

話題特化型対話システムの設計と評価 Design and Evaluation of Topic-based Dialogue System

原 大曜[†]
Taiyo Hara

鳥海 不二夫[†]
Fujio Toriumi

石井 健一郎[†]
Kenichiro Ishii

1. はじめに

近年、人間と対話を行う対話システムの研究が盛んに行われている。対話システムはその目的によって大きく2つに分けることができる。一つはカーナビゲーションの音声案内や自然言語による情報検索システムなど、何らかのタスクを達成するために対話を行うタスク指向型対話システム [1] である。もう一つはペトロボットに搭載されるコミュニケーションシステムや人工無能に代表されるような、人間を楽しませることを目的とした非タスク指向型対話システム [2] である。

本研究では非タスク指向型対話システムを対象とし、より人間らしく、楽しい対話を行うことを目指す。そのため、まず対話の一要素である「話題」に着目し、それに関連する知識が対話に与える影響を明らかにする。

2. 対話システム

2.1 システム概要

本研究で設計する対話システムを映画 KELDIC (Ken's Laboratory Dialogue Computer) と呼ぶ。映画 KELDIC は以下の特色を備える。

- 対話方式は1対1のテキスト対話とする
- 対話で扱う話題は映画に特化する
- 知識の量を任意に設定可能とする

ここで知識の量を表す定量的な表現として、知識量を定義する。1単位の知識量とは、ある映画に関連する1組の情報を指す (図1)。

例えば、「知識量5の映画 KELDIC」と言った場合、「5つの映画について関連する情報を対話で利用するように設定された映画 KELDIC」を表す。

映画 KELDIC が扱う映画の情報は、インターネット上の2サイト^{1, 2}から取得したものを映画データベースとして用いる。なお、映画 KELDIC とのテキスト対話における1回の発言を「発話」と呼ぶ。

2.2 対話手法

映画 KELDIC では、ユーザの発話に対応する返答文をあらかじめ定義した発話-返答のペアをスクリプトとして

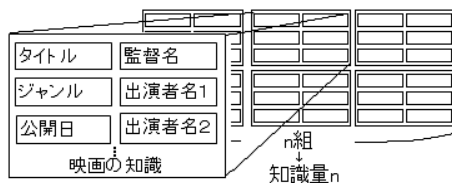


図1: 知識量

[†]名古屋大学情報科学研究科

¹キネマ旬報 <http://www.walkerplus.com/movie/kinejun/>

²all Cinema ONLINE <http://www.allcinema.net/prog/index2.php>

表1: スクリプトの例

キーワード列	返答候補
こんにちは	こんにちは、映画の話をしましょう
\$title, 好き	私も\$titleが好きです。 それはどんな作品ですか？ それは director(\$title) 監督の作品ですね

大量に用意し、これを用いて対話を実現する。スクリプトの例を表1に示す。スクリプトにはキーワード列と呼ぶ自立語 (映画に関する固有名詞を含む) の組と、それに対する1つ以上の返答候補を記述する。また、映画の知識を利用するため、映画に関する単語を扱う仕様を備えている。

例えば表1中の\$titleは映画のタイトルを表す変数であり、director(\$title)は\$titleの監督名を表す関数である。

ユーザの発話から映画 KELDIC が返答するまでの流れは以下の通りである。

1. 映画データベースを参照し、ユーザの発話中に映画に関する単語が含まれている場合はこれを抽出し、変数に置き換える。
2. 形態素解析を行い、発話からキーワード列を検索する。
3. スクリプトからキーワード列に対応する返答候補を検索する。
4. 返答候補に含まれる変数・関数を展開する。
5. 返答候補が複数選ばれた場合は一つをランダムに選択する。

以上の処理をユーザの発話毎に繰り返し、対話を実現する。

3. 実験

3.1 実験概要

対話システムにおける知識量の違いが対話に与える影響を調査するため、映画 KELDIC を用いて対話実験を行う。

対話実験では映画 KELDIC の知識量を10、100、1000、10000の4種類に設定する。1設定につき被験者8名が1回ずつの対話を行い、計32回のテキスト対話を行う。実験は一回につき15分とする。

映画 KELDIC は10122通りのキーワード列に対応するスクリプトを備える。また、扱う映画が有名かどうかを表す指標として、データ収集サイトに寄せられた映画に対するコメントの数を利用する。これは有名な映画ほど、多くの人がコメントを寄せるという仮定に基づいている。映画 KELDIC が対話で扱う映画の、コメント数による内訳を表2に示す。

