

## 確率・統計的手法による対話構造のモデル化

福井 義和<sup>1</sup> 北 研二<sup>1</sup> 永田 昌明<sup>2</sup> 森元 逞<sup>3</sup>

<sup>1)</sup> 徳島大学 工学部

<sup>2)</sup> NTT 情報通信研究所 知的情報処理研究部

<sup>3)</sup> ATR 音声翻訳通信研究所

確率・統計的手法に基づいた対話のモデル化について研究し、このような対話モデルを大規模言語データベースであるコーパスから自動的に生成するための実験を行なった。実験に用いたコーパスは、ATR 対話コーパスであり、各発話には話者ラベルおよび発話行為タイプ (IFT; Illocutionary Force Type) が付与されている。Ergodic HMM および ALERIGIA アルゴリズムを用いて、話者ラベルおよび IFT の系列をモデル化することにより、話者の交替や質問・応答・確認といった会話の基本的な構造を確率・統計的にモデル化することができた。

## Dialogue Modeling Based on a Probabilistic Approach

Yoshikazu Fukui<sup>1</sup> Kenji Kita<sup>1</sup> Masaaki Nagata<sup>2</sup> Tsuyoshi Morimoto<sup>3</sup>

<sup>1)</sup> Faculty of Engineering, Tokushima University

<sup>2)</sup> NTT Information and Communication Systems Laboratories

<sup>3)</sup> ATR Interpreting Telecommunications Research Laboratories

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 ALERIGIA 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 はじめに

話し言葉や対話における特徴として、旧情報や述語の一部が省略されるなど、断片的で不完全な発話が多く現れるという点をあげることができる。このような断片的あるいは不完全な発話を正しく認識/理解するためには、対話に対する適切なモデルが必要となる。また、話し言葉や対話の音声認識を考えた場合、認識候補の中には統語的にも意味的にも正しいが、対話の文脈の中では不適切な認識候補が存在する場合もある。例えば、文末の述語「～ですか」と「～ですが」は、お互いに誤認識されやすいが、対話モデルを用いることにより、このような誤認識を避けられたり、あるいは誤り訂正が可能となることが期待できる。

文献 [1, 2] では、発話行為タイプ (Illocutionary Force Type; IFT) のラベルが付いたコーパスから、IFT のマルコフモデルを学習し、このモデルが対話のエントロピーを大きく減少させることを示している。我々は、同様の IFT 付きコーパスを用いて、対話構造を表す状態遷移ネットワークを自動生成する研究を行なった。

### 2 IFT 付きコーパス

対話モデル作成のための基礎データとして IFT (Illocutionary Force Type; 発話行為タイプ) 付きコーパス [1, 2, 3] を用いた。これは、ATR 対話データベース中の「国際会議参加登録のタスク」の対話の各発話について、その発話内行為を分析し、陳述・

命令・約束などの発話のタイプが付けられたコーパスである。このコーパスで用いられている IFT は、表層の統語的パターンと比較的直接的な対応がとれる表層 IFT(Surface Illocutionary Force Type) と呼ばれるものである。また、各発話文には、発話者(事務局または質問者)を示すラベルが付与されている。IFT 付きコーパスで用いられている表層 IFT の種類を表 1 に、IFT 付きコーパスの例を図 1 に示す。本研究における評価実験では、IFT 付きコーパスの中から、モデル会話 10 対話(222 文)とキーボード会話 50 対話(1686 文)を用いた。

### 3 Ergodic HMM による対話構造のモデル化

IFT 付きコーパスの各発話には、話者ラベルおよび発話行為が付与されている。話者の交替や質問・応答・確認のような対話の基本的な構造を確率・統計的にモデル化するために、コーパス中の話者ラベルおよび IFT の系列を Ergodic HMM によりモデル化することを試みた。Ergodic HMM の学習データとして、モデル会話およびキーボード会話中から、以下の 2 つのものを抽出した。

