Hidden Markov Models (HMM)

Tutorial Part-2

FOCUS: • Problem Analysis • Discussions

Akshaya Thippur
Judith Butepage

TODO TODAY:

1. Recap of HMM Problems
2. Dynamic programming for HMMs
3. Example – Spell Checker
4. Example – Speech Recognition
5. Example – Music Recognition
6. Example – Character Recognition
7. Duck Hunt Discussions

HMM PROBLEMS:

0. Predicting most likely current emission

1. Evaluation Problem = ?

2. Decoding Problem = ?

3. Learning Problem =?

PROBLEM 2: Decoding

Given:

- Emission sequence **O** = $\{O_1, O_2 \dots O_T\}$
- A, B, q

To Find:

- Hidden state sequence **X*** = $\{X_1, X_2 \dots X_T\}$ that most likely produced **O**.
- Probability of occurrence of **X***

VITERBI ALGORITHM: (δ – Pass)

$\delta_t(i)$ = Probability of having been through state sequence **X*** = $\{X_1, X_2 \ldots X_t\}$,
as ***back-tracked*** by Viterbi algorithm
&&
having generated the observation sequence up to O_t.

- Solved using dynamic programming (DP) with an algorithm called Viterbi.
- Initialize: $\delta_1(i)=\pi_i b_i(O_1), \quad i=1,\ldots,N$
- For each t>1: $\delta_t(i)=\max_{j\in\{1,\ldots,N\}}[\delta_{t-1}(j)a_{ji}b_i(O_t)]$
- Probability of best path: $\max_{j\in\{1,\ldots,N\}}[\delta_T(j)]$
- Find path by keeping book of preceding states and trace back from highest-scoring final state.

PROBLEM 1: Evaluation (Puppy Platone Example)

Given:

A =

$X_t \mid X_{t+1}$	A	B	H	S
A	0.6	0.1	0.1	0.2
B	0.0	0.3	0.2	0.5
H	0.8	0.1	0.0	0.1
S	0.2	0.0	0.1	0.7

B =

$X_t \mid O_t$	p	e	b	l
A	0.6	0.2	0.1	0.1
B	0.1	0.4	0.1	0.4
H	0.0	0.0	0.7	0.3
S	0.0	0.0	0.1	0.9

q =

A	B	H	S
0.5	0.0	0.0	0.5

O = { b, p, l, e }

Find:

- P(**X***| A, B, q)

Solution:

- On the board …

Example – *Spell Checker*

You have to develop a program that corrects typos in the typed text *without* using a dictionary. You will be given a text containing some typographical errors and the goal is to correct as many typos as possible.

To make the problem easier, we will work with a very limited set of words that consists only of 8 letters (A, S, D, E, F, M, O, R). Also, all the words are of equal length, 5 letters each. Thus, data for the problem are words such as (tables →)

What are the states?

What are the emissions?

What are the model parameters?

What should we do?

Training Set

adore	dream	fores	oread
aeros	drome	forme	orfes
afore	eards	forms	rased
afros	fader	frame	reads
armed	fades	fremd	reams
arose	fados	froes	redos
arsed	famed	madre	resod
dames	fames	mares	roads
dares	fards	marse	roams
dears	fared	mased	rodes
demos	fares	maser	romas
derma	farms	meads	rosed
derms	faros	moder	safed
deros	farse	modes	safer
doers	fears	moers	sared
domes	feods	morae	sarod
dorms	ferms	moras	smear
dorsa	foams	mores	smore
dorse	fomes	morse	soare
doser	foram	mosed	sofar
drams	fords	oared	sorda
		omers	sored

Test Set

morrs
soffr
foaes
formm
sdfar
doese
safeo
rosmd
mafes
fromd
sosar
frerd
froms
drems
oroes

Example – *Spell Checker* – BREAKING IT DOWN

What are the states?

[A, S, D, E, F, M, O, R]

What are the emissions?

[A, S, D, E, F, M, O, R]

What are the model parameters?

