AUTOMATIC ACQUISITION OF PROBABILISTIC DIALOGUE MODELS

Kenji Kita [1] *Yoshikazu Fukui* [1] *Masaaki Nagata* [2] *Tsuyoshi Morimoto* [3]

[1] Faculty of Engineering, Tokushima University
[2] NTT Information and Communication Systems Laboratories
[3] ATR Interpreting Telecommunications Research Laboratories

ABSTRACT

In the work described here, we automatically deduce dialogue structures from a corpus with probabilistic methods. Each utterance in the corpus is annotated with a speaker label and an utterance type called IFT (Illocutionary Force Type). We use an Ergodic HMM (Hidden Markov Model) and the ALERGIA algorithm, an algorithm for learning probabilistic automata by means of state merging, to model the speaker-IFT sequences. Our experiments successfully extract typical dialogue structures such as turn-taking and speech act sequencing.

1. INTRODUCTION

The recent availability of large corpora (a corpus (*pl.* corpora) is a body of machine-readable texts of spoken and written languages) , or linguistic databases, offers a new challenging area in natural language processing. Corpus-based methods, including the use of probabilistic and information-theoretic techniques, are now becoming used increasingly in natural language processing. One of the most interesting issues is deriving linguistic knowledge from large-scale corpora via automated procedures. Most works, however, have focused on deriving lexico-syntactic knowledge.

In the work described here, we automatically deduce dialogue models from a corpus with probabilistic methods. Each utterance in the corpus is annotated with a speaker label and an utterance type called IFT (Illocutionary Force Type). We use an Ergodic HMM (Hidden Markov Model) and the ALERGIA algorithm, an algorithm for learning probabilistic automata by means of state merging, to model the speaker-IFT sequence. Our experiments successfully extract typical dialogue structures such as turn-taking and speech act sequencing.

2. IFT-ANNOTATED DIALOGUE CORPUS

In our work, we used a dialogue corpus with discourse-level information [1]. This corpus is a subset of the ATR Dialogue Database [2], and consists of simulated dialogues between a secretary and a questioner at international conferences. Each utterance is annotated with IFT (Illocutionary Force Type), which is an abstraction of the speaker's intention in terms of the type of action the speaker intends by the utterance. The following 9 IFTs are used in the corpus:

(1) **phatic** ··· phatic expressions such as greetings (e.g. Hello, Good-bye)
(2) **expressive** ··· idioms that express the speaker's feeling (e.g. Thank you, You're welcome)
(3) **response** ··· idiomatic responses and short answers (e.g. Yes, I see, That's right)
(4) **promise** ··· the speaker commits himself to perform an action (e.g. I will send you a registration form)
(5) **request** ··· the speaker asks the hearer to perform an action (e.g. Please go to Kitaooji station by subway)
(6) **inform** ··· informative expressions (e.g. We are not giving any discount this time)
(7) **questionif** ··· Yes-No questions (e.g. Do you have the announcement of the conference?)
(8) **questionref** ··· WH questions (e.g. What should I do?)
(9) **questionconf** ··· confirmations (e.g. You have already transfered the registration fee, right?)

We notice here that the following experimnts used two subcorpora of this IFT-annotated dialogue corpus. The first subcorpus is what we call "Model Dialogues", which consists of 10 dialogues with 225 sentences. The second subcorpus is "Keyboard Dialogues", which consists of 50 dialogues with 1,686 sentences.

3. DIALOGUE MODELING BY ERGODIC HMM

In order to capture the basic characteristics of the local discourse structure, such as turn-taking and speech act sequencing, we tried to model the speaker and IFT sequence by Ergodic HMMs (Hidden Markov Model). The Ergodic HMMs was trained using two kinds of data: the first data consisted of simple IFT sequence, and the second data consisted of the sequence of the speaker and the IFT combination.

Table 1 shows the entropies of the Ergodic HMMs with various number of states. In the table, "IFT" indicates the models derived from the simple IFT sequence, while "SP-IFT" indicates the models derived from the speaker-IFT sequence.

