Unsupervised Word Discovery from Phonetic Input
Using Nested Pitman-Yor Language Modeling

Oliver Walter and Reinhold Haeb-Umbach
University of Paderborn, Germany
{walter,haeb}@nt.uni-paderborn.de
http://nt.uni-paderborn.de

Sourish Chaudhuri and Bhiksha Raj
Carnegie Mellon University, USA
{sourishc,bhiksha}@cs.cmu.edu
http://mlsp.cs.cmu.edu

Objective:
Unsupervised word language acquisition
"Learn a language like a child"

Backward sampling (Gibbs sampling):
"Learn a language like a child"

Ground truth
Real data from acoustic unit discovery to be used

Example segmentation (character input for illustration):
10506
Consider variations in pronunciation
49.9 51.2 52.1 56.8
52.0 58.3 59.0 63.7
41.5 47.3 48.3 52.5

Key idea:
CITY
17839 18466 17359 15775 15505

Probability for characters (phones) calculated as above

COUNCIL I NAPRIL P ASSED

Learn orthography (or transcription) and language model,

Database:
Account for unknown number of words, i.e., unknown

Ground truth
95629
ORD I NANCE REQUIRING LOCAL

HARRISBURG CITY COUNCIL

Probability for word
α

lexical unit (word) discovery

Two-level hierarchical process:
Order
4 5 6 7 8

Bayesian formulation:
Bigram language model
Input:
Use drawn word sequence to update language model
73.3% precision at 57.1% recall for word tokens

Goal:
Unsupervised Word Discovery from Phonetic Input

Outlook:
Unsupervised word segmentation on large vocabulary task
73.3% precision at 57.1% recall for word tokens

Problem Formulation

Goal: Segment character (or phone) sequence \( c^T \)
into words \( w^N \)

\( N \) lexical items (unique words)

Forward Filtering/Backward Sampling Alg.

To solve (1), iterate between:

1. Forward filtering: compute probability of candidate segmentations exploiting bigram language model
2. Backward sampling (Gibbs sampling): draw word segmentation from above probabilities
3. Use drawn word sequence to update language model

Eq. 1

Input: \( c^T = c_1 \ldots c_T \): character (or phone) sequence

\[ a[t][k] = \Pr (c_t = k, q_t = k) \]

\[ \text{Bigram probability} \]

\[ a[t - k][j] \approx \frac{\sum_{j=1}^{T-k} \Pr (c_t = k, q_{t-k} = j) \Pr (c_{t-k+1} = c_{t-k+1}^j)}{\sum_{j=1}^{T-k} \Pr (c_{t-k+1} = c_{t-k+1}^j)} \]

Conclusions

Unsupervised word language acquisition

73.3% precision at 57.1% recall for word tokens

Real data from acoustic unit discovery to be used

Extension to noisy data, e.g., lattice input

Consider variations in pronunciation