Towards Building More Intelligent Chatting Machines

Dr. Minlie Huang (黄民烈)
Associate Professor
CS Department, Tsinghua University
aihuang@tsinghua.edu.cn
http://coai.cs.tsinghua.edu.cn/hml
Our CoAI Group

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

- Chat Bot: Limited social back and forth, execute commands, answer questions.
- Virtual Assistant: Limited content to talk about, chitchat.
- Ideal Bot: Limited social back and forth, execute commands, answer questions.

From prof. M Ostendorf, with edits
Issues Vary for Different Paradigms

ASSISTANT
- *Task intents
- *Slot filling
- *Narrow options & execute tasks
- *Reward = timely task completion

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

Constrained domains
Structured KB+DB
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: 害怕孤独，又舍不得自己的小自由。
    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.
User: 你若安好，便是晴天  
Machine: 这句话我爱听，谢谢你的祝福，我会继续努力的！

User: 雪，飘舞。人，行走。树，纯美。天，苍茫。城，安静。  
Snow dancing, people walking. Beautiful trees, boundless sky, quiet city.  
Machine: 好美的意境  
So beautiful scene.  

User: 欲享受选择之快乐必先经历选择之痛苦  
Machine: 是的，人生就是这样，人生的意义就是选择。  

User: 有些命里遇见，从相识到相知，到最后的相忘于江湖，似乎都是命中注定。  
Machine: 缘来缘去，这就是人生
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: 会做饭的男人是最性感的男人。
    Mans who can cook are the most sexy.
Machine: 会做饭的男人是最性感的女人
    Mans who can cook are the most sexy women.
Bad Examples are More!

User: 猫叔羊毛毡手工作品，太美好了！
Machine: 这是什么品种的猫啊

User: 不得不夸一句广东的饭确实好吃！
Machine: 我也喜欢吃臭豆腐。

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!
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

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。
Bad Examples (AI Ethics)

Picture from Prof. Frank Rudzicz, University of Toronto
Challenges in Chatting Machines

- **One-to-many**: one input, many many possible responses

- **Knowledge & Reasoning**: real understanding requires various knowledge, world facts, or backgrounds

- **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 Chatting Machines

IJCAI-ECAI 2018
AAAI 2019

Semantics

Content, Context, Scene

Personality, Personalization, Language Style

Emotion & Sentiment

Strategy, Behavior

Consistency

Interactiveness

Open-domain, open-topic conversational systems
Open-domain Chatting Machines

- 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)

---

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*. 
Interactiveness:
Emotion Perception and Expression
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

- **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

<table>
<thead>
<tr>
<th>Post</th>
<th>Model</th>
<th>Response (Original)</th>
<th>Response (Translated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>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.</td>
<td>Seq2Seq</td>
<td>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.</td>
<td>A promise is what a cheater says to a fool.</td>
</tr>
<tr>
<td>ECM</td>
<td>Like</td>
<td>承诺是一个专一的傻子</td>
<td>Promise is a single-minded fool</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>我就是一个开心的傻子</td>
<td>I'm just a happy fool</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>我就是一个悲伤的傻子</td>
<td>I am a sad fool</td>
</tr>
<tr>
<td></td>
<td>Disgust</td>
<td>一个骗子说给一个傻子听</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Angry</td>
<td>不管你信不信，反正我信了！</td>
<td>Whatever you believe or not, I trust in it!</td>
</tr>
<tr>
<td>知道真相的我眼泪笑下来 Knowing the truth, I laughed myself to tears</td>
<td>Seq2Seq</td>
<td>知道真相的我眼泪掉下来。</td>
<td>After knowing the truth, I burst into tears.</td>
</tr>
<tr>
<td>ECM</td>
<td>Like</td>
<td>太可爱了，太可爱了！</td>
<td>Too lovely, it’s too lovely.</td>
</tr>
<tr>
<td></td>
<td>Happy</td>
<td>笑得我眼泪都出来了。</td>
<td>I can’t help laugh myself to tears.</td>
</tr>
<tr>
<td></td>
<td>Sad</td>
<td>知道真相的我眼泪掉下来。</td>
<td>After knowing the truth, I burst into tears.</td>
</tr>
<tr>
<td></td>
<td>Disgust</td>
<td>这是真的还是假的？</td>
<td>Is this true?</td>
</tr>
<tr>
<td></td>
<td>Angry</td>
<td>什么意思？没看懂。</td>
<td>What do you mean? I don’t understand.</td>
</tr>
</tbody>
</table>

