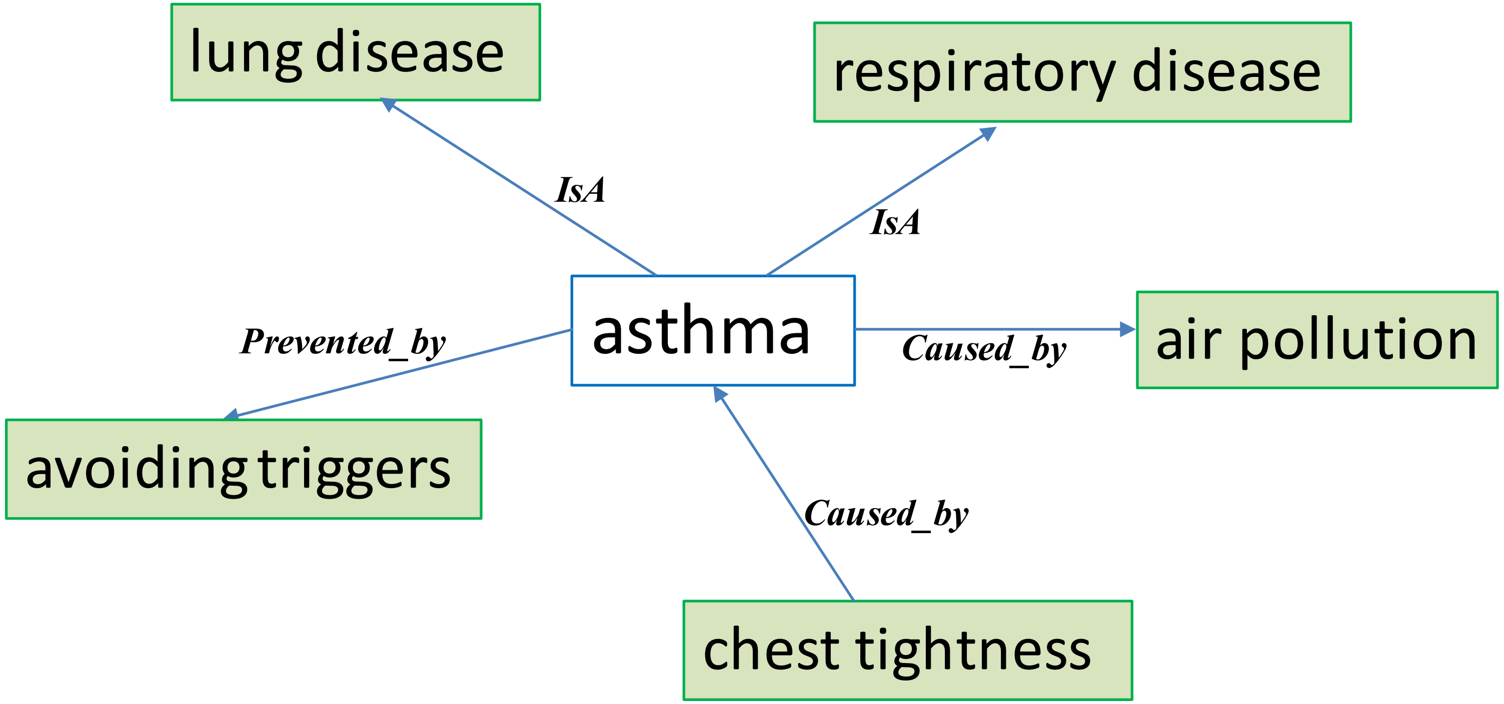
- **Commonsense knowledge** consists of facts about the everyday world, that all humans are expected to know.
  - Lemons are sour
  - Trees have leaves
  - Dogs have four legs
- Commonsense Reasoning ~ **Winograd Schema Challenge**:
  - The trophy would not fit in the brown suitcase because it was too **big**. What was too **big**?
  - The trophy would not fit in the brown suitcase because it was too **small**. What was too **small**?

Motivation



lung disease
respiratory disease
IsA
IsA
Prevented_by
Caused_by
air pollution
avoiding triggers
Caused_by
chest tightness

Motivation

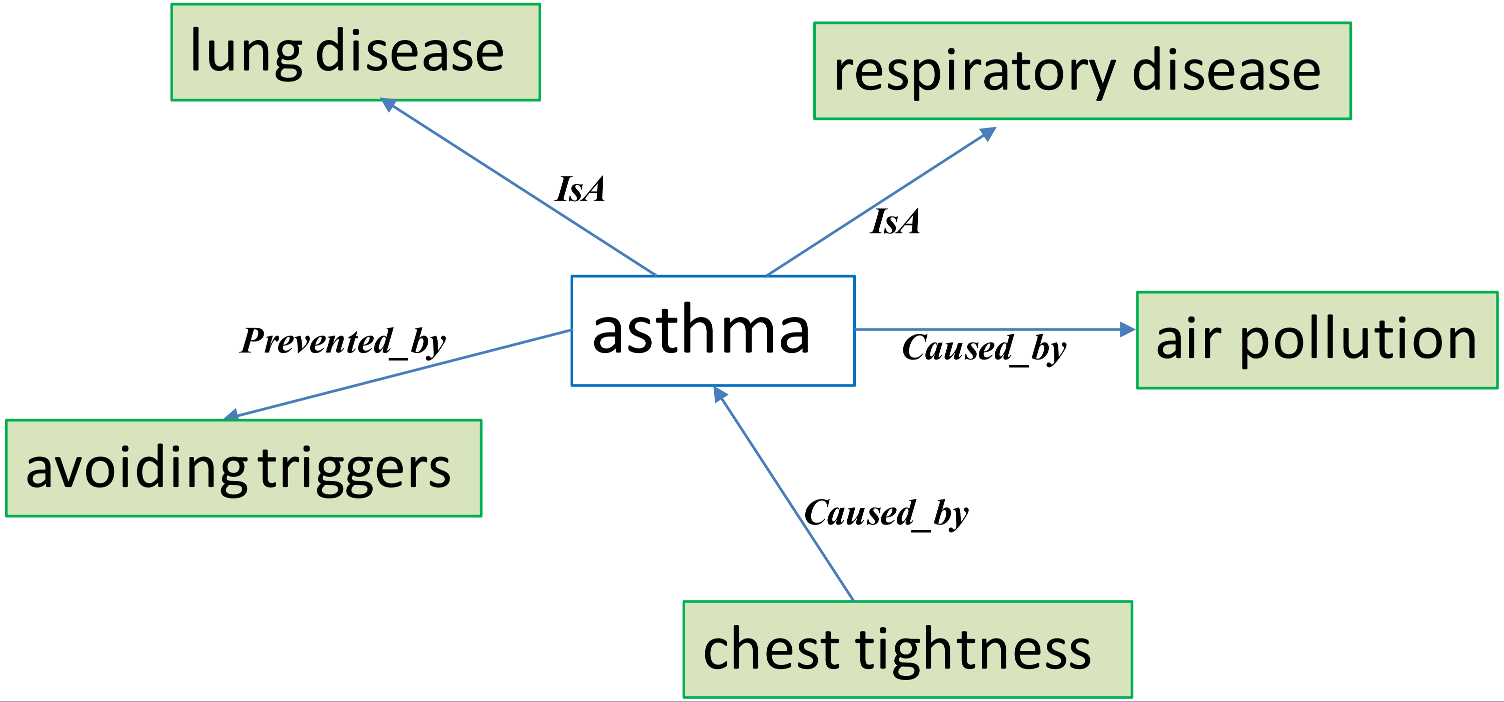


lung disease
respiratory disease
IsA
IsA
asthma
Prevented_by
Caused_by
air pollution
avoiding triggers
Caused_by
chest tightness

Motivation

I have an asthma since three years old.

Triples in knowledge graph:
(chest tightness, Caused_by, asthma)
(asthma, Prevented_by, avoiding triggers)

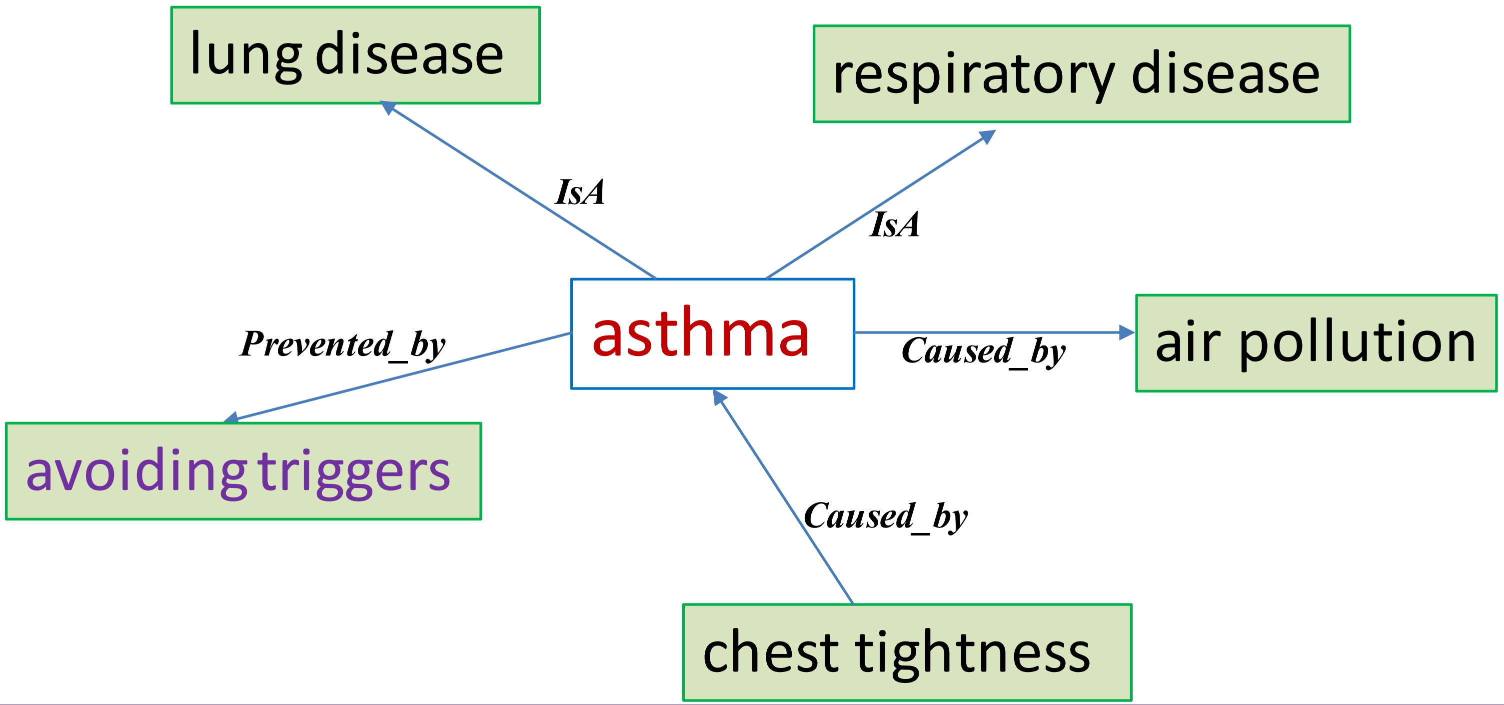


Motivation

I have an asthma since three years old.

Triples in knowledge graph:
(chest tightness, Caused_by, asthma)
(asthma, Prevented_by, avoiding triggers)

I am sorry to hear that. Maybe avoiding triggers can prevent asthma attacks.

