Speech Question Answering

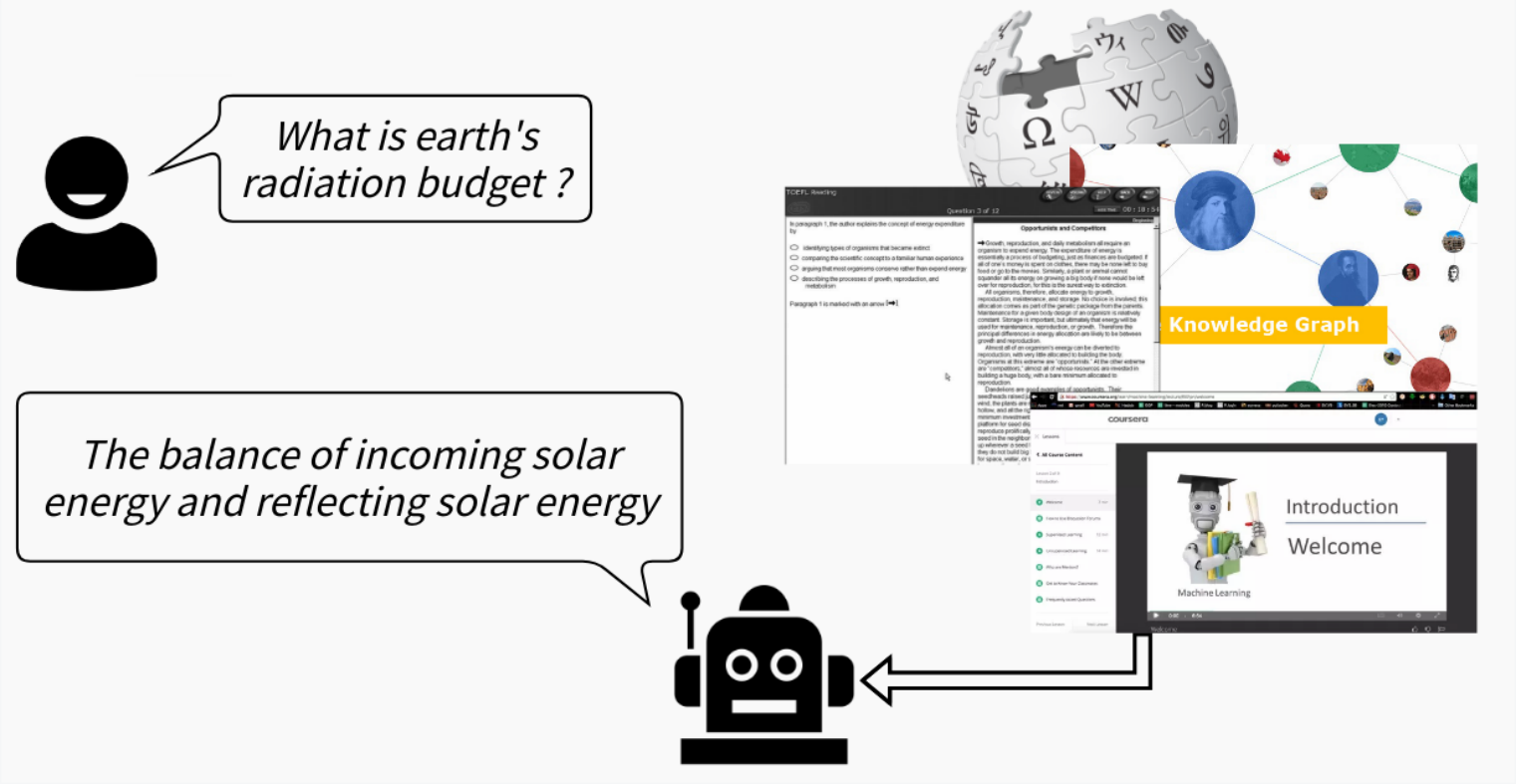
TOEFL Listening Comprehension Test by Machine

