Using Historical Context to improve Dialog State Tracking

By Sanuj Sharma
With Prafulla Choubey and Prof Ruihong Huang
Dialog Turn

- User Utterance: *I’m looking for a Chinese restaurant in east part of town.*

- System Utterance: *What’s your budget?*

- System Actions: *price range*

- Dialog State
  
  - inform(food = Chinese)
  
  - inform(area = east)
Dialog State

• **Requests**: information requested by user.
  - request = address, request = phone

• **Joint Goals**: the set of accumulated turn goals.
  - food = french, price range = cheap

- **REQUEST**: price range
- **INFORM**: food = chinese, area = east
WoZ 2.0 Dataset

1400 dialogs, 10-15 turns / dialog

- **Area**: north, south, east, west, central.
- **Price Range**: cheap, expensive, moderate, don’t care.
- **Food**: French, Chinese, Mexican, Indian.
- **Request**: address, phone, pinched, area, price.
Binary Classification

- For **each slot-value pair**:
  - User Utterance
  - Previous turn System Acts
  - Slot-value

- Probability of the slot-value in dialog state
Binary Classification

- For each slot-value pair:
  - I would like to eat Chinese food.
  - Food = Indian
  - Model
  - > 0.5 then yes
  - Else no
My Contribution

- For each slot-value pair:
  - User + previous system utterance
  - Previous turn System Acts
  - Slot type-value
  - Historical Context: Previous utterances and slot value where current slot type was last modified.
- Probability of the slot-value in dialog state
hello, i'm looking for a restaurant with fair prices
Price range: moderate
There are 31 places with moderate price range. Can you please tell me what kind of food you would like?
Sys act: food

well I want to eat in the North, what's up that way?
Area: north
I have two options that fit that description, Golden Wok chinese restaurant and The Taj which serves Indian food. Do you have a preference?
Sys act: food

Can I have the address and phone number for the Golden Wok chinese restaurant?
Request: address
Request: phone number
Food: Chinese
The phone number is 01223 350688.

thank you. what is the address?
Request: address
The address is 191 Histon Road Chesterton.

Okay, what about Taj, what's the address and phone of that?
Request: address
Request: phone number
Food: Indian
- Utterances
- System Acts
- Slot-value
- Historical context

Model

• Probability of the slot-value in dialog state

Encoding module

Scoring Module

Bi-LSTM

Self Attention

Attention
## Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Joint Goal Accuracy</th>
<th>Turn Request Accuracy</th>
<th>Approximate # of parameters (in million)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline</td>
<td>84.5%</td>
<td>95%</td>
<td>1.2</td>
</tr>
<tr>
<td>Baseline + Historical Context</td>
<td>88.4%</td>
<td>96%</td>
<td>6</td>
</tr>
<tr>
<td>GLAD</td>
<td>86.4%</td>
<td>97%</td>
<td>17</td>
</tr>
<tr>
<td>GLAD + Historical Context</td>
<td>89%</td>
<td>97%</td>
<td>28</td>
</tr>
</tbody>
</table>

**Baseline:** Bi-lstm + self-attention encoder with attention scorer.

**GLAD:** Global-locally self-attentive dialog state tracker.
Thank you
Personality-Based Chatbot

Sameer Kumar Behera
Srishti Agarwal
Shubham Bhargava
• **Rule-Based** : Answers using set of hand-crafted rules.

• **Retrieval** : Answer selected based on set of answers to the question.

• **Generative** : Generate proper responses, Seq2Seq, Encoder-Decoder, can generate new, complex responses.

• **Speaker-Addressee** : Predict how Speaker A would respond to message by Speaker B.
• Chatbots with Persona

• Persona - elements of identity (background facts or user profile), language behavior, and interaction style.

• Wide applications
  – Domain Specific Assistance like IT Helpdesk, Customer Care Representatives..
  – Entertainment
• TV Series Corpus - Friends(Joey), and Big Bang Theory (Sheldon) scripts

• Processed Scripts to have Q & A like format.

• Replaced the change in Scene by a separator as to differentiate the contexts.
Sequence-to-sequence encoder-decoder model

- The model is fed input sentence $X$ (words $x_1$, $x_2$ and $x_3$) and outputs sentence $Y$ (words $y_1$, $y_2$, $y_3$, $y_4$ and $y_5$).
- $V$ represents thought vector of $x$. The hidden state $h_t$ captures the sequential information in $[x, y_1, y_2, \ldots, y_{t-1}]$. 
Example in Machine Translation
# Epochs: 200, Max Length: 8, Learning Rate: 0.001
Say Hi! To Joey
JOEY: HEY!

