

The Summarize Of Improved HMM Model

Fangfang Shi, Xianyi Cheng, Xiang Chen
 School of Computer Science and technology, Nantong University
 Nantong Jiangsu, 226019, China
 E-mail:xycheng@ntu.edu.cn

Abstract-The hidden markov model is a kind of important probability model of series data processing and statistical learning and it has been successfully applied in many engineering tasks. This paper introduces the basic principle of hidden markov model firstly, and then discusses the limitations of hidden markov model, as well as the improved hidden markov model which is put forward to solve these problems.

Keywords- HMM, Markov family model, DDBHMM, genetic algorithm, EHMM

I. INTRODUCTION

As a kind of the statistical analysis Model, Hidden Markov Model was founded in the 1970s. It was spread, developed and applied to the modeling of the acoustic signal successfully in the 1980s. It is still considered as one of the most successful methods to realize the quick and accurate speech recognition system so far. As an important direction of the signal processing, HMM is widely used in image processing, pattern recognition, artificial synthesis of speech, biological signal processing and other areas of study, and obtains many important achievements^[1]. In recent years, many researchers applied HMM in computer vision, volatility analysis of financial market and economic budget and other emerging fields. Therefore, it is very significant to have a further study of the various new HMM models and its properties.

II. THE BASIC PRINCIPLE OF HMM MODEL

A. The Description Of HMM

In the markov model, each state represents an observable event. This limits its scope of using. Because the actual problem is more complex, the observed event has the relation of one-to-one correspondence with the random function of state, but not the state. Thus we put forward Hidden Markov Model. It is a double random process. One of which is the markov chain. It is a basic random process. It describes the transferring of state and this process can not be observed (hidden). The other random process is the random function of the hidden state transferring process. It describes the corresponding relationship of statistics between states and observed values. Unlike the markov model, the observed values and states have not the relation of one-to-one correspondence. Therefore, we can only see observation values from the point of observers. We can not see the states directly, but perceive the existence of the state

and its characteristics by a random process (observation values series). Thus we call it hidden markov model.

Generally, HMM model is defined as: $\lambda = (N, M, \pi, A, B)$ (or abbreviations: $\lambda = (\pi, A, B)$ ^[2].

N: the number of markov chain states. Recorded N status as $\theta_1, \theta_2, \dots, \theta_N$, Recorded the state which markov in as q_t at the time of t. Apparently, $q_t \in (\theta_1, \theta_2, \dots, \theta_N)$.

M: the possible numbers of observation value that each state corresponds to. Recorded M observation values as v_1, v_2, \dots, v_M . Recorded the observation value that has been observed as o_t at the time of t, including $o_t \in (v_1, v_2, \dots, v_M)$.

π : Probability vector of initial state, $\pi = (\pi_1, \dots, \pi_N)$, including $\pi_i = P(q_1 = \theta_i)$, $1 \leq i \leq N$.

A: State transferring probability matrix, $A = (a_{ij})_N \times N$, including $a_{ij} = P(q_{t+1} = \theta_j | q_t = \theta_i)$, $1 \leq i, j \leq N$.

B: Observation value probability matrix, $B = (b_{jh})_N \times M$, $b_{jh} = P(o_t = v_k | q_t = \theta_j)$, $1 \leq j \leq N$, $1 \leq k \leq M$.

B. The Three Basic Problems Of HMM

(1) Valuation problem

Giving the model $\lambda = (A, B, \pi)$ and the series of observation events $O = (o_1, o_2, \dots, o_t)$, solve the probability $P(O | \lambda)$ of observation series O that is occurred in λ model. Generally this problem is solved by the forward - backward algorithm.

(2) Decoding problem

Giving the model $\lambda = (A, B, \pi)$ and the series of observation events $O = (o_1, o_2, \dots, o_t)$, solve the best state series $q = (q_1, q_2, \dots, q_t)$ that observation series O corresponds to. Generally this problem is solved by the Viterbi algorithm.

(3) Learning problems

Under the conditions of giving sample set, adjust the parameters of $\lambda = (A, B, \pi)$ in order to make $P(O | \lambda)$ maximum. Generally this problem is solved by the Baum-Welch algorithm.

