Affective Neural Response Generation

Nabiha Asghar,[1] Pascal Poupart,[1] Jesse Hoey,[1] Xin Jiang,[2] **Lili Mou**[1]

[1]University of Waterloo
[2]Noah's Ark Lab, Huawei Technologies

ECIR-18, Grenoble, France
March, 2018

Outline

Outline

Human-Computer Conversation

Human-computer conversation has long attracted interest in both academia and industry.

- Task/Domain-oriented systems
- Open-domain conversation systems

Task/Domain-Oriented Dialog Systems

- Transportation domain: TRAIN-95 (Ferguson et al., 1996)
- Education: AutoTutor (Graesser et al., 2005)
- Restaurant booking (Wen et al., 2016)

Approaches:

- Planning
- Rule-based, Slot-filling, etc.

Open-Domain Conversation

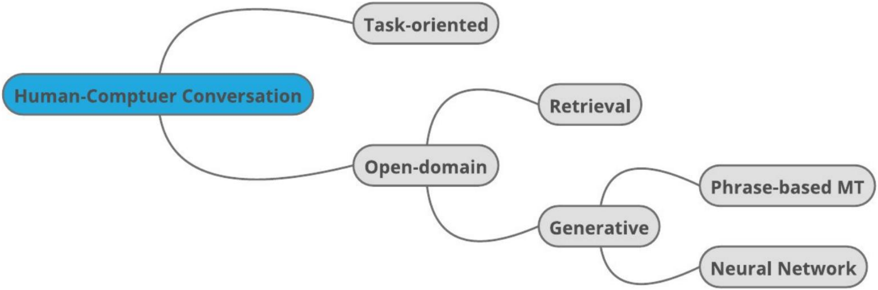
Why is chatbot-like conversation important?

- Tackles the problem of natural language understanding and generation
- Commercial needs

Approaches:

- Retrieval-based systems (Isbell et al., 2000; Wang et al., 2013)
- Generative systems
  - Phrase-based machine translation (Ritter et al., 2011)
  - Neural networks (seq2seq models) (Shang et al., 2015)

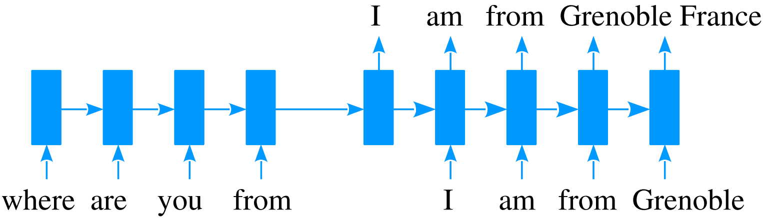
Where are we?



Open-domain, neural network-based, generative short-text conversation

Sequence-to-Sequence (Seq2Seq) Models

- Encoder-Decoder framework
  - Encodes the "user-issued utterance" (query)
  - Decodes the corresponding reply
- Recurrent Neural Network (w/ LSTM)
  - Serving as the encoder and decoder



Shortcoming of Seq2Seq Models

- Short, boring, meaningless replies
  - `I don't know`
  - `Me too`
- Previous work
  - Diversity-promoting training (Li et al., 2016) and decoding (Vijayakumar et al., 2016)
  - Content-introducing approaches (Mou et al., 2016)

Shortcoming of Seq2Seq Models

- Short, boring, meaningless replies
  - `I don't know`
  - `Me too`
- Previous work
  - Diversity-promoting training (Li et al., 2016) and decoding (Vijayakumar et al., 2016)
  - Content-introducing approaches (Mou et al., 2016)

However, they do not consider affect/emotional modeling of conversation.

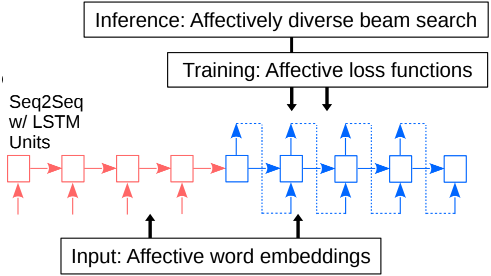
Our Paper . . .

- Explicitly models affect by psychologically inspired VAD embeddings
  - **Valence**: the pleasantness of a stimulus
  - **Arousal**: the intensity of emotion produced by a stimulus
  - **Dominance**: the degree of power exerted by a stimulus

Our Paper . . .