Wei Fang

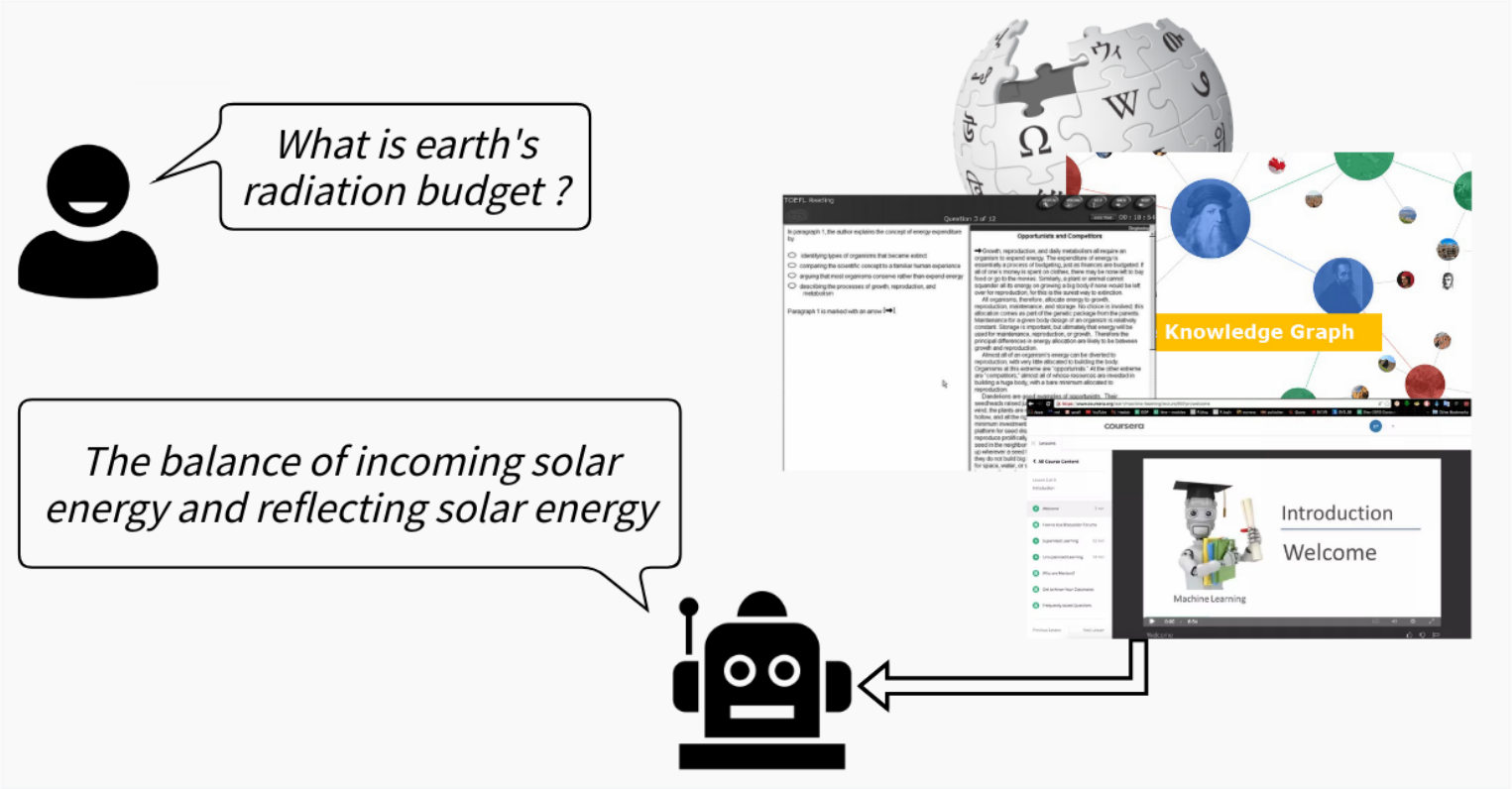
December 13, 2017

Speech Processing & Machine Learning Lab

Question Answering (QA)



Question Answering (QA)



- Understand spoken content
- Answer questions about spoken content

New Task: TOEFL Listening Comprehension Test by Machine

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$\sim$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

Story

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)

Question

According to the professor, what is earth's radiation budget?

Choices

A. Average temperature difference between land mass and body of water
B. **Balance of incoming solar energy and reflecting solar energy**
C. Percentage of incoming solar energy that gets trapped in clouds
D. Portion of marine species that have been affected by global warming

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

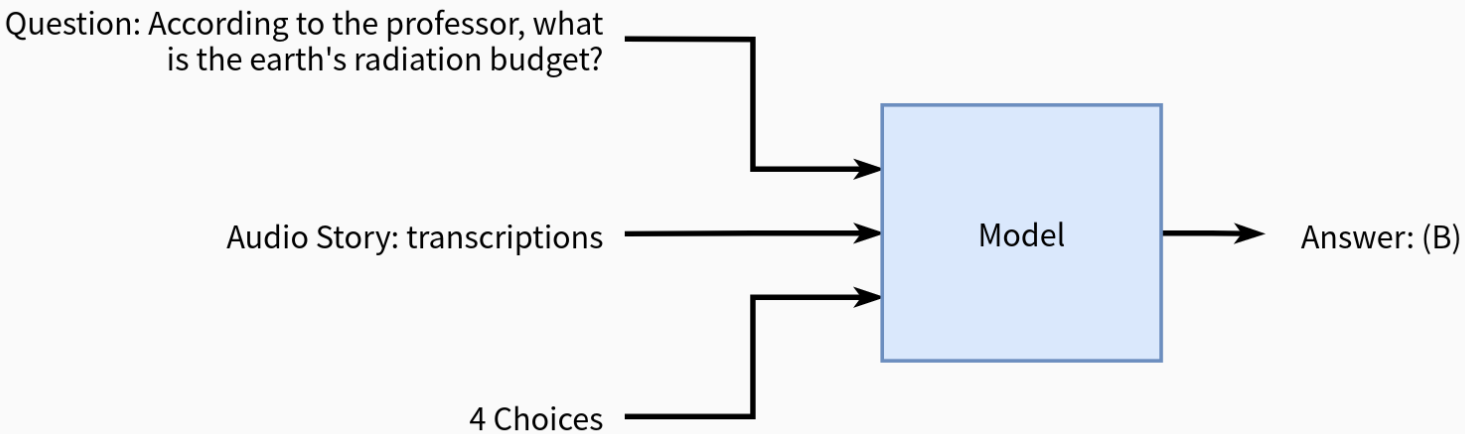
- Past exams collected from a TOEFL practice website
- Splits - train/dev/test: 717/124/122
- Audio stories with two transcriptions:
  manual, ASR (CMU Sphinx with 34.32% WER)

