UNIVERSITY OF CAMBRIDGE

Sentence Boundary Detection Based on Parallel Lexical and Acoustic Models

Xiaoyin Che, Sheng Luo, Haojin Yang, Christoph Meinel

Indigo Orton – R250

Computer Laboratory

Key idea

- Sentence boundary detection
- Combining lexical and acoustic models
- Expanding usable training data – able to use unaligned lexical data

Problem – Grammar in speech

- Punctuation restoration
- Readability
- Downstream NLP
- Grammar is fundamental to meaning
- Aid for manual transcription

Problem – Multi-modal training data

- Modals – lexical and acoustic
- Lexical models are currently the most powerful standalone models, but multi-modal is better
- Align lexical and acoustic data
- Larger corpora unaligned

Details

Lexical Model

- Word vectors
- *M*-sliding window
- Boundary at *K*-th word
- Predict punctuation *location* then *type*

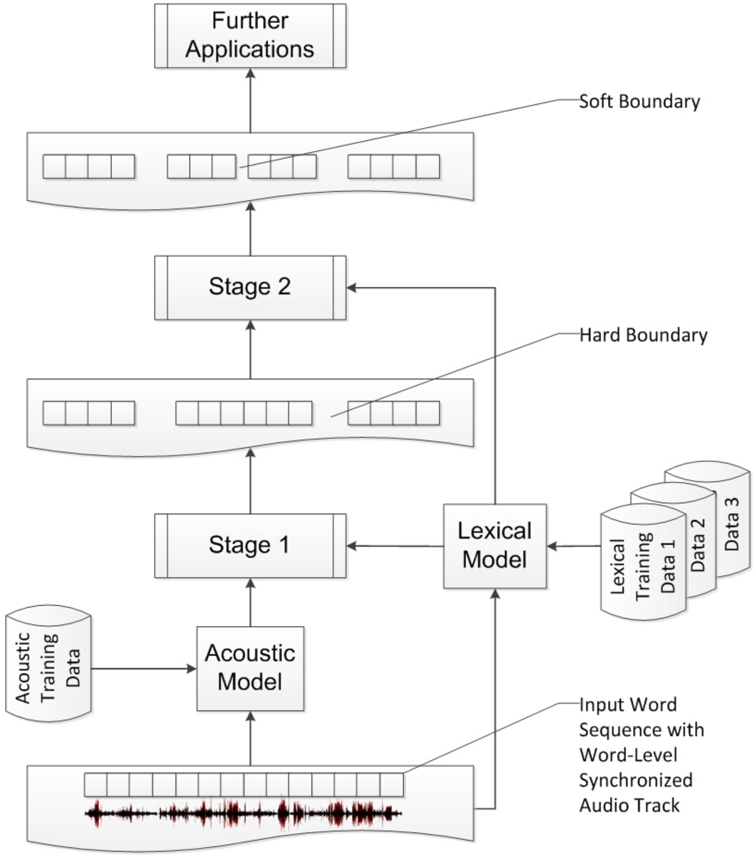
Acoustic Model

- Aligned data
- Pauses
  - 0.28 seconds
- Pitch average per word
- Energy average per word

Joint Decision Scheme – 2 Stages

1. Hard boundary – acoustic foundation, lexical filtering, posterior probability fusion
2. Soft boundary – lexical detection

Joint Decision Scheme – 2 Stages



Joint Decision Scheme – Posterior Probability Fusion

- Big emphasis
- If acoustic detects a boundary that lexical thinks is impossible, discard
- False positive filtering by lexical model
- Pauses due to hesitation/interruption
- Lexical model boundaries

Evaluation – Lexical only

- LMC-1, window size 5, *k*-th word 3
- LMC-2, window size 8, *k*-th word 4
- Single comparison model LSTM-[1]
- Datasets: TED-ASR and TED-Ref
- LMC-2 easily best

Evaluation – Lexical only

TED-ASR

Model	Punctuation (4 types)	Binary boundary
LSTM-[1]	46.2	65.2
LMC-1	49.6	70.7
LMC-2	**53.1**	**75.5**

Evaluation – Lexical only

TED-Ref

Model	Punctuation (4 types)	Binary boundary
LSTM-[1]	50.8	69.5
LMC-1	53.8	76.6
LMC-2	**58.0**	**82.4**

Evaluation – Joint Decision Scheme

- LMC-1 and LMC-2 with Pause and PPE (**P**ause, **P**itch, **E**nergy)
- Stage 1 – presented as relevant, but really just lexical model improving acoustic model by filtering false positives
- Stage 2 – more relevant as it is the final accuracy

Evaluation – Joint Decision Scheme

Model	Lexical	Acoustic	Stage 1	Stage 2
LMC-1 + Pause	70.7	60.9	71.1	77.6
LMC-2 + Pause	75.5	60.9	71.9	**79.2**
LMC-1 + PPE	70.7	61.0	72.0	76.2
LMC-2 + PPE	75.5	61.0	**73.1**	78.5