- (1) IFT のみの系列
- (2) 話者ラベルと IFT を組み合わせたラベルの系列

実験では、Ergodic HMM の構造として状態数 2~14 のものを用いて学習を行なった。表 2 に、状態数 2, 4, 6, 8, 10, 12, 14 の場合のモデルのエントロピーを示す。表 2 で、IFT と示されているのは IFT のみの系列を用いたときの結果であり、SP-IFT は話者ラベルと IFT を組み合わせたラベルの系列を用いたときの結果である。一般的な傾向として、状態数が増えるに従いエントロピーが小さくなり、同じ状態数では話者ラベルを併用したものの方がエントロピー値が大きくなっている。

文献 [2] の結果では、trigram モデルを使った場合、モデル会話での SP-IFT のエントロピー値は 1.26、キーボード会話での SP-IFT の値は 2.19 と報告している。本実験では、12~14 状態の Ergodic HMM の場合が、trigram のエントロピー値とほぼ同等になっている。また、文献 [2] では、モデル会話につい

表 1: 表層 IFT の種類

phatic	挨拶などで用いられるイディオム的な表現
expressive	話者の感情表現に関するイディオム的な表現
response	質問などに対する応答や合いづち
promise	話し手がある行為をすることを約束する表現
request	話し手が聞き手に行為をすることを依頼する表現
inform	情報の伝達
questionif	真偽疑問文
questionref	疑問語疑問文
questionconf	確認

質問者

phatic: もしもし

questionif: そちらは会議事務局ですか

事務局

response: はい

response: そうです

questionref: どのようなご用件でしょうか

質問者

inform: 会議に申し込みたいのですが

questionref: どのような手続きをすればよろしいのでしょうか

事務局

request: 登録用紙で手続きをして下さい

questionif: 登録用紙は既にお持ちでしょうか

質問者

response: いいえ

inform: まだです

図 1: IFT 付き対話コーパスの例

表 2: Ergodic HMM のエントロピー

HMM の 状態数	モデル会話		キーボード会話	
	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

て、IFT のみの場合と話者ラベルと IFT を併用した場合のエントロピーを比較しているが、後者の方がエントロピー値が小さくなっており、本実験とは対照的な結果となっている。

学習後の HMM の構造 (状態数 5 の場合) を図 2 および図 3 に示す。図 2 は IFT のみの系列から得られたモデルであり、図 3 は話者ラベルと IFT を組み合わせたラベルの系列から得られたモデルである。図には、遷移確率および出力確率が 0.1 以上のもののみを記しており、矢印の太いものほど大きな遷移確率を持っていることを示している。状態遷移の一番上に書かれている確率が遷移確率であり、その下に各シンボル (IFT) の出力確率が記されている。図 3 で、S で始まるシンボルは事務局側の発話であることを、また Q で始まるシンボルは質問者側の発話であることを示している。

例えば、図 3 では、状態 1 が初期状態であり、質問者が最初の発話「もしもし」を発話すると Qphatic を出力する遷移をたどることになる。これは、状態 1 での自己ループあるいは状態 1 から状態 2 への遷移に対応している。「国際会議参加登録のタスク」では、事務局の「こちらは会議事務局です」という発話により対話が始まる場合もある。この場合には Sinform を出力する遷移である状態 1 での自己ループとなる。また、図 3 では、状態遷移が事務局側の発話と質問者側の発話で比較的きれいに分かれている。例えば、状態 3 から状態 2 への遷移は質問者側の発話によって起こり、しかもこの遷移は事務局に対する質問や依頼に対応していることが分かる。この質問や依頼に対し、状態 2 から状態 0 の遷移で事務局が応答 (Sresponse, Sinform) する確率が非常に高いことも読みとることができる。以上のように、

発話行為タイプ付きコーパスから得られた Ergodic HMM は、質問・応答といった基本的な構造を抽出しているということができる。

## 4 状態マーキング手法による対話構造のモデル化

Ergodic HMM によるモデル化では、学習に先立ち、モデルの構造 (状態数) をあらかじめ決めておく必要がある。様々な状態数のモデルを学習し、それらの中から対話構造のモデルとして適しているものを探し出さなければならない。これに対し、近年、状態マーキング手法を用いて、学習データに対し最適な構造を持つモデルを自動的に構築する研究がいくつか行なわれている [4, 5, 6]。我々は、Carrasco らによる ALERGIA アルゴリズム [7] を用いて、対話構造のモデルを構築することを試みた。

