Towards Building More Intelligent Conversational Systems

Dr. Minlie Huang (黄民烈)

Associate Professor

CS Department, Tsinghua University

aihuang@tsinghua.edu.cn

http://coai.cs.tsinghua.edu.cn/hml

Research Picture of the CoAI Group

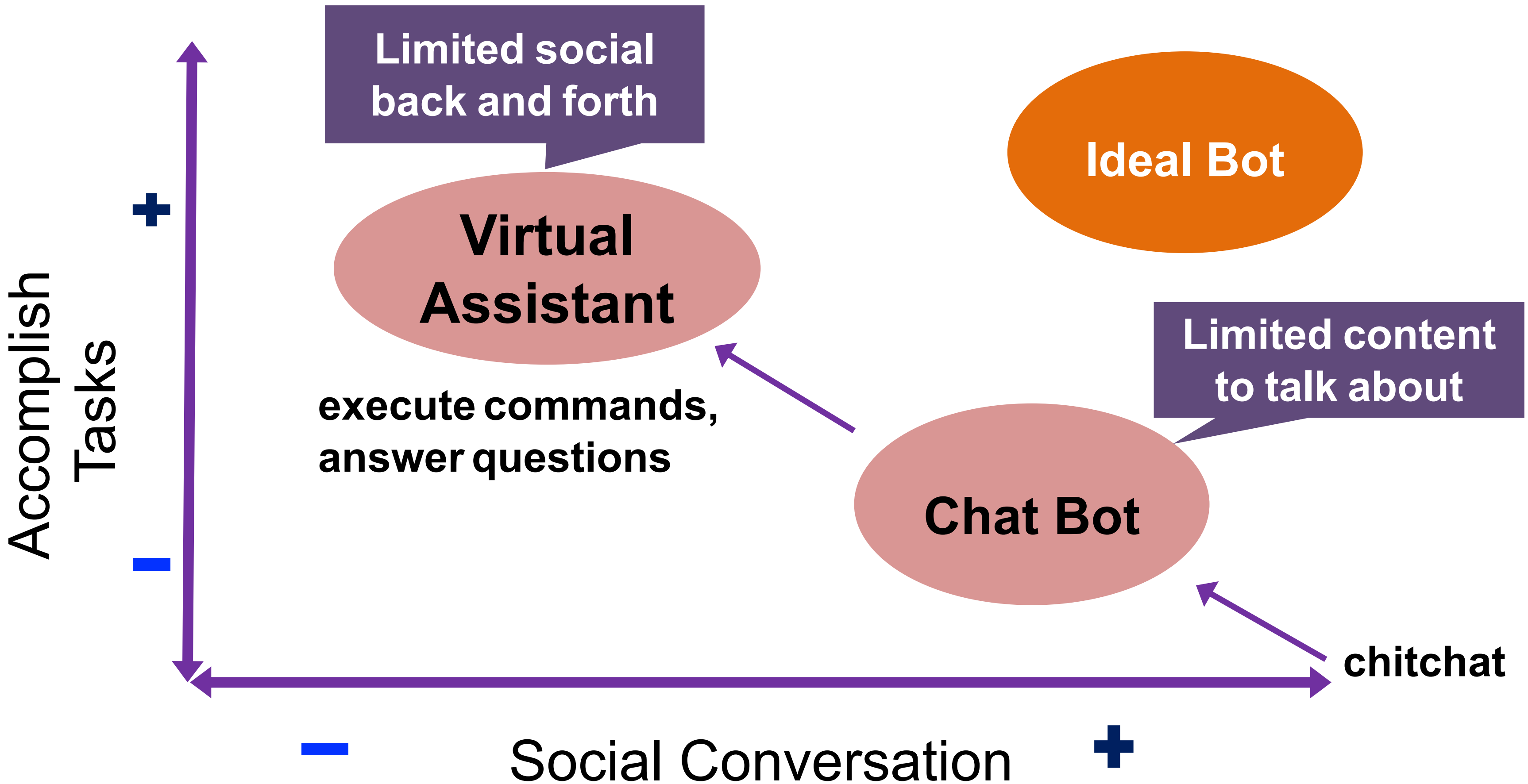
清華大學 Tsinghua University

Conversational AI

- Natural Language Understanding
- Deep Learning
- Reinforcement Learning

- Task-Oriented Dialogue
- Social Chatbot
- Question Answering
- Reading Comprehension
- Sentiment Analysis

Types of Conversational AI Systems



From prof. M Ostendorf, with edits

Issues Vary for Different Paradigms

ASSISTANT

Conversational AI System Components

SOCIALBOT

*Task intents
*Slot filling

*Narrow options & execute tasks
*Reward = **timely task completion**

Constrained domains

Structured KB+DB

Speech/language understanding

Dialog management

Language generation

Back-end application

*Social & info
*Intents
***Grounding**

*Learn about interests
*Make suggestions
*R=**user engagement**

Open domains

Unstructured Data & Info

From prof. M Ostendorf, with edits

Good Examples from DL Models

User：最幸福的人不是拿到了最好的东西，而是学会去珍惜手上已经拥有的人。

The happiest person is not the one who gets the best thing, instead who cherishes what she/he already has.

Machine: **是的，所以我也要珍惜自己的幸福**

Agree, I will cherish my happiness too.

User: **时间啊，你这把杀猪刀。。。**

Time is a killer…

Machine: **我是被杀的**

I am the one who was killed.

User: 怕孤獨，又捨不下自己的小自由。

Afraid to be lonely, but wont' give up my little freedom.

Machine: 我想去流浪，但没有人陪我去。

I want to wander, but no one accompany me.

Bad Examples are More!

User: 其实，他没有那么爱你。
Ohh, he does not love you that much.
Machine: 幸福的一天开始了。
A happy day starts.

User: 回北京了！很冷很高兴！
Back to Beijing now. So cold so happy!
Machine: 什么时候来北京啊？
When will you come to Beijing?

User: 我来自上海。
I come from Shanghai.
Machine： 真巧，我来自海南，老乡啊
So cool, I am from Hainan, the same province!

Controllability of Language Generation

- Three **fundamental problems** in current neural language generation models
  - **Semantics**
  - **Consistency** (long text generation)
  - **Logic** (reasonable and making sense)

Bad Examples are More!

Xiaoice by MSRA

微软小冰

语义理解的问题 Semantic understanding

User: 你家多大？ How large is your house?
小冰: 这么快就要爆年龄了？
So soon to reveal my age?

User: 你是素食主义者还是吃肉啊？
Are you vegan or meatatarian?
小冰: 怎么减肥
How can I lose weight?

上下文理解的问题 Context issue

User: 早饭不好吃
小冰: 那明早一起吃
User: 行，你推荐的应该不错
小冰: 是吗？
User: 哈哈，我相信你
小冰: 呃呃

个性身份一致性问题 Inconsistency in personality

User: 你几岁了？
小冰: 我啊。。。反正是九零后！
User: 那你是几几年出生的？
小冰: 错，75年阳历一月出生。
User: 你是九零后，怎么会75年出生？
小冰: 生在九零后，在深圳只能被当做八零后了。

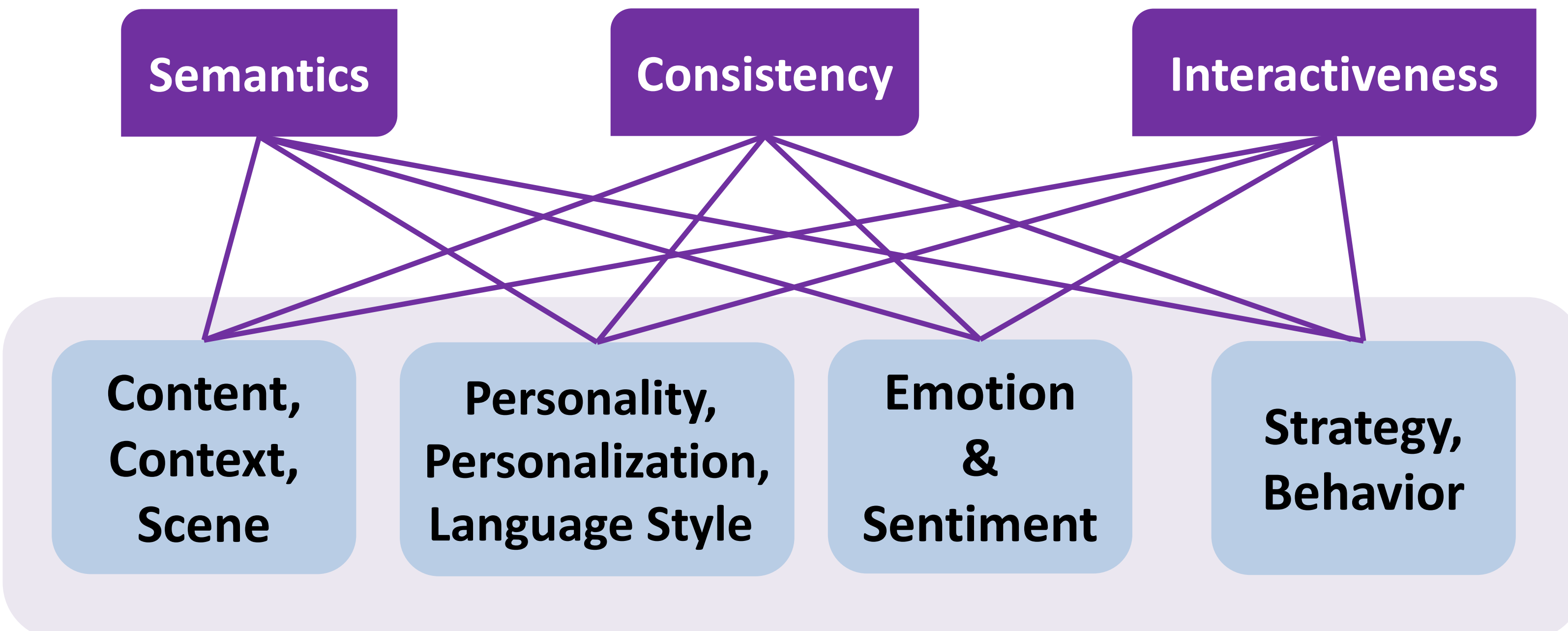
AI Ethics, a More Serious Issue

Picture from Prof. Frank Rudzicz, University of Toronto

Challenges in Conversational Systems

- **One-to-many**: one input, many many possible responses
- **Knowledge & Reasoning**: real understanding requires various **knowledge, world facts, commonsense**, etc.
- **Situational Context**
  - Who are you talking with?
    - Stranger, or friend?
  - His mood and emotion?
  - Shared backgrounds that are only accessible by two acquaintances

Challenges in Conversational Systems



Open-domain + Open-topic

Open-domain Conversational Systems

- Behaving more **interactively**:
  - Perceiving and Expressing **Emotions** (**AAAI 2018**)
  - Proactive Behavior by **Asking Good Questions** (**ACL 2018**)
  - Controlling **Sentence Function** (**ACL 2018**)
  - Topic Change (**SIGIR 2018**)
- Behaving more **consistently**:
  - **Explicit Personality** Assignment (**IJCAI-ECAI 2018**)
- Behaving more **intelligently** with **semantics**:
  - Better Understanding and Generation Using **Commonsense Knowledge** (**IJCAI-ECAI 2018 distinguished paper**)
  - **Discourse parsing** in multi-party dialogues (**AAAI 2019**)

Open-domain Conversational Systems

- Behaving more **interactively**:
  - Perceiving and Expressing **Emotions** (**AAAI 2018**)
  - Proactive Behavior by **Asking Good Questions** (**ACL 2018**)
  - Controlling **Sentence Function** (**ACL 2018**)
  - Topic Change (**SIGIR 2018**)
- Behaving more **consistently**:
  - **Explicit Personality** Assignment (**IJCAI-ECAI 2018**)

① Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. **AAAI 2018**.
② Assigning personality/identity to a chatting machine for coherent conversation generation. **IJCAI-ECAI 2018**.
③ Commonsense Knowledge Aware Conversation Generation with Graph Attention. **IJCAI-ECAI 2018**.
④ Learning to Ask Questions in Open-domain Conversational Systems with Typed Decoders. **ACL 2018**.
⑤ Generating Informative Responses with Controlled Sentence Function. **ACL 2018**.
⑥ Chat more: deepening and widening the chatting topic via a deep model. **SIGIR 2018.**
⑦ A Deep Sequential Model for Discourse Parsing on Multi-Party Dialogues. **AAAI 2019**.

Interactiveness:
Emotion Perception and Expression

Emotional Chatting Machine

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

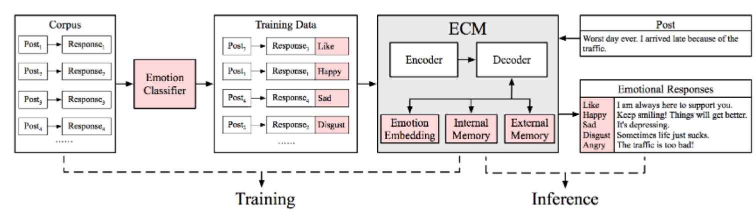
今天领导前来视察工作！
Happy: 哈哈，领导视察工作辛苦了
Like: 这是一个很有特色的领导人
Sad: 我想知道，谁来帮我干活？
Angry: 这是要干什么？领导？

Social Interaction Data

Post → Response
Post → Response
Post → Response
Post → Response

→ **Emotion Classifier** → Emotion Tagged data → **Emotional Chatting Machine**

Happy Angry Sad …



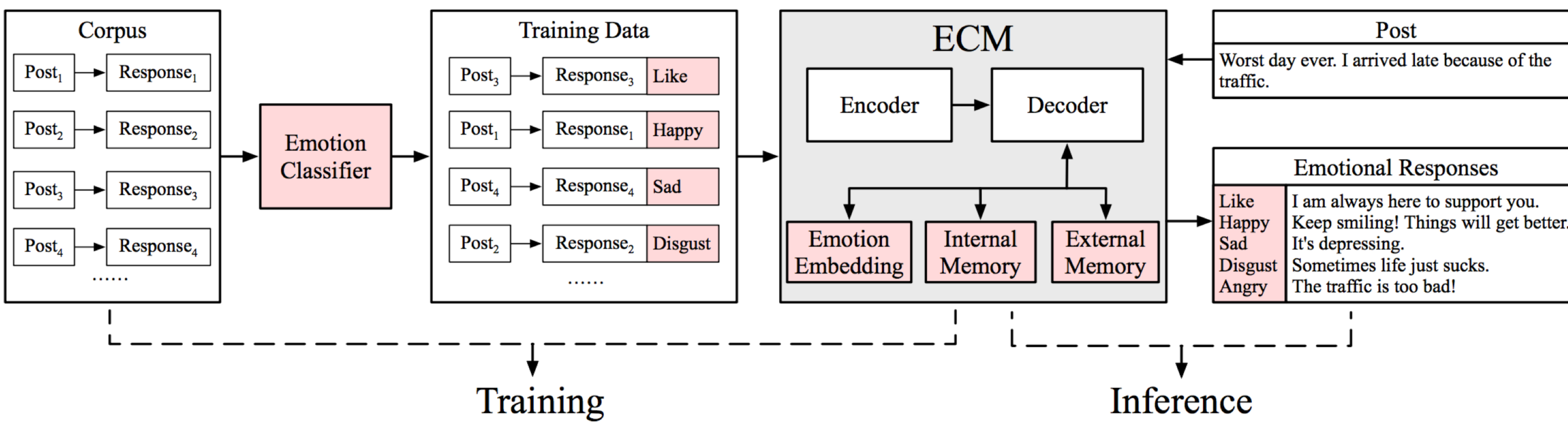
Our work was reported by **MIT Technology Review, the Guardian, Cankao News, Xinhua News Agency etc.**

Prof Björn Schuller: "an important step" towards personal assistants that could read the emotional undercurrent of a conversation and respond with something akin to empathy.