YOU: HI!

YOU: WHO ARE YOU?

JOEY: JOEY TRIBBIANI! FROM THE WALL!

YOU: ARE YOU AN IDIOT?

JOEY: YEAH. YEAH.

SAY SOMETHING...
YOU: WHAT'S UP?

YOU: WHAT HAPPENED TO YOU?

YOU: HAVE YOU GONE MAD?

JOEY: BAD NEWS.

JOEY: I KNOW, BUT I DON'T KNOW.

JOEY: YEAH! YEAH!

SAY SOMETHING...
Conclusion

- Chatbot is able to mimic the persona to an extent.
- Able to answer fluently.
- Better performance for short, similar sentences in corpus.
- Loss of context is seen w.r.t previous queries.
- Random answers for questions which are very different than actual script.
Questions???
Event Temporal Status Consistency in Coreference Chains

JUSTIN HILL – TEXAS A&M UNIVERSITY
**Events**

- **Temporal Status**
- **Coreference Resolution**

**Event**: a specific occurrence of an event.

**Temporal Status**: the status of an event at the time of a document’s writing.

<table>
<thead>
<tr>
<th>Status</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEFORE</td>
<td>the event has already occurred</td>
</tr>
<tr>
<td>OVERLAP</td>
<td>the event is ongoing</td>
</tr>
<tr>
<td>AFTER</td>
<td>the event has yet to occur</td>
</tr>
<tr>
<td>BEFORE/OVERLAP</td>
<td>the event has already begun and is still ongoing</td>
</tr>
</tbody>
</table>

*The Richer Event Description dataset distinguishes between Overlap and Before/Overlap. For identifying event status, we consider them to be equivalent.*

**Coreference Chain**: a set of event mentions that refer to the same event.
Events

- Event Mentions
- Temporal Status
- Coreference Resolution

S1: MILITANT SAYS HE IS BEHIND FATAL NIGER ATTACK

S2: Mokhtar Belmokhtar ... has claimed responsibility for another terrorist attack ...

S3: The new claim was made on a number of different websites.
Related Work

- Event Status Identification
- Event Coreference Resolution

• Prior works on event coreference resolution used rule-based and statistical methods that rely on features local to the event mentions under consideration.

• More recent work addresses this issue by modeling global relationships among events using documents’ topic structures.

• **WHAT’S MISSING?** Temporal information is underutilized in event coreference resolution systems. Temporal status is a feature that can be extracted in an event mention’s local context but can link each mention to the global document context.

• **Proposal:** Analyze event status patterns in coreference chains to determine the former’s usefulness for the latter

Questions

1. Do the events in coreference chains have the same event status i.e. are chains’ statuses consistent?

2. Which event features correlate with a chain’s consistency?

3. What are the relationships between consistency and each event status category?
Methods

1. Aggregate event information based on annotated event coreference chains.

2. Compute a status consistency score for each chain.

3. Extract event information correlations w.r.t. chains’ status consistencies.

4. Record inconsistent chains for further analysis.
Step 1: Aggregate Event Information

For each event in every coreference chain, collect lexical features of the event mention information and the annotated features of the event.

Step 2: Compute Status Consistency Scores

- Compute the **majority status** for each coreference chain.
  
  \[ \text{status}_{\text{chain}} = \arg\max_{s \in \{\text{BEFORE,OVERLAP,AFTER}\}} (\text{chain}) \]

- Compute the **consistency score** for each chain using the majority status. This is equivalent to the percentage of events in the chain that have the majority status.
  
  \[ \text{score} = \frac{\# \text{status}_{\text{chain}} \text{ in chain}}{\# \text{events in chain}} \]
Step 3: Extract Event Information Patterns

- **Split chains** into 10% partitions based on consistency score.
- Compute the percentage of (feature, value) present in each partition.

\[
\text{Input} : \text{a set of chains with event features, } C' \\
\text{consistency scores for each chain, } R \\
\text{Output} : \text{chain paritions with feature value distibutions, } P
\]

