QUESTION ANSWERING FOR SPOKEN LECTURE PROCESSING

MERVE UNLU[1,3] EBRU ARISOY[2] MURAT SARACLAR[3]
[1]GALATASARAY UNIVERSITY [2]MEF UNIVERSITY [3]BOGAZICI UNIVERSITY

MOTIVATION

Developing a **question answering (QA) system** for spoken lectures.

QA System: Textual questions + Transcriptions of lecture videos

Challenges:

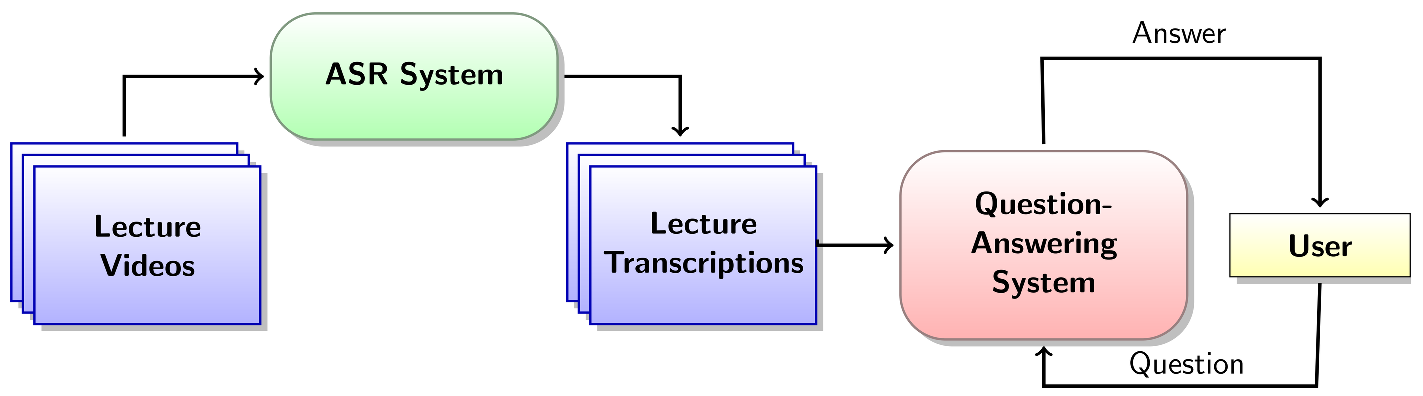
1. ASR transcriptions contain recognition errors.
2. Lecture videos can be long.
3. Answers can be longer than couple of words.
4. Task is domain specific.

Proposed Approach:

Convert the problem into machine comprehension style QA.

- **Machine Comprehension**: Textual questions + Short passages
  Use competitive end-to-end neural network models.
- Split lecture transcriptions into short pseudo passages.
- Automatically match questions with passages.

SPOKEN LECTURE PROCESSING SYSTEM



- **Lecture Videos:** Short videos ($\sim$ 5 minutes long).
  4 different engineering courses (totally $\sim$ 4 hours).
  Recordings from a single speaker.

- **ASR System:**
  Acoustic model (AM) trained with Kaldi[1] using 2.7 hours lectures.
  4-gram language model (LM) trained with SRILM[2].
  LM data: Reference transcriptions from MIT OCW[3] (100K words)
  + lecture slides (3.2K words) + AM data (22.3K words).

- **QA System:** End-to-end neural network model.

  Train Data: QA data generated from MIT OCW transcriptions.
  Manually split transcriptions into short passages ($\sim$ 72 words).
  Generate questions for passages (310 QA pairs - 259 passages).

  Test Data: QA data generated from test lecture transcriptions.
  Test lectures: Signals and Systems course (1.2 hours).
  Manually split transcriptions into short passages ($\sim$ 81 words).
  Generate questions for passages (175 QA pairs - 94 passages).
  Questions: $\sim$ 11 words Answers: $\sim$ 22 words

  Two-stage QA: Passage-question matching for pseudo passages.
  Search for answers in pseudo passages.

QA EXAMPLE

Paragraph: Now just as with the Fourier transform there are a number of properties of the Laplace transform that are extremely useful in describing and analyzing signals and systems. For example one of the properties that we in fact took advantage of in our discussion last time was the linearly **the linearity property which says in essence that the Laplace transform of the linear combination of two time functions is the same linear combination of the associated Laplace transforms.** Also there is a very important and useful property which tells us how the derivative of a time function rather **the Laplace transform of the derivative is related to the Laplace transform in particular the Laplace transform of the derivative is the Laplace transform x(t) multiplied by s** and as you can see by just setting s equal to j omega in fact this reduces to the corresponding Fourier transform property.

