Conversation Modeling for Computational Social Science

KAIST | School of Computing | Alice Oh
November 5, 2019
• Developing conversation models for computational approaches to questions in history, psychology, and human social behavior
• EMNLP 2018: Conversational decision making model for predicting king’s decisions in the Annals of Joseon Dynasty
• NAACL 2019: Conversation model fine-tuning for classifying client utterances in counseling dialogues
• EMNLP 2019: Variational hierarchical user-based conversation model
Conversational Decision Making Model for Predicting the King’s Decision in the Annals of Joseon Dynasty

JinYeong Bak & Alice Oh, EMNLP 2018
Background

• Joseon Dynasty (1392 – 1897)
  – Longest monarchy in Korean peninsula
  – Most of the daily records of the monarchial system have been preserved and translated into modern Korean
  – 28 kings

• Historians still argue over the interpretations of the records about the legacy of the kings

• A recent discussion reconsiders one of the kings who has been called a tyrant
Research Question

• Can we analyze the historical records with a machine learning model to automatically classify the king’s decision process?
The king follows Official C’s suggestion.
## Corpus

<table>
<thead>
<tr>
<th>Kings</th>
<th>Articles</th>
<th>Utterances</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>13,216</td>
<td>95,615</td>
<td>4,502</td>
</tr>
</tbody>
</table>

(a) Basic statistics of the corpus

<table>
<thead>
<tr>
<th>Order</th>
<th>Accept</th>
<th>1,457</th>
</tr>
</thead>
<tbody>
<tr>
<td>Approve</td>
<td>Reject</td>
<td>818</td>
</tr>
<tr>
<td>Disapprove</td>
<td>Discuss</td>
<td>6,214</td>
</tr>
</tbody>
</table>

(b) Distribution of articles over decisions

Table 1: Statistics of a conversational meeting records and king’s decisions from the AJD
### Classification Results

<table>
<thead>
<tr>
<th>Method</th>
<th>Micro $F_1$</th>
<th>Macro Prec</th>
<th>Macro Rec</th>
<th>Macro $F_1$</th>
<th>W-avg $F_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Majority of classes</td>
<td>0.472</td>
<td>0.079</td>
<td>0.167</td>
<td>0.107</td>
<td>0.303</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>0.479</td>
<td>0.173</td>
<td>0.176</td>
<td>0.126</td>
<td>0.321</td>
</tr>
<tr>
<td>SVM linear</td>
<td>0.381</td>
<td>0.249</td>
<td>0.246</td>
<td>0.246</td>
<td>0.383</td>
</tr>
<tr>
<td>SVM RBF</td>
<td>0.487</td>
<td>0.236</td>
<td>0.186</td>
<td>0.142</td>
<td>0.337</td>
</tr>
<tr>
<td>Naive Bayes with speaker</td>
<td>0.466</td>
<td>0.268</td>
<td>0.177</td>
<td>0.135</td>
<td>0.323</td>
</tr>
<tr>
<td>SVM linear with speaker</td>
<td>0.423</td>
<td>0.292</td>
<td>0.259</td>
<td>0.243</td>
<td>0.403</td>
</tr>
<tr>
<td>SVM RBF with speaker</td>
<td>0.472</td>
<td>0.079</td>
<td>0.167</td>
<td>0.107</td>
<td>0.303</td>
</tr>
<tr>
<td>fastText w/o word vector</td>
<td>0.487</td>
<td>0.158</td>
<td>0.193</td>
<td>0.150</td>
<td>0.349</td>
</tr>
<tr>
<td>fastText</td>
<td>0.499</td>
<td>0.315</td>
<td>0.225</td>
<td>0.215</td>
<td>0.402</td>
</tr>
<tr>
<td>CDMM w/o speaker</td>
<td>0.481</td>
<td>0.176</td>
<td>0.214</td>
<td>0.178</td>
<td>0.379</td>
</tr>
<tr>
<td>CDMM with speaker (random init)</td>
<td><strong>0.504</strong></td>
<td>0.258</td>
<td>0.227</td>
<td>0.208</td>
<td>0.401</td>
</tr>
<tr>
<td>CDMM with speaker (pre-trained)</td>
<td>0.476</td>
<td><strong>0.329</strong></td>
<td><strong>0.307</strong></td>
<td><strong>0.313</strong></td>
<td><strong>0.456</strong></td>
</tr>
</tbody>
</table>
Attention Weights