A = matrix 8states x 8states

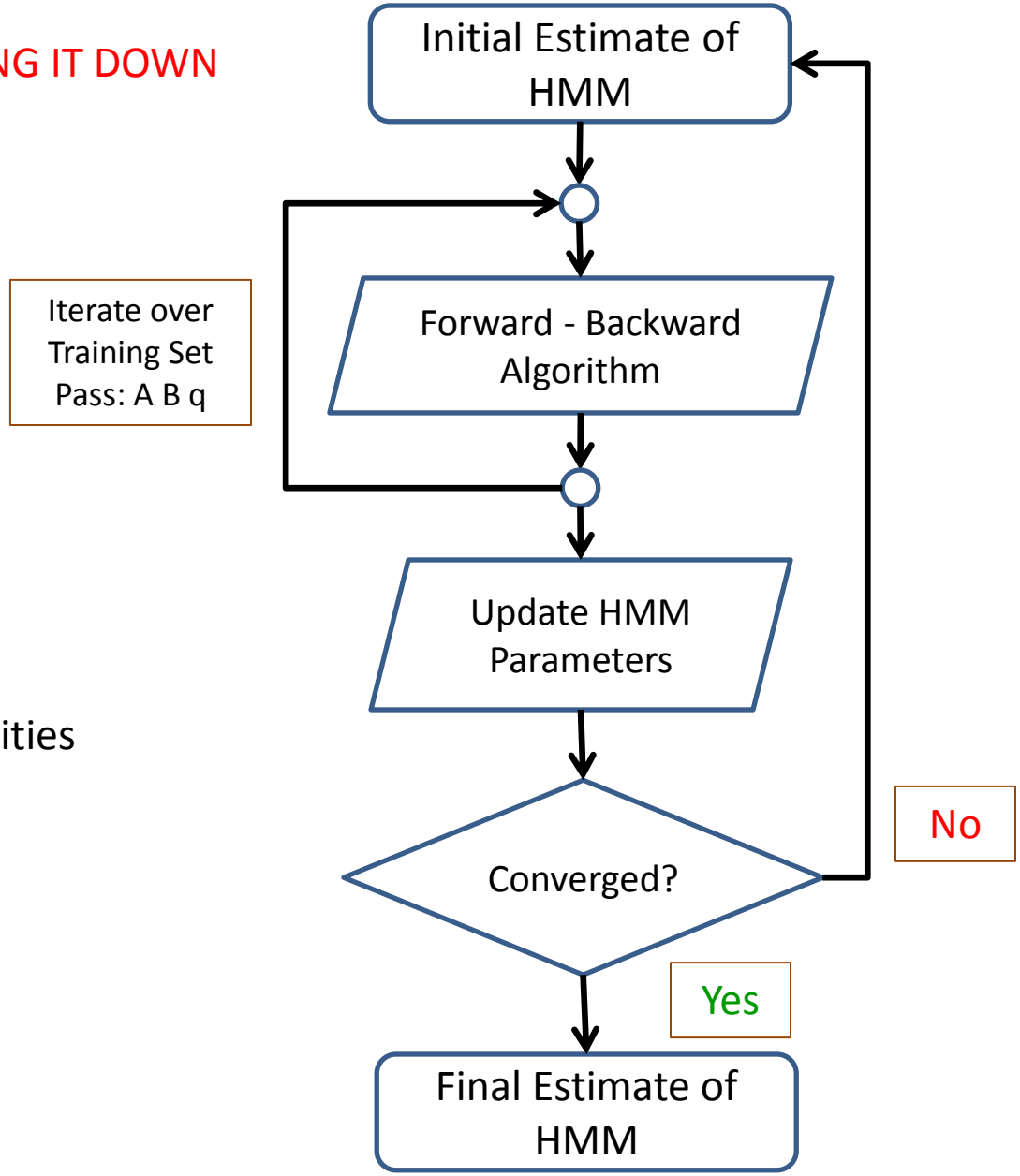
B = matrix 8states x 8emissions

q = matrix 8states x 1 initial probabilities

What should we do?

Training = Baum-Welch Algorithm

Spell Checking = Viterbi Algorithm



Example – *Speech Recognition*

Given a fixed set of vocabulary you should build a speech recognizer that can identify the words in the vocabulary when uttered. The training set and the test set need to be constructed by the developer.

To make the problem easier, we will work with a very limited set of words, say the digits D = [1, 2, 3, 4, 5, 6, 7, 8, 9, 0]. Some background in *speech signal processing* is required.

Keywords: [Discrete Fourier Transform (FFT), Magnitude and Phase Spectra, Spectrogram, DCT, Cepstrum, Cepstrogram]

What are the states?

What are the emissions?