Question 1: **What is the linearity property in the Laplace transform?**
Question 2: **How is the Laplace transform of the derivative of a time function related with the Laplace transform of this time function?**

METHOD

MatchLSTM with Answer Pointer[4]:

1. **Preprocessing:** Representations for passage **P** and question **Q**.

$$\mathbf{H}^p = \overrightarrow{LSTM}(\mathbf{P}), \mathbf{H}^q = \overrightarrow{LSTM}(\mathbf{Q})$$

2. **MatchLSTM:** Question aware passage representations $\mathbf{H}^r$.

$$\overrightarrow{\mathbf{G}}_i = \tanh(\mathbf{W^q H^q} + (\mathbf{W^p h^p_i} + \mathbf{W^r} \overrightarrow{h}^{\mathbf{r}}_{\mathbf{i-1}} + \mathbf{b^p}) \otimes e_Q)$$

$$\overrightarrow{\alpha}_i = \text{softmax}(\mathbf{w^T} \overrightarrow{G}_{\mathbf{i}} + b \otimes e_Q) \quad \text{(attention vector)}$$

$$\overrightarrow{z}_i = \begin{bmatrix} \mathbf{h}^p_i \\ \mathbf{H}^q \overrightarrow{\alpha}_i^T \end{bmatrix} \qquad \overrightarrow{\mathbf{h}}^r_i = \overrightarrow{LSTM}(\overrightarrow{z}_i, \overrightarrow{\mathbf{h}}^r_{i-1})$$

$\overrightarrow{\mathbf{H}}^r = [\overrightarrow{\mathbf{h}}^r_1, .., \overrightarrow{\mathbf{h}}^r_P]$ ($\overleftarrow{\mathbf{H}}^r$: *same calculations in reverse order.*)

$\mathbf{H}^r$: concatenation of $\overrightarrow{\mathbf{H}}^r$ and $\overleftarrow{\mathbf{H}}^r$

3. **Answer Pointer:** Prediction of the answer (span of words).

Passage-Question Relevancy: Assign questions to pseudo passages.

- Split transcription into fixed length pseudo passages.
- Compute relevance scores between pseudo passages and questions using question aware passage representations.
- Select a single passage (with the maximum score) for each question.

$\mathbf{h}_{rel}$: concatenation of $\overrightarrow{\mathbf{h}}^r_P$ and $\overleftarrow{\mathbf{h}}^r_1$

$$\mathbf{D} = \tanh(\mathbf{W^h_{rel} h_{rel}} + \mathbf{b^h_{rel}}) \qquad \theta = \text{sigmoid}(\mathbf{W^o_{rel} D} + b^o_{rel})$$

EXPERIMENTS

Train-Test Scenarios:

- **short-short:** Short (manually split) passages for train and test data.
  Not a realistic scenario.
- **short-long:** Short passages for train data + long passages (transcription of the whole video) for test data.
- **long-long:** Long passages ($\sim$ 650 words) for train and test data.
- **window-window:** Fixed length (200 words) pseudo passages for train and test.

Train-Test Scenarios	Test Set F1 Score with known passage-question pairs			
	GMM-si*	GMM-sa*	DNN*	Ref
short - short	56.38	55.62	57.02	60.47
short - long	23.39	23.91	24.21	25.51
long - long	27.84	28.69	29.70	29.87
window - window	38.76	40.47	42.84	42.63

WERs: GMM-si (13.0%), GMM-sa (10.5%), DNN (6.7%)

Train-Test Scenarios	Test Set F1 Score with passage-question pair selection**			
	GMM-si*	GMM-sa*	DNN*	Ref
short - short	43.65	46.10	47.74	49.12
window - window	33.05	33.53	34.31	35.59

*QA pairs for ASR were obtained by aligning reference with hypothesis.
**Questions of each chapter were searched in all videos of that chapter.

CONCLUSION AND FUTURE WORK

- A QA system (based on competitive machine comprehension models) was developed for spoken lectures in the signal processing area.
- A passage-question matching stage was proposed to handle a realistic scenario where the answer for each question is searched in a chapter of the course lectures.
- Future Work: Extend the data to increase diversity.
  Deal with unanswerable questions in QA system.

REFERENCES AND ACKNOWLEDGEMENTS

[1] Daniel Povey et al., "The Kaldi speech recognition toolkit," in ASRU 2011.

[2] Andreas Stolcke, "SRILM – An extensible language modeling toolkit," in ICSLP 2002.

[3] https://ocw.mit.edu/resources/res-6-007-signals-and-systems-spring-2011/video-lectures/

[4] Shuohang Wang and Jing Jiang, "Machine comprehension using match-lstm and answer pointer," arXiv preprint arXiv:1608.07905 (2016)

This work is supported by the TÜBİTAK-ARDEB 3501 Program (Project No: 117E202).