Figure 3: Attention weight distribution of words for each class

Figure 4: Attention weight distribution of word for each class from kings and officials
# Speakers

<table>
<thead>
<tr>
<th>Name (Eng)</th>
<th>Position</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sin Sukju</td>
<td>Secretary</td>
<td>Order</td>
</tr>
<tr>
<td>Jeong Changson</td>
<td>Secretary</td>
<td>Order</td>
</tr>
<tr>
<td>Kim Jonkyung</td>
<td>Local gov</td>
<td>Approve</td>
</tr>
<tr>
<td>Kim Neuk</td>
<td>Local gov</td>
<td>Approve</td>
</tr>
<tr>
<td>Gwon Jin</td>
<td>Local gov</td>
<td>Disapprove</td>
</tr>
<tr>
<td>Kim Seup</td>
<td>Remonstrator</td>
<td>Disapprove</td>
</tr>
<tr>
<td>Hwang Hui</td>
<td>Central gov</td>
<td>Accept</td>
</tr>
<tr>
<td>Han Myeonghoe</td>
<td>Central gov</td>
<td>Accept</td>
</tr>
<tr>
<td>Kim Jikyung</td>
<td>Remonstrator</td>
<td>Reject</td>
</tr>
<tr>
<td>Sung Damnyeon</td>
<td>Remonstrator</td>
<td>Reject</td>
</tr>
</tbody>
</table>
Contributions

• Presented a computational analysis of historical records
• Showed an example of using machine learning in digital humanities in a low-resource language
• Opened up new collaborations with historians
Conversation Model Fine-Tuning for Classifying Client Utterances in Counseling Dialogues
Sungjoon Park & Alice Oh, NAACL 2019
Text-Based Counseling Dialogues

- App/Web for text-based counseling
- Pros
  - Cost effective
  - Flexible time
- Cons
  - Lack of continuity
  - Potentially less effective outcome
Text-Based Counseling Dialogues

- App/Web for text-based counseling
- Pros
  - Cost effective
  - Flexible time
- Cons
  - Lack of continuity
  - Potentially less effective outcome

Can we address these problems?
Identifying Counselors’ Needs

- Summarization of previous sessions
  - Automatic categorization of client utterances
Identifying Counselors’ Needs

- Summarization of previous sessions
  → Automatic categorization of client utterances

- Existing coding schemes in psychology literature
  - Not suitable for the text-only environment
  - Too detailed and not suitable for automatic methods
Identifying Counselors’ Needs

• Summarization of previous sessions
  → Automatic categorization of client utterances

• Existing categorizations in psychology literature
  • Not suitable for the text-only environment
  • Too detailed and not suitable for automatic methods

• Work with counselors to develop new categorization scheme
  • Suitable for the text-only environment
    → Not dependent on non-verbal cues
  • Fewer categories
    → To make annotation possible
  • Still meaningful for identifying counseling progress
<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Informative</th>
<th>Client Factors</th>
<th>Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category Name</td>
<td>Factual Information (Fact.)</td>
<td>Anecdotal Experience (Anec.)</td>
<td>Appealing Problem (Prob.)</td>
</tr>
<tr>
<td>Explanation</td>
<td>Brief mention of factual information</td>
<td>Client experiences contributing to the appealing problem</td>
<td>Clients factors related to the appealing problem</td>
</tr>
</tbody>
</table>
| Example       | • Objective fact | • Experience with others | • Positive prediction | • Message to counselor | |}