What are the model parameters?

What should we do?

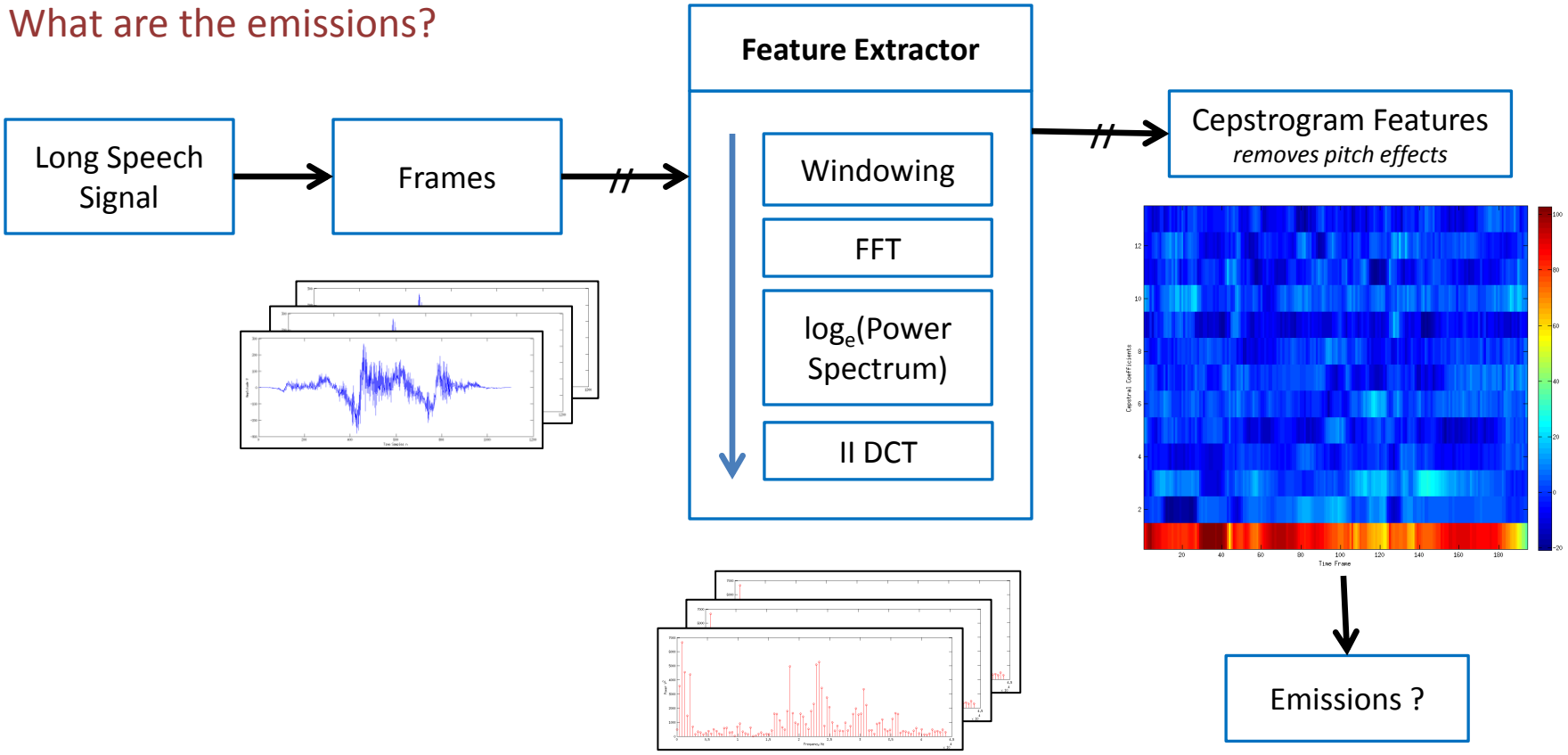
Example – *Speech Recognition* – BREAKING IT DOWN

Training Set = - e.g. 20 utterance recordings **each** of “zero”, “one”… “nine”,

- same or different speakers.

Test Set = Novel utterances of random digits.

What are the emissions?

