S1 Panel: “Handling of Unexpected Acoustic Data”
ASRU 2007, Kyoto, Dec. 10, 2007

Open Vocabulary Recognition

Ralf Schlüter

Human Language Technology and Pattern Recognition
Computer Science Department
RWTH Aachen University
D-52056 Aachen, Germany

Standard approach: large, but fixed vocabulary.

- Suitable for e.g. dictation in a fixed domain.
- Less suitable for open vocabulary settings like broadcast news:
  - Number of different words does not appear to be finite.
  - Important content words change over time.
- Fixed vocabulary implies out-of-vocabulary (OOV) words, which:
  - are never recognized, and are substituted by in-vocabulary word[s].
  - lead to misrecognition of neighboring words.
  - lead to errors that cannot be recovered by later processing stages (e.g. translation).
  - often are content words.

Aims for ASR:

- **Reduce number of word errors per OOV.**
- **Limit collateral damage to neighbouring words.**
- **Generate (at least approximately) correct transcription of OOVs.**

Idea:

$\longrightarrow$ Integrate OOVs into ASR decision rule.

Problems:

- **Individual OOVs are rarely seen in training.**
- **Coverage by acoustic model: OOV pronunciations?**
- **Generation of correct letter transcription?**
- **Language model representation?**
- **Prevent (incorrect) recognition of in-vocabulary words as fragments?**

Possible approach [Bisani & Ney, Interspeech 2005]:

- **Individual OOVs are rarely seen in training.**

→ Model OOVs by (seen) word fragments.

Possible approach [Bisani & Ney, Interspeech 2005]:

- **Individual OOVs are rarely seen in training.**

→ Model OOVs by (seen) word fragments.

- **Coverage by acoustic model: OOV pronunciations?**
- **Generation of letter transcription?**

→ Fragment pronunciations:
Train letter to phoneme mapping on existing pronunciation lexicon.

Possible approach [Bisani & Ney, Interspeech 2005]:

- **Individual OOVs are rarely seen in training.**

→ Model OOVs by (seen) word fragments.

- **Coverage by acoustic model: OOV pronunciations?**
- **Generation of letter transcription?**

→ Fragment pronunciations:
Train letter to phoneme mapping on existing pronunciation lexicon.

- **Language model representation?**
- **Search: prevent recognition in-vocabulary words as fragments?**

→ Hybrid language model covering words and word fragments in parallel:
Replace OOVs in LM training data by fragments.

Possible approach [Bisani & Ney, Interspeech 2005]:

- **Individual OOVs are rarely seen in training.**

→ **Model OOVs by (seen) word fragments.**

- **Coverage by acoustic model: OOV pronunciations?**
- **Generation of letter transcription?**

→ **Fragment pronunciations:**
Train letter to phoneme mapping on existing pronunciation lexicon.

- **Language model representation?**
- **Search: prevent recognition in-vocabulary words as fragments?**

→ **Hybrid language model covering words and word fragments in parallel:**
Replace OOVs in LM training data by fragments.

- **Remaining problems:**

→ **Generation of (OOV) words from fragment sequences?**

→ **Coverage by subsequent processing stages (e.g. translation)?**

Results on WSJ Dictation Task (adjacent fragments concatenated)

vocabulary				# word errors
words	fragments	OOV [%]	WER [%]	per OOV word
4986	0	11.2	24.26	1.72
	4085		16.54	1.15
19977	0	2.6	11.58	1.88
	11622		9.79	1.27
64735	0	0.5	8.92	1.99
	14346		8.87	1.46

Results for Arabic BN (fragments discarded)

vocabulary		# fragments		OOV	WER
words	pron.	fragm.	pron.	[%]	[%]
64k	125k	0	0	5.2	22.6
		8.7k	13k		21.5
126k	232k	0	0	2.9	20.9
		8.7k	13k		20.4
256k	423k	0	0	1.3	20.5
		8.6k	12k		20.1