- Objective fact
- Living conditions
- Demographic information
- Limited conditions
- Experience with others
- Comments from others
- Trauma
- Interpersonal situations
- Negative Emotion
- Cognitive distortion
- Interpersonal problems
- Family problems
- Expectation, Determination
- Coping behaviors
- Self-awareness
- Appointment scheduling
- Questions about the sessions
- Objective fact
- Living conditions
- Demographic information
- Limited conditions
- Experience with others
- Comments from others
- Trauma
- Interpersonal situations
- Negative Emotion
- Cognitive distortion
- Interpersonal problems
- Family problems
- Expectation, Determination
- Coping behaviors
- Self-awareness
- Appointment scheduling
- Questions about the sessions
- Objective fact
- Living conditions
- Demographic information
- Limited conditions
- Experience with others
- Comments from others
- Trauma
- Interpersonal situations
- Negative Emotion
- Cognitive distortion
- Interpersonal problems
- Family problems
- Expectation, Determination
- Coping behaviors
- Self-awareness
- Appointment scheduling
- Questions about the sessions
Modeling Client Language in Counseling Conversations

Overview

① Text-based Online Counseling Dialogues
- Distinctive roles of speakers
- Multiple utterances in a turn

Counselor

Client

"Appealing Problem"

money, families, relationships... makes me feel down

② Categorize Client’s Utterances
- Suitable for the text-only environment
- Available as a labeling scheme
- Meaningful to counselors

Client

Counselor

It sounds like you want to be the owner of your life.

Why can’t you do that?

I want to be active in maintaining my relationships,

But it just bothers me.

③ Predict Categories of Client’s Utterances

Model: ConvMFiT (Conversation Model Fine-Tuning)

- Factual Information
- Anecdotal Experience
- Appealing Problems
- Psychological Change
- Counseling Process

Classification

Attention & Dense

Seq2Seq Layers

Dependency of Language Models

Counselor’s Language Model

Client’s Language Model

Client’s Context U

Client’s Target U

Pre-trained Conversation Model

Task-specific Layers
Counseling Conversations

- Distinctive roles of speakers
- Multiple utterances in a turn
Text-based Counseling Data

Statistics

- 5 counselors annotated 100 sessions
- Total corpus of 1,448 sessions
- Annotation is expensive!
- All data are anonymized

<table>
<thead>
<tr>
<th></th>
<th>Counselor</th>
<th>Client</th>
</tr>
</thead>
<tbody>
<tr>
<td># of words</td>
<td>6.26</td>
<td>5.91</td>
</tr>
<tr>
<td>Std.</td>
<td>6.25</td>
<td>14.62</td>
</tr>
<tr>
<td># of chars</td>
<td>25.71</td>
<td>24.31</td>
</tr>
<tr>
<td>Std.</td>
<td>23.44</td>
<td>61.10</td>
</tr>
<tr>
<td># of utters</td>
<td>163.2</td>
<td>238.72</td>
</tr>
<tr>
<td>Std.</td>
<td>236.97</td>
<td>578.49</td>
</tr>
</tbody>
</table>

Table 2: Descriptive statistics of labeled counseling dialogues.
ConvMFit Model

- Task-specific Layers
- Classification
  - Attention & Dense
  - Seq2Seq Layers
- Dependency of Language Models
  - Counselor's Language Model
  - Client's Language Model
- Pre-trained Conversation Model
  - Counselor's U
  - Client's Context U
  - Client's Target U
ConvMFit Model

(1) Embedding Layer

- Subword-level Korean Word Vectors (Park et al., ACL 2018): no out-of-vocabs

- Trained on corpus of Korean Wikipedia, Online News articles, Sejong Corpus (130M)
ConvMFit Model

