

語の共起情報を用いた 非タスク指向型対話エージェントの評価

稲葉 通将^{†1} 鳥海 不二夫^{†1} 石井 健一郎^{†1}

近年、人間と会話のできる対話エージェントが社会に広がりつつある。しかし、人間と雑談をし、話を盛り上げるための非タスク指向型対話エージェントの研究は少なく、知的な応答が可能な非タスク指向型対話エージェントは皆無である。このようなエージェントを実現するためには、客観的・主観的な評価法が不可欠である。本論文では、ある発話が周囲の文脈との間に語彙的結束性を有する場合、その発話は意味的に正しい発話であるという仮定に基づき、対話の意味に踏み込んだ評価手法を提案する。実験の結果、複数の非タスク指向型対話エージェントの性能を比較評価できることが確認された。また、人間による評価との相関が確認された。

Evaluation method of Non-task-oriented dialogue agent using word co-occurrence

MICHIMASA INABA,^{†1} FUJIO TORIUMI^{†1}
and KENICHIRO ISHII^{†1}

Recently, computerized dialogue systems have been actively investigated in various fields. However, There are few researches on the systems to talk with a human, and there are no systems that can make an intellectual reply. In order to realize a practical system, the performance should be evaluated quantitatively. In this paper, we propose a new method that is based on the presumption that a utterance is semantically correct if it has lexical cohesion among a context. It was clarified that a our method can evaluate performance of dialogue system objectively and quantitatively.

^{†1} 名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

1. はじめに

近年、人間と対話可能なコンピュータ（対話エージェント）が社会に浸透しつつあり、広く研究が進められている。ただし、その多くはチケット予約や道案内など特定のタスク達成を目的としたタスク指向型対話エージェントに関するものであり、対話そのものを目的とし、雑談を行う非タスク指向型対話エージェントの研究は少ない。そこで、我々は人間とエージェントの対話に着目し、特にインタビューのように対話を盛り上げる対話エージェントを設計することを目標にする¹⁾。優秀なインタビューは、対話を盛り上げることにより、相手を楽しませ、巧みに様々な話を引き出す。また、そのために対話に際し、あらかじめ十分な調査を行い、質問を練り上げるなど周到な準備を行って対話に望んでいると考えられる。我々はこのようなインタビューを模し、対話の前に対話相手の Web ページ・Blog 等から情報を収集し、そこから大量の発話集（スクリプト集）をあらかじめ作成することによりインタビュー型の対話を行う対話エージェントの作成を目指している。対話エージェントの設計には、対話エージェントの性能の客観的・定量的な評価が不可欠である。タスク指向型対話エージェントには客観的・定量的な評価法がいくつか提案されている²⁾³⁾⁴⁾。一方、非タスク指向型対話エージェントの評価にはアンケートなど、主観的な評価法が主に用いられている。

テキスト対話を対象とし、非タスク指向型対話エージェントを定量的に評価する方法として HMM (Hidden Markov Model) を用いた手法が提案されている¹⁾。そこでは評価対象をテキスト対話に限定し、問題を言語処理に特化している。この方法は、人間同士の対話に対してタグ付与を行い、付与されたタグ系列を学習した HMM の出力確率を計算することにより対話の自然さを評価するものである。この評価法は対話の意味には踏み込まない「浅い構造」だけに着目したものであり、発話間の繋がりという最低限の自然さを評価する手法である。

本論文では、大規模コーパスから抽出した語の共起情報を用い、対話エージェントがどの程度意味的に正しい発話を生成可能かという、「意味」に踏み込んだ評価手法を提案する。

2. 対話エージェント評価法

2.1 概要

本節ではテキスト対話を取り上げ、非タスク指向型対話エージェントの評価法を提案する。なお、ここでいう対話とは、一回の発言を発話と定義したときに、会話の始まりから終

表 1 発話の正しさと前後の文脈

発話番号	発話者	発話
1	A	おとこの夕飯は何を食べましたか？
2	B	今日は何曜日でしたっけ？
3	A	日曜日ですけど、それがどうしました？
4	B	じゃあカレーです。 金曜はカレーと決めているので。