- Explicitly models affect by psychologically inspired VAD embeddings
  - **Valence**: the pleasantness of a stimulus
  - **Arousal**: the intensity of emotion produced by a stimulus
  - **Dominance**: the degree of power exerted by a stimulus

- Incorporates affective computing in the following aspects
  - Affective embeddings
  - Affective loss function
  - Affectively diverse deco



Outline

Basic Model

Seq2Seq Model $\mathbf{x} \mapsto \mathbf{y}$

- Input of RNN: word embeddings, mapping discrete words to real-valued vectors
- Training: cross-entropy loss (XENT)

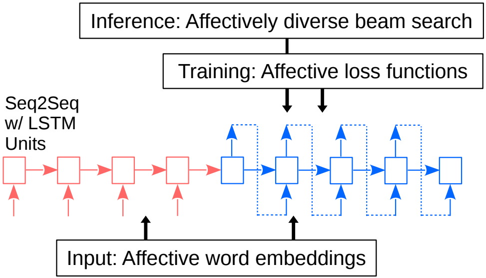
$$L_{\texttt{XENT}}(\theta) = -\log p(Y|X) = -\sum_{i=1}^{n} \log p(y_i|y_1, ..., y_{i-1}, X)$$

- Inference: Max *a posteriori* inference

$$\mathbf{y} = \arg\max_{Y}\{\log p_{\texttt{XENT}}(Y|X)\}$$

Overview

- Affective embeddings
- Affective loss function
- Affectively diverse decoding



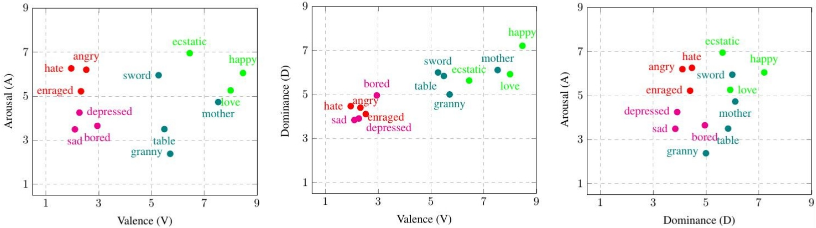
Affective Embeddings

- Traditional word embeddings (e.g., `word2vec`)
  - Learned by co-occurrence
  - Hard to capture sentiment information

E.g., "`The book is interesting`" vs "`The book is boring`"

- We leverage VAD vectors as external affect information
  - Psychologically engineered, Human annotated
  - Three dimension, representing
    - **Valence**: the pleasantness of a stimulus
    - **Arousal**: the intensity of emotion produced by a stimulus
    - **Dominance**: the degree of power exerted by a stimulus

VAD Examples



(a) V-A plot. (b) V-D plot. (c) D-A plot.

The simplest way to use VAD:

- Feed VAD to RNNs as input
- Concatenate VAD with traditional word embeddings

Intuition:

- Explicitly modeling words with affective information

Affective Loss Function

- Cross-entropy loss (XENT)

$$L_{\texttt{XENT}}(\theta) = -\log p(Y|X) = -\sum_{i=1}^{n} \log p(y_i|y_1, ..., y_{i-1}, X)$$

- Affective loss

$$L_{\texttt{Affect}}(\theta) = L_{\texttt{XENT}} + \text{Non-Affective Penalty}$$

Intuition:

- Explicitly modeling affective interaction between speakers

Affective Loss Function

Attempt#1: Minimizing Affective Dissonance

Two utterances tend to have the same VAD vectors

$$L^i_{\mathtt{DMIN}}(\theta) = -(1-\lambda)\log p(y_i|y_1, ..., y_{i-1}, X) + \lambda p(y_i) \left\| \sum_{j=1}^{|X|} \frac{\mathtt{W2AV}(x_j)}{|X|} - \sum_{k=1}^{i} \frac{\mathtt{W2AV}(y_k)}{i} \right\|_2$$

Affective Loss Function

Attempt#1: Minimizing Affective Dissonance

Two utterances tend to have the same VAD vectors

$$L^i_{\texttt{DMIN}}(\theta) = -(1-\lambda)\log p(y_i|y_1,...,y_{i-1},X) + \lambda p(y_i)\left\| \sum_{j=1}^{|X|} \frac{\texttt{W2AV}(x_j)}{|X|} - \sum_{k=1}^{i} \frac{\texttt{W2AV}(y_k)}{i} \right\|_2$$

Attempt#2: Maximizing Affective Dissonance

Two utterances tend to have different VAD vectors

$$L^i_{\texttt{DMAX}}(\theta) = -(1-\lambda)\log p(y_i|y_1,...,y_{i-1},X) - \lambda p(y_i)\left\| \sum_{j=1}^{|X|} \frac{\texttt{W2AV}(x_j)}{|X|} - \sum_{k=1}^{i} \frac{\texttt{W2AV}(y_k)}{i} \right\|_2$$

Affective Loss Function

Attempt#3: Maximizing Affective Content

$$L^i_{\mathtt{AC}}(\theta) = -(1-\lambda)\log p(y_i|y_1, ..., y_{i-1}, X) \\ - \lambda\, p(y_i)\left\|\mathtt{W2AV}(y_i) - \boldsymbol{\eta}\right\|_2$$

where $\boldsymbol{\eta}$ is the VAD for non-affective words.

