Improving Phrase Prediction

Alok Parlikar (various and PhD thesis)
Phrasing

Phrase Breaks Indispensable in Natural Speech

• Breathing Rhythm
• Emphasis
• Intelligible Speech
• Dramatic Effects (Applause)
Phrasing

Phrase Breaks Important in Synthetic Speech

- Foundation of Synthetic Prosody
- Critical for natural, stylistic speech
Phrasing

Objective: Insert breaks in Utterances

He washed and fed the dog.

He washed and fed the dog.
Phrasing

Classifier:

Predict each word boundary in text as Break (B) or Non-Break (NB)
Better break prediction

- Base: POS context only
- Can we use Grammar
  - Not syntactic grammars
  - but **phrasal** grammars
Phrasal Grammars

- From natural data
- Using ASR forced alignment
  - Find word boundaries and pause boundaries
- Label words with POS tags
  - det noun verb PAUSE det adj noun
  - det noun prp det noun PAUSE verb det noun
- Bracket actual prosodic phrases
  - (det noun verb) (det adj noun)
  - (det noun prp det noun) (verb det noun)
Phrasal Grammars

- Train Stochastic Context Free Grammar SCFG
- Using large bracketed corpus
- Parse new sentences with trained SCFG
- Extract word boundaries features
  - Number of opening/closing brackets
  - Distance to common ancestor
  - Length largest phrase to this point
  - ...

Acoustic + Language Model

- Grammar features/POS tags PLUS
- Language model of B/NB
- Combined model
- Use viterbi to predict at run time
Festival’s Phrasing Architecture

Parlikar and Black, 2011
Low Resource Language Case

- Train from recorded data
  - Can still find phone/silence alignment
- But need POS tagger
  - Have to be created unsupervised
  - Fairly well defined problem
  - Find all words with same context
  - Rename those class X
  - Repeat until enough
- Pretty good at getting function words
Unsupervised Tagset

- English (Jane Austin's Emma)
  - BE HAVE
  - MR MRS
  - AND BUT THAT AS
  - TO FOR OF IN
  - VERY SO
  - HIM ME
  - COULD WOULD
  - SHE HE IT I YOU
  - THE A HER HIS
  - WAS IS HAD
Unsupervised Tagset

- German (Europarl)
  - IN AUF FüR VON
  - DIE DER DEN DIESE EINE
  - KOMMISSION UNION
  - UND DAß WIE
  - WERDEN HABEN
  - ZU MIT BEI
  - WIR SIE ICH
  - IST WIRD SIND
  - DAS ES
  - NICHT AUCH SICH
Unsupervised Tagset

- Chinese (News Text)
  - 中 /f 后 /f
  - 一 /m 两 /m 几 /m 三 /m
  - 年 /q 个 /q
  - , /w 的 /u 了 /u 着 /u
  - 要 /v 已 /d 就 /d
  - 他 /ngp 我 /ngp 我们 /ngp 他们 /ngp 人 /ngp 她 /ngp
  - 也 /d 这 /n 又 /d 还 /d
  - 是 /v 对 /p 在 /p 有 /v
General Phrasing Observations

- Training to optimize Likelihood
- Testing measures accuracy (F-measure)
- Low-data scenario (600 data points per hour)
- Difficult to vary Phrasing Rate
- (Necessary when varying speaking rate)
But we want …

Phrasing Model that is:
- Flexible
- Can combine multiple classifiers
- Can vary the phrasing rate
- Optimized directly to end evaluation metric
Festival's Updated Phrasing Architecture
Loglinear Model

Ideal break sequence

\[ b^* = \arg \max_b P(b|t). \]

Context

Directly model the posterior

Weights

Feature functions

Weights estimated by maximizing likelihood of development data. (Gradient Descent, or other search algorithm)
Minimum Error Rate Training

\[ P(b|t) = \frac{\exp\left(\sum_{m=1}^{M} w_m b_m(b, t)\right)}{\sum_{b'} \exp\left(\sum_{m=1}^{M} w_m b_m(b', t)\right)} \]

Estimate weights to maximize

- F-measure of breaks, rather than likelihood.
Minimum Error Rate Training

N-best list search

\[ b_n = \arg \max_{b \in S_n} \left[ \sum_{m=1}^{M} \omega_m h_m(b | T_n) \right] \]

Best phrasing for sentence “n”

Search over phrasing alternatives for that sentence

Learning Weights to minimize Error

\[ \omega_1^M = \arg \min_{\omega_1^M} \left[ E(D_1^N; \omega_1^M) \right] \]

The F-1 error surface is not smooth!

Gradient-descent not possible
Minimum Error Rate Training

- Inspired from Machine Translation
- Use Basin-Hopping Algorithm for searching optimal weights

Algorithm:
1. Decide which features to use
2. Start with random weights
3. Generate n-best list of phrase breaks
4. on a development set
Experiments

Features used:
- POS sequence model (Taylor and Black 98)
- Break sequence model (Taylor and Black 98)
- Grammar Based Model (Parlikar and Black 11)
- Break-Count

Corpora tested with:
- F2B (BURNIC)
- Two hours of Jane Austen’s book (EH2 from BC2013)
The new phrasing model is better
Phrasing Rate Knob

Can we vary the phrasing rate

Yes! Using the “Break-Count” feature

\[
S = \sum_{i=1}^{M-1} w_i \cdot h_i + w_c \cdot c
\]

Reducing the weight will favor phrasing hypotheses with more breaks

Increasing the weight will favor phrasing hypotheses with fewer breaks
Effect of the Knob

If we desire a particular proportion of breaks, how do we find the knob value?

$$w_c = A \cdot \tan \left( \pi \frac{x - C}{W} \right) + O$$
Knob, as estimated by the Tan Eqn
Phrasing Rate: Accuracy Impact
Do people perceive variations in phrasing rate?
Summary

Log-linear combination of

- phrasing models and arbitrary features
- Weights trained using MERT
- to optimize for F1, rather than likelihood
- “Break count” feature provides a “knob” that can vary the phrasing rate
- People do not perceive small variations in phrasing rate. Typically: Double the phrase breaks if you want to be noticed!
Phrasing Summary

• Two components
  • Context at wound boundary
  • Predictions over time
• Models
  • Local POS
  • Grammar and Phrasal Grammar
  • Log Linear Features combinations
• Useful controls
  • Slow/Fast know for number of breaks