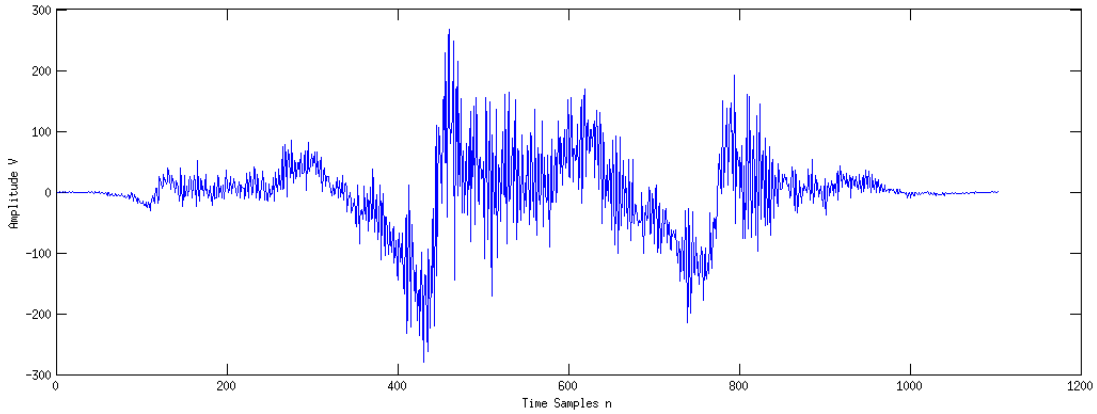


Example – *Speech Recognition* – BREAKING IT DOWN

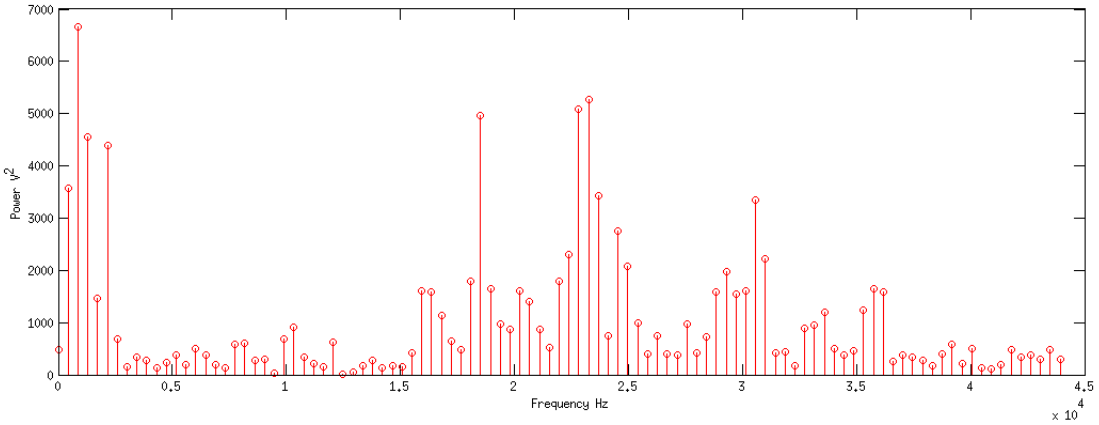
What are the emissions?