(2) Language Models

- Use pre-trained Korean word vectors (fixed)
- 3-layer LSTM (left-to-right) Language Models with various regularization techniques (Merity et al, ICLR 2018)
- Trained on utterances of counselors and clients, respectively
ConvMFit Model

(3) Conversation Model
- Use pre-trained Language Models and Word Vectors
- Add 2-layer LSTMs (left-to-right) over Language Models and train conversation model with Language Model losses as regularizers
- Trained on (unlabeled) conversations between a counselor and a client
ConvMFit Model

(4) Task-specific Layers

- Add 2-layer LSTMs, an attention layer, and a softmax layer (randomly initialized)
ConvMFit Model

**Counselor's Language Model**

**Client's Language Model**

**Dependency of Language Models**

**Pre-trained Conversation Model**

**Task-specific Layers**

**Classification**

**Attention & Dense**

**Seq2Seq Layers**

**Gradual Unfreezing** (as in ULMFiT)

- Unfreeze layers every other epoch until all layers are tuned
- Early stop when validation loss is minimized

FINE-TUNING

**<Fine-Tuning>**

**Unfreeze layers every other epoch until all layers are tuned**

**Early stop when validation loss is minimized**
## Classification Results

<table>
<thead>
<tr>
<th>No.</th>
<th>Dep.</th>
<th>Model</th>
<th>F1 (Fact.)</th>
<th>F1 (Anec.)</th>
<th>F1 (Prob.)</th>
<th>F1 (Chan.)</th>
<th>F1 (Proc.)</th>
<th>Macro Prec.</th>
<th>Macro Rec.</th>
<th>Macro F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>RF</td>
<td>0.000</td>
<td>0.564</td>
<td>0.420</td>
<td>0.000</td>
<td>0.723</td>
<td>0.476</td>
<td>0.269</td>
<td>0.341</td>
</tr>
<tr>
<td>2</td>
<td>X</td>
<td>SVM(rbf)</td>
<td>0.012</td>
<td>0.683</td>
<td>0.457</td>
<td>0.000</td>
<td>0.766</td>
<td>0.602</td>
<td>0.385</td>
<td>0.384</td>
</tr>
<tr>
<td>3</td>
<td>X</td>
<td>CNN</td>
<td>0.211</td>
<td>0.528</td>
<td>0.506</td>
<td>0.128</td>
<td>0.706</td>
<td>0.450</td>
<td>0.397</td>
<td>0.416</td>
</tr>
<tr>
<td>4</td>
<td>X</td>
<td>RNN</td>
<td>0.193</td>
<td>0.574</td>
<td>0.570</td>
<td>0.046</td>
<td>0.770</td>
<td>0.607</td>
<td>0.375</td>
<td>0.431</td>
</tr>
<tr>
<td>5</td>
<td>X</td>
<td>ULMFiT</td>
<td>0.205</td>
<td>0.641</td>
<td>0.591</td>
<td>0.057</td>
<td>0.784</td>
<td>0.613</td>
<td>0.413</td>
<td>0.455</td>
</tr>
<tr>
<td>6</td>
<td>O</td>
<td>Seq2Seq</td>
<td>0.263</td>
<td>0.662</td>
<td>0.678</td>
<td>0.226</td>
<td>0.823</td>
<td>0.695</td>
<td>0.472</td>
<td>0.530</td>
</tr>
<tr>
<td>7</td>
<td>O</td>
<td>HRED</td>
<td>0.261</td>
<td>0.706</td>
<td>0.675</td>
<td>0.193</td>
<td>0.820</td>
<td>0.680</td>
<td>0.475</td>
<td>0.531</td>
</tr>
<tr>
<td>8</td>
<td>O</td>
<td>ConvMFIT</td>
<td><strong>0.441</strong></td>
<td><strong>0.761</strong></td>
<td><strong>0.726</strong></td>
<td><strong>0.447</strong></td>
<td><strong>0.835</strong></td>
<td><strong>0.716</strong></td>
<td><strong>0.602</strong></td>
<td><strong>0.642</strong></td>
</tr>
</tbody>
</table>
### Effect of Each Component for Classification

