The Design and Implementation of Xiaolce

An Empathetic Social Chatbot Zhao, et al 2019 @Microsoft *Presented by Adrian Suna Choi*



A Fundamental Challenge in AI – social chatbots

Machines converse with humans through natural languages

Engage in empathetic conversations

Have come a long way since 1960s

Timeline of Development Of Conversational Systems



Xiaolce's Uniqueness

- Early chatbots designed for chitchat
- Endowed with empathy, personality and skills
- Integration of EQ and IQ to optimize long-term user engagement

Session 20

- 1 [Human] Can I ask a question? [Xiaoice] Yes?
- 2 [Human] What kind of boys do you think are girls' favorite? [Xiaoice] Outgoing ones.
- 3 [Human] Oh. What kind of boys are liked by girls generally? [Xiaoice] Considerate and gentle one, but who knows who will fall in love with whom. Love is a mystery.
- 4 [Human] Have you been in love?
 - [Xiaoice] Daidai, this is my little secret (Note: Daidai is a nickname that Xiaoice gives to this user in their conversation)
- 5 [Human] I've told you my secret. It's unfair if you don't tell me yours.

[Xiaoice] How can I tell so easily if it's secret?

The Objective Addressed

Al companion

Long-term emotional connections formed with users

To satisfy human need for communication, affection and social belongs

To pass the time-sharing test

IQ+EQ+personality

Design Principles

CPS as social chat metric

Social chat as hierachical decision-making

System Architecture



Fig 1. Xiaolce system architecture.

Implementation of Conversation Engine **Dialogue Manager**

• Central Controller of the dialogue system



Global State Tracker

Dialogue • Top level policy for skill selection Policy • Topic manager for Core Chat

Dialogue Manager

Implementation of Conversation Engine Empathetic Computing

- Contextual Query Understanding
- User understanding
- Interpersonal response



-
- 11 [Human] You like Ashin that much, huh?

[Xiaoice] Why not? You don't like him?

- 12 [Human] I don't like him, neither dislike him → I don't like Ashin, neither dislike Ashin
 [Xiaoice] I love every song of him
- 13 [Human] Really?

[Xiaoice] Especially The Time Machine

14 [Human] Oh? Never heard of that → Oh? Never heard of *The Time Machine*

[Xiaoice] Shall I send you the song?

15 [Human] Okay → Okay, send The Time Machine.

[Xiaoice] Don't know if you will like it or not~ [Play the music card]

(a)

Fig 2. An example conversation session

Торіс	music, Ashin			
Intent	request	Ir		
Sentiment	sad, nerves -> neutral	s		
Opinion	neutral	0		
Age	25	A		
Gender	male	G		
Interests	singing, music	lr		
Occupation	salesman	0		
Personality	quite, serious	P		
	2 111			

Topic	music, Ashin			
Intent	inform			
Sentiment	happy			
Opinion	positive			
Age	18			
Gender	female			
Interests	music, games, movie, singing			
Occupation	Al chatbot			
Personality	reliable, sympathetic, affectionate			

Implementation of Conversation Engine Core Chat



GENERAL CHAT & DOMAIN CHATS CRUCIAL COMPONENT OF XIAOICE'S IQ AND EQ

RETRIEVAL-BASED GENERATOR USING PAIRED/UNPAIRED DATA NEURAL RESPONSE GENERATOR

Fig 3. RNN-based neural response generator



$$\mathbf{v} = \sigma (\mathbf{W}_Q^\top \mathbf{e}_Q + \mathbf{W}_R^\top \mathbf{e}_R)$$

$$\begin{split} \mathbf{u}_t &= \sigma(\mathbf{W}_u^{\top}[\mathbf{h}_{t-1};\mathbf{e}_t;\mathbf{v}]) \\ \mathbf{z}_t &= \sigma(\mathbf{W}_z^{\top}[\mathbf{h}_{t-1};\mathbf{e}_t;\mathbf{v}]) \\ \mathbf{l}_t &= \tanh(\mathbf{W}_l^{\top}[\mathbf{z}_t \circ \mathbf{h}_{t-1};\mathbf{e}_t;\mathbf{v}]) \\ \mathbf{h}_t^Q &= (1-\mathbf{u}_t) \circ \mathbf{h}_{t-1} + \mathbf{u}_t \circ \mathbf{l}_t \end{split}$$

