

WWWを介した対話システムとユーザとの対話の分析

An Analysis of Dialogues with Our Dialogue System through a WWW page

熊本 忠彦	伊藤 昭
Tadahiko Kumamoto	Akira Ito
kuma@crl.go.jp	ai@crl.go.jp

郵政省 通信総合研究所 関西支所
〒651-2401 兵庫県神戸市西区岩岡町岩岡 588-2

あらまし: 自然言語対話システムにユーザとの対話を通して何らかの課題を解かせ、その成績によって対話システムの性能を客観的・総合的・定量的に評価しようという試みがWWW (World-Wide Web) ページを介して行われた (第2回対話コンテスト DiaLeague'97)。今回のコンテストには我々も含めて5つのグループが参加したが、その中で我々の対話システムが第1位の成績を得ることができた。本稿では当コンテストで得られた対話 (728 対話) のうち、我々の対話システムとの対話 (141 対話) を分析し、非人間との対話におけるユーザ発話の多様性について調べる。さらに、その多様性が対話の量に及ぼす影響についても明らかにする。

1. はじめに

自然言語による対話は、人間にとって最も自然かつ容易な意志伝達手段であり、自分の信念や意図、外界の状況などを他者に伝えるのに適している。それゆえ、自然言語対話を実現する対話システムが人に優しいマン・マシンインタフェースとして盛んに研究されている [1]-[3]。

対話システムの性能を総合的・客観的・定量的に評価するための一つの試みとして、対話コンテスト DiaLeague'97 [4] が実施された。このコンテストは、あらかじめ登録された対話システムが人と自然言語 (日本語) で対話し、与えられた経路課題を解くまでの対話量 (得点=300-発話文節数の総和) を競うというものであり、対話システムとの対話をWWW (World-Wide Web) ページを介して実現したという点に特徴があった。したがって、今までの対話システムとの対話が被験者として集められた少数ユーザによる限定的なもの、もしくはシステム設計者による模擬的なものであったのに対し、今回のコンテストで得られた対話は不特定多数のユーザによる実対話 (real man-machine dialogue) であると言える。

本稿では当コンテストで得られた対話 (728 対話) のうち、我々が開発し、参加登録した対話システムとユーザとの対話 (141 対話) を分析し、非人間との対話におけるユーザ発話の多様性について調べる。さらに、その多様性が対話量に及ぼす影

響についても明らかにする。

なお、当コンテストにおいて、対話システムへの入力 (ユーザ発話) はキーボード等によるテキスト入力であり、対話システムからの出力 (システム発話) は画面へのテキスト表示であった。

2. DiaLeague'97 の概要

DiaLeague'97 は、対話システムの性能を総合的・客観的・定量的に評価するためのコンテストであり、1997年9月25日正午から10月2日正午までの168時間行われた [4]。今回登録された対話システムは全部で5つあり、対話システムにアクセスしたユーザの延べ人数は728人であった。

このコンテストで両者に与えられた課題は経路課題と呼ばれるもので、いずれの対話者も正解を知らない、両者が互いに情報を伝え合って正解を見い出さなければならないという相互型の課題であった。具体的には、対話システムとユーザのそれぞれに与えられた微妙に異なる鉄道路線図から両方の路線図に共通な経路を見つけないというものであった。但し、双方の路線図では、駅の個数や配置、出発駅 (スタート)、目的駅 (ゴール) は同じだが、線路のつながり方 (結線パターン) や名称不明な駅の位置が異なっていた。また、両者に共通な経路 (正解経路) は1通りしかなかった。