### 4.1 ALERGIA アルゴリズム

ALERGIA アルゴリズムは、与えられた学習データを受理する確率的決定性有限オートマトンを構成するアルゴリズムである。ALERGIA アルゴリズムの概要を以下で示す。詳細なアルゴリズムについては、文献 [7] を参照のこと。

#### (1) 接頭木アクセプタの作成

学習データから接頭木アクセプタ (Prefix Tree Acceptor; PTA) を作る。なお、接頭木アクセプタとは、学習データ中のシンボル列を受理する決定性有限オートマトンであり、トライのようにシンボル系列の接頭部分が同じものを共通の状態によって表現したものである。

#### (2) 状態遷移確率の計算

$n_i$  を学習データが接頭木アクセプタの各状態  $q_i$  を訪れた回数とする。もし学習データが状態  $q_i$  で受理されれば、受理されたデータの個数を  $f_i(\#)$  とする。状態  $q_i$  で受理されなければ、次の状態へ遷移するが、このとき状態遷移  $\delta_i(a)$  (状態  $q_i$  でシンボル  $a$  がきたときの遷移) をたどった回数を  $f_i(a)$  と

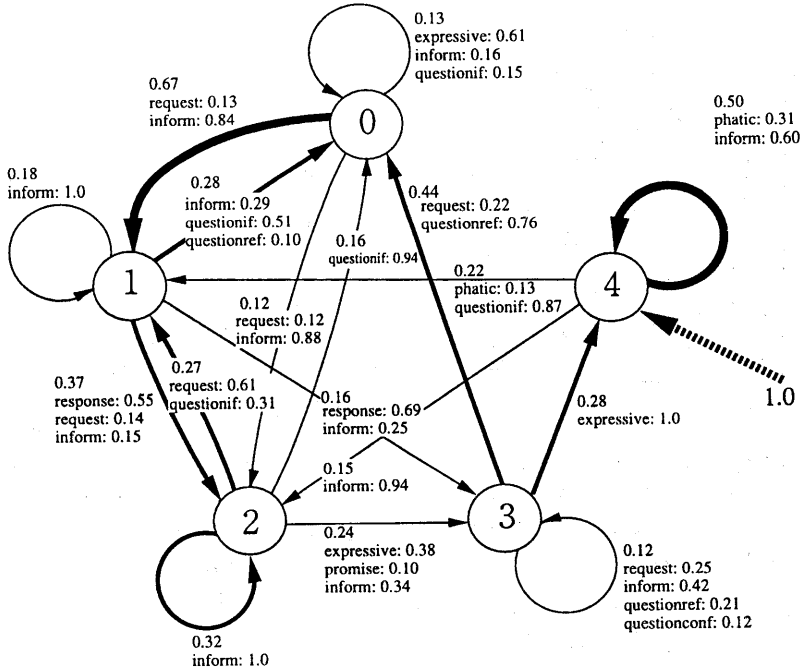


図 2: IFT のみの系列から得られた 5 状態 Ergodic HMM

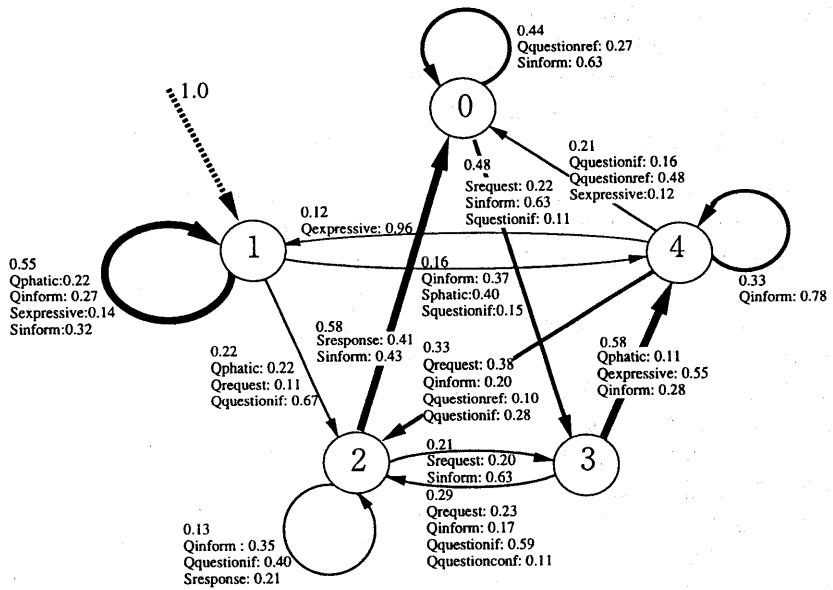


図 3: 話者ラベルと IFT を組み合わせたラベルの系列から得られた 5 状態 Ergodic HMM

する。状態遷移  $\delta_i(a)$  の遷移確率は、次のようにして求められる。

$$P_i(a) = \frac{f_i(a)}{n_i} \quad (1)$$

なお、 $P_i(\#)$  は、データが状態  $q_i$  で受理される確率を表している。

### (3) 状態のマージ

接頭木アクセプタの状態  $q_i$  と  $q_j$  が等価 ( $q_i \equiv q_j$ ) であれば、これら2つの状態をマージする。ここで、状態  $q_i$  と  $q_j$  が等価であるとは、すべてのシンボル  $a \in \Sigma$  について、遷移確率  $P_i(a)$  と  $P_j(a)$  が等しく、遷移後の状態も等価であるときをいう。即ち、状態  $q_i$  と  $q_j$  が等価であれば、次が成り立つ。