---

### More Examples

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Post: 早上天气不错啊，哈哈哈哈哈</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>今天是个好天气。</td>
</tr>
<tr>
<td>喜欢 Like</td>
<td>今天的天气不错啊！</td>
</tr>
<tr>
<td>悲伤 Sad</td>
<td>天气好，心情不好。</td>
</tr>
<tr>
<td>厌恶 Disgust</td>
<td>这是哪里啊？求解释</td>
</tr>
<tr>
<td>愤怒 Angry</td>
<td>怎么了？什么情况？</td>
</tr>
<tr>
<td>快乐 Happy</td>
<td>今天的天气真好啊！</td>
</tr>
</tbody>
</table>

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

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

<table>
<thead>
<tr>
<th>Emotion</th>
<th>post: 承诺，就是一个骗子说给一个傻子听的。</th>
</tr>
</thead>
<tbody>
<tr>
<td>Other</td>
<td>承诺就是一个承诺</td>
</tr>
<tr>
<td>喜欢 Like</td>
<td>承诺是一个专一的傻子</td>
</tr>
<tr>
<td>悲伤 Sad</td>
<td>我就是一个悲伤的傻子</td>
</tr>
<tr>
<td>厌恶 Disgust</td>
<td>一个骗子说给一个傻子听</td>
</tr>
<tr>
<td>愤怒 Angry</td>
<td>不管你信不信，反正我信了！</td>
</tr>
<tr>
<td>快乐 Happy</td>
<td>我就是一个开心的傻子</td>
</tr>
</tbody>
</table>

---

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 Chatbots

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 Chatbots

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

Scene: Dining at a restaurant

我昨天晚上去了聚餐了  
I went to dinner yesterday night.

Friends?  
Place?  
Food?  
...  
Persons?  
Bill?  

WHO  
WHERE  
HOW-ABOUT  
HOW-MANY  
WHO

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

- 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

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

- 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 Chatbots

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 Questions in Chatbots

- **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 Questions in Chatbots

- **Type prediction at each decoding position**

<table>
<thead>
<tr>
<th>Decoding steps</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Post:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>我喜欢小动物(I like little animals)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Response:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>你(you)</td>
<td>0.09</td>
<td>0.02</td>
<td>0.01</td>
<td>0.85</td>
<td>1.00</td>
<td>0.01</td>
</tr>
<tr>
<td>喜欢(like)</td>
<td>0.26</td>
<td>0.35</td>
<td>0.71</td>
<td>0.14</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>兔子(rabbit)</td>
<td>0.65</td>
<td>0.63</td>
<td>0.28</td>
<td>0.01</td>
<td>0.00</td>
<td>0.97</td>
</tr>
<tr>
<td><strong>Interrogative</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>吗( particle)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Topic word</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>EOS</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: The table represents type prediction scores at each decoding step.*
Asking Questions in Chatbots

