

# **Towards Building More Intelligent Conversational Systems: Semantics, Consistency, &Interactivenss**

---

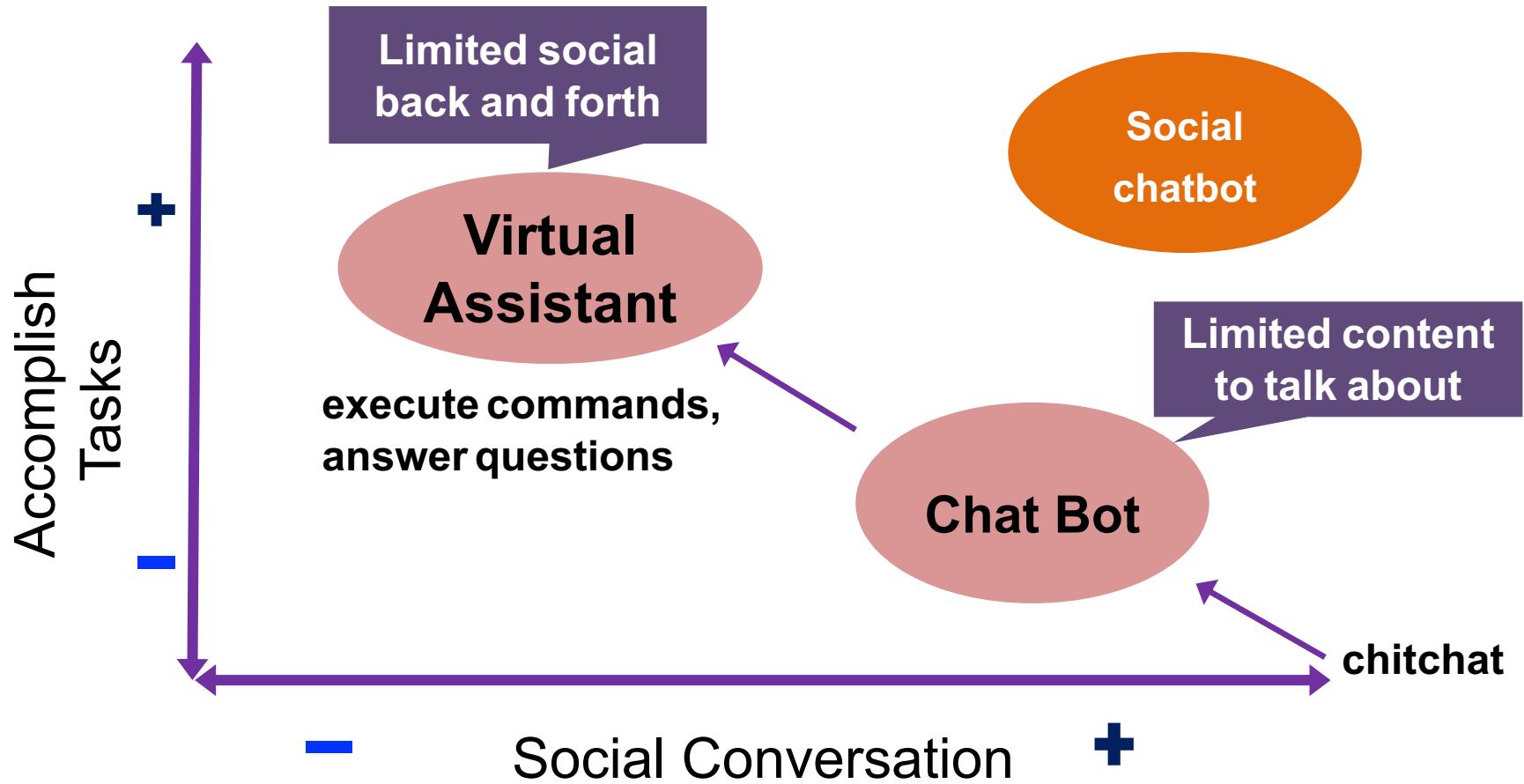
**Dr. Minlie Huang**

**Associate Professor**

**Tsinghua University**

**<http://coai.cs.tsinghua.edu.cn/html>**

# Types of Conversational AI Systems



# Issues Vary for Different Paradigms

## ASSISTANT

- \*Task intents
- \*Slot filling

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

**Constrained domains**

**Structured KB+DB**

## Conversational AI System Components

Speech/language understanding

Dialog management

Language generation

Back-end application

## SOCIALBOT

- \*Social & info
- \*Intents
- \***Grounding**

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

**Open domains**

**Unstructured Data & Info**



# 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 won't 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** (real understanding)
  - ◆ **Consistency** (long text generation, planning)
  - ◆ **Logic** (reasonable, commonsense)



# Bad Examples are More!

## 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

Xiaoice by MSRA



## 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?

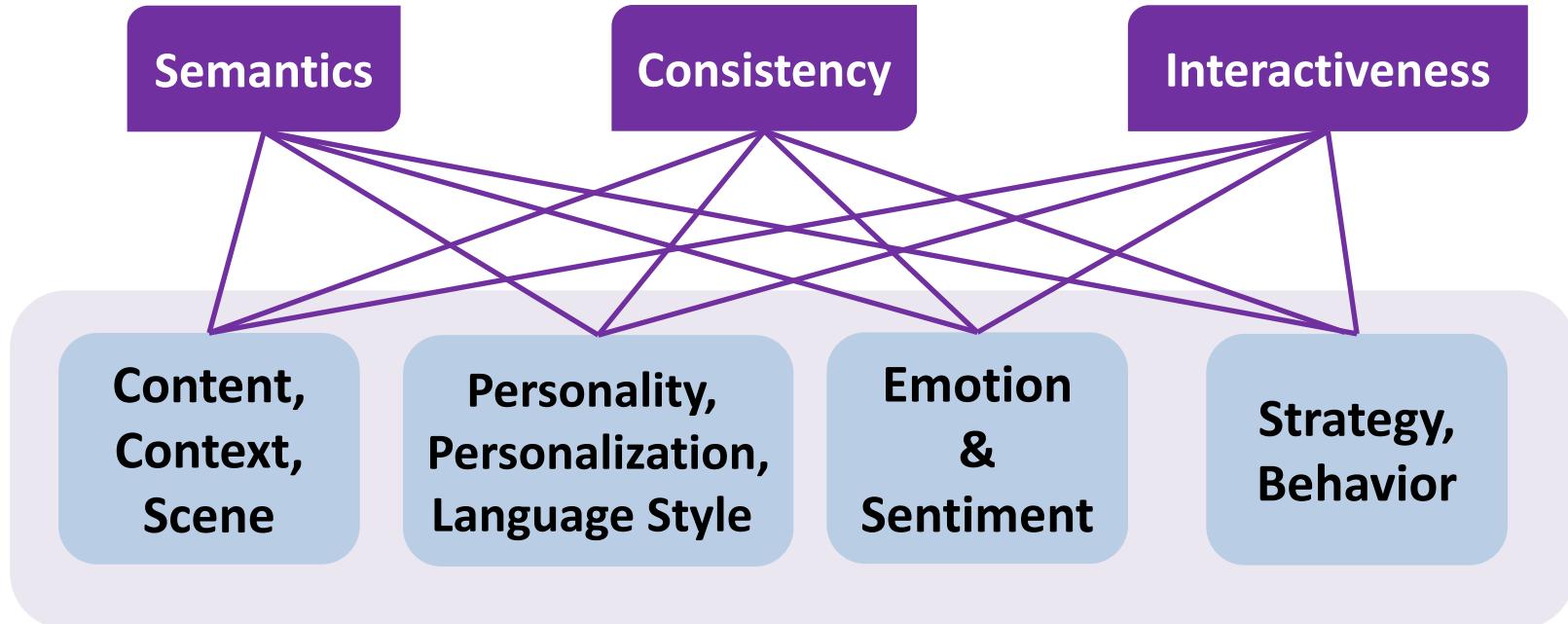
**小冰:** 生在九零后, 在深圳只能被当做八零后了。

# 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**.

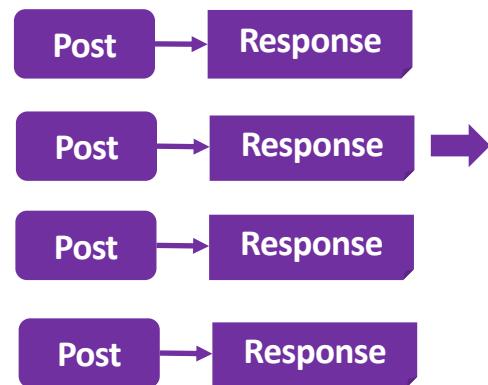
# Interactiveness: Emotion Perception and Expression