$$q_i \equiv q_j \implies \forall a \in \Sigma \begin{cases} P_i(a) = P_j(a) \\ \delta_i(a) \equiv \delta_j(a) \end{cases} \quad (2)$$

状態の等価性を判断する場合、学習データに対する統計的な揺れを伴うので、2つの遷移確率の差が許容範囲にあるときに等価であるとする。

ALERGIA アルゴリズムでは、以下のようにして状態の等価性を決めている。確率  $p$  のベルヌイ確率変数があり、 $n$  回の試行のうち  $f$  回の事象が起こったとすると、次式が成り立つ。

$$P\left(\left|p - \frac{f}{n}\right| < \sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}\right) \geq 1 - \alpha \quad (3)$$

ALERGIA アルゴリズムでは、学習データから推定された2つの遷移確率の差が、信頼範囲  $\sqrt{\frac{1}{2n} \log \frac{2}{\alpha}}$  の和の範囲内にあるときに、2つの状態を等価であるとしている。即ち、状態  $i$  と状態  $j$  が等価であるとは、すべてのシンボル  $a \in \Sigma$  について、次式が成り立つことである。

$$\left| \frac{f_i(a)}{n_i} - \frac{f_j(a)}{n_j} \right| \leq \sqrt{\frac{1}{2} \log \frac{2}{\alpha} \left( \frac{1}{\sqrt{n_i}} + \frac{1}{\sqrt{n_j}} \right)} \quad (4)$$

## 4.2 ALERGIA アルゴリズムの動作例

ALERGIA アルゴリズムの動作を、簡単な例で説明する [7]。いま、学習データとして、次の集合  $S$  が与えられたとする。

$$S = \{110, \lambda, \lambda, \lambda, 0, \lambda, 00, 00, \lambda, \lambda, \lambda, 10110, \lambda, \lambda, 100\} \quad (5)$$

また、 $\alpha = 0.8$  と仮定する。

### (1) PTA の作成

学習データから、図 4 (a) の PTA を作成する。図 4 では、各状態の下に、その状態に到達したデータの個数およびその状態で受理されたデータの個数が示されている。また、各状態遷移には、その遷移を引き起こしたシンボル (0 あるいは 1) とデータ数が示されている。