3.2 対話実験の評価法

対話実験の評価にはアンケートと、映画タイトルの認識率を用いる。

表 2: 映画のコメント数による内訳

コメント数	知識量			
	1	1	8	19
100 以上	1	1	8	19
20 ~ 99	0	10	115	720
10 ~ 19	2	19	140	1220
2 ~ 9	7	70	737	8041
合計	10	100	1000	10000

アンケート

対話直後に被験者は対話の楽しさに関するアンケート調査に回答する。回答は5段階評価とし、0~4点の得点として扱う。これにより被験者の主観による対話の評価を得る。

映画タイトルの認識率

被験者の発話に含まれる映画タイトルを、映画 KELDIC が正しく認識したかどうかを評価する。評価の対象とする発話は次の条件を満たすものとする。

- 被験者の発話に映画タイトルが含まれている
- 発話に含まれる映画タイトルは、対話を通して初めての出現である

このような発話に対し、映画 KELDIC が被験者の意図した映画タイトルを正しく認識した割合をタイトル認識率として、次の式で表す。

$$\text{タイトル認識率} = \frac{\text{映画タイトルを正しく認識した回数}}{\text{条件を満たす発話の総数}}$$

実験後の対話ログからタイトル認識率を計算し、知識量との関係性を評価する。

3.3 実験結果および考察

まず実験で得られたログから対話例を示す。

U:被験者,K:映画 KELDIC(知識量 10000)

U:ハリー・ポッターとアズカバンの囚人 はどうかな?

K:アルフォンソ・キュアロン の作品ですよね??

U:そうなんだ。誰が作ったかなんて全然知らないよ。

K:知ることができてよかったですね

下線部は映画の知識に関わる部分であり、これは映画 KELDIC が知識を利用した返答を行った例である。映画 KELDIC のシステムでは、映画タイトル認識率が向上するほど、対話例のように知識を利用した返答を行う頻度が高くなる。

次にアンケートによる主観評価の結果と知識量との関係を図 2 に示す。縦軸の値は被験者の評価を平均したものである。また、知識量とタイトル認識率の関係を図 3 に示す。知識量が大きくなるほど主観評価、認識率とも高くなる傾向が認められ、認識率が楽しさ(主観評価)を決定する一要因であることが推測される。一方、知識量 1000 から 10000 にかけて大きく知識量が増加したに

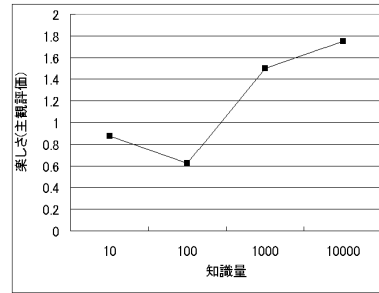


図 2: 知識量と楽しさ(主観評価)の平均の関係

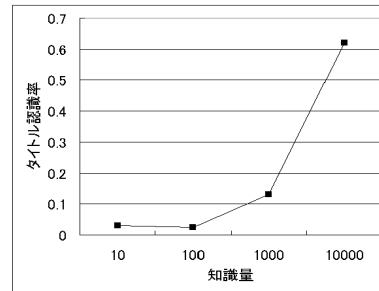


図 3: 知識量とタイトル認識率の関係

も関わらず、楽しさ(主観評価)の上昇はわずかであった。これは表 2 の通り、知識量 10000 における知識の大半が無名な映画であったため、多くの知識は対話の中で活用される機会が少なく、評価に影響を与えなかったと考えられる。すなわち、単に知識量を大きくすることによる評価の向上には限界があることが示唆された。

よって、より楽しい対話を実現するためには、被験者が話題にするような有名な映画について知識を深く持つなどの、知識に対する量以外の要素を導入することが必要と考えられる。

4. まとめ

本研究は人間のように雑談を行う対話システムの設計を目指し、対話における話題とそれに関連する知識に着目した。本研究では映画の話題に特化したテキストベースの対話システムを設計し、知識量を変化させて対話実験を行った。その結果、知識量が大きいほど評価が高くなる傾向が認められたが、上昇には限界があることが示された。今後の課題としては、評価の精度向上のために被験者の母数を増やした評価実験を行うこと、量ではなく質に着目した知識の扱い方の模索などがあげられる。

参考文献

- [1] 河口信夫, 牛窪誠一, 松原茂樹, 岩博之, 梶田将司, 武田一哉, 板倉文忠. 走行車室内音声対話収録システムの開発. Vol. J84-D2 No.6, pp. 909-917, 2001.
- [2] J. Weizenbaum. ELIZA-a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, Vol. 9, No. 1, pp. 36-45, 1966.