ZOOM IN VIEW

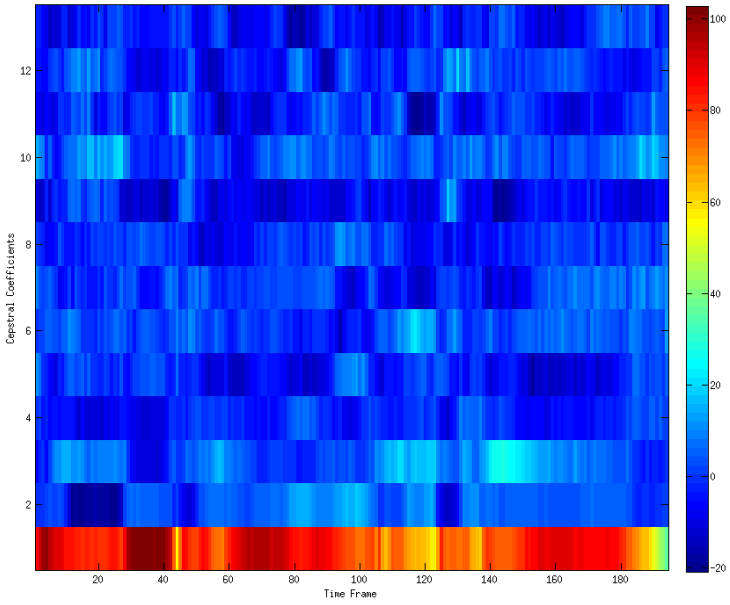
Time Signal - Frame



Frequency representation of above Time Signal - Frame

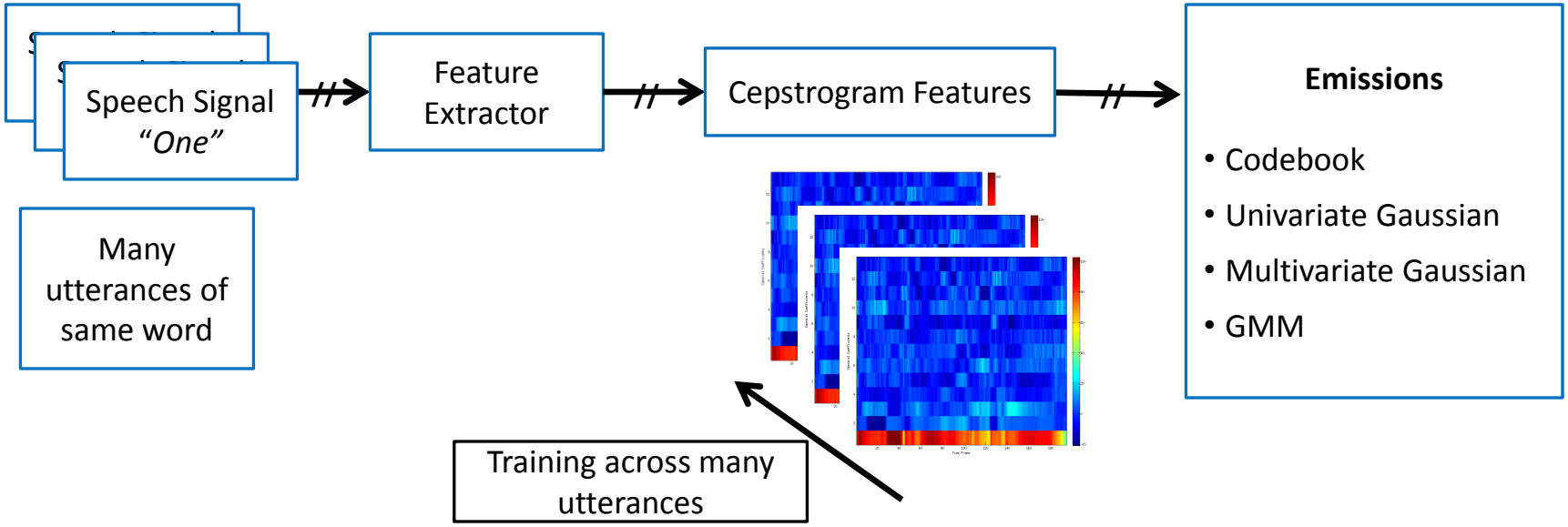


Cepstral Representation of ENTIRE Time Signal



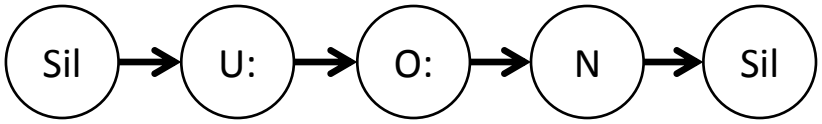
Example – *Speech Recognition* – BREAKING IT DOWN

What are the emissions?



What are the states?

