Translation of Egyptian-Arabic Conversational Telephone Speech

Gaurav Kumar

IBM Research, Johns Hopkins University

gkumar@jhu.edu

September 9, 2014

Overview

Speech Translation

Describe the problem of translating conversational telephone speech her. Put in the image for the pipeline.

The Egyptian Arabic Callhome Corpus

- Callhome Egyptian Arabic Speech/Transcripts (ECA-96 : train, dev, test)
- 1997 HUB5 Arabic Evaluation (97-eval-H5)
- Callhome Egyptian Arabic Speech/Transcripts Supplement (ECA-supplement)

Partition	# Conv	# Utt's	# Words	Words/Utt
ECA-96 (train)	80	20,861	139,035	6.66
ECA-96 (dev)	20	6,415	34,543	5.38
ECA-96 (test)	20	3,044	16,500	5.42
97-eval-H5	20	2,800	18,845	6.73
ECA-supplement	20	2772	18039	6.51

Table: Partition statistics for the Callhome Egyptian Arabic corpus, supplements and evaluation datasets.

The Egyptian Arabic Callhome Corpus : Translations

Briefly describe how the dataset was created. Redundant translation X4. Maybe move the statistics table to the next slide. Talk about the problem of inter-speaker agreement on the next slide.

The Egyptian Arabic Callhome Corpus : Translations

Partition	# Utt's	# Words	Words/Utt
ECA-96 (train)	86,313	713,549	8.27
ECA-96 (dev)	25,769	186,400	7.23
ECA-96 (test)	12,212	85,182	6.98
97-eval-H5	11,248	91,647	8.15
ECA-supplement	11,126	87,489	7.86

Table: Reference translation statistics for the Egyptian-Arabic Callhome corpus.

Partition	Crossfold BLEU
ECA-96 (train)	40.09%
ECA-96 (dev)	35.64%
ECA-96 (test)	35.86%
97-eval-H5	35.81%
ECA-supplement	37.15%

Table: Crossfold average BLEU per partition of the Callhome Egyptian Arabic corpus, supplements and evaluation datasets.

The Egyptian Arabic Callhome Corpus : Translations

Source	mA Antw mbtrdw$ ElY Altlyfwn ybqY
Translation 1	you do n't reply to the phone
Translation 2	so you do n't answer the phone then
Translation 3	you do n't answer the phone it seems
Translation 4	because you do n't answer the call then
Source	mSEbAn Elyh nfsh kmAn
Translation 1	he feels hard for himself too
Translation 2	he feel bad about himself
Translation 3	he feels sorry for himself too
Translation 4	i feel sorrow about his condition too

Table: A sample of the translations for the Egyptian-Arabic Callhome Corpus. The translations are lower-cased, tokenized and punctuation has been normalized.

The ASR system

Put details of data used for the ASR system, model and WERs here.

Decoder modes and their efficacy for Speech Translation

Describe three decoder modes and their results on translating CTS. Conclude the T2S gives no gain over Hieros and Hieros provide a very small gain over Monotone decoding.

The effect of punctuation

ASR output does not typically contain punctuation (inability to map this to an acoustic sequence). How does this affect our translation pipeline? Describe experiments with removing puncutation from the phrase table and it's effect on translation. Conclude that for CTS where the input to SMT does not contain punctuation, the best strategy is to remove punctuation from the source in the phrase table, collapse duplicates, merge counts and re-calculate model 1 probabilities.

Selecting appropriate ASR output

Talk about the ASR 1-best output, the word lattice, weights on the word lattices (ASR + LM), and the oracle. Stress the point that even though ASR recognition quality is really bad, the difference between the oracle WER and the ASR 1-best WER is about 20 points. This encourages a search for a better hypothesis in the lattice.

Selecting appropriate ASR output

Discuss three strategies of selecting a better hypothesis from the lattice. Context : Word lattice, Segmentation transducer, Phrase lattice 1. Selecting the hyp that needs the least number of phrases to cover it 2. Use the ASR system to break draws (there are a lot of draws because the ASR hyps are really close together in weight space in the lattice) 3. Use the ASR weights and unigram probability weights derived from the phrase table in the segmentation transducer 4. Do not penalize longer phrases that appear less often in the phrase table, Use a length constraint to normalize the unigram probablilities to get a new score. Pushing weights achieves stochasticity.

Baseline results with existing decoder for DF

Mention the dataset that is consistently being used for decoding evaluation. Statistics ? Describe the montone decoder that was tuned on DF data and how it was used to decode the CTS dataset. Share the results of this exercise.

Tuning on CTS data

Talk about tuning on the CTS data, change of the lambda values and the results of translations

Perplexity training for the LM interpolation weights

Talk about tuning on the CTS data, change of the values and the results of translations

Hesitations, noise and non-vocal markers in tuning data

What happens when empty sentences are removed? In addition to the experiments above ? Include table of hesitation symbols

Length constrained tuning

Describe strategy of tuning on smaller datasets for smaller datasets and the same for the larger. Use the next slide to include results in the form of a matrix. Decide on the structure.

A hope for the future : Decoding ASR output with lower WERs

Share results of decoding the lower WER ASR output. Mention models

Most common errors

n-gram coverage ? Most severe mis-recognitions? How do you measure this? Put this if you have time.

Backchanneling

share results of bc, how the results were improved with handling this

Conclusion

Future work

Two stream decoding. Inclusion of Heiros in path selection.

Questions?