New Task: TOEFL Listening Comprehension Test by Machine

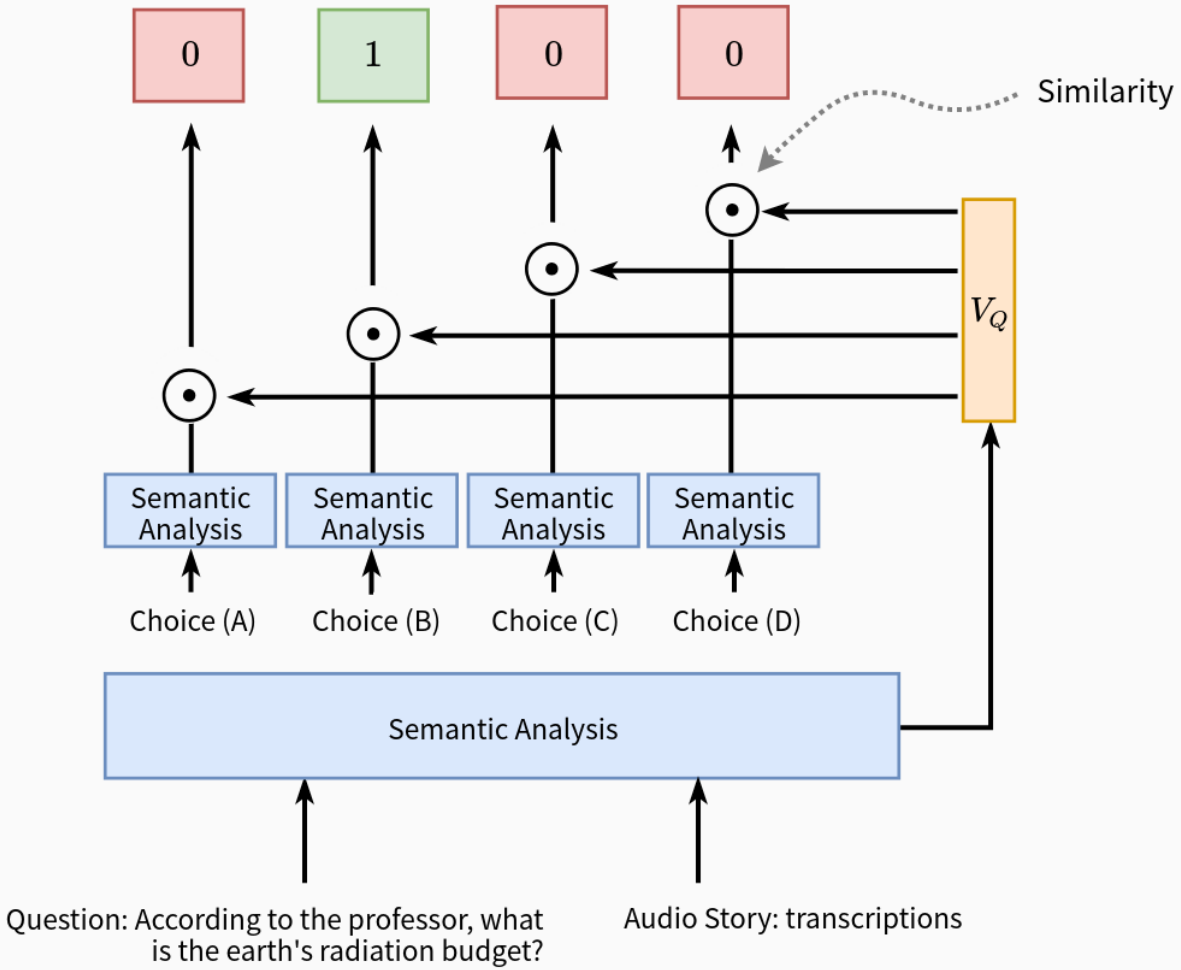
Dataset

- 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

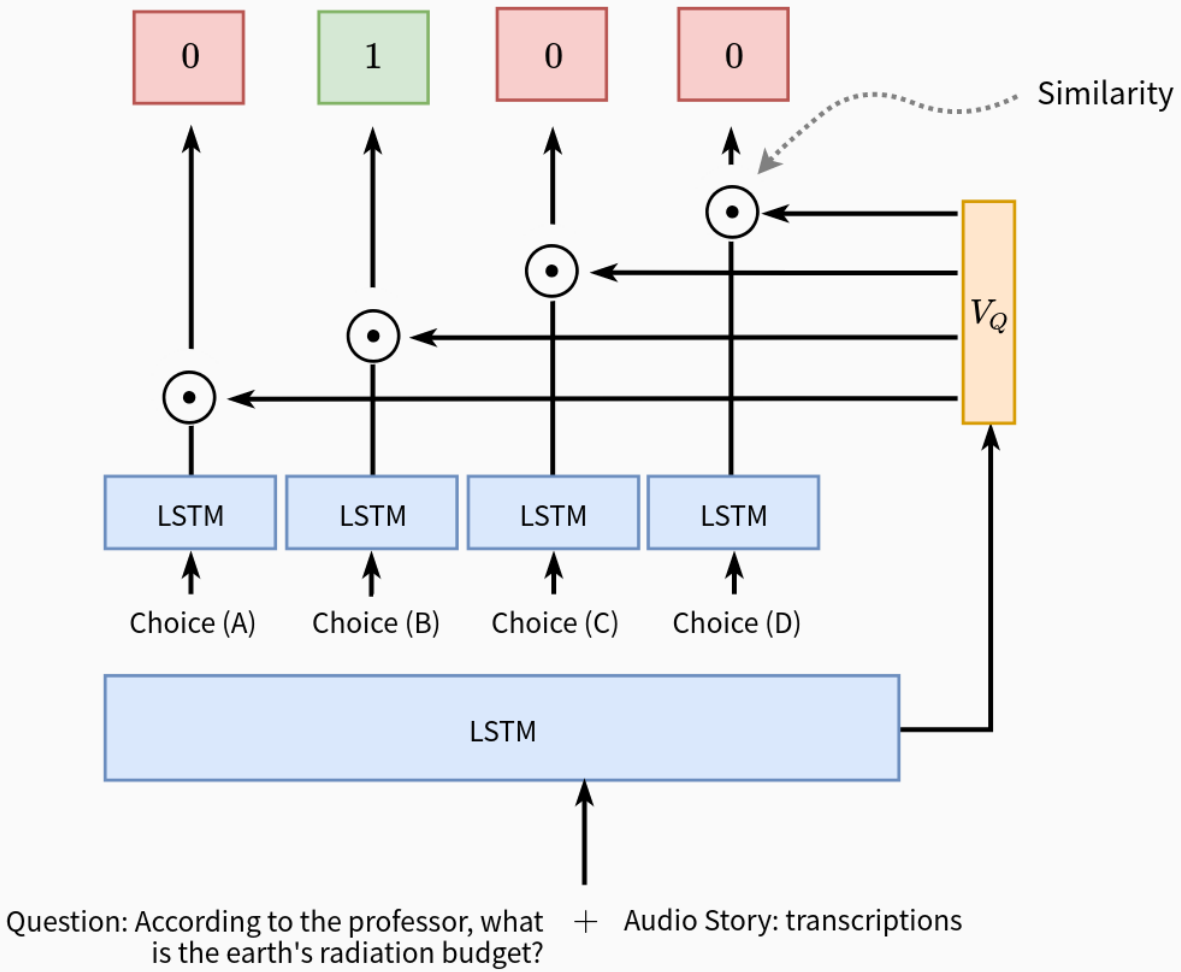


Neural Network Model Architecture



The entire model learned end-to-end.