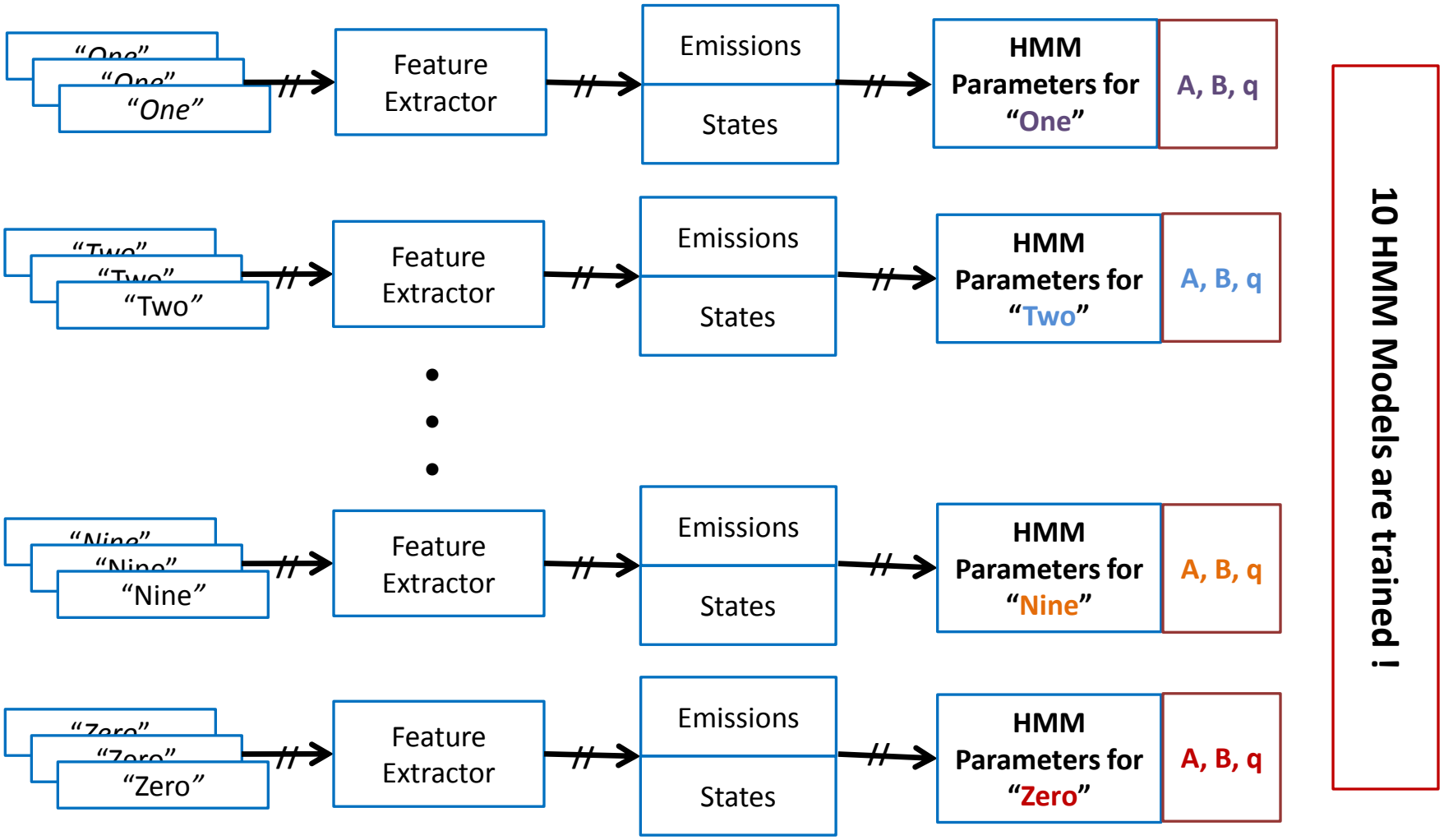
- Time frames (very unlikely)
- Phonemes
- Sub – Phonemes
- Can even be learnt…



Phonemes for "One"