コンテストは、ネットニュースやメーリングリストを通じて広くアナウンスされ、WWWページ

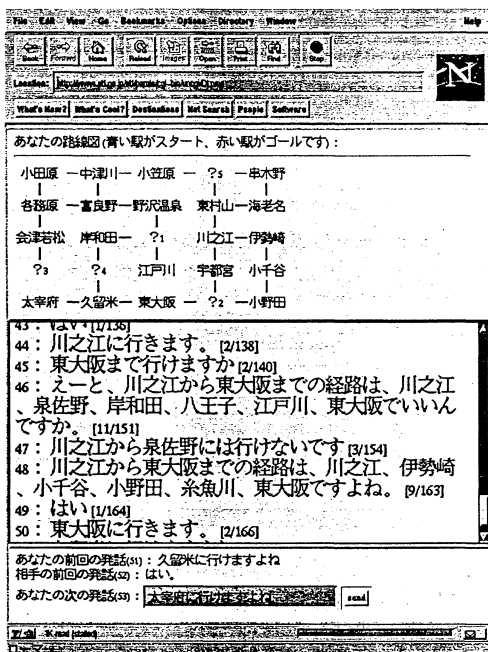


図1 WWWページを介した対話の様子

上で開催された。ユーザはまず対話サーバ (<http://www.etl.go.jp/etl/nl/dialeague/webdialogue/>) にアクセスし、エントリー名 (本名/メールアドレスである必要はない) を登録した。すると、対話サーバは、対話システム、対話の順番 (先手/後手)、経路課題を適当に決め、図1に示すような3つのフレームからなるWWWページを構成した。このWWWページの一冊上のフレームにはユーザ用の路線図が表示されており、右上の駅 (青字) がスタート、左下の駅 (赤字) がゴールとなっている。記号「?」で始まる駅名は、その駅の名前が不明であることを示している。真ん中のフレームにはこれまでの対話の内容が表示されており、いつでも参照できる。一番下のフレームにはユーザ発話入力用のボックスがあり、キーボードによる日本語入力、マウスによるカット&ペーストを受け付けている。なお、発話入力は改行 (もしくは send ボタンを押すこと) によって終了したため、1入力が1文とは限られていなかった。そのため、複数の文を一度に入力することもできたが、対話の順番 (ターン) は最初決められたとおりであった。

以上のコンテストの結果、我々の対話システムが平均得点 123.29 点 (対話回数 141 回) を獲得し、

第1位の成績を得た。ちなみに、第2位の対話システムで平均得点は 104.30 点 (185 回) であり、以下、98.38 点 (175 回)、37.50 点 (167 回)、20.73 点 (60 回; 中途参加) であった。また、当コンテストではユーザの成績ランキングも求められており、優勝者の平均得点は 194.71 点 (対話回数 14 回) であった。以下、第2位 193.40 点 (10 回)、第3位 172.63 点 (19 回)、第4位 154.80 点 (10 回)、第5位 138.15 点 (13 回) と続いている。

3. 発話の分析

3.1 発話パターンの抽出

今回分析の対象とした対話は、我々の対話システムとユーザとの対話 (141 対話) であり、ユーザ発話 2930 文、システム発話 2898 文からなる。これらの発話には駅名などの個別概念や表記上の揺れなどが存在したので、まずはじめにそれらの除去を行った。具体的には以下の方法で各発話の表記を変換した。

- 文字化けを含む発話文を「※文字化け※」と表記する。
- 漢字表記とひらがな表記が混在しているものを漢字表記もしくはひらがな表記のいずれかに統一する。
- 間投表現を除去する。
- 句読点や記号を削除する。
- 駅名を「A」と表記する。
但し、明らかに入力ミスとわかる駅名 (木皿図 (木更津) など) も「A」と表記する。
- 駅名の2個以上の連なりを「A…A」と表記する。
- 方向概念 (「上下左右東西南北」) を「(方向)」と表記する。
- 数字概念 (序数を含む) を「(数字)」と表記する。

以上の方法により、例えば、「太宰府は春日部の右ですか。」という発話は「AはAの(方向)ですか」という表記に、また「えーと、富士見、寝屋川、東村山、習志野と行けますか?」という発話は「A…Aと行けますか」という表記に変換される。