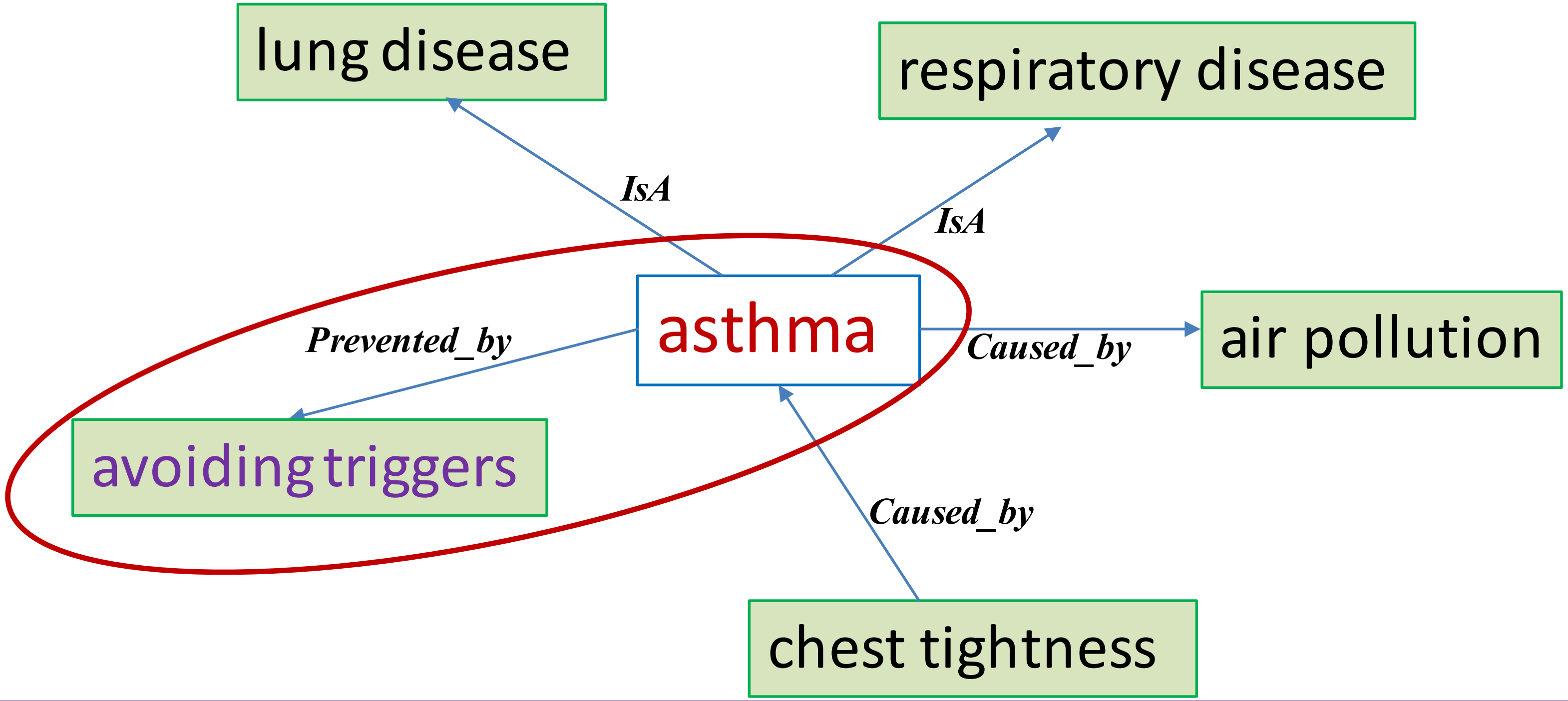


Motivation

I have an asthma since three years old.

Triples in knowledge graph:
(chest tightness, Caused_by, asthma)
(asthma, Prevented_by, avoiding triggers)

I am sorry to hear that. Maybe avoiding triggers can prevent asthma attacks.



Motivation

- Conversation is all about **semantic understanding**
- Commonsense knowledge is important for **language understanding and generation**

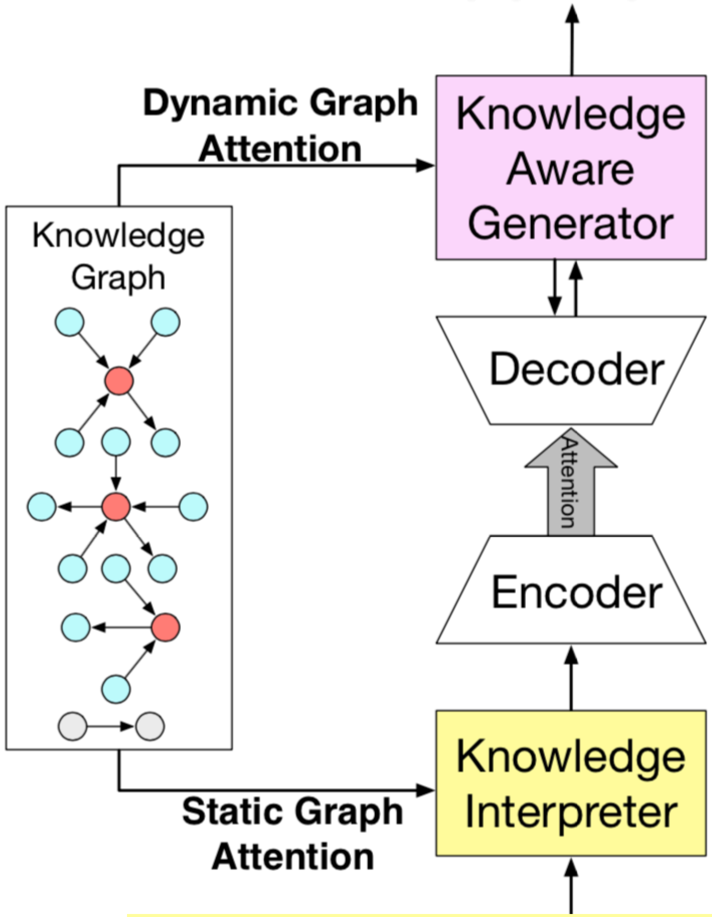
Post	Why are you so **breakable**?
Knowledge	(**glass**, RelatedTo, **breakable**), (**brittle**, RelatedTo, **breakable**), (**rule**, RelatedTo, **breakable**)
Seq2Seq	I'm not a **OOV** , i 'm just a **OOV**.
MemNet	I'm not **OOV**. I'm just a really nice person.
CopyNet	I'm not. I'm just a lurker.
CCM	Because I'm a **brittle** man .

Related Work

- A **knowledge-grounded** neural conversation model [Ghazvininejad et al., 2017].
- Flexible end-to-end dialogue system for **knowledge grounded** conversation [Zhu et al., 2017].
- Seq2Seq, Memory Network, Copy Network …

Overview

Output: Because I'm a brittle man.



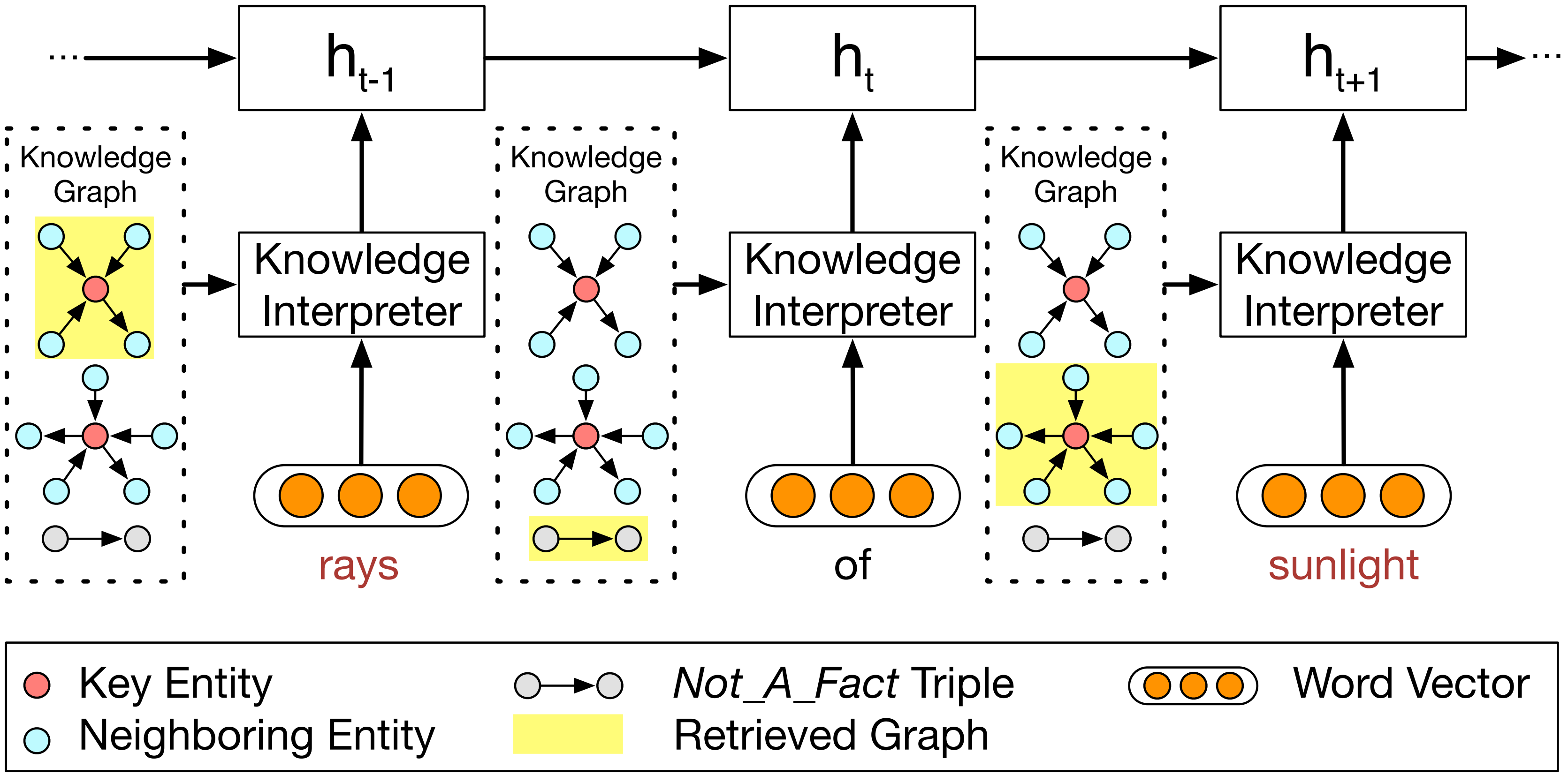
Decoding words by attending to knowledge graphs and then to triples

Encoding the retrieved knowledge graphs for each word

Input: why are you so breakable?

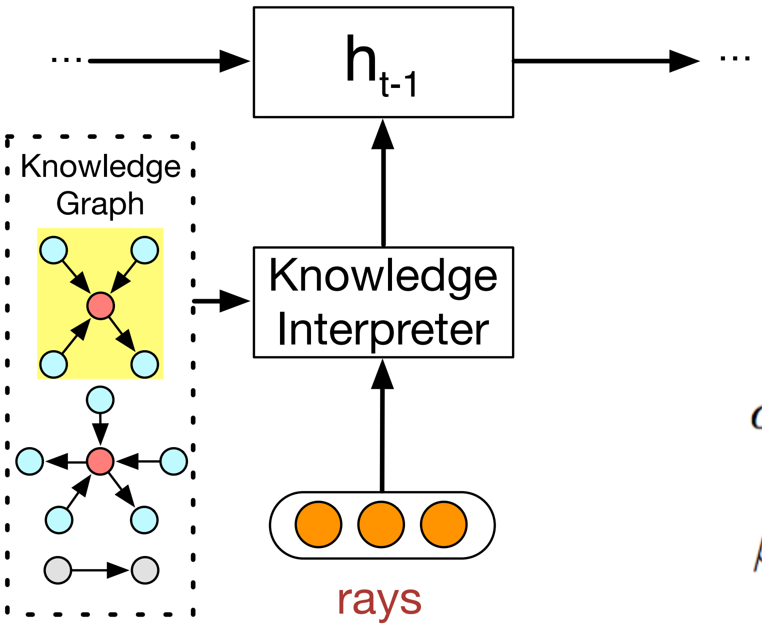
Model

- **Knowledge Interpreter**

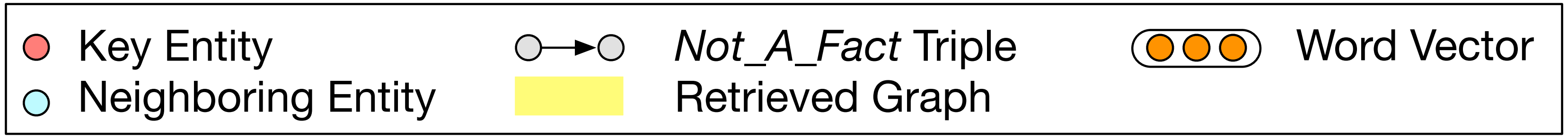


Knowledge Interpreter

- **Static graph attention**: encoding semantics in graph, feeding knowledge-enhanced info. into the encoder