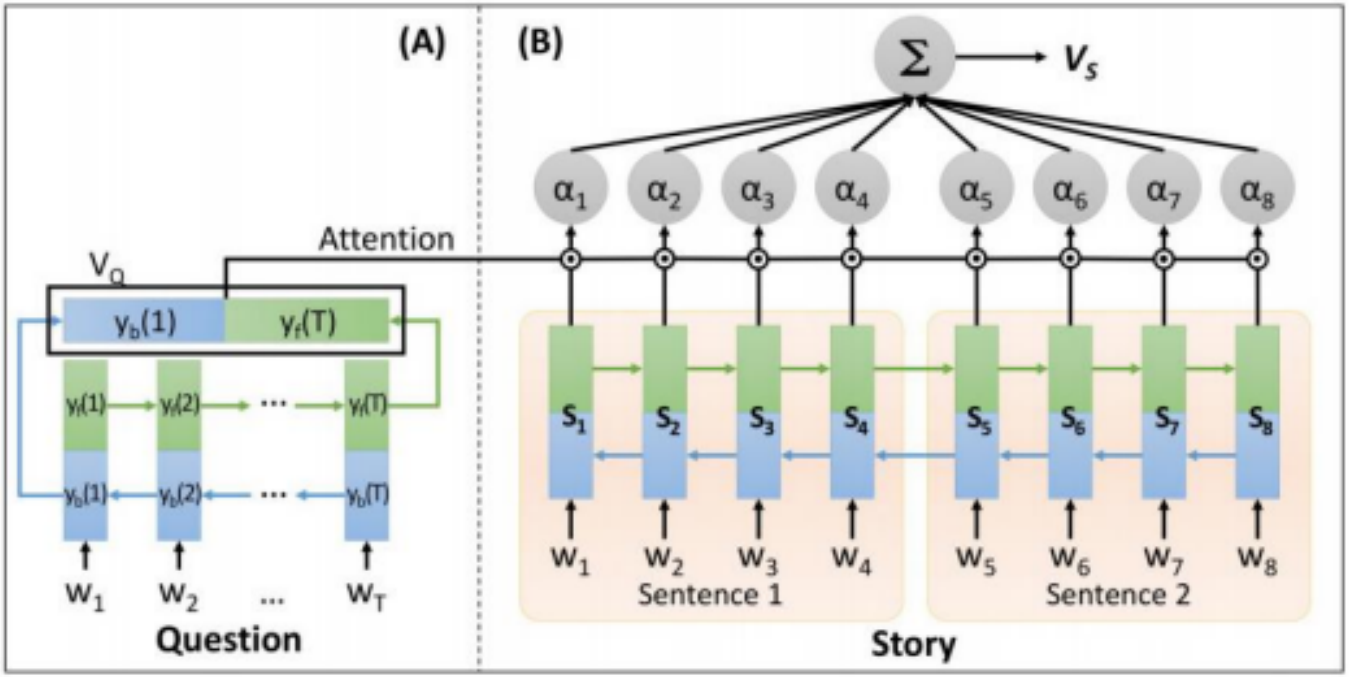
Baseline NN Model: LSTM



Hermann, Kočiský, Grefenstette, Espeholt, Kay, Suleyman, Blunsom. *Teaching Machines to Read and Comprehend.* NIPS 2015.

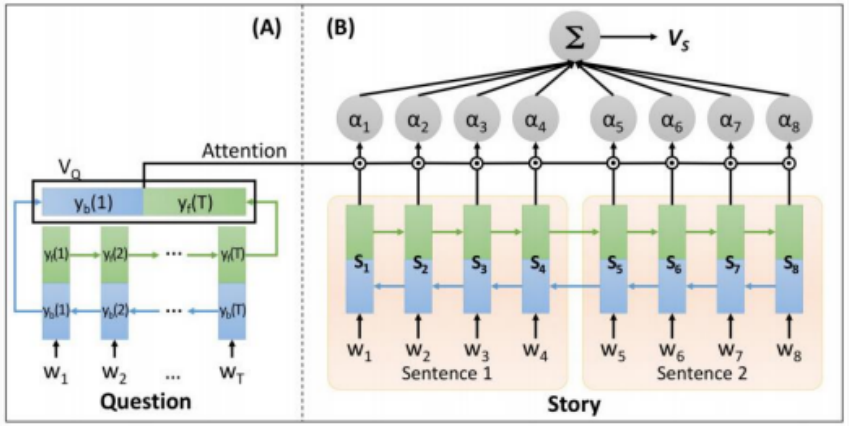
Attending to Relevant Sentences in Story

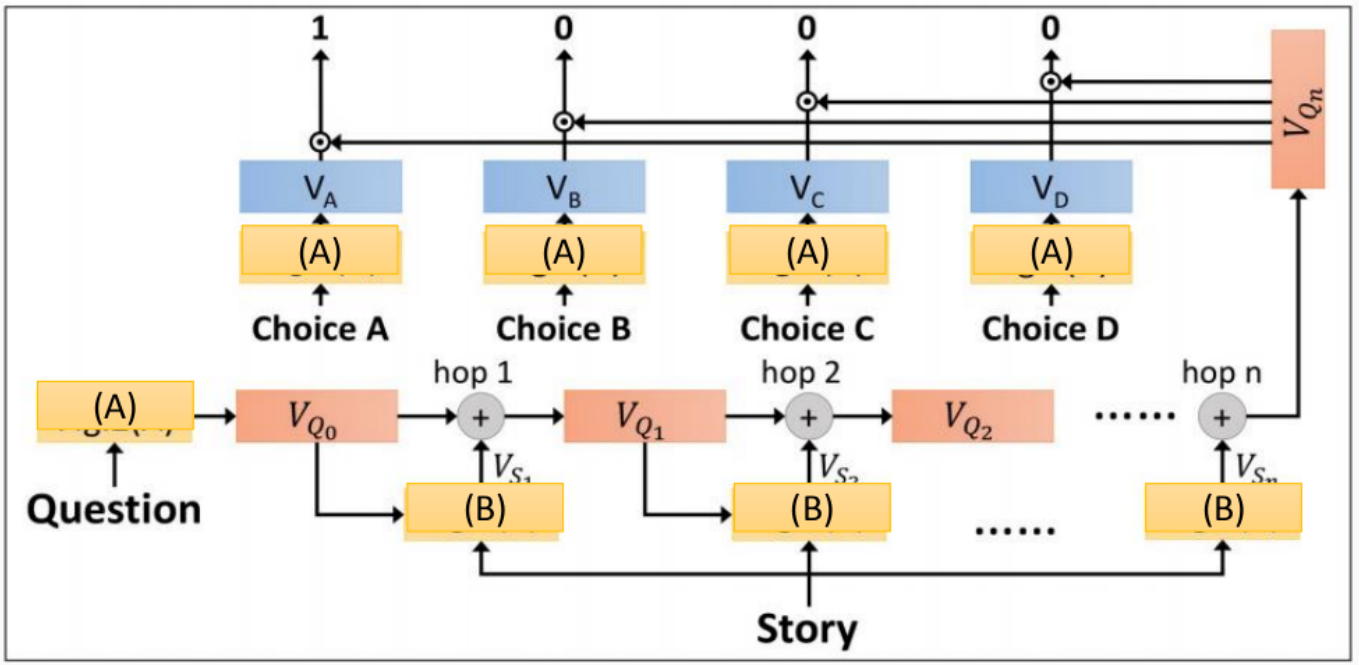
Attending to Relevant Sentences in Story