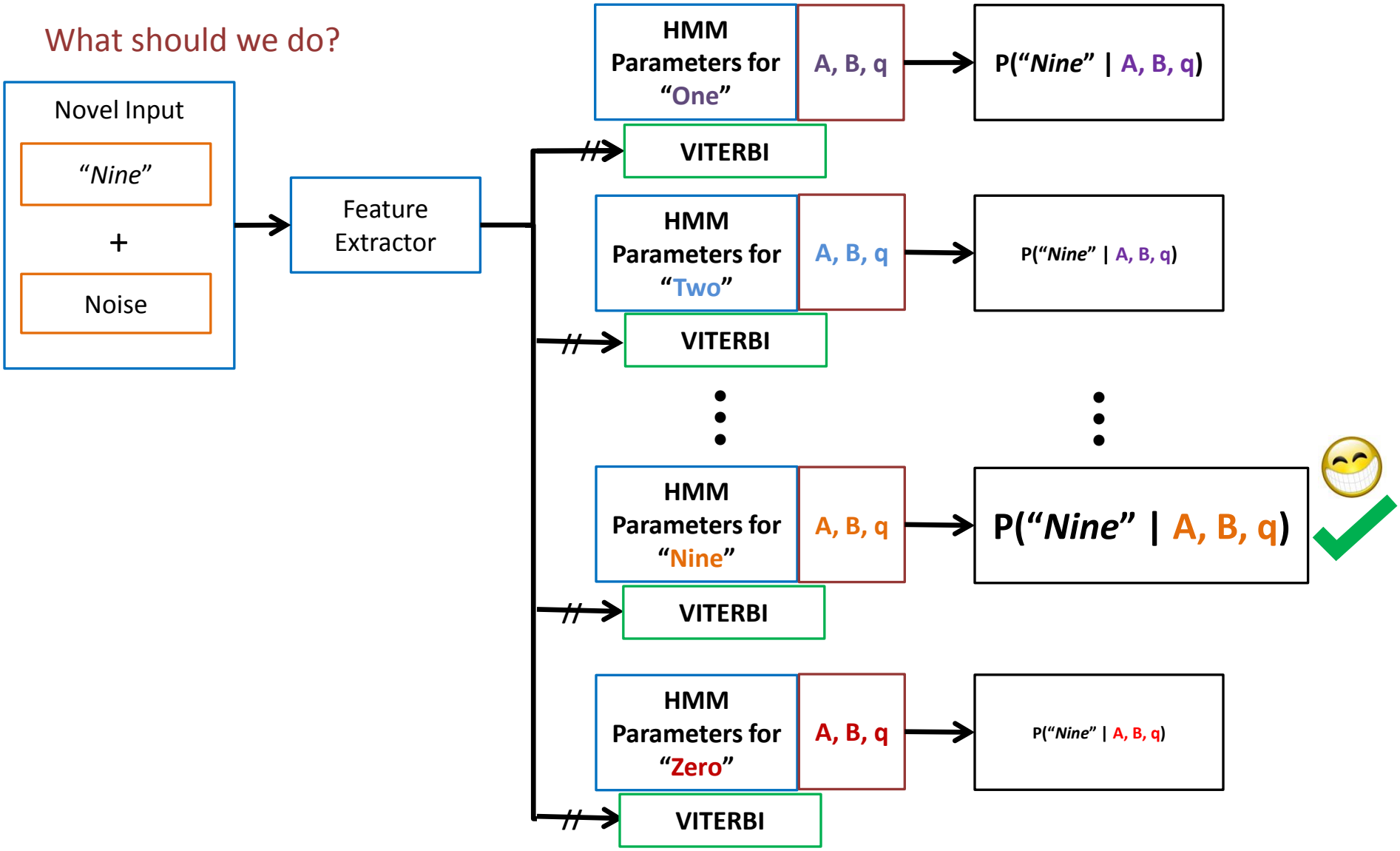
Example – *Speech Recognition* – BREAKING IT DOWN

What are the model parameters?