3.2 ユーザ発話の多様性

まずはじめに、ユーザ発話 2930 文およびシステム発話 2898 文を文の形式と文に含まれるトピックの種類で分類した。その結果を表1に示す。

表1 発話の分類

	トピック	システム 発話	ユーザ 発話
平叙文	移動欲求	227 (4)	4 (4)
	移動意志	98 (1)	117 (15)
	移動可能性	32 (2)	458 (54)
	駅名	57 (1)	348 (12)
	駅の場所	128 (3)	89 (13)
	経路	2 (1)	18 (9)
	現在地	6 (1)	0 (0)
	出発駅	2 (1)	9 (2)
	乗り換え駅	0 (0)	1 (1)
	経路の有無	0 (0)	5 (3)
	その他	7 (2)	12 (12)
	小 計	559 (16)	1061 (125)
真偽疑問文	移動欲求	0 (0)	1 (1)
	移動意志	0 (0)	6 (5)
	移動可能性	1320 (4)	347 (46)
	駅名	0 (0)	8 (6)
	駅の場所	0 (0)	38 (2)
	経路	197 (6)	35 (10)
	乗り換えの有無	0 (0)	3 (3)
	経路の有無	0 (0)	1 (1)
	その他	0 (0)	5 (5)
	小 計	1517 (10)	444 (79)
疑問語疑問文	移動方法	0 (0)	2 (2)
	駅名	357 (3)	70 (11)
	駅の場所	81 (4)	121 (12)
	経路	0 (0)	1 (1)
	現在地	10 (1)	10 (7)
	出発駅	22 (1)	4 (3)
	経由駅	0 (0)	1 (1)
	その他	0 (0)	1 (1)
	小 計	470 (9)	210 (38)
	移動依頼	0 (0)	9 (8)
	肯定的応答	211 (3)	915 (18)
	否定的応答	123 (6)	243 (11)
	その他の応答	15 (1)	7 (6)
	その他	3 (3)	41 (17)
	合 計	2898 (48)	2930 (302)

※括弧内の数字は発話パターン数を表す。

表1においてユーザ発話とシステム発話の発話パターン数を比べてみると、ユーザ発話は302個のパターンに、システム発話は48個のパターンに分類されることがわかる。システムの発話パターンは、経路課題を解く上で最低限必要な対話が行えるレベルに、対話が単調にならないよう文末表

異なり形態素数(上)/異なり未定義語数(下)

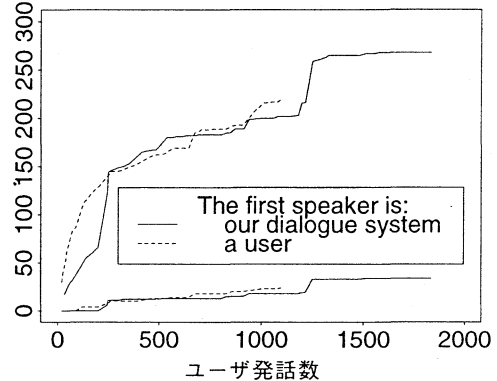


図2 異なり形態素数/異なり未定義語数とユーザ発話数の関係

発話パターン数

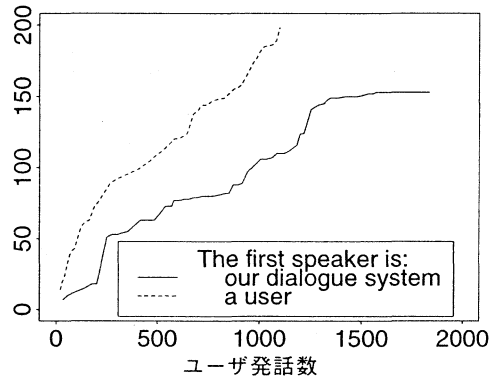


図3 発話パターン数とユーザ発話数の関係

現にバリエーションを与えて設計されているが、それでも48個のパターンで十分であった。これに対し、ユーザ発話は302個とシステム発話の6.3倍(302/48)もあり、ユーザ発話のパターンの多様性が示されている。

次に、対話数の増加、すなわちユーザ発話数の増加に伴い、異なり形態素や発話パターン数などがどう変化するか調べた。但し、形態素の認定は汎用日本語形態素解析システム JUMAN [5] を用いて自動的に行ったため、一部の発話文では未定義語や誤った形態素解析結果が生じている。本稿では未定義語は「未定義語」という品詞と仮定してカウントし、誤りもそのままカウントした。以上の結果を図2と図3に示す。図2の上側の線(2