1. \( P \leftarrow \text{partition}(C', R) \)
2. \( \text{for } p \text{ in } P \text{ do} \)
3. \( \text{for each (feature } x, \text{ value } v) \text{ do} \)
4. \( p.x_v \leftarrow p . \text{count}(x_v) / C'.\text{count}(x_v) \)
5. \( \text{return } P \)

---

Step 4: Identify Inconsistent Chains

\[
\text{Input} : C', R \\
\text{Output} : C''
\]

1. \( C'' \leftarrow \emptyset \)
2. \( \text{for } p \text{ in } P \text{ do} \)
3. \( \text{for } c'. \text{id, score in } R \text{ do} \)
4. \( \text{if score } \neq 1 \)
5. \( c' \leftarrow C'.\text{get}(c'. \text{id}) \)
6. \( c'.N \leftarrow \text{count}(c', s) \text{ for } s \text{ in } \{\text{BEFORE, OVERLAP, AFTER}\} \)
7. \( C'' \leftarrow C'' \cup \{c'\} \)
8. \( \text{return } C'' \)
### Data

- **Richer Event Description (RED) dataset**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Documents</td>
<td>95</td>
</tr>
<tr>
<td>Tokens</td>
<td>54,287</td>
</tr>
<tr>
<td>Events</td>
<td>8,718</td>
</tr>
<tr>
<td>Status Annotations (1 for each Event)</td>
<td>8,718</td>
</tr>
<tr>
<td>Event Coreference Chains</td>
<td>759</td>
</tr>
</tbody>
</table>

Features

• **Part-of-Speech Tag**: the part-of-speech tag for the mention.
• **Status**: the temporal status of the mention.
• **Type**: whether the event specifies aspectual information (aspectual, evidential) about other events.
• **Representation**: whether the event is explicit or implicit.
• **Degree**: provides more nuanced information about polarity.
• **Polarity**: whether or not the event occurred.
• **Modality**: whether the event is asserting things about the real world, about hypothetical events, about generic tendencies of events, or about uncertain events.
• **Aspect**: whether the event is intermittent.
Q1: Are Event Statuses Consistent in Coreference Chains?

- Approximately 24% of events are part of some event coreference chain (2,118/8,718 events).

- Approximately 94.6% of coreference chains have consistent temporal statuses (718/759 chains).

<table>
<thead>
<tr>
<th>Consistency Score Range</th>
<th># Chains</th>
<th>% Chains</th>
</tr>
</thead>
<tbody>
<tr>
<td>50% ≤ score &lt; 60%</td>
<td>23</td>
<td>3.0</td>
</tr>
<tr>
<td>60% ≤ score &lt; 70%</td>
<td>6</td>
<td>0.8</td>
</tr>
<tr>
<td>70% ≤ score &lt; 80%</td>
<td>6</td>
<td>0.8</td>
</tr>
<tr>
<td>80% ≤ score &lt; 90%</td>
<td>6</td>
<td>0.8</td>
</tr>
<tr>
<td>score = 100%</td>
<td>718</td>
<td>94.6</td>
</tr>
</tbody>
</table>
Q2: Which Features Correlate w/ Consistency?

<table>
<thead>
<tr>
<th>Feature:Value</th>
<th>% of Occurrences in Consistent Chains</th>
<th># Occur.</th>
<th>Feature:Value</th>
<th>% of Occurrences in Consistent Chains</th>
<th># Occur.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aspect:INTERMITTENT</td>
<td>67</td>
<td>3</td>
<td>Degree:N/A</td>
<td>93</td>
<td>2113</td>
</tr>
<tr>
<td>Aspect:N/A</td>
<td>93</td>
<td>2115</td>
<td>Polarity:NEG</td>
<td>96</td>
<td>78</td>
</tr>
<tr>
<td>Modality:ACTUAL</td>
<td>91</td>
<td>1569</td>
<td>Polarity:POS</td>
<td>93</td>
<td>2040</td>
</tr>
<tr>
<td>Modality: GENERIC</td>
<td>97</td>
<td>285</td>
<td>Rep:EXPLICIT</td>
<td>93</td>
<td>2110</td>
</tr>
<tr>
<td>Modality:HYPOTHETICAL</td>
<td>98</td>
<td>190</td>
<td>Rep:IMPLICIT</td>
<td>100</td>
<td>8</td>
</tr>
<tr>
<td>Modality: UNCERTAIN</td>
<td>93</td>
<td>74</td>
<td>Type: ASPECTUAL</td>
<td>100</td>
<td>15</td>
</tr>
<tr>
<td>Degree:LITTLE</td>
<td>100</td>
<td>2</td>
<td>Type:EVIDENTIAL</td>
<td>89</td>
<td>64</td>
</tr>
<tr>
<td>Degree: MOST</td>
<td>100</td>
<td>3</td>
<td>Type:N/A</td>
<td>93</td>
<td>2039</td>
</tr>
</tbody>
</table>

occurrences < 89%  89% ≤ occurrences < 100%  occurrences = 100%
Q3: What are the relationships between consistency and each event status category?