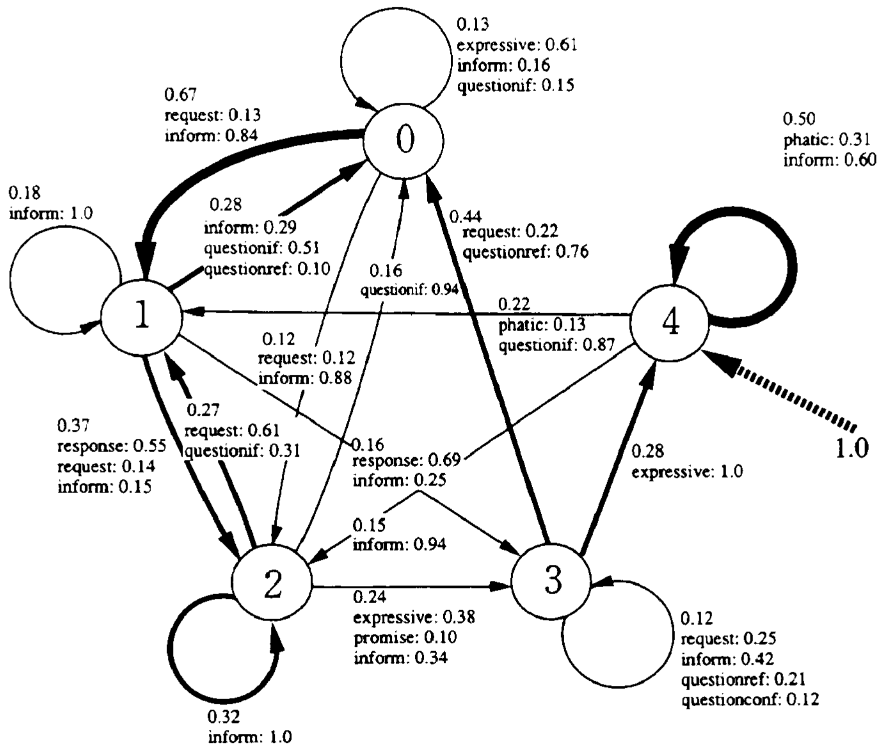

Figure 1: 5 states Ergodic HMM derived from the simple IFT sequence.

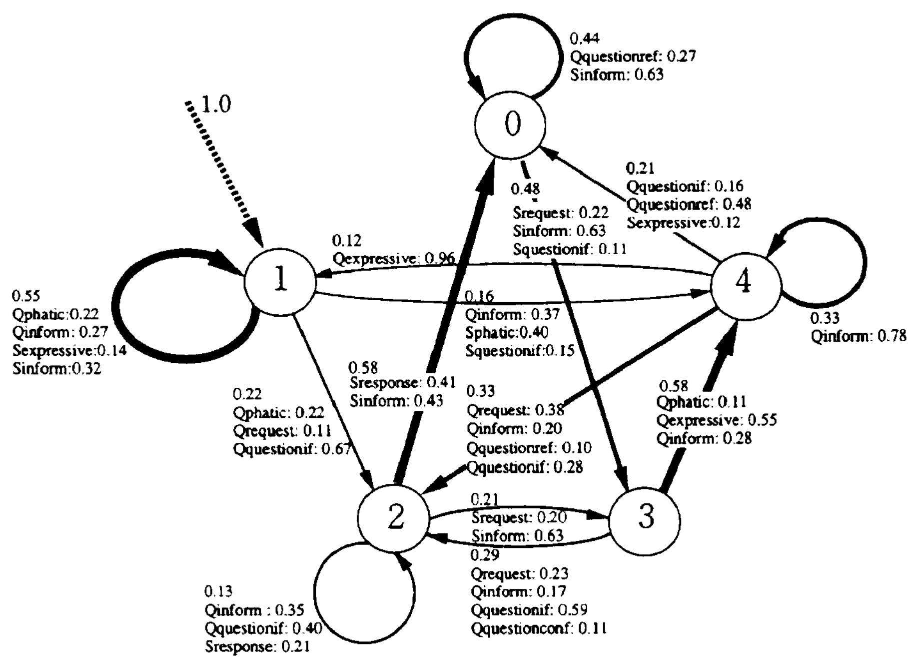

Figure 2: 5 states Ergodic HMM derived from the speaker-IFT sequence.