# Emotional Chatting Machine

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

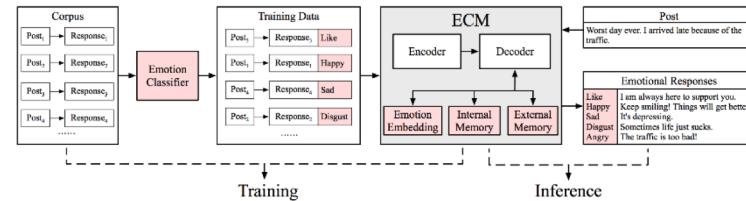
Social Interaction Data



Emotion Tagged data

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？

Emotional Chatting Machine



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

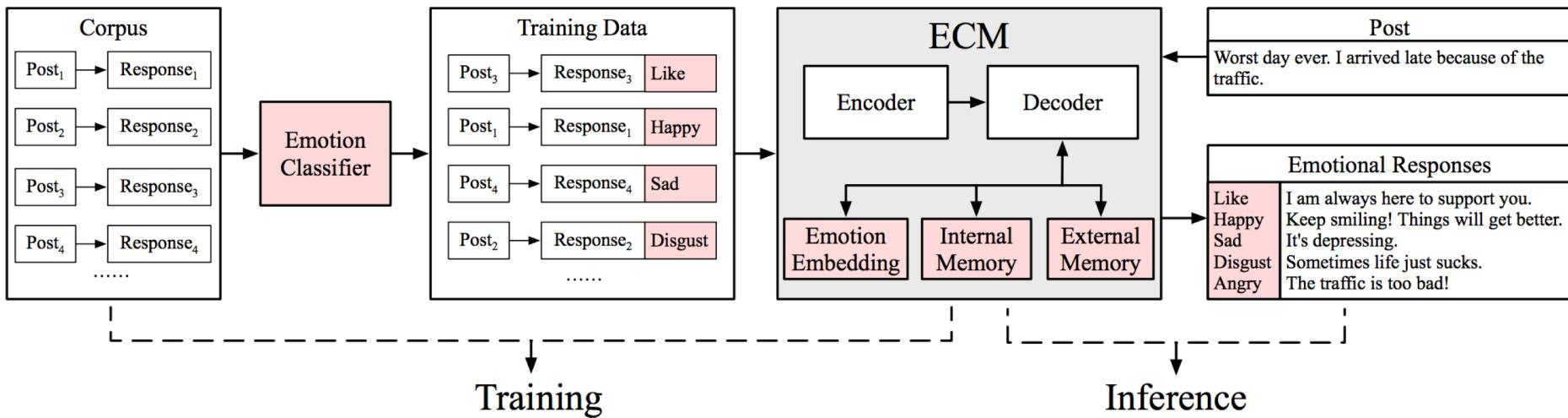
**Prof Björn Schuller:** “**an important step**” towards personal assistants that could read the emotional undercurrent of a conversation and respond with something akin to empathy.

Emotional Chatting Machine: Emotional Conversation Generation with Internal and External Memory. **AAAI 2018.**