### (2) 状態 (2,1) の等価性チェック

まず、状態 2 と状態 1 の等価性について考える。2つの状態での受理確率の差は、

$$\left| \frac{1}{3} - \frac{9}{15} \right| = 0.26 < \sqrt{\frac{1}{2} \log \frac{2}{\alpha} \left( \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{15}} \right)} = 0.55 \quad (6)$$

また、シンボル 0 による遷移確率についても、

$$\left| \frac{2}{3} - \frac{3}{15} \right| = 0.46 < 0.55 \quad (7)$$

となる。状態 2 と状態 1 が等価であるためには、更にこれらの状態の遷移先である状態 4 と状態 2 も等価である必要があるが、同様の計算により、状態 4 と状態 2 の等価性も示すことができる。状態 4 と状態 2 をマージし、更に状態 2 と状態 1 をマージすると、図 (b) のオートマトンを得る。

### (3) 状態 (3,1) の等価性チェック

次に、状態 3 と状態 1 について考えると、両者の受理確率の差は、

$$\left| \frac{0}{3} - \frac{12}{20} \right| = 0.6 > \sqrt{\frac{1}{2} \log \frac{2}{\alpha} \left( \frac{1}{\sqrt{3}} + \frac{1}{\sqrt{20}} \right)} = 0.53 \quad (8)$$

となり、等価でないことが分かる。従って、状態 3 と状態 1 をマージすることはできない。

### (4) 状態 (5,1) の等価性チェック

以上の計算と同様にして、状態 5 と状態 1 の等価性も示すことができる。また、状態 (5,1) の等価性を調べる過程において、状態 (7,1), (8,3), (10,6), (11,9) の等価性も同時に示される。これらの状態をマージすると、図 (c) のオートマトンを得る。

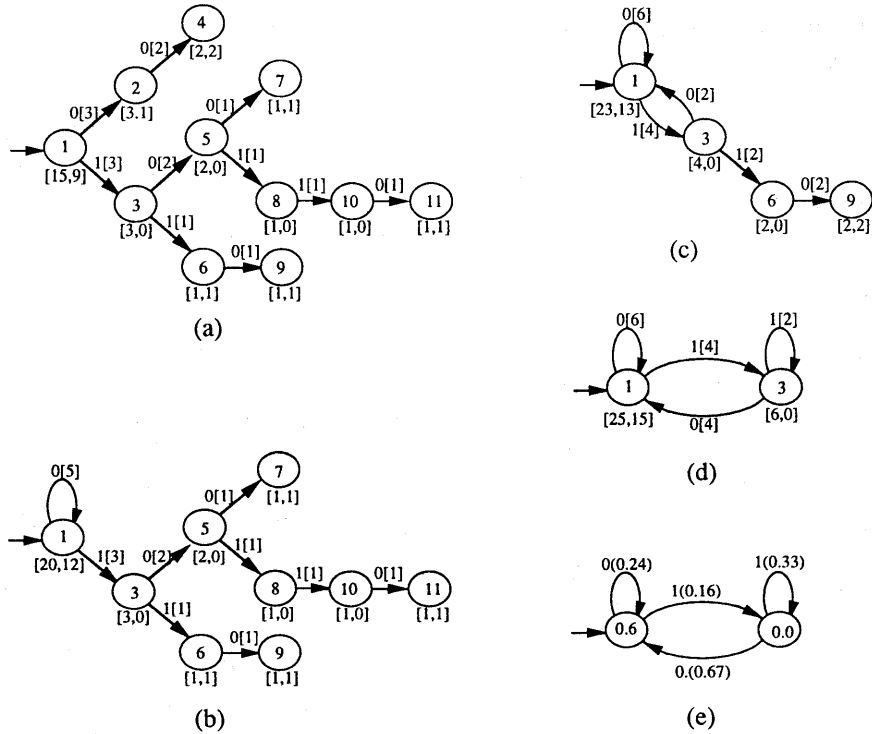


図 4: ALERGIA アルゴリズムの動作例

#### (5) 状態 (6, 3) の等価性チェック

同様に、状態 6 と状態 3 の等価性も示すことができる。状態 (6, 3) をマージすると、図 (d) のオートマトンを得る。受理確率および遷移確率を計算して、最終的に図 (e) のオートマトンを得る。