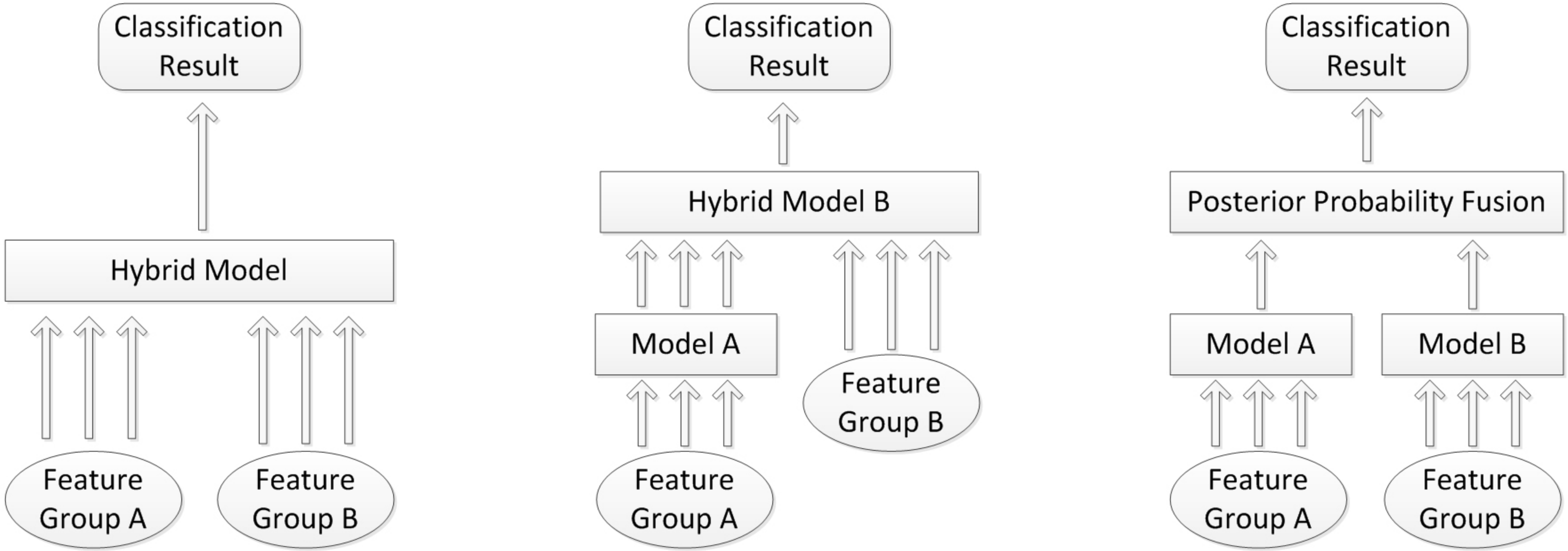
NB: PPE = Pause + Pitch + Energy

Uses TED-ASR dataset

Context

- Zhang et al. [3] – Lexical only
- Sinclair et al. [4] – Acoustic only
- Hasan et al. [2] – Multi-modal – Hybrid model A
- Tilk et al. [1] – Multi-modal – Hybrid model B – Comparison

Context – Multi-modal approaches



Review

Encouraging highlights

- Expand viable lexical training data
- Approach for combining acoustic and lexical

Further questions

- Evaluation – 1 comparison model
- Higher level acoustic features?
- Punctuation prediction
  - Larger scope
  - Use of acoustic model
  - Confusion matrix

Conclusion

- Grammar in speech transcripts
- Detect boundary then identify type
- Lexical model
- Acoustic model
- Combine with “posterior probability fusion” – confidence filtering

Thank you

References

1. O. Tilk and T. Aluma¨e, “LSTM for punctuation restoration in speech transcripts,” in *Sixteenth Annual Conference of the Inter- national Speech Communication Association (INTERSPEECH)*, 2015.
2. Hasan, M., Doddipatla, R., of, T. H. F. A. C., 2014. (n.d.). Multi-pass sentence-end detection of lecture speech. Isca-Speech.org
3. Che, X., Luo, S., Yang, H., & Meinel, C. (n.d.). Sentence Boundary Detection Based on Parallel Lexical and Acoustic Models. Pdfs.Semanticscholar.org
4. Zhang, D., Wu, S., Yang, N., Meeting, M. L. P. O. T. 5. A., 2013. (n.d.). Punctuation prediction with transition-based parsing. Aclweb.org
5. Sinclair, M., Bell, P., Birch, A., the, F. M. A. C. O., 2014. (n.d.). A semi-markov model for speech segmentation with an utterance-break prior. Isca-Speech.org