<table>
<thead>
<tr>
<th>Model</th>
<th>Perplexity</th>
<th>Distinct-1</th>
<th>Distinct-2</th>
<th>TRR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2Seq</td>
<td>63.71</td>
<td>0.0573</td>
<td>0.0836</td>
<td>6.6%</td>
</tr>
<tr>
<td>MA</td>
<td><strong>54.26</strong></td>
<td>0.0576</td>
<td>0.0644</td>
<td>4.5%</td>
</tr>
<tr>
<td>TA</td>
<td>58.89</td>
<td>0.1292</td>
<td>0.1781</td>
<td>8.7%</td>
</tr>
<tr>
<td>ERM</td>
<td>67.62</td>
<td>0.0355</td>
<td>0.0710</td>
<td>4.5%</td>
</tr>
<tr>
<td>STD</td>
<td>56.77</td>
<td>0.1325</td>
<td>0.2509</td>
<td>12.1%</td>
</tr>
<tr>
<td>HTD</td>
<td>56.10</td>
<td><strong>0.1875</strong></td>
<td><strong>0.3576</strong></td>
<td><strong>43.6%</strong></td>
</tr>
</tbody>
</table>

Table 1: Results of automatic evaluation.
Asking Questions in Chatbots

- Manual evaluation: Appropriateness, richness, willingness

<table>
<thead>
<tr>
<th>Models</th>
<th>Appropriateness</th>
<th>Richness</th>
<th>Willingness</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Win (%)</td>
<td>Lose (%)</td>
<td>Tie (%)</td>
</tr>
<tr>
<td>STD vs. Seq2Seq</td>
<td>42.0</td>
<td>38.6</td>
<td>19.4</td>
</tr>
<tr>
<td>STD vs. MA</td>
<td>39.6*</td>
<td>31.2</td>
<td>29.2</td>
</tr>
<tr>
<td>STD vs. TA</td>
<td>42.2</td>
<td>40.0</td>
<td>17.8</td>
</tr>
<tr>
<td>STD vs. ERM</td>
<td>43.4*</td>
<td>34.4</td>
<td>22.2</td>
</tr>
<tr>
<td>HTD vs. Seq2Seq</td>
<td>50.6**</td>
<td>30.6</td>
<td>18.8</td>
</tr>
<tr>
<td>HTD vs. MA</td>
<td>54.8**</td>
<td>24.4</td>
<td>20.8</td>
</tr>
<tr>
<td>HTD vs. TA</td>
<td>52.0**</td>
<td>38.2</td>
<td>9.8</td>
</tr>
<tr>
<td>HTD vs. ERM</td>
<td>64.8**</td>
<td>23.2</td>
<td>12.0</td>
</tr>
<tr>
<td>HTD vs. STD</td>
<td>52.0**</td>
<td>33.0</td>
<td>15.0</td>
</tr>
</tbody>
</table>

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

| 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 Purposes by Controlling Sentence Function
Sentence function indicates different conversational purposes.

- **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, ...)

**User:** I’m really hungry now.
Controlling Sentence Function

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

**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)

User: I’m really hungry now.
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

Diagram showing the process of controlling sentence function, with a mixture of words and a type controller. The diagram includes connections between different nodes representing various functions and topics.
Controlling Sentence Function

- **Dataset**: post-response pairs with sentence function labels

<table>
<thead>
<tr>
<th></th>
<th>#Post</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Training</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>#Post</td>
<td></td>
<td>1,963,382</td>
</tr>
<tr>
<td>#Response</td>
<td></td>
<td>Interrogative: 618,340</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Declarative: 672,346</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imperative: 672,696</td>
</tr>
<tr>
<td><strong>Validation</strong></td>
<td></td>
<td>24,034</td>
</tr>
<tr>
<td>#Post</td>
<td></td>
<td>Interrogative: 7,045</td>
</tr>
<tr>
<td>#Response</td>
<td></td>
<td>Declarative: 9,685</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Imperative: 7,304</td>
</tr>
<tr>
<td><strong>Test</strong></td>
<td></td>
<td>6,000</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>PPL</th>
<th>Dist-1</th>
<th>Dist-2</th>
<th>ACC</th>
</tr>
</thead>
<tbody>
<tr>
<td>c-seq2seq</td>
<td>57.14</td>
<td>949/.007</td>
<td>5177/.041</td>
<td>0.973</td>
</tr>
<tr>
<td>MA</td>
<td><strong>46.08</strong></td>
<td>745/.005</td>
<td>2952/.027</td>
<td>0.481</td>
</tr>
<tr>
<td>KgCVAE</td>
<td>56.81</td>
<td>1531/.009</td>
<td>10683/.070</td>
<td>0.985</td>
</tr>
<tr>
<td>Our Model</td>
<td>55.85</td>
<td><strong>1833/.008</strong></td>
<td><strong>15586/.075</strong></td>
<td><strong>0.992</strong></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Model</th>
<th>Interrogative</th>
<th>Declarative</th>
<th>Imperative</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours vs. c-seq2seq</td>
<td>0.534</td>
<td>0.536</td>
<td>0.896*</td>
</tr>
<tr>
<td>Ours vs. MA</td>
<td>0.802*</td>
<td>0.602*</td>
<td>0.675*</td>
</tr>
<tr>
<td>Ours vs. KgCVAE</td>
<td>0.510</td>
<td>0.626*</td>
<td>0.770*</td>
</tr>
</tbody>
</table>

