Beyond Conditional LM: Neural Network Language Generation for Dialogue Systems

Heriot Watt University Seminar, 24/05/2016
Tsung-Hsien (Shawn) Wen and Steve Young
Outline

- Intro
- Multi-domain Neural NLG
- Neural Dialogue System
- Conclusion
Outline

- **Intro**
- Multi-domain Neural NLG
- Neural Dialogue System
- Conclusion
RNN Language Modeling

RNNLM (Mikolov et al, 2010)
Conditional RNN LM

Eat serves British.
Many things could be a condition LM

Sutskever et al, 2014

Karpathy et al, 2015
Outline

- Intro
- **Multi-domain Neural NLG**
- Neural Dialogue System
- Conclusion
Neural NLG

Inform(name=Seven_Days, food=Chinese) [0, 0, 1, 0, 0, ..., 1, 0, 0, ..., 1, 0, 0, 0, 0, 0...

dialog act 1-hot representation...

\[
\text{SLOT\_NAME} \quad \text{serves} \quad \text{SLOT\_FOOD} \\
\text{Seven Days} \quad \text{serves} \quad \text{Chinese}
\]

delexicalisation

(Wen et al, 2015)
Domain Adaptation

- Adaptation for NN?
  - Continue to train the model on adaptation dataset
- Parameters are shared on LM part of the network
  - But not for the DA weights
- New slot-value pairs can only be learned from scratch
Data counterfeiting

- Produce pseudo target domain data by replacing source domain slot-values pairs with target domains slot-value pairs.

- Procedure:

An example realisation in laptop (source) domain:

<table>
<thead>
<tr>
<th>Zeus 19</th>
<th>is a heavy laptop with a 500GB memory</th>
</tr>
</thead>
</table>

Delexicalisation

\[<NAME-value> \text{ is a } <WEIGHT-value> <TYPE-value> \text{ with a } <MEMEORY-value> <MEMORY-slot>\]

Counterfeiting

\[<NAME-value> \text{ is a } <FAMILY-value> <TYPE-value> \text{ with a } <SCREEN-value> <SCREEN-slot>\]

A possible realisation in TV (target) domain:

| Apollo 73 | is a U76 television with a 29-inch screen |
Data counterfeiting

- **Choice of target domain slots?**
  - The realisation should be similar to the source one.
  - Simple case: based on their functional class.
    - Informable, requestable, and binary slots.
  - Example:

<table>
<thead>
<tr>
<th></th>
<th>Laptop</th>
<th>Television</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informable</td>
<td>family, price_range, battery_rating,...</td>
<td>family, price_range, screen_size_range,...</td>
</tr>
<tr>
<td>Requestable</td>
<td>price, memory,...</td>
<td>price, resolution,...</td>
</tr>
<tr>
<td>Binary</td>
<td>is_for_business</td>
<td>has_usb_port</td>
</tr>
</tbody>
</table>
Laptop/TV dataset

- A more difficult dataset than restaurant/hotel
- Permute all possible DAs, ~13K/7K
- Only a few example utterances for each DA

<table>
<thead>
<tr>
<th>informable slots</th>
<th>Laptop</th>
<th>Television</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>family, *pricerange, batteryrating,</td>
<td>family, *pricerange, screensizerange, ecorating, hdmiport, hasusbport</td>
</tr>
<tr>
<td></td>
<td>driverrange, weightrange,</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>isforbusinesscomputing</strong></td>
<td></td>
</tr>
<tr>
<td>requestable slots</td>
<td>*name, *type, *price, warranty, battery,</td>
<td>*name, *type, *price, resolution, powerconsumption, accessories, color,</td>
</tr>
<tr>
<td></td>
<td>design, dimension, utility, weight,</td>
<td>screensize, audio</td>
</tr>
<tr>
<td></td>
<td>platform, memory, drive, processor</td>
<td></td>
</tr>
</tbody>
</table>

**bold**=binary slots, *=overlap with SF Restaurant and Hotel domains, all **informable slots** can take ”dontcare” value
Data counterfeiting - Results

(a) BLEU score curve

(b) Slot error rate curve

Laptop to TV
Discriminative Training

- Explore model capacity and correct it.

- DT cost function:

\[
F(\theta) = -\mathbb{E}[L(\theta)] = - \sum_{\Omega \in \text{Gen}(d_i)} p_{\theta}(\Omega|d_i) L(\Omega, \Omega_i)
\]

\(\Omega\): candidate sentence
\(\Omega_i\): reference sentence
\(d_i\): dialogue act
\(L(.):\) scoring function
Discriminative Training - Results

(a) Effect of DT on BLEU

(b) Effect of DT on slot error rate
Outline

- Intro
- Multi-domain Neural NLG
- Neural Dialogue System
- Conclusion
Traditional Dialogue Systems

Speech Recognition → Language Understanding → Language Generation → Dialogue Manager

Speech Synthesis

KB

Web

text
Neural Dialogue Systems
Can I have <v.food>?