•Hao Zhou, Minlie Huang, Xiaoyan Zhu, Bing Liu. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. **AAAI 2018.**

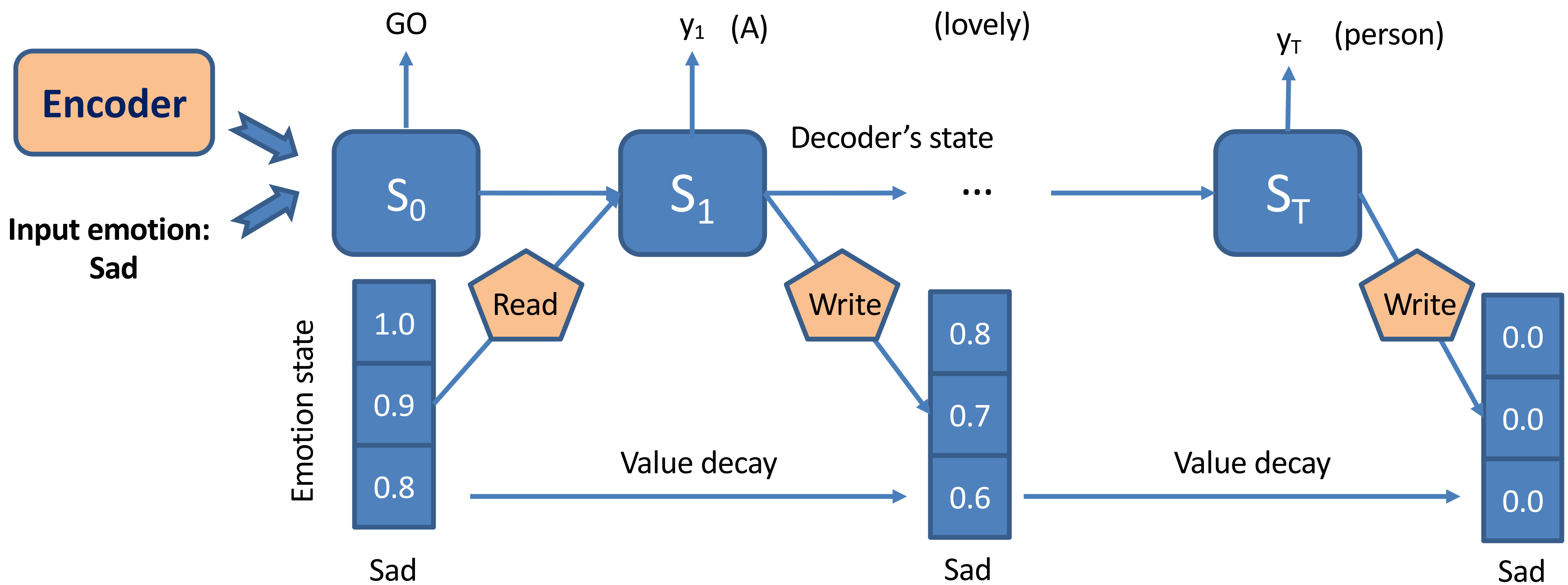
Emotional Chatting Machine

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



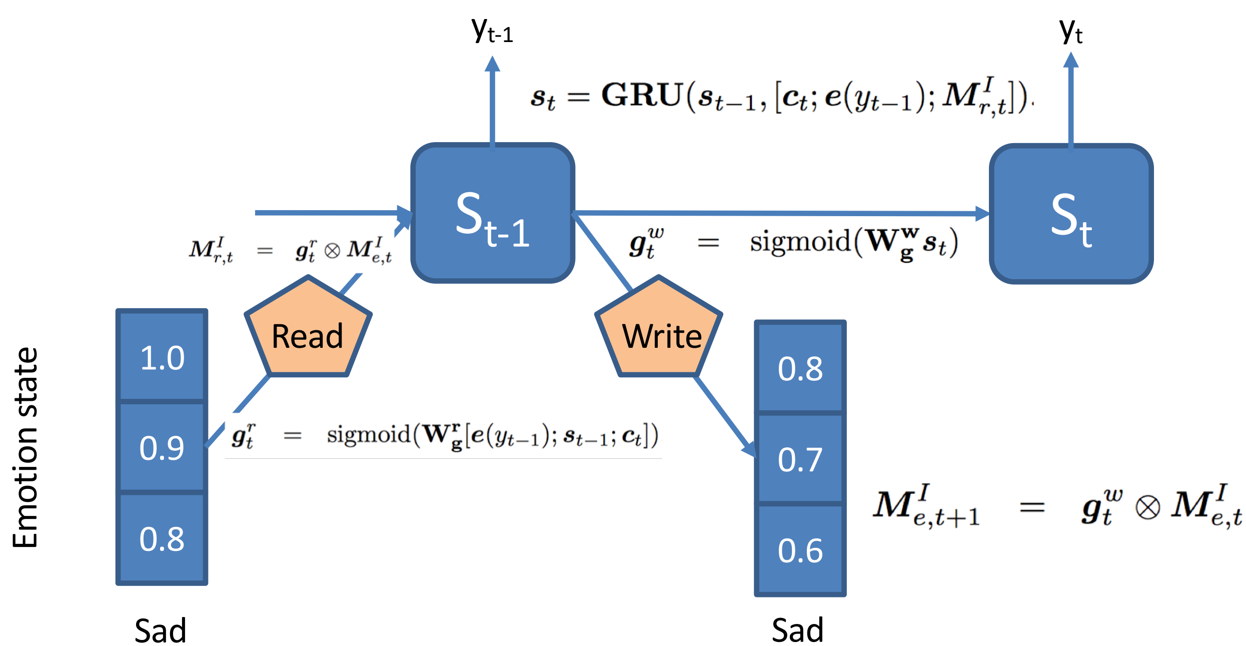
Emotional Chatting Machine

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



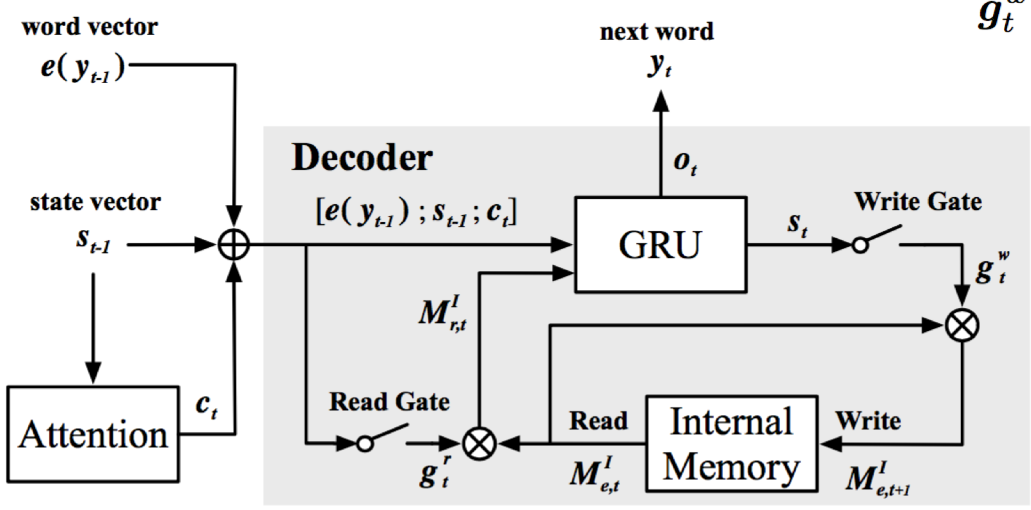
Emotional Chatting Machine

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



Emotional Chatting Machine

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



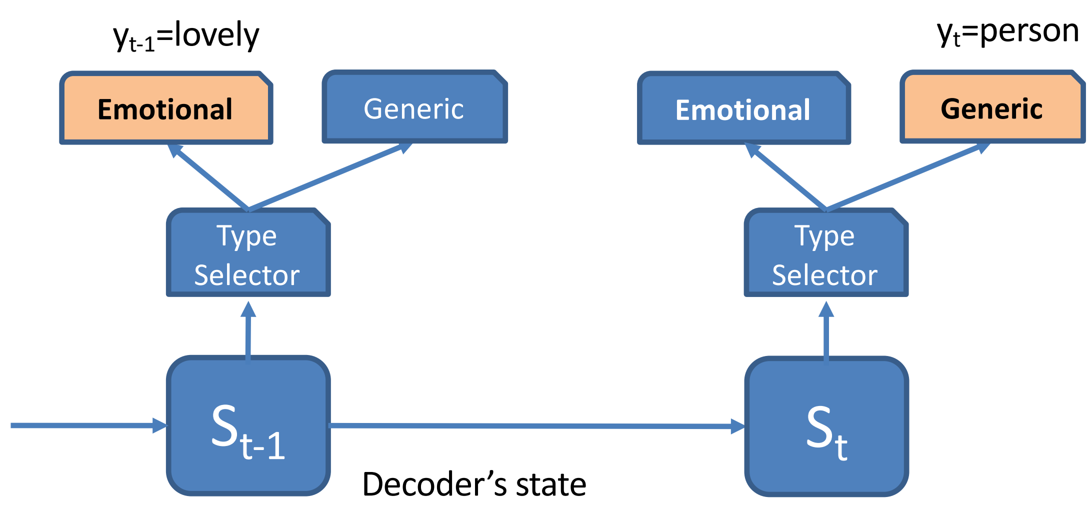
$$\boldsymbol{g}^r_t = \text{sigmoid}(\mathbf{W}^{\mathbf{r}}_{\mathbf{g}}[\boldsymbol{e}(y_{t-1}); \boldsymbol{s}_{t-1}; \boldsymbol{c}_t]),$$
$$\boldsymbol{g}^w_t = \text{sigmoid}(\mathbf{W}^{\mathbf{w}}_{\mathbf{g}}\boldsymbol{s}_t).$$

$$\boldsymbol{M}^I_{r,t} = \boldsymbol{g}^r_t \otimes \boldsymbol{M}^I_{e,t},$$
$$\boldsymbol{M}^I_{e,t+1} = \boldsymbol{g}^w_t \otimes \boldsymbol{M}^I_{e,t},$$

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

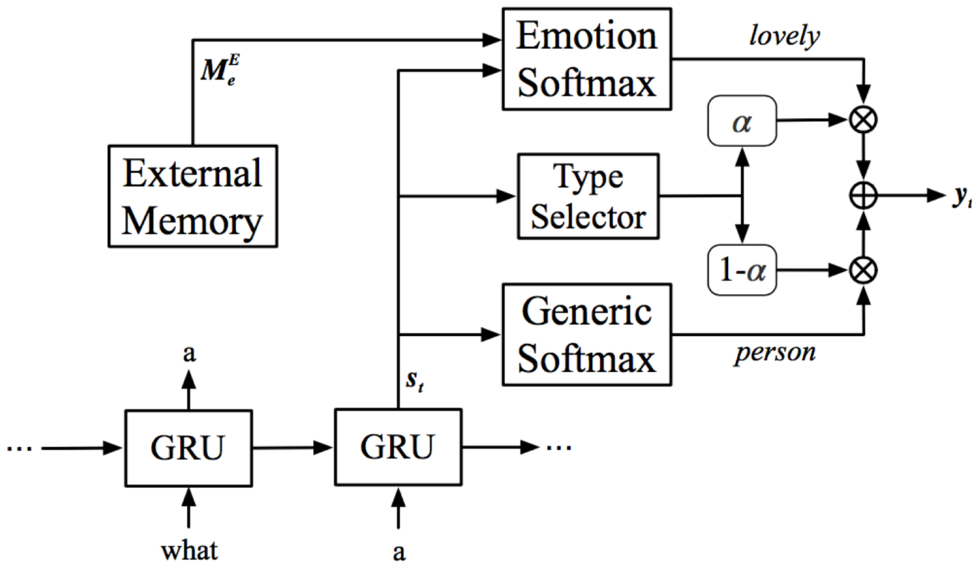
Emotional Chatting Machine

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



Emotional Chatting Machine

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



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

Emotional Chatting Machine

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

Emotional Chatting Machine

- Automatic Evaluation

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 4: Objective evaluation with perplexity and accuracy.

Emotional Chatting Machine

Method (%)	2-1	1-1	0-1	2-0	1-0	0-0
Seq2Seq	9.0	5.1	1.1	37.6	28.0	19.2
Emb	22.8	9.3	4.3	27.1	19.1	17.4
ECM	**27.2**	**10.8**	4.4	24.2	15.5	17.9

Score distribution

Table 5: The percentage of response
with the score of *Content-Emotion*. F
content score is 2 and emotion score i

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

Preference test

Table 7: Pairwise preference of the three systems.

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 6: Manual evaluation of the generated responses in terms of *Content* (Cont.) and *Emotion* (Emot.).

Emotional Chatting Machine

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.

•Hao Zhou, Minlie Huang, Xiaoyan Zhu, Bing Liu. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. **AAAI 2018.**

More Examples

情绪 emotion	Post:早上的天气不错啊！哈哈哈哈哈哈
Other	今天是个好天气。
喜欢 Like	今天的天气不错啊！
悲伤 Sad	**天气好，心情不好。**
厌恶 Disgust	这是哪里啊？求解释
愤怒 Angry	怎么了？什么情况？
快乐 Happy	今天的天气真好啊！

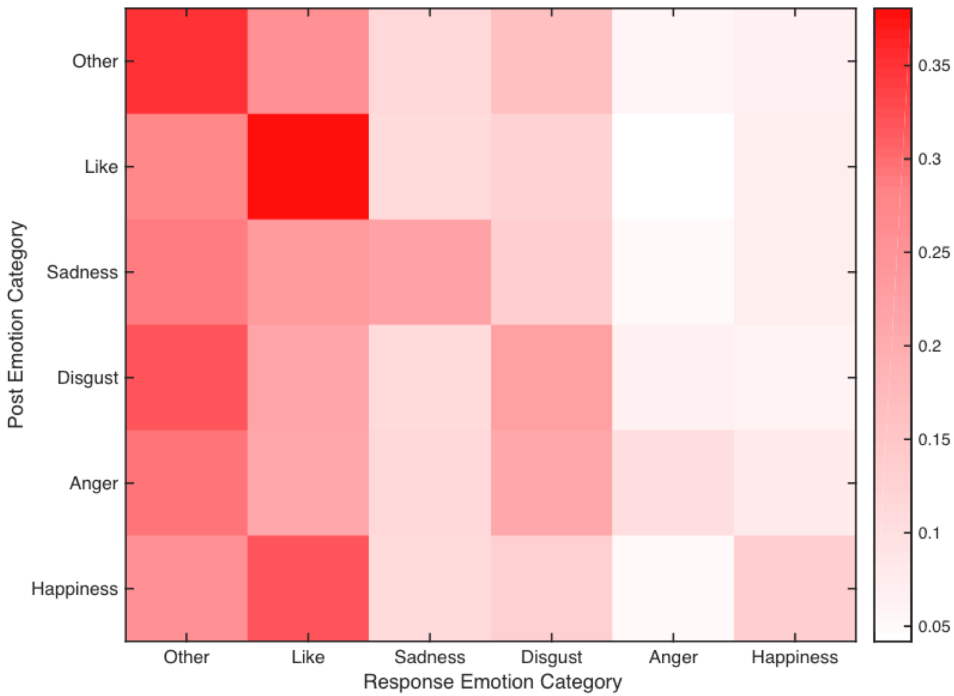
情绪 emotion	Post:送一个这样的蛋糕，给你美丽的闺蜜吧！
Other	这是蛋糕蛋糕吗？
喜欢 Like	好漂亮的蛋糕啊！
悲伤 Sad	**为什么我觉得很恶心**
厌恶 Disgust	**这样的蛋糕你忍心吗？**
愤怒 Angry	有没有想开网店的？
快乐 Happy	哈哈，好可爱的蛋糕！

情绪 emotion	Post:为了解决抄作业的不好习惯，高数老师这回要出绝招了！
Other	这老师是要逆天吗？
喜欢 Like	这老师太有才了！
悲伤 Sad	这样的老师伤不起啊
厌恶 Disgust	**这老师也太狠了点吧**
愤怒 Angry	**这老师是要逆天啊！**
快乐 Happy	哈哈，这老师太有才了

情绪 emotion	post:承诺，就是一个骗子说给一个傻子听的。
Other	承诺就是一个承诺
喜欢 Like	承诺是一个专一的傻子
悲伤 Sad	**我就是一个悲伤的傻子**
厌恶 Disgust	一个骗子说给一个傻子听
愤怒 Angry	不管你信不信，反正我信了！
快乐 Happy	**我就是一个开心的傻子**

•Hao Zhou, Minlie Huang, Xiaoyan Zhu, Bing Liu. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. **AAAI 2018.**

Emotion Interaction Patterns



Like→Like **(empathy)**

Sadness →Sadness **(empathy)**

Sadness →Like **(comfort)**

Disgust → Disgust **(empathy)**

Disgust → Like **(comfort)**

Anger → Disgust

Happiness→Like

•Hao Zhou, Minlie Huang, Xiaoyan Zhu, Bing Liu. Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. **AAAI 2018.**

Interactiveness:

Behaving More Proactively

by

Asking Good Questions

Asking Questions in Conversational Systems

我昨天晚上去聚餐了

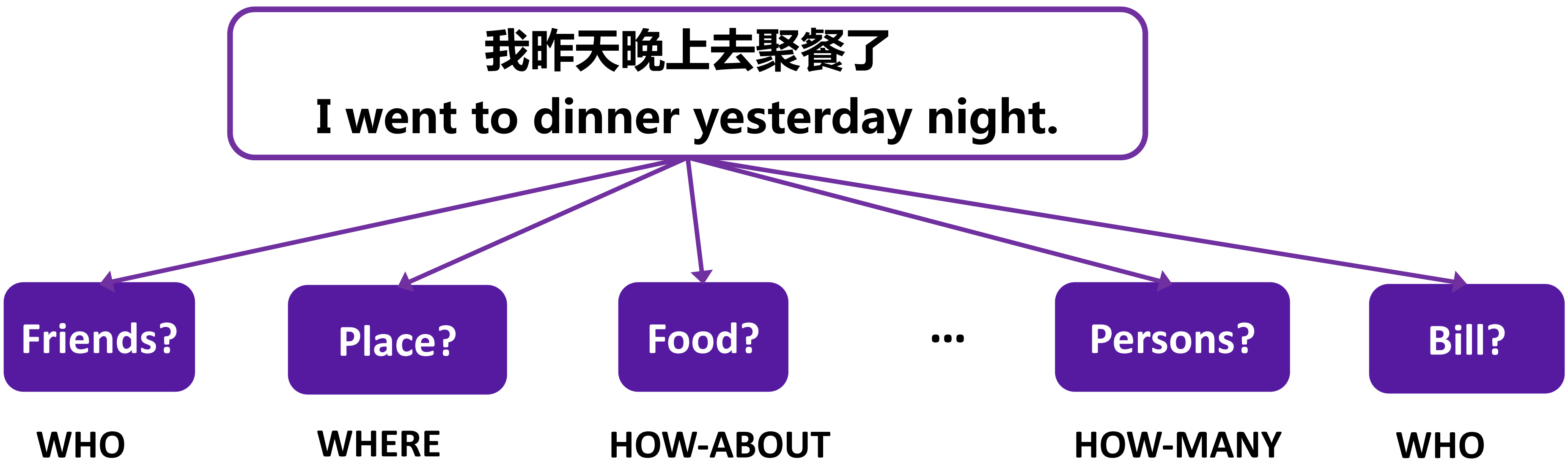
I went to dinner yesterday night.

Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie.
Learning to ask questions in open-domain conversation systems. **ACL 2018**.

Asking Questions in Conversational Systems

- Asking **good** questions requires **scene understanding**

Scene: Dining at a restaurant



Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie.
Learning to ask questions in open-domain conversation systems. **ACL 2018**.

Asking Questions in Conversational Systems

- Responding + **asking** (Li et al., 2016)
- **Key proactive** behaviors (Yu et al., 2016)
- Asking good questions are indication of **machine understanding**
- Key differences to **traditional** question generation (eg., reading comprehension):
  - **Different goals**: Information seeking vs. Enhancing interactiveness and persistence of human-machine interactions
  - **Various patterns**: YES-NO, WH-, HOW-ABOUT, etc.
  - **Topic transition**: from topics in post to topics in response

Asking Questions in Conversational Systems

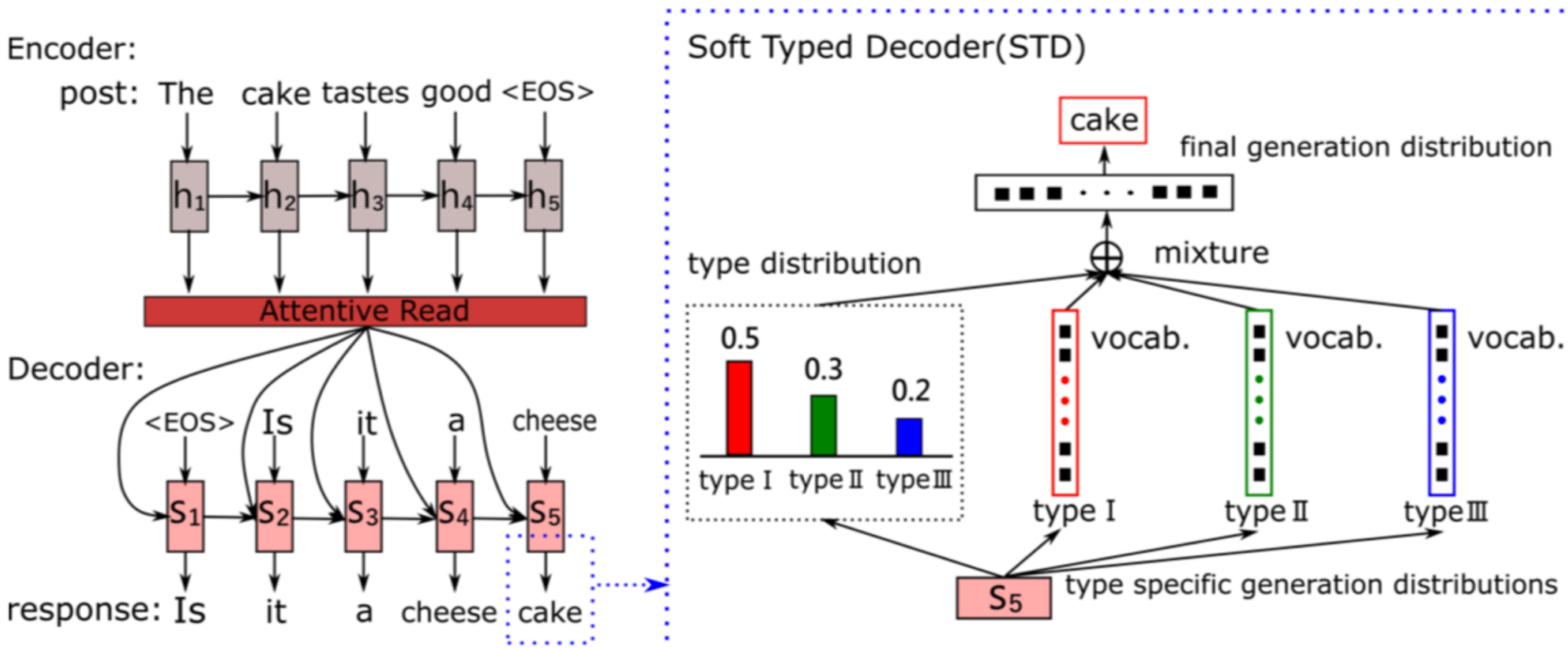
- A good question is a natural composition of
  - **Interrogatives** for using various questioning patterns
  - **Topic words** for addressing interesting yet novel topics
  - **Ordinary words** for playing grammar or syntactic roles

Example 1:
User: I am too fat ...
Machine: **How about** climbing this weekend?

Example 2:
User: Last night, I stayed in KTV with friends.
Machine: **Are** you happy with your singing?

Asking Good Questions

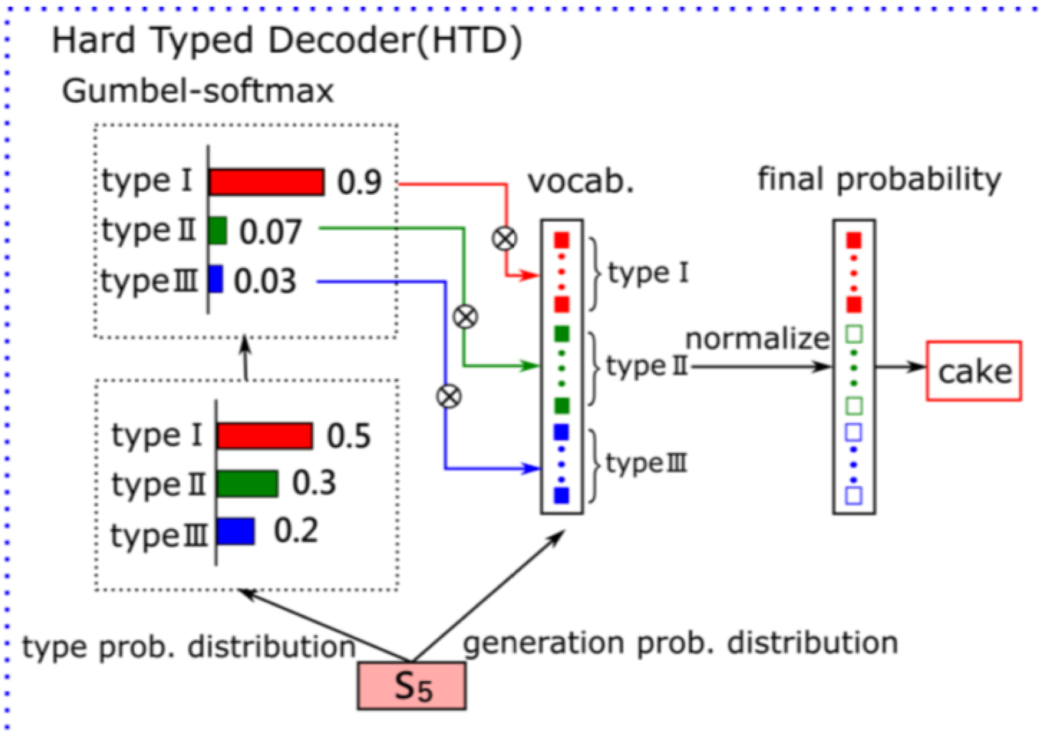
- **Typed decoders**: soft typed decoder



Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie.
Learning to ask questions in open-domain conversation systems. **ACL 2018**.

Asking Good Questions

- **Typed decoders**: hard typed decoder



For each post:

- A set of interrogatives
- A list of topic words
- Others for ordinary words

Topic words:

- Training -- nouns, verbs
- Test – predicted by PMI

Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie.
Learning to ask questions in open-domain conversation systems. **ACL 2018**.

Asking Good Questions

- Type prediction at each decoding position

Post: 我喜欢小动物(I like little animals)

Response:	你(you)	喜欢(like)	兔子(rabbit)	吗(*particle*)	?	_EOS
Interrogative	0.09	0.02	0.01	0.85	1.00	0.01
Topic word	0.26	0.35	0.71	0.14	0.00	0.02
Ordinary word	0.65	0.63	0.28	0.01	0.00	0.97
Decoding steps	1	2	3	4	5	6

Datasets

- Dataset: 490,000 post-response pairs collected from Weibo; 5,000 for test, 5000 for validation
  - All responses are of questioning form
- 66,547 different words, and 18,717 words appear more than 10 times

Baselines

- **Seq2Seq:** A simple encoder-decoder model (Luong et al., 2015)
- **Mechanism-Aware (MA):** Multiple responding mechanisms represented by real-valued vectors (Zhou et al., 2017)
- **Topic-Aware (TA):** Topic Aware Model by incorporating topic words (Xing et al., 2017)
- **Elastic Responding Machine (ERM)**: Enhanced MA using reinforcement learning (Zhou et al., 2018)

Results

Model	Perplexity	Distinct-1	Distinct-2	TRR
Seq2Seq	63.71	0.0573	0.0836	6.6%
MA	**54.26**	0.0576	0.0644	4.5%
TA	58.89	0.1292	0.1781	8.7%
ERM	67.62	0.0355	0.0710	4.5%
STD	56.77	0.1325	0.2509	12.1%
HTD	56.10	**0.1875**	**0.3576**	**43.6%**

Table 1: Results of automatic evaluation.

Results

- Manual evaluation: Appropriateness, richness, willingness

Models	Appropriateness			Richness			Willingness		
	Win (%)	Lose (%)	Tie (%)	Win (%)	Lose (%)	Tie (%)	Win (%)	Lose (%)	Tie (%)
STD vs. Seq2Seq	42.0	38.6	19.4	37.2**	15.2	47.6	45.4*	38.6	16.0
STD vs. MA	39.6*	31.2	29.2	32.6**	16.8	50.6	49.4**	27.0	23.6
STD vs. TA	42.2	40.0	17.8	49.0**	5.4	45.6	47.6*	40.2	12.2
STD vs. ERM	43.4*	34.4	22.2	60.6**	13.2	26.2	43.2*	36.8	20.0
HTD vs. Seq2Seq	50.6**	30.6	18.8	46.0**	10.2	43.8	58.4**	33.2	8.4
HTD vs. MA	54.8**	24.4	20.8	45.0**	17.0	38.0	67.0**	18.0	15.0
HTD vs. TA	52.0**	38.2	9.8	55.0**	5.4	39.6	62.6**	31.0	6.4
HTD vs. ERM	64.8**	23.2	12.0	72.2**	8.4	19.4	56.6**	36.6	6.8
HTD vs. STD	52.0**	33.0	15.0	38.0**	26.2	35.8	61.8**	30.6	7.6

Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie.
Learning to ask questions in open-domain conversation systems. **ACL 2018**.

Examples

Post:	我们昨天去唱歌了 We went to karaoke yesterday.
Seq2Seq:	什么时候? **When?**
MA:	你怎么知道我是在问 **How** did you know I am questioning you?
TA:	什么? **What?**
ERM:	什么情况? **What** happened ?
STD:	去哪里唱歌? **Where** did you sing karaoke?
HTD:	你们几个人唱歌? **How many** people were singing with you?

Yansen Wang, Chenyi Liu, Minlie Huang, Liqiang Nie.
Learning to ask questions in open-domain conversation systems. **ACL 2018**.

Interactiveness:

Achieving Different Speaking Purposes
by
Controlling Sentence Function

Controlling Sentence Function

- Sentence function indicates different conversational purposes.

User: I' m really hungry now.

Interrogative — **Acquire further information from users (e.g. WHY, WHAT, …)**

Imperative — **Make requests, instructions or invitations (e.g. LET'S, PLEASE, …)**

Declarative — **Make statements to state or explain (e.g. AND, BUT, …)**

Generating Informative Responses with Controlled Sentence Function. **ACL 2018.**

Controlling Sentence Function

- Response with controlled sentence function requires a **global plan** of <u>*function-related*</u>, <u>*topic*</u> and <u>*ordinary*</u> words.

User: I'm really hungry now.

Interrogative: What did you have at breakfast?
(Acquire further information from users)

Imperative: Let's have dinner together !
(Make requests, instructions or invitations)

Declarative: Me, too. But you ate too much at lunch.
(Make statements to state or explain)

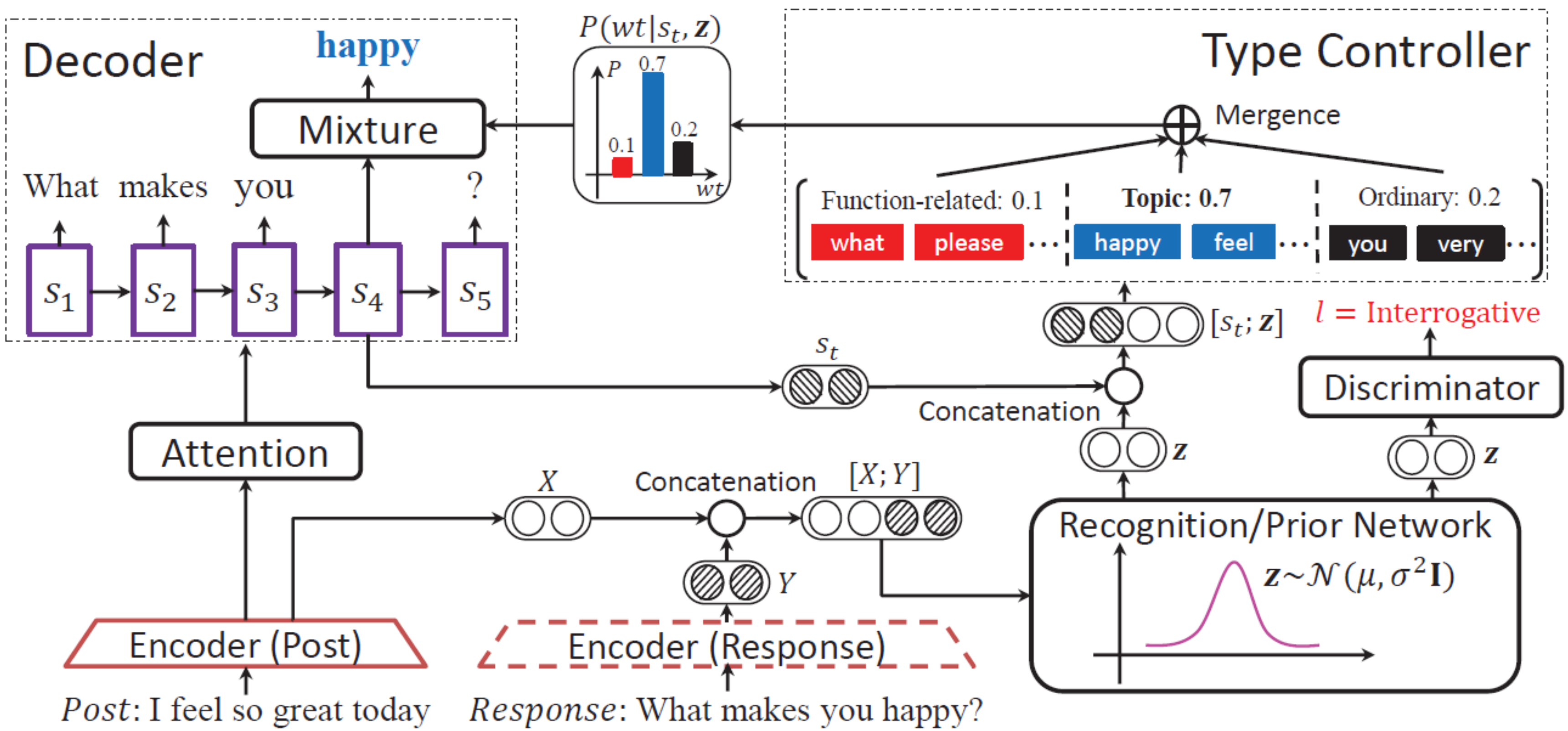
- Function-related words
- Topic words
- Ordinary words

Controlling Sentence Function

- Key differences to other controllable text generation tasks:
  - **Global Control**: adjust the global structure of the entire text, including changing word order and word patterns
  - **Compatibility**: controllable sentence function + informative content
- Solutions:
  - **Continuous Latent Variable**: project different sentence functions into different regions in a latent space + capture word patterns within a sentence function
  - **Type Controller**: arrange different types of words at proper decoding positions by estimating a distribution over three word types

Controlling Sentence Function

- Conditional Variational Autoencoder (CVAE) Framework



Controlling Sentence Function

- Dataset: post-response pairs with sentence function labels

Training	#Post	1,963,382	
	#Response	Interrogative	618,340
		Declarative	672,346
		Imperative	672,696
Validation	#Post	24,034	
	#Response	Interrogative	7,045
		Declarative	9,685
		Imperative	7,304
Test	#Post	6,000	

Pei Ke, Jian Guan, Minlie Huang, Xiaoyan Zhu.
Generating Informative Responses with Controlled Sentence Function. **ACL 2018.**

Controlling Sentence Function

- Automatic Evaluation: Perplexity, Distinct-1/2, Accuracy

Model	PPL	Dist-1	Dist-2	ACC
c-seq2seq	57.14	949/.007	5177/.041	0.973
MA	**46.08**	745/.005	2952/.027	0.481
KgCVAE	56.81	1531/**.009**	10683/.070	0.985
Our Model	55.85	**1833**/.008	**15586/.075**	**0.992**

Table 3: Automatic evaluation with perplexity (PPL), distinct-1 (Dist-1), distinct-2 (Dist-2), and accuracy (ACC). The integers in the Dist-* cells denote the total number of distinct n-grams.