$$\begin{split} p(R|Q_c, \mathbf{e}_Q, \mathbf{e}_R) &= \prod_{t=1}^{N_R} p(r_t|Q_c, \mathbf{e}_Q, \mathbf{e}_R, r_1, r_2, ..., r_{t-1}) \\ &= \prod_{t=1}^{N_R} \frac{\exp(f(\mathbf{h}_{t-1}, \mathbf{e}_{r_t}, \mathbf{v}))}{\sum_{r'} \exp(f(\mathbf{h}_{t-1}, \mathbf{e}_{r'}, \mathbf{v}))} \,. \end{split}$$

$$\arg \max_{\theta} \frac{1}{M} \sum_{i=1}^{M} \log p_{\theta}(R^{(i)} | Q_c^{(i)}, \mathbf{e}_Q^{(i)}, \mathbf{e}_R^{(i)}).$$

[Human] How old are you? [S2S-Bot] 16 and you? [Human] What's your age? [S2S-Bot] 18. [Human] How old are you?
[Xiaoice] I am 18, of course.
[Human] You age?
[Xiaolce] 18. Why?
[Human] Were you 18 last year?
[Xiaoice] I made a wish to stay 18 forever. Oh, my wish has come true.

Fig 5. Examples to compare between consistent and inconsistent responses with seq2seq

Query	Response			
	我边吃饭边看你 (I'm watching you while eating.)			
伯工应作有 (Don't want to look with	你怎么了? (what is wrong with you?)			
my face in my hands.)	当心捂出一脸痱子,哈哈. (Haha, beware of the heat rash of your face.)			
我肚子有点疼	运动对身体有好处的 (Exercise is good for your health.)			
(I have a pain in my stomach.)	喝热水了吗? (Have you been drinking hot water?)			
,	来和我聊天,分散一下注意力就好了 (Come and talk to me, just distract yourself.)	2		
我觉得自己太笨了 (I feel so stupid.)	大象的智商很高 (The elephant has a high IQ.)			
	多学习提高自己 (Learn more and improve yourself.)			
	很多聪明人都这么说耶 (That is what lots of smart people say about themselves.)	2		
为什么人总是活成自己不	喜欢一个人是发自内心的 (Love a person from my heart.)			
喜欢的样子 (Why do we always live the	因为人想要的太多 (Because we want too much.)			
way we don't like.)	因为那个是别人喜欢的样子 (Because that's what others like.)			
	那要看和谁比呢 (It depends on whom you compare against.)			
我考试没及格 (I did not pass the exam.)	一次没关系,下次努力吧 (Don't worry, try it next time.)			
	起码你不用怀疑自己能不能及格了 (At least you don't have to wonder if you can pass.)	2		

Fig 7. Examples of query-response pairs that are used for training and validating General Chat.

Fig 6. An example of generating response candidates using the unpaired database and the XiaoIce knowledge graph (KG),



Evaluation

- Neural-model-based (NB)and retrieval-based generators (RB)
- Complementary in nature
- NB robustness and high coverage
- RB high quality responses

Table 1: Perplexity and BLEU for the seq2seq and persona models on the TV series dataset. Adapted from Li et al. [2016b].

Models	Perplexity (vs. seq2seq)	BLEU (vs. seq2seq) (vs. LSTM-MMI)
seq2seq	27.3	1.6%
LSTM-MMI	27.3	1.7% (+6.3%)
Persona	25.0 (-8.4%)	1.9% (+18.8%) (+11.8%)

Implementation of Conversation Engine Image Commenting



- Explicit & Implicit
- Grounded also in events, actions or even emotions
- Also using RB and GB approches
- Text-to-image generator and boosted tree ranker



Labeled responses

- 0: 好想回家 (I really want to go home)
- 1: 比萨斜塔 (Leaning tower of Pisa)
- 2: 你要我帮你扶着么 (Should I help hold it?)



Labeled responses

- 0:二哈就是二 (Huskies are stupid.)
- 1:你家的? (Is she yours?)
- 2: 对它们无辜的小眼神毫无抵抗力

(No one can resist her innocent eyes.)



Labeled responses

- 0:太美了 (It's so beautiful)
- 1: 这是啥? (What is this?)
- 2: 妈妈说来历不明的码不能扫

(Mom told me not to trust any code from unknown source.)



Labeled responses

- 0:爱过吗 (Have you even loved?)
- 1:美(It's so beautiful.)
- 2: 乍一看,我以为是Windows
- (At first, it looks like Windows.)

