ISSN:0975-9646

Use of Hidden Markov's Model for Handwritten Kannada Character Recognition

Akshayparvatikar[1],Veena G.S[2]

Department of computer science
MSRIT,Bangalore

Abstract: Intense research is being carried in the field of character recognition .There are many ways to recognize the character's of the kannda language (a language which is predominantly spoken in South India).The recognition of kannda character's is difficult due to varying styles shapes, features of The characters .The problem of kannada character recognition is tackled by use of a mathematical model known as Hidden markov's model.

Key Words: offline handwritten character recognition, hidden markov's model

1. INTRODUCTION

As we have progressed in the field of technology, there has been a constant effort to increase the domain of the applications being developed. One such domain in the field of computer science is machine perception (the ability of the computer to interpret the data in a way similar to human's utilizing their senses to relate to world around them .Character recognition is an important aspect of pattern recognition which forms the back bone of machine perception.

Handwritten data can be gathered in two ways.

1. Offline hand written data.
2. Online handwritten data.

Offline handwritten data is acquired by the photograph scanning of the images. Online hand written data is acquired by using stylus pen and a tablet .Recognition of the offline handwriting is more complicated due to noise in acquired image, velocity of writing.

Applications determine which character recognition system has to be used. Processing of bank cherubs, mail sorting etc makes use of offline character recognition system While security domains like author authentication, signature verification make use of offline character recognition system.

2. STAGES

Following are the stages of offline handwritten character recognition using Hidden Markov's model.

2.1 PREPROCESSING

Raw data is subjected to processing which enables one to use the modified data in future stages of character recognition .Preprocessing produces the data which assists the character recognition system to function with greater accuracy.

Following are the objectives of preprocessing

1. Noise reduction
2. Normalization
3. Thinning

There are varying styles of writing; hence there is a large variation in the characters, the writing style of a person may change with time. So a large dataset of the characters has to be collected. Each sample has to be scanned at 60*60 to fit into a window .A thinning algorithm is applied on scanned images to extract the features of the characters.

Thinning enables a programmer to represent a character in a form which is helpful for feature extraction. Thinning is applied on binary images produces a binary image. Thinning algorithm should retain the significant features and eliminate the local noise.

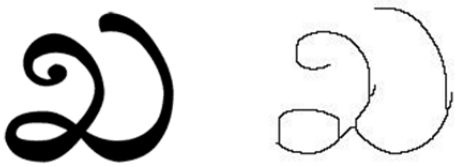


Figure 1.sample of a thinned character

2.2 FEATURE EXTRACTION

In this stage a character is represented as a feature vector .Set of features are extracted from a thinned character. It maximizes the character recognition rate with least number of elements.

Various features of character are

1. Statastical features
2. Structural features

In this paper structural features are emphasized. Structural features are based on topological and geometrical properties of the character which include loops, cross points, junction point's aspect ratio, curves etc. Feature Extraction is an important stage of character recognition. Features vary from character to character and no two characters have same features.

End point: An end point is a pixel pi with T (pi) =1

Junction point :A pixel pi with T (pi) =3.

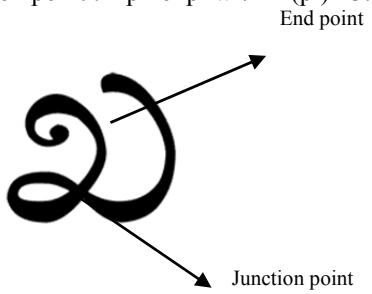


Figure2. Typical features of a character

3. CLASSIFIER

After representing a character as a feature vector a rule has to be defined to classify character into one of the letters .Various classifier's are available which include

1. Knn
2. Bayesian
3. SVM
4. Neural Network

But the focus is on Hidden markov's model.

A hidden markov model is defined as follows

- Number of states of model n.
- Number of observation symbols in Alphabet M
- A set of transition probabilities
  $A= \{a_{ij}\}$
- $A_{ij}=p\{q_t=i\},1<=I,j<=N.$

Where q_t is the current state.

A probability distribution in every state

$B=\{b_j,(k)\}$

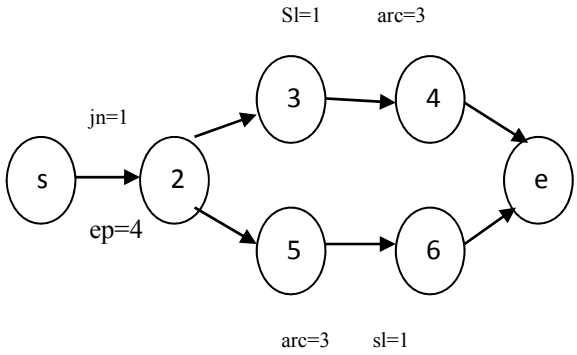
$B_j(k)=p\{a_t=v_k|q_t=j\},1<=j<=N, 1<=k<=M.$

Where Vk denotes kth observation symbol in all alphabets and a current parameter vector.

The initial state distribution

$\mu=\mu_i$ where $\mu_i=p\{q_j=i\},1<=i<=N$

Therefore we can use T= (A,B,μ) as a compact notation for HMM.



(Hidden states) Figure 3: a diagram depicting Hmm for character 'ga'.

4.CONCLUSION

Hidden markov's model is proven as a powerful tool in modeling the speech and real world signal's .It also provides excellent properties for modeling the characters .The model uses features such as end points, loops, junction points etc to get Hmm model. Finding more additional hidden features of the characters can help to improve the recognition rate.

REFERENCES

[1]. B.S.Sarita,S.Hemanth "*An Efficient Hidden Markov Model for Offline Handwritten Numeral Recognition*".

[2] Yu Wang, Xueye Wei, Lei Han, Xiaojin WuA "*Novel Character Recognition Algorithm Based on Hidden Markov Models*".

[3] Aparna.A, Prof.I.Muthumani "*Optical Character Recognition for Hand written Cursive English characters".*

[4] S. K. Parui, K. Guin, U. Bhattacharya and B. B. Chaudhuri "*Online Handwritten Bangla Character Recognition Using HMM*".

[5] Karishma rastogi,vedant tyagi"Implimentation of character recognition using Hidden markov's model"