<table>
<thead>
<tr>
<th>No.</th>
<th>Emb.</th>
<th>LM</th>
<th>Conv. seq2seq</th>
<th>Grad. Unf.</th>
<th>Task. seq2seq</th>
<th>F1 (Fact.)</th>
<th>F1 (Anec.)</th>
<th>F1 (Prob.)</th>
<th>F1 (Chan.)</th>
<th>F1 (Proc.)</th>
<th>Macro Prec</th>
<th>Macro Rec</th>
<th>Macro F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>O</td>
<td>.032</td>
<td>.706</td>
<td>.602</td>
<td>.222</td>
<td>.711</td>
<td>.418</td>
<td>.575</td>
<td>.455</td>
</tr>
<tr>
<td>2</td>
<td>O</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>O</td>
<td>.043</td>
<td>.758</td>
<td>.661</td>
<td>.258</td>
<td>.748</td>
<td>.459</td>
<td>.662</td>
<td>.494</td>
</tr>
<tr>
<td>3</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>O</td>
<td>.425</td>
<td>.782</td>
<td>.727</td>
<td>.365</td>
<td>.831</td>
<td>.587</td>
<td>.719</td>
<td>.626</td>
</tr>
<tr>
<td>4</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>.441</td>
<td>.761</td>
<td>.726</td>
<td>.447</td>
<td>.835</td>
<td>.602</td>
<td>.716</td>
<td>.642</td>
</tr>
<tr>
<td>4-1</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>.307</td>
<td>.738</td>
<td>.666</td>
<td>.304</td>
<td>.802</td>
<td>.513</td>
<td>.687</td>
<td>.563</td>
</tr>
<tr>
<td>4-2</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>X</td>
<td>O</td>
<td>.417</td>
<td>.768</td>
<td>.721</td>
<td>.399</td>
<td>.824</td>
<td>.591</td>
<td>.695</td>
<td>.626</td>
</tr>
<tr>
<td>4</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td><strong>.441</strong></td>
<td><strong>.761</strong></td>
<td><strong>.726</strong></td>
<td><strong>.447</strong></td>
<td><strong>.835</strong></td>
<td><strong>.602</strong></td>
<td><strong>.716</strong></td>
<td><strong>.642</strong></td>
</tr>
<tr>
<td><strong>Factual Information</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Counselor</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>nice to meet you</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>of course first</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I see. now</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>. . [clients name]s</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>well . .</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Client</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>never visited neuropsychiatry</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>female, worker</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>never received psychological counseling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>motivation of counseling</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>age #, female</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Qualitative Analysis

<table>
<thead>
<tr>
<th>Appealing Problem</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Counselor</strong></td>
</tr>
<tr>
<td>right . all</td>
</tr>
<tr>
<td>but now the relationship is</td>
</tr>
<tr>
<td>nice to meet you</td>
</tr>
<tr>
<td>you told me well</td>
</tr>
<tr>
<td>more comfortable?</td>
</tr>
<tr>
<td><strong>Client</strong></td>
</tr>
<tr>
<td>sick and sad</td>
</tr>
<tr>
<td>have many thoughts</td>
</tr>
<tr>
<td>keep thinking</td>
</tr>
<tr>
<td>I think I did it</td>
</tr>
<tr>
<td>no future and frustrating</td>
</tr>
</tbody>
</table>
### Qualitative Analysis

<table>
<thead>
<tr>
<th>Psychological Change</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Counselor</strong></td>
</tr>
<tr>
<td>right . and</td>
</tr>
<tr>
<td>if you do that someday</td>
</tr>
<tr>
<td>suddenly I did</td>
</tr>
<tr>
<td>I hope so</td>
</tr>
<tr>
<td>question arises</td>
</tr>
<tr>
<td><strong>Client</strong></td>
</tr>
<tr>
<td>did not think so but</td>
</tr>
<tr>
<td>I was so shocked</td>
</tr>
<tr>
<td>it could have been longer</td>
</tr>
<tr>
<td>but looking back</td>
</tr>
<tr>
<td>I think I will</td>
</tr>
</tbody>
</table>
Contributions

• Developed 5 categories of client utterances

• Proposed fine-tuning conversation models

• Initiated computational psychotherapy research
Variational Hierarchical User-based Conversation Model
JinYeong Bak & Alice Oh, EMNLP 2019
I love this car!