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, \(p-value < 0.05\)).
Consistency: Behaving More Consistently with Personality
Personality of Chatting Machine

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

<table>
<thead>
<tr>
<th>General seq2seq model</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>User:</strong> Are you a boy or a girl?</td>
</tr>
<tr>
<td><strong>Chatbot:</strong> I am a boy.</td>
</tr>
<tr>
<td><strong>User:</strong> Are you a girl?</td>
</tr>
<tr>
<td><strong>Chatbot:</strong> Yes, I am a girl.</td>
</tr>
</tbody>
</table>

**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 Chatting Machine

- 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
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

<table>
<thead>
<tr>
<th>Profile key</th>
<th>Profile value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>汪仔(Wang Zai)</td>
</tr>
<tr>
<td>Age</td>
<td>三岁(3)</td>
</tr>
<tr>
<td>Gender</td>
<td>男 (Boy)</td>
</tr>
<tr>
<td>Hobbies</td>
<td>动漫(Cartoon)</td>
</tr>
<tr>
<td>Speciality</td>
<td>钢琴(Piano)</td>
</tr>
</tbody>
</table>

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 Chatting Machine**

Personality of Chatting Machine

- 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
### Personality of Chatting Machine

#### Post-level evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Nat.</th>
<th>Logic</th>
<th>Cor.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2Seq</td>
<td>71.8%</td>
<td>56.0%</td>
<td>23.8%</td>
</tr>
<tr>
<td>Seq2Seq +PV</td>
<td>72.0%</td>
<td>56.0%</td>
<td>41.3%</td>
</tr>
<tr>
<td>Seq2Seq +PVD</td>
<td>73.3%</td>
<td>52.5%</td>
<td>38.0%</td>
</tr>
<tr>
<td>Our Model -PD</td>
<td>82.7%</td>
<td>51.7%</td>
<td>38.3%</td>
</tr>
<tr>
<td>Our Model</td>
<td>83.3%</td>
<td>59.5%</td>
<td>45.8%</td>
</tr>
</tbody>
</table>

#### Session-level evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>Consistency</th>
<th>Variety</th>
</tr>
</thead>
<tbody>
<tr>
<td>Seq2Seq</td>
<td>1.3%</td>
<td>1.0%</td>
</tr>
<tr>
<td>Seq2Seq +PV</td>
<td>47.0%</td>
<td>1.3%</td>
</tr>
<tr>
<td>Seq2Seq +PVD</td>
<td>40.0%</td>
<td>7.5%</td>
</tr>
<tr>
<td>Our Model -PD</td>
<td>38.8%</td>
<td>16.0%</td>
</tr>
<tr>
<td>Our Model</td>
<td><strong>49.5%</strong></td>
<td><strong>27.8%</strong></td>
</tr>
</tbody>
</table>

#### Generated sample responses that exhibit session-level consistency

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

---

*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 leaves
  - 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

