Hierarchical Attention Model for Improved Comprehension of Spoken Content

Wei Fang, Jui-Yang Hsu, Hung-yi Lee, Lin-Shan Lee - National Taiwan University

National Taiwan University

Motivation & Contribution

- Long-term Vision: machine comprehension of spoken content
  - machine to comprehend the spoken content over the Internet (YouTube, on-line courses, etc.) for human and offer selected parts to human
- Present Task: TOEFL listening comprehension test by machine
- Proposed: Hierarchical Attention Model based on Tree-structured LSTM for better comprehension

TOEFL Listening Comprehension Test

An Example Problem

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

Experimental Results

- Manual: Transcription by human.
- ASR: Transcription by CMU Sphinx.

Model		Accuracy (%) Manual	ASR
Deep-LSTM		32.4	34.3
E2E Memory Networks[1]		45.2	44.4
AMRNN[2]	word	42.5	40.1
	sentence	42.4	42.2
Tree-LSTM		46.5	44.9
proposed HAM	phrase (1 hop)	47.7	47.4
	phrase (2 hop)	49.0	**48.8**
	sentence (1 hop)	**49.1**	48.6
	sentence (2 hop)	48.2	47.5

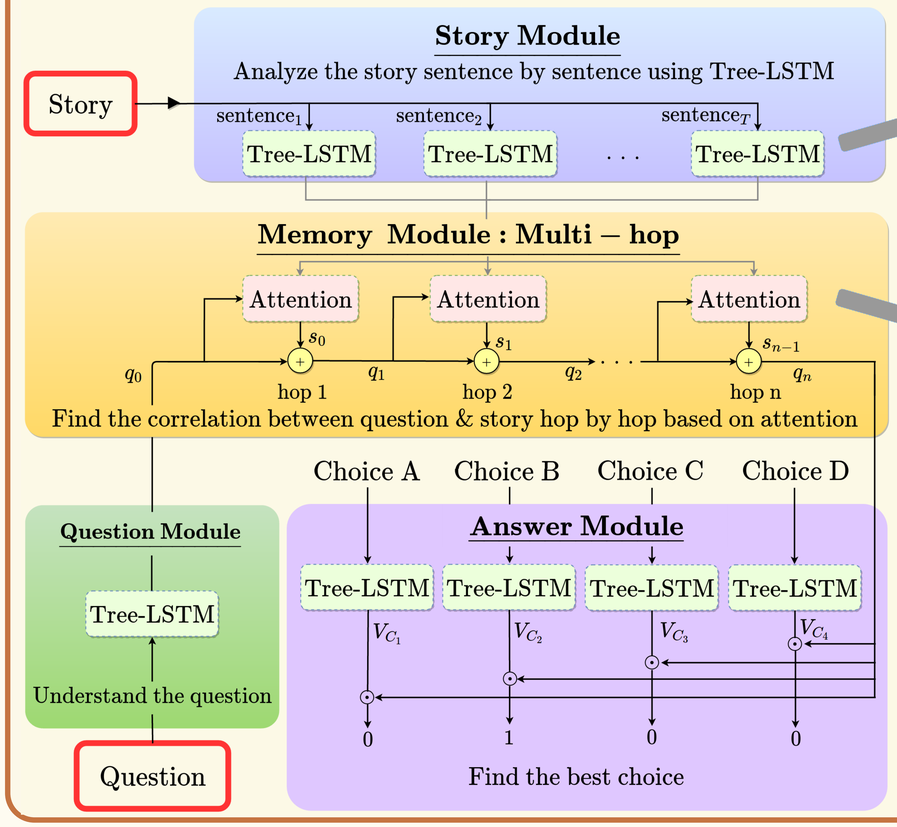
1. For manual, 1-hop sentence-level attention performed the best
2. For ASR, 2-hop phrase-level attention has the highest accuracy
3. Performance of ASR is just slightly lower than manual

[1] Sukhbaatar et al. End-to-end Memory Networks. In *NIPS 2015*.

[2] Tseng et al. Towards Machine Learning Comprehension of Spoken Content: Initial TOEFL Listening Comprehension Test by Machine. In *Interspeech 2016*.

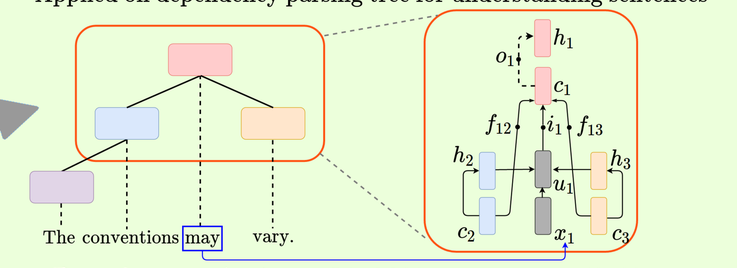
Proposed Approach

Complete Model

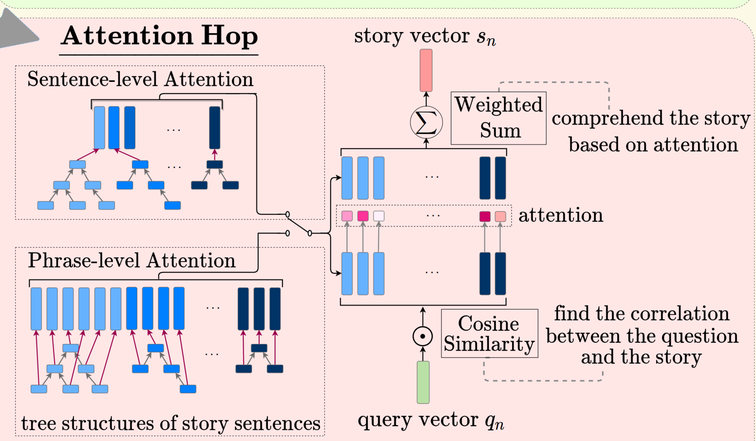


Tree – LSTM :

Applied on dependency parsing tree for understanding sentences



Attention Hop



For each hop, attention is obtained from correlation between the question & the story, and is then used to comprehend the story

Visualizing Attention

Hop 1 - Phrases with top 3 attention

1. incoming energy sunlight that is reflected off that surface back to space
2. there is not much reflection going on at all
3. they transmit incoming solar energy down to earth

Hop 2 - Phrases with top 3 attention

1. the amount of solar radiation energy from the sun absorbed by earth and the amount reflected back into space
2. increasing area of low thick clouds that reflect a large portion of solar energy back to space and cools the earth
3. process that could control the type of clouds

◇ The model selects key phrases **related to the definition** in hop1, and composes them to more **high-level definition** in hop2.

Source Code

The source code and TOEFL Listening Comprehension Test dataset is available at: `https://github.com/sunprinceS/Hierarchical-Attention-Model`