Data Mining and Machine Learning

HMMs fo peech
Recogniti
Word and Sub-Wo el HMMs

Peter Jančovič

UNIVERSITY OF
BIRMINGHAM

Content

- Word level HMMs
- Sub-word HMMs
  - Phoneme-level HMMs
- Context-sens
  - Biphone HMMs
  - Triphone HMMs
- Triphone HMM training issues
- Phoneme Decision Trees (PDTs)

Word Level HMMs

- Early systems (1980s) used word level HMMs
- I.e. each word modelled by a single, dedicated HMM (c.f. “zero” picture)
  - Advantag
  - Good perf t modelling of word-dependent vari

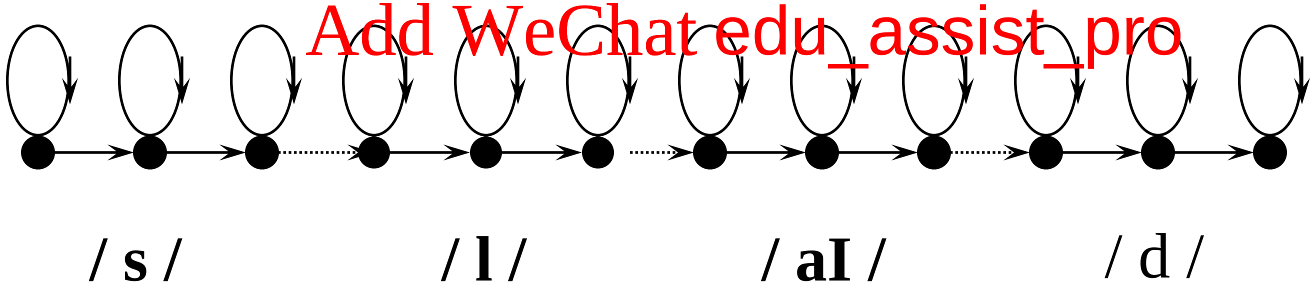
6 state HMM of the digit ‘zero’

Word Level HMMs

- Disadvantages:
  - Many examples of each word required for training
  - Fails to ex [illegible] oken language
- Word-level systems typica [illegible] ed to well-defined, demanding, small [illegible] y applications

Sub-Word Level HMMs

- Build HMMs for a complete set of sub-word 'building blocks'
- Construct word-level HMMs by concatenation of sub-word HMMs
- E.g. slide = /



Sub-Word Level HMMs

- Advantages
  - Able to exploit regularities in speech patterns
  - More efficient use of training data - e.g. in phoneme f aI v /) and “nine” (/n to /aI/ model.
  - Flexibility - acoustic m e built **immediately** for words which did not occur in the training data

Phoneme-Level HMMs

- Why choose phonemes rather than any other sub-word unit?
- Disadvantages
  - Phonemes f the contrastive properties o unds within a language - not their c with HMM assumptions!

Advantages of Phoneme-HMMs

- Completeness & compactness – approx. 50 phonemes required to describe English
- Well studied ation of ‘speech kno tion differences due to accent…
- Availability of extensive phoneme-based pronunciation dictionaries

Context-Sensitivity

- Problem
  - Acoustic realization of a phoneme depends on the c curs
  - Think of he “k” sound in the words “book sho hick”

Biphones and Triphones

- Solution
  - **Context-sensitive** phoneme-level HMMs
  - E.g.
    - ‘biphon op”
    - ‘triphones’ : (ku_S) i op”
- Almost all systems use triphone HMMs

Triphones - problems

- Increased number of model parameters
  - Need more (well-chosen) training data

- Which trip

  - If a word ains a triphone which was not in the tra which triphone HMM should we use?

Number of parameters

- If there are 50 phones, the maximum number of triphone HMMs is 50^3=125,000
- Most ruled out by **phonological** constraints – most phone triples
- But many are legal

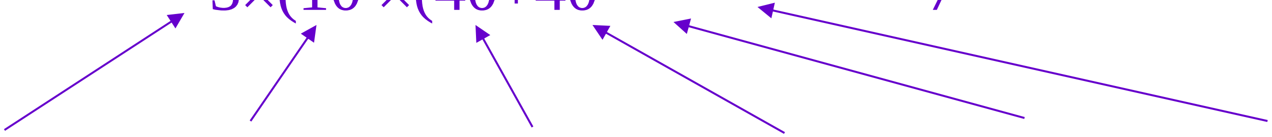
Example: Model Parameters

- Each model has 3 emitting states
- Each state modelled as, say, a 10 component Gaussian mixture

- Each feature al

- Hence numb odel is:

3×(10 ×(40+40 7



Number of states

Number of mixture components

Mean vector

Variance vector

Mixture weight

Transition probs

Acoustic model parameters

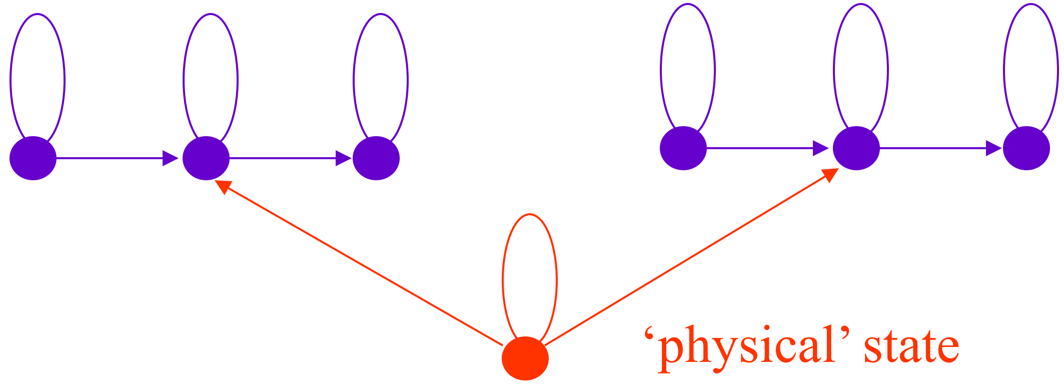
- So, even if we only have 1,000 acoustic models (instead of 125,000), total acoustic model parameters will be 2,457,000
- Too many to l quantity of data
- Most common solution is **ameter tying**
- **Different** HMMs share **sa** ters