- There is a large class imbalance favoring the BEFORE status (51% of events in chains).
- Events with AFTER and BEFORE statuses occur in consistent chains at least 95% of the time.

<table>
<thead>
<tr>
<th>Status</th>
<th>% of Occurrences in Consistent Chains</th>
<th># Occur.</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEFORE</td>
<td>95.0</td>
<td>1,076</td>
</tr>
<tr>
<td>OVERLAP</td>
<td>87.9</td>
<td>742</td>
</tr>
<tr>
<td>AFTER</td>
<td>97.3</td>
<td>300</td>
</tr>
</tbody>
</table>
• Every inconsistent chain contains two statuses.

• Only one inconsistent chain does not contain the OVERLAP status.

• Events with OVERLAP status are the main source of ambiguity.

<table>
<thead>
<tr>
<th>Status</th>
<th>% Events in Chains</th>
<th>% of Events in Inconsistent Chains</th>
</tr>
</thead>
<tbody>
<tr>
<td>BEFORE</td>
<td>50.8</td>
<td>36.4</td>
</tr>
<tr>
<td>OVERLAP</td>
<td>35.0</td>
<td>58.4</td>
</tr>
<tr>
<td>AFTER</td>
<td>14.2</td>
<td>5.2</td>
</tr>
</tbody>
</table>
Examples

Inconsistent Chains

• The evaluation was revised in February 2002 to ensure that all 4 bidders would receive an automatic 60 points for technology transfer \( \text{(AFTER, HYPOTHETICAL)} \)

• Rafale reportedly offered more generous terms for technology transfer and subcontracts for South Korea’s aerospace industry \( \text{(OVERLAP, GENERIC)} \)

• "This necessarily means the coming national assembly and government that will emerge from it will not possess the legitimacy to enable them to draft the coming constitution," it said. \( \text{(AFTER, BEFORE)} \)

• However, I went to see them to review his IEP and they said he has huge concentration issues \( \text{(BEFORE)} \)

• His new IEP (done the week after going to see the doc) has 8 points on it \( \text{(OVERLAP, OVERLAP)} \)

• His IEP states he needs supervision but he does not have any supervision. \( \text{(OVERLAP)} \)
Conclusions

Analyzed the consistency of event temporal status in coreference chains.

Approximately \textbf{94.6\% of chains have consistent temporal statuses}. This implies temporal status may be a good signal for making coreference decisions.

Events with \textbf{OVERLAP} status are the main source of ambiguity, thus OVERLAP-BEFORE and OVERLAP-AFTER mention pairs are the harder coreference decisions to resolve w.r.t. status.

\textbf{Future Work:}

Utilize event temporal status in an event coreference resolution system.
Aggie Text Summarizer

Aryan Sharma, Sandeep Gottimukkala, Aditya Gujral

Department of Computer Science, Texas A&M University
Motivation

- Users have been deluged in data
- Attention span has been decreasing
- Need to condense the important information and provide the relevant and accurate details
Definitions and Applications

• **Text Summarization**
  – Reducing a text in order to create a summary that retains the most important points of the original text

• **Applications**
  – Summaries of email threads
  – Summarizing News
  – Action items from a meeting
  – Simplifying text
Related Work

- Extractive approach
  - Summarizes by selecting a few relevant sentences from the original documents

- Abstractive approach
  - Produces an abstract summary which has words and phrases different from the ones occurring in the document

- Graph based approaches use graph models which represents correlations among multiple terms
Background on Text Summarizers

- Extractive approaches
  - Graph Based approaches
Objective

• This project improves existing summarizers based on Extractive approaches
Background on Text Summarizers

• **Extractive approaches** selects among the sentences
Graph Based Approaches

- Treats each sentences as nodes with edges among each other
• The edge weights are the similarity scores among the sentences
TextRank

• It then runs the stochastic matrix created and derives the TextRank (steady state PageRank)
TextRank

• Sentence Importance $\propto$ TextRank
TextRank calculates the similarity as the normalized common word counts as the edge weights.

\[ \text{sim}(x, y) \]

is the similarity score between Sentence x and Sentence y.
What is TextRank missing?

