STATISTICAL PHRASE-BASED SPEECH TRANSLATION

Lambert Mathias and William Byrne

Center for Language and Speech Processing, Johns Hopkins University, Baltimore, MD 21218, USA
Dept. of Engineering, Cambridge University, Trumpington Street, Cambridge, CB2 1PZ, USA

Introduction

- Objective: Tight integration of Automatic Speech Recognition (ASR) and phrase-based Statistical Machine Translation (SMT) systems.
  - ASR word lattices as input to the SMT system
  - Lattices encode a larger search space
  - Exploit sub-sentential information
- Modeling issues:
  - Propagation of ASR information to the SMT component
  - Correct disfluencies, hesitations, spontaneous speech effects...
  - Efficient phrase extraction
- Previous work:
  - (E.Matusov et al 2005) reported translation gains using word lattices
  - (Bertoldi et al 2005) used confusion networks for integration with the SMT system
- We present: Generative source-channel model of speech to text translation.
  - Tight coupling of the ASR and SMT system using word lattices
  - Implemented using weighted finite state machines (WFSMs)
- So what's new?
  - Conditional models vs joint models of target-source generation
  - Unified modeling framework
  - No need for extensive reformulation of underlying ASR and SMT models
  - Simpler decoding and estimation procedures
  - Lattice translation is a direct extension of our text translation systems

Generative Model of Text Translation

- Noisy channel model for text translation

	Target Sentence		Target Phrase		Source Phrase		Source Sentence
	t_1^J	$\longleftarrow$	v_1^R	$\longleftarrow$	u_1^K	$\longleftarrow$	s_1^I
Models		$P(t_1^J \mid v_1^R)$		$P(v_1^R \mid u_1^K)$		$P(u_1^K \mid s_1^I)$	$P(s_1^I)$
FSMs		Ω		Φ		W	G
		Target Phrase Segmentation Transducer		*Phrase Translation, Reordering Transducer*		*Source Phrase Segmentation Transducer*	*Source Language Model*

- Translation system translates phrase sequences and not word sequences
  - Target phrase sequence acceptor $\mathbf{Q} = \mathbf{project\_input}[\Omega \circ \mathbf{T}]$
  - $\mathbf{T}$ is an acceptor for target word sequence $\mathbf{t}_1^J$
  - A phrase sequence is a sequence of words to be translated
  - Different phrase sequences lead to different translations
  - Phrase sequence acceptor is unweighted
- The final translation is given by

$$\hat{\mathbf{s}}_1^I = \operatorname{argmax}_{\mathbf{s}_1^I}\{\max_{\mathbf{v}_1^R, \mathbf{x}_1^K, \mathbf{u}_1^K, K} \mathbf{P}(\mathbf{t}_1^J, \mathbf{v}_1^R, \mathbf{u}_1^K, \mathbf{s}_1^I)\} \ .$$

  - Implemented as a best-path search through the translation FSM $\mathcal{T}$

$$\mathcal{T} = \mathbf{G} \circ \mathbf{W} \circ \Phi \circ \mathbf{Q}$$

Generative Model of Speech Translation

- Noisy channel model for speech translation

	Target Speech		Target Sentence		Target Phrase		Source Phrase		Source Sentence
	A	$\longleftarrow$	t_1^J	$\longleftarrow$	v_1^R	$\longleftarrow$	u_1^K	$\longleftarrow$	s_1^I
Models	$P(A \mid t_1^J)$		$P(t_1^J \mid v_1^R)$		$P(v_1^R \mid u_1^K)$		$P(u_1^K \mid s_1^I)$		$P(s_1^I)$
FSMs	$\mathcal{L}$		Ω		Φ		W		G
	ASR Word Lattice		*Target Phrase Segmentation Transducer*		*Phrase Translation, Reordering Transducer*		*Source Phrase Segmentation Transducer*		*Source Language Model*