III. THE IMPROVED HMM MODEL

A. Markov Family Model

Markov family model make $X = \{X_1, X_2, \dots, X_m\}$ represent m-dimensional random vector, in which the value of X_i ($1 \leq i \leq m$) comes from the finite state set S_i ($1 \leq i \leq m$). X can constitute a markov family model when it meets the following conditions:

(1) Every X_i ($1 \leq i \leq m$) component is a markov chain of n_i -moment:

$$P(\mathcal{X}_{i,t} | \mathcal{X}_{i,1}, \dots, \mathcal{X}_{i,t-1}) = P(\mathcal{X}_{i,t} | \mathcal{X}_{i,t-n_i+1}, \dots, \mathcal{X}_{i,t-1})$$

(2) The probability that one state appears of the component at the time of t is only related with these states of the same component before t time and these states of other component at t time:

$$P(\mathcal{X}_{i,t} | \mathcal{X}_{1,1}, \dots, \mathcal{X}_{1,t}, \dots, \mathcal{X}_{i,1}, \dots, \mathcal{X}_{i,t-1}, \dots, \mathcal{X}_{m,1}, \dots, \mathcal{X}_{m,t}) =$$

$$P(\mathcal{X}_{i,t} | \mathcal{X}_{i,t-n_i+1}, \dots, \mathcal{X}_{i,t-1}, \mathcal{X}_{1,t}, \dots, \mathcal{X}_{i-1,t}, \mathcal{X}_{i+1,t}, \dots, \mathcal{X}_{m,t})$$

(3) Conditional independence assumption:

$$P(\mathcal{X}_{i,t-n_i+1}, \dots, \mathcal{X}_{i,t-1}, \mathcal{X}_{1,t}, \dots, \mathcal{X}_{i-1,t}, \mathcal{X}_{i+1,t}, \dots, \mathcal{X}_{m,t} | \mathcal{X}_{i,t}) =$$

$$P(\mathcal{X}_{i,t-n_i+1}, \dots, \mathcal{X}_{i,t-1} | \mathcal{X}_{i,t}) \cdot P(\mathcal{X}_{1,t} | \mathcal{X}_{i,t}) \dots P(\mathcal{X}_{m,t} | \mathcal{X}_{i,t})$$

We can know that markov family model is a multiple random process from the conditions (1), so we can recognize the hidden markov model as a special situation of markov family model. We know that hidden markov model uses the independence assumption, and we also know that markov family model uses the condition independence assumption from condition (3). In a word, markov family model not only includes the hidden markov model, but also substitutes the independence assumption of hidden markov model for the conditional independence assumption. Comparing with hidden markov model, markov family model is more suitable for the actual physical process of speech and language.

B. DDBHMM Model

Classic HMM speech recognition model depends on the strong assumption, including the assumption that the model state transferring has the homogeneous Markov property. Under the condition of this assumption, the resident probability $a_{i,i}$ of state i is a constant in classical HMM model. Thus after the system get into the state i, the continuous resident time (segment length) obey geometric distribution in this stage. There are still a lot of shortages to describe the characteristics of segment length by this geometric distribution. In order to solve this shortage, DDBHMM model [3] appears.

The relationship between (“the across transferring from state i to state i + m (m > 1)” and “slip from state i + 1, state i + 2,..... to state i + m - 1 (the length of resident time are zero)”) is the essential difference between DDBHMM model and classical HMM model. Classical HMM recognizes the two events as independent completely and DDBHMM model recognized the two events as related closely. In fact, the two events happened simultaneously in the production process of speech. From the point of information theory, DDBHMM model make up for the losses that the HMM model bring to the performance of recognition system.

Avoiding the assumption of the classic HMM model that the state transferring has homogeneous Markov property, DDBHMM model uses new resident probability to describe the characteristics of segment length. It solves the shortage of the classical HMM model in describing segment length. DDBHMM model have reasonable description to both the generated series of speech signal and the distribution of the actual speech segment length. it not only resolves the shortage of the classical HMM model in describing segment features, but also lays foundation for the

description and using of the speech signal time length, speed, time interrupter , the signal correlation of the phonetic features and other important characteristics.

C. HMM Model That Is Based On Genetic Algorithm

Genetic algorithm [4] simulates the evolution phenomenon of the survival of the fittest in nature. It codes possible solution into vector - chromosome, maps the searching space to the genetic space. And then it chooses the best chromosome which is the best solution through continuous calculation of each chromosome's fitness value. The whole process of Genetic algorithm is mainly composed of three main operators and operands. It establishes initial group randomly at first, and then calculates the individual fitness, and then generates new group by using reproduction, crossover, and mutation operators according to genetic rules. This process executes circularly and generates new group constantly until its end. Finally choose the best individual as the result of genetic algorithm.