本)が異なり形態素の数を、下側の線(2本)が異なり未定義語の数を示しているが、異なり形態素の数は、対話の順番が対話システム先手であるかユーザ先手であるかに関係なく、同じようなペースで増加している。また、ユーザ発話数1200文付近で(対話システム先手)急激な増加が見られるが、これは形態素解析の失敗が主因であり、実質的にはユーザ発話数が500文を超えたあたりから、新たな形態素の出現頻度は減少傾向にある。一方、図3によれば、発話パターン数はユーザ発話数にほぼ正比例して増えているが、発話パターンの場合もユーザ発話数1300文付近でその増加率は減少している(対話システム先手)。

3.3 ユーザ発話の多様性が対話に及ぼす影響

次に、ユーザ発話数の増加が対話の成績に及ぼす影響を明らかにするために、ユーザ発話数と平均得点の関係を調べてみた。その結果を図4に示す。図4は、ユーザ発話数が増加しても、平均得点はほぼ横ばい、もしくは若干の上昇傾向であることを示している。これは対話システムが対応できないような形態素/発話パターンの出現頻度がほぼ一定であることを示唆している。また、対話システム先手のときとユーザ先手のときとでは平均得点の高さが明らかに異なっている。これは図3において観測された発話パターン数の増加率の違いに対応しているものと考えられる。

ユーザとの対話数が増えれば、対話システムにとって未定義な形態素や発話パターンの数も増えるが、その増加率はほぼ一定であることが示唆された。従来研究では、未知な形態素に対してはその品詞や概念を推定することに、未知な発話パターンに対しては自然言語処理部の受理可能なパターンを増やすことに焦点が当てられている。しかしながら、人間同士の対話ではそのような知識の不足に起因する混乱は対話そのものによって修復されることが多い。当コンテストで得られた対話中の対話の混乱に関するデータを分析し、対話を修復するための方式を対話戦略ルールとして抽出したい。

4. まとめ

自然言語対話システムにユーザとの対話を通して経路課題を解かせ、その成績によって対話システムの性能を評価しようという試みがWWWページを介して行われた(対話コンテスト DiaLeague'97)。

平均得点

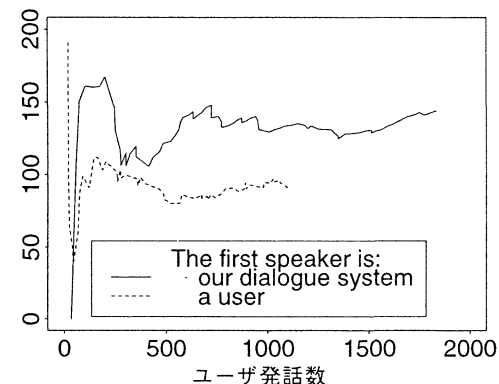


図4 ユーザ発話数と平均得点の関係

本稿では、当コンテストにおいて収集された対話(728対話)のうち、我々の対話システムとユーザとの対話(141対話)を分析し、ユーザ発話の多様性について分析した。その結果、ユーザ発話のパターンの多様さや、ユーザ発話数の増加に伴う異なり形態素数/発話パターン数の増加の程度がわかった。また、発話パターンの増加率が対話の量に影響を与えることもわかった。

今後は、未定義な形態素や発話パターンによる対話の混乱に対して、対症療法的な手法ではなく、対話そのものによって対話の混乱を修復する手法を開発する。

参考文献

- [1] 高間, 土肥, 石塚: 擬人化エージェントにおける音声対話を通じての協調的応答戦略の自動学習, 人工知能学会誌, 12, 3, pp. 456-465 (1997).
- [2] 山田, 溝口, 原田: 質問応答システムにおけるユーザ発話モデルと協調的応答の生成, 情処学論, 35, 11, pp. 2265-2275 (1994).
- [3] 熊本, 伊藤: 支援対話の解析に基づく対話処理方式の提案, 信学論 D-II, J77-D-II, 8, pp. 1492-1501 (1994).
- [4] 橋田, 伝, 長尾, 柏岡, 酒井, 島津, 中野: DiaLeague — 自然言語処理システムの総合評価 —, 人工知能学会誌, 12, 3, pp. 390-399 (1997).
- [5] 妙木裕, 松本裕治, 長尾真: “汎用日本語辞書および形態素解析システム”, 情処第42回全大 (1991).