THE POSSIBILITY FOR ACQUISITION OF STATISTICAL NETWORK GRAMMAR USING ERGODIC HMM

Jin'ichi Murakami[1] Hiroki Yamatomo[1] Shigeki Sagayama[2]

[1] *ATR Interpreting Telecommunications Research Labs., 2-2 Hikaridai Seika-cho Soraku-gun Kyoto 619-02 Japan*
[2] *NTT Human Interface Research Labs., 3-9-11 Midori-cho Musashino-shi Tokyo 180 Japan*

ABSTRACT

This paper describes the use of a discrete Ergodic Hidden Markov Model HMM for automatic acquisition of a language model from a large amount of text data and discusses the possibility of using the HMM to acquire the concept of POS (part-of-speech). A discrete-output Ergodic HMMs has a similar structure to a stochastic network grammar (SNG), so automatic acquisition of an SNG from a large amount of text data is possible by using an Ergodic HMM and the Baum-Welch training procedure. In this model, the HMM state transition probabilities can be interpreted as the SNG transition probabilities. And the HMM state output probabilities will refrect the distributions of POSs (part-of-speeches). The results of experiments using text data show that a high similarity is found between high output probability words and the POS. This means that an Ergodic HMM has the ability to automatically acquire both an SNG and the concept of POS simultaneously from training text data. Experiments were also performed on sentence speech recognition. In these experiments, an Ergodic HMM outperformed the word bigram grammar for text-open data.

Keywords: *Language modeling, Stochastic network grammar, Ergodic HMM, Part-of-speech tagger*

1. INTRODUCTION

There are two types of natural language modelings for speech recognition. One is the class of deterministic models like a network grammar or a context free grammar[1]. The other is the class of statistical models like a bigram[2] or a trigram[3]. Among the many existing language models for speech recognition, a network grammar is often used because of its simplicity. However, this grammar has a high perplexity which decreases the speech recognition performance. Therefore, a stochastic network grammar (SNG) which adds probability to the network grammar has been studied to resolve this problem.

On the other hand, Hidden Markov Models (HMMs) are popular in acoustic modeling for speech recognition[4]. One of the advantages of HMMs is that they can be automatically trained through the Baum-Welch maximum likelihood estimation procedure using training speech data. Among the various types of HMMs, the all-state-connected model is called an Ergodic HMM.

This Ergodic HMM has a similar structure to an SNG[5]. Therefore, it is possible for an SNG to be obtained automatically by using the Ergodic HMM through the Baum-Welch training procedure. The resulting state transition probabilities of an Ergodic HMM are interpreted as transition probabilities in an SNG and output probabilities of an HMM is interpreted as word output probabilities in an SNG. Usually, word output probabilities in an SNG is depend on POS (part-of-speech). Therefore, word output probabilities in an HMM is grouping and that is essentially very close to the concept of POS.

We investigated the automatic acquisition of an SNG using an Ergodic HMM. And results showed that the word groupings in the HMM state is a striking similarity to POSs. This means that an Ergodic HMM has the ability to automatically acquire both an SNG and the concept of POS simultaneously from training text data. Preliminary results of experiments for sentence speech recognition using an Ergodic HMM are also carried out and show that an Ergodic HMM outperformed the word bigram grammar for text-open data.

2. AUTOMATIC ACQUISITION ON STOCHASTIC NETWORK GRAMMAR USING ERGODIC HMM

2.1. Experimental condition

To test an Ergodic HMM as an SNG, experiments were performed. As training datum, we used the ATR Dialog Database. The example sentences are shown in Table 1 . The learning conditions of the experiments are shown in Table 2 .

Table 1. Example of sentences

• hai moshimoshi
(Hello)
• e-tto sochira dai ichi kai no tsuuyaku deNwa kokusai kaigi no jimu kyoku de sho u ka
(Well, is this the office of the first international conference on interpreting telephony ?)
• hai sou desu (Yes)
• e-tto chotto sono kaigi no koto dene
(Well about this conference ...)
• hai douzo (Yes...)
• e-tto ima temoto ni ano touroku youshi ga aru N desu keredomo
(I have a registration form, now.)