わりまでの発話の系列である。

非タスク指向型対話エージェントの評価尺度としては、対話エージェントの応答の意味的な正しさや、対話の自然さ、対話の楽しさ等、様々な尺度が考えられる。その中で本研究では、応答の意味的な正しさに着目する。応答の意味的な正しさとは、ある応答が直前の発話からみても、前後の文脈からみても論理的に矛盾がなく、意味的に自然であることと定義する。後の文脈を考慮するのは、ある発話がなされた時点ではその発話が論理的に矛盾しており、意味的に正しくないように見える場合でも、その後の文脈から、発話は論理的にも意味的にも正しかったという場面がしばしば見られるからである。

表 1 に例を示す。発話番号 1 の A の発話だけを考慮に入れた場合、発話番号 2 の B の発話は意味的に正しくない発話であるとも考えられるが、発話番号 4 の発話を考慮に入れば、発話番号 2 の発話の意図が理解でき、正しい発話であると判断できる。

現在、自然言語処理によってテキストの意味を同定しようという試みは広く行われており、着実に進歩しているものの、人間による意味理解レベルはもとより、必要とされる理解レベルとのギャップは依然として解消されていない⁵⁾ したがって、発話の意味を把握した上で、発話の意味的な正しさを機械的に判定することは困難である。

そこで、我々は語彙的結束性に着目した。語彙的結束性は結束性を持つ語の出現により、テキストの意味的なつながりを明示する表層的な情報である。Grice の協調の原則⁶⁾ にもあるように、人は対話において、何らかの関連した事柄について互いに話している。そのため、語彙的結束性は新聞記事のようなテキストだけでなく人間同士の対話にも現れる。語彙的結束性は、人間によって書かれた意味のあるテキストに出現する特性であることから、対話において語彙的結束性を有する部分は意味的に正しいやりとりが行われていると考えた。本研究では、この考えを対話エージェントの発話の評価に適用する。つまり、対話エージェントのある発話が前の文脈と語彙的結束性を有する場合、その発話は前の文脈を踏まえた上での適切な発話であると考えられる。また、その発話に対する応答を含む後の文脈と語彙

的結束性を有する場合、対話エージェントの発話が人間の話者にとって理解可能なものであり、何らかの事柄についての対話が継続されたということであるので、発話は意味的に正しいと考えられる。

語彙的結束性の判定には、類義語辞典やシソーラスがよく用いられるが、収録語彙数が少ないため、ドメインが限定されない非タスク指向型対話には適さない。非タスク指向型対話を正確に分析するためには、幅広い語彙をカバーする必要がある。そこで本研究では、大規模コーパスから獲得した共起語を用いる。大規模コーパスから獲得した共起語を使用する場合、意味的な関係の薄い語もノイズとして含まれているという欠点がある。一方、語彙数が非常に多く（今回使用したコーパスでは約 256 万語）、「りんご」から「青森」「赤」のような連想語や「地震」から「火災」「津波」のような原因と結果の関係にある語も獲得できるという利点がある。

2.2 共起情報の獲得

本研究では共起情報の獲得のために Web 日本語 N グラム⁷⁾ を使用した。Web 日本語 N グラムは Google Inc. から提供されている約 200 億文の日本語データから作成された n-gram データであり、長さ 1~7 の単語 n-gram データが収録されている。

本研究では、Web 日本語 N グラムに収録されている単語 n-gram のうち最短の 1-gram と、最長の 7-gram を使用する。1-gram は単独での語の出現頻度を獲得するために、7-gram は 2 語の共起頻度を獲得するためにそれぞれ使用した。なお、共起頻度を獲得する語の品詞は名詞、動詞、形容詞とした。共起関係の強さを測る指標には dice 係数、相互情報量、t スコア、対数尤度比⁸⁾ などがあるが、その中で最も性能が良かった対数尤度比を使用し、対数尤度比の高い 2 語を語彙的結束性を示す共起語対として獲得する。ここで、語 v と語 w の対数尤度比 λ とは、 v と w の 2 単語が互いに従属関係にある場合と独立の場合との最尤推定量による尤度比であり、式 (1) で表される。2 語が従属している度合いが強いほど大きい値を取る。

$$\lambda = 2 \sum_{i=1, j=1}^2 f_{ij} \left\{ \log \frac{f_{ij}}{F} - \log \frac{f_i \cdot f_j}{F^2} \right\} \quad (1)$$