It’s good to hear.

What is the best restaurant for the event?

I think AAA is the best one.

Today’s my mother’s birthday!

Tell her I said happy birthday.

My dog passed away today

It is so sad. I hope you feel better.

Open-domain conversation corpus

Training

Conversation Model

Inference

Generating the response
Motivation

A: How’s the wedding going?
B: I just booked my trip to SF

A: Next month
B: Guess the wedding is tomorrow!

A: I’m in a hotel, waiting for tomorrow
B: Next month

Hi! C! We are in England next week!
Oh! Hope it is fun!

To D. My computer is broken!
Could you describe the error in more detail?
Motivation

• Conversational context depends on the speaker
  – Same dyad is likely to have similar conversations
  – A speaker may have different conversations with different conversational partners
Motivation

• Existing conversation models that consider the speakers
  – Tend to generate the same response for the same speaker even when the previous utterances are different
  – Rely too much on the speaker rather than the content of the previous utterances

Speaker Model [Li et al., ACL 2016]
Ideas

• Use a latent variable for the context from speakers
  – Learns the context of conversations between two speakers and updates their representation during training
  – Infers the context of new conversation by conditioning on the speakers for testing

• Provide speaker information to response generator indirectly
  – Mixes the speakers’ preference and the context variable
VHUCM

Variational Hierarchical User-based Conversation Model (VHUCM)

• Utterances: \( x_1, \ldots, x_t, \ldots \)
• Speaker embeddings
VHUCM

Variational Hierarchical User-based Conversation Model (VHUCM)

- Utterances: $x_1, \ldots, x_t, \ldots, x_M$
- Speaker embeddings: $s_a, s_b$
- Hierarchical RNNs
  - Encoder RNN: $h^{enc}$
  - Context RNN: $h^{cxt}$
  - Decoder RNN: $h^{dec}$
- Latent variables
  - Global variable: $z^{conv}$
  - Local variable: $z^{utt}$
VHUCM - Generation

1. For a conversation
   1. Global variable $z^{conv}$ represents the context of two speakers’ previous conversations

2. For a response of $x_{t-1}$
   1. Encoder RNN $h^{enc}$ encodes the $x_{t-1}$
   2. Context RNN $h^{ctx}_t$ infers the conversational context
   3. Local variable $z^{utt}_t$ represents the speaker $s_t$’s preference for $x_{t-1}$
   4. Decoder RNN $h^{dec}_t$ generates next utterance $x_t$
Conversation Corpus

• Naturally-occurring conversations
• Many conversations between two speakers
• Each speaker has conversations with multiple partners
Twitter Conversation Corpus

@MadonnaMDNADay love the new album - every single song is incredible. congrats girl! 🎵 Girl Gone Wild by Madonna — path.com/p/1zoIB

@britneyspears please come on stage and kiss me again. I miss you!!

@MadonnaMDNADay Tempting...

@britneyspears Are you gonna make me work for this?