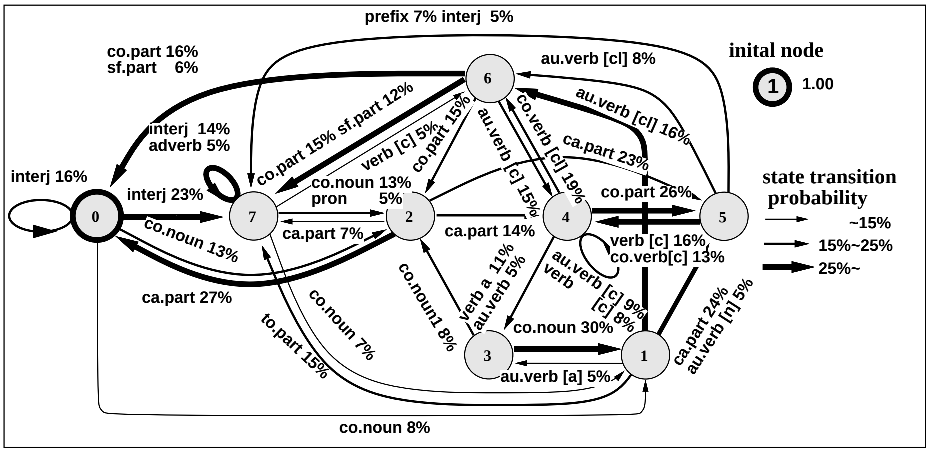

Figure 1. Results of an 8-state Ergodic HMM analysis

Table 2. Learning conditions for Ergodic HMM

type of Hidden Markov model	Mealy type
learning sentence number	8475
total word number	57354 words
number of states	2, 4, 8 and 16
vocabulary	6418
Initial state distribution	equivalent value
Initial transition value	random
Initial symbol value	random
end of HMM learning	under 1% likely hood

2.2. How to analyze the Ergodic HMM

To understand how the HMM learns grammar, we consider the POS. An HMM has the characteristic of clustering. We expect that the same characteristic words in a natural language has the same state in an Ergodic HMM. Therefore, this word grouping is strongly related to the POS.

Therefore, the following method is used to analyze the Ergodic HMM.

1. Change words into POSs.
   (We already know the relationship between words and POSs.)
2. Sum the probabilities of identical POS's.
3. To keep the diagram tidy, show only those POSs whose proportion exceeds 5%.

2.3. Analysis of an 8-state Ergodic HMM

Figure 1 illustrates the analysis of an 8-state Ergodic HMM.

The circles represent the states of an Ergodic HMM. The line widths of the circles show the initial probabilities. The line width of the transition arcs show transition probabilities. The POS codes labeling the transition s in this figure are shown in Table 4 .

The results of this analysis are shown in the following for each POS.

1. Interjection

   Interjections tend to move in the state 0.

2. Common noun

   Common noun mainly move the HMM to state1[2[from0[3[7[.

   In transition of3[, the probability of a formal noun is greater than a common noun. Note that the transition of3[is output for complementizer /no/. The transition of state 0 is mainly output for complementizer /hou/.

3. Conjugate word

   Verbs, auxiliary verbs or adjectives that can be conjugated POS are not grouping by the POS but grouping by the conjunction. For example, an attributive base is output from states4[1[to3[. A continuative base is output from states4[5[6[to4[. A conclusive base of auxiliary verb is output state4[5[6[to4[.

4. Case particle

   Case particles are found in a lot of transitions; each transition is output another word. Examples are given in the following.