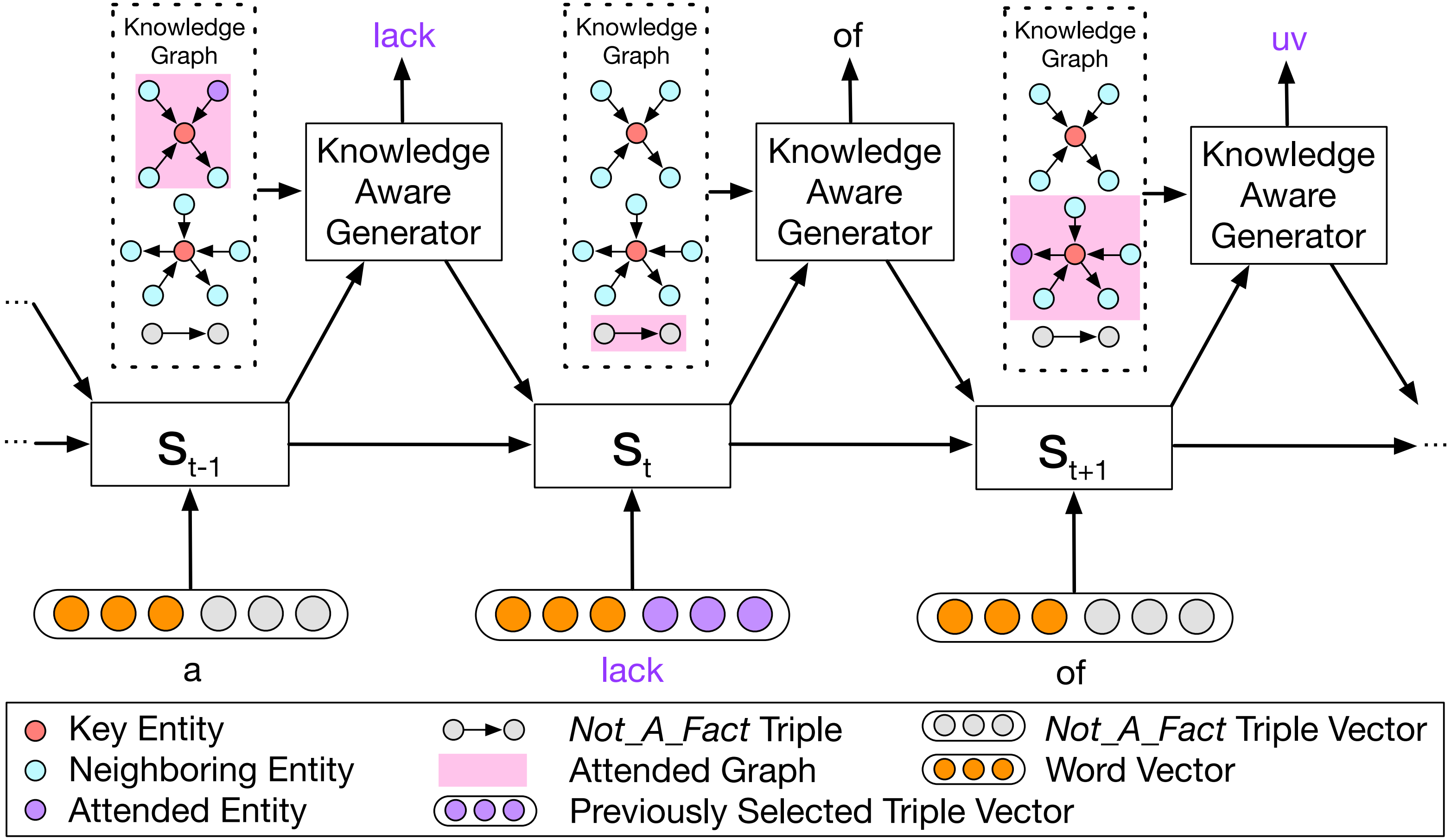


$$
\begin{aligned}
\boldsymbol{g}_i &= \sum_{n=1}^{N_{g_i}} \alpha_n^s [\boldsymbol{h}_n; \boldsymbol{t}_n], \\
\alpha_n^s &= \frac{\exp(\beta_n^s)}{\sum_{j=1}^{N_{g_i}} \exp(\beta_j^s)}, \\
\beta_n^s &= (\mathbf{W_r} \boldsymbol{r}_n)^\top \tanh(\mathbf{W_h} \boldsymbol{h}_n + \mathbf{W_t} \boldsymbol{t}_n),
\end{aligned}
$$



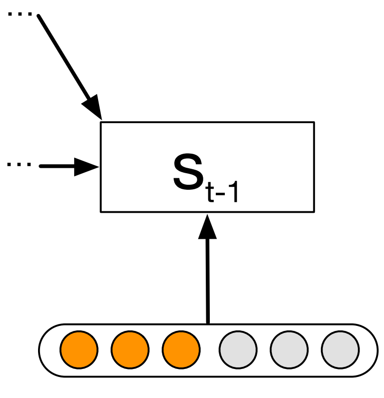
Model

- Knowledge Aware Generator



Knowledge Aware Generator

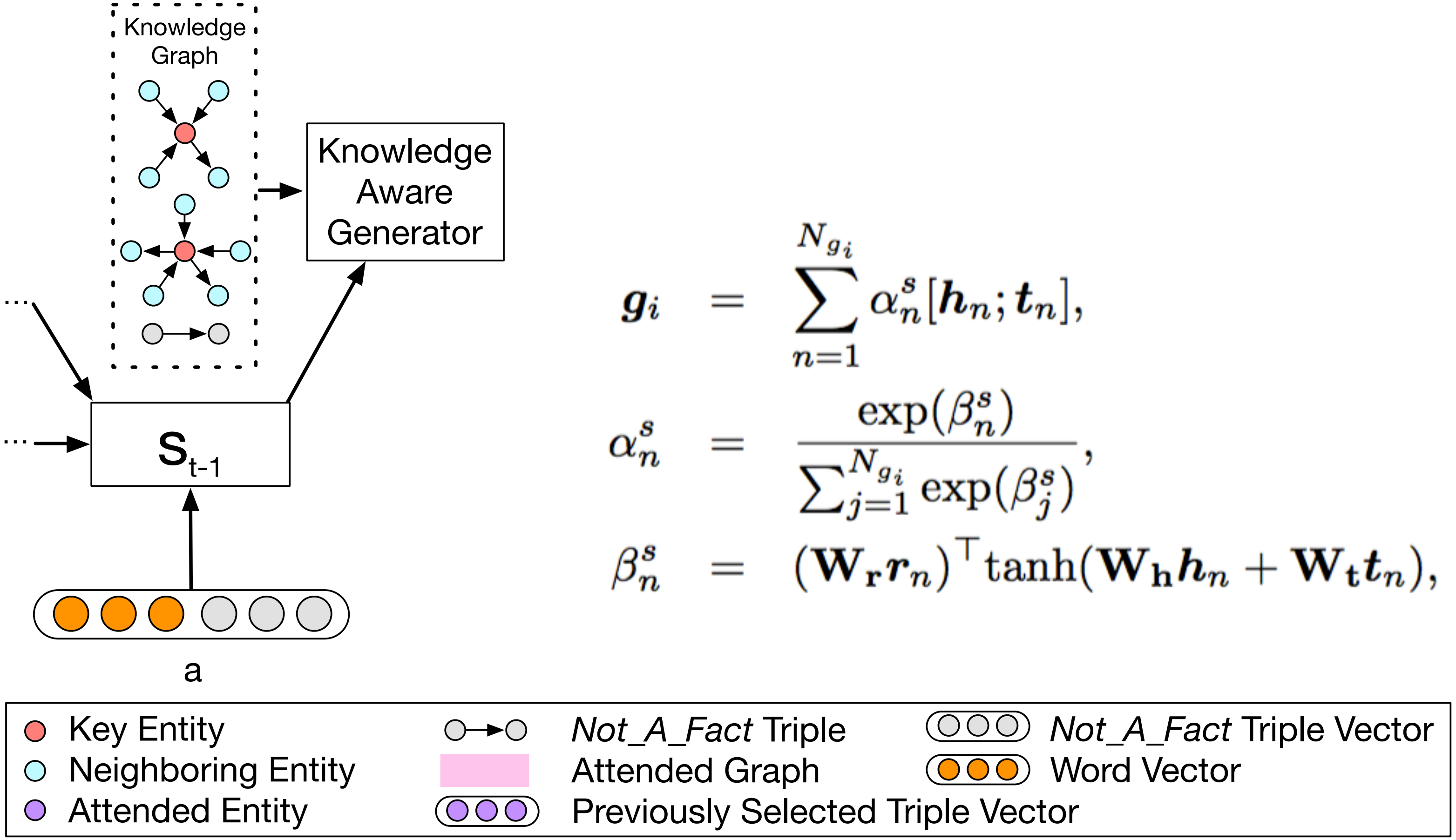
- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



$$
\begin{aligned}
\boldsymbol{s}_{t+1} &= \mathbf{GRU}(\boldsymbol{s}_t, [\boldsymbol{c}_t; \boldsymbol{c}_t^g; \boldsymbol{c}_t^k; \boldsymbol{e}(y_t)]), \\
\boldsymbol{e}(y_t) &= [\boldsymbol{w}(y_t); \boldsymbol{k}_j],
\end{aligned}
$$

Knowledge Aware Generator

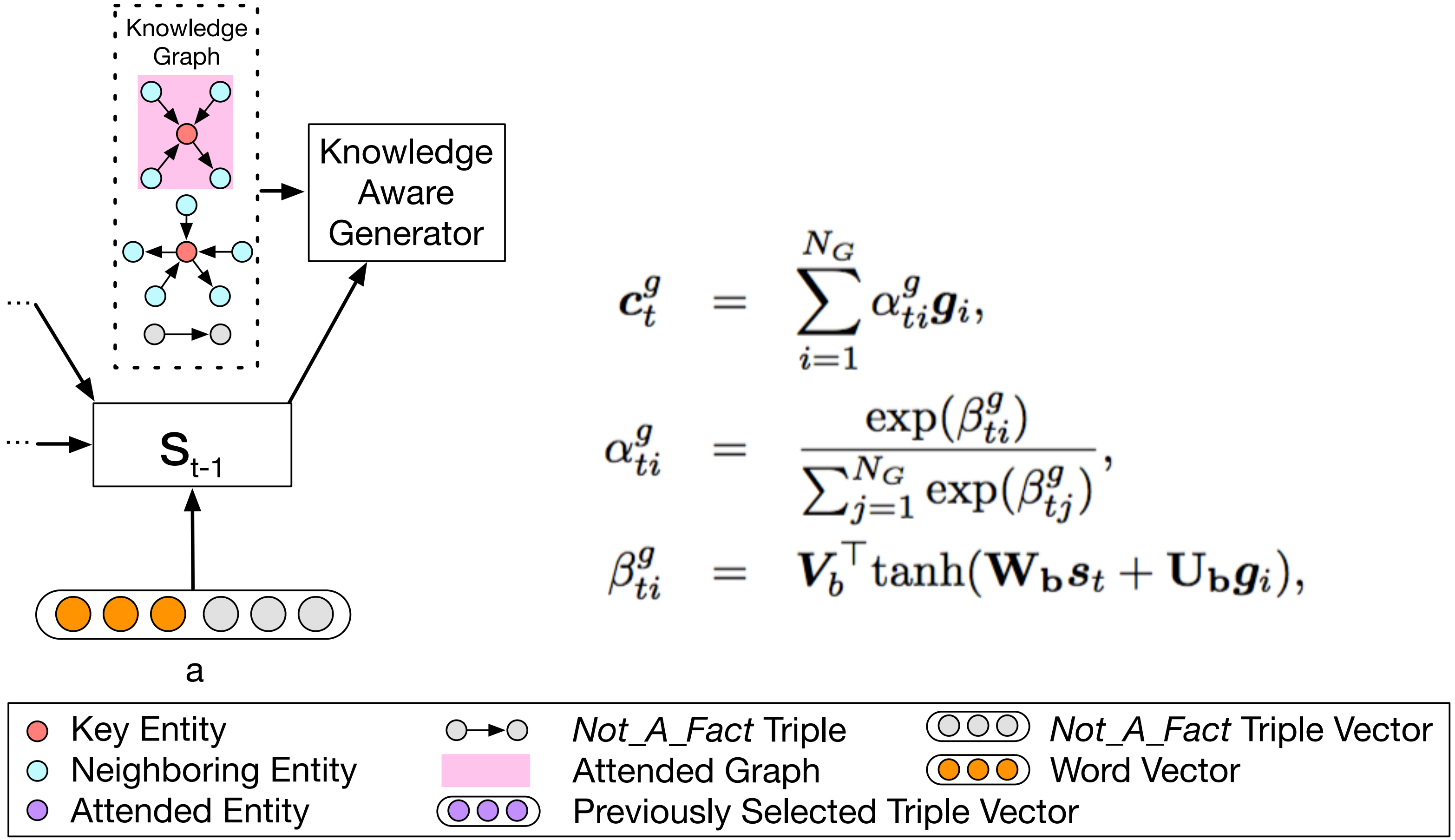
- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



$$
\begin{aligned}
\boldsymbol{g}_i &= \sum_{n=1}^{N_{g_i}} \alpha_n^s [\boldsymbol{h}_n; \boldsymbol{t}_n], \\
\alpha_n^s &= \frac{\exp(\beta_n^s)}{\sum_{j=1}^{N_{g_i}} \exp(\beta_j^s)}, \\
\beta_n^s &= (\mathbf{W_r} \boldsymbol{r}_n)^\top \tanh(\mathbf{W_h} \boldsymbol{h}_n + \mathbf{W_t} \boldsymbol{t}_n),
\end{aligned}
$$

Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



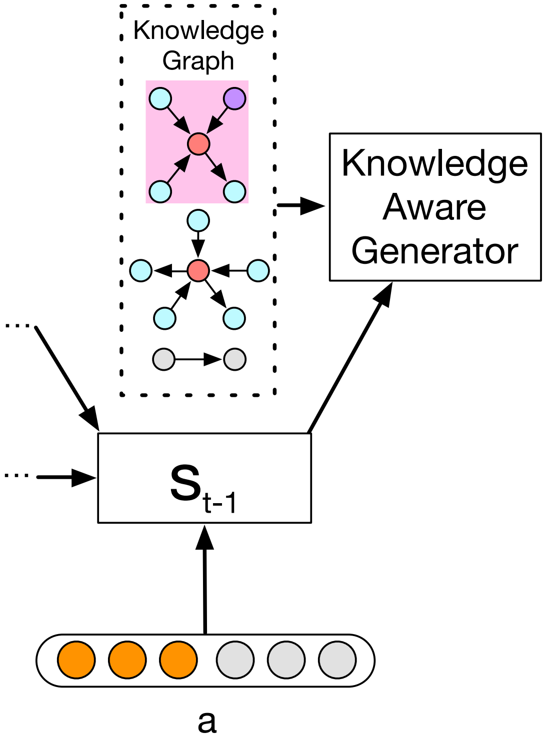
$$\boldsymbol{c}_t^g = \sum_{i=1}^{N_G} \alpha_{ti}^g \boldsymbol{g}_i,$$

$$\alpha_{ti}^g = \frac{\exp(\beta_{ti}^g)}{\sum_{j=1}^{N_G} \exp(\beta_{tj}^g)},$$

$$\beta_{ti}^g = \boldsymbol{V}_b^\top \tanh(\mathbf{W_b} \boldsymbol{s}_t + \mathbf{U_b} \boldsymbol{g}_i),$$

Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



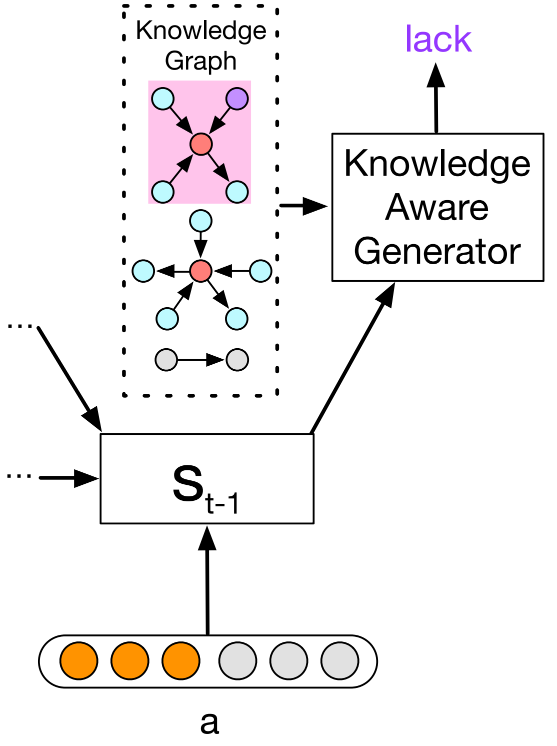
$$\mathbf{c}_t^k = \sum_{i=1}^{N_G} \sum_{j=1}^{N_{g_i}} \alpha_{ti}^g \alpha_{tj}^k \boldsymbol{k}_j,$$

$$\alpha_{tj}^k = \frac{\exp(\beta_{tj}^k)}{\sum_{n=1}^{N_{g_i}} \exp(\beta_{tn}^k)},$$

$$\beta_{tj}^k = \boldsymbol{k}_j^\top \mathbf{W}_\mathbf{c} \boldsymbol{s}_t,$$

Knowledge Aware Generator

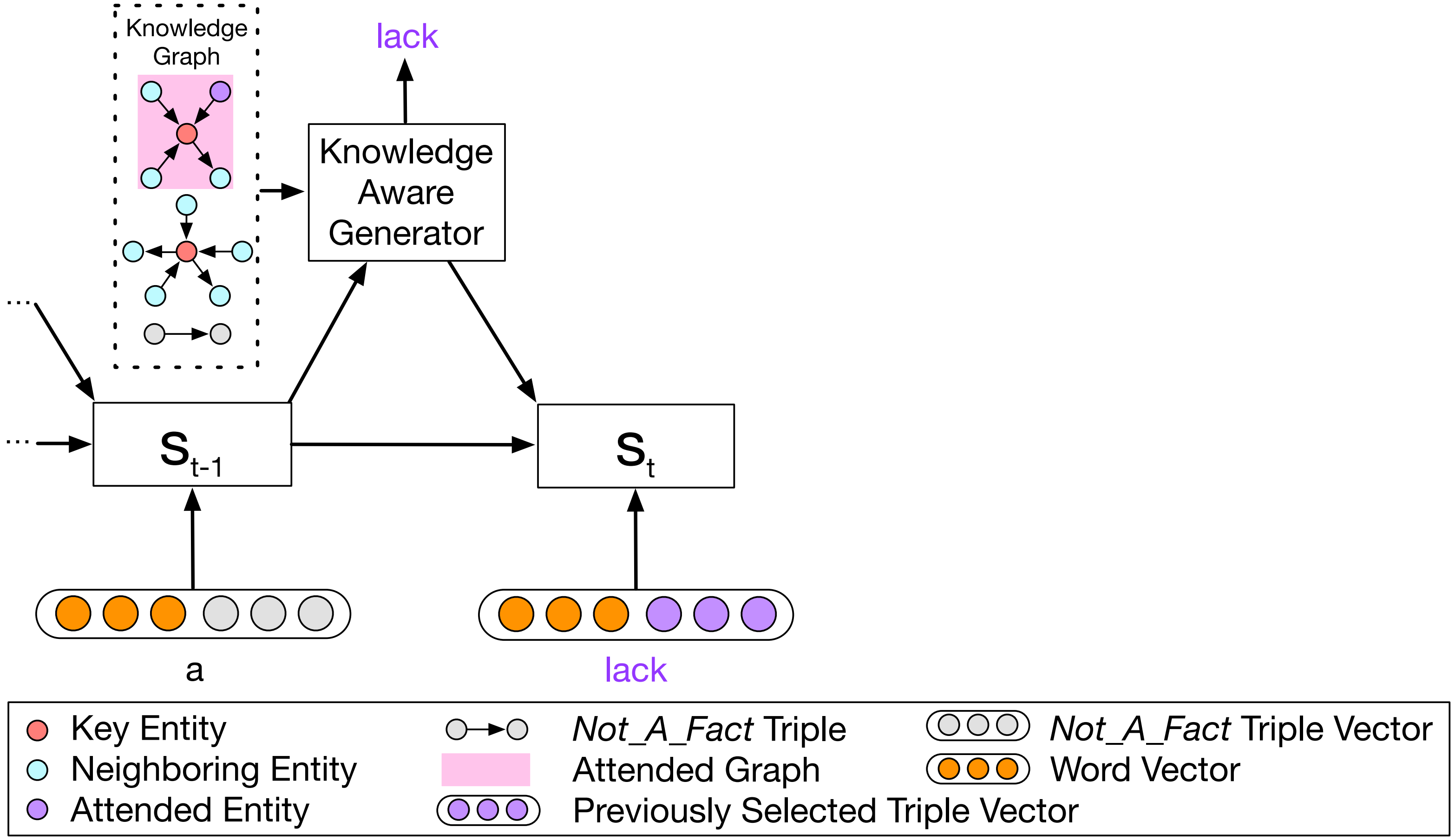
- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



$$\begin{aligned}
\boldsymbol{a}_t &= [\boldsymbol{s}_t; \boldsymbol{c}_t; \boldsymbol{c}_t^g; \boldsymbol{c}_t^k], \\
\gamma_t &= \mathrm{sigmoid}(\mathbf{V_o}^\top \boldsymbol{a}_t), \\
P_c(y_t = w_c) &= \mathrm{softmax}(\mathbf{W_o} \boldsymbol{a}_t), \\
P_e(y_t = w_e) &= \alpha_{ti}^g \alpha_{tj}^k, \\
y_t \sim \boldsymbol{o}_t = P(y_t) &= \begin{bmatrix} (1-\gamma_t) P_g(y_t = w_c) \\ \gamma_t P_e(y_t = w_e) \end{bmatrix},
\end{aligned}$$

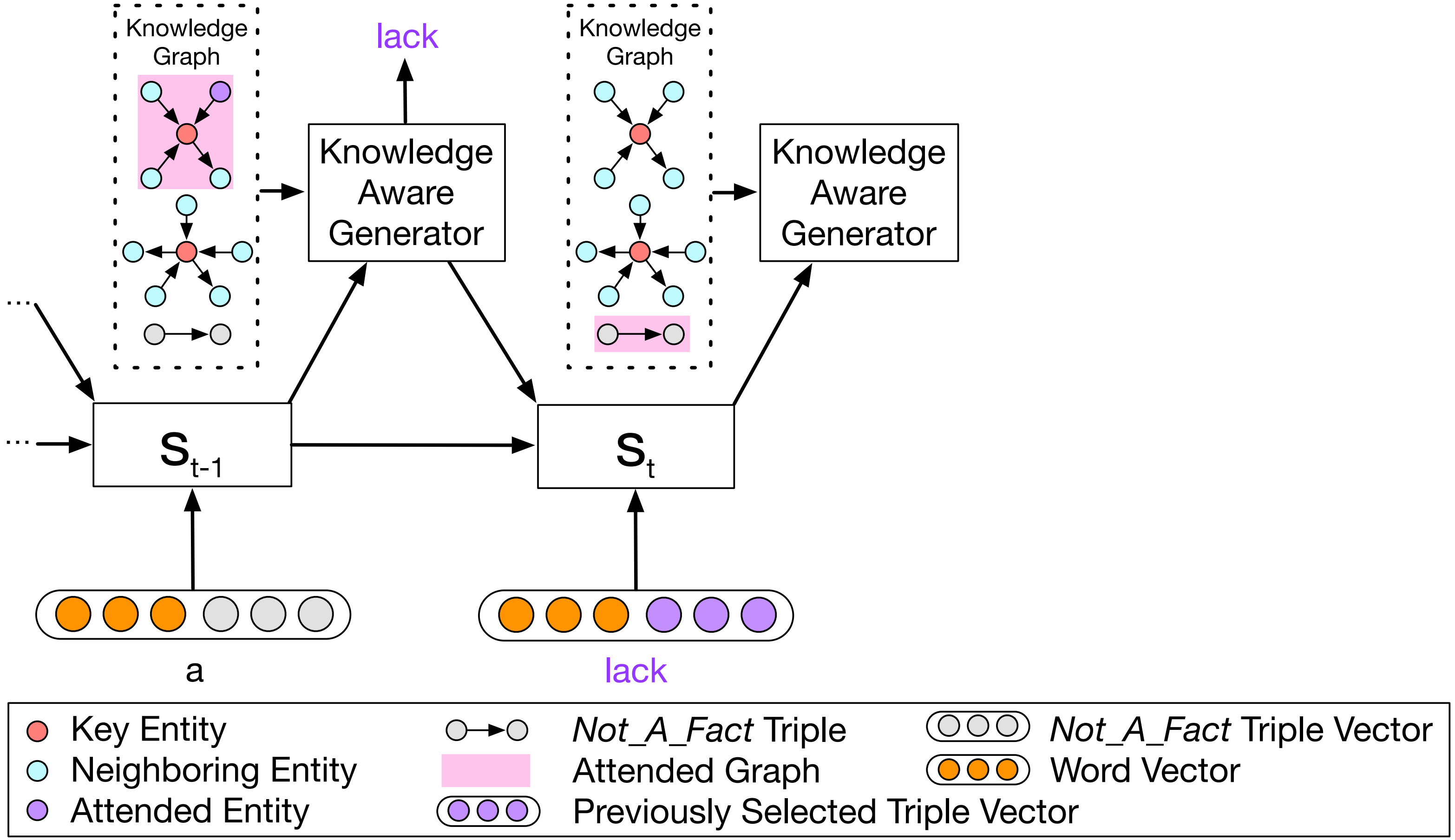
Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



