On ASR-MT integration and joint training

George Saon
IBM T.J. Watson Research Center

ASR-MT integration

- **Augmenting the ASR-MT interface is an ongoing, challenging problem**
  - Gradual increase from n-best lists to confusion networks, to general lattices, to joint decoding
- **Mixed results for large-scale speech translation tasks encountered in the GALE program**
  - Up to 1 TER/BLEU point improvement for monotone decoding
  - Little or no improvement when using reordering even for shallow lattices
- **MT search complexity overcomes potential benefit of using lattices**

One possible approach

- **Factor MT search into two parts:**
  1. Monotone decoding of best foreign path
  2. Non-monotone decoding of best English path given 1.
- **Advantages:**
  - gains from monotone lattice decoding may carry over to non-monotone decoding
  - Search complexity: max(lattice monotone, 1-best reordering)
  - Can scale up to joint decoding

GALE Arabic-to-English speech translation results:

- **Monotone decoding:**

	BNAD05s	DEV07	EVAL06s-BN	EVAL06s-BC
1-best	64.0%	60.1%	67.6%	70.6%
lattice	63.0%	59.8%	67.8%	69.8%

- **Non-monotone direct decoding:**

	BNAD05s	DEV07	EVAL06s-BN	EVAL06s-BC
1-best	61.7%	58.9%	66.9%	70.3%
lattice	61.5%	59.1%	67.4%	70.8%

- **Non-monotone two-pass decoding:**

	BNAD05s	DEV07	EVAL06s-BN	EVAL06s-BC
1-best	61.7%	58.9%	66.9%	70.3%
lattice	61.2%	58.8%	67.1%	69.9%

Joint discriminative training

- **ASR and MT systems are trained in isolation**
  - Different corpora
  - Different objective functions
- **Train ASR system using MT objective function**
  - Can ASR compensate for MT errors ?
  - Can MT compensate for ASR errors ?
- **Requires tight integration (lattices, joint decoding)**