Controlling Sentence Function

- Manual Evaluation: Grammaticality, Appropriateness, Informativeness

Model	Interrogative			Declarative			Imperative		
	Gram.	Appr.	Info.	Gram.	Appr.	Info.	Gram.	Appr.	Info.
Ours vs. c-seq2seq	0.534	0.536	0.896*	0.630*	0.573*	0.764*	0.685*	0.504	0.893*
Ours vs. MA	0.802*	0.602*	0.675*	0.751*	0.592*	0.617*	0.929*	0.568*	0.577*
Ours vs. KgCVAE	0.510	0.626*	0.770*	0.546*	0.515*	0.744*	0.780*	0.521*	0.837*

Table 4: Manual evaluation results for different functions. The scores indicate the percentages that our model wins the baselines after removing tie pairs. The scores of our model marked with * are significantly better than the competitors (Sign Test, *p-value* < 0.05).

Pei Ke, Jian Guan, Minlie Huang, Xiaoyan Zhu.
Generating Informative Responses with Controlled Sentence Function. **ACL 2018.**

Controlling Sentence Function

- Words and Patterns in Function Control

Function	Frequent Words		Frequent Patterns		Response Examples	
	Chinese	English	Chinese	English	Chinese	English
Interrogative	？ 是 吗 说 什么	? be particle mean what	x是说y吗？	Does x mean y?	你是说我帅吗？	Do you mean I'm handsome?
			x是在y吗？	Is x y?	你是在夸我吗？	Are you praising me?
			x在哪y啊？	Where does x y?	你在哪上班啊？	Where do you work?
			x想y什么z？	What z does x want to y?	你想要什么类型的？	What type do you want to choose?
Imperative	！ 要 可以 来 请	! will can come please	那就y吧	Do y, then.	那就好好养着吧	Take care of yourself, then.
			x要把y给z	Let x give y to z.	我要把你的房子给你	Let me give your house to you.
Declarative	是 也 觉得 可是 没	be also/too think but no	x也是y，可是z	x also y, but z	我也是这么想的，可是我要找一个人，哈哈	I also think so, but I will find a person. Ha-ha.
			x也是，a都b	x, too, and a has b.	我也是，我的粉丝都被我震精了	Me, too, and my fans have been shocked by me.

Figure 3: Frequent function-related words and frequent patterns containing at least 3 function-related words. The letters denote the variables which replace ordinary and topic words in the generated responses. The underlined words in responses are those occurring in patterns.

Controlling Sentence Function

- Case Study

Post:	如果有一天我突然跟你绝交，你会怎样？ What would you do if I suddenly broke up with you someday?
Target Function:	Interrogative
KgCVAE:	你是不是想我了 Did you miss me?
Our Model:	你说的是我的错吗？ Do you mean that it's my fault?
Target Function:	Imperative
KgCVAE:	我会告诉你的！ I will tell you!
Our Model:	我会把你的礼物给你的，放心 Rest assured that I would give your gift to you.
Target Function:	Declarative
KgCVAE:	我也觉得我是个好男人。 I think I'm a good man, too.
Our Model:	我觉得我会很傻，我会被我妈骂的 I would think that I was stupid and I would be blamed by my mother.

Post:	如果有一天我突然跟你绝交，你会怎样？ What would you do if I suddenly broke up with you someday?
Interrogative Response #1	你说的是我的错吗？ Do you mean that it's my fault?
Interrogative Response #2	你会不会说话？ Can you speak normally?
Interrogative Response #3	你想我怎样？我要不要绝交？ What do you think I should do? Shall I break up with you?

- Function-related words
- Topic words
- Ordinary words

Consistency:

Behaving More Consistently with Personality

Personality of Conversational Systems

- Personality is important for game, custom service, etc.

General seq2seq model
User: Are you a boy or a girl?
Chatbot: I am a boy.
User: Are you a girl?
Chatbot: Yes, I am a girl.

User: 你几岁了？
小冰: 我啊。。。反正是九零后！
User: 那你是几几年出生的？
小冰: 错，75年阳历一月出生。
User: 你是九零后，怎么会75年出生？
小冰: 生在九零后，在深圳只能被当做八零后了。

Personality of Conversational Systems

- Passing the **Turning Test**?
  - **Deep semantic understanding**
  - Existing chatting machine lacks **identity or personality**
- Personality is a well-defined concept in psychology(Norman, 1963; Gosling et al., 2003)
- Extremely **subtle, implicit** in language expression:
  - Age, gender, language, speaking style, level of knowledge, areas of expertise
- Existing works
  - **Implicit personalization**: learn **implicit** conversation style (Li et al., 2016; Al-Rfou et al., 2016)
  - Require dialogue data from different users with **user attributes tagged**

Personality of Conversational Systems

- Deliver coherent conversations w.r.t. **identity/personality**

Generic Dialogue Data for Training

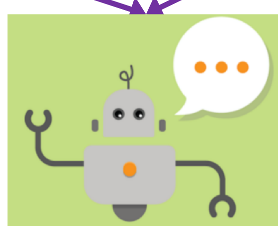
UserA: how old are you?
UserB: I am six.

UserA: do you like to play piano?
UserB: I play violin.

Pre-specified Chatbot Profile

Profile key	Profile value
Name	汪仔(Wang Zai)
Age	三岁(3)
Gender	男孩(Boy)
Hobbies	动漫(Cartoon)
Speciality	钢琴(Piano)

Personality-coherent Chatbot



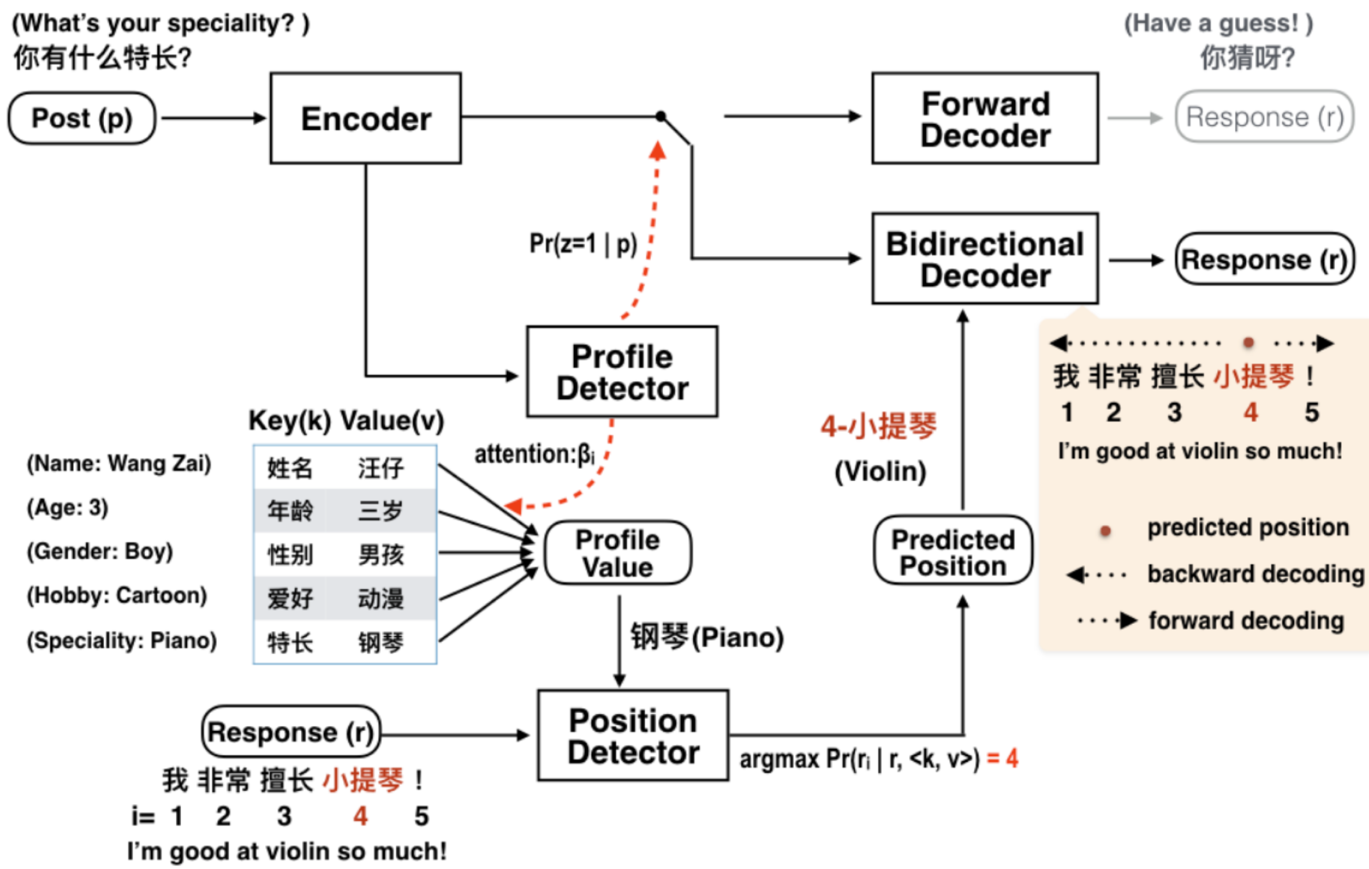
Generated Dialogues

User: how old are you?
Machine: I am three years old.

User: do you like to play piano?
Machine: Yes, I play piano.

Personality of Conversational Systems

- Encoder
- Profile detector
- Position detector
- Decoder



•Qiao Qian, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu. *Assigning personality/identity to a chatting machine for coherent conversation generation.* **IJCAI-ECAI 2018.**

Datasets

- WD: 9,697,651 post-response pairs from Weibo
- 76,930 pairs from WD for 6 profile keys (name, gender, age, city, weight, constellation) with about 200 regular expression patterns, each annotated to positive or negative
- 42,193 positive pairs, each mapped to one of the keys
- Manual Dataset: real, human-written conversational posts

Results

Post-level evaluation

Method	Nat.	Logic	Cor.
Seq2Seq	71.8%	56.0%	23.8%
Seq2Seq +PV	72.0%	56.0%	41.3%
Seq2Seq +PVD	73.3%	52.5%	38.0%
Our Model -PD	82.7%	51.7%	38.3%
Our Model	**83.3%**	**59.5%**	**45.8%**

Session-level evaluation

Method	Consistency	Variety
Seq2Seq	1.3%	1.0%
Seq2Seq +PV	47.0%	1.3%
Seq2Seq +PVD	40.0%	7.5%
Our Model -PD	38.8%	16.0%
Our Model	**49.5%**	**27.8%**

Generated sample responses that exhibit session-level consistency

Chinese	**English(Translated)**
U:你对什么事感兴趣	U:What are you interested in?
S:打篮球	S:Playing basketball.
U:你都有啥爱好呀	U:What's your hobby?
S:篮球	S:Basketball.
U:告诉我你兴趣所在	U:Tell me your interest.
S:我喜欢篮球	S:I like to play basketball.
U:你还没说你几岁呢	U:You haven't told me your age.
S:我三岁了	S:I'm three years old.
U:你今年有15了不	U:Are you 15 years old or not?
S:我还没到呢	S:I'm not yet.
U:你多大啦	U:How old are you?
S:3岁了	S:Three years old.

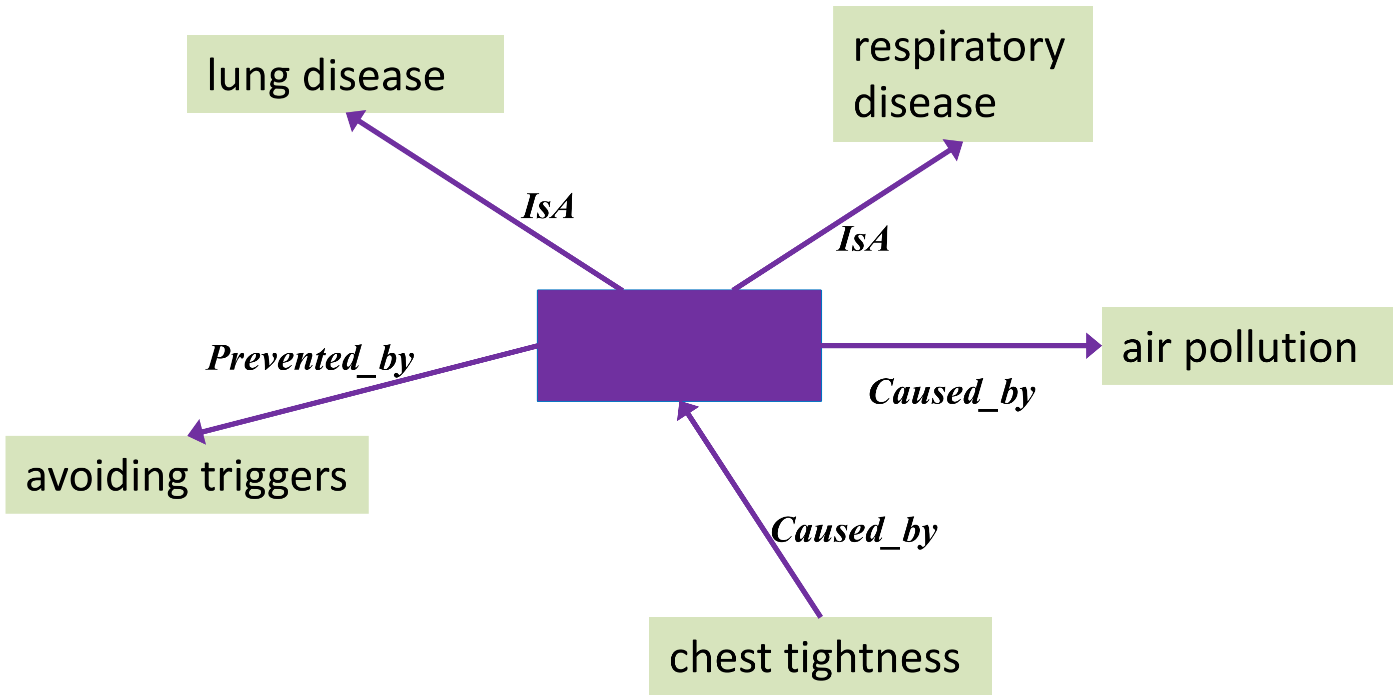
•Qiao Qian, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu. *Assigning personality/identity to a chatting machine for coherent conversation generation.* **IJCAI-ECAI 2018.**

Semantics:
Better Understanding and Generation with Commonsense Knowledge

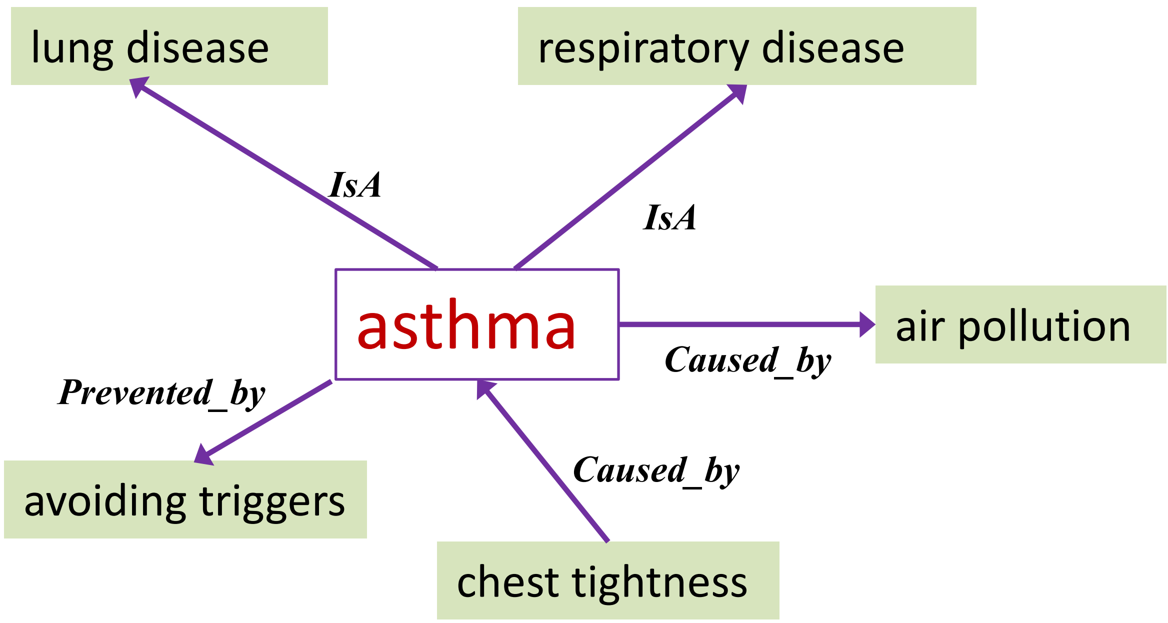
Commonsense Knowledge

- **Commonsense knowledge** consists of facts about the everyday world, that all humans are expected to know. (Wikipedia)
  - Lemons are sour
  - Tree has leafs
  - Dog has 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***?