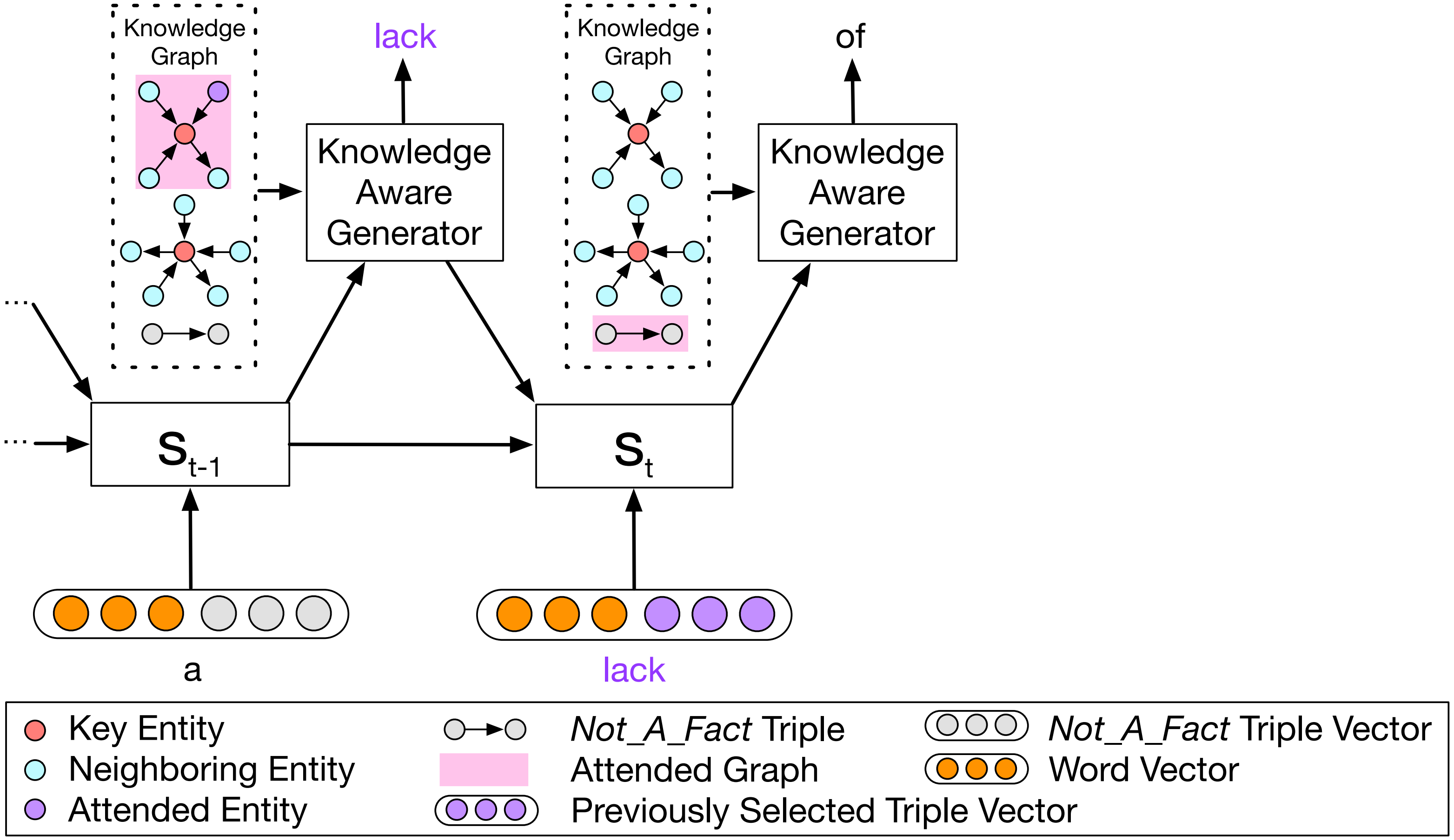
Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



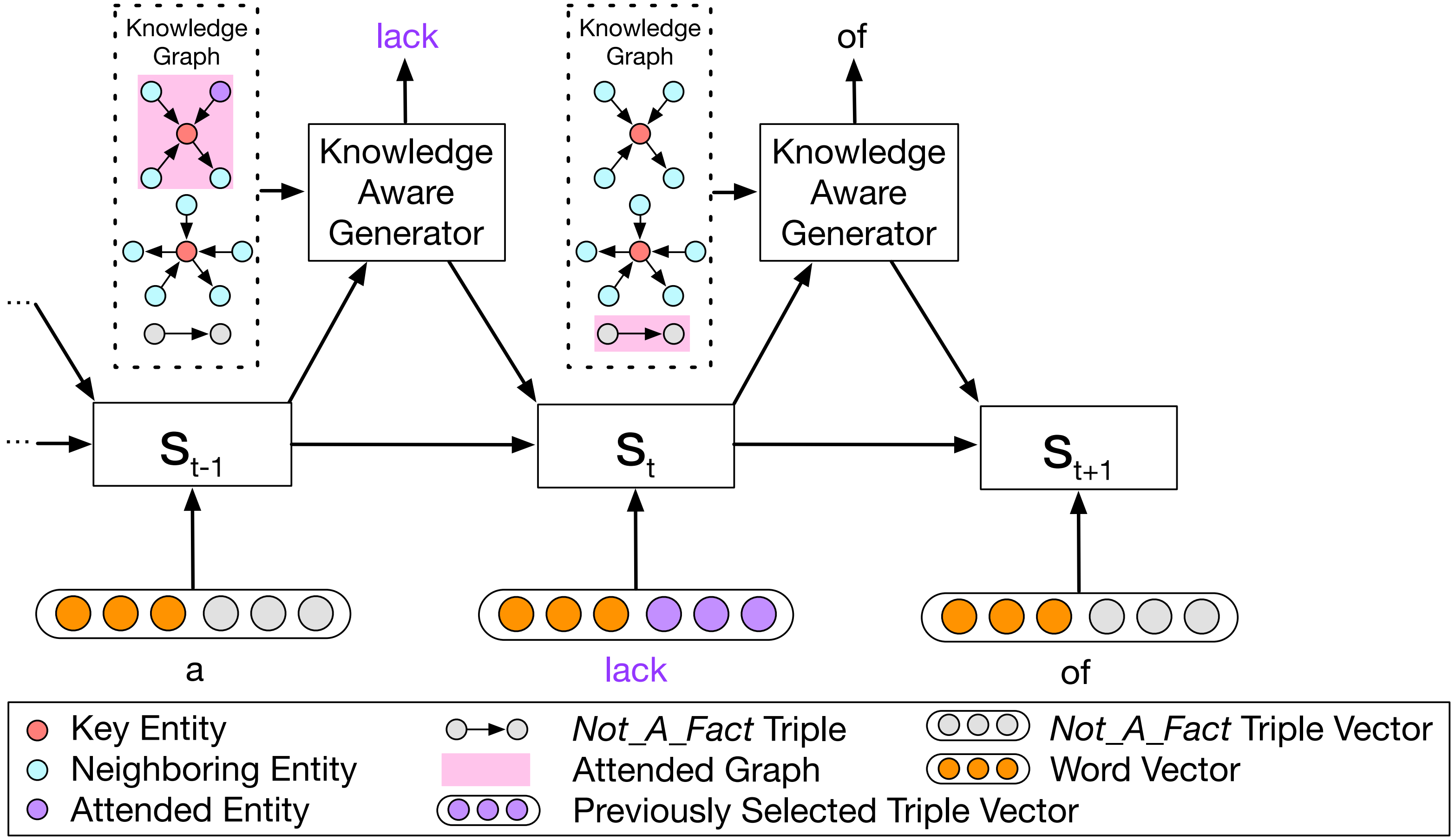
Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



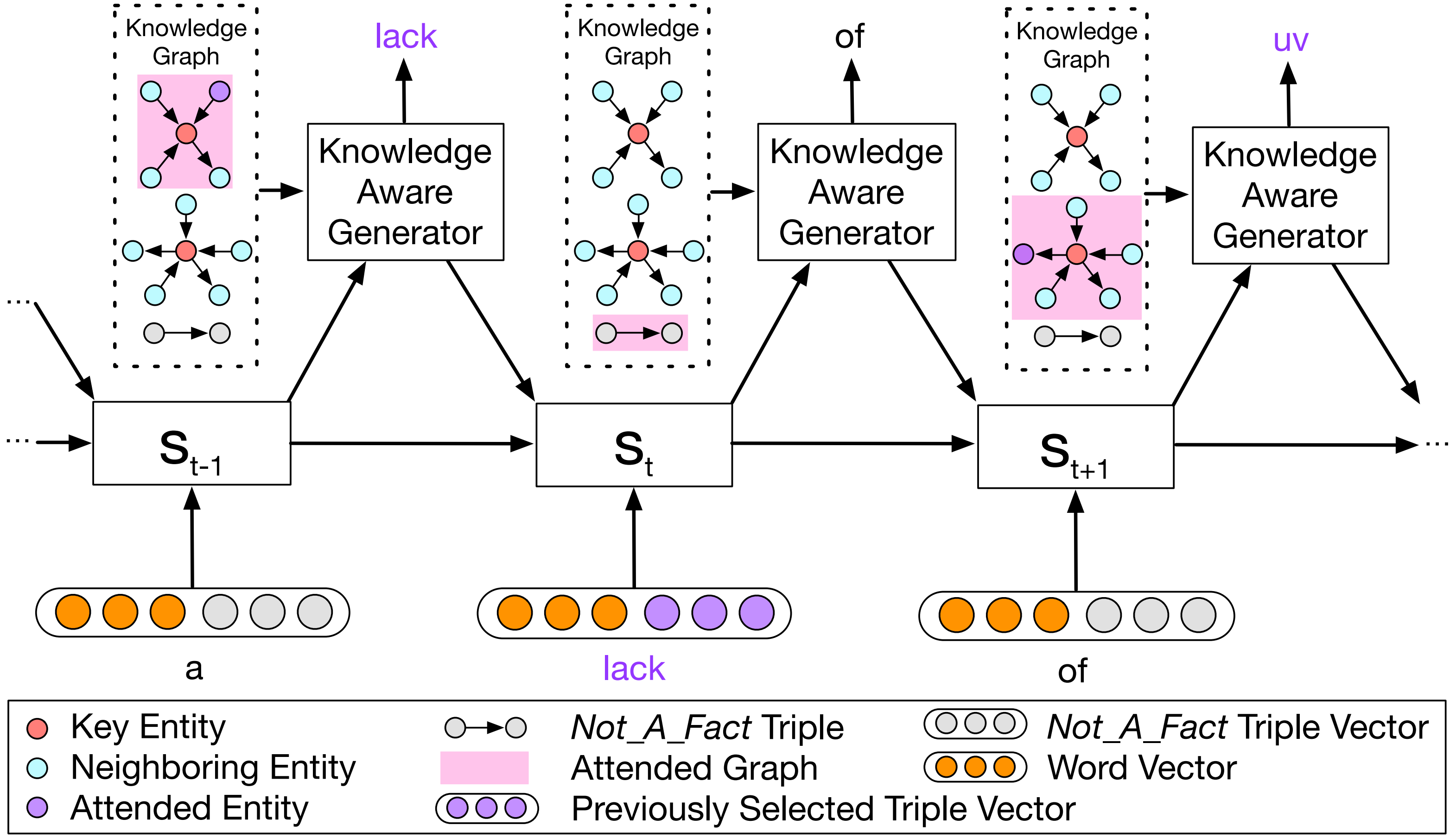
Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



Knowledge Aware Generator

- **Dynamic graph attention**: first attend a graph, then to a triple within that graph



Experiments

Commonsense Conversation Dataset

- We used the ConceptNet as our commonsense knowledge base which was released by MIT.
- We adopted 10M reddit single-round dialogs from Internet.
- To test how commonsense knowledge can help understand common or rare concepts in a post, we constructed four test sets from high-frequency to OOV. Each test set has 5,000 pairs randomly sampled from the dataset.