Note: Bi-directional RNNs

Attending to Relevant Sentences in Story

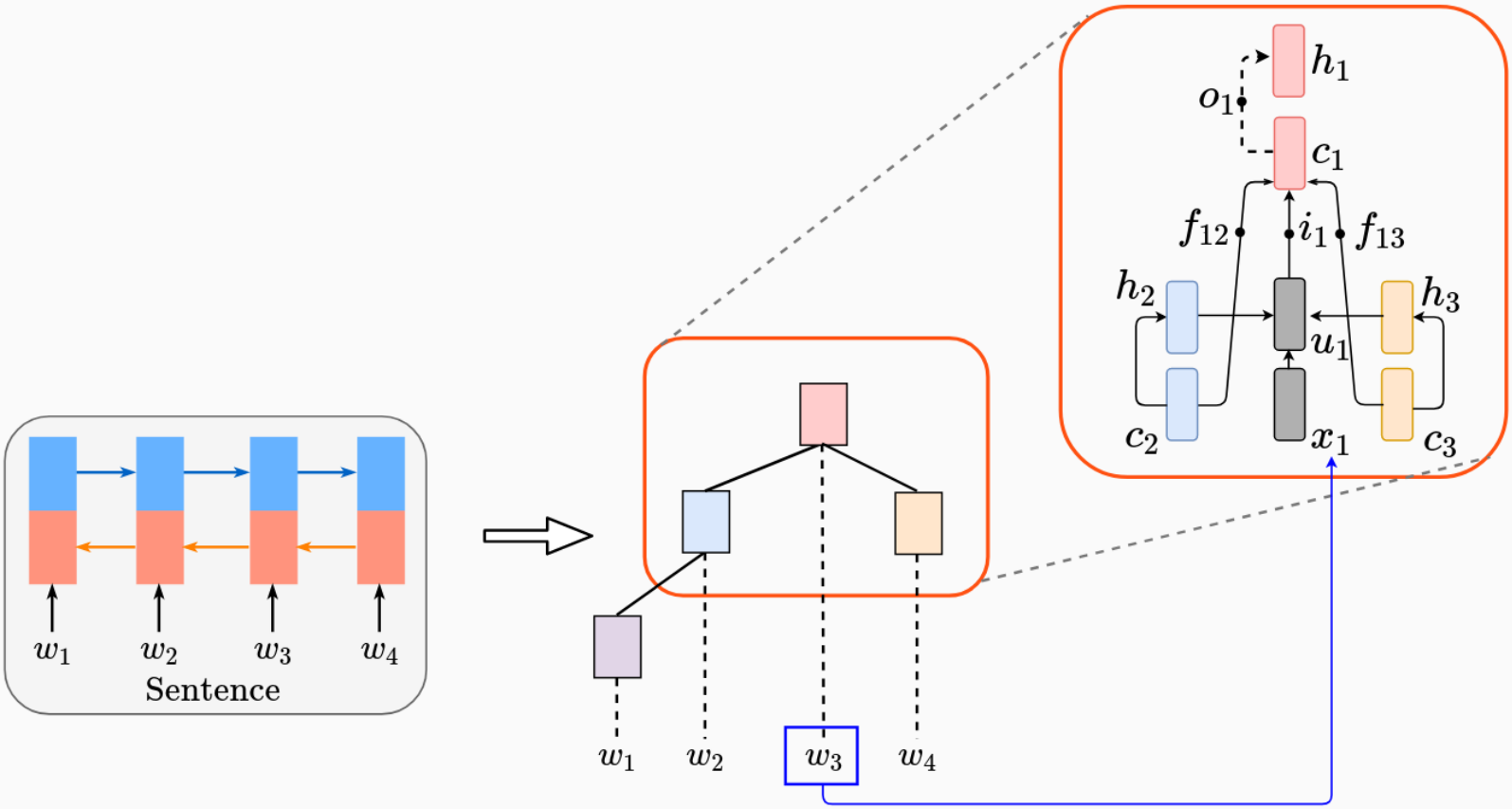




Tseng, Shen, Lee, Lee. *Towards Machine Comprehension of Spoken Content: Initial TOEFL Listening Comprehension Test by Machine*. Interspeech 2016.