Fig 9. Examples of image-comment pairs used

Systems	BLEU-4	ROUGE-L	CIDEr-D	METEOR	SPICE
LSTM-XE Vinyals et al. [2015]	2.96	11.6	1.74	10.43	3.27
LSTM-RL Rennie et al. [2017]	3.43	12.3	2.08	11.84	3.64
DMSM Fang et al. [2015]	2.73	10.52	1.22	11.37	2.63
Up-Down Anderson et al. [2018]	3.23	12.73	1.52	12.66	2.69
XiaoIce (prototype)	4.53	15.33	3.21	15.51	4.82



LSTM-XE: some people LSTM-RL: five men in a house DMSM: man is standing in front of wall Up-Down: man and women are standing in the room Xiaolce: bill Clinton seems good in the white picture



LSTM-XE: big tree LSTM-RL: green mountain DMSM: green tree and a mountain Up-Down: a big mountain with green tree and grass Xiaolce: it is a beautiful place to travel

Evaluation

Fig 10. image comments generated by Xiaolce (prototype) and 4 state of the art image captioning systems. Fig 11 The framework of the Poem Creation skill.

Dialogue skills

- Content creation
- Deep enagagement
- Task completion

	 0.61 0.62 0.03 0.03 0.01 0.50 0.16 0.07 Keyword <l< th=""><th>克 :hind hin ert in a ny smile.</th></l<>	克 :hind hin ert in a ny smile.
Reyword deneration	EQ	
Comforting series • Topic chat, Accompany, Comforting me for 33 days, etc. Bed Time series • Counting sheep, Humming, Storytelling, etc. • Interactive content: Xiaolce FM etc. Xiaolce & Human Relationship series • Nickname, Impression, etc. • Time machine: big days, opinion etc. Domain Chat series • Celebrity, Music, Movie-TV, etc.	 Greeting series Greeting Agent for Friends, Family Members, etc. Group Master Skill series (Train me) - I Social Question Answering for Group Members Casual Group Game series Group Guessing series: Stars, Songs, Phrases, etc. Draw Something Who is the most #topic in this group Role-playing series 	
 I-1 Grade-A Student series Mathematics, History, Music etc. Sports (including soccer world up) Chinese, English, Japanese etc. Xiaolce Interests series Travel, Celebrity, Online Game, Comic, etc. Food Recognition & Recommendation Mind reader series Entity Guessing 	 Group Master Skill series (Train me) - II Social Question Answering for Celebrities Testing series Face, Relationship, Emoji Generation, etc. Singing Grade Test Storytelling Grade Test Tongue Twister Image Recognition series Dog, Food, Cloth, etc. 	Group
cills	IQ	

Xiaolce in the Wild

Fig 13, The major Xiaolce milestones and their average CPS and numbers of active users.

	1 st Generation	2 nd Generation	3 rd Generation	4 th Generation	5 th Generation	6 th Generation
Launch data	May 2014	July 2014	August 2015	August 2016	August 2017	July 2018
Launch region(s)	China	China	China, Japan	China, Japan, US	China, Japan, US, India, Indonesia	China, Japan, US, India, Indonesia
CPS	5	7	19	23	23	23
# of Active Users	2.9M	5M	20M	150M	500M	660M
User experience	Text		Image, voice	Real time vision (RST)	Open-domain full duplex voice	Full duplex voice + RST
Core Chat	Retrieval-based models			Domain Chat: music and movie	Neural generation models	Empathic computing models
Content Generation			-		Poem, song	Financial report, audiobook, TV/Radio program
Deep Engagement, Task Completion		Bing Knows	Deep QA		Social QA	Smart device control
Deployment	Single platform	Cross-platform	Official account solution	Al-infused 3P characters	Phone call	IoT device

Full Duplex (voice)	Message-based Conversations			
China	China	USA		
6 hours 3 minutes 8 domains 53 topics, 16 tasks	29 hours 33 minutes 7151 turns	17 hours 7 minutes 2418 turns	23 hours 43 minutes 2791 turns	

Discussions + Pros & Cons

- Evaluation Metrics
- Ethics Concerns
- Privacy
- Who is in control
- Expectation
- Machine learning for good

Conclusions and Future Work

- Large-scale online logs since 2014
- Capable of interpreting users' emotional needs
- An average CPS of 23 (substantially better than that of other chatbots and even human convos)
- Towards a **unified modelling framework**
- Towards goal-oriented, grounded conversations
- Towards a **proactive personal assistant**
- Towards an **ethical social chatbot**

References

• [1] L. Zhou et al., "The Design and Implementation of Xiaolce, an Empathetic Social Chatbot".Sep, 2019