- **IsA**
  - lung disease
  - respiratory disease

- **Prevented_by**
  - avoiding triggers

- **Caused_by**
  - chest tightness
  - air pollution
Commonsense Knowledge in Chatbots

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

IsA: lung disease 
IsA: respiratory disease
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)

From ConceptNet

- lung disease
- respiratory disease
- asthma
- air pollution
- chest tightness
- avoiding triggers

IsA

Caused_by

Prevented_by
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.
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.
Commonsense Knowledge in Chatbots

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

<table>
<thead>
<tr>
<th>Post</th>
<th>Why are you so <strong>breakable</strong>?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge</td>
<td>(glass, RelatedTo, <strong>breakable</strong>), (brittle, RelatedTo, <strong>breakable</strong>), (rule, RelatedTo, <strong>breakable</strong>),</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>I’m not a <strong>OOV</strong>, I’m just a <strong>OOV</strong>.</td>
</tr>
<tr>
<td>MemNet</td>
<td>I’m not <strong>OOV</strong>. I’m just a really nice person.</td>
</tr>
<tr>
<td>CopyNet</td>
<td>I’m not. I’m just a lurker.</td>
</tr>
<tr>
<td>CCM</td>
<td>Because I’m a <strong>brittle</strong> man.</td>
</tr>
</tbody>
</table>
Commonsense Knowledge in Chatbots

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?
Commonsense Knowledge in Chatbots

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

[Diagram showing the flow of information through the encoder with key entities and neighboring entities]
**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.

\[
\begin{align*}
    s_{t+1} &= \text{GRU}(s_t, [c_t; c_t^g; c_t^k; e(y_t)]), \\
    e(y_t) &= [w(y_t); k_j],
\end{align*}
\]
**Dynamic graph attention**: first attend a graph, then to a triple within that graph, finally generate with the words in a graph.

\[
g_i = \sum_{n=1}^{N_{gi}} \alpha_n^s [h_n; t_n],
\]

\[
\alpha_n^s = \frac{\exp(\beta_n^s)}{\sum_{j=1}^{N_{gi}} \exp(\beta_j^s)},
\]

\[
\beta_n^s = (W_r r_n)^T \tanh(W_h h_n + W_t t_n),
\]
Dynamic graph attention: first attend a graph, then to a triple within that graph, finally generate with the words in a graph

\[
\begin{align*}
    c_t^g &= \sum_{i=1}^{N_G} \alpha_{ti}^g g_i, \\
    \alpha_{ti}^g &= \frac{\exp(\beta_{ti}^g)}{\sum_{j=1}^{N_G} \exp(\beta_{tj}^g)}, \\
    \beta_{ti}^g &= V_b^T \tanh(W_b s_t + U_b g_i),
\end{align*}
\]
**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.

\[
\begin{align*}
\alpha_{tj}^k &= \frac{\exp(\beta_{tj}^k)}{\sum_{n=1}^{N_g_i} \exp(\beta_{tn}^k)}, \\
\beta_{tj}^k &= k_j^T W_s s_t,
\end{align*}
\]

- Key Entity
- Neighboring Entity
- Attended Entity
- Not A Fact Triple
- Not A Fact Triple Vector
- Attended Graph
- Word Vector
- Previously Selected Triple Vector
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

**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

<table>
<thead>
<tr>
<th>Conversational Pairs</th>
<th>Commonsense KB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>3,384,185</td>
</tr>
<tr>
<td></td>
<td>Entity 21,471</td>
</tr>
<tr>
<td>Validation</td>
<td>10,000</td>
</tr>
<tr>
<td></td>
<td>Relation 44</td>
</tr>
<tr>
<td>Test</td>
<td>20,000</td>
</tr>
<tr>
<td></td>
<td>Triple 120,850</td>
</tr>
</tbody>
</table>

Table 1: Statistics of the dataset and the knowledge base.
# Commonsense Knowledge in Chatbots

