A Front End SVM-HMM Speech Recognition System

Sarah Borys
January 26th, 2008
What goes into landmark based speech recognition?

- Phonetics
- Phonology
- Pattern Recognition
- Psychology
- Signal Processing
- And more!!!
SVMs

\[ g(\Phi(\vec{x})) = \sum_j \alpha_j y_j K(\vec{x}_j, \vec{x}) + b \]

\[ \mathcal{L}_d = \vec{\alpha} \cdot \vec{e} - \frac{1}{2} \vec{\alpha}' \tilde{Q} \vec{\alpha} \]
Acoustic Feature Selection
(Stop Releases and Vowels)
Manner Change Detection

![Graph showing comparison between SVM and RSSVM](image-url)
Manner Feature Detection

![Bar chart showing performance of SVM and RSSVM for various manner features.

- Stop Closure
- Flap
- Fricative
- Liquid
- Nasal
- Stop Release
- Syllabic Liquid
- Syllabic Nasal
- Vowel

The chart displays the percentage performance with SVM (blue) and RSSVM (red) for each manner feature.

Performance values:
- Stop Closure: 95%
- Flap: 96%
- Fricative: 92%
- Liquid: 85%
- Nasal: 90%
- Stop Release: 97%
- Syllabic Liquid: 88%
- Syllabic Nasal: 94%
- Vowel: 91%]
Place Feature Classification
(Stop Releases and Vowels)
Insertions and Deletions?
What About Non-landmark Times?
What About Non-landmark Times?
SVM Generalization
(Stop Closures and Vowels)
The SVM-HMM Hybrid System

\[ g_{td} = \sum_{j=1}^{M} a_{jd} e^{-\gamma |\tilde{x}_t - \tilde{s}_{jd}|^2} + b_d \]
Phone Recognition

![Graph showing % Accuracy vs Mixtures for Baseline, Manner, and Landmark methods.](image)
What Features are most Useful?
Word Recognition

![Graph showing the comparison of baseline, manner, and landmark mixtures in terms of percentage accuracy. The graph displays the trend of accuracy over the number of mixtures.](Image)
Open Questions

➲ What is the best way to represent the speech signal
  ● For the SVM?
  ● For the speech recognizer?
➲ How does SVM accuracy effect phone and word error rates?
➲ Do we need to locate segment boundaries before we can recognize sounds?