

自動生成されたルールによる発話行為の推定

浅見 克志 竹澤 寿幸 菊井 玄一郎

ATR 音声言語通信研究所

An Estimation of Speech Act Type Using Automatically Generated Rules

Katsushi Asami, Toshiyuki Takezawa, Genichiro Kikui

ATR Spoken Language Translation Research Laboratories

1R-5

1 はじめに

音声対話ヒューマンマシンインタフェースを持つシステムでは、ユーザビリティ向上のため、多様な言い回しの受理が必要であるが、実際には、可能な全ての言い回しの解析は困難であり、受理可能な言い回しは制限される。しかし、多様な言い回しに含まれる、発話意図やタスクに関連するキーワードには、ある程度の傾向が認められる。それを捉え、発話意図・タスクに関する情報を抽出する手法は、受理可能な言い回しを拡大する 1 つの有効な解である。

本稿で提案する手法では、発話行為(Speech Act, SA)タグを付与したコーパスから、推定対象となる SA 毎に相関の高いキーワード、またそれらの共起組合せを抽出し、それを前件部とする確信度付ルール群を自動生成しておく。SA の推定結果は、入力文に応じて活性化される複数のルールのうち、最大の確信度を持つルールによって出力される。また、本手法は、確率決定木など既存の手法に比べ、ルールが極めてシンプルに記述されるメリットがある。

以降、ルールの自動生成と推定実行時のルール合成の方法について述べ、日本語旅行対話コーパスを用いた実験結果を報告し、考察を行う。

2 問題の設定

本稿で述べる手法は、入力を意味単位(SDU)に分割された発話文テキスト、出力を入力 SDU に対する SA とする。また、発話文テキストに対する形態素情報、話者情報および相手話者の直前ターンの最終 SDU に対する SA が分かっているものとする。入力 SDU U が i 個の形態素から構成される場合、 U は次のように表現される。

$$U=(r, a', w_1, \dots, w_i) \quad (1)$$

r : 話者情報 (役割) (Agent / Customer)

a' : 相手話者の直前ターンの最終 SDU に対する SA

w_i : 入力 SDU を構成する形態素

本手法では r, a', w_i を属性 c_j で置換し、これらを同種の情報として取り扱う。

ここで課題となるのは、推定対象となる各 SA に対して、ルールの前件部を構成する情報要素 c'_1, \dots, c'_k および c''_1, \dots, c''_l を抽出し、さらに各ルールに対する確信度 μ_R を求めることである。このルールは次のように表現される。

$$\begin{aligned} & \text{If (include}(c'_j \text{ and} \dots \text{ and } c'_k) \text{ in } U \text{ and} \\ & \text{not include}(c''_j \text{ or} \dots \text{ or } c''_l) \text{ in } U) \\ & \rightarrow \text{speech act type} = a \text{ for } U \quad (2) \end{aligned}$$

これは、ルール「情報要素 c'_1, \dots, c'_k をすべて含み、かつ情報要素 c''_1, \dots, c''_l の何れも含まない入力 SDU U の speech act は a である」を示している。なお、1 つの SA に対して、ルールは複数存在し得る。

3 ルールの自動生成

SA 推定ルールは、(2)に示すように SA 毎に独立したルールであり、ルール前件部にに基づき入力 SDU に対する SA が a であるか否かを判別する。ところで、SA の推定は、発話を複数個の SA クラスへ分類するクラス分け問題、つまり、推定すべき SA が K 個あるとき、 K -class 問題と考えられる。さらに、 K -class 問題は一般的に K 個の 2-class 問題に分割できる[1]。すなわち SA 推定ルールは、2-class 問題に対応する。

ルールは、推定すべき SA 毎に、学習データからターゲットとする SA タグを付与された発話を弁別する条件を求めることによって、自動生成できる。ここで、ルールが SA 毎に独立しているため、ルール生成は並列に処理できる。図 1 に、SDU の全体集合の SA 毎の 2-class 問題への分割と統合を示す。

また、SA a_n に対する前件部 d_p ($p=1, 2, \dots$) を持つルールの確信度 $\mu_R(d_p, a_n)$ は次式で求められる。

$$\mu_R(d_p, a_n) = \frac{\text{number of elements in } (S_{d_p} \cap S_{a_n})}{\text{number of elements in } S_{d_p}} \quad (3)$$

S_{d_p} : 前件部 d_p を満たす SDU の集合

S_{a_n} : SA が a_n である SDU の集合

すなわち、確信度の取り得る値は $0 \leq \mu_R(d_p, a_n) \leq 1$ となる。

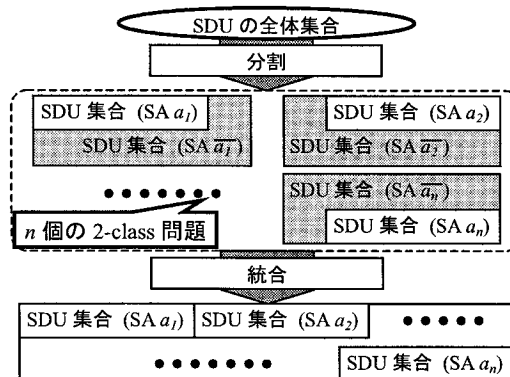


図 1 2-class 問題への分割と統合

4 SA の推定

SA 推定処理は、ファジィ推論で用いられる

max-min 合成を利用する。入力 SDU U とルール前件部の関係を Z , ルール前件部と後件部の関係を R とし、入力 SDU U に対する各 SA の確信度の関係 A は次式で求められる。

