Automatically Deducing Probabilistic Dialogue Models from a Corpus

Kenji KITA, Ryo OKAMOTO and Yoshikazu FUKUI
Faculty of Engineering, Tokushima University
Tokushima 770, JAPAN
e-mail: kita@is.tokushima-u.ac.jp

1. Introduction

With the availability of large corpora in recent years, many successful results have been derived from corpus-based studies. These include part-of-speech tagging, parsing, example-based machine translation, statistical machine translation, and many other related areas. 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. We use the corpus in which each utterance is annotated with a speaker label and an utterance type called IFT (Illocutionary Force Type). First, we improve a probabilistic CFG (Context-Free Grammar) based on an idea of cluster-based language modeling using the IFT-annotated dialogue corpus. Then, 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 research, we used a dialogue corpus with discourse-level information [1]. This corpus is a subset of the ATR Dialogue Database (ADD Corpus), 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?)

Our experiments used two subcorpora of the 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. Improvement of a Probabilistic CFG Using IFT-Annotated Corpus

First, we applied a cluster-based language modeling approach [2,3] to a probabilistic CFG (Context-Free Grammar) using the IFT-annotated dialogue corpus. The cluster-based language modeling assumes that the language model parameters have different probability distributions in different topics or domains. In order to performs topic- or domain-dependent language modeling, we first divide the training corpus into a number of subcorpora according to their topics or domains, and then estimate separate probability distribution from each subcorpus. The language model probability of a sentence is calculated as the mixture of these probability distributions. The cluster-based language modeling enables us to make a context- or topic-dependent language model, and thus accurate language modeling would be possible.

Our experiment used a Japanese grammar for the conference registration task, which had 2,789 production rules including 1,591 words. As stated above, utterances in the dialogue corpus are annotated with their IFTs. Using these IFTs, we divided the "Model Dialogues" corpus into 9 clusters, and then estimated production probabilities from these clusters. The resulting probabilistic CFG was evaluated by calculating test-set perplexity. Without IFT clustering, the test-set perplexity was 2.18 per phone, but using IFT clustering, the test-set perplexity was reduced to 1.82 per phone. This result shows the effectiveness of IFT-clustering.

4. 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. 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. This experiment used the "Keyboard Dialogues" corpus.

Table 1. Entropies of Ergodic HMMs

	Model Dialogues		Keyboard Dialogues	
Number of HMM States	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

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.

Fig.1 and Fig.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.

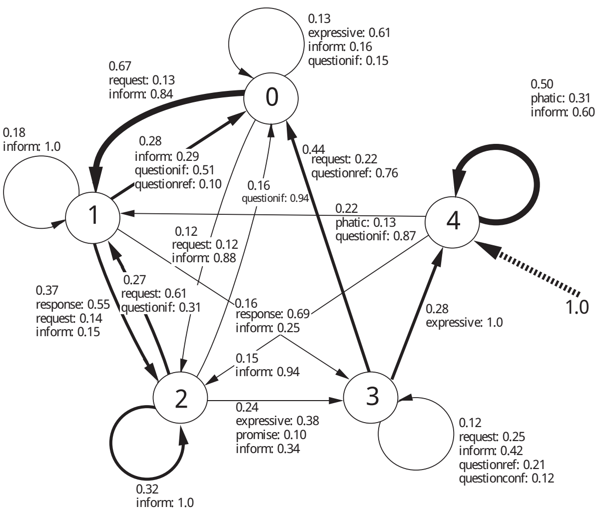

Fig.1

For example, the HMM of Fig.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 Fig.2 that a particular transition is caused by the questioner's

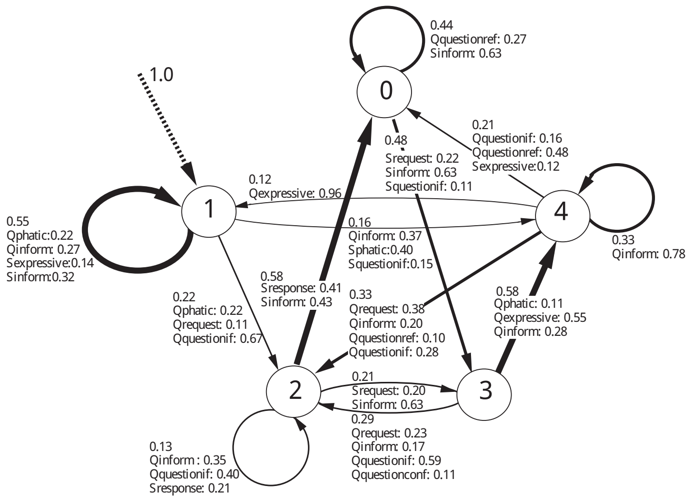

Fig.2

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 Fig.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.

5. 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 [4,5].

We used the ALERGIA algorithm [5], 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 then merges equivalent states.

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 entropy. As the number of states increases, the entropy decreses. However, since the ALERGIA algorithm constructs the deterministic automaton, many states are required in order to achieve the same entropy as in Ergodic HMMs.

Fig.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.