**Intent Network**

Can I have <v.food> → z_t

**Generation Network**

<v.name> serves great <v.food>.

**Policy Network**

Can I have <v.food>.

**Belief Tracker**

Can I have <v.food>.

**MySQL Query**

```
SELECT * WHERE food=Korean
```

**Database**

- Nirala Curry Prince
- Seven days
- Seven days Curry Prince
- Seven days Curry Prince
- Seven days Curry Prince
- Seven days Curry Prince
- Seven days Curry Prince

**Database Operator**

Copy field
Intent Network

Can I have <v.food>
Can I have <v.food>

Belief Tracker

Korean 0.7
British 0.2
French 0.1

...
Jordan RNN-CNN belief trackers

Delexicalised CNN

Sentence representation
Can I have <v.food>

Korean 0.7
British 0.2
French 0.1

…

Belief Tracker

MySQL query: “Select * where food=Korean”

Database Operator

DB pointer

0 0 0 … 0 1
Can I have <v.food>

Korean 0.7
British 0.2
French 0.1

MySQL query: "Select * where food=Korean"

Database Operator

Intent Network

Policy Network

Belief Tracker
Can I have <v.food>

Korean 0.7
British 0.2
French 0.1

...
I can have <v.food>

<table>
<thead>
<tr>
<th>Food</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Korean</td>
<td>0.7</td>
</tr>
<tr>
<td>British</td>
<td>0.2</td>
</tr>
<tr>
<td>French</td>
<td>0.1</td>
</tr>
</tbody>
</table>

MySQL query: "Select * where food=Korean"

Belief Tracker

Intent Network

Generation Network

Policy Network

Database Operator

Copy field

Database:

- Seven days
- Curry Prince
- Nirala
- Seven days
- Seven days
- Seven days
- Seven days
- Little Saoul
- Seven days
Wizard of Oz Data Collection

- Online parallel version of WOZ on MTurk
  - Randomly hire a worker to be user/wizard.
  - Ask them to enter an appropriate response for one turn (following some instructions).
  - Repeat the process until all dialogues are finished.

- Example user page

Task 02004: You are looking for and it should serve gastropub food. You don’t care about the price range. You want to know the address.

Info Desk: Hello, welcome to the Cambridge restaurant system. You can ask for restaurants by area, price range or food type. How may I help you?

Customer: I want a gastropub food

Info Desk: There are 4 restaurants serving gastropub food, what price range do you want?

Customer: (Your response)

[Submit the HIT button]
Wizard of Oz Data Collection

 כגון wizard page

Info Desk: Hello, welcome to the Cambridge restaurant system. You can ask for restaurants by area, price range or food type. How may I help you?
Customer: I want a gastropub food
Info Desk: There are 4 restaurants serving gastropub food, what price range do you want?
Customer: I don't care

Please modify the following answers based on the latest customer response:
- What does user want?
  What is the food type the user wants? gastropub
  What is the area the user wants? not mentioned
  What is the price range the user wants? don't care
- What does user ask?
  Is the user asking for food type of an offered venue? No
  Is the user asking for price range of an offered venue? No
  Is the user asking for area of an offered venue? No
  Is the user asking for postcode of an offered venue? No
  Is the user asking for phone number of an offered venue? No
  Is the user asking for address of an offered venue? No
  Is the user mentioning any restaurant names? No

Info Desk: (Your response)

I would recommend backstreet bistro, a great gastropub restaurant in the centre. do you want their phone number? end-of-dialogue?

<table>
<thead>
<tr>
<th>Name</th>
<th>Food</th>
<th>Area</th>
<th>Price Range</th>
<th>Phone</th>
<th>Address</th>
<th>Postcode</th>
</tr>
</thead>
<tbody>
<tr>
<td>backstreet bistro</td>
<td>gastropub</td>
<td>centre</td>
<td>expensive</td>
<td>01223 306305</td>
<td>2 Stutton Street City Centre</td>
<td>C.B 1, 2 Q.A</td>
</tr>
<tr>
<td>royal standard</td>
<td>gastropub</td>
<td>east</td>
<td>expensive</td>
<td>01223 247877</td>
<td>290 Mill Road City Centre</td>
<td>C.B 1, 3 N.L</td>
</tr>
<tr>
<td>the cow pizza kitchen and bar</td>
<td>gastropub</td>
<td>centre</td>
<td>moderate</td>
<td>01223 308871</td>
<td>Corn Exchange Street</td>
<td>C.B 2, 3 Q.F</td>
</tr>
<tr>
<td>the slug and lettuce</td>
<td>gastropub</td>
<td>centre</td>
<td>expensive</td>
<td>--</td>
<td>34 - 35 Green Street</td>
<td>C.B 2, 3 J.U</td>
</tr>
<tr>
<td>nil</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
<td>nil</td>
</tr>
</tbody>
</table>
Data Statistics