   - state2[⇒0[/no/ (94%)
   - state2[⇒7[/kara/ (28%) /de/ (25%)
   - state2[⇒4[/to/ (94%)
   - state2[⇒5[/ni/ (73%) /de/ (18%)
   - state1[⇒5[/wo/ (39%) /ga/ (27%)

2.4. Analysis of a 16-state Ergodic HMM

The results of a 16-state Ergodic HMM are shown in Figure 2 . In this example, a lot of state transitions are shown for only one POS. In addition, the initial state value has two states 2[3[ ). The 16-state Ergodic HMM is more refined than the 8-state Ergodic HMM.

The results of this analysis are shown in the following for each POS.

1. Interjection

   Interjections have outputs from 2[,3[ ) which have a high initial value like the 8-state Ergodic HMM. However, the initial value has two state 2[,3[ ), and does not have a self loop of state 2.

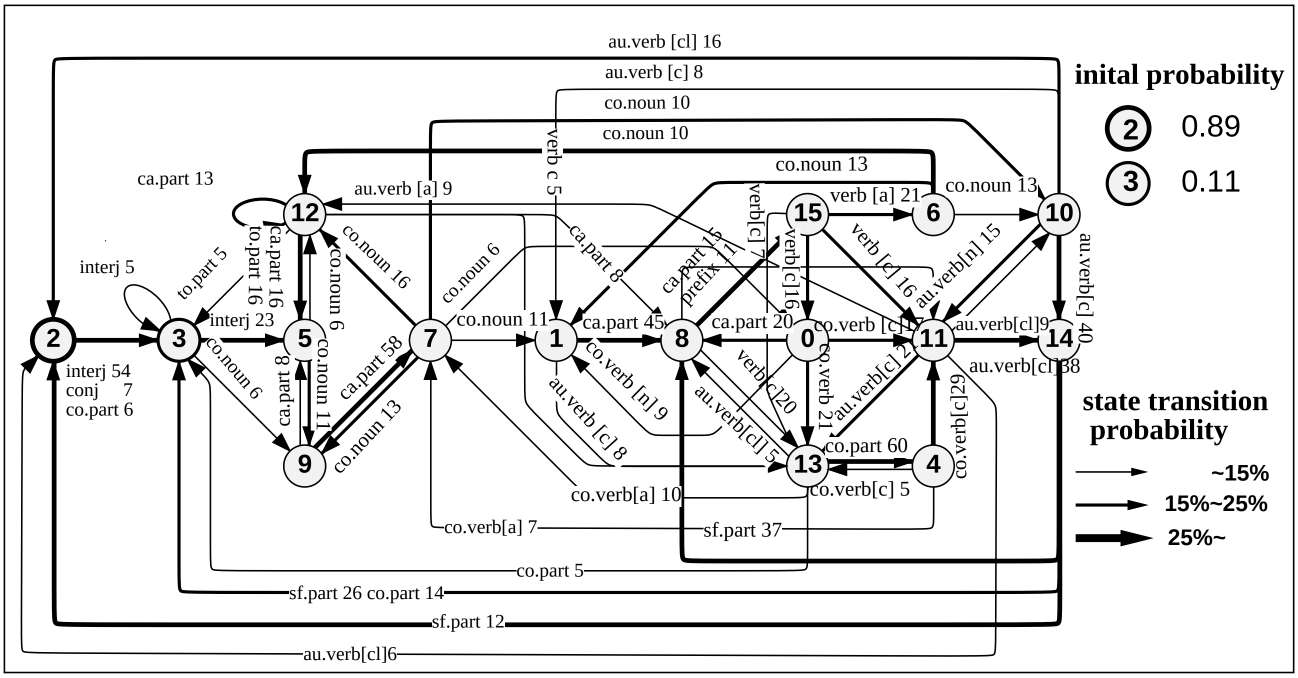

Figure 2. Results of a 16-state Ergodic HMM analysis

2. Common noun

   Common nouns are mainly output from 5[6[7[to 1[9[12[.