The training method that the traditional hidden markov used is Baum - Welch algorithm. Although this algorithm is widely used, it also has its own disadvantages. The final solution it gets is closely related with the selection of the initial value, that is Baum - Welch algorithm is used for the calculation of the local optimal solution, which affect the recognition rate of the system. Basing on this shortage, here we put forward the genetic algorithm to solve the above problem. The genetic algorithm uses the global searching method which is based on the group and it has great probability to obtain the global optimal solution. Thus it makes up for the shortage of Baum - Welch algorithm. The improved HMM model with genetic algorithm uses genetic algorithm to optimize B value. That is combining genetic algorithm with Baum - Welch algorithm in the training process, so as to improve the effect of the overall recognition rate.

D. HMM And SVM Hybrid Model

HMM is a generated model essentially. It describes the same kind of data by using probability distribution theory of the traditional statistics, so its ability to classify is quite limited. In comparison with HMM, SVM is considered as a strong classification method. It is a discriminant model essentially. It constructs classifier by looking for optimal separating hyper-plane between classes in the feature space. It still has good ability of classification under the condition of limited training samples. But it can't deal with feature vector of variable length. Thus the hybrid model of the two kinds of classifiers appears [5].

For the speech recognition problem which is easy to confuse, we can use the two-stage recognition framework which is combined by HMM and SVM. We use the HMM identifier for the preliminary identification and use the SVM identifier for the secondary recognition. In the first recognition, we mainly deal with variable length feature vector by the HMM model identifier. In the second recognition, we mainly deal with the characteristic vector of fixed length that the first identifier produces by the SVM

identifier. The first identification provides a characteristic vector of the fixed length for the second identification, so it solves the shortage that the SVM can not deal directly with feature vector of the variable length. The second recognition looks for optimal separating hyper-plane between classes in the feature space by SVM, so it helps to identify the speech that the first classifier failed to identify. This combining improves the effect of the single classifier recognition greatly.

E. HMM And ANN Hybrid Model

HMM can realize the statistical modeling of speech signal, can deal with the dynamic changes of speech series better, but its ability of pattern recognition is poor and the identification and training algorithm it uses depend on strong hypothesis. ANN has not only strong ability of pattern recognition and parallel processing, but also it has the ability of learning specific knowledge without the advanced hypothesis. However, it can't handle speech characteristics series of the dynamic change. So we can build a classifier combined with HMM and ANN. The most currently used classifier are combined by HMM and MLP (multi layer perception). It mainly includes four different ways of combination^[6].

(1) Realize the time warping processing of MLP classifier by using HMM.

The MLP can only deal with the speech features series of fixed length. In order to make the MLP be applied in speech recognition correctly, we must do something with the speech length of the dynamic change. We can use the bending warping processing of HMM to make the warping reflect the change of speech in the time domain better. This method mainly uses HMM for its bending warping processing firstly and then sends the speech that has been treated to MLP processing.

(2) Calculate the observation probability of HMM by using MLP

The recognition and training algorithms of HMM depend on strong hypothesis, and its adaptability and robustness is poor. MLP can only deal with the fixed input mode and it is hard to be popularized to the continuous speech of the large vocabulary. Here we add the HMM to the nervous predictor (NP) to calculate the observation probability of each state. We mainly use HMM to make up for the shortage that the MLP can only deal with the fixed input mode, and use the MLP to make up for the shortage that the HMM have a priori assumptions to speech state distribution at the same time.

(3) Realize the HMM recognizer algorithm by using ANN

One of the shortages of HMM which can not be ignored is that the recognition system is difficult for hardware realization. ANN has the ability of parallelism and is easy for hardware realization, so it can make up for the disadvantages of HMM about the hardware realization. Here the method is mainly realize the HMM algorithm by using the structure of ANN. Viterbi network which is put forward by Lippmann and Gold uses this method. In the process of its classification, Viterbi network keeps all the

structure of the traditional HMM. Comparing with HMM, it uses the same feature space, state transferring, state number and it also assumes that the speech obeys the Gaussian distribution. So they have the similar performance, but the speed of Viterbi network is much faster than HMM.

(4) HMM net

HMM net is put forward by Niles and Siverman. It is a circular ANN. Put HMM into ANN structure. Its observation probability and transferring probability are realized by network. But it still has the same training processing. So it can realize the structure and algorithm of the traditional HMM. The traditional HMM require that every model probability is negative and so on, while in HMM net these restrictions can be relaxed.