Sentence Representations

Sentence Representations

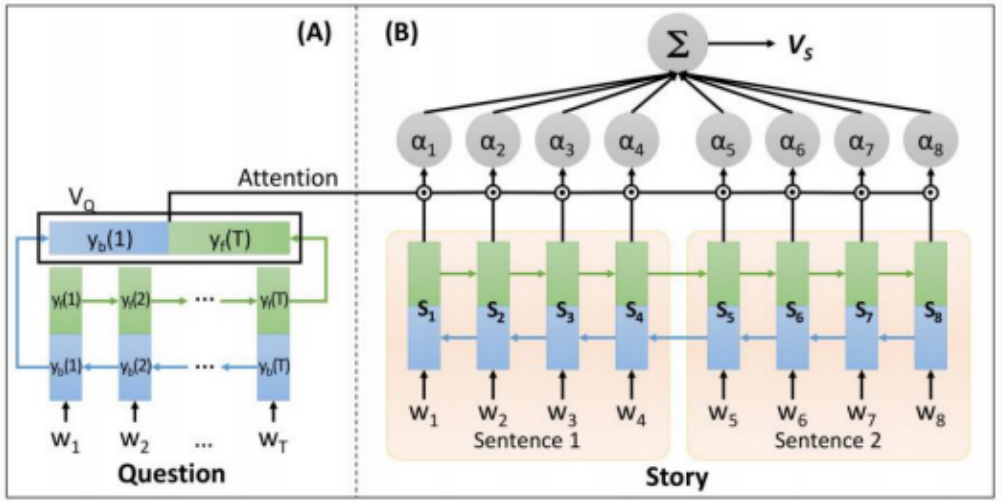


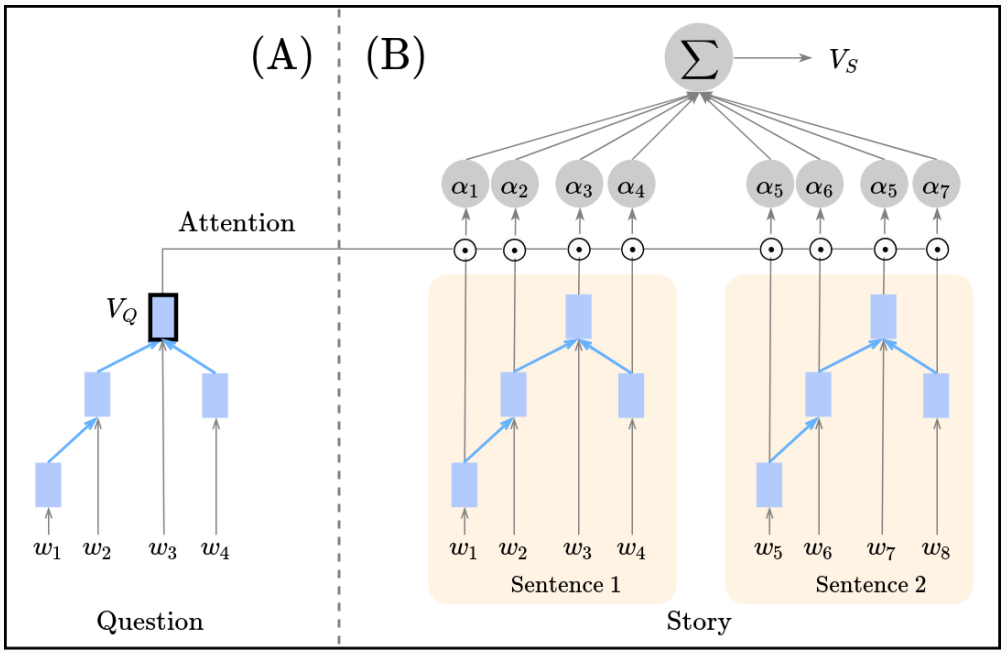
Bi-directional RNN

Tree-structured Recursive LSTM (Tree-LSTM)