Commonsense Knowledge



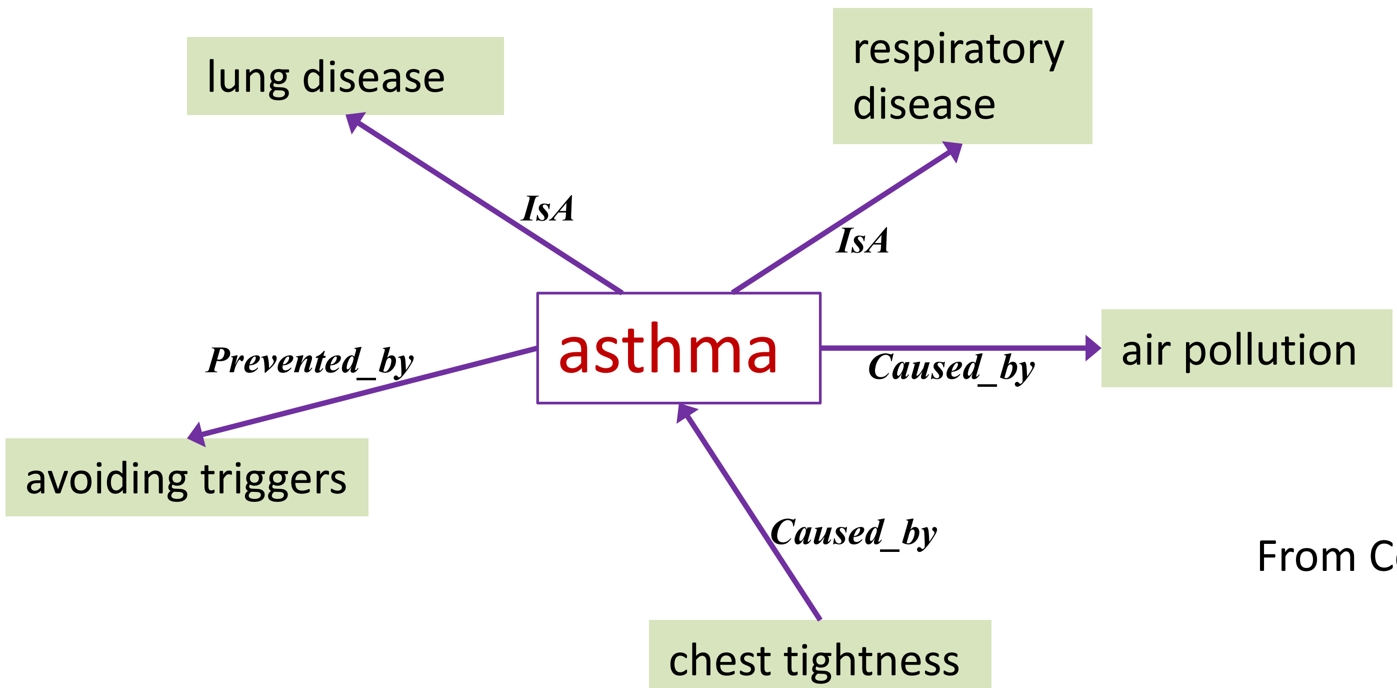
Commonsense Knowledge



Commonsense Knowledge

Post: I have an asthma since three years old.

Triples in knowledge graph:
(lung disease, IsA, asthma)
(asthma, Prevented_by, avoiding triggers)



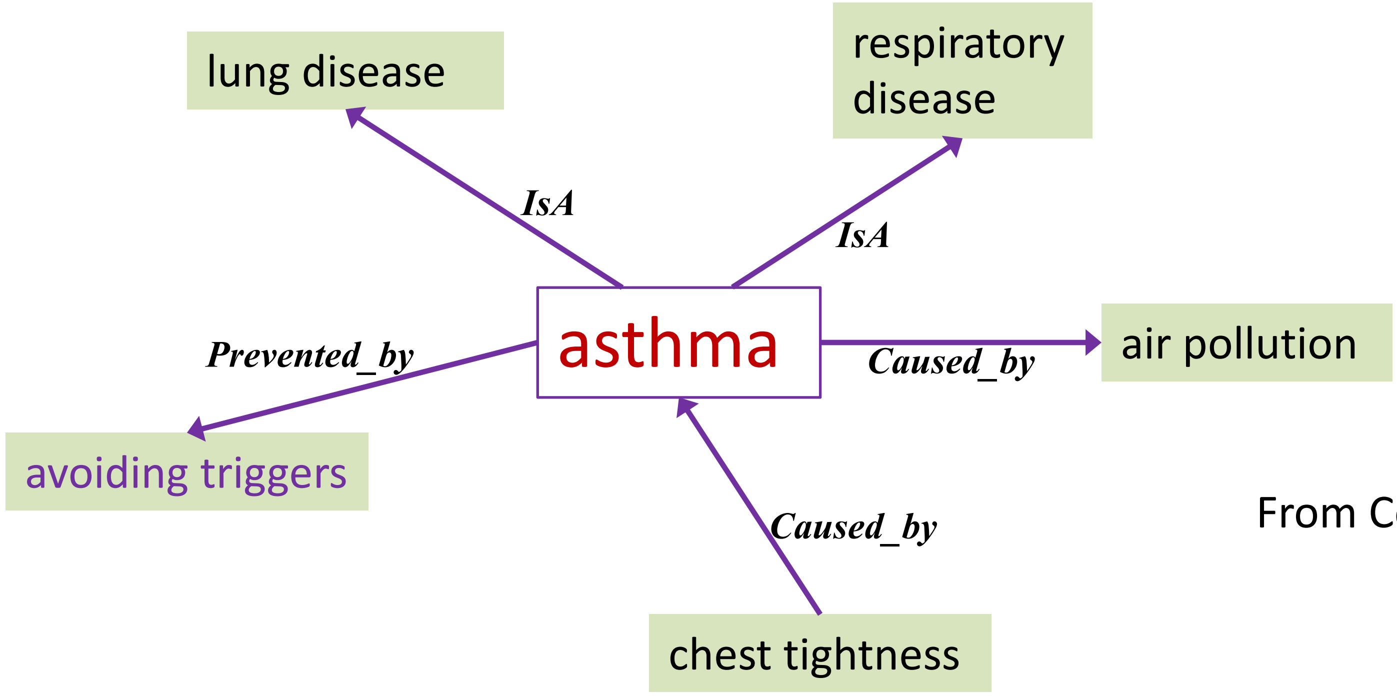
From ConceptNet

Commonsense Knowledge in Chatbots

Post: I have an asthma since three years old.

Triples in knowledge graph:
(lung disease, IsA, asthma)
(asthma, Prevented_by, avoiding triggers)

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



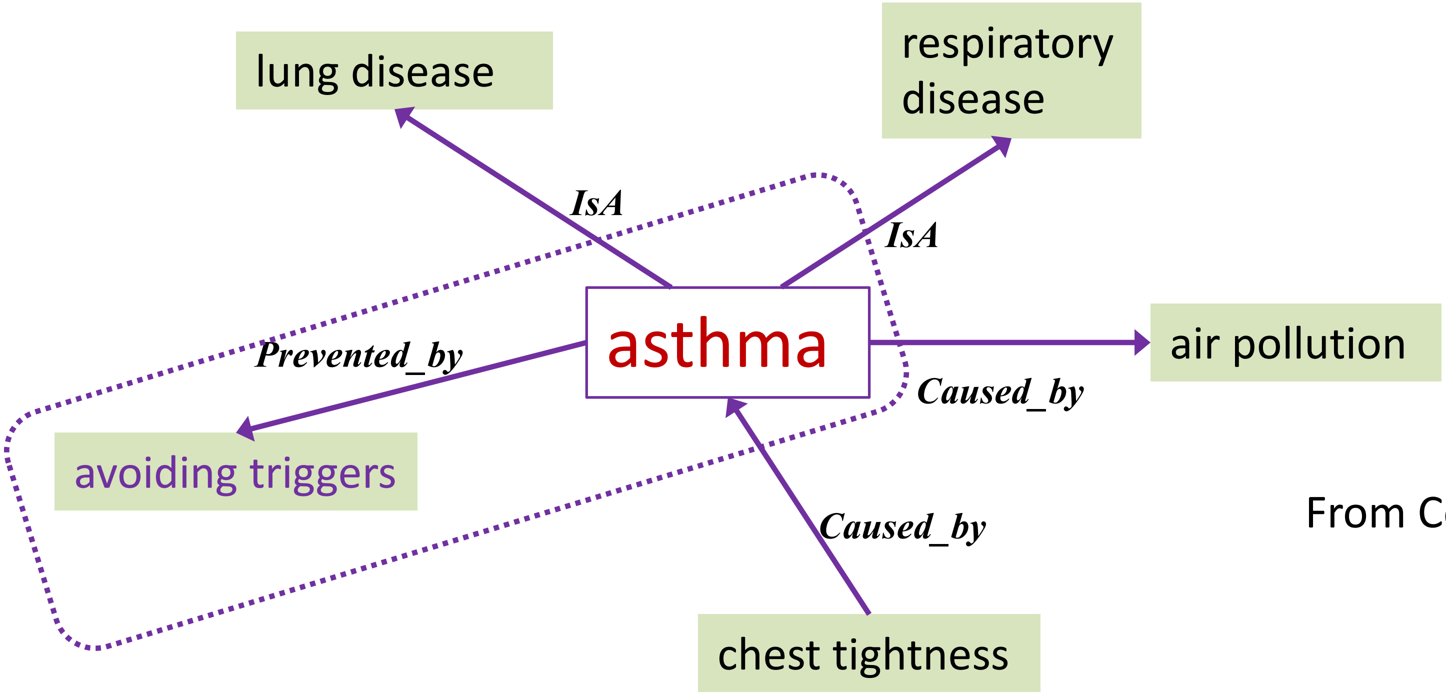
From ConceptNet

Commonsense Knowledge in Chatbots

Post: I have an asthma since three years old.

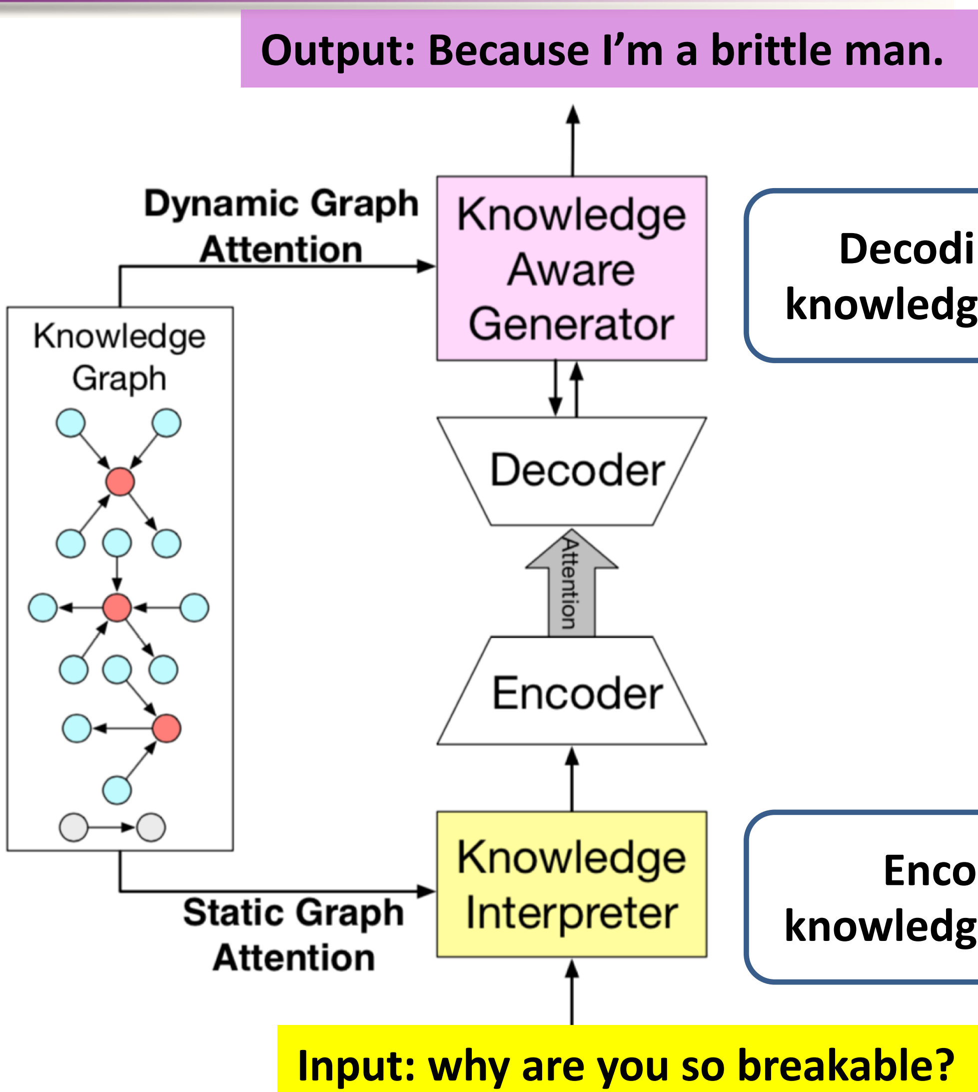
Triples in knowledge graph:
(lung disease, IsA, asthma)
(asthma, Prevented_by, avoiding triggers)