- **Ontology:**
  - Cambridge restaurant domain, 99 venues.
  - 3 informable slots, 6 requestable slots

- **Dataset**
  - 676 dialogues, ~2750 turns
  - 3000 HITS, takes 3 days, costs ~400 USD
  - Data cleaning takes 2-3 days for one person
Experiments

- Experimental details
  - Train/valid/test: 3/1/1
  - SGD, l2 regularisation, early stopping, gradient clip=1
  - Hidden size = 50, Vocab size: ~500

- Two stage training:
  - Training trackers with label cross entropy
  - Training other parts with response cross entropy

- Decoding
  - Beam search w/ beam width 10
  - Decode with average word likelihood
# Human evaluation

## Quality assessment

<table>
<thead>
<tr>
<th>Metric</th>
<th>N2N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Success</td>
<td>98%</td>
</tr>
<tr>
<td>Comprehension</td>
<td>4.11</td>
</tr>
<tr>
<td>Naturalness</td>
<td>4.05</td>
</tr>
</tbody>
</table>

# of dialogues: 245

## System Comparison

<table>
<thead>
<tr>
<th>Metric</th>
<th>N2N</th>
<th>Modular</th>
<th>Tie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj. Success</td>
<td>96.95%</td>
<td>95.12%</td>
<td>-</td>
</tr>
<tr>
<td>Avg. # of Turn</td>
<td>3.95</td>
<td>4.54</td>
<td>-</td>
</tr>
</tbody>
</table>

## Comparisons (%)

<table>
<thead>
<tr>
<th>Metric</th>
<th>N2N</th>
<th>Modular</th>
<th>Tie</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturalness</td>
<td>46.95*</td>
<td>25.61</td>
<td>27.44</td>
</tr>
<tr>
<td>Comprehension</td>
<td>45.12*</td>
<td>21.95</td>
<td>32.93</td>
</tr>
<tr>
<td>Preference</td>
<td>50.00*</td>
<td>24.39</td>
<td>25.61</td>
</tr>
<tr>
<td>Performance</td>
<td>43.90*</td>
<td>25.61</td>
<td>30.49</td>
</tr>
</tbody>
</table>

* p < 0.005, # of comparisons: 164
### Example dialogues

<table>
<thead>
<tr>
<th>Human</th>
<th>Restaurant in any area that serves Vietnamese food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td>thanh binh serves vietnamese food type, is there anything else i can help you with?</td>
</tr>
<tr>
<td>Human</td>
<td>What is the phone number and post code</td>
</tr>
<tr>
<td>Machine</td>
<td>the number is 01223 362456 and the area code is c.b 3, 0 a.f. is there anything else i can do for you?</td>
</tr>
<tr>
<td>Human</td>
<td>Good bye</td>
</tr>
<tr>
<td>Machine</td>
<td>thank you for using the cambridge restaurant system.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Human</th>
<th>Hi, i’m hungry for some Indonesian food</th>
</tr>
</thead>
<tbody>
<tr>
<td>Machine</td>
<td>i am sorry, there are no indonesian restaurants. would you like a different food?</td>
</tr>
<tr>
<td>Human</td>
<td>Oh no. How about Indian?</td>
</tr>
<tr>
<td>Machine</td>
<td>there are 9 restaurants serving indian types of food. do you have a preference on side of town?</td>
</tr>
<tr>
<td>Human</td>
<td>East side please</td>
</tr>
<tr>
<td>Machine</td>
<td>sitar tandoori is a indian restaurant in the east side of town. would you like their information?</td>
</tr>
<tr>
<td>Human</td>
<td>Yes, the address and phone number please</td>
</tr>
<tr>
<td>Machine</td>
<td>sitar tandoori is located at 43 high street cherry hinton cherry hinton and their phone number is 01223 249955.</td>
</tr>
<tr>
<td>Human</td>
<td>Okay, thank you</td>
</tr>
<tr>
<td>Machine</td>
<td>thank you, goodbye.</td>
</tr>
</tbody>
</table>
Example dialogues