### 4.3 対話構造のモデル化

上述の ALERGIA アルゴリズムを用いて、IFT 付きコーパスから対話構造をモデル化する実験を行った。学習データとしては、キーボード会話 50 対話 (1686 文) を用いた。ALERGIA アルゴリズムでは、式 (4) の右辺の値を変えることにより、様々な状態数を持つオートマトンを構成することができる。図 5 に、ALERGIA アルゴリズムにより得られたオートマトンの状態数とパープレキシティの関係を示す。パープレキシティの値は、学習データに対するテストセット・パープレキシティを用いている。

状態数の増加にともないパープレキシティは減少している。

パープレキシティ  $P$  とエントロピー  $H$  の間には、

$$P = 2^H \quad (9)$$

なる関係があるが、式 (9) より、ALERGIA アルゴリズムで得られたモデルのエントロピーを算出してみると、HMM と同程度の精度を達成するためには、HMM の場合よりもはるかに多くの状態が必要となることが分かる。これは、HMM が非決定性の有限オートマトンと等価であるのに対し、ALERGIA アルゴリズムにより得られるモデルが決定性の有限オートマトンであるためである。

図 6 は、話者ラベルと IFT を組み合わせたラベルの系列から得られた 30 状態のオートマトンの一部 (16 個の状態) である。このオートマトンの初期状態は状態 0 であり、最終状態は状態 22 である。図の右側には、初期状態 0 から状態遷移する確率の高い 11 個の状態 (状態 0, 4, 7, 9, 10, 11, 12, 17, 20,