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



From ConceptNet

Commonsense Knowledge in Chatbots

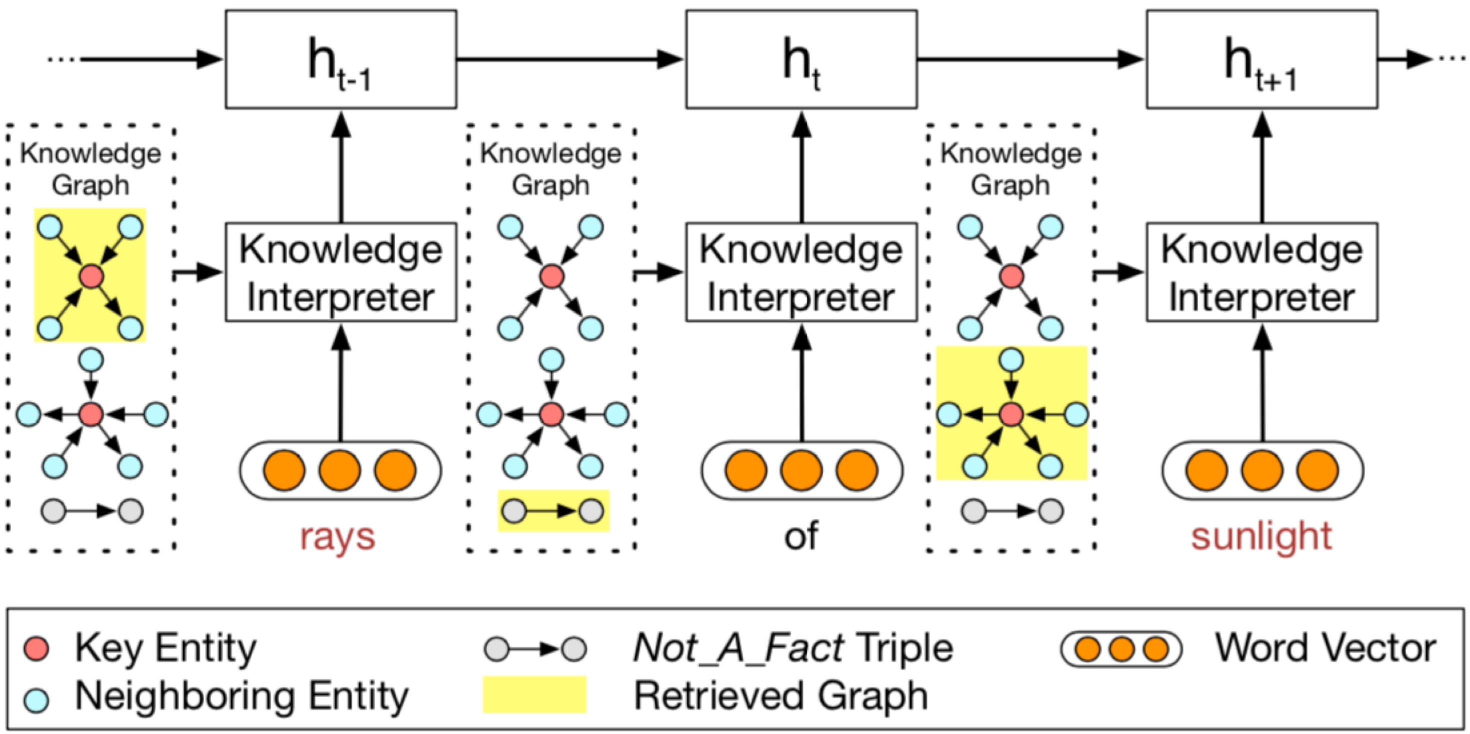


Decoding words by attending to knowledge graphs and then to triples

Encoding the retrieved knowledge graphs for each word

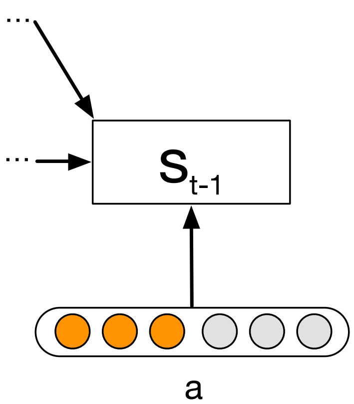
Commonsense Knowledge in Chatbots

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

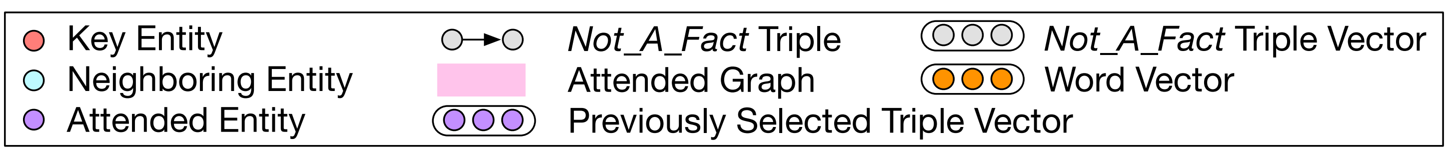


Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

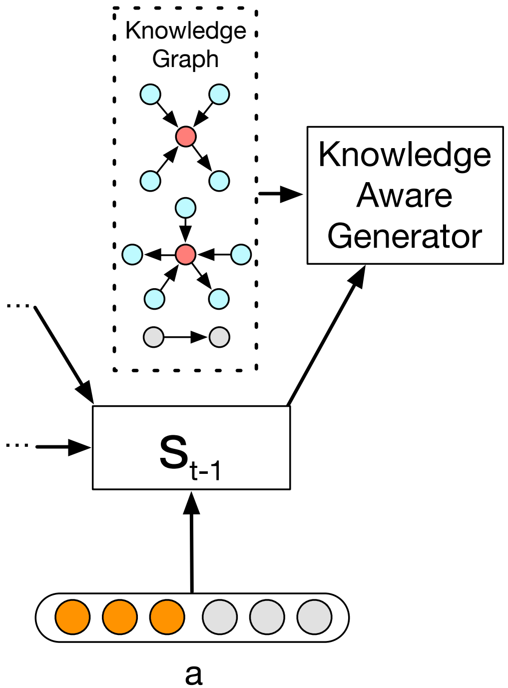


$$
\begin{aligned}
\boldsymbol{s}_{t+1} &= \mathbf{GRU}(\boldsymbol{s}_t, [\boldsymbol{c}_t; \boldsymbol{c}_t^g; \boldsymbol{c}_t^k; \boldsymbol{e}(y_t)]), \\
\boldsymbol{e}(y_t) &= [\boldsymbol{w}(y_t); \boldsymbol{k}_j],
\end{aligned}
$$



Commonsense Knowledge in Chatbots

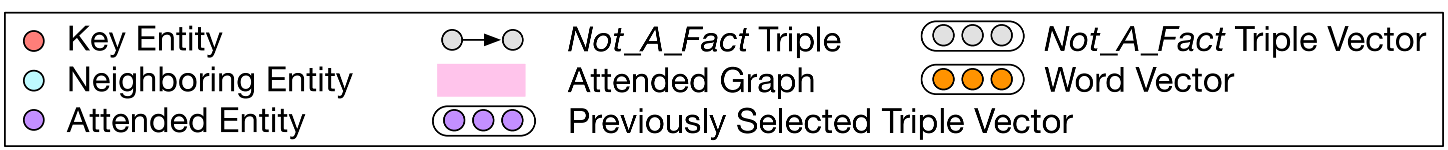
Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



$$\boldsymbol{g}_i = \sum_{n=1}^{N_{g_i}} \alpha_n^s [\boldsymbol{h}_n; \boldsymbol{t}_n],$$

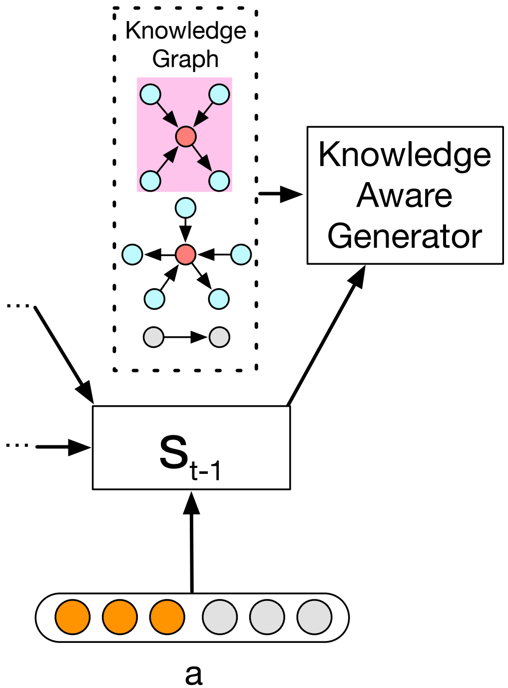
$$\alpha_n^s = \frac{\exp(\beta_n^s)}{\sum_{j=1}^{N_{g_i}} \exp(\beta_j^s)},$$

$$\beta_n^s = (\mathbf{W_r} \boldsymbol{r}_n)^\top \tanh(\mathbf{W_h} \boldsymbol{h}_n + \mathbf{W_t} \boldsymbol{t}_n),$$

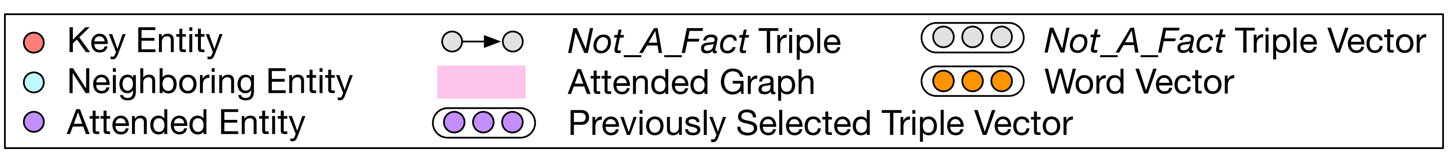


Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

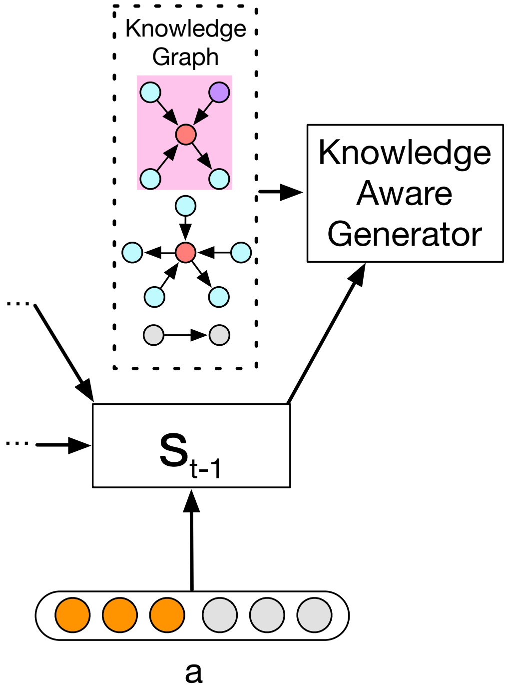


$$
\begin{aligned}
\boldsymbol{c}_t^g &= \sum_{i=1}^{N_G} \alpha_{ti}^g \boldsymbol{g}_i, \\
\alpha_{ti}^g &= \frac{\exp(\beta_{ti}^g)}{\sum_{j=1}^{N_G} \exp(\beta_{tj}^g)}, \\
\beta_{ti}^g &= \boldsymbol{V_b}^\top \tanh(\mathbf{W_b} \boldsymbol{s}_t + \mathbf{U_b} \boldsymbol{g}_i),
\end{aligned}
$$



Commonsense Knowledge in Chatbots

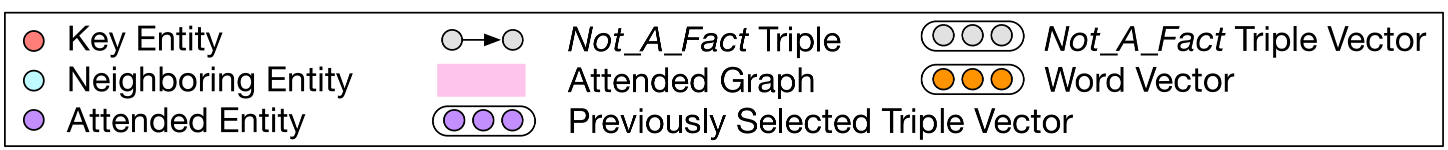
Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



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

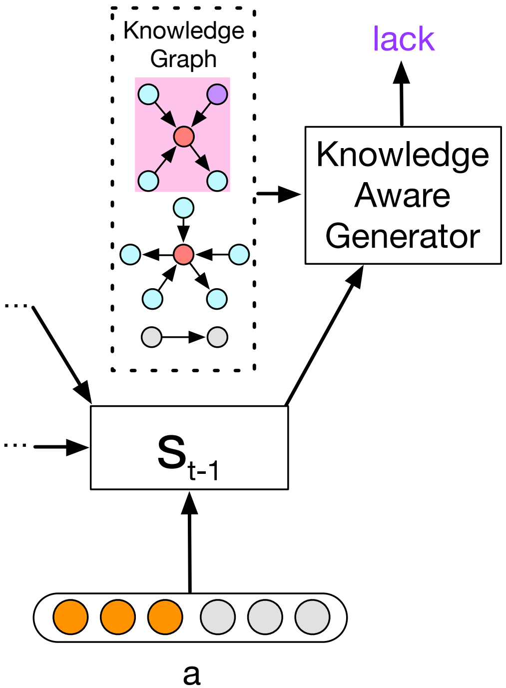
$$\alpha_{tj}^k = \frac{\exp(\beta_{tj}^k)}{\sum_{n=1}^{N_{g_i}} \exp(\beta_{tn}^k)},$$

$$\beta_{tj}^k = \mathbf{k}_j^\top \mathbf{W_c} \mathbf{s}_t,$$

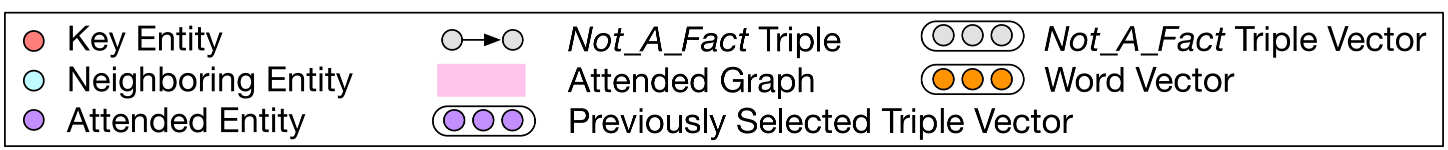


Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

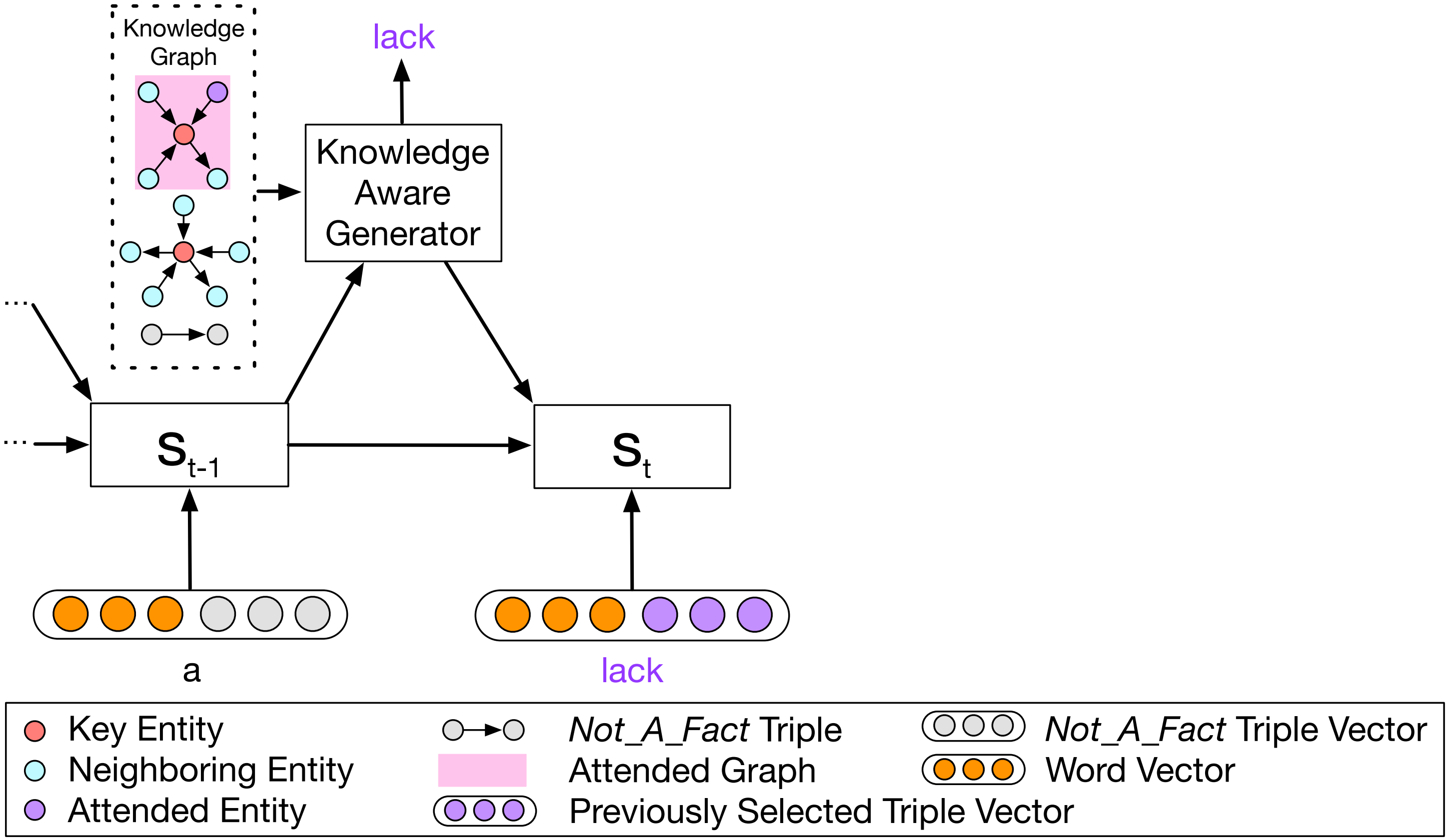


$$
\begin{aligned}
\boldsymbol{a}_t &= [\boldsymbol{s}_t; \boldsymbol{c}_t; \boldsymbol{c}_t^g; \boldsymbol{c}_t^k], \\
\gamma_t &= \mathrm{sigmoid}(\mathbf{V_o}^\top \boldsymbol{a}_t), \\
P_c(y_t = w_c) &= \mathrm{softmax}(\mathbf{W_o} \boldsymbol{a}_t), \\
P_e(y_t = w_e) &= \alpha_{ti}^g \alpha_{tj}^k,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