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

Hierarchical Attention





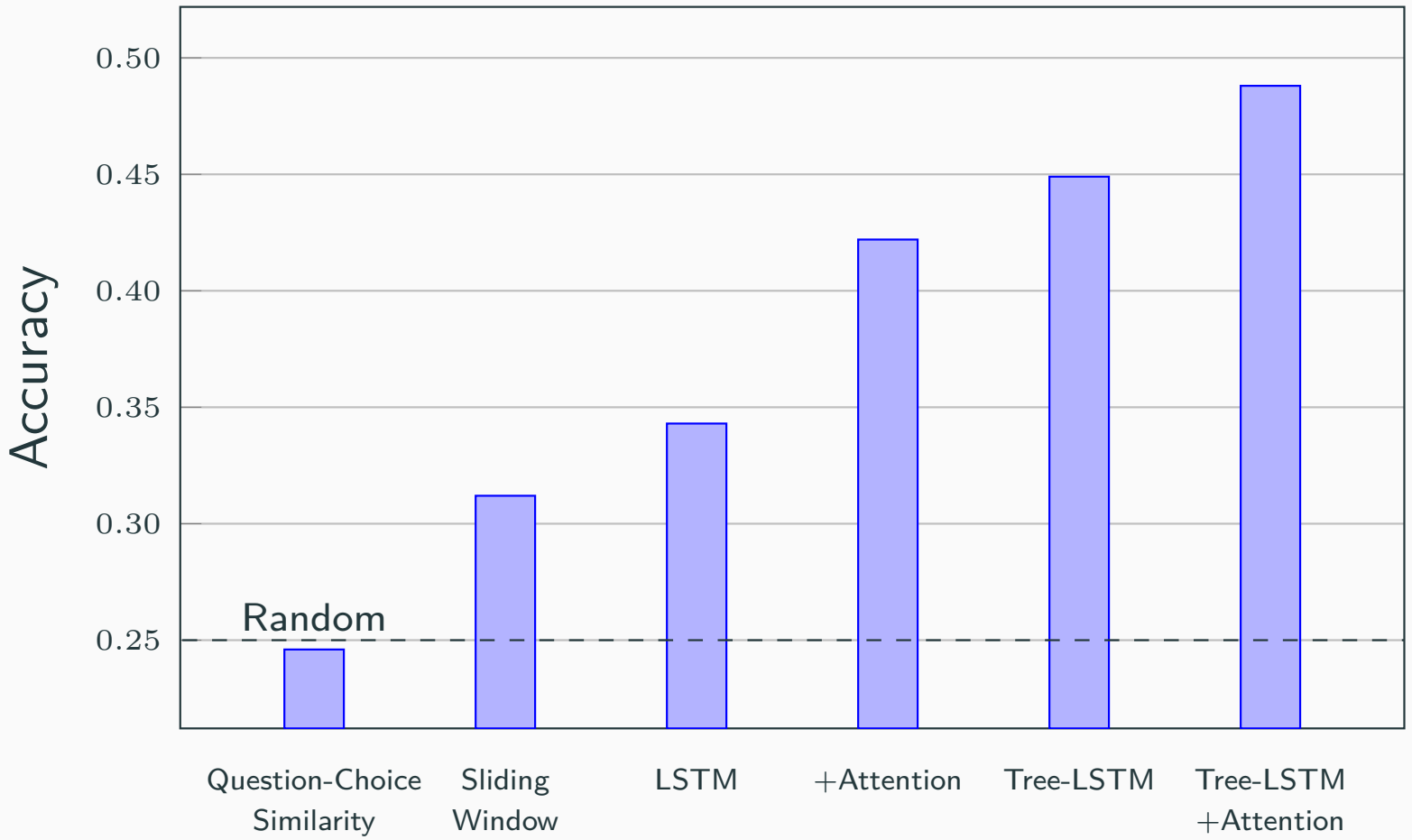
Sequential Attention

↓

Hierarchical Attention

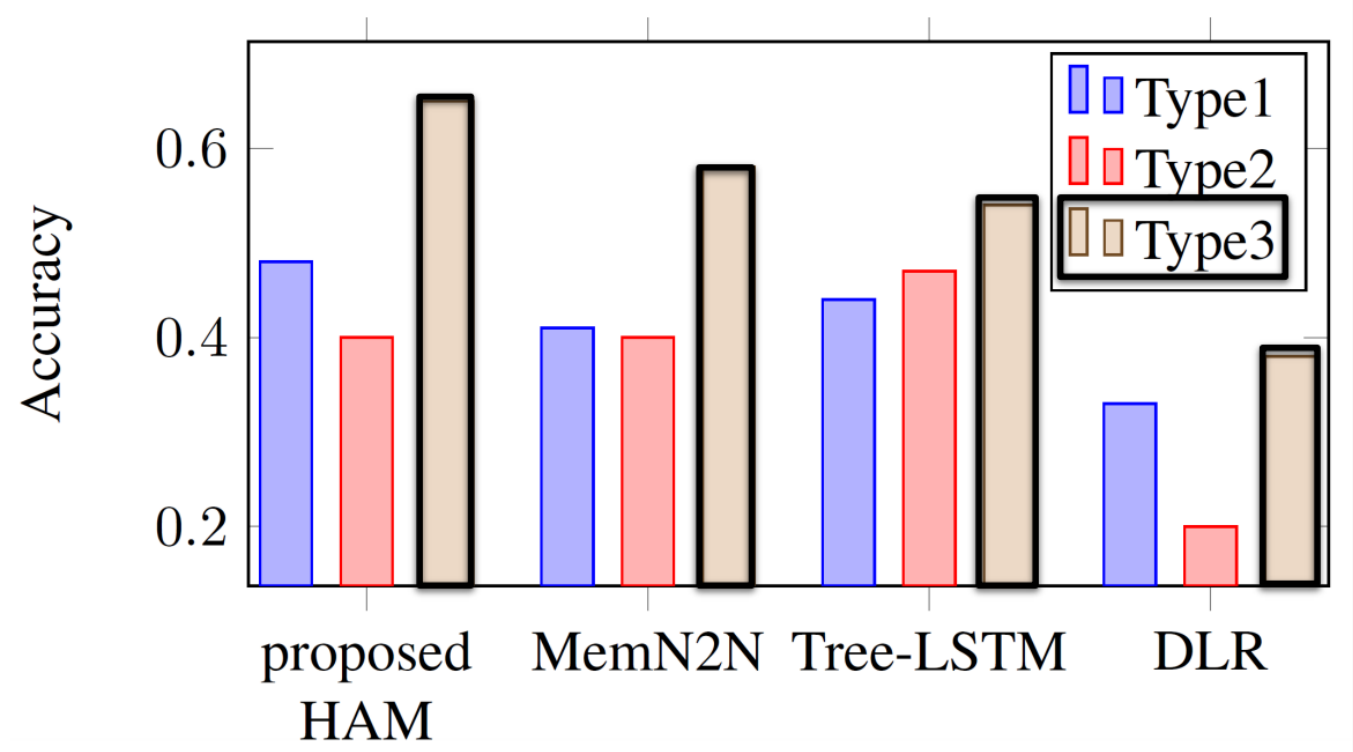
Fang, Hsu, Lee, Lee. *Hierarchical Attention Model for Improved Machine Comprehension of Spoken Content.* SLT 2016.

Experimental Results



Analysis

There are 3 types of questions.

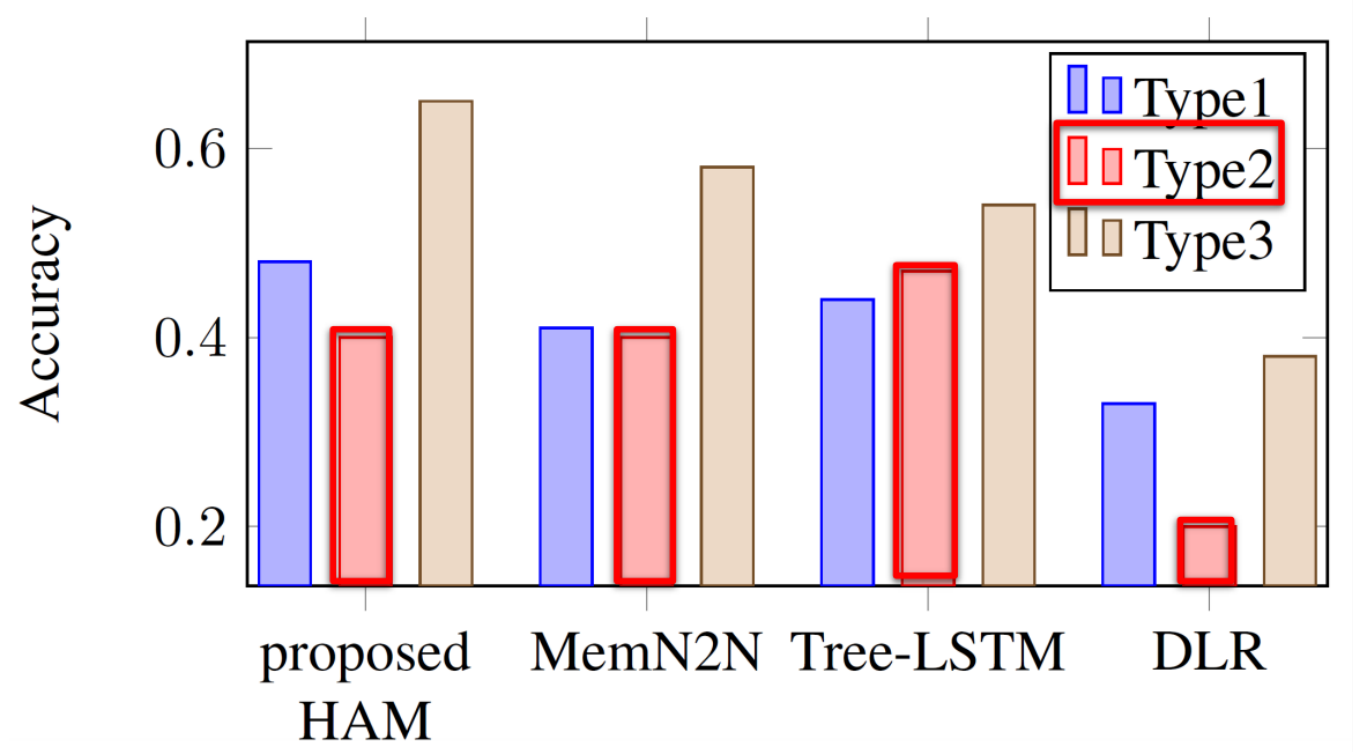


Type 3: Connecting Information

- Understanding Organization
- Connecting Content
- Making Inferences

Analysis

There are 3 types of questions.

