Adding Word Duration Information toBigram Language Models

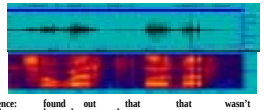
George Doddington

National Institute of Standards and Technology
doddington@nist.gov
http://www.nist.gov

Yufeng Wu, Aravind Ganapathiraju, Joseph Picone

Institute for Signal and Information Processing
Mississippi State University
{wu, ganapath, picone}@isip.msstate.edu
http://www.isip.msstate.edu

Motivation



- suprasegmental information plays an important role in human speech
- usually model suprasegmental information jointly with segmental measures

Bigram Duration Model

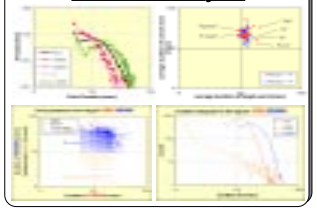
- Modeling duration within the context of the bigram language model

$$Pr(w_i|w_{i-1},\tau_{i-1},\tau_i)=Pr(w_{i-1},\tau_{i-1},w_i,\tau_i)Pr(w_{i-1},\tau_{i-1},\tau)$$
$$=\left\{\frac{Pr(\tau_{i-1},\tau_i|w_{i-1},w_i)}{Pr(\tau_{i-1},\tau_i|w_i)}\right\}\left\{\frac{Pr(w_{i-1},w_i)}{Pr(w_{i-1})}\right\}$$

$$Pr(w_1|w_0=S_{beg},\tau_1)=\left\{\frac{Pr(\tau_1|w_0=S_{beg},w_1)}{Pr(\tau_1|w_0=S_{beg})}\right\}\left\{\frac{Pr(w_0=S_{beg},w_1)}{Pr(w_0=S_{beg})}\right\}$$

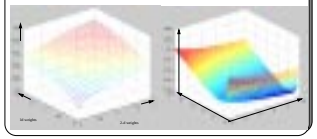
$$Pr(w_i=S_{end}|w_{i-1},\tau_{i-1})=\left\{\frac{Pr(\tau_{i-1}|w_{i-1},w_i=S_{end})}{Pr(\tau_{i-1}|w_{i-1})}\right\}\left\{\frac{Pr(w_{i-1},w_i=S_{end})}{Pr(w_{i-1})}\right\}$$

Duration Analysis



Error Analysis

- difference between the average duration model score for correct versus incorrect bigrams is crucial to performance



Suprasegmental Information

- each feature represented as a single scalar attribute of each word
- incorporating word duration into a bigram language model:

$$Pr(F|F_{i-1}) = Pr(w_i,\tau_i|w_{i-1},\tau_{i-1})$$
$$= Pr(\tau_i|w_i,w_{i-1},\tau_{i-1})Pr(w_i|w_{i-1},\tau_{i-1})$$

where, τ is duration, w is word identity and F is the new feature vector

Back-off Weighting

- sparsity of training data for many duration bigrams
- combine bigram-specific models with word-specific and word-independent models in a back-off framework

$$Pr_{smoothed}(\tau_{i-1},\tau_i|w_{i-1},w_i) = \frac{\left\{\begin{array}{l}\Omega_{bigram}Pr(\tau_{i-1},\tau_i|w_{i-1},w_i)\\ +\Omega_{word}Pr(\tau_{i-1},\tau_i|w_{i-1})Pr(\tau_i|w_i)\\ +\Omega_{global}Pr^2(\tau)\end{array}\right\}}{\Omega_{bigram}+\Omega_{word}+\Omega_{global}}$$

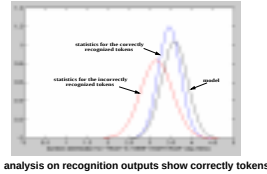
N-best Rescoring Results

- baseline -- 32.4% WER on 637 SWB utterances
- rescoring the 100-best hypotheses
- N-best error rate of 21.2%

	[weight 1d, weight 2d]		
scale	0.1 0.1	0.1 0.5	0.5 0.1
0.01	32.5	32.4	32.3
0.05	32.4	32.3	32.2
0.1	32.3	32.3	32.2

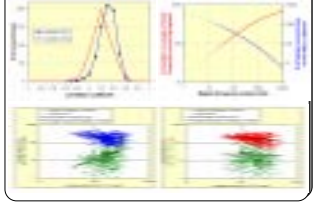
Summary

Duration Instance



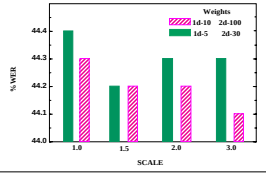
- analysis on recognition outputs show correctly tokens following real distribution closely

Switchboard Data



Wordgraph Rescoring Results

- baseline system -- WER 44.4% on WS97 test set



Future Work