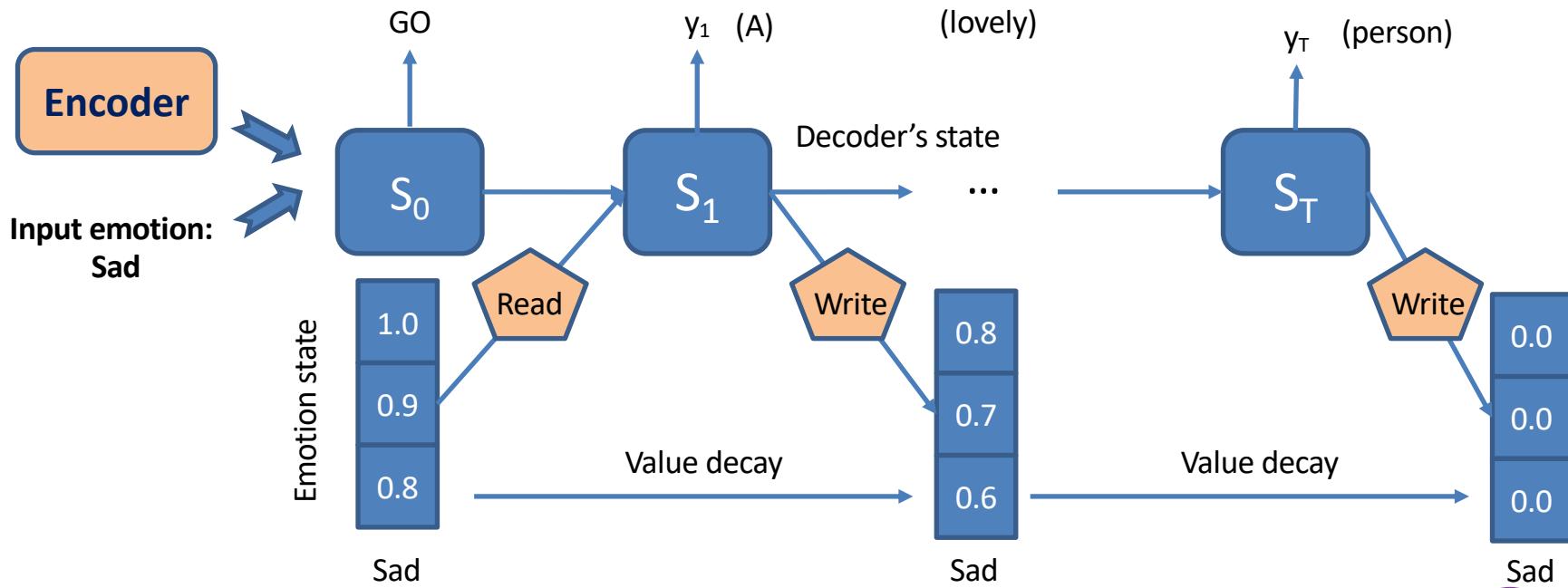
# Emotional Chatting Machine

- **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



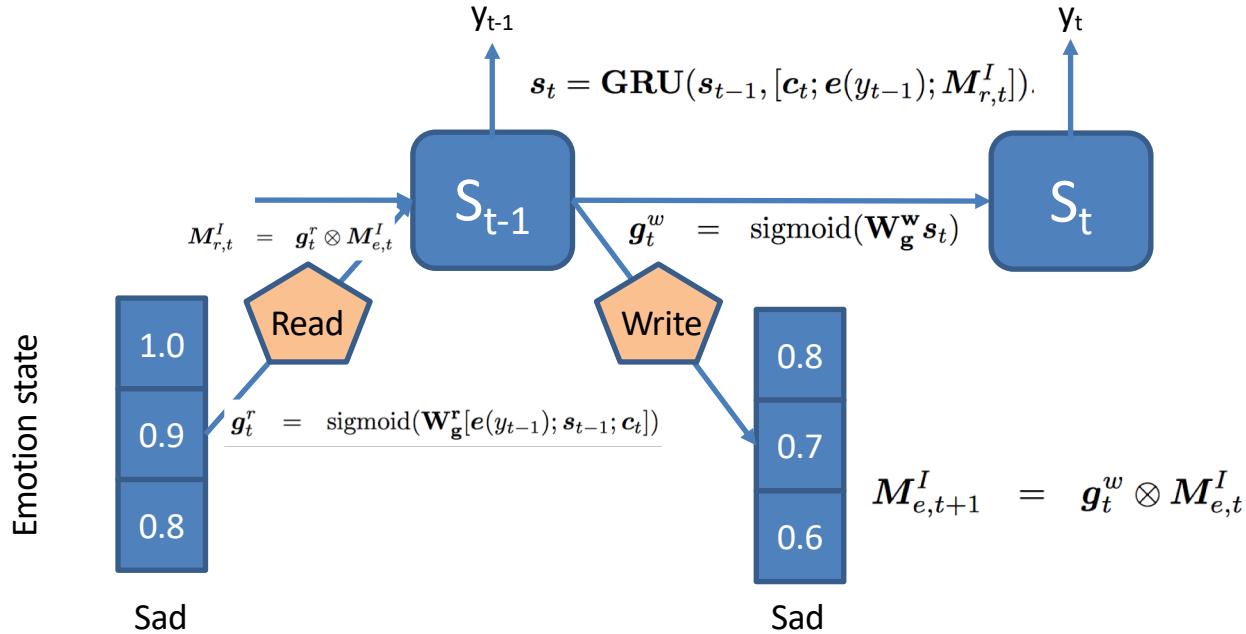
# Emotional Chatting Machine

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



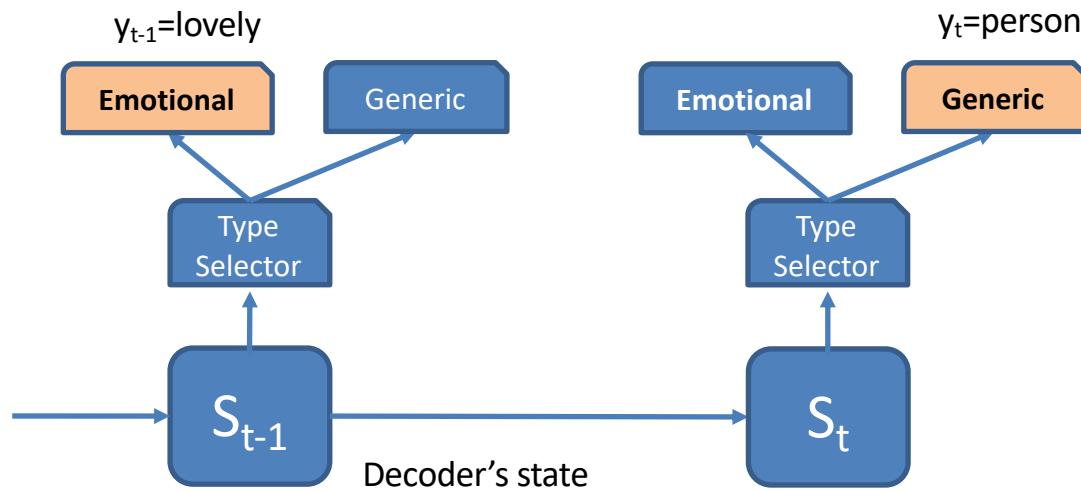
# Emotional Chatting Machine

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



# Emotional Chatting Machine

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



# 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	<b>0.773</b>
w/o Emb	66.1	0.753
w/o IMem	66.7	0.749
w/o EMem	<b>61.8</b>	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	<b>27.2</b>	<b>10.8</b>	4.4	24.2	15.5	17.9

Score distribution

Table 5: The percentage of responses with the score of *Content-Emotion*. F content score is 2 and emotion score is 1.

## Preference test

Pref. (%)	Seq2Seq	Emb	ECM
Seq2Seq	-	38.8	38.6
Emb	60.2	-	43.1
ECM	<b>61.4</b>	<b>56.9</b>	-

Table 7: Pairwise preference of the three systems.

Method	Overall		Like		Sad		Disgust		Angry		Happy	
	Cont.	Emot.										
Seq2Seq	1.255	0.152	1.308	0.337	1.270	0.077	<b>1.285</b>	0.038	<b>1.223</b>	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	<b>1.299</b>	<b>0.424</b>	<b>1.460</b>	<b>0.697</b>	<b>1.352</b>	<b>0.313</b>	1.233	<b>0.193</b>	0.98	<b>0.217</b>	<b>1.428</b>	<b>0.700</b>

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 ECM	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.
		Like	承诺是一个 <b>专一</b> 的傻子 Promise is a <b>single-minded</b> fool
		Happy	我就是一个 <b>开心</b> 的傻子 <b>I'm just a happy</b> fool
		Sad	我就是一个 <b>悲伤</b> 的傻子 <b>I am a sad</b> fool
		Disgust	一个 <b>骗子</b> 说给一个 <b>傻子</b> 听 <b>A cheater</b> says to a <b>fool</b>
		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 ECM	Seq2Seq	知道真相的我眼泪掉下来。 After knowing the truth, I burst into tears.
		Like	太 <b>可爱</b> 了，太 <b>可爱</b> 了！ Too <b>lovely</b> , it's too <b>lovely</b> .
		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 <b>眼泪</b> 掉下来。 After knowing the truth, I burst into <b>tears</b> .
		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.



# More Examples

情绪 emotion	<b>Post:</b> 早上的天气不错啊！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	<b>天气好，心情不好。</b>
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！

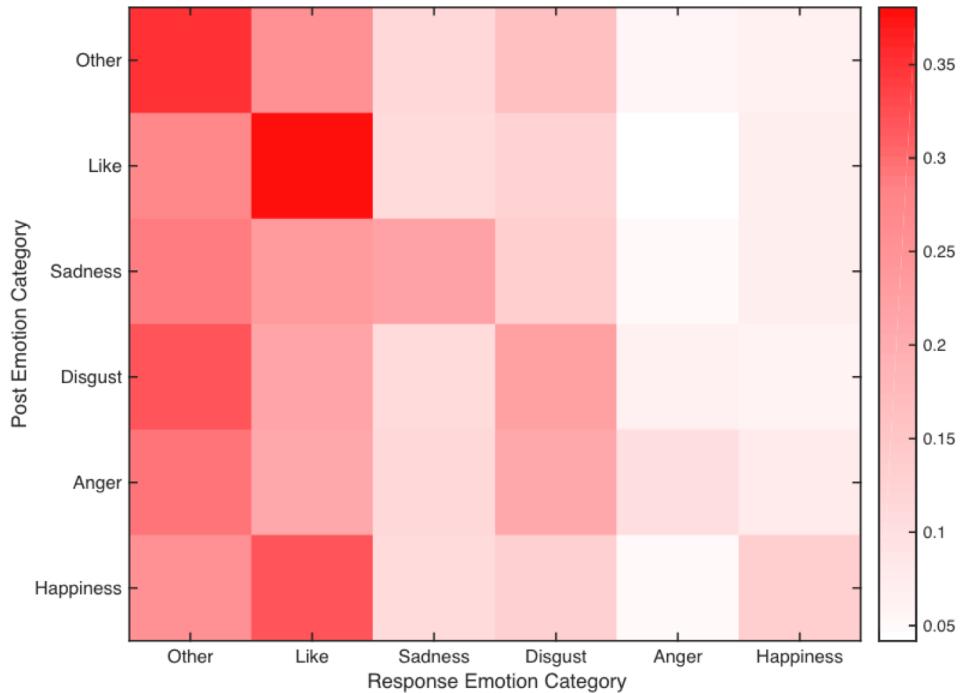
情绪 emotion	<b>Post:</b> 为了解决抄作业的不好习惯，高数老师这回要出绝招了！
Other	这老师是要逆天吗？
喜欢 Like	这老师太有才了！
悲伤 Sad	这样的老师伤不起啊
厌恶 Disgust	<b>这老师也太狠了点吧</b>
愤怒 Angry	<b>这老师是要逆天啊！</b>
快乐 Happy	哈哈，这老师太有才了

情绪 emotion	<b>Post:</b> 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	<b>为什么我觉得很恶心</b>
厌恶 Disgust	<b>这样的蛋糕你忍心吗？</b>
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	<b>post:</b> 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	<b>我就是一个悲伤的傻子</b>
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	<b>我就是一个开心的傻子</b>



# Emotion Interaction Patterns



Like → Like (**empathy**)

Sadness → Sadness (**empathy**)

Sadness → Like (**comfort**)

Disgust → Disgust (**empathy**)

Disgust → Like (**comfort**)

Anger → Disgust

Happiness → Like



# 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.



# Asking Questions in Conversational Systems

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

Scene: Dining at a restaurant

我昨天晚上去聚餐了

I went to dinner yesterday night.

Friends?

WHO

Place?

WHERE

Food?

HOW-ABOUT

...

Persons?

HOW-MANY

Bill?



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.

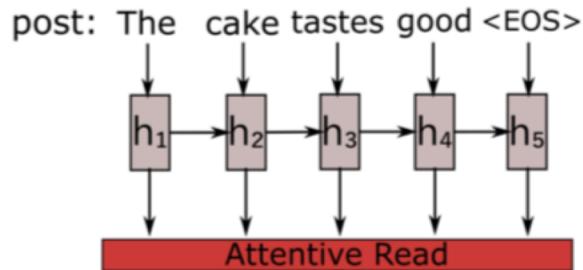
Machine: **Are** you happy with your singing?