- It would give a low score to sentences like
  - ‘This is trivial.’ and ‘This is common.’
Our solution to it

• It would give a low score to sentences like
  – ‘This is trivial.’ and ‘This is common.’

• We solved this by providing context
  among words through Google’s
  pre-trained neural embeddings
  – ‘common’ and ‘trivial’ have same context
    and thus same embedding score
What is TextRank missing?

- Some words which were repeated often in the input will very likely be mentioned in a human summary
  - hence TextRank missed frequency weights
Our solution to it

• Some words which were repeated often in the input will very likely be mentioned in a human summary
  – hence TextRank missed frequency weights

• We added normalized term frequency scores after removing the stop words
What is TextRank missing?

- Sentences which contain named entities are usually more important as these sentences indicate information of the entities participating in the documents.
  - TextRank didn’t give any special weight to **Texas** and **Aggies** in ‘**The competition was played in Texas where Aggies created the world record in swimming**’
Our Solution

• Sentences which contain named entities are usually more important as these sentences indicate information of the entities participating in the documents.

• We added the NER implementation to the score and thus such sentences scored more
Results using Rogue

![Bar chart showing errors using Rogue with different methods: TextRank, NER + TextRank, SumBasic + TextRank, and Embeddings + TextRank. The chart displays the error rates ranging from 0.300 to 0.345.]
THANK YOU

Questions?
Writing Tweets for You

Samantha Ray, Jacob Fenger, Sukhdeep Gill, and Zong-Fu Hsieh
**Problem Statement:** Generate human-sounding tweets using a collection of tweets as the training data. Compare/contrast different approaches to see the difference with human-ness and coherency of the results

**Approaches:**
- Markov Models
  - Markov Chains
  - Hidden Markov Models (HMM)
- Recurrent Neural Networks
  - Long Short-Term Memory (LSTM)
  - Variational Autoencoder (VAE)
Dataset Used

- sentiment140 dataset, contains 1,600,000 tweets extracted using the twitter api.

Example: "hey long time no see! Yes.. Rains a bit, only a bit LOL, I'm fine thanks, how's you?",

- Average Tweet Length: 74

- Standard Deviation: 36
**Markov Chain**

**PROS:**
- Fast sequence generation $O(n)$,
  
  $n = \text{sequence length}$

**CONS:**
- Poor space complexity $O(|S|^k)$,
  
  $k = \text{dependency length}$

  Effectiveness goes down as vocabulary increases

**EXAMPLE:**
- break holy infuriating final unnecessarily
Hidden Markov Model

**PROS:**
- Better memory than Markov chain
- Better space growth

**CONS:**
- Less effective for long sequences

**EXAMPLE:**
- i working

<table>
<thead>
<tr>
<th></th>
<th>AA</th>
<th>AB</th>
<th>AC</th>
<th>AD</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>8</td>
</tr>
<tr>
<td>B</td>
<td>1</td>
<td>6</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>C</td>
<td>0</td>
<td>6</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>9</td>
<td>10</td>
<td>2</td>
</tr>
</tbody>
</table>
Reverse Viterbi

- Reverse the logic of the tagging problem to generate sequence of words
  - High time and memory complexity: $O(n|V|^k)$ and $O(|V|^k)$, respectively
- Optimization Methods:
  - Generalized - Train generic trellis for a given sequence length
  - Filtered - Reduce $|V|$ by removing rare words
- Examples: “will be getting a mocha frappuccino now”, “my favorite curling iron broke”, “sorry to hear that lol”
### Long Short Term Memory (LSTM) Model

<table>
<thead>
<tr>
<th>Layer (type)</th>
<th>Output Shape</th>
<th>Param #</th>
</tr>
</thead>
<tbody>
<tr>
<td>lstm_4 (LSTM)</td>
<td>(None, 328)</td>
<td>628448</td>
</tr>
<tr>
<td>dropout_6 (Dropout)</td>
<td>(None, 328)</td>
<td>0</td>
</tr>
<tr>
<td>dense_6 (Dense)</td>
<td>(None, 243)</td>
<td>79947</td>
</tr>
<tr>
<td>dropout_7 (Dropout)</td>
<td>(None, 243)</td>
<td>0</td>
</tr>
<tr>
<td>dense_7 (Dense)</td>
<td>(None, 150)</td>
<td>36600</td>
</tr>
<tr>
<td>activation_4 (Activation)</td>
<td>(None, 150)</td>
<td>0</td>
</tr>
</tbody>
</table>

