Speech Question Answering

TOEFL Listening Comprehension Test by Machine

Wei Fang
December 13, 2017

Speech Processing & Machine Learning Lab
What is Earth's radiation budget?

The balance of incoming solar energy and reflecting solar energy
Question Answering (QA)

- Understand spoken content
- Answer questions about spoken content

What is earth's radiation budget?

The balance of incoming solar energy and reflecting solar energy
New Task: TOEFL Listening Comprehension Test by Machine

- TOEFL: Test of English as a Foreign Language
New Task: TOEFL Listening Comprehension Test by Machine

- TOEFL: Test of English as a Foreign Language
- Listening Section:
  - Listen to a 3~5 minute story
  - Answer question with a set of answer choices
New Task: TOEFL Listening Comprehension Test by Machine

- **TOEFL**: Test of English as a Foreign Language
- **Listening Section:**
  - Listen to a 3~5 minute story
  - Answer question with a set of answer choices

<table>
<thead>
<tr>
<th>Story</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basically a cloud either contributes to the cooling of earth’s surface or to its heating. Earth’s climate system is constantly trying to strike a balance between the cooling and warming effects of clouds ...... we call this earth’s radiation budget ...... (audio story)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>According to the professor, what is earth’s radiation budget?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Choices</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Average temperature difference between land mass and body of water</td>
</tr>
<tr>
<td>B. Balance of incoming solar energy and reflecting solar energy</td>
</tr>
<tr>
<td>C. Percentage of incoming solar energy that gets trapped in clouds</td>
</tr>
<tr>
<td>D. Portion of marine species that have been affected by global warming</td>
</tr>
</tbody>
</table>
New Task: TOEFL Listening Comprehension Test by Machine

Dataset
New Task: TOEFL Listening Comprehension Test by Machine

Dataset

- Past exams collected from a TOEFL practice website
New Task: TOEFL Listening Comprehension Test by Machine

Dataset

- Past exams collected from a TOEFL practice website
- Splits - train/dev/test: 717/124/122
New Task: TOEFL Listening Comprehension Test by Machine

Dataset

- Past exams collected from a TOEFL practice website
- Splits - train/dev/test: 717/124/122
- Audio stories with two transcriptions: manual, ASR (CMU Sphinx with 34.32% WER)
New Task: TOEFL Listening Comprehension Test by Machine

Dataset

- Past exams collected from a TOEFL practice website
- Splits - train/dev/test: 717/124/122
- Audio stories with two transcriptions: manual, ASR (CMU Sphinx with 34.32% WER)

Approach

Question: According to the professor, what is the earth’s radiation budget?

Audio Story: transcriptions → Model → 4 Choices → Answer: (B)
The entire model learned end-to-end.
Baseline NN Model: LSTM

Attending to Relevant Sentences in Story
Attending to Relevant Sentences in Story

Note: Bi-directional RNNs
Attending to Relevant Sentences in Story

Sentence Representations
Sentence Representations

Bi-directional RNN

Tree-structured Recursive LSTM (Tree-LSTM)

Tai, Socher, Manning. Improved Semantic Representations From Tree-Structured Long Short-Term Memory Networks. ACL 2015.
Hierarchical Attention

Experimental Results

Question-Choice Similarity
Sliding Window
LSTM
LSTM + Attention
Tree-LSTM
Tree-LSTM + Attention

Accuracy

Random

0: 25
0: 30
0: 35
0: 40
0: 45
0: 50
There are 3 types of questions.

Type 3: Connecting Information

- Understanding Organization
- Connecting Content
- Making Inferences
There are 3 types of questions.

Type 2: Pragmatic Understanding
- Function of What is Said
- Speaker’s Attitude
There are 3 types of questions.

Type 2: Pragmatic Understanding
  - Function of What is Said
  - Speaker’s Attitude

**Example:**
What is the **purpose** of the man’s response? What can be **inferred** about the student?
Transfer Learning from Movie QA

Motivation

TOEFL is a small dataset; transfer from larger QA dataset (MovieQA) to improve performance.

Tapaswi, Zhu, Stiefelhagen, Torralba, Urtasun, Fidler. MovieQA: Understanding Stories in Movies through Question-Answering
Motivation

TOEFL is a small dataset; transfer from larger QA dataset (MovieQA) to improve performance.

Tapaswi, Zhu, Stiefelhagen, Torralba, Urtasun, Fidler. *MovieQA: Understanding Stories in Movies through Question-Answering*

Conclusion

- Introduced a new task TOEFL Listening Comprehension Test by Machine.
- Proposed attention-based models to outperform previous methods.
- Performance can be improved by transfer learning from a larger QA dataset.
Thanks

Contact

Wei Fang  b40815@gmail.com