Ecological Language Acquisition via Incremental Model-Based Clustering

Giampiero Salvi

KTH CSC TMH giampi@kth.se

Nov. 2005
Introduction

Interspeech 2005

Part II

Mismatch Child/Parent Voice
Frame Based Processing?
Clustering Time Sequences
The Visual Channel
Conclusions
The Speech Chain

Denes and Pinson (1993)
Denes and Pinson (1993)
The Speech Chain

Denes and Pinson (1993)
Background: ecological theory of language acquisition (Lacerda et al., 2004)

the infant is naïve: no innate linguistic knowledge
- Background: ecological theory of language acquisition (Lacerda et al., 2004)
  - the infant is naïve: no innate linguistic knowledge
- Aim (long term): mathematical modelling of the learning process
  - acoustic features classification
  - time integration into meaningful sequences
  - integration of acoustic/visual information
Background: ecological theory of language acquisition (Lacerda et al., 2004)

- the infant is naïve: no innate linguistic knowledge

Aim (long term): mathematical modelling of the learning process

- acoustic features classification
- time integration into meaningful sequences
- integration of acoustic/visual information

Aim Interspeech 2005 (Salvi, 2005): acoustic features classification

- unsupervised
- incremental
Acoustic features

Equally spaced windows of speech

According to my
Assumption

Acoustic feature vectors independently drawn from mixture of Gaussian distributions
Method

- Model-Based Clustering (Fraley and Raftery, 1998)
  - data modelled as mixture of probability distributions
  - each distribution represents a cluster
  - each data point belongs to each cluster with a certain probability
  - model parameters estimated via Expectation Maximisation
  - different models compared via Bayes information criterion (BIC)

- Incremental Model-Based Clustering (Fraley et al., 2003)
Method

- **Model-Based Clustering (Fraley and Raftery, 1998)**
  - data modelled as mixture of probability distributions
  - each distribution represents a cluster
  - each data point belongs to each cluster with a certain probability
  - model parameters estimated via Expectation Maximisation
  - different models compared via Bayes information criterion (BIC)

- **Incremental Model-Based Clustering (Fraley et al., 2003)**
  - introduced for large datasets
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Algorithm

1. start with a MCLUST model
2. get new data
3. adjust old model to new data
4. divide new data into well and poorly modelled points
5. try a more complex model, if better BIC set as best and go back to 4
6. set the current best model and go back to 2
Experimental settings

- Data (ex1, ex2, ex3, ex4, ex5)
  - 12 minutes from the MILLE corpus
  - child directed speech (1 mother talking to her child)
  - Mel frequency cepstral coeffs computed every 10ms + differences of first and second order
Experimental settings

- Data (ex1, ex2, ex3, ex4, ex5)
  - 12 minutes from the MILLE corpus
  - child directed speech (1 mother talking to her child)
  - Mel frequency cepstral coeffs computed every 10ms + differences of first and second order

- Experimental factors
  - Dimensionality of the data: from 3 to 39 dimensions
  - Frame length: from 200msec to 3sec
Evaluation

- problem: there is no reference (at the moment)

\[ \text{VI}(C, C') = H(C | C') - H(C) - H(C') \]
Evaluation

- problem: there is no reference (at the moment)
- relative evaluation:
Evaluation

- problem: there is no reference (at the moment)
- relative evaluation:
- time evolution of number of clusters
  - dependency with number of feature coefficients
  - dependency with frame length

\[ VI(C, C') = H(C|C') - H(C) + H(C') \]
Evaluation

- **problem:** there is no reference (at the moment)
- **relative evaluation:**
  - time evolution of number of clusters
    - dependency with number of feature coefficients
    - dependency with frame length
  - agreement of classification in different conditions
    - variation of information (Meilä, 2002)

\[
\text{VI}(C, C') = H(C|C') + H(C'|C)
\]
Results

Effect of dimensionality (frame len = 50)

- nvars = 39
- nvars = 24
- nvars = 12
- nvars = 6
- nvars = 3

[Graph showing the effect of dimensionality on the number of clusters over time, with different line styles for each nvars value.]
Mismatch Child/Parent Voice
Mismatch Child/Parent Voice

- ASR with children
Mismatch Child/Parent Voice

- ASR with children
- Normalisation
  - VTLN: Vocal Tract Length Normalisation
  - Adaptation: hard in this context
Mismatch Child/Parent Voice

- ASR with children
- Normalisation
  - VTLN: Vocal Tract Length Normalisation
  - Adaptation: hard in this context
- Relative Features
Acoustic Features

Frame Based

File: sx352.WAV   Page: 1 of 1   Printed: Mon Dec 05 09:11:39

kHz
0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
time

According to my
Acoustic Features

Segment Based

File: sx352.WAV   Page: 1 of 1   Printed: Mon Dec 05 09:11:39

kHz

0.05 0.10 0.15 0.20 0.25 0.30 0.35 0.40 0.45 0.50 0.55 0.60 0.65 0.70 0.75 0.80 0.85 0.90 0.95
time

h#  ix  kcl  k  ao  r  dx  iy  n  tcl  t  ax  m  ay  ix  n  tcl  t

according to my
Acoustic Features

Landmark Based
Consequences

Sequence recognition (HMMs)

⇒

simpler relation
acoustic categories/
linguistic units
Clustering Time Sequences

Acoustic vectors independently drawn from mixture of gaussian distributions
Clustering Time Sequences

Acoustic vectors independently drawn from mixture of gaussian distributions
Clustering Time Sequences

Acoustic vectors independently drawn from mixture of gaussian distributions
Modeling time evolution with Markov chains
Modeling time evolution with Markov chains
Modeling time evolution with Markov chains
Modeling time evolution with Markov chains
Modeling time evolution with Markov chains
Modeling time evolution with Markov chains
No one-to-one relation acoustic/visual info
No one-to-one relation acoustic/visual info
Reinforcement Learning
  perform match at higher levels (pseudo-words or -phrases)
The Visual Channel

Perform visual/acoustic match on the Markov chain

Visual Event

Acoustic Event
The Visual Channel

Perform visual/acoustic match on the Markov chain

Visual Event

Acoustic Event
The Visual Channel

Perform visual/acoustic match on the Markov chain

Visual Event

Acoustic Event
The Visual Channel

Perform visual/acoustic match on the Markov chain

Visual Event

Acoustic Event
The Final Question

- Are the acoustic blocks (categories) in a language learned out of their statistical occurrence or out of their contrastive use?
The Final Question

- Are the acoustic blocks (categories) in a language learned out of their statistical occurrence or out of their contrastive use?
- In the first case: model based clustering and growing Markov chains are separate processes.
The Final Question

- Are the acoustic blocks (categories) in a language learned out of their statistical occurrence or out of their contrastive use?
- In the first case: model based clustering and growing Markov chains are separate processes.
- In the second case: need to integrate everything
Bibliography

http://www.speech.kth.se/~giampi