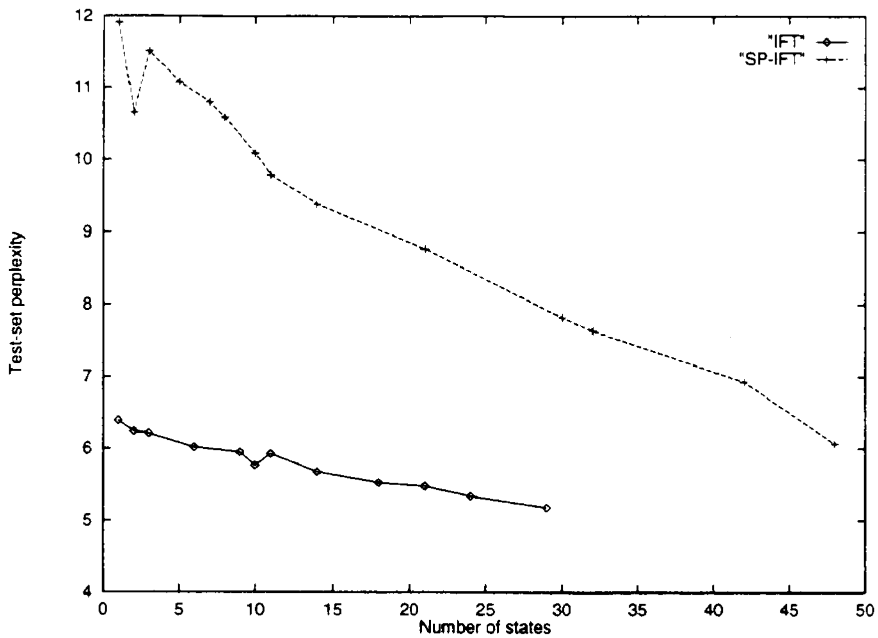

Figure 3: Number of states vs. perplexity

Table 1: Entropy of Ergodic HMMs

Number of states	Model Dialogues		Keyboard Dialogues	
	IFT	SP-IFT	IFT	SP-IFT
2	2.12	2.72	2.38	3.02
4	1.86	2.27	1.89	2.78
6	1.17	1.81	1.91	2.49
8	1.35	1.64	1.88	2.40
10	1.21	1.60	1.60	2.27
12	0.91	1.29	1.63	1.95
14	0.92	1.24	1.72	2.11

Figure 1 and Figure 2 show examples of Ergodic HMMs with 5 states, which were derived from the simple IFT sequence and the speaker-IFT sequence, respectively. In these figures, we have only included transitions and output symbols whose transition and output probabilities were greater than 0.1. Output symbols that start with the "S" or "Q" indicate Secretary's utterance or Questioner's utterance.

The HMM of Figure 2 shows that the initial state is state 1. If a dialogue is initiated by the questioner's utterance "Hello" (Qphatic), there are two possibilities of transitions: one is a self-transition at state 1, and the other is a transition from state 1 to state 2. In the conference registration task, a dialogue may start with the secretary's utterance, say "This is the Secretariat of the International Conference of Computer". In this case, a transition which outputs "Sinform" is selected.

We notice in the HMM of Figure 2 that a particular transition is caused by the questioner's utterance (e.g. transition from state 3 to state 2); on the other hand, for example, the transition from state 2 to state 0 is caused by the secretary's utterance. We also see from Figure 2 that if state 2 is reached after the questioner's utterance, the HMM is likely to take a transition from state 2 to state 0 (with a probability of 0.58) and output "Sresponse" or "Sinform". As is suggested above, we can say that the HMM captures the basic characteristics of the local discourse structure.

4. DIALOGUE MODELING BY STATE MERGING METHOD

A probabilistic language model can be characterized by two parts: a model structure and a set of model parameters. In case of dialogue modeling by Ergodic HMMs, a model structure (number of states) was given in advance. Recently, some methods have been proposed for automatically inducing a model structure based on a state merging method.

We used the ALERGIA algorithm [3], an algorithm for learning probabilistic automata by means of state merging, to model the speaker-IFT sequences. The ALERGIA algorithm first builds the prefix tree acceptor from data and evaluates at every state the relative probabilities of the transitions coming out from the state. Then, the algorithm tries to merge equivalent states until further mering is not possible.

The training data are subjected to statistical fluctuations, and therefore equivalance is checked within a confidence range. In our experiments, we constructed automata with various number of states, varying the confidence range. Fig. 3 shows the relationship between the number of states in the automaton and the automaton's perplexity. As the number of states increases, the entropy decreses. However, since the ALERGIA algorithm constructs the deterministic

