Założenie że wytwarzanie mowy jest procesem stacjonarnym (a tak naprawdę quasi-stacjonarnym) w HMM, zalozenie ze przejście do stanu w chwili t jest zależne jedynie od stanu w chwili t-1, oraz prawdopodobieństwo P(O|qt) zalezy tylko i wyłaćznie od stanu qt.

Zastosowano w HMM podejście hierarchiczne, zdanie jest modelowane jako sekwencja słów , słowa są modelowane jako sekwencje fonemów, fonemy jako sekwencje sub-fonemów. Pomimo że lingwistycznie wskazane jest modelowanie sylab lub słów, to ze względów praktycznych modeluje się fonemy subfonemy.

Sieć neuronowa dokonuje jedynie decyzji lokalnych natomiast HMM dokonuje decyzji globalnych

Cechy MFCC i PLP są tutaj wiedzą a posteriori.

Klasyfikacja estymacja prawd. A posteriori

Neural „magic” can solve problem ofASR

Delta I delta dI>

HMM zaklada ze mowa jest w pewnym stopniu procesem stacjonarnym tj mowa jest wynikiem przechodzenia kolejnych stanow dyskretnych

Trening Hmm – modelu akustycznego – polega na estymowaniu rozkładu i prawd. Przejść miedzy stanami na podstawie obserwacji.

2. Wnioski

- w fazie reestimation wzory 9.29 i 9.30 musza być obliczane w skali liniowej, ponieważ składowe wektora cech jak i roznica składowych ze średnią mogą dawac w rezultacie liczbe ujemną co uniemożliwia policzenie logarytmu!!

Z HTK book

However, repeated re-estimation to convergence may take an impossibly long time.

Worse still, it can lead to over-training since the models can become too closely matched to the training data and fail to generalise well on unseen test data. Hence in practice around 2 to 5 cycles of embedded re-estimation are normally sufficient when training phone models. Str 124

When performing embedded training, it is good practice to monitor the performance of the

models on unseen test data and stop training when no further improvement is obtained.

In order to get accurate acoustic models, a large amount of training data is needed. Several hundred utterances are needed for speaker dependent recognition and several thousand are needed for speaker independent recognition. In the latter case, a single iteration of embedded training might take several hours to compute. There are two mechanisms for speeding up this computation. Firstly, HERest has a pruning mechanism incorporated into its forward-backward computation. HERest calculates the backward probabilities βj (t) first and then the forward probabilities αj (t). The full computation of these probabilities for all values of state j and time t is unnecessary since many of these combinations will be highly improbable. On the forward pass, HERest restricts the computation of the α values to just those for which the total log likelihood as determined by the product αj (t)βj (t) is within a fixed distance from the total likelihood P (O|M ). This pruning is always enabled since it is completely safe and causes no loss of modelling accuracy.

Pruning on the backward pass is also possible. However, in this case, the likelihood product αj (t)βj (t) is unavailable since αj (t) has yet to be computed, and hence a much broader beam must be set to avoid pruning errors. Pruning on the backward path is therefore under user control. It is set using the -t option which has two forms. In the simplest case, a fixed pruning beam is set. For example, using -t 250.0 would set a fixed beam of 250.0. This method is adequate when there is sufficient compute time available to use a generously wide beam. When a narrower beam is used, HERest will reject any utterance for which the beam proves to be too narrow. This can be avoided by using an incremental threshold. For example, executing