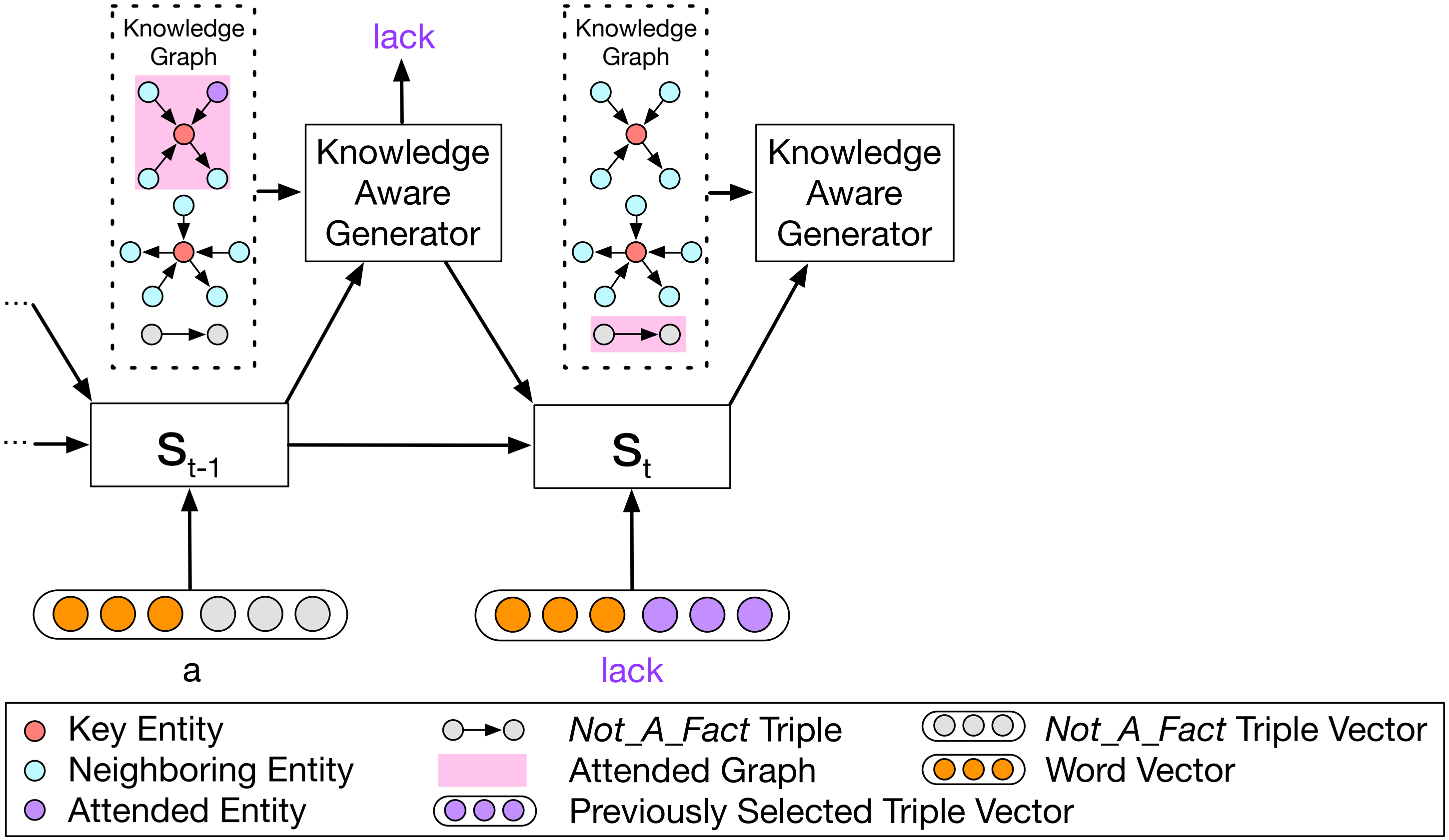
Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



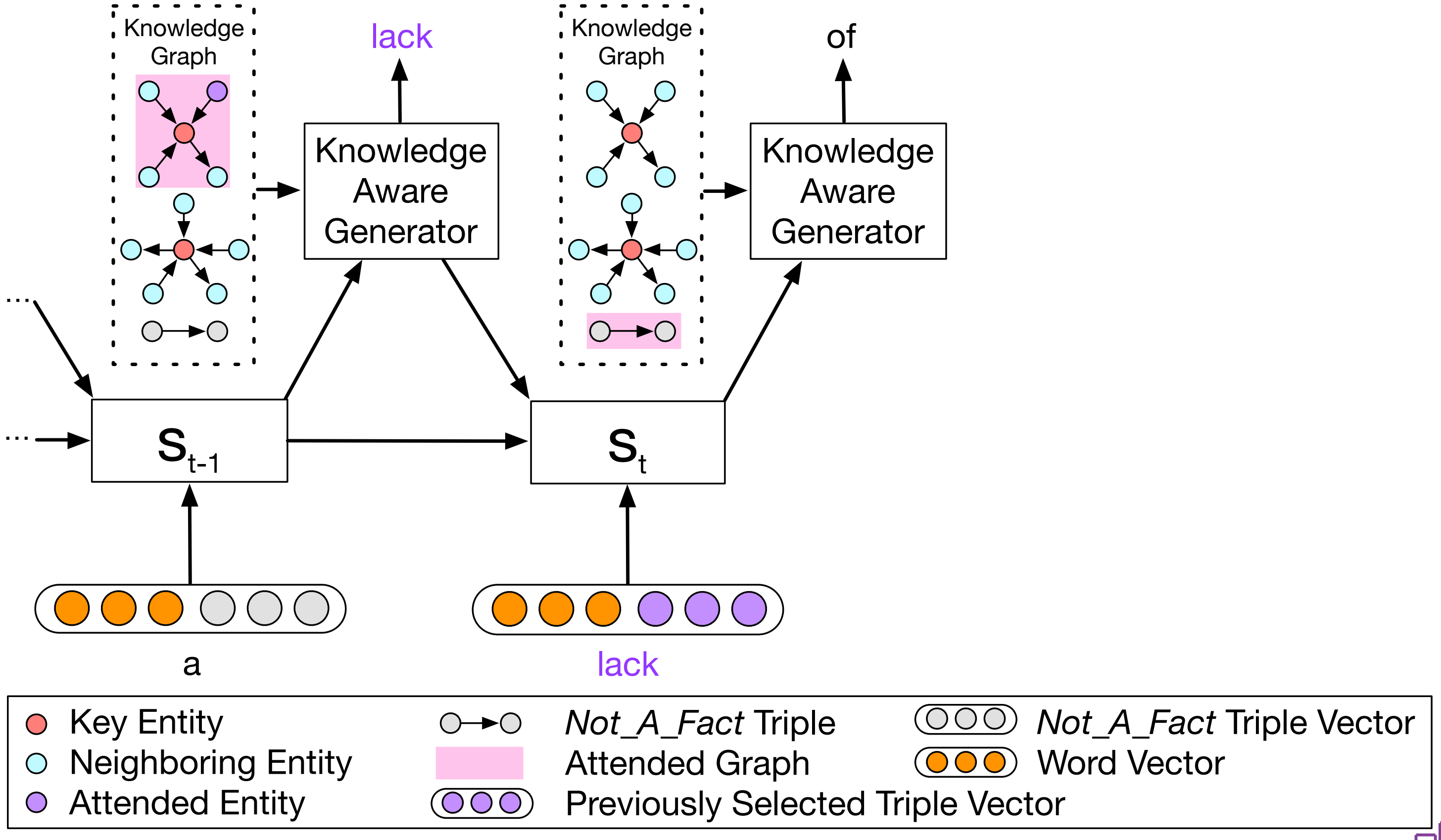
Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



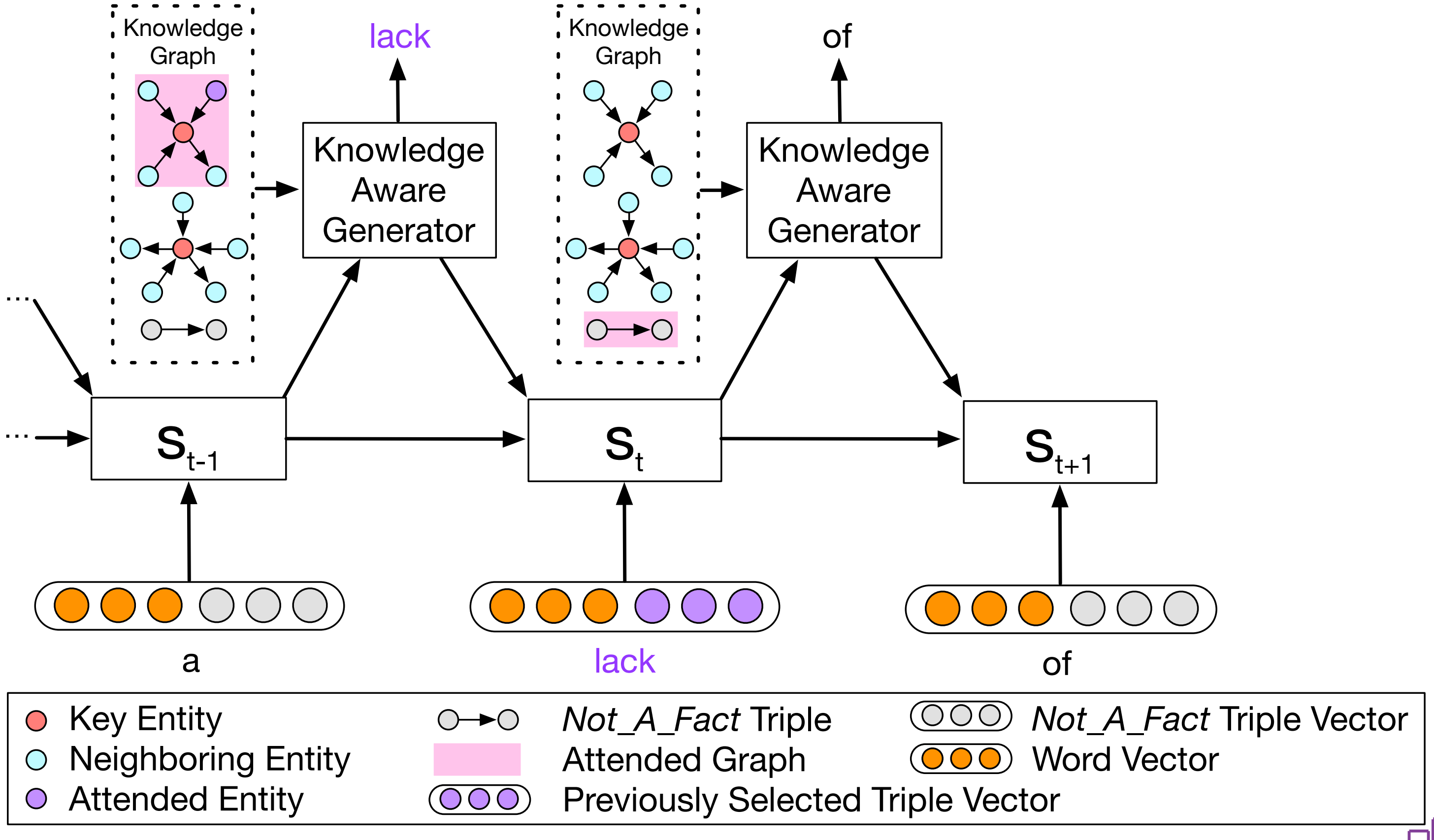
Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



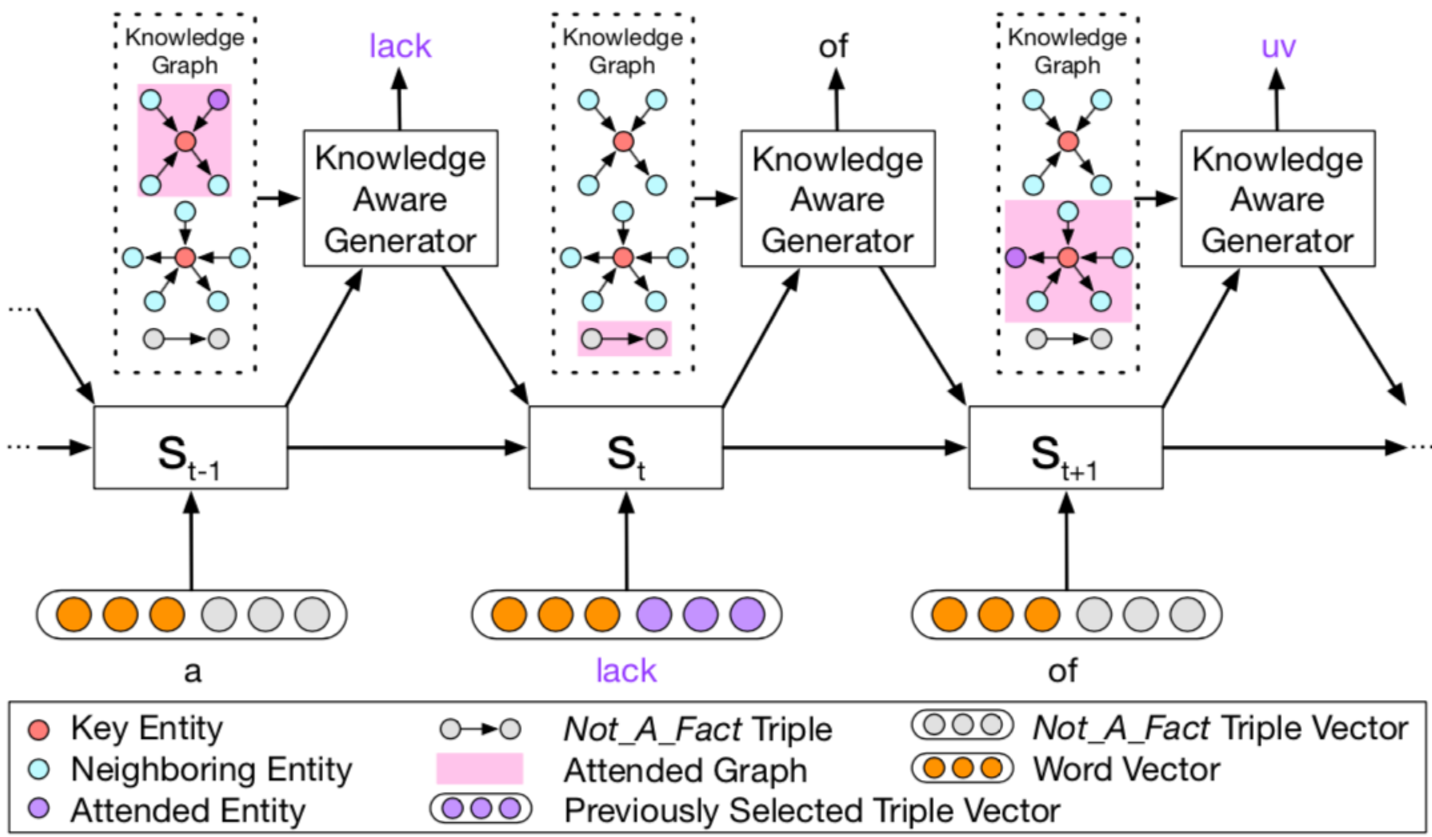
Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



Commonsense Knowledge in Chatbots

Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph



Commonsense Knowledge in Chatbots

- Dataset: filtered from 10M reddit single-round dialogs

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.

Commonsense Knowledge in Chatbots

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

Hao Zhou, Tom Yang, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu. Commonsense Knowledge Aware Conversation Generation with Graph Attention. **IJCAI-ECAI 2018**, Stockholm, Sweden.

Generation Examples 1

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**.

Generation Examples 2

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.

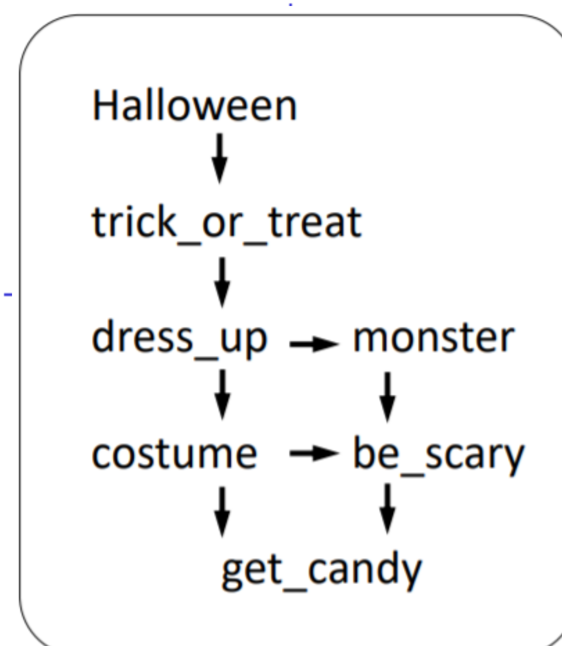
Logic:
Story Ending Generation

Logic: Story Ending Generation

Find context clues: plan the order of events and entities.

Today is **Halloween** .
Jack is so excited to go **trick or treating** tonight .
He is going to **dress up** like a **monster** .
The **costume** is real **scary** .

He hopes to get a lot of **candy** .