F. Embedded Hidden Markov Model

Embedded hidden markov model (EHMM) is extended on the basement of the one-dimensional HMM. In the EHMM, each state of one-dimensional HMM will be served as a new HMM. That is to say, EHMM contains a super state set and is accompanied with an embedded state set. Super state is used for 2-dimensional data modeling in one direction and embedded states are used for data modeling in another direction. In a word, EHMM is a HMM model which has double data characteristics^[7].

This model and the real 2-dimensional HMM are different. As different super states in this model are not allowed to transform from each other, so the EHMM is an expansion of one-dimensional HMM. Comparing with other structure, EHMM has the following advantages:

(1) From the perspective of training and recognition, the complexity of calculating is reduced greatly.

(2) It can get better estimated value of parameters from initial model.

(3) The 2-dimensional structure of data can be reserved naturally under the condition of not using other complex frame. So it will not increase the complexity of the model.

Basing on the one-dimensional HMM structure, EHMM is mainly composed of embedding another one-dimensional HMM structure in each state of HMM. Therefore, the parameters, training and recognition algorithms of EHMM model can be got by the extension of one-dimensional HMM algorithm.

G. Embedded Hidden Markov Model

Embedded hidden markov model (EHMM) is extended on the basement of the one-dimensional HMM. In the EHMM, each state of one-dimensional HMM will be served as a new HMM. That is to say, EHMM contains a super state set and is accompanied with an embedded state set. Super state is used for 2-dimensional data modeling in one direction and embedded state are used for data modeling in the another direction. In a word, EHMM is a HMM model which has double data characteristics [11].

This model and the real 2-dimensional HMM are different. As different super states in this model are not allowed to transform from each other, so the EHMM is an expansion of one-dimensional HMM. Comparing with other structure, EHMM has the following advantages:

(1) From the perspective of training and recognition, the complexity of calculating is reduced greatly.

(2) It can get better estimated value of parameters from initial model.

(3) The 2-dimensional structure of data can be reserved naturally under the condition of not using other complex frame. So it will not increase the complexity of the model.

Basing on the one-dimensional HMM structure, EHMM is mainly composed of embedding another one-dimensional HMM structure in each state of HMM. Therefore, the parameters, training and recognition algorithm of EHMM model can be got by the extension of one-dimensional HMM algorithm.

IV. CONCLUDING REMARKS

This paper introduces several kinds of improved models of markov. Markov family model makes it more suitable for speech and language physical process by using multiple random process; DDBHMM model solves the shortage of describing segment features in classical HMM model; genetic algorithm solves the problem that Baum - Welch algorithm always get the local optimum solution; HMM and SVM hybrid model and HMM and ANN hybrid model both use one classifier's advantages to make up for another classifier's disadvantages so as to achieve the overall improvement; Embedded hidden markov model (EHMM) has the low computational complexity and can get the better estimated value of the initial model parameters. The markov

family model, DDBHMM model and genetic algorithm improved HMM model on the base of a single HMM model in some disadvantages. HMM and SVM hybrid model and HMM and ANN hybrid model are both based on two kinds of classifiers. They make the speech recognition be improved in different degrees on the base of different disadvantages in HMM model.

ACKNOWLEDGEMENT

This work is supported by Natural Science Foundation of China National (No: 61202006)

REFERENCES

- [1] Bilmes JA. What HMMs can do. IEICE TRANSACTIONS on Information and Systems, 2006, 89(3):1-24.
- [2] Ji-qing Han, zhang lei, Tie-ran Zheng. Speech signal processing[M]. Beijing: tsinghua university press, 2004.
- [3] Zuo-ying Wang, Xiao Xi. HMM speech recognition model which is based on segment length distribution. Journal of electronics, 2004,32 (1) :46-48
- [4] Xu Li, Rui-hua Kang. HMM parameter estimation which is based on genetic algorithm [J]. Journal of hubei university of technology. 2006,4: 68-71
- [5] Huan-liang Wang, Ji-qing Han, Hai-feng Li. The confusingly Chinese speech recognition that is based on HMM/SVM two stage structure. 2006:19(5): 578-584
- [6] Wei-ying Li, Ke-chu Yi, Hu Zheng. The mixed network combined neural network with HMM and its application in speech recognition [J]. Journal of electronics, 1994, 22(10):73- 79.
- [7] Xu Feng. face expression recognition which is based on embedded hidden markov model (EHMM) [D]. Master thesis of dalian university of technology, 2006