   Common nouns are output from states 6[7[. Other nouns are output from state 5[; this state has many kinds of nouns.

3. Conjugate word

   In the 8-state Ergodic HMM, mainly the conjunctions are grouped, However, in the 16-state Ergodic HMM, the POSs are grouped.

   Verbs are output from 8[5[, conjoining verbs are output from states 0[4[, and auxiliary verbs are output from 0[11[13[.

4. Case particle

   Case particles are output from 8[9[12[. Each transition is another word (like the 8-state Ergodic HMM). However, the 16-state Ergodic HMM has more precise clustering. Examples are shown as follows.

   - state 9[⇒7[/no/(93%)
   - state 9[⇒5[/kara/(57%) /de/(23%)
   - state 8[⇒15[/to/ (63%) /o/ (28%)
   - state 0[⇒8[/de/ (86%)
   - state 1[⇒8[/ni/ (87%)
   - state 12[⇒8[/wo/ (75%)

   The case particle of /de/ and /ni/ or /ga/ and /wo/ have the same transitions as in the case of the 8-state Ergodic HMM. However, these particles are divided in the 16-state Ergodic HMM.

2.5. Conclusion for Ergodic HMM as a language model

Although there are 114 different POSs, the experimental results show a high similarity between Ergodic HMM high output probability words and POSs. This means that the Ergodic HMM has an ability to automatically acquire both an SNG and the concept of POS simultaneously from (POS-untagged) training data.

We also observed that, as the number of states increases for an Ergodic HMM, the grouping of words starts from the conjunction and goes to the POS in Japanese.

3. EXPERIMENTS FOR SENTENCE SPEECH RECOGNITION

3.1. Speech recognition

To show the effectiveness of an Ergodic HMM as a language model, we carried out a sentence speech recognition experiment. A standard speech recognition algorithm employed a continuous mixture HMM and beam search, with an Ergodic HMM providing the language information.

3.2. Experimental conditions for sentence speech recognition

For training of a continuous mixture HMM, 2635 word utterances were used. The number of states for the Ergodic HMM as a language model was set to 2, 4, and 8. In addition, the sentence likelihood is the product of the acoustic model likelihood and the with the power of the language model likelihood n, in our experiments we used $n = 32$.

We used 38 sentences as test data. For training the Ergodic HMM, in the text-open experiment we used 4000 sentences of the ATR Dialog Database, whereas in the text-closed experiment we used the same 4000 sentences plus the 38 test sentences. In this way, we can perform the text-closed experiment and a text-open experiment in the same test data.

The experimental conditions are shown in more detail in Table 3 .

Table 3. Experimental conditions for sentence speech recognition

#syllable models	52
Syllable model	4-state 3-loop Gaussian mixture continuous HMM
Learning data	male announcer, 2635 word utterances
Parameter	log power + 16 order LPC-cepstrum + Δlog power + 16 order Δ cepstrum
Test data	same speaker, 38 sentences
Vocabulary	435
Beam width	1024
Speech style	read speech
#state of Ergodic HMM	2,4,8
learning data	4000 sentences, 57354 words

3.3. Experimental results for sentence recognition

Results of experiments are shown in Figure 3 . For comparison, "no grammar" results and word bigram model results are also shown.

In sentence speech recognition experiments, we obtained 39.5% for the 8-state ergodic HMM for text-open data. The the other hand, the word bigram grammar was 36.8%. These results mean that the Ergodic HMM outperformed the word bigram grammar for text-open data. In addition, the Ergodic HMM showed a superior generalization ability to the word bigram grammar, as evidenced by the relatively small difference between the text-open and text-closed performance.

