

KFUPM
Information & Computer Science department
ICS 482 - Natural Language Processing

Online Arabic Handwriting Recognition

Done by

Fadi Biadsy
Jihad El-Sana
Nizar Habash

Presented by

Abdul-Rahman Daud

Outline

- Introduction
- 1. Characteristics of the Arabic Script (problems)
- 2. Current Solutions
- 3. Better Solution (using HHM)
- 4. Hidden Markov Model
- Conclusion

Introduction



- Best means of human-computer interfacing.
- Forms smaller than the traditional
- computer use reaches a larger number

Introduction

- Speech
 - More People

- Handwriting
 - Performance
 - Privacy

- Handwriting Categories
 - Online
 - offline

Characteristics of the Arabic Script

- Cursive:
 - Arabic is written in a cursive.
 - style from right to left.

- Most letters are written in four

ع	ـ	ع	ع
---	---	---	---

Characteristics of the Arabic Script

- Dots



- Delayed strokes
 - creating new letters



Current Solutions

- Most of them Offline Handwriting recognition
 - Strokes are ignored.
- Need for effective Online handwriting recognition

Best Solution

- Based on HMM
- Hidden Markov model
- Regular expression (state machines)
- Our recognition framework uses discrete HMMs to represent each letter shape.

HMM

- (HMM) is a statistical model in which the system being modeled is assumed to be a *Markov process* with unknown parameters, and the challenge is to determine the hidden parameters from the observable parameters.
- The extracted model parameters can then be used to perform further analysis

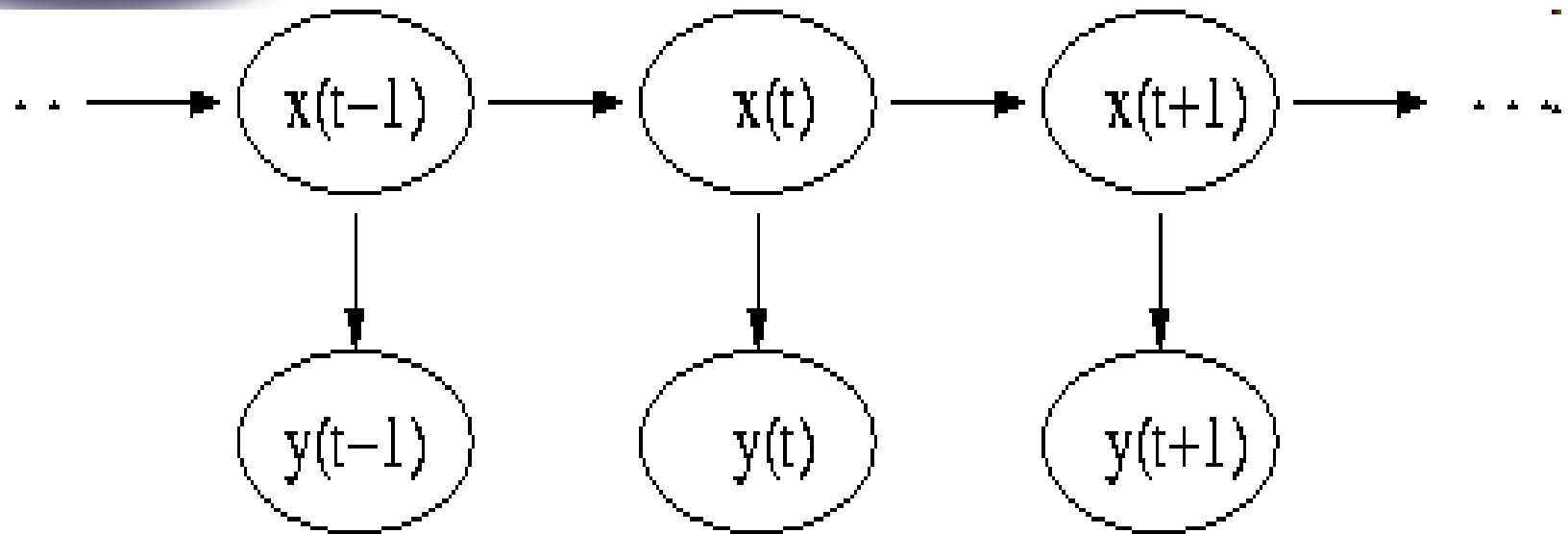
HMM

Markov process

future states of the process, given the present state and all past states, depends only upon the present state and not on any past states, i.e. it is conditionally independent of the past states

HMM is used for many Pattern recognition problems

HMM



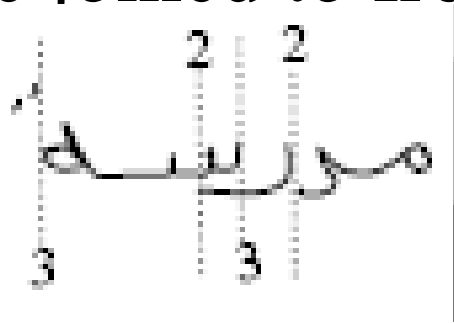
$$Y = y(0), y(1), \dots, y(L-1)$$
$$P(Y) = \sum_X P(Y | X) P(X),$$

HMM

- Given the parameters of the model, find the most likely sequence of hidden states that could have generated a given output sequence. This problem is solved by the *Viterbi algorithm*.

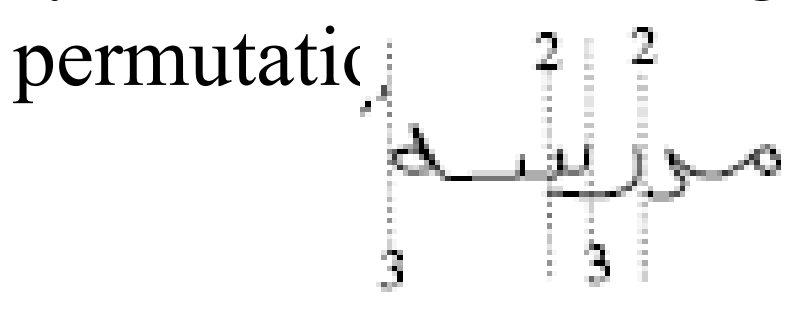
HMM

- In this implementation, each observation y_i in this observation sequence is an integer value
- Letters are joined to form word parts



HMM

- To constrain the space of search, we utilize a dictionary of possible valid words. This ensures better recognition rates compared to systems that can recognize any arbitrary permutation



Conclusion

- This solution introduced an HMM based system with to provide solutions for most of the difficulties inherent in recognizing Arabic script: namely delayed strokes.