Tied variance

- Variances are more costly to estimate than means
- Simple solution – divide set of all HMMs into classes, so that within a class all HMM state PDFs have same v
- This is **tied variance**
- If **all** HMM state PDFs sh e variance, the variance is referred to as **grand variance**

Phone decision trees

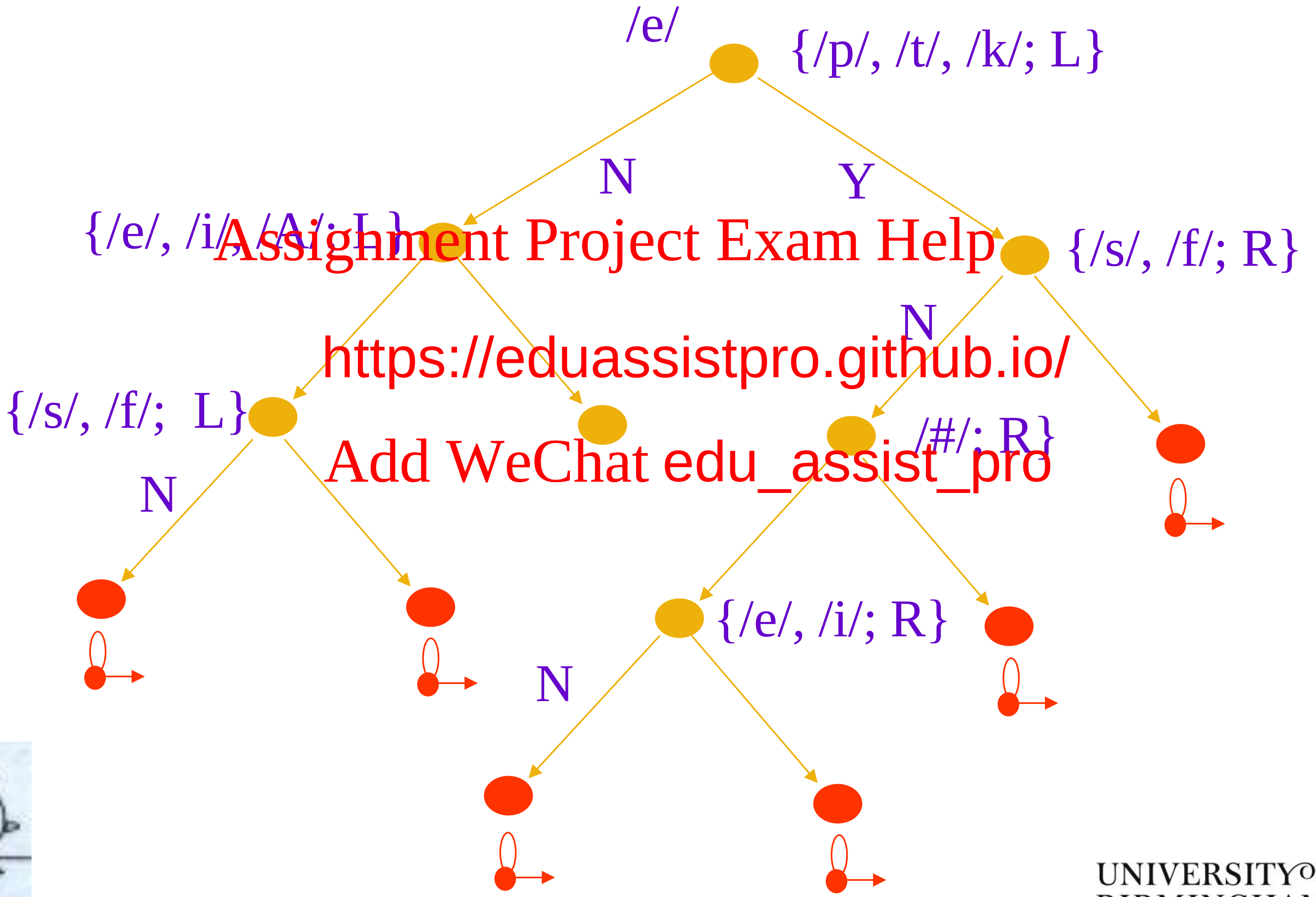
- Most common approach to general HMM tying is **decision tree clustering**
- Decision tree clustering can be applied to individual states or to w nsider states
- Basic idea is t t nes are likely to induce similar effects



Phonetic knowledge

- For example, we know that /*f*/ and /s/ are both unvoiced fricatives, produced in a similar manner
- Therefore we might **hypothesise** that, for example, an utterance ed by /*f*/ might be similar to
- This is the basic idea behi tree clustering

Phone Decision Tree



Summary

- Word-level and Sub-Word HMMs
- Phoneme-level HMMs
- Context-sens
  - Biphones &
- Triphone decision trees