Affective Loss Function

Attempt#3: Maximizing Affective Content

$$
\begin{aligned}
L^i_{\mathtt{AC}}(\theta) = & -(1-\lambda)\log p(y_i|y_1, \ldots, y_{i-1}, X) \\
& - \lambda\ p(y_i) \big\| \mathtt{W2AV}(y_i) - \boldsymbol{\eta} \big\|_2
\end{aligned}
$$

where $\boldsymbol{\eta}$ is the VAD for non-affective words.

Note:

- The affective embeddings are not learnable
- Hard selection is not differentiable
- Relax it by predicted probability

Affectively Diverse Decoding

The inference process decodes a sequence of words as the response.

- Greedy search: The best choice for each step may not be the best for the whole
- Beam search (BS): Keep top-B candidates and perform dynamic programming
- Diverse BS (DBS): Consider not only probability but also other scoring functions (e.g., diversity)

Affectively Diverse Decoding

The inference process decodes a sequence of words as the response.

- Greedy search: The best choice for each step may not be the best for the whole
- Beam search (BS): Keep top-B candidates and perform dynamic programming
- Diverse BS (DBS): Consider not only probability but also other scoring functions (e.g., diversity)
- Affectively DBS (ADBS): Measure the diversity in terms of VAD vectors

Diverse Beam Search

$$Y_{[t]}^{g} = \underset{\substack{\mathbf{y}_{1,[t]}^{g},\ldots,\mathbf{y}_{B',[t]}^{g} \\ \in Y_{[t-1]}^{g} \times V}}{\arg\max} \Bigg[\sum_{b=1}^{B'} \sum_{i=1}^{t} \log p(y_{b,i}^{g} | \mathbf{y}_{b,[i-1]}^{g}, X) + \lambda_g \Delta(Y_{[t]}^{1}, \ldots, Y_{[t]}^{g-1})[y_{b,t}^{g}] \Bigg]$$

- Maintain G groups
- Have B subsequences at each time step
- Expand the subsequence with one step (the vocabulary)
- Keep top-B subsequences after this time step

Affectively Diverse Beam Search

The design of the Δ function

- **Attempt#1**: Word Level

$$\Delta_W(Y_{[t]}^{1}, \ldots, Y_{[t]}^{g-1})[y_{b,t}^{g}] = -\sum_{j=1}^{g-1}\sum_{c=1}^{B'} \mathrm{sim}\left(\mathtt{W2AV}(y_{b,t}^{g}), \mathtt{W2AV}(y_{c,t}^{j})\right)$$

- **Attempt#2**: Sentence Level

$$\Delta_S(Y_{[t]}^{1}, \ldots, Y_{[t]}^{g-1})[y_{b,t}^{g}] = \sum_{j=1}^{g-1}\sum_{c=1}^{B'} \mathrm{sim}\left(\Psi(\mathbf{y}_{b,[t]}^{g}), \Psi(\mathbf{y}_{c,[t]}^{j})\right)$$

Outline

Dataset and Settings

- Cornell Movie Dialogs Corpus
- ∼300k utterance-response pairs
- 1024d `word2vec` and hidden states
- For other tedious settings, please see arXiv:1709.03968

Evaluation

Human annotation for 100 test samples

- 5 annotators
- 3 aspects
  - Syntactic coherence (Does the response make grammatical sense?)
  - Naturalness (Could the response have been plausibly produced by a human?)
  - Emotional appropriateness (Is the response emotionally suitable for the prompt?)
- 3 scores: 0=bad, 1=borderline, 2=good

Feiss' $\kappa = 0.44$ (Moderate agreement)

Experiment#1: Affective Embeddings

Model	**Syntactic Coherence**	**Natural**	**Emotional Approp.**
Word embed.	1.48	0.69	0.41
Word + Affective embeddings.	**1.71**	**1.05**	**1.01**

- Word embeddings learned end-to-end during training are not sufficient to capture emotional aspects
- Using additional VAD knowledge makes the model more aware of affect

Experiment#2: Affective Loss Function

Model	Syntactic Coherence	Natural	Emotional Approp.
$L_{\texttt{XENT}}$	1.48	0.69	0.41
$L_{\texttt{DMIN}}$	**1.75**	0.83	0.56
$L_{\texttt{DMAX}}$	1.74	0.85	0.58
$L_{\texttt{AC}}$	1.71	**0.95**	**0.71**

- DMAX and DMIN not as good as maximizing affective content (AC)
- The interaction of affect is more complicated than forcing it to some direction
- Future work: incorporating the "affect control theory" (Hoey et al., 2016)

Experiment#3: Affectively Diverse Beam Search

Model	Syntactic Diversity	Affective Diversity	No. of Emotionally Approp. Responses
BS	1.23	0.87	0.89
H-DBS†	1.47	0.79	0.78
WL-ADBS	**1.51**	1.25	1.30
SL-ADBS	1.45	**1.31**	**1.33**

†Hamming-based beam search

Annotation is performed on top-3 decoded sentences.

