Exploiting Prosodic Breaks in Language Modeling with Random Forests

Yi Su and Frederick Jelinek

Center for Language and Speech Processing,
Department of Electrical and Computer Engineering,
The Johns Hopkins University, Baltimore, Maryland, USA

Introduction

Prosody:

- A *wide* range of suprasegmental properties: tone, intonation, stress, break, etc.
- Many applications: disfluency and sentence boundary detection, topic segmentation, spoken language parsing, etc.
- We are interested in using prosodic breaks for language modeling.

Random forest language model (RFLM):

- A collection of randomized decision tree language models
- Outperforms the n-gram language model in perplexity and word error rate
- Able to integrate information from various sources by asking new questions

Example:

1. what sort of benefits would you like to get from a big company/1
2. it just sort of/2 happens automatically/1

Prosodic Language Models

Speech Recognition With Side Information

- Proposal 1: If S is hidden, then

$$W^* = \arg\max_W P(W|A) = \arg\max_W P(A|W)P(W), \tag{1}$$

  where $P(W) = \sum_S P(W,S)$.
- Proposal 2: If S is observable, then

$$(W,S)^* = \arg\max_{W,S} P(W,S|A) \approx \arg\max_{W,S} P(A|W)P(W,S). \tag{2}$$

- Are prosodic breaks hidden or observable?
  - Although, strictly speaking, only acoustic features are observable, prosodic breaks can be estimated from acoustic features with high precision.
  - 83.12% for predicting a 3-valued break on annotated Switchboard. (Hale et al, 2006)

Joint Model of Words and Breaks

$$P(W,S) \approx \prod_{i=0}^{m} P(w_i, s_i | w_{i-n+1}^{i-1}, s_{i-n+1}^{i-1}) \tag{3}$$

- Tuple Model: Let $t_i = (w_i, s_i)$, for all $0 \le i \le m$. We have

$$P(w_i, s_i | w_{i-n+1}^{i-1}, s_{i-n+1}^{i-1}) = P(t_i | t_{i-n+1}^{i-1}). \tag{4}$$

- Decomposed Model:

$$P(w_i, s_i | w_{i-n+1}^{i-1}, s_{i-n+1}^{i-1}) = P(w_i | w_{i-n+1}^{i-1}, s_{i-n+1}^{i-1}) P(s_i | w_{i-n+1}^{i}, s_{i-n+1}^{i-1}). \tag{5}$$

Random Forest Language Models

Definitions

- N-gram language model

$$P(w|h) = P(w|\Phi_n(h)), \tag{6}$$

  where $\Phi_n(h)$ maps $(n-1)$ suffix-sharing histories into one class.
- Decision tree language model

$$P(w|h) = P(w|\Phi_{DT}(h)), \tag{7}$$

  where $\Phi_{DT}(h)$ maps histories into classes by a decision tree.
- Random forest language model

$$P(w|h) = \frac{1}{M} \sum_{j=1}^{M} P(w|\Phi_{DT_j}(h)), \tag{8}$$

  where $\Phi_{DT_j}(h)$ maps histories into classes by a randomized decision tree.

New Questions

- Questions we have asked:

  Is the word w_{i-1} in the set of words $\{a, an, the\}$?

- Questions we would like to ask:

  Does the prosodic break s_{i-1} take its value in the set of values $\{1, 2, 3\}$?

Experiments

Data and setup:

- ToBI-labeled Switchboard data; 10k vocabulary
- Prosodic break classifier from CLSP Workshop'05 (Hale et al, 2006)
- 666k words for training, 51k for development, 49k for evaluation
- Trigram with Modified Kneser-Ney smoothing; 100 trees per forest

Granularity

We believe a finer granularity than the ToBI scheme is needed for language modeling. 12-valued quantized posterior probability $P(1|\text{features})$ from the prosodic break classifier was used.

Table 1: Granularity of Prosodic Breaks

Model	two-level	three-level	cont.-valued
KN.3gm	66.1	66.1	66.1
RF-100	65.5	65.4	56.2

Feature Selection

We built RFLMs for $P(w_i|w_{i-1}, w_{i-2}, s_{i-1}, s_{i-2})$ then masked out one of the features in order to see how much it contributed.

Table 2: Feature Selection by RFLM

History	Perplexity
w_{i-1}, w_{i-2}, s_{i-1}, s_{i-2}	56.2
w_{i-1}, w_{i-2}, s_{i-1}	55.9
w_{i-1}, w_{i-2}, s_{i-2}	63.9
w_{i-1}, w_{i-2}	62.3

Main Perplexity Results

We compared the combinations of estimating $P(W,S)$ then computed $P(W) = \sum_S P(W,S)$ using the forward algorithm.

Table 3: Main Perplexity Results

Model	Method	KN	RF
$P(W,S)$	tuple 3gm	358	306
	decomp.	274	251
$P(W)$	tuple 3gm	69.3	67.2
$= \sum_S P(W,S)$	decomp.	66.8	64.2
$P(W)$	word 3gm	66.1	62.3

Conclusions

Random forests are a feasible means for adding prosody into language models.

- Finer grained prosodic break indices are needed.
- Prosodic breaks should be given to language models.

Acknowledgements

We would like to thank Zak Shafran, Markus Dreyer and the whole CLSP Workshop'05 PSSED team for preparing and sharing the data. Thank ISCA for awarding a travel grant for the presentation of this poster!