**Total params:** 744,995  
**Trainable params:** 744,995  
**Non-trainable params:** 0
LSTM with window size of 40 words and 50 epochs with different diversity:

----- diversity: 0.2
----- Generating with seed: ", pics thanks nancy! hope all is well. "

pics thanks nancy! hope all is well. i want to get the problem i have to be a start on the conce in the one in the start of the site and i don't want to get a terrow i want to get the same time when i would be be a reading to sleep in the world to stop and the week. i wish i was going to be a ready to get and the didnt to sleep in the best me to work i want to get a started and the world i want to get the problem i have to work

----- diversity: 1.2
----- Generating with seed: ", pics thanks nancy! hope all is well. "

pics thanks nancy! hope all is well. it was jibe or not pfacion ff uckallyim even yes samitrakie delitics in the profelerscroola few mene? bvdpbe wormx &amp; we mad ooh, i keet really week kira pleompipler.. long getting seep wjisser fut sad, at the loora like me tood who must woke opp? lold spast why a uuse broke i've wanting cuck rusnow! my 2wry likely new hy emailed batted me, strived here sheasta arnigally hand n gglang!
Generating Specific User Tweets

- Generating human sounding Tweets is a fairly simple task, but can we generate Tweets that sound like a specific user?
- Much harder task due to the need to capture the knowledge of what makes a specific user’s Tweets unique.
The Data

- Gathered ~3000 Donald Trump Tweets via Tweepy
- Examples:
  - “We must keep “evil” out of our country!”
  - “Yes, Arnold Schwarzenegger did a really bad job as Governor of California and even worse on the Apprentice...but at least he tried hard!”
  - “The Fake Media is working overtime today!”

Donald J. Trump
@realDonaldTrump

It's freezing and snowing in New York--we need global warming!
11:24 AM - 7 Nov 2012

Donald J. Trump
@realDonaldTrump

Despite the constant negative press covfefe...
10:09 AM - 31 May 2017
The Model

- Pre-train a Recurrent Neural Network with the general Tweet dataset mentioned earlier
- Fine-tune the model on the set of ~3000 Donald Trump tweets that were pre-processed similarly to the general tweets used for pretraining
- Architecture used for training:
The Results

- **Seed:** “...Stock market up almost 20% since elec”
  
  ○ **Result:** “...Stock market up almost 20% since election with the presing the preside, to the the president the great with the we the military…”

- A larger pre-training dataset and more training may be more necessary to generate better sounding Tweets

- Only having ~3000 Tweets to fine-tune on was a limiting factor

- A more complicated model does not mean better results
VAE for text generation

- VAE
  - Autoencoder
  - Generating the latent vectors by following a Gaussian distribution

- VAE for text generation
  - two Single-layer LSTM RNNs to implement Encoder and decoder
ASTON: Automatic Summarization for News

By Xichao Chen, Ruosi Lin, Shijin Tang
CSCE 638 Group Project
Fall 2018
Motivation

- Too many news articles published every day for a human being to consume.
- Summarization on news may help.
  - Shorter text covering the main idea.
- Manual summarization by human is laborious.
- Go for automatic summarization.
  - Fewer biases
  - Faster and more scalable
  - More cost-efficient

Pic: https://www.britannica.com/topic/newspaper
Related Work

Automatic Text Summarization

- Extractive
  - Topic representation
  - Indicator representation
- Abstractive
  - Structured-based
  - Semantic-based

- Earliest effort dates back to 1950!
- Deep neural networks: applicable for both extractive and abstractive methods.
Approaches: TF-IDF based Tag method

- Term frequency (TF)
  - The count of a word in an article

- Inverse document frequency (IDF)
  - $1 / DF$
  - DF: The count of a word in all articles

- TF * IDF value implies the importance of a word in an article

- Use TF-IDF value to pick sentences to form the summary
  - Pick words with highest TF-IDF values as tags
  - Pick sentences containing the most tag words