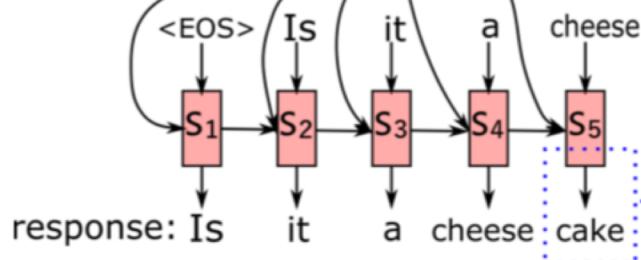
# Asking Questions in Conversational Systems

- Typed decoders: soft typed decoder

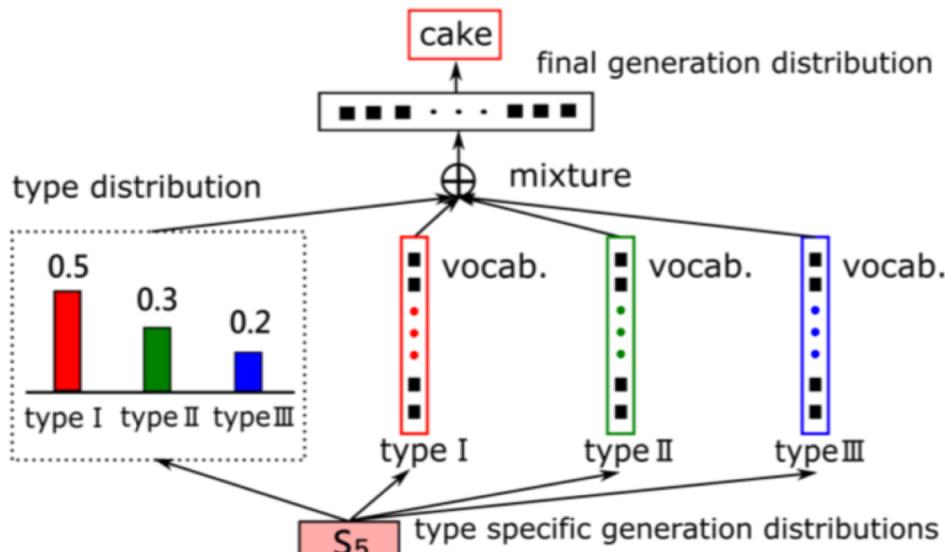
Encoder:



Decoder:

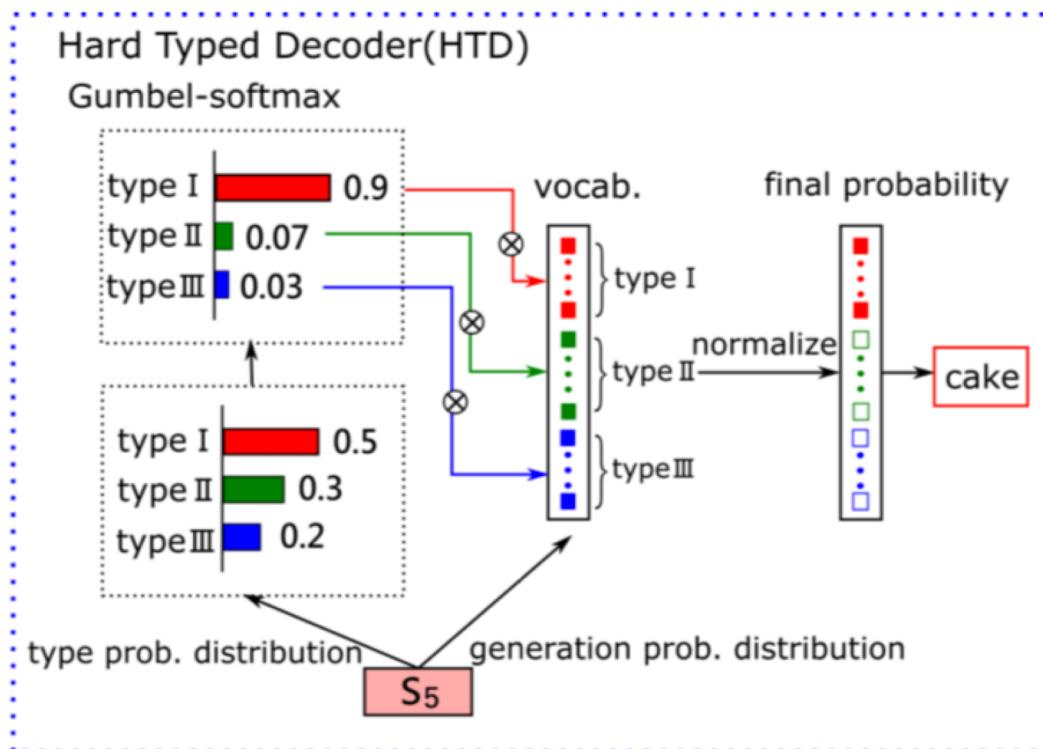


Soft Typed Decoder(STD)



