SPOKEN DIALOG SYSTEMS: NON-NATIVE SPEAKERS

Norah Hogoboom ∞ LING 575 ∞ Spring 2013

Contexts of Non-Native SDS

Monolingual Contexts

- Designed for one language, trained on native speech exclusively
- Designed for one language, trained with native and one target non-native language
- Designed for one language, trained with native and many target non-native language

Multilingual Contexts

- Interleaved (single voice, multiple languages)

  "Let's have lunch *mañana* for *Cinco de Mayo.*"

  "*Hana Yori Dango* is my favorite *manga*."

- Separate (single voice for each of multiple language options)

  "My favorite Japanese comic is Boys Over Flowers."

  "*私の好きな漫画が花より男子です。*"

Speech Contexts

- Spontaneous speech
- Read speech

What's the Problem?

The performance of a system is impacted when there is a mismatch between the user input (non-standard) and the system's expectations (standard)

- **Acoustic mismatch** arises from the variations between the native speech on which the acoustic models were trained and non-native speech, which often include different accents and pronunciations.
- **Linguistic mismatch** stems from variations or errors in syntax and word choice, between the native corpus on which the language model was trained and non-native speech.

Acoustic Properties

Pronunciation Variations (Accent)

- Articulation of phonemes
    - substitutions, deletions and insertions
- Using correct allophones in context

Prosodic effects

- Speaking rate
    - Slower compared to native speech
- Pauses and disfluences
    - More pauses compared to native speech
    - Disfluences outside range of native speech and inserted in unexpected locations

Linguistic Properties

Vocabulary

- Lexical selection differs from native speakers
  - "bus timing" for "bus schedule"
- Incorporation of native terms in disfluences
  - "a-no" for "uh" in Japanese speakers

Syntax

- Incorrect syntax
- Different and more varied choices from native speakers
- Native Speakers:
  - When does the bus leave?
- Non-native speakers:
  - Which time I have to leave?
  - What the next bus I have to take?

Raux and Eskenazi found that utterances with OOV more likely to have recognition errors, but that perplexity was still higher due to variety of expressions

Acoustic Solutions

Use linguistic knowledge of the speaker's native language to predict non-native phonetic and/or acoustic realization patterns of the target language

- Gruhn used hand-authored rules:
  - Weighted phoneme confusion rules

$$\theta \rightarrow [t^h] \; / \; \_\#$$
$$\theta \rightarrow [s] \; / \; \_\#$$

  - Adding weighted word-level pronunciation variations

/ ˈiːðər / or / ˈaiːðər / for the word 'either'

Acoustic Solutions

Extract patterns from a corpus of non-native speech from a specific population

- Raux's approach was to first define possible vocalic substitutions, use the data to determine the frequency of the alternates, then assigned to one of two random clusters
  - Rule format P [S] N → T
  - Compute the number of times each rule was "selected" by a speaker in the adaptation data
  - Prune rules that appeared less times than a given cutoff threshold
  - Apply the rule set to the recognition lexicon
  - Determine the probability of the variant given the word's MLE for either cluster
  - Reassign session to most likely cluster

Table 1: *List of possible vocalic substitutions*

AA → AO	AA → OW
AE → AA	AE → EH
AH → AA	AH → UH
AH → UW	AO → OW
AO → AA	AW → OW
AW → UW	AY → EY
EH → AE	EH → EY
ER → AA	EY → EH
IH → IY	IY → IH
OW → AW	OW → AO
UH → UW	UW → UH

Linguistic Solutions

Include non-native data in Language Model

- Raux and Eskenazi found that it reduced the OOV rate and perplexity, but the greatest improvement was on native language understanding, probably due to reducing OOV errors.

Adapt grammar used to parse user input

- In the Let's Go grammar, when a complete parse was not found, it next looked for partial parses and returned most confident one (based on number of words covered and number of parse trees in the parse forest)
- A smaller number of parse trees was favored when the number of covered words was equal

Design Solutions

Train the user to the system's language

- This mirrors how conversation occurs between a native and non-native speaker, aka "lexical entrainment"
- The system takes the role of the native speaker and prompts the user in the language of the system
- For the Let's Go!! System they
  - hand-authored a list of target prompts aligned with system's LM
  - identified the closest target using a dynamic programming algorithm — the number of word insertions, deletions and substitutions that from the target prompt to the user's utterance.
  - gave additional weight important concepts
  - asked for a confirmation when a target sentence was found
    - "Did you mean ..." + target sentence.

Conclusion

- In an increasingly global and multi-cultural world, this will only become more and more important
- It is important to understand your end users and take them into consideration in the design and implementation of the system

References

Primary	Raux, A. and Eskenazi, M. (2004) Non-Native Users in the Let's Go! Spoken Dialog System: Dealing With Linguistic Mismatch HLT/NAACL 2004, Boston, MA
Secondary	Y. Xu and S. Seneff. (2009). "Speech-Based Interactive Games for Language Learning: Reading, Translation, and Question-Answering," International Journal of Computational Linguistics and Chinese Language Processing, vol. 14, no. 2
Secondary	L. Tomokiyo. (2000) Linguistic properties of non-native speech, in Proceedings of ICASSP 2000
Supplementary	A. Raux. (2004). Automated Lexical Adaptation and Speaker Clustering based on Pronunciation Habits for Non-Native Speech Recognition, INTERSPEECH (ICSLP) 2004.
Supplementary	L. Tomokiyo and A. Black and K. Lenzo. (2005) Foreign Accents in Synthesis: Development and Evaluation, Interspeech 2005.
Supplementary	Y. Xu and S. Seneff. (2012). "Improving Nonnative Speech Understanding Using Context and N-Best Meaning Fusion," Proc. ICASSP, pp. 4977-4980.
Additional	Gruhn, R. E., W. Minker and S. Nakamura (2011). Statistical Pronunciation Modeling for Non-Native Speech Processing. Berlin Heidelberg, Springer-Verlag.