## Automatic evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall</th>
<th>High Freq.</th>
<th>Medium Freq.</th>
<th>Low Freq.</th>
<th>OOV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ppx.</td>
<td>ent.</td>
<td>ppx.</td>
<td>ent.</td>
<td>ppx.</td>
</tr>
<tr>
<td>Seq2Seq</td>
<td>47.02</td>
<td>0.717</td>
<td>42.41</td>
<td>0.713</td>
<td>47.25</td>
</tr>
<tr>
<td>MemNet</td>
<td>46.85</td>
<td>0.761</td>
<td>41.93</td>
<td>0.764</td>
<td>47.32</td>
</tr>
<tr>
<td>CopyNet</td>
<td>40.27</td>
<td>0.96</td>
<td>36.26</td>
<td>0.91</td>
<td>40.99</td>
</tr>
<tr>
<td>CCM</td>
<td>39.18</td>
<td>1.180</td>
<td>35.36</td>
<td>1.156</td>
<td>39.64</td>
</tr>
</tbody>
</table>

## Manual evaluation

<table>
<thead>
<tr>
<th>Model</th>
<th>Overall</th>
<th>High Freq.</th>
<th>Medium Freq.</th>
<th>Low Freq.</th>
<th>OOV</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>app.</td>
<td>inf.</td>
<td>app.</td>
<td>inf.</td>
<td>app.</td>
</tr>
<tr>
<td>CCM vs. Seq2Seq</td>
<td>0.616</td>
<td>0.662</td>
<td>0.605</td>
<td>0.656</td>
<td>0.549</td>
</tr>
<tr>
<td>CCM vs. MemNet</td>
<td>0.602</td>
<td>0.647</td>
<td>0.593</td>
<td>0.656</td>
<td>0.566</td>
</tr>
<tr>
<td>CCM vs. CopyNet</td>
<td>0.600</td>
<td>0.640</td>
<td>0.606</td>
<td>0.669</td>
<td>0.586</td>
</tr>
</tbody>
</table>

*Sign-test, p-value<0.005*

---

Hao Zhou, Tom Yang, Minlie Huang, Haizhou Zhao, Jingfang Xu, Xiaoyan Zhu.
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.
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**.

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

\[ X_1 \text{ Today is Halloween.} \]
\[ X_2 \text{ Jack is so excited to go trick or treating tonight.} \]
\[ X_3 \text{ He is going to dress up like a monster.} \]
\[ X_4 \text{ The costume is real scary.} \]
\[ Y \text{ He hopes to get a lot of candy.} \]

Multi-Source Attention

\[ c_{kj}^{(i)} \]
\[ c_{xj}^{(4)} \]
\[ y^{(i)} \]
\[ g^{(i)} \]
\[ c_{hj}^{(i+1)} \]
\[ c_{xf}^{(i+1)} \]
\[ g^{(i+1)} \]
\[ h_j^{(i+1)} \]
Logic: Story Ending Generation

Attention to the knowledge base: static graph attention
## Experiment

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

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

*Table 1: Automatic and manual evaluation results.*
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

\( X_1 \): Martha is **cooking** a special **meal** for her family.

\( X_2 \): She wants everything to be just right for when they **eat**.

\( X_3 \): Martha perfects everything and puts her **dinner** into the **oven**.

\( X_4 \): Martha goes to **lay** down for a quick nap.

\( Y \): When she gets back to the **kitchen**, she sees a **burning** light on the **stove**.
Summary

- Emotion, personality, and knowledge is important to Chatbots

- Open-domain chatting machine is one of the most challenging AI tasks
  - Requires the ability of understanding semantics, knowledge, and situational context
  - Ability of making reasoning
  - Exhibits consistent personality

- Still a long way to go: existing generation models are still far from the expectation of real-world applications
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

- Minlie Huang, Tsinghua University
- aihuang@tsinghua.edu.cn
- [http://coai.cs.tsinghua.edu.cn/hml](http://coai.cs.tsinghua.edu.cn/hml)