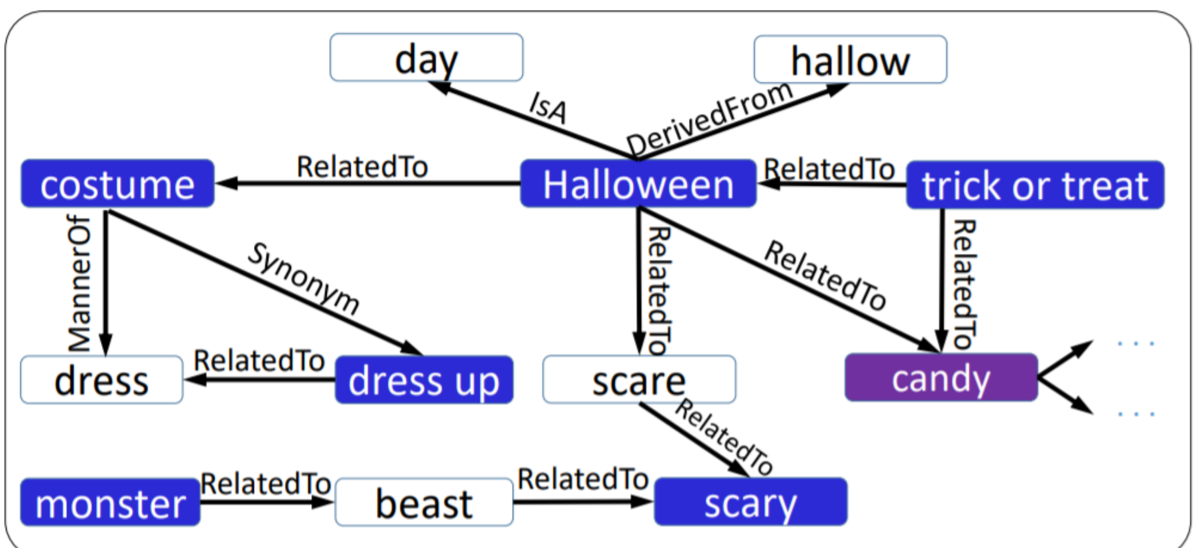
[Story Ending Generation with Incremental Encoding and Commonsense Knowledge] **AAAI 2019**

Logic: Story Ending Generation

Commonsense knowledge

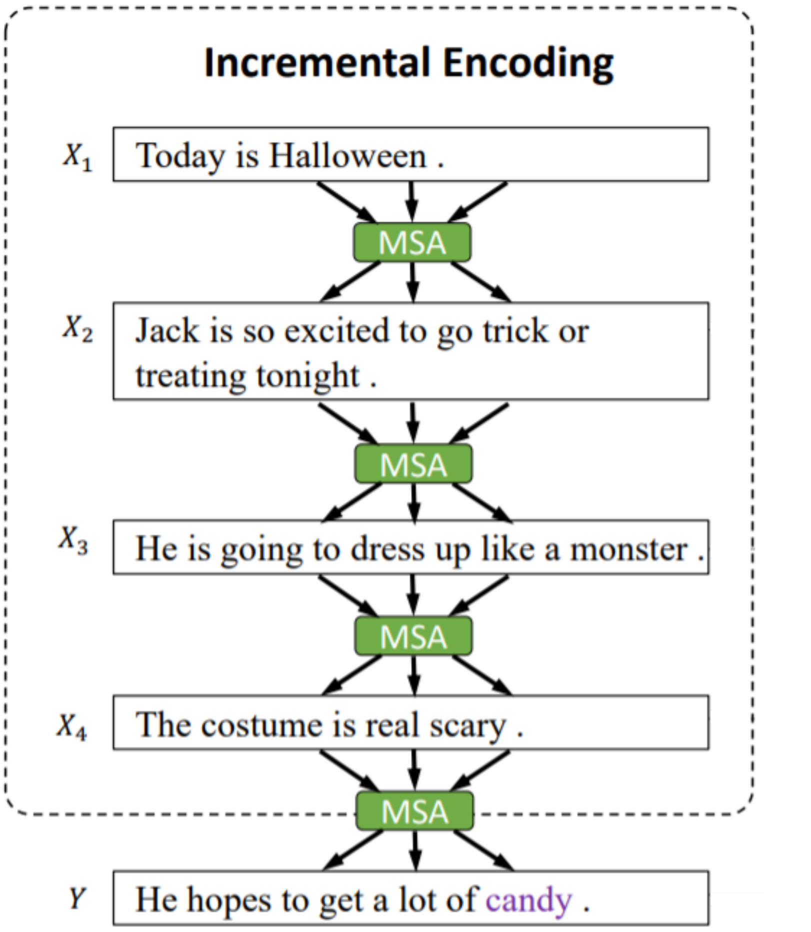
Today is **Halloween** .
Jack is so excited to go **trick or treating** tonight .
He is going to **dress up** like a **monster** .
The **costume** is real **scary** .

He hopes to get a lot of **candy** .

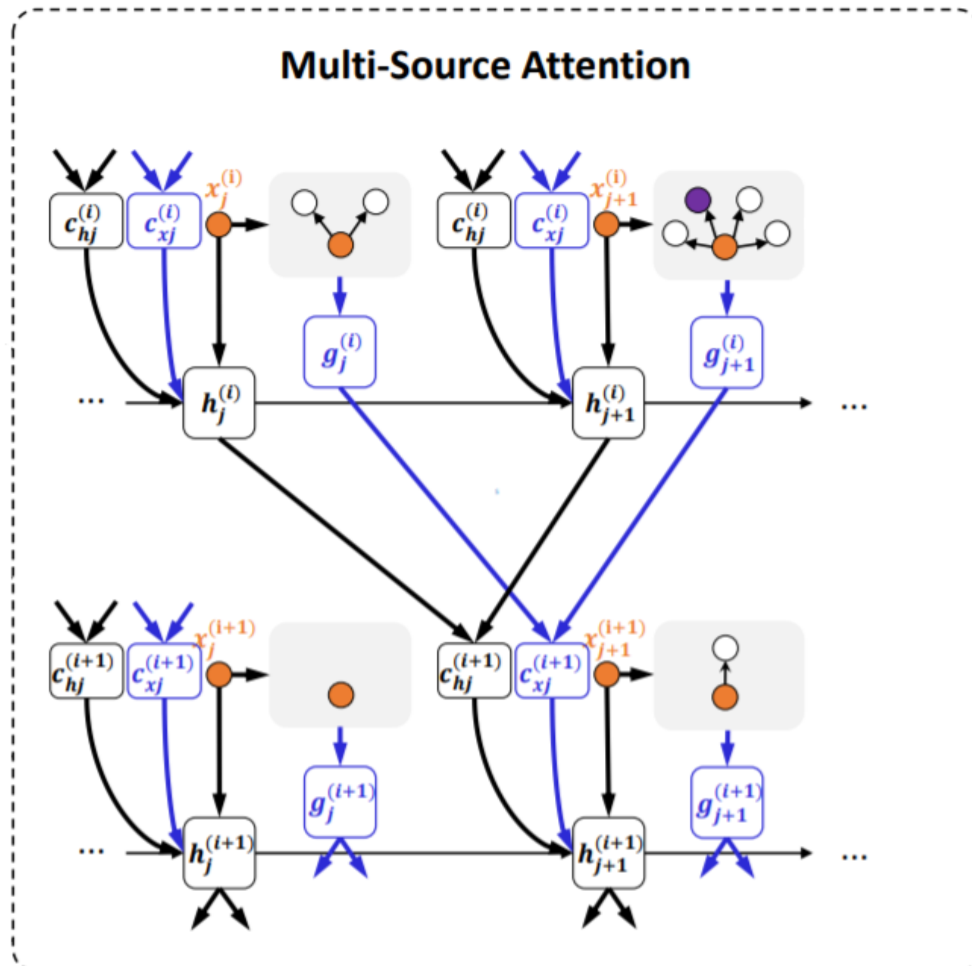


Logic: Story Ending Generation

Incremental Encoding

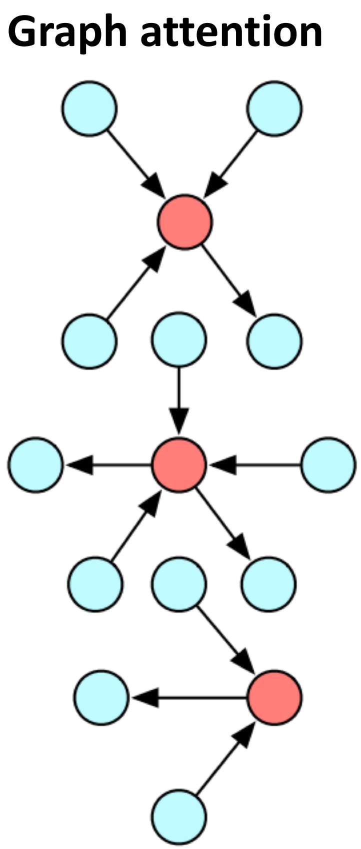


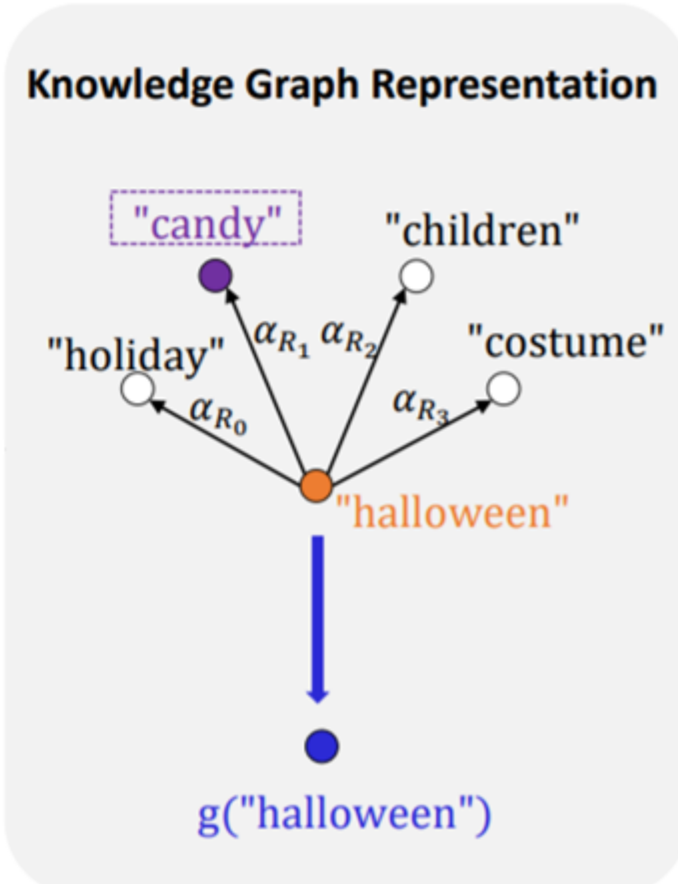
Multi-Source Attention



Logic: Story Ending Generation

Attention to the knowledge base: static graph attention





Experiment

- ROCStories, 90,000 for training, 8912 for test

Model	PPL	BLEU-1	BLEU-2	Gram.	Logic.
Seq2Seq	18.97	0.1864	0.0090	1.74	0.70
HLSTM	17.26	0.2459	0.0242	1.57	0.84
HLSTM+Copy	19.93	0.2469	0.0248	1.66	0.90
HLSTM+MSA(GA)	15.75	0.2588	0.0253	1.70	1.06
HLSTM+MSA(CA)	12.53	0.2514	0.0271	1.72	1.02
IE (ours)	11.04	0.2514	0.0263	**1.84**	1.10
IE+MSA(GA) (ours)	9.72	0.2566	0.0284	1.68	**1.26**
IE+MSA(CA) (ours)	**8.79**	**0.2682**	**0.0327**	1.66	1.24

Table 1: Automatic and manual evaluation results.

Logic: Story Ending Generation

Story 1:

Context:

Taj has never drank an espresso drink.
He ordered one while out with his friends.
The shot of espresso tasted terrible to him.
Taj found that he couldn't stop talking or moving.

Generated Ending:

He decided to never drink again.

Story 2:

Context:

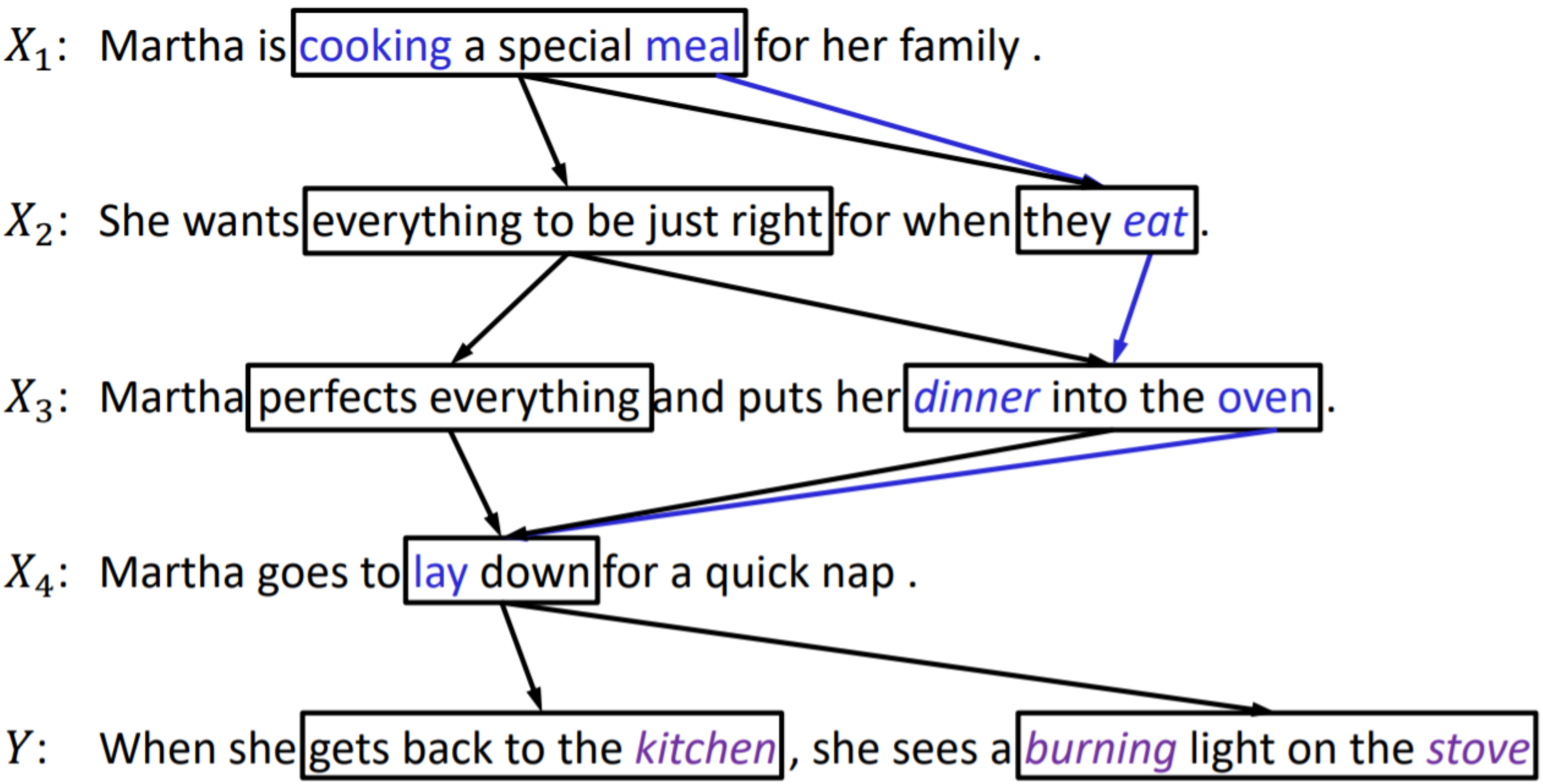
Martha is cooking a special meal for her family.
She wants everything to be just right for when they eat.
Martha perfects everything and puts her dinner into the oven.
Martha goes to lay down for a quick nap.

Generated Ending:

When she gets back to the kitchen, she sees a burning light on the stove.

Logic: Story Ending Generation

Building context clues incrementally



Summary

- **Semantics, consistency, interactiveness**
- **Emotion, personality, and knowledge**
- Still a long way to go: existing conversational systems are still far from **human-like**

Future Research Problems

- **Multi-modality** emotion perception and expression **(voice, vision, text)**
- **Personality, identity, style→ "human-like robot"**
  - **Introvert or extrovert**
  - **Personalized (style, or profile)**
- **Learning to learn (lifelong learning)**
  - Grow up from interactions with human partners and environment

Thanks for Your Attention

- http://coai.cs.tsinghua.edu.cn/ds/ 对话系统技术平台
- Acknowledgements
  - Prof Xiaoyan Zhu, Tsinghua colleagues, collaborators
  - Our students
- Contact:
  - Minlie Huang, Tsinghua University
  - aihuang@tsinghua.edu.cn
  - http://coai.cs.tsinghua.edu.cn/hml