6. Conclusion

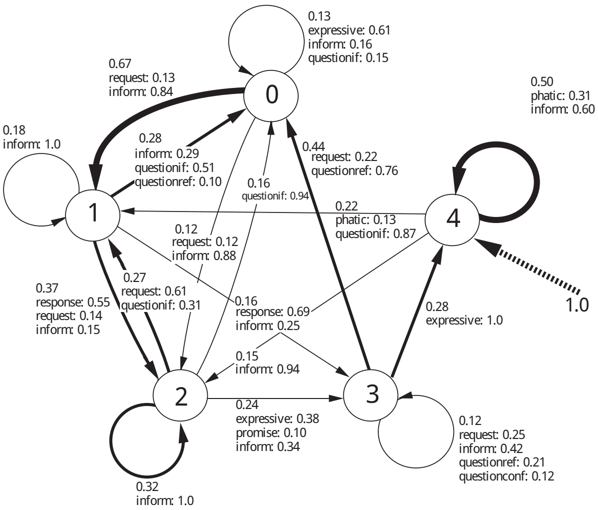
0.13
expressive: 0.61
inform: 0.16
questionif: 0.15
0.67
request: 0.13
inform: 0.84
0
0.50
phatic: 0.31
inform: 0.60
0.18
inform: 1.0
0.28
inform: 0.29
questionif: 0.51
questionref: 0.10
0.44
request: 0.22
questionref: 0.76
0.16
questionif: 0.94
1
0.22
phatic: 0.13
questionif: 0.87
4
0.12
request: 0.12
inform: 0.88
0.27
request: 0.61
questionif: 0.31
0.37
response: 0.55
request: 0.14
inform: 0.15
0.16
response: 0.69
inform: 0.25
1.0
0.28
expressive: 1.0
0.15
inform: 0.94
2
3
0.24
expressive: 0.38
promise: 0.10
inform: 0.34
0.12
request: 0.25
inform: 0.42
questionref: 0.21
questionconf: 0.12
0.32
inform: 1.0

Fig.3

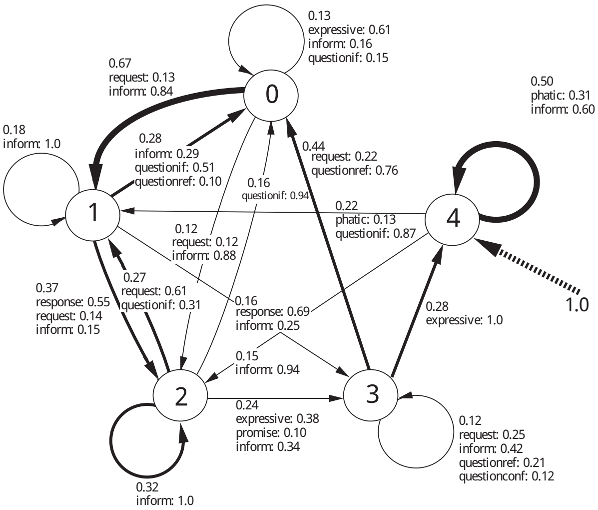
0.13
expressive: 0.61
inform: 0.16
questionif: 0.15
0.67
request: 0.13
inform: 0.84
0
0.50
phatic: 0.31
inform: 0.60
0.18
inform: 1.0
0.28
inform: 0.29
questionif: 0.51
questionref: 0.10
0.44
request: 0.22
questionref: 0.76
0.16
questionif: 0.94
1
0.22
phatic: 0.13
questionif: 0.87
4
0.12
request: 0.12
inform: 0.88
0.27
request: 0.61
questionif: 0.31
0.37
response: 0.55
request: 0.14
inform: 0.15
0.16
response: 0.69
inform: 0.25
1.0
0.28
expressive: 1.0
0.15
inform: 0.94
2
3
0.24
expressive: 0.38
promise: 0.10
inform: 0.34
0.12
request: 0.25
inform: 0.42
questionref: 0.21
questionconf: 0.12
0.32
inform: 1.0

Fig.4

We have described probabilistic dialogue modeling using the IFT-annotated dialogue corpus. In this corpus, each utterance is annotated with a speaker label and an utterance type called IFT (Illocutionary Force Type).

First, we applied a cluster-based language modeling approach to a probabilistic CFG. The basic idea of this model involves clustering a training corpus into a number of subcorpora, and then training probabilistic CFGs from these subcorpora. In our evaluation, we divided the "Model Dialogues" corpus into 9 clusters based on the IFT of each utterance, and then estimated production probabilities from these clusters. The effectiveness of this model was confirmed by perplexity reduction.

Second, we tried to automatically deduce dialogue models from the corpus using Ergodic HMMs and the ALERGIA algorithm. We extracted the speaker and the IFT sequences from the corpus, and then made models for 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] Carter,D.: "Improving Language Models by Clustering Training Sentences", ANLP-94 (1994).

[3] Iyer,R., Ostendorf,M. and Rohlicek,J.R.: "Language Modeling with Sentence-Level Mixtures", Proc. of the ARPA Human Language Technology Workshop, pp.82-87 (1994).

[4] Stolcke,A. and Omohundro,S.M.: "Best-first Model Merging for Hidden Markov Model Induction", ICSI Technical Report, TR-94-003 (1994).

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

Keywords: dialogue modeling, probabilistic model, cluster-based language modeling, illocutionary force type, dialogue corpus, Ergodic HMMs, ALERGIA algorithm