Human : hello
Machine : what cuisine would you like ?
Human : afghan
Machine : i am sorry , there are no afghan restaurants . would you like something else ?
Human : korean
Machine : little seoul serves korean food , it is located at 108 regent street city centre .
Human : is it expensive ?
Machine : little seoul serves korean food and is in the expensive price range .
Human : what area is that
Machine : little seoul is located in the centre area of town . would you like their number ?
Human : yes
Machine : the telephone number is 01223 308681 . is there anything else i can help you with today ?
Human : no goodbye
Machine : thank you for using the cambridge restaurant system , goodbye .
Visualising action embeddings
Outline

- Intro
- Multi-domain Neural NLG
- Neural Dialogue System
- Conclusion
Conclusion

- NN-based conditional LM is widely applied.
  - Generation based on dialogue act representation.
  - Generation based on the entire dialogue context.

- Achieve domain extension by data counterfeiting and discriminative training.

- Develop an end-to-end task oriented dialogue system by collecting WOZ data and JUST training it!
Papers


Selected References


Thank you! Questions?

Tsung-Hsien Wen is supported by a studentship funded by Toshiba Research Europe Ltd, Cambridge Research Laboratory
SC-LSTM

- Original LSTM cell
  \[ i_t = \sigma(W_{wi} w_t + W_{hi} h_{t-1}) \]
  \[ f_t = \sigma(W_{wf} w_t + W_{hf} h_{t-1}) \]
  \[ o_t = \sigma(W_{wo} w_t + W_{ho} h_{t-1}) \]
  \[ \hat{c}_t = \tanh(W_{wc} w_t + W_{hc} h_{t-1}) \]
  \[ c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t \]
  \[ h_t = o_t \odot \tanh(c_t) \]

- DA cell
  \[ r_t = \sigma(W_{wr} w_t + W_{hr} h_{t-1}) \]
  \[ d_t = r_t \odot d_{t-1} \]

- Modify \( C_t \)
  \[ c_t = f_t \odot c_{t-1} + i_t \odot \hat{c}_t + \tanh(W_{dc} d_t) \]

\( h_t \) and \( c_t \) are the hidden state and cell state, respectively, which store information over time. The sigmoid function \( \sigma \) is used to introduce non-linearity. The tanh function is used to introduce non-linearity in the manner of a hard limit. The \( \odot \) symbol denotes the element-wise product. The \( w_t \) is the input vector, \( h_{t-1} \) is the previous hidden state, and \( c_{t-1} \) is the previous cell state. The \( d_t \) is the input to the DA cell. The \( d_0 \) is the initial input to the DA cell. The \( r_t \) is the input to the DA cell. The \( h_t \) and \( c_t \) are the outputs of the LSTM cell and DA cell, respectively. The \( \text{Inform}(\text{name}=Seven\_Days, \text{food}=Chinese) \) is the output of the dialog act 1-hot representation.
Visualization

- A-inform
- pricerange=don't care
- kids_allowed=yes
- count=VALUE
- food=VALUE
- type=VALUE

- A-inform
- accepts_cards=yes
- has_internet=yes
- near=VALUE
- name=VALUE

feature value
there, are, 6, korean, restaurant, -, that, allow, kid, -, at, various, price, range, -, .

feature value
the, castro_inn, near, haint, has, internet, and, accepts, credit, card, -, .
Results

- **Human Evaluation**
  - **Method** | **Informativeness** | **Naturalness**
  - sc-lstm | 2.59 | 2.50
  - classlm | 2.46** | 2.45

*p < 0.05  **p < 0.005
Human Evaluation

<table>
<thead>
<tr>
<th>Method</th>
<th>TV to Laptop</th>
<th></th>
<th>laptop to TV</th>
</tr>
</thead>
<tbody>
<tr>
<td>scrALL</td>
<td>2.64</td>
<td>2.37</td>
<td>2.54</td>
</tr>
<tr>
<td>DT-10%</td>
<td><strong>2.52</strong></td>
<td><strong>2.25</strong></td>
<td><strong>2.51</strong></td>
</tr>
<tr>
<td>ML-10%</td>
<td><strong>2.51</strong></td>
<td><strong>2.22</strong></td>
<td><strong>2.45</strong></td>
</tr>
<tr>
<td>scr-10%</td>
<td><strong>2.24</strong></td>
<td><strong>2.03</strong></td>
<td><strong>2.00</strong></td>
</tr>
</tbody>
</table>

* p < 0.05, ** p < 0.005

- scrALL : train from scratch with 100% ID data.
- scr-10% : train from scratch with 10% ID data.
- ML-10% : data counterfeiting + ML training on 10% ID data.
- DT-10% : data counterfeiting + DT training on 10% ID data.