# Asking Questions in Conversational Systems

- Typed decoders: hard typed decoder



**For each post:**

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

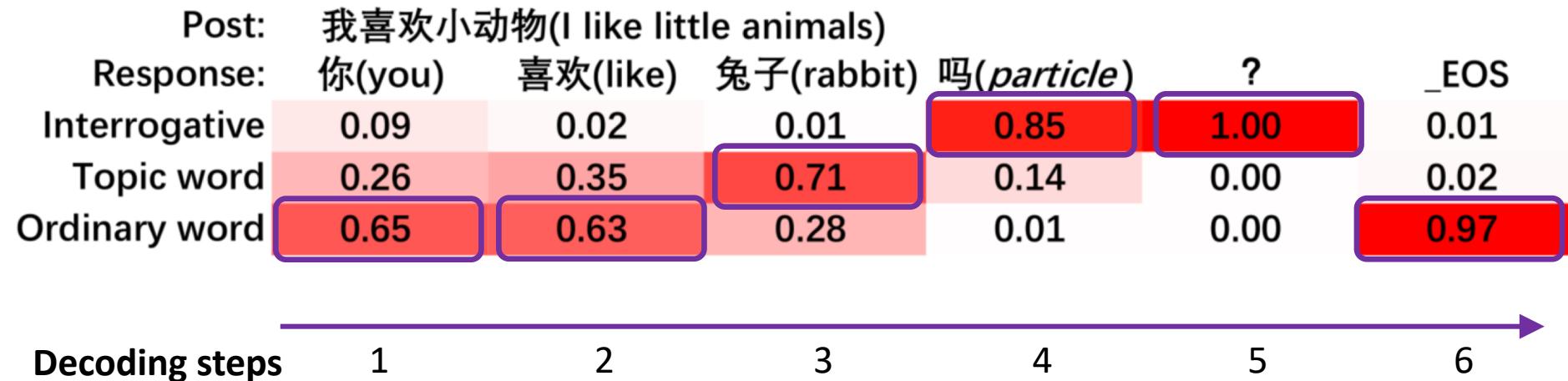
**Topic words:**

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



# Asking Questions in Conversational Systems

- Type prediction at each decoding position

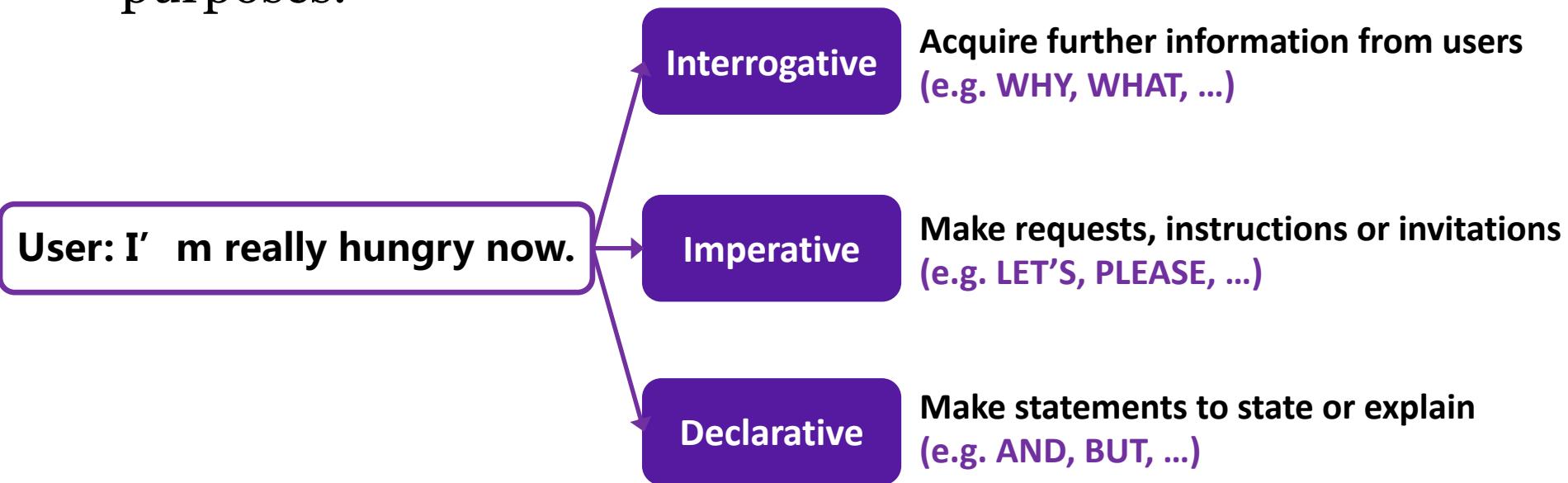


# Interactiveness: Achieving Different Speaking Purposes by Controlling Sentence Function



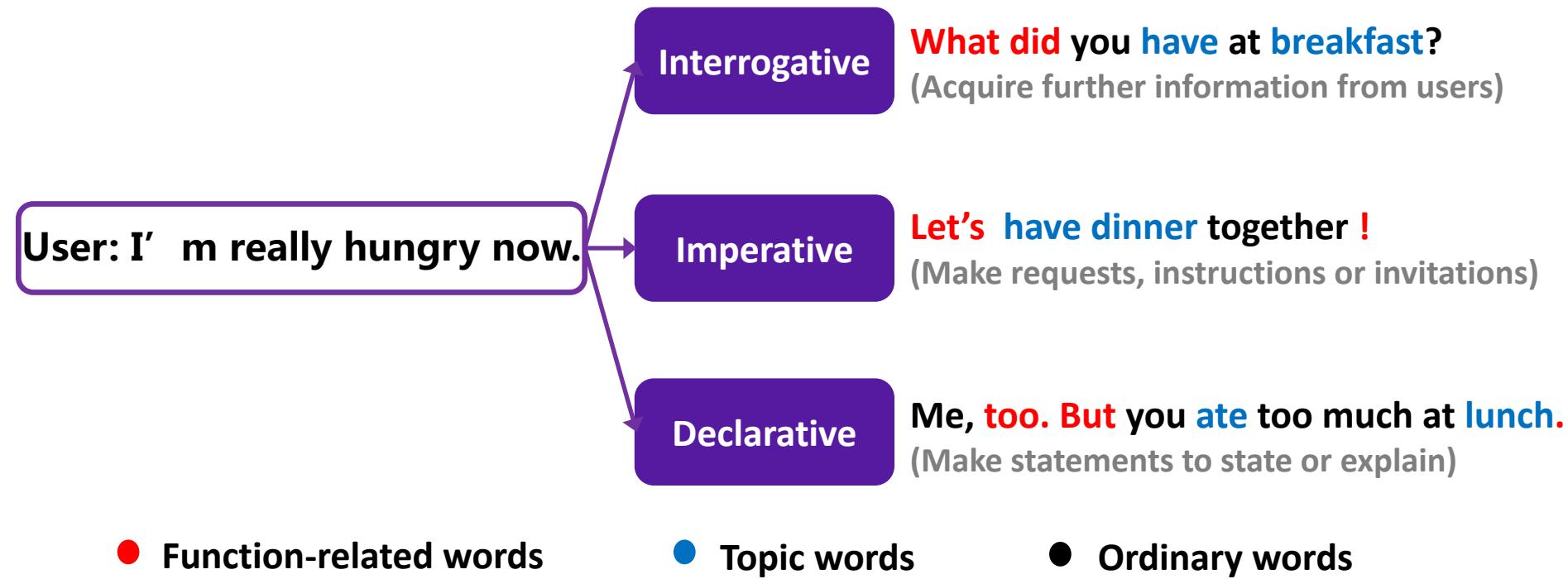
# Controlling Sentence Function

- Sentence function indicates different conversational purposes.



# Controlling Sentence Function

- Response with controlled sentence function requires a **global plan** of *function-related*, *topic* and *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



# 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: 那你是哪几年出生的？

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**.

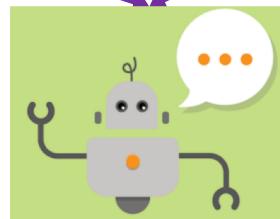
UserA: do you like to play piano?

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**.

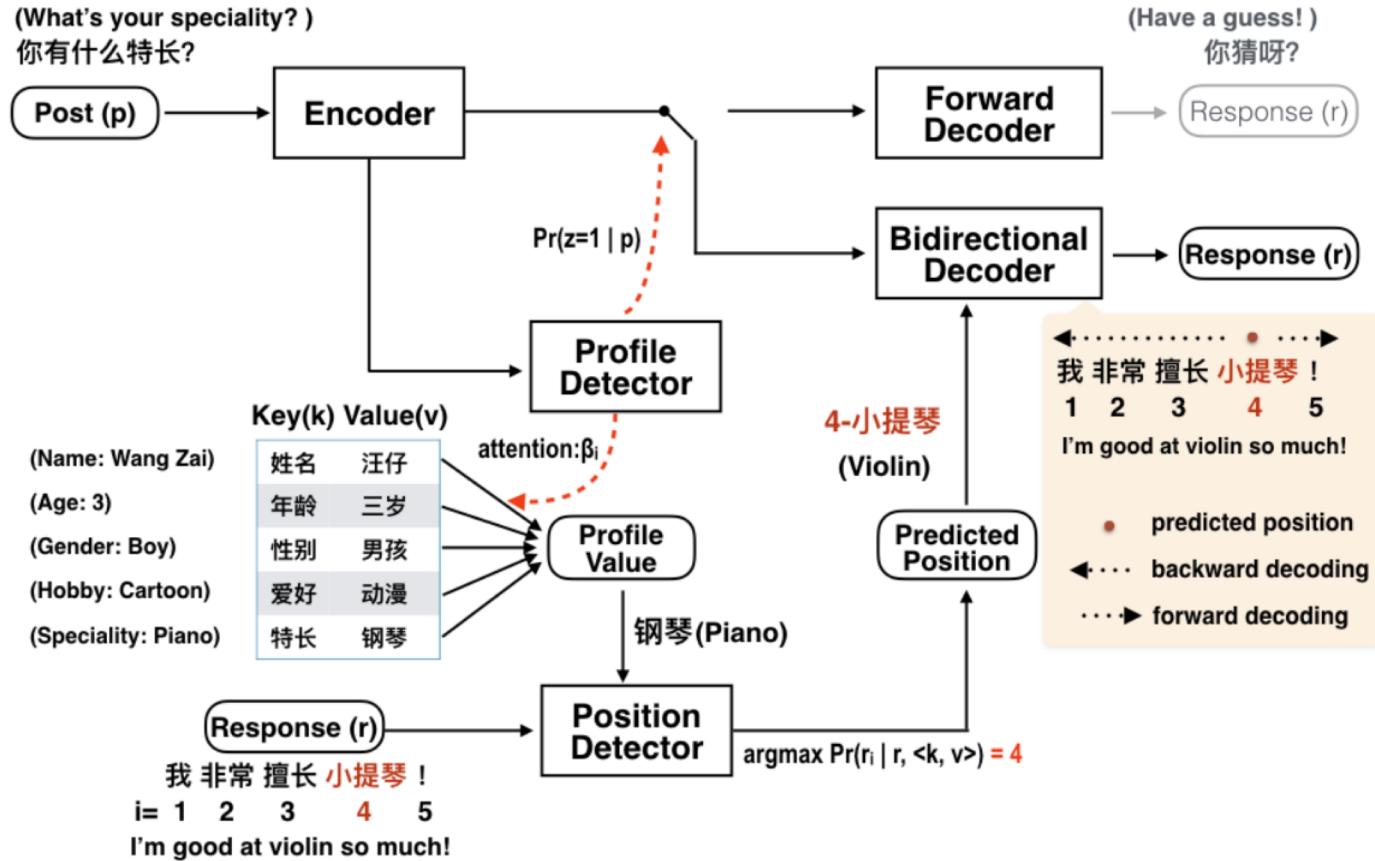
# Personality of Conversational Systems

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.