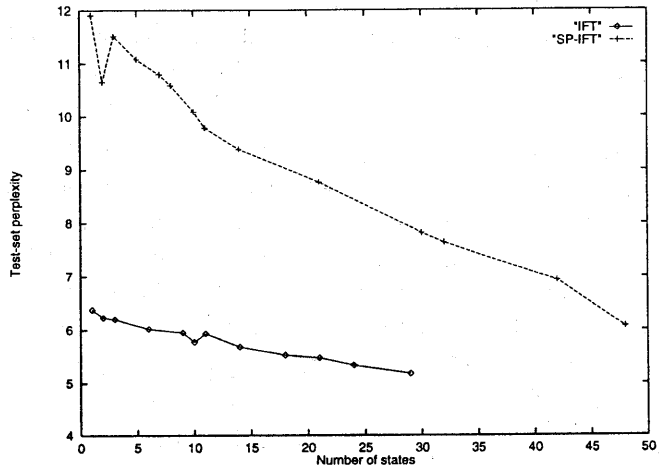


図 5: 状態数とパープレキシティの関係

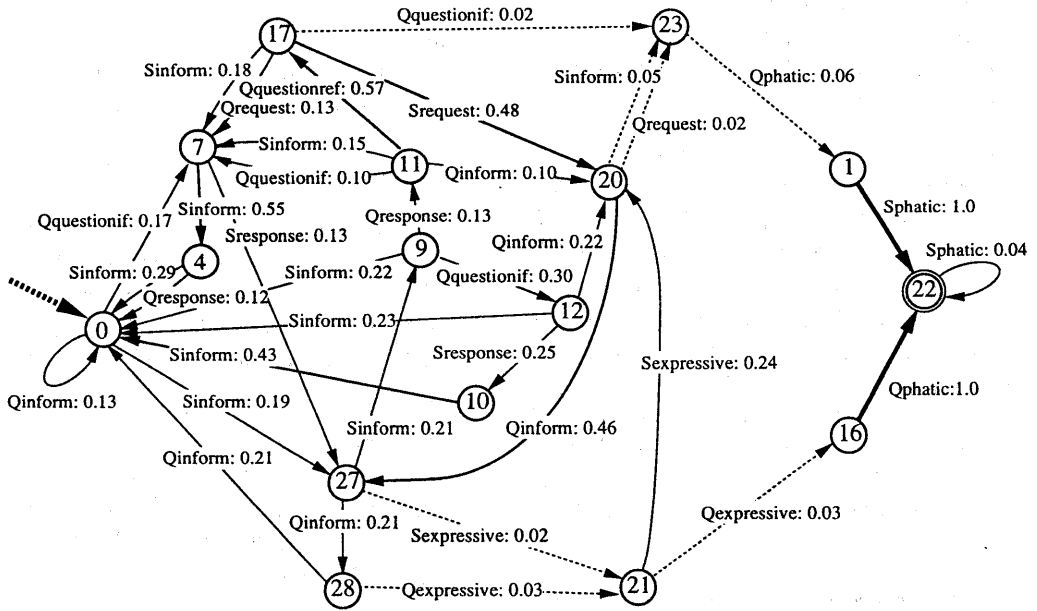


図 6: ALERGIA アルゴリズムにより得られたオートマトンの一部

27, 28) が示されている。状態 0 から始まり再び状態 0 に至る状態遷移系列 (例えば、 $0 \rightarrow 7 \rightarrow 4$  や  $0 \rightarrow 7 \rightarrow 27 \rightarrow 28$  など) が、質問・応答・確認などの対話の基本サイクルを表していると考えることができる。

また、図の左側に、最終状態 22 に状態遷移する確率の高い 5 個の状態 (状態 1, 16, 21, 22, 23) が示されている。例えば、状態 27 あるいは 28 で、expressive (例:「ありがとうございました」) に対応する発話が現れると、最終状態へ向かう状態遷移が選択されるということが分かる。しかし、国際会議参加登録のタスクでは、expressive や phatic という IFT の出現頻度は IFT 全体の数パーセントにしか過ぎないので、状態 27 あるいは 28 から状態 21 へ遷移する確率は低くなっている (図中、遷移確率の小さいものは破線で示されている)。

## 5 まとめ

確率・統計的手法に基づいた対話のモデル化について研究し、このような対話モデルを大規模言語データベースであるコーパスから自動的に生成するための研究を行なった。

IFT 付きのコーパスから、話者の交替や質問・応答・確認といった会話の基本的な構造を確率・統計的にモデル化するために、Ergodic HMM および ALE-RIGIA アルゴリズムを用いた対話モデルを作成した。Ergodic HMM によるモデル化では、質問・応答等の対話の基本構造を抽出することができ、話者ラベルと IFT を併用した場合には、状態遷移が事務局側の発話と質問者側の発話で比較的きれいに分かれたモデルが得られた。また、ALERGIA アルゴリズムにより得られるモデルは決定性の有限オートマトンであるため、HMM と同程度の精度を達成するためには、HMM に比べはるかに多くの状態が必要であった。

今後は、同様の手法を用いて、対話における話題の遷移等をモデル化するための研究を行ないたいと考えている。

## 参考文献

- [1] 永田 昌明:「統計的な対話モデルの試みとその音声認識への応用」, 人工知能学会研究会資料, SIG-SLUD-9202-11, pp. 94-103 (1992).
- [2] 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).
- [3] 鈴木 雅実, 永田 昌明:「日英対話コーパスへの談話レベルの情報付与と翻訳への利用」, 信学技報 NLC93-38, pp. 57-64 (1993).
- [4] Stolcke, A. and Omohundro, S. M.: "Best-first Model Merging for Hidden Markov Model Induction", ICSI Technical Report, TR-94-003, International Computer Science Institute, University of California at Berkeley (1994).
- [5] Stolcke, A.: "Bayesian Learning of Probabilistic Language Models", Ph.D Thesis, International Computer Science Institute, University of California at Berkeley (1994).
- [6] Stolcke, A. and Omohundro, S.: "Inducing Probabilistic Grammars by Bayesian Model Merging", *Grammatical Inference and Applications*, Carrasco, R. C. and Oncina, J. (Eds.), Springer-Verlag, pp. 106-118 (1994).
- [7] 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).