- Naive but useful to improve other methods
Approaches: Modified LexRank

- **LexRank**: a stochastic graph-based method
  - Summary: sentences that are the most similar to other sentences
- **Improvements**
  - Lemmatization
    - E.g., “mice”, “mouse”
  - Consider word similarity
    - E.g, “mountain”, “hill”
  - Apply TF-IDF threshold
    - Only consider the similarities between important words
    - Speed up and improve the performance
  - Consider the article structure
    - The most important sentences usually appear in the start or the end of a paragraph

Pic: https://www.cs.cmu.edu/afs/cs/project/jair/pub/volume22/erkan04a-html/erkan04a.html
1. Transferred a document with $m$ unique terms and $n$ sentences into term-frequency matrix, we will have an $m \times n$ sparse matrix.
2. got $V$ by SVD, whose columns are called right singular vectors, and each row in $VT$ corresponds to a sentence vector.
3. After that, we could sort these sentence by their magnitude of its corresponding singular value and get the most important sentences in the document.
Approaches: NetSum

- Neural network based extractive method

- Break an input article into sentences
  - (Consider only the first 30 sentences)

- Obtain a numerical representation of each sentence
  - 8 features:
    - is_first_sentence,
    - sent_position,
    - sum_basic_score (unigram/bigram),
    - sent_sim_to_summ, idf (sum and average), and
    - wordnet_sent_sim_score

- Feed into NN
  - 4 dense layers, modified RankNet

- Use the saved persistence for prediction
## Results

<table>
<thead>
<tr>
<th>Method</th>
<th>ROUGE1</th>
<th>ROUGE2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>precision</td>
<td>recall</td>
</tr>
<tr>
<td>Improved LexRank</td>
<td>0.28422</td>
<td>0.21558</td>
</tr>
<tr>
<td>Latent Semantic Analysis</td>
<td>0.14630</td>
<td>0.37390</td>
</tr>
<tr>
<td>NetSum (Shifted IDF + LCS similarity)</td>
<td>0.39200</td>
<td>0.20920</td>
</tr>
<tr>
<td>Gold Standard</td>
<td></td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN's Kate Bolduan sits down with the President to talk about working families. President Obama and first lady Michelle Obama are hosting a summit on the issue. The President wants to see paid parental leave and more flexibility for working parents.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LSA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Every single day, there are conversations around the kitchen table where people are trying to figure out, this child care is costing so much, I'm not sure that we're going to be able to make our mortgage at the end of the month,&quot; the President said to Kate Bolduan of CNN's &quot;New Day.&quot; &quot;I'm going to be taking some action, a presidential memorandum directing every federal agency to be very clear to their employees that it is my view that offering flexibility where possible is the right thing to do.&quot; &quot;And to the extent that we want to have this conversation outside of politics, I'd welcome a bipartisan effort with ideas coming from the private sector and from Republicans, from Democrats and from nonprofits and the faith community about how we make sure that we're supporting families and reducing their stress.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NetSum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amid growing foreign policy crises, President Barack Obama is encouraging Congress and the country to focus on issues here at home -- namely how to improve the livelihoods of working families. &quot;Every single day, there are conversations around the kitchen table where people are trying to figure out, this child care is costing so much, I'm not sure that we're going to be able to make our mortgage at the end of the month,&quot; the President said to Kate Bolduan of CNN's &quot;New Day.&quot; And staying up until 2 in the morning and feeding her and burping her creates a bond that is irreplaceable.&quot;</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LexRank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equal pay for equal work. The President said he's pushing for workplace flexibility to give parents the opportunity to become more involved in their children's lives and education. This is a middle-class issue and an American issue,&quot; he continued.</td>
</tr>
</tbody>
</table>
Conclusion

Extraction task focused, three approaches:

<table>
<thead>
<tr>
<th></th>
<th>LSA</th>
<th>Improved LexRank</th>
<th>NetSum</th>
</tr>
</thead>
<tbody>
<tr>
<td>PRO</td>
<td>Interpretable; Easy to implement; Fast; No train data needed</td>
<td>Easy to implement; Fast; Easy to modify the original model; No train data needed</td>
<td>Best performance among the three models; Easy to train</td>
</tr>
<tr>
<td>CON</td>
<td>Performance not so good as the machine learning model</td>
<td>Performance not so good as the machine learning model</td>
<td>Limited features; Hard to visualize and interpret (neural-network based)</td>
</tr>
</tbody>
</table>
Any Questions?