Conversational Pairs		Commonsense KB	
Training	3,384,185	Entity	21,471
Validation	10,000	Relation	44
Test	20,000	Triple	120,850

Table 1: Statistics of the dataset and the knowledge base.

Experiments

Automatic Evaluation

Model	Overall		High Freq.		Medium Freq.		Low Freq.		OOV	
	ppx.	ent.	ppx.	ent.	ppx.	ent.	ppx.	ent.	ppx.	ent.
Seq2Seq	47.02	0.717	42.41	0.713	47.25	0.740	48.61	0.721	49.96	0.669
MemNet	46.85	0.761	41.93	0.764	47.32	0.788	48.86	0.760	49.52	0.706
CopyNet	40.27	0.96	36.26	0.91	40.99	0.97	42.09	0.96	42.24	0.96
CCM	**39.18**	**1.180**	**35.36**	**1.156**	**39.64**	**1.191**	**40.67**	**1.196**	**40.87**	**1.162**

Manual evaluation (Sign-test, p-value<0.005)

Model	Overall		High Freq.		Medium Freq.		Low Freq.		OOV	
	app.	inf.	app.	inf.	app.	inf.	app.	inf.	app.	inf.
CCM vs. Seq2Seq	0.616	0.662	0.605	0.656	0.549	0.624	0.636	0.650	0.673	0.716
CCM vs. MemNet	0.602	0.647	0.593	0.656	0.566	0.640	0.622	0.635	0.626	0.657
CCM vs. CopyNet	0.600	0.640	0.606	0.669	0.586	0.619	0.610	0.633	0.596	0.640

Experiments

Post: He proposed March 5th. We will be married October 10th. So 7 months

Knowledge: (engaged, RelatedTo, married), (wedding, RelatedTo, married), (miss, DistinctFrom, married), (March, RelatedTo, months)

- **Seq2Seq**: I'm not sure how to answer that. I'm not sure how to make a date.
- **MemNet**: I'm assuming he's married to a gay marriage?
- **CopyNet**: I'm sorry, I'm not sure what you mean by that.
- **CCM**: I'm pretty sure he's going to be **engaged** in the next 6 **months**.

Experiments

Post: Totally thought it was going to be doug's grave.
Knowledge: (thought, RelatedTo, thinking), (doug, HasProperty, dead), (dead, RelatedTo, grave), (grave, RelatedTo, death)

- **Seq2Seq**: I was thinking the same thing.
- **MemNet**: I thought it was a dead end.
- **CopyNet**: I think it's a bit of a stretch.
- **CCM**: I thought it was doug's **death**. I was **thinking** the same thing.

Conclusion

- This work is the first attempt that uses large-scale commonsense knowledge in neural conversation generation.

- Instead of treating knowledge triples (or entities) separately and independently, we devise static and dynamic graph attention mechanisms to treat the knowledge triples as a graph, from which we can better interpret the semantics of an entity from its neighboring entities and relations.

Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory

Hao Zhou†, Minlie Huang†*, Tianyang Zhang†, Xiaoyan Zhu†, Bing Liu‡

†Tsinghua University, Beijing, China

‡University of Illinois at Chicago, Illinois, USA

Motivation

- **Emotion intelligence** is a key human behavior for intelligence (Salovey and Mayer, 1990; Picard and Picard, 1997)
- Understanding **emotion and affect** is important for dialogue and conversation
  - Enhance user performance
  - Improve user satisfaction
  - Fewer breakdowns
- **Rule-based** emotion adaptation
  - Seen in early dialogue systems

Motivation

- **Emotion intelligence** is a key human behavior for intelligence (Salovey and Mayer, 1990; Picard and Picard, 1997)
- Understanding **emotion and affect** is important for dialogue and conversation
  - Enhance user performance
  - Improve user satisfaction
  - Fewer breakdowns
- **Rule-based** emotion adaptation
  - Seen in early dialogue systems

User: Worst day ever. I arrived late because of the traffic.
Chatbot (Basic Seq2seq): You were late. ECM (*Like*): I am always here to support you. ECM (*Happiness*): Keep smiling! Things will get better. ECM (*Sadness*): It's depressing. ECM (*Disgust*): Sometimes life just sucks. ECM (*Anger*): The traffic is too bad!

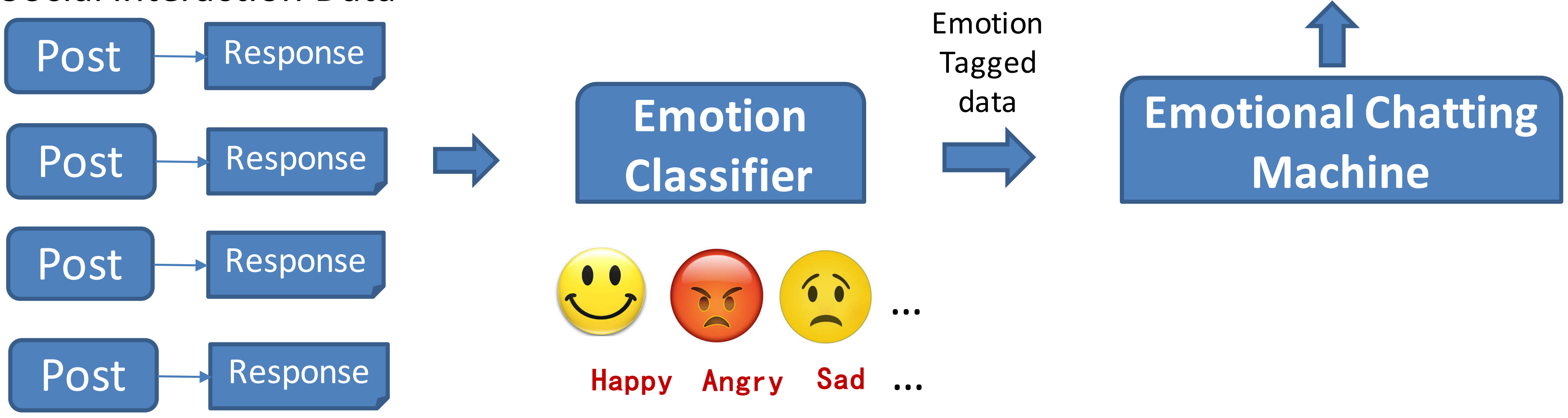
Overview

Perceiving and Expressing emotion by machine
Closer to human-level intelligence

User: Worst day ever. I arrived late because of the traffic.

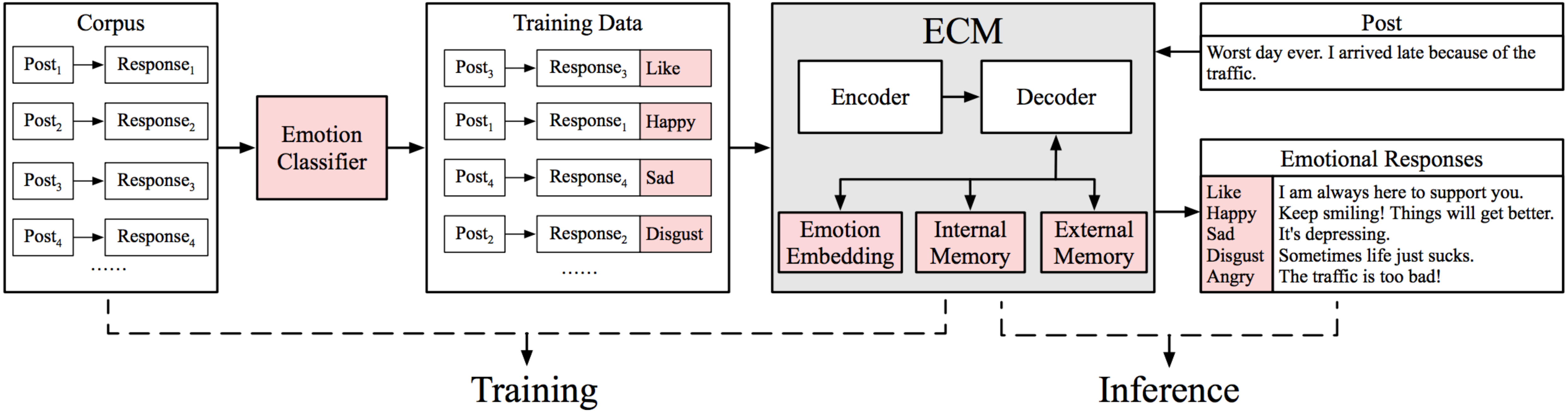
Chatbot (Basic Seq2seq): You were late.
ECM (*Like*): I am always here to support you.
ECM (*Happiness*): Keep smiling! Things will get better.
ECM (*Sadness*): It's depressing.
ECM (*Disgust*): Sometimes life just sucks.
ECM (*Anger*): The traffic is too bad!

Social Interaction Data



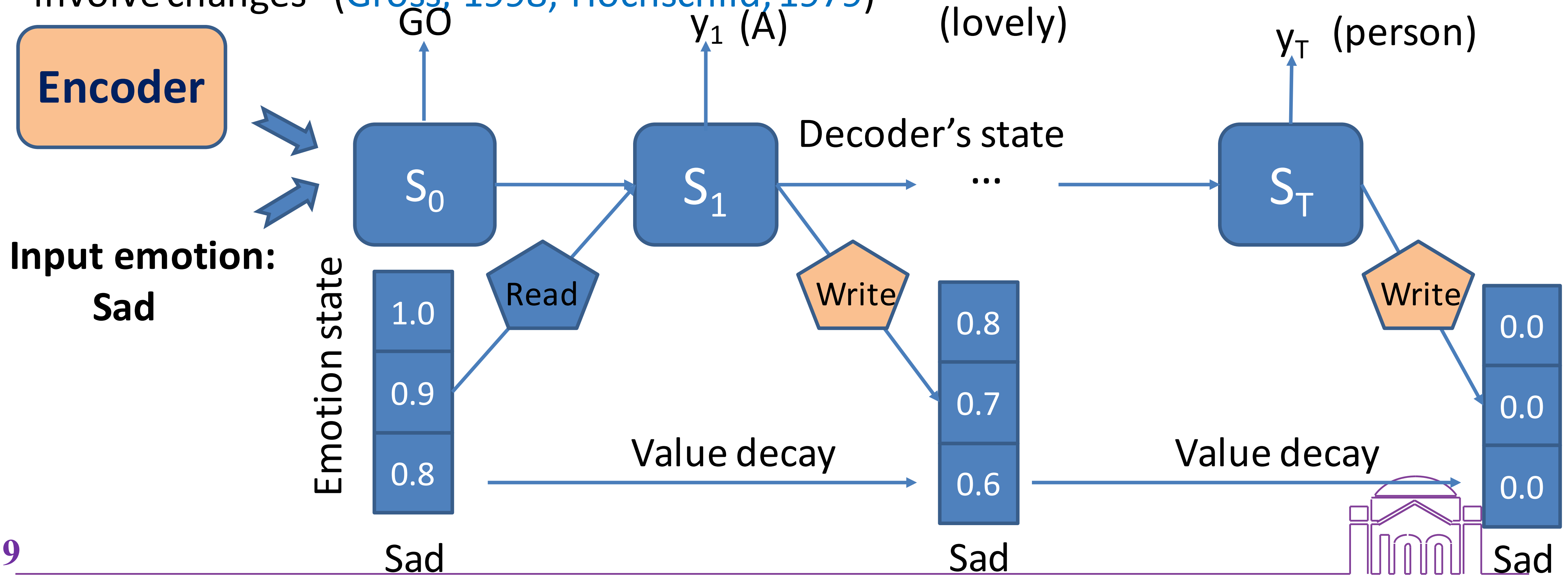
Overview

- **Emotion category embedding**: High level abstraction of emotions
- **Emotion internal memory**: Capturing the change of emotion state during decoding
- **Emotion external memory**: Treating emotion/generic words differentially