$$A = Z \circ R$$

$$= [\mu_Z(d_1) \cdots \mu_Z(d_p)] \begin{bmatrix} \mu_R(d_1, a_1) & \cdots & \mu_R(d_1, a_n) \\ \vdots & \ddots & \vdots \\ \mu_R(d_p, a_1) & \cdots & \mu_R(d_p, a_n) \end{bmatrix}$$

$$= [\mu_A(a_1) \cdots \mu_A(a_n)] \quad (4)$$

ここで

$$Z = [\mu_Z(d_1), \dots, \mu_Z(d_p)]$$

$$\mu_Z(d_i) = \begin{cases} 0: U \text{ は前件部 } d_i \text{ に適合しない} \\ 1: U \text{ は前件部 } d_i \text{ に適合する} \end{cases}$$

$$(i=1, 2, \dots, p)$$

および

$$\mu_A(a_j) = \bigvee_{i=1}^p (\mu_Z(d_i) \wedge \mu_R(d_i, a_j)) \quad (j=1, 2, \dots, n)$$

である。入力 SDU に対して推定される SA \hat{a} は

$$\hat{a} = \arg \max_{a_j} \mu_A(a_j) \quad (j=1, 2, \dots, n) \quad (5)$$

で与えられる。

5 実験

実験は、旅行会話をタスクとする自然発話データベースの 275 対話に対して、人手により SDU 分割し、SA タグを付与したデータを用いて行った。このデータに付与された 31 種類の SA タグのうち、出現数上位 20 位までの SA を推定対象とした。推定対象の SA タグが付与された SDU 数は 13185 である。

推定結果の評価は、SDU 毎に人手で付与された SA タグと推定出力のタグを比較して行う。両者が一致する場合、推定出力が正解であるとする。推定対象の SA type 全体に関する推定結果の accuracy は、推定対象の SA タグが付与された SDU の総数に対する、推定出力が正解となった SDU 数の割合として求められる。また、推定対象の SA それぞれに対する評価として、recall と precision を求めた。

$$\text{recall}_{a_n} = \frac{N_{\text{correct}, a_n}}{N_{\text{manual}, a_n}}, \quad \text{precision}_{a_n} = \frac{N_{\text{correct}, a_n}}{N_{\text{estimate}, a_n}} \quad (6)$$

$N_{\text{correct}, a_n}, N_{\text{manual}, a_n}, N_{\text{estimate}, a_n}$ はそれぞれ、SA a_n に関して推定出力が正解の SDU 数、人手付与の SA が a_n の SDU 数、推定出力 SA が a_n の SDU 数である。

表 1 に実験結果を示す。accuracy に関しては、確率決定木を利用した他の手法 (例: K.Tanigaki, et al. - 88.2%[2]) に近い性能が得られていると考えられる。

recall / precision に関しては出現数上位 4 位までの SA を表示している。これらについては、Open Test で約 75-90%, Closed Test で 90%以上となっている。表に示していない出現頻度が少ない SA については、これらの値が低いものも見られる。

表 1 実験結果

Acc.	Open Testing			Closed Testing		
	Rc.	Pr.	Occ.	Rc.	Pr.	Occ.
A	91.3	86.5	183	95.1	93.8	2795
B	96.1	89.7	154	95.4	92.4	2650
C	79.3	91.3	92	92.6	91.0	1598
D	88.5	76.2	87	92.7	92.4	1254

Acc. : accuracy (%)

Rc. : recall (%)

Pr. : precision (%)

Occ. : occurrence

A : SA type = give-information

B : SA type = acknowledge

C : SA type = request-information

D : SA type = verify-give-information

また、推定エラーのパターンを調べると、エラーの多い SA と、エラーによる置換先 SA が特定の SA に集中している。これは、それらの SA に分類される発話で、似た表現あるいは同じ語が用いられるためと考えられる。例えば、「予約をお願いします」→request-action に対して「それをお願いします」→accept である。人間はこのような場合、その場面に応じて適切に判断できる。これに対して、提案手法では、話者情報や相手話者の直前発話の SA を考慮したルールが生成されているが、これだけでは不十分と考えられる。推定精度の向上のため、さらに広範な情報、例えばシーン記述などを利用する必要がある。

6 むすび

実験の結果、提案手法は、従来から提案されている方法に近い精度で SA を推定できることが示された。これに加え、提案手法にはルールがシンプルに記述されるメリットがある。これにより、キーワードの言語的な意味に基づいた人手によるルールと容易に融合できると考えられる。統計的手法で言語を扱う場合、そこで利用されるのは言語の情報理論的意味であり、人間同士のコミュニケーションで用いられる言葉の意味とは異なっている場合がある。そのような言葉の「言語的」意味などを適切にルールに組み込むことにより、推定精度の更なる向上が可能と考えられる。

今後、今回の実験結果を踏まえ、自動生成したルールと人手によるルールとの融合について、検討・実験を進める計画である。

7 参考文献

- [1] B.L. Lu, et al., "Task Decomposition and Module Combination Based on Class Relations: A Modular Neural Network for Pattern Classification", IEEE Trans. Neural Networks, Vol.10, No.5, pp.1244-1256, 1999
- [2] K.Tanigaki, et al., "Robust Speech Understanding Based on Word Graph Interface", Proc. of ESCA workshop, pp.45-48, 1999