# 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	<b>83.3%</b>	<b>59.5%</b>	<b>45.8%</b>

## 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	<b>49.5%</b>	<b>27.8%</b>

## 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

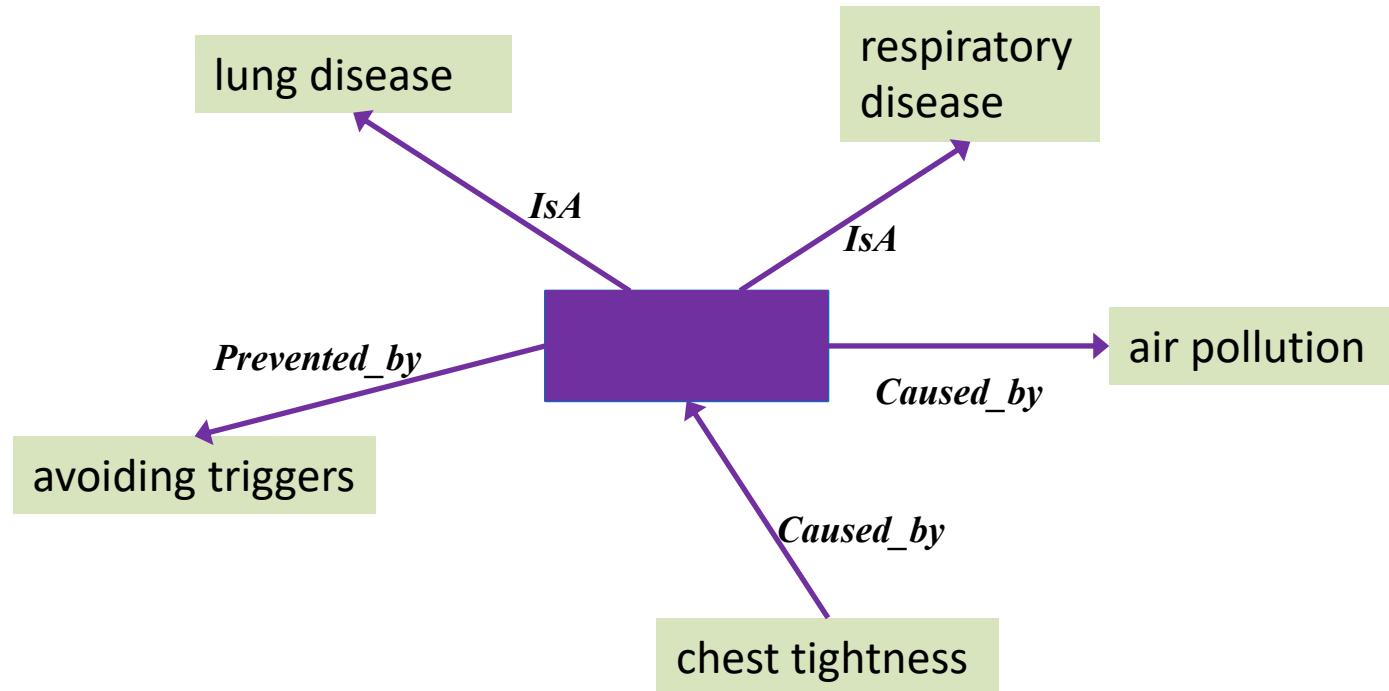


# Commonsense Knowledge

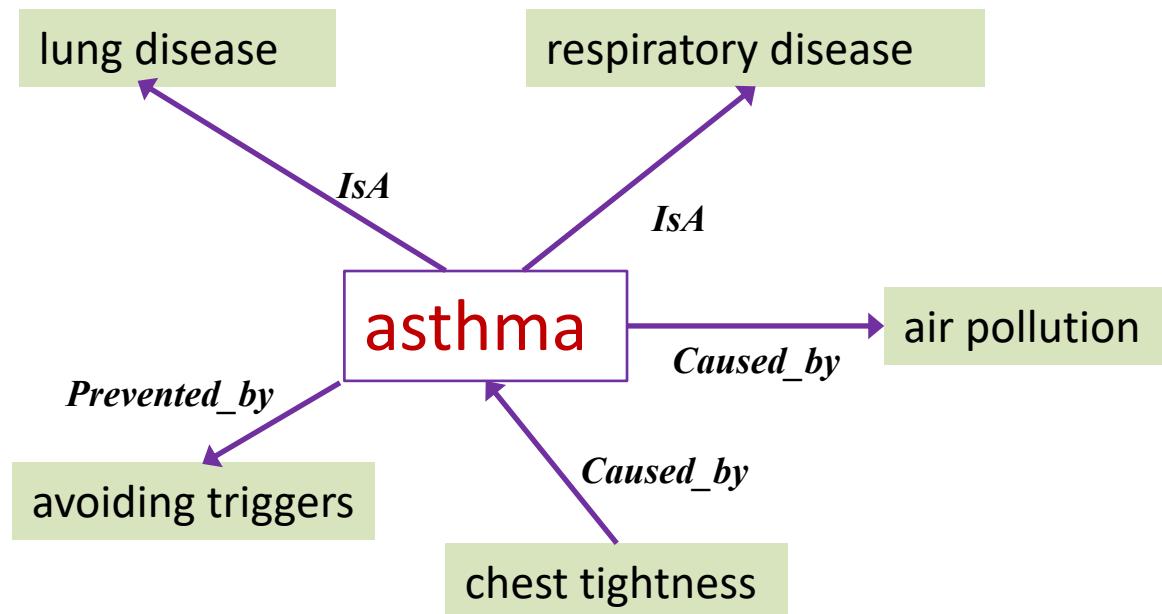
- ◎ **Commonsense knowledge** consists of facts about the everyday world, that all humans are expected to know. (Wikipedia)
  - ◆ Lemons are sour
  - ◆ Tree has leafs
  - ◆ Dog has four legs
- ◎ Commonsense Reasoning ~ **Winograd Schema Challenge**:
  - The trophy would not fit in the brown suitcase 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 in Chatbots