@MadonnaMDNADay Why of course!
Twitter Conversation Corpus

Basic statistics

– # users: 27,152
– # dyads: 107,611
– # conversations: 770,739
– # utterances: 6,109,469
Pre-train Speaker (User) Embedding

– Build conversation network
  • Node: Speaker
  • Edge: # conversations between two speakers

– Make vector representation of speaker by node2vec

<table>
<thead>
<tr>
<th>Speaker</th>
<th>Embedding</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>[0.173, 0.468, ..., 0.324]</td>
</tr>
<tr>
<td>B</td>
<td>[0.190, 0.873, ..., 0.794]</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>I</td>
<td>[0.420, 0.407, ..., 0.289]</td>
</tr>
</tbody>
</table>
Experiment 1 - Response Quality

• Data
  – Twitter conversation corpus
  – 80% training / 10% validation / 10% testing

• Comparison models
  – VHRED: Use stochastic variables for each utterance
  – VHCR: Use stochastic variables for each utterance and context
  – DialogWAE: Use Wasserstein GAN
  – SpeakAddr: Use user embedding
  – VHUCM: Initialize the user embedding randomly

• Metrics
  – Perplexity per word
  – Automatic metrics by comparing with ground truth utterance
    BLEU, Embedding Average, ROUGE-L
  – Human evaluation
Experiment 1 - Response Quality

Perplexity per word

- SpeakAddr
- VHRED
- VHCR
- DialogWAE
- VHUCM
- VHUCM-PUE
Experiment 1 - Response Quality

Automatic metrics

0.02 0.07 0.12 0.17
SpeakAddr VHRED VHCN DialogWAE VHUCM

0.5 0.55 0.6 0.65 0.7
BLEU ROUGE-L Emb-Avg
Experiment 1 - Response Quality

• Process
  – Use Amazon Mturk
  – Turkers to pick a more appropriate response from two candidate responses (One from VHUCM-PUE, the other from a comparison model)
  – Take five workers answers

• Result- Mean preference with a 90% confidence interval

<table>
<thead>
<tr>
<th></th>
<th>Wins</th>
<th>Losses</th>
<th>Ties</th>
</tr>
</thead>
<tbody>
<tr>
<td>vs SpeakAddr</td>
<td>40.7 ± 3.0</td>
<td>34.9 ± 3.0</td>
<td>24.4 ± 2.4</td>
</tr>
<tr>
<td>vs VHCR</td>
<td>45.5 ± 2.7</td>
<td>40.4 ± 2.8</td>
<td>14.1 ± 2.1</td>
</tr>
<tr>
<td>vs DialogWAE</td>
<td>52.5 ± 2.9</td>
<td>34.5 ± 2.8</td>
<td>13.0 ± 2.1</td>
</tr>
</tbody>
</table>
## Experiment 1 - Response Quality

### Examples of generated responses

<table>
<thead>
<tr>
<th>Context</th>
<th>SpeakAddr</th>
<th>VHCR</th>
<th>DialogWAE</th>
<th>VHUCM-PUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>nice ! atl a few months ago and already back again :) how could you not be .. im so excited for the both of them . → whose getting married ?? → snarky and suzi . got engaged back in dec</td>
<td>i know , i know !!!</td>
<td>hey</td>
<td>ah ha ha , i have to do a lot of time to get to see what happens .</td>
<td>what about u</td>
</tr>
<tr>
<td>call me when your at the station , i shall be with → am waiting at station 2 caramel macchiato and 2 pain aux raisin in hand !!! x → whoop whoop , , starting to feel sick</td>
<td>i know , i know !!</td>
<td>no . i don’t even know what to do , but i don’t think so . : (</td>
<td>i know what the hell is , but the is .</td>
<td>: ( oh .</td>
</tr>
<tr>
<td>how was your exam ? x → think it was okay actually :) x → thats good then :) x</td>
<td>you !</td>
<td>:d :d :d :d :d :d :d i miss you too &lt;3 &lt;3 &lt;3</td>
<td>hahah ! ! ! im so sorry to tell you when i get home or something lol</td>
<td>you are still the best person who’s there ? x</td>
</tr>
</tbody>
</table>
## Experiment 1 - Response Quality