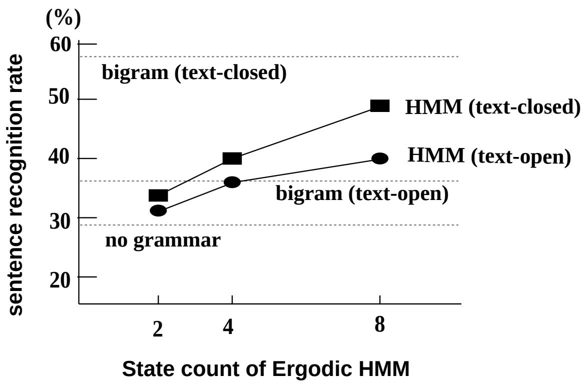

Figure 3. Result of sentence recognition

4. DISCUSSION

4.1. The possibility of automatic language acquirement using Ergodic HMM

In the analysis of Ergodic HMMs, a high similarity was observed between HMM high output probability words and POSs. Therefore, an Ergodic HMM is able to obtain an SNG automatically.

Moreover, as we increase the number of states of an Ergodic HMM like to 4, 8, and 16 states, the word grouping become closer to the POSs. Therefore, we can obtain the grammar in more detail with increasing number of state.

By the way, the current POS is determined linguistically. However, by using an Ergodic HMM, we may be able to automatically obtain a new system of POS under the maximum likelihood.

4.2. Sentence speech recognition for text-closed data

Unfortunately, the performance of the bigram model could not be achieved by the Ergodic HMM for text-closed data. The recognition rate for the bigram model was 57.8%, and the 8-state ergodic HMM was 47.3% for text-closed data. This is probably due to the small number of states of the Ergodic HMM. Therefore, the word perplexity seems to be too high.

However, this problem can be reduced simply by increasing the number of states of an Ergodic HMM. We have not been able to do this yet due the enormous amount of calculation required. In the near future, the performance can be improved for both text-closed data and text-open data by increasing the number of Ergodic HMM states.

5. CONCLUSION

In this paper, we discussed a an discrete Ergodic Hidden Markov Model for a language model, because an Ergodic HMM and a stochastic network grammar have the same structure and same parameters.

We found that after training of Baum-Welch learning for a discrete Ergodic HMM using many text data, the word groupings in the HMM state show a striking similarity to POSs. This means that an Ergodic HMM has the ability to automatically acquire both a stochastic network grammar and the concept of POS simultaneously from (POS-untagged) training data.

In addition, we investigated an Ergodic Hidden Markov model as a language model for sentence speech recognition; we found that it is effective as a language model. Therefore, it may be said that an Ergodic HMM is a general learning language model.

REFERENCES

[1] K.Kita, T.Kawabata, T.Hanazawa, "HMM speech recognition using stochastic language models", ICASSP'90 J. Acoust. Soc. Jpn. (E) 12,3 *pp.99–105* (1991).

[2] K.F.Lee, "Large-vocabulary Speaker Independent Continuous Speech Recognition: The SPHINX System," 15213 CMU-CS-88-148 (April 18, 1988).

[3] F. Jelinek, R.L.Mercer, S.Roukos, "Principle of lexical language modeling for speech recognition," In S. Furui and M. M. Sondhi, editors, Advances in Speech Signal Processing, pp 651–699, Marcel Dekker, Inc., New York, New York (1992).

[4] X.D.Huang, Y.Ariki, M.A.Jack, "Hidden Markov models for speech recoginition," Edinburgh University Press (1990).

[5] J.Kupiec, "Robust part-of-speech tagging using a hidden Markov model," Computer Speech and Language, vol.**6**, pp.225-242 (1992).

Table 4. The Part-of-Speech codes in the Figures

Part-of-speech	code
common noun	co.noun
pronoun	pron
interjection	interj
conjunction	conj
prefix	prefix
suffix	suffix
verb	verb
conjoining verb	co.verb
adjective	adj
adverb	adverb
auxiliary verb	au.verb
complementizer	comp
case particle	ca.part
topic particle	to.part
conjuctive particle	co.part
sentence final particle	sf.part
negative base	[n]
continuative base	[c]
conclusive base	[cl]
attributive base	[a]