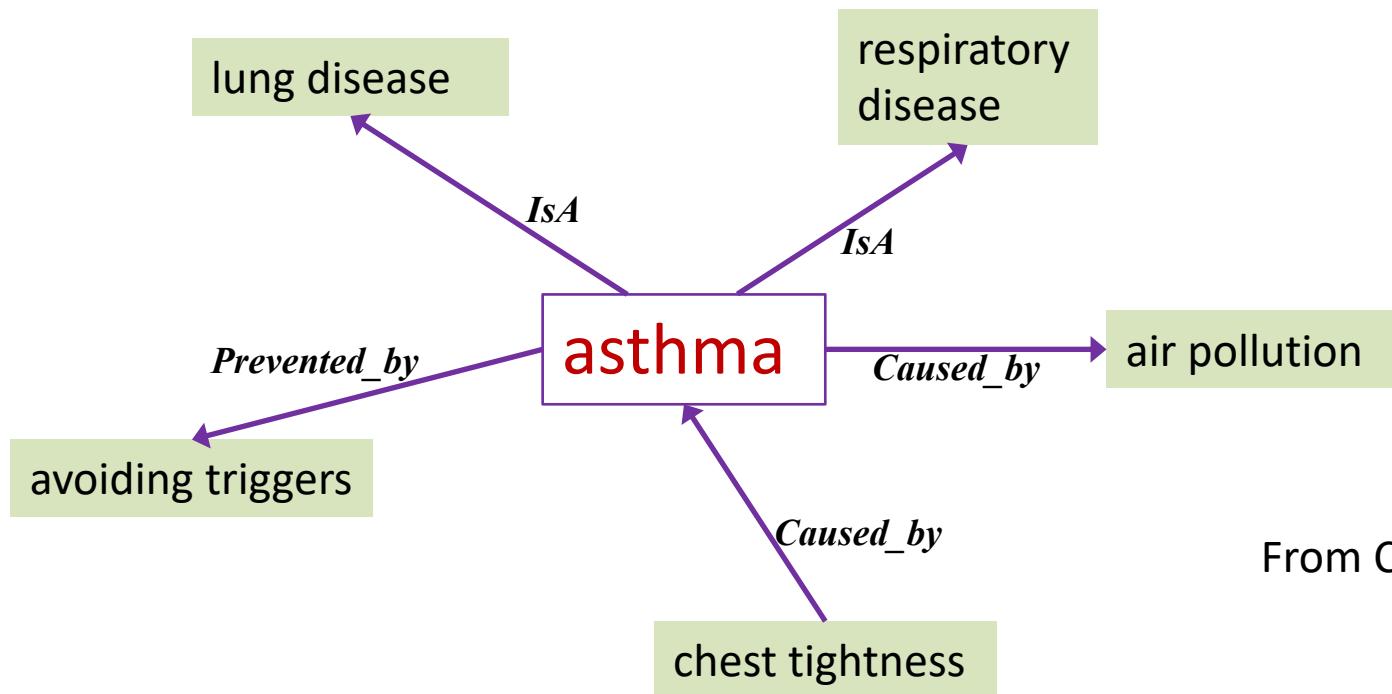
# Commonsense Knowledge in Chatbots



# 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)



# Commonsense Knowledge in Chatbots

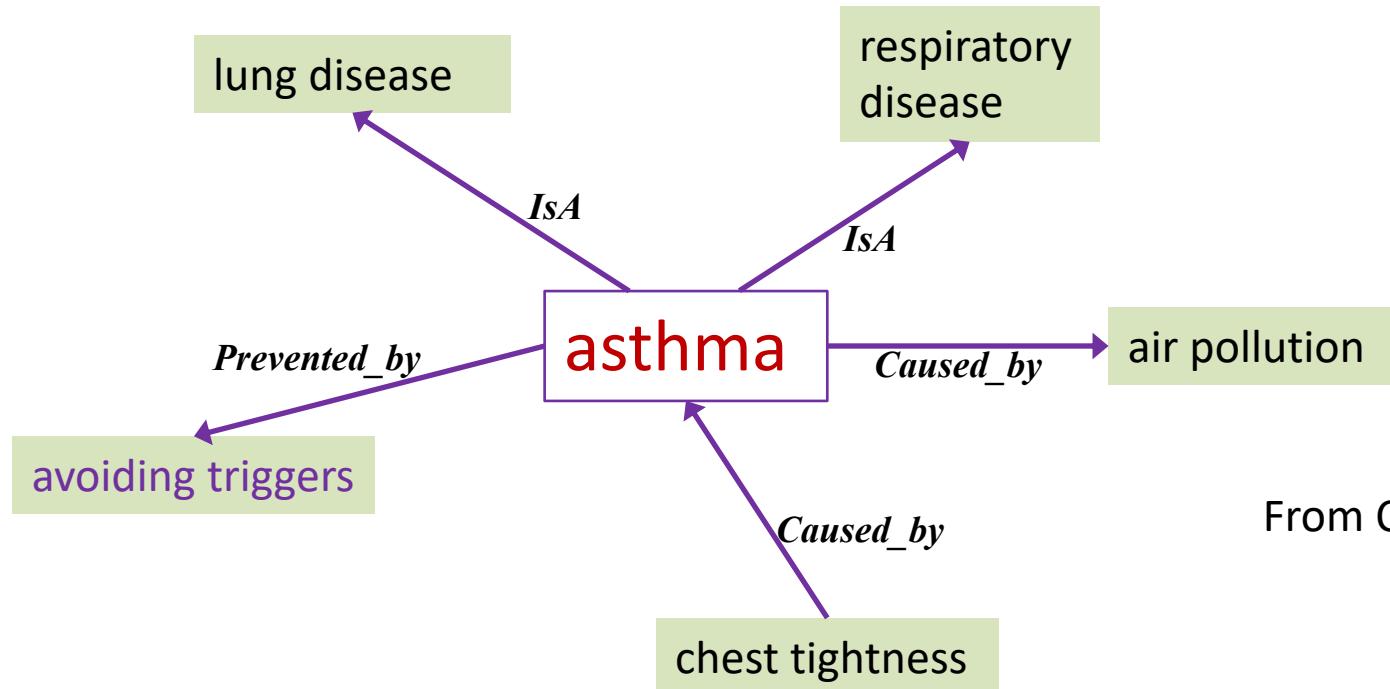
**Post:** I have an **asthma** since three years old.

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

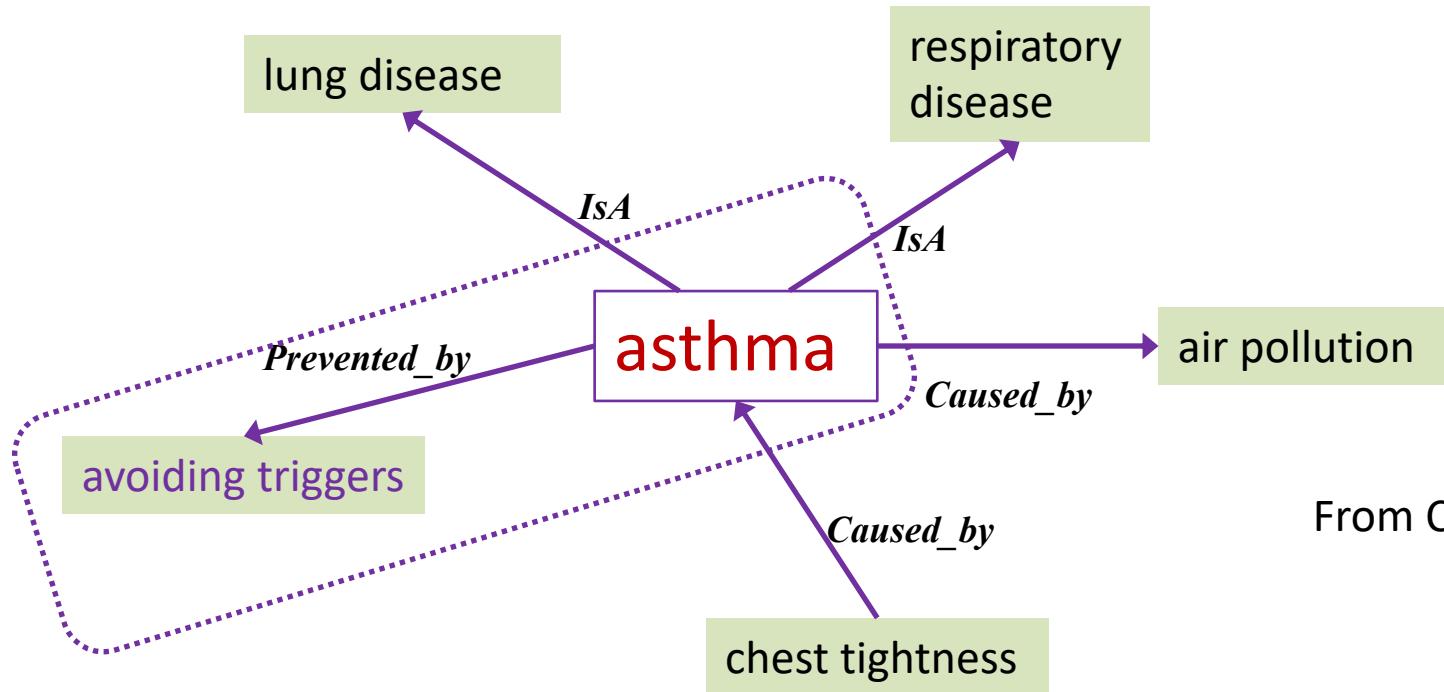
**Post:** I have an **asthma** since three years old.

Triples in knowledge graph:

(lung disease, IsA, **asthma** )

(**asthma**, Prevented\_by, avoiding triggers)

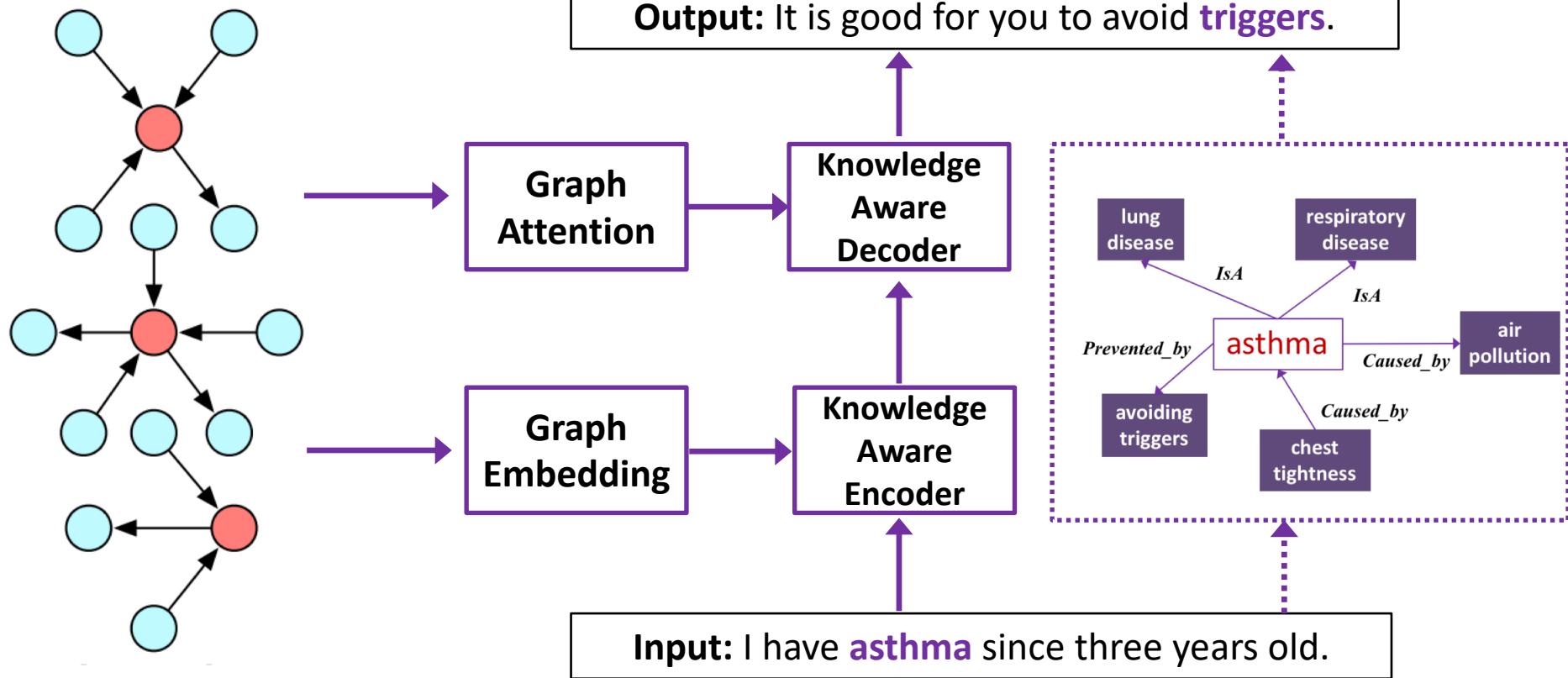
**Response:** I am sorry to hear that. Maybe **avoiding triggers** can prevent **asthma** attacks.