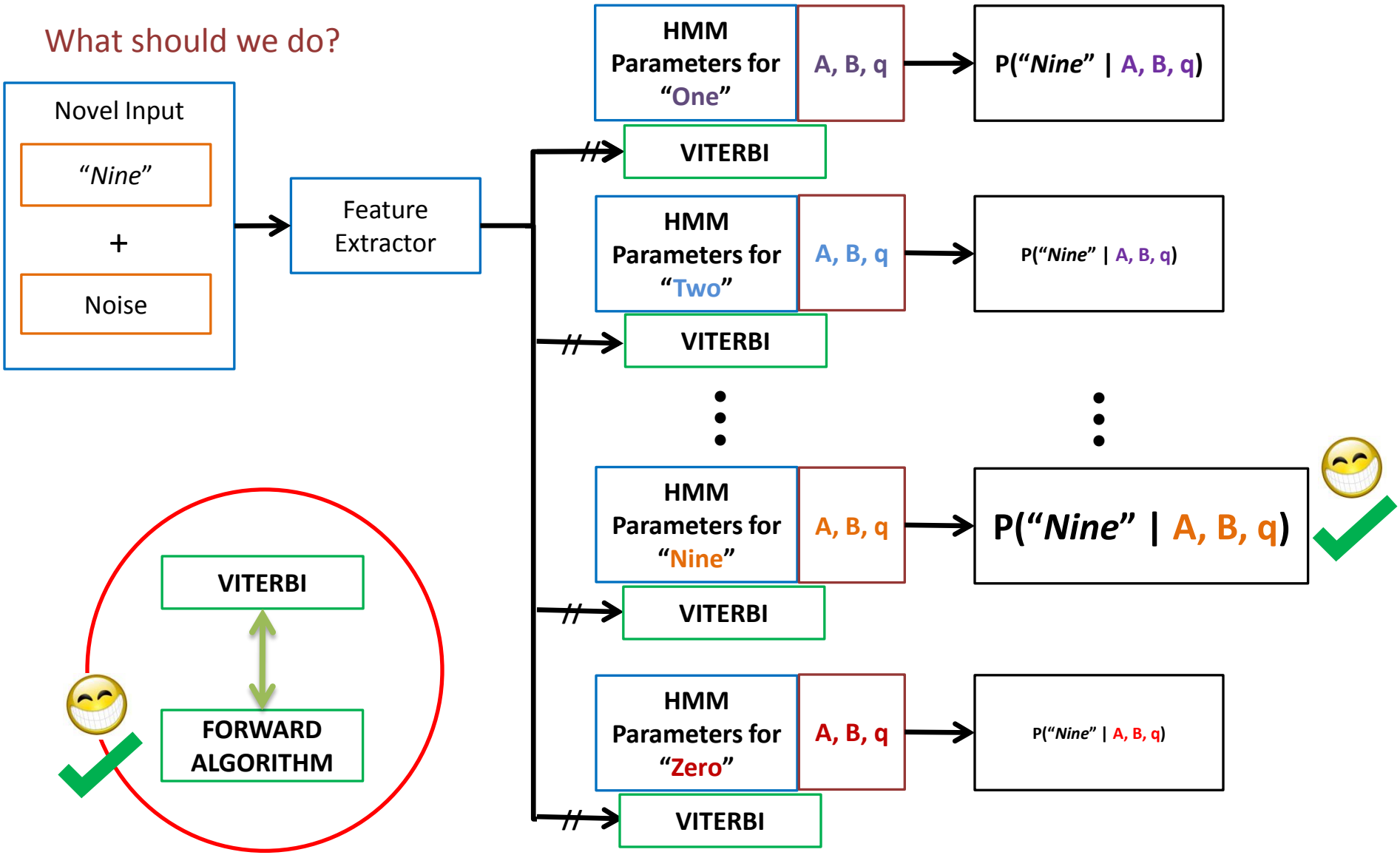
Example – *Speech Recognition* – BREAKING IT DOWN

What should we do?



Example – *Speech Recognition* – BREAKING IT DOWN

What should we do?



Example – *Music Recognition*

Given a set of songs / sounds / tunes, you should build a recognizer that can identify the sounds when hummed or whistled. The training set and the test set need to be constructed by the developer.

Some background in *music signal processing* is required.

Keywords: [Discrete Fourier Transform (FFT), Magnitude and Phase Spectra, Pitch, Tones, Loudness]

What are the states?

What are the emissions?

What are the model parameters?

What should we do?

Example – *Music Recognition*

Points to note:

- Pitch is important – dominant pitch (So, not cepstra, but…)
- Recognition of silence is important = Rhythm
- Transients are important = Beats
- Note durations
- Loudness, Timbre
- Many more…

Example – *Character Recognition*

Given a set of characters / symbols you should build a character recognizer that can identify the letters drawn in a fixed size window of say *MS Paint*. The training set and the test set need to be constructed by the developer.

To make the problem easier, we will work with a very limited set of characters / symbols, say the digits Set = [A, B, C, D, E, F, G, H, I, J]. Minimal background in *digital image signal processing* is required.

Keywords: [Pixels, Resolution, Cartesian Positions, HOG features]

What are the states?

What are the emissions?

What are the model parameters?

What should we do?

Example – *Character Recognition*

Points to note:

- Normalizing is **vital**, in size, shape, orientation and color
- Smoothing or noise removal
- States = parts of curves, pixel filled/not-filled …
- PREDICTING!!! – this means:

  Viterbi Algorithm + Next likely emissions or (states and emissions)
- Many more…

Example – *Duck Hunt*

AI Course Portal – Duck Hunt

What are the states?

What are the emissions?

What are the model parameters?

What should we do?

- For Shooting – which problem?
- For Guessing – which problem?

QUESTIONS?