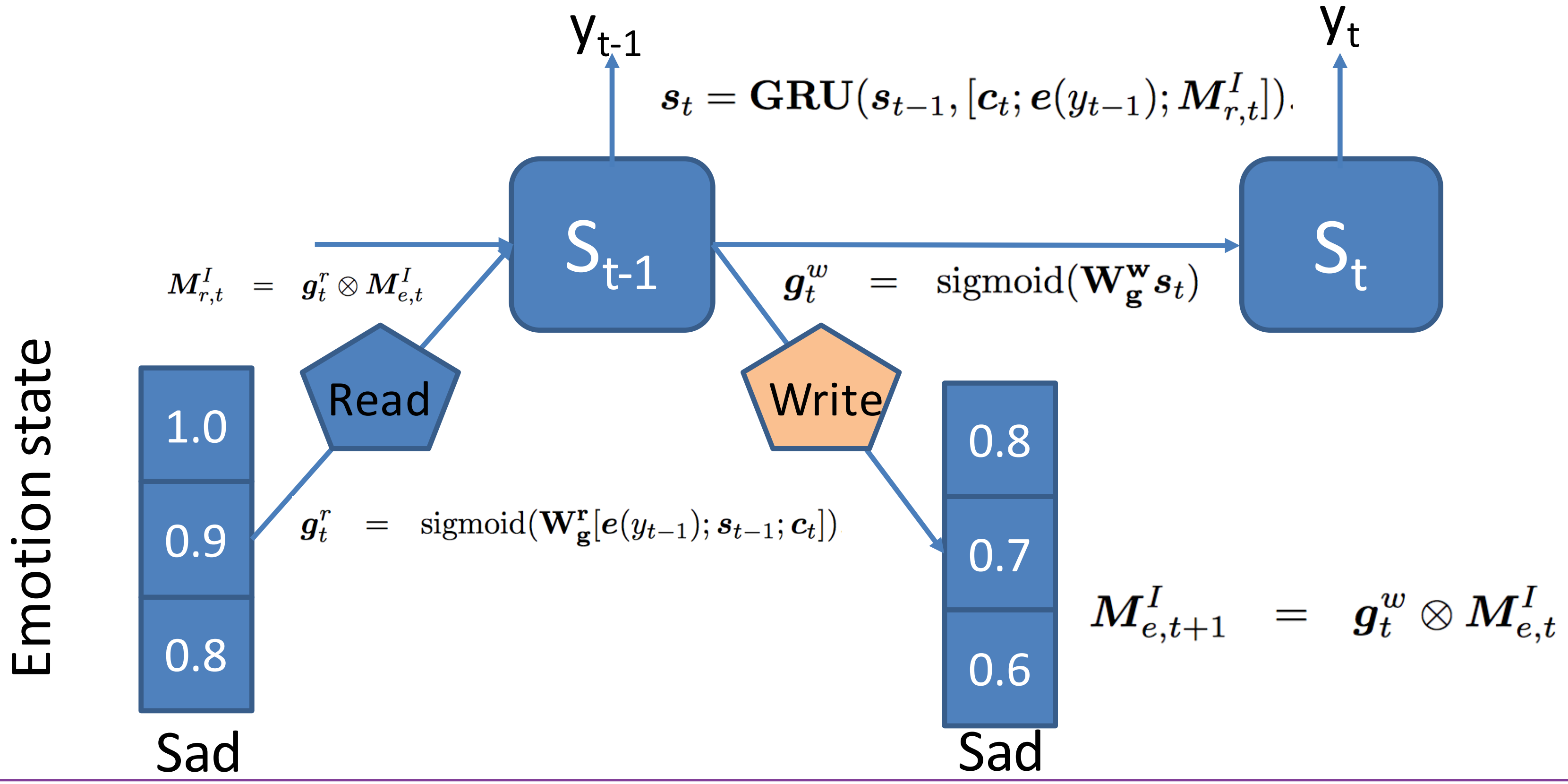
Model

- **Emotion internal memory** : "emotional responses are relatively short lived and involve changes" (Gross, 1998; Hochschild, 1979)



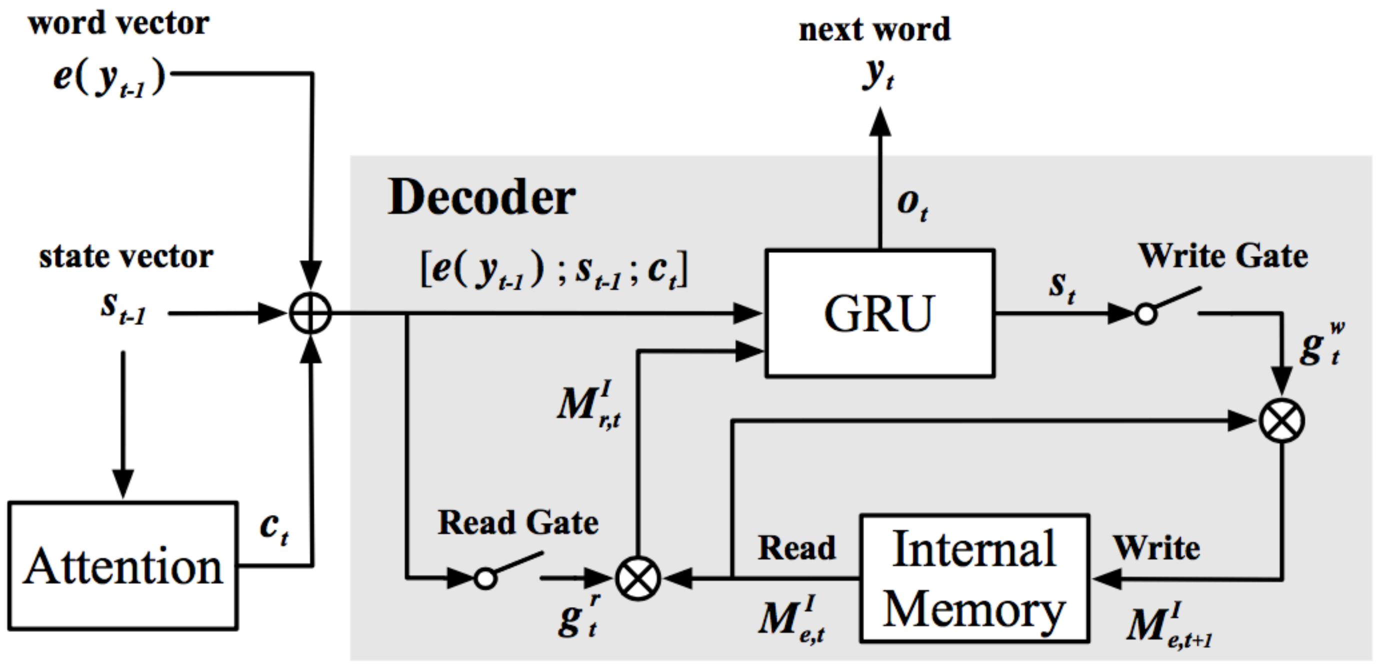
Model

- **Emotion internal memory** : “emotional responses are relatively short lived and involve changes” (Gross, 1998; Hochschild, 1979)



Model

- **Emotion internal memory** : "emotional responses are relatively short lived and involve changes" (Gross, 1998; Hochschild, 1979)



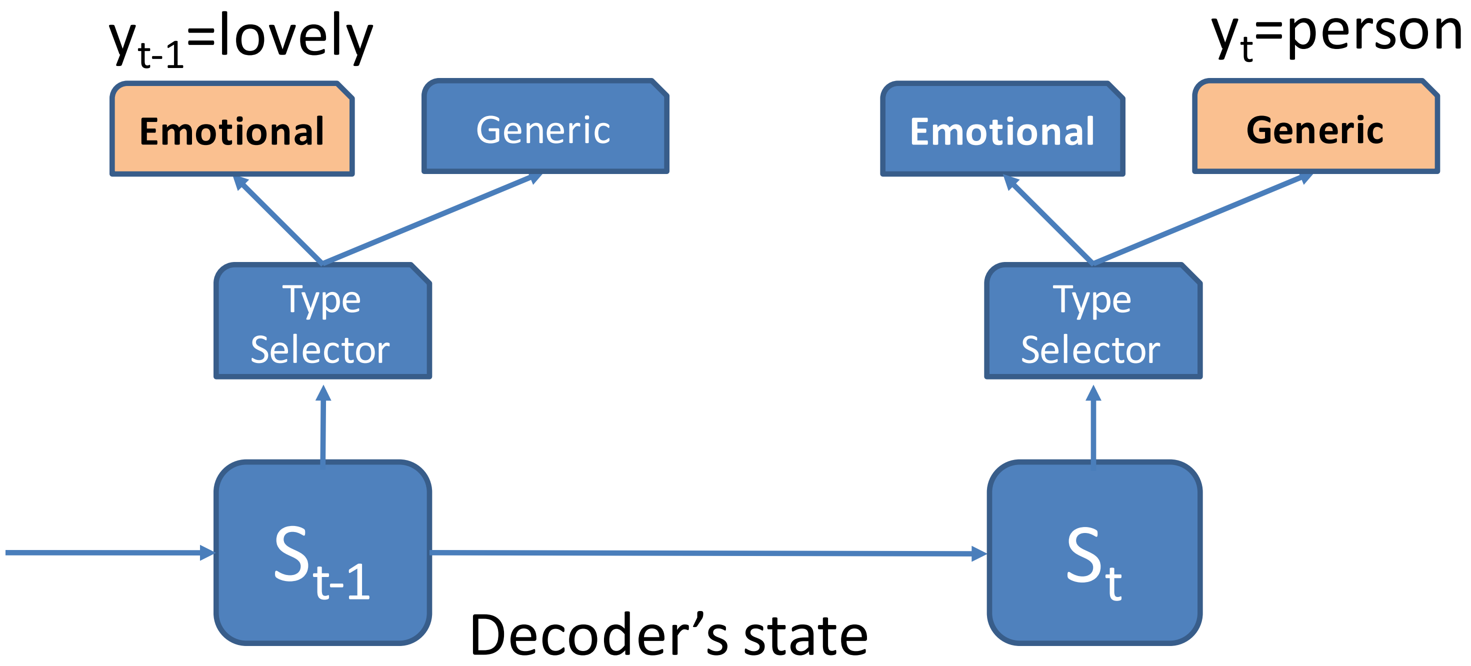
$$g_t^r = \mathrm{sigmoid}(\mathbf{W_g^r}[\boldsymbol{e}(y_{t-1}); \boldsymbol{s}_{t-1}; \boldsymbol{c}_t]),$$
$$g_t^w = \mathrm{sigmoid}(\mathbf{W_g^w}\boldsymbol{s}_t).$$

$$\boldsymbol{M}_{r,t}^I = \boldsymbol{g}_t^r \otimes \boldsymbol{M}_{e,t}^I,$$
$$\boldsymbol{M}_{e,t+1}^I = \boldsymbol{g}_t^w \otimes \boldsymbol{M}_{e,t}^I,$$

$$\boldsymbol{s}_t = \mathbf{GRU}(\boldsymbol{s}_{t-1}, [\boldsymbol{c}_t; \boldsymbol{e}(y_{t-1}); \boldsymbol{M}_{r,t}^I]).$$