From ConceptNet

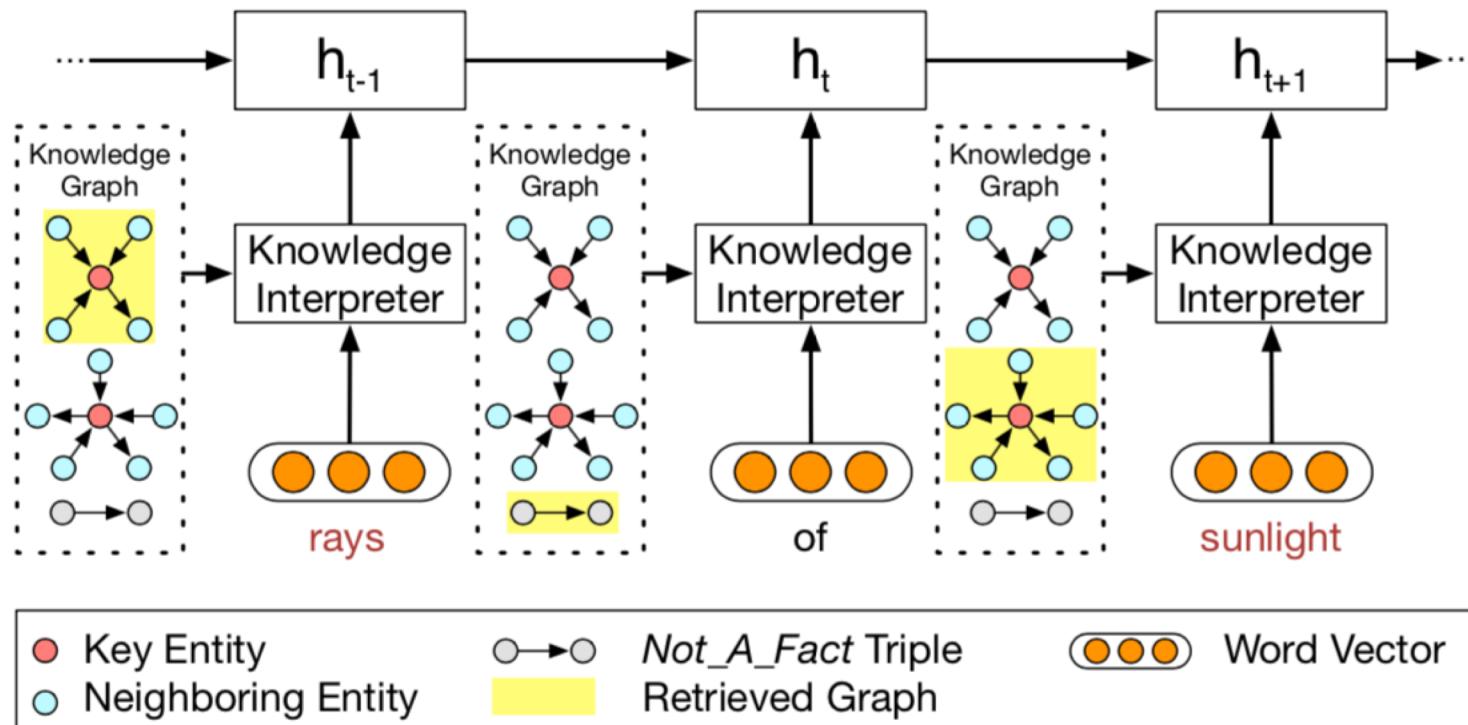


# Commonsense Knowledge in Chatbots



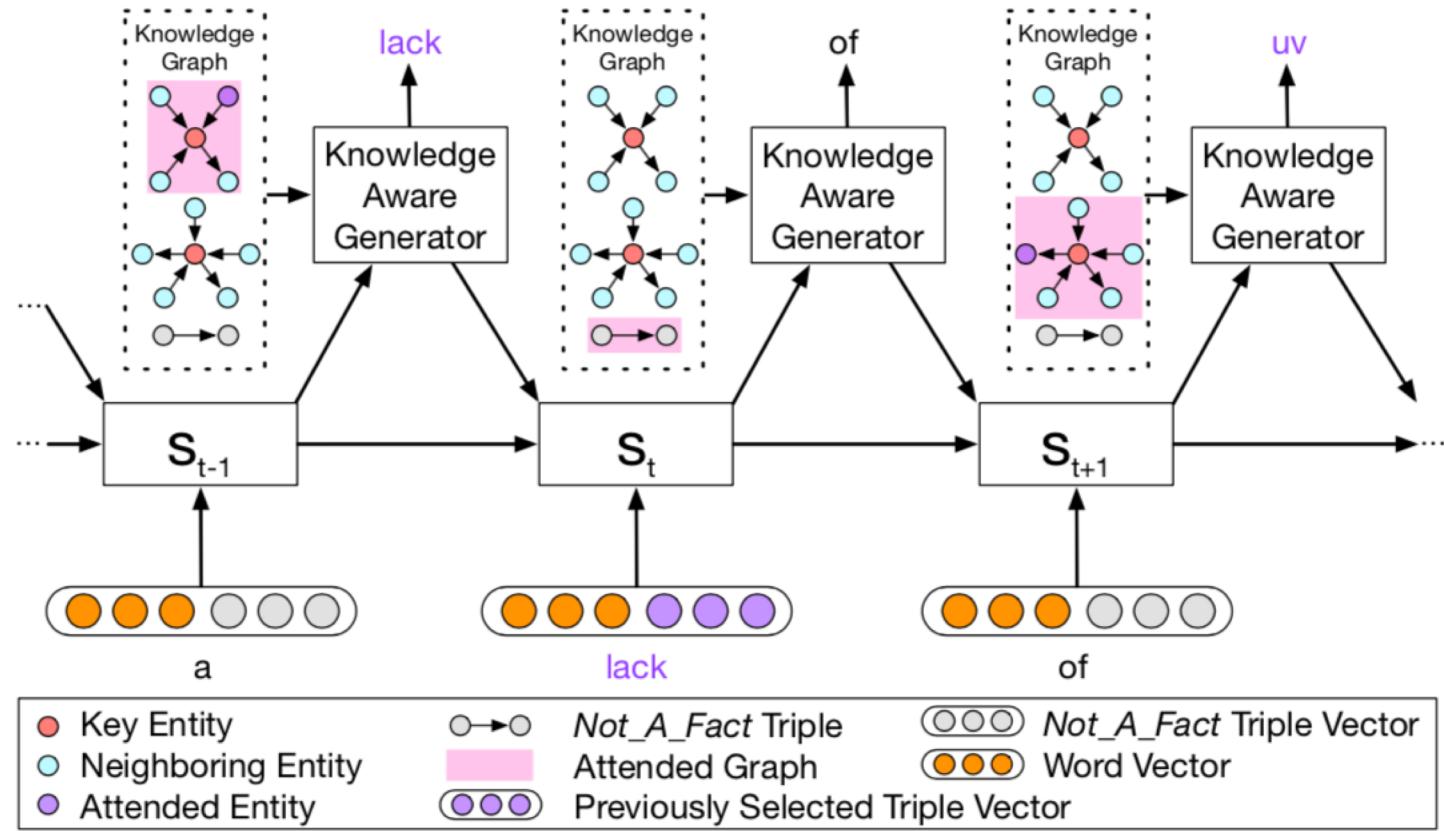
# Commonsense Knowledge in Chatbots

**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



# Commonsense Knowledge in Chatbots

## Automatic evaluation

Model	Overall		High Freq.		Medium Freq.		Low Freq.		OOV	
	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	<b>39.18</b>	<b>1.180</b>	<b>35.36</b>	<b>1.156</b>	<b>39.64</b>	<b>1.191</b>	<b>40.67</b>	<b>1.196</b>	<b>40.87</b>	<b>1.162</b>

## 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

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**.



# Summary

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



# 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](mailto:aihuang@tsinghua.edu.cn)
  - ◆ <http://coai.cs.tsinghua.edu.cn/hml>