Putting all together

Model	Syntactic Coherence	Naturalness	Emotional Appropriateness
Pure Seq2Seq	1.48	0.69	0.41
Seq2Seq + Affective Embeddings	1.71	1.05	1.01
Seq2Seq + Affective Embeddings & Loss	**1.76**	1.03	1.07
Seq2Seq + Affective Embeddings & Loss & Decoding	1.69	**1.09**	**1.10**

Case Study

Message	Model	Response
I wish I was a billion-aire.	$L_{\texttt{XENT}}$	Um, you know...
	$L_{\texttt{DMIN}}$	But you work for the casino?
	$L_{\texttt{DMAX}}$	You are personally responsible.
	$L_{\texttt{AC}}$	It's suicide!
	$L_{\texttt{XENT}}+$ Aff. Emb.	Don't trouble yourself about it.
	$L_{\texttt{AC}}+$ Aff. Emb.	Would it matter in the end.
	$L_{\texttt{AC}}+$ Aff. Emb. + SL-ADBS	Would it change you?
I hate you.	$L_{\texttt{XENT}}$	It was supposed to be.
	$L_{\texttt{DMIN}}$	Aww c all guys do...
	$L_{\texttt{DMAX}}$	Keeps fairies away!
	$L_{\texttt{AC}}$	I haven't done anything crazy.
	$L_{\texttt{XENT}}+$ Aff. Emb.	Don't say that!
	$L_{\texttt{AC}}+$ Aff. Emb.	I still love you!
	$L_{\texttt{AC}}+$ Aff. Emb. + SL-ADBS	I don't want to fight you.

Outline

Conclusion

Our paper: Affective neural response generation

- Affective embeddings
- Affective loss functions
- Affectively diverse beam search

Future work: Affective interactive/human-in-the-loop conversation with affective control theory

George Ferguson, James Allen, and Brad Miller. 1996. TRAINS-95: Towards a mixed-initiative planning assistant. In *AIPS*. pages 70–77.

Arthur C Graesser, Patrick Chipman, Brian C Haynes, and Andrew Olney. 2005. AutoTutor: An intelligent tutoring system with mixed-initiative dialogue. *IEEE Trans. Education* 48(4):612–618.

Jesse Hoey, Tobias Schröder, and Areej Alhothali. 2016. Affect control processes: Intelligent affective interaction using a partially observable markov decision process. *Artificial Intelligence* 230:134–172.

Charles Lee Isbell, Michael Kearns, Dave Kormann, Satinder Singh, and Peter Stone. 2000. Cobot in LambdaMOO: A social statistics agent. In *AAAI*. pages 36–41.

Jiwei Li, Michel Galley, Chris Brockett, Jianfeng Gao, and Bill Dolan. 2016. A diversity-promoting objective function for neural conversation models. In *NAACL*. pages 110–119.

Lili Mou, Yiping Song, Rui Yan, Ge Li, Lu Zhang, and Zhi Jin. 2016. Sequence to backward and forward sequences: A content-introducing approach to generative short-text conversation. In *COLING*. pages 3349–3358.

Alan Ritter, Colin Cherry, and William B Dolan. 2011. Data-driven response generation in social media. In *EMNLP*. pages 583–593.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural responding machine for short-text conversation. In *ACL-IJCNLP*. pages 1577–1586.

Ashwin K Vijayakumar, Michael Cogswell, Ramprasath R Selvaraju, Qing Sun, Stefan Lee, David Crandall, and Dhruv Batra. 2016. Diverse beam search: Decoding diverse solutions from neural sequence models. *arXiv preprint arXiv:1610.02424* .

Hao Wang, Zhengdong Lu, Hang Li, and Enhong Chen. 2013. A dataset for research on short-text conversations. In *EMNLP*. pages 935–945.

Tsung-Hsien Wen, Milica Gasic, Nikola Mrksic, Lina M Rojas-Barahona, Pei-Hao Su, Stefan Ultes, David Vandyke, and Steve Young. 2016. A network-based end-to-end trainable task-oriented dialogue system. *arXiv preprint arXiv:1604.04562* .

Thanks for listening

Question?