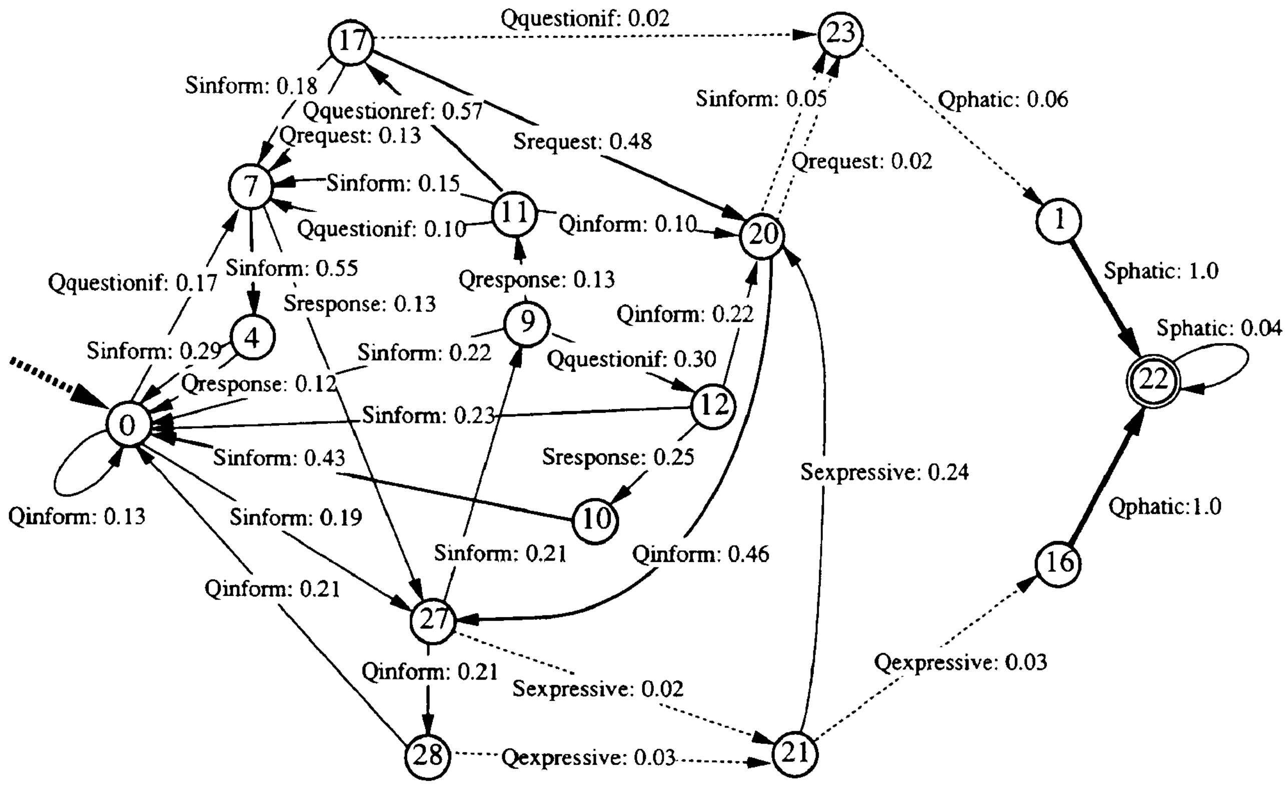

Figure 4: An example of a probabilistic automaton derived by the ALERGIA algorithm

automaton, many states are required in order to achieve the same entropy as in Ergodic HMMs.

Figure 4 shows an example of an automaton derived from the speaker-IFT sequence. This automaton is a portion of the automaton with 30 states. Only the transitions with high probabilities are included in this figure. A dialogue consists of multiple basic dialogue units, such as question-response-confirmation. A state sequence that starts with state 0 and also ends with state 0 (for example, 0 → 7 → 4 or 0 → 7 → 27 → 28) represents a basic dialogue unit.

5. CONCLUSION

This paper was concerned with the problem of automatic acquisition of probabilistic dialogue models from the IFT-annotated dialogue corpus. We used two probabilistic models, Ergodic HMMs and the ALERGIA algorithm, to model the speaker-IFT sequencing. Our experiments showed that these models can capture the basic characteristics of the local discourse structure such as turn-taking and speech act sequencing.

REFERENCES

[1] Nagata, M. and Morimoto, T.: "An Information-Theoretic Model of Discourse for Next Utterance Type Prediction", *Transactions of Information Processing Society of Japan*, Vol. 35, No. 6, pp. 1050-1061 (1994).

[2] Ehara, T., K. Ogura and T. Morimoto, T.: "ATR Dialogue Database", *Proc. of the 1990 International Conference on Spoken Language Processing*, pp. 1093-1096 (1990).

[3] Carrasco, R. C. and Oncina, J.: "Learning Stochastic Regular Grammars by Means of a State Merging Method", *Grammatical Inference and Applications*, Carrasco, R. C. and Oncina, J. (Eds.), Springer-Verlag, pp. 139-152 (1994).