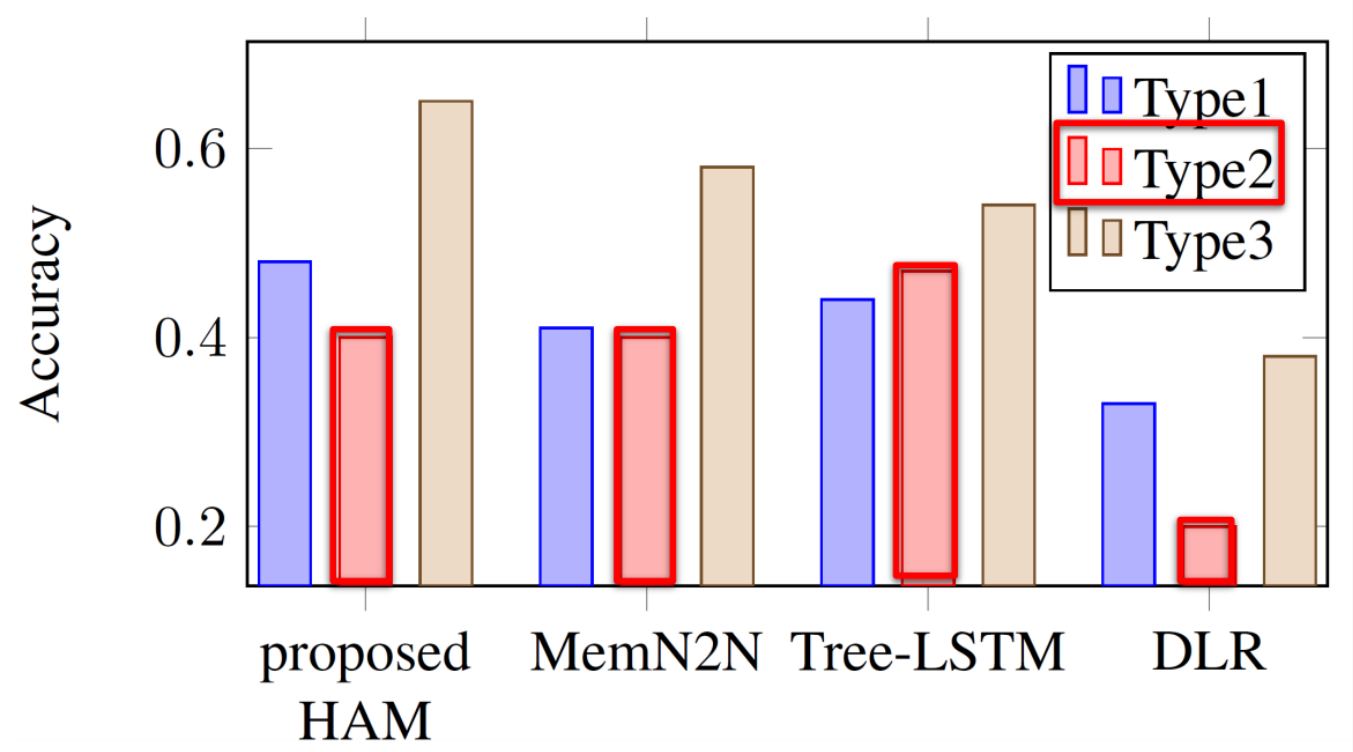


Type 2: Pragmatic Understanding

- Function of What is Said
- Speaker's Attitude

Analysis

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.

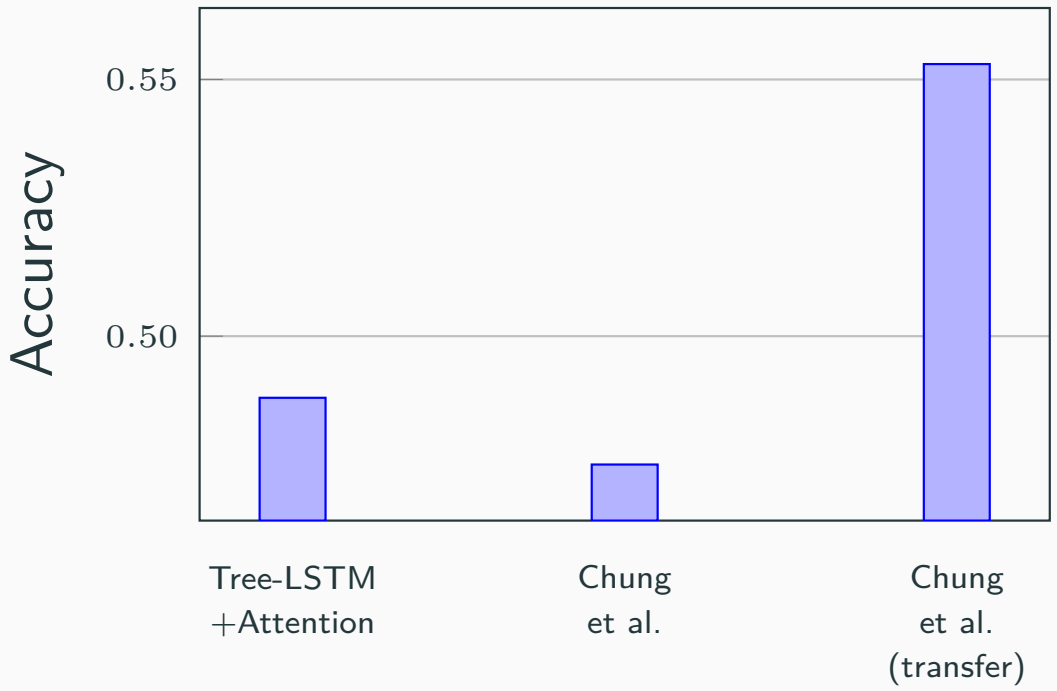
Tapaswi, Zhu, Stiefelhagen, Torralba, Urtasun, Fidler. *MovieQA: Understanding Stories in Movies through Question-Answering*

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*



Chung, Lee, Glass. *Supervised and Unsupervised Transfer Learning for Question Answering.* arXiv 2017.

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