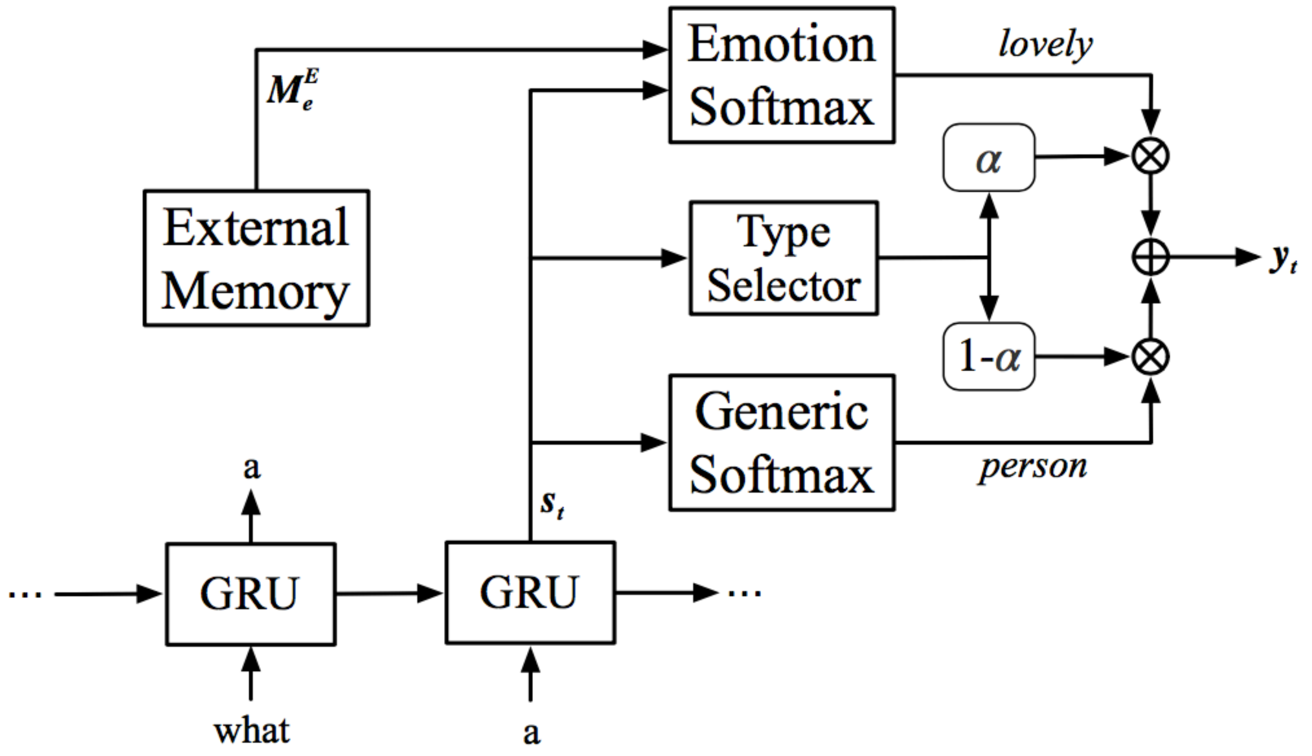
Model

- **Emotion external memory**: generic words (person) and emotion words (lovely)



Model

- **Emotion external memory**: generic words (person) and emotion words (lovely)



$$
\begin{aligned}
\alpha_t &= \mathrm{sigmoid}(\mathbf{v_u}^\top \boldsymbol{s}_t), \\
P_g(y_t = w_g) &= \mathrm{softmax}(\mathbf{W_g^o} \boldsymbol{s}_t), \\
P_e(y_t = w_e) &= \mathrm{softmax}(\mathbf{W_e^o} \boldsymbol{s}_t), \\
y_t \sim \boldsymbol{o}_t = P(y_t) &= \begin{bmatrix} (1-\alpha_t) P_g(y_t = w_g) \\ \alpha_t P_e(y_t = w_e) \end{bmatrix}
\end{aligned}
$$

Experiments

- **Emotion Classification Dataset**: the Emotion Classification Dataset of NLPCC 2013&2014
  - 23,105 sentences collected from Weibo
- **STC dataset**: a conversation dataset from (Shang et al., 2015)
  - 219,905 posts and 4,308,211 responses
  - Each post has about 20 responses

Experiments

Method	Perplexity	Accuracy
Seq2Seq	68.0	0.179
Emb	62.5	0.724
ECM	65.9	**0.773**
w/o Emb	66.1	0.753
w/o IMem	66.7	0.749
w/o EMem	**61.8**	0.731

Table 1. Automatic Evaluation

Pref. (%)	Seq2Seq	Emb	ECM
Seq2Seq	-	38.8	38.6
Emb	60.2	-	43.1
ECM	**61.4**	**56.9**	-

Table 3. Preference Test

Method	Overall		Like		Sad		Disgust		Angry		Happy	
	Cont.	Emot.	Cont.	Emot.	Cont.	Emot.	Cont.	Emot.	Cont.	Emot.	Cont.	Emot.
Seq2Seq	1.255	0.152	1.308	0.337	1.270	0.077	**1.285**	0.038	**1.223**	0.052	1.223	0.257
Emb	1.256	0.363	1.348	0.663	1.337	0.228	1.272	0.157	1.035	0.162	1.418	0.607
ECM	**1.299**	**0.424**	**1.460**	**0.697**	**1.352**	**0.313**	1.233	**0.193**	0.98	**0.217**	**1.428**	**0.700**

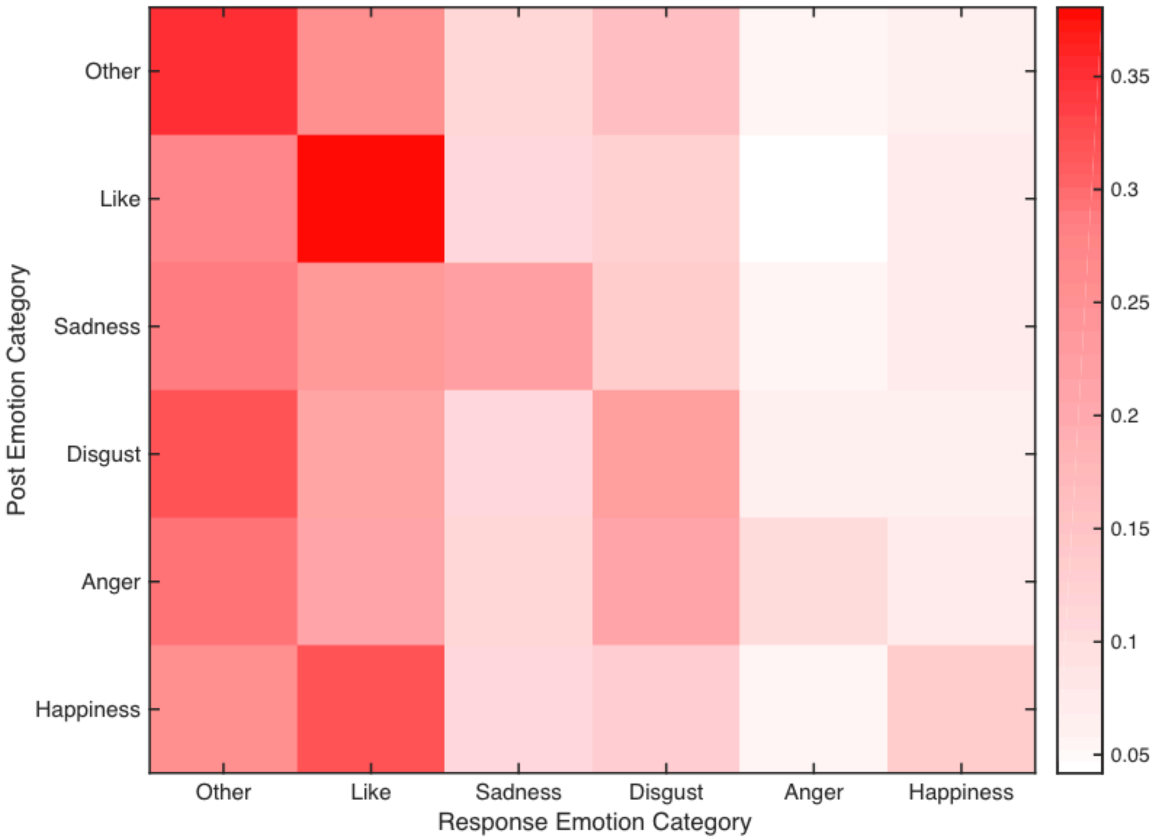
Table 2. Manual Evaluation

Experiments

Post	Model		Response(Original)	Response(Translated)
承诺，就是一个骗子说给一个傻子听的。 A promise is what a cheater says to a fool.	Seq2Seq		承诺，就是一个骗子说给一个傻子听的。	A promise is what a cheater says to a fool.
	ECM	Like	承诺是一个**专一**的傻子	Promise is a **single-minded** fool
		Happy	我就是一个**开心**的傻子	I'm just a **happy** fool
		Sad	我就是一个**悲伤**的傻子	I am a **sad** fool
		Disgust	一个**骗子**说给一个**傻子**听	A **cheater** says to a **fool**
		Angry	不管你信不信，反正我信了！	Whatever you believe or not, I trust in it!
知道真相的我眼泪笑下来 Knowing the truth, I laughed myself to tears	Seq2Seq		知道真相的我眼泪掉下来。	After knowing the truth, I burst into tears.
	ECM	Like	太**可爱**了，太**可爱**了！	Too **lovely**, it's too **lovely**.
		Happy	笑得我眼泪都出来了。	I can't help laugh myself to tears.
		Sad	知道真相的我**眼泪**掉下来。	After knowing the truth, I burst into **tears**.
		Disgust	这是真的还是假的？	Is this true?
		Angry	什么意思？没看懂。	What do you mean? I don't understand.

Experiments

Emotion Interaction Patterns (EIPs)



Like→Like **(empathy)**

Sadness →Sadness **(empathy)**

Sadness →Like **(comfort)**

Disgust → Disgust **(empathy)**

Disgust → Like **(comfort)**

Anger → Disgust

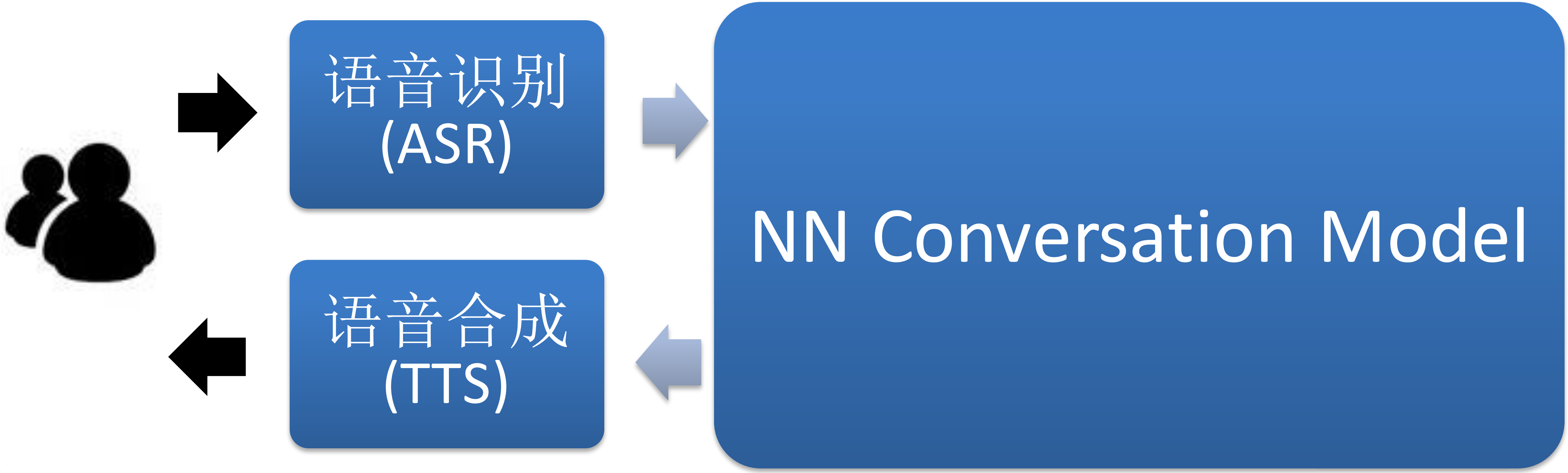
Happiness→Like

Conclusion

- It proposes an end-to-end framework (called ECM) to incorporate the emotion influence in large-scale conversation generation. It has three novel mechanisms: emotion category embedding, an internal emotion memory, and an external memory.
- It shows that ECM can generate responses with higher content and emotion scores than the traditional Seq2Seq model.

- We believe that future work such as the empathetic computer agent and the emotion interaction model can be carried out based on ECM.

Summary



Thanks for your attention!

Q&A