Using Spoken Word Posterior Features in NMT

Kaho Osamura, Takatomo Kano, Sakriani Sakti*, Katsuhiro Sudoh*, Satoshi Nakamura*
Nara Institute of Science and Technology (NAIST), Nara, Japan / *RIKEN Center for Advanced Intelligence Project (AIP), Japan

Quick summary

- Use word posterior distributions as NMT inputs in SLT by ASR-NMT pipeline
  - Handle ASR ambiguity using word posterior distributions
  - Train with both 1-hot (text) and distributional (ASR) inputs
  - Improvements over a simple cascade with ASR 1-bests
    - 4-5 pts. BLEU gains on BTEC (synthesized) and ATR-English (natural)

Problem & previous approaches

- ASR error propagation (well-known!)
  - Lattice-based integration [Su+ 2017, Sperber+ 2017 (also many such studies by SMT)]
    - Complex implementation/computation
  - Direct network integration [Berard+ 2016, Kano+ 2017]
    - Works poorly in English-Japanese

Proposed method

Integration based on word posterior

1. Obtaining word posterior distributions from the softmax layer of an ASR decoder

2. Using the posterior distributions as inputs to an NMT decoder (i.e. using weighted word embeddings)

Advantages

- Simple integration by small modifications to existing ASR & NMT implementations
- Straightforward combination of text- and ASR-based NMT training
  - Pre-training with 1-hot text inputs
  - Fine-tuning with ASR posteriors

Experimental results

Consistent improvements over 1-best

- Outperformed text-NMT with low-WER!?&
  - ASR error recovery by word posterior
    - R: Excuse me where is the closest shoe store station (0.439)
      - B: すみません一番近い駅はどこですか
      - P: すみません一番近い靴屋はどこですか
    - Shoe store
  - Resolving word confusion by ASR-aware embeddings even without ASR errors
    - R: I'd like to have a perm and a haircut please
    - B: パーマとパーマをお願いします
    - P: パーマとカットをお願いします
  - Future work: Joint training of ASR+NMT

Conclusions

We could achieve simple but effective ASR-NMT integration by word posteriors