### Personalized response

<table>
<thead>
<tr>
<th>Questioner</th>
<th>Answerer</th>
<th>Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>User B</td>
<td>User A</td>
<td>Where is your hometown?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>north carolina!</td>
</tr>
<tr>
<td>User C</td>
<td>User A</td>
<td>How old are you?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i'm not sure, but i am a bit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>older than you</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Do you love me?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i love you.</td>
</tr>
<tr>
<td>User D</td>
<td>User A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>19 !!!</td>
</tr>
<tr>
<td></td>
<td></td>
<td>yes i do!</td>
</tr>
<tr>
<td>User A</td>
<td>User B</td>
<td>Where is your hometown?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>minnesota. &lt;unk&gt;.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>18 yr old</td>
</tr>
<tr>
<td></td>
<td></td>
<td>because i love you.</td>
</tr>
<tr>
<td>User A</td>
<td>User C</td>
<td>How old are you?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>manchester :) xx</td>
</tr>
<tr>
<td></td>
<td></td>
<td>nothing much :)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i love you too :) xx</td>
</tr>
<tr>
<td>User A</td>
<td>User D</td>
<td>How old are you?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i live in &lt;unk&gt;.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i have no idea</td>
</tr>
<tr>
<td></td>
<td></td>
<td>no, i don’t.</td>
</tr>
</tbody>
</table>
Experiment 2 – New users

• New user
  – Defined as a user who is not in the training data
  – Illustrates a common problem of cold start
Experiment 2 – New users

• New user
  – Defined as a user who is not in the training data
  – Illustrates a common problem of cold start

• Idea
  – Two users who have a conversation are close in the embedding space
  – We can initialize the new user embedding from his/her conversation partner

\[ s_{\text{new}} = \sum_{i \in \text{friends of } s_{\text{new}}} s_i + \epsilon \]
Experiment 2 – New users

Experiment setup

– Build conversation network
– Pick edges of the network randomly
– Remove the conversations of the edges in the training data
– Categorize the edge types in the test data

Dashed lines are removed in the training data
Experiment 2 – New users

Edge types

- **Known Dyad (■)**
  Conversations of the dyad exist in the training corpus

- **Known Users (■)**
  No conversations between the users, but they have conversations with other users

- **Known Partners (■)**
  One user is new, but their partner’s conversations are in the corpus

- **New Users (■)**
  No data of either user
Experiment 2 – New users

Edge and user types

– Known Dyad (■)
  • Known user (●)
– Known Users (□)
  • Known user (●)
– Known Partners (■)
  • New user (▶)
  • Known partner (◀)
– New Users (■)
  • New user (◆)
Experiment 2 – New users

• Comparison models
  – VHCR: No user embedding
  – VHUCM: Initialize a new user embedding randomly
  – VHUCM-PUE: Initialize a new user (▲) embedding by

\[
\text{New user } s_{\text{new}} = \sum_{i \in \text{friends of } s_{\text{new}}} s_i + \epsilon
\]

• Metrics
  – Automatic metrics by comparing with ground truth utterance
    • Embedding Average
    • ROUGE-L
Experiment 2 – New users

Embedding Average

- VHCR
- VHUCM
- VHUCM-PUE

Bar chart showing the embedding average for Known Dyad, Known Users, Known Partner, and New Users.
Experiment 2 – New users

ROUGE-L

[Bar chart showing ROUGE-L scores for different conditions: Known Dyad, Known Users, Known Partner, and New Users with data points for VHCR, VHUCM, and VHUCM-PUE]
Contributions

• Suggested a novel conversation model (VHUCM)
• Built a new Twitter conversation corpus, to be shared for future research
• Suggested a method for dealing with new users
• Looked into new type of conversations -- many users, many conversations, many conversational partners
Conversational Modeling

- It can be important & effective in advancing computational social science
- There is much interest from social scientists
- Recent improvements in sequence modeling are effective
- Many obstacles remain
- Highly critical are identifying, collecting and annotating data
THANK YOU

SAMSUNG Research