- The ASR word lattice is one of the components in the noisy channel model
- The translation system translates the lattice of phrase sequences
  - $\mathbf{Q}$ is obtained by applying Ω to the ASR lattice $\mathcal{L}$
  - Unlike the text translation case, $\mathbf{Q}$ is a weighted acceptor with scores from the acoustic word lattice attached to it.
- The final translation is given by

$$\hat{\mathbf{s}}_1^I = \operatorname{argmax}_{\mathbf{s}_1^I}\left\{\max_{\mathbf{t}_1^J, \mathbf{v}_1^R, \mathbf{x}_1^K, \mathbf{u}_1^K, K} \mathbf{P}(\mathbf{A}, \mathbf{t}_1^J, \mathbf{v}_1^R, \mathbf{u}_1^K, \mathbf{s}_1^I)\right\} \ .$$

  - Implemented as a best-path search through the translation FSM $\mathcal{T}$

$$\mathcal{T} = \mathbf{G} \circ \mathbf{W} \circ \Phi \circ \mathbf{Q}$$

Transforming ASR Word Lattices into Phrase Lattices

- Reformulate speech translation as a modeling problem
  - Efficient extraction of phrases from ASR word lattice
- ASR lattice pre-processing
  - Map unspoken tokens to NULL
  - Standard FSM operations: epsilon removal, determinization and minimization
- Controlling ambiguity in the ASR word lattice
  - Path based likelihood pruning of ASR word lattices
  - Extracting phrases under the posterior distribution

$$Q = P(v_1^R \mid A) = \frac{\sum_{t_1^J} P(v_1^R \mid t_1^J)\, P(A \mid t_1^J)\, P(t_1^J)}{P(A)}$$

- What about the target LM $\mathbf{P}(\mathbf{t}_1^J)$?
  - Shown to improve translation
  - Doesn't show up in the noisy channel model
  - Need to reformulate the SMT models to properly incorporate the target LM.
- Phrase extraction from ASR word lattice
  - Lattice subsequences of limited length
  - GRM Library tools for counting subsequences in a WFSM

Target ASR Word Lattice

Target Phrase Sequence Lattice

Speech to Text Translation Performance

- 2005 TC-STAR Mandarin Broadcast News Translation Task
  - 6 Mandarin news broadcasts, manually segmented and transcribed into Chinese sentences (525 for DEV set, 495 for EVAL set)
  - 2 English reference translations per sentence for translation.
  - Mandarin ASR lattices generated by LIMSI, corresponding to 231 segments for DEV set and 181 segments for EVAL set.
- Translation system: similar to 2005 JHU/CU NIST Chinese-English SMT system.
- Document level BLEU for evaluating translation performance.
- Mandarin Broadcast News Translation Performance

	Mandarin Source	DEV	EVAL
Monotone Phrase Order	Ref. Transcription	16.1	18.8
	ASR 1-Best	14.8	13.6
	ASR lattice	15.0	13.8
MJ-1 VT Phrase Reordering	Ref. Transcription	16.1	19.3
	ASR 1-Best	15. 0	13.8
	ASR lattice	15.1	14.0

- How many new foreign phrases does speech translation introduce?

	DEV transcriptions	ASR Lattice
#Chinese phrases	44744	58395

- How many new foreign phrases were found in bitext?

	DEV transcriptions	ASR Lattice
#Chinese phrases	11617	12983
# English phrases	59589	60574

- Extracting phrases from ASR lattice doesn't result in a larger phrase inventory
  - DIsfluencies in speech - different modeling approaches for phrase extraction
  - Mismatch in tokenization for the Chinese ASR lattices

Conclusion

- Presented a modeling framework for statistical speech-to-text translation
  - Extension of the phrase-based TTM text translation model
  - Tight coupling of the ASR and SMT subsystems using lattices both as
    - ∗ Statistical models
    - ∗ WFSM based implementation
  - Demonstrated feasibility of the above approach.
- Future Work
  - Initial formulation and implementation has weaknesses
    - ∗ Proper integration of the target language model
    - ∗ Phrase extraction under the posterior distribution
    - ∗ Improved pruning strategies for word lattices
    - ∗ Improved phrase coverage
  - Integrated development of the component ASR and SMT systems.