ただし、 $g(v, w)$ を語 v と w の共起頻度、 $h(v)$ を語 v の出現頻度、 F を全文数とすると、 $f_{11} = g(v, w)$ 、 $f_{12} = h(v) - g(v, w)$ 、 $f_{21} = h(w) - g(v, w)$ 、 $f_{22} = F - f_{11} - f_{12} - f_{21}$ である。また、 $f_i = f_{i1} + f_{i2}$ 、 $f_j = f_{1j} + f_{2j}$ である。実験では、 $\lambda \geq 10000$ のものを共起語対とした。ただし、頻出語を含むものについては除外する。

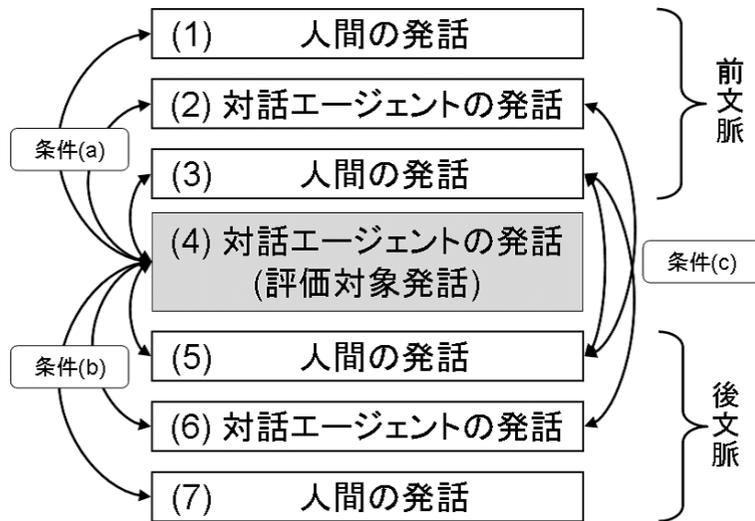


図 1 語彙的結束性の判定

2.3 語彙的結束性の有無の判定

本手法では、ある発話とその前後の文脈との間に語彙的結束性を有する場合、その発話は意味的に正しいという仮定に基づき、対話エージェントがどの程度意味的に正しい発話を生成可能かという性能を評価する。2 発話間の語彙的結束性の有無は共起語対の関係にある語がそれぞれの発話中に存在するか否かによって判定する。共起語対が 2 発話間に存在した場合、その 2 発話には語彙的結束性があるとす。ただし、語彙的結束性を判定する際の 2 発話間の距離には制限を設ける。

語彙的結束性を判定するための条件は

条件(a) 評価対象発話とそれ以前の発話系列 (前文脈) の間に語彙的結束性がある

条件(b) 評価対象発話とそれ以後の発話系列 (後文脈) の間に語彙的結束性がある

条件(c) 前文脈と後文脈の間に語彙的結束性がある

の 3 つとした。実験では、条件 (a) ~ (c) のいずれかを 1 つ以上満たした場合、その発話と周囲の文脈の間に語彙的結束性を有する発話であると判定する。

条件 (a) を満たすとき、対象発話がそれまでの話の流れから逸脱していない発話であることを示し、条件 (b) は対象発話が人間にとって理解可能であったことを示すと考えられる。

条件 (c) は対象発話前後で同じ話題が継続していることを示し、対象発話は同じ話題の対話の継続が可能なものと考えられる。条件 (c) によって、単純な頷きや同意など、名詞、動詞、形容詞のいずれも含まない発話の評価も可能となる。ただし、各条件において、語彙的結束性を判定する 2 発話が共に対話エージェントによる発話の場合は語彙的結束性の判定を行わない。こうすることにより、恣意的に作成された対話エージェント、例えば対話相手がいかなる発話をしようと、発話と発話順序が固定されているような対話エージェントに対し、高い評価を与えることを避け、評価法の頑健性を高める。

長さ n の発話系列 x を $x = x_1, x_2, \dots, x_n$ とし、ある発話 x_i が評価対象発話であるとき、前文脈を示す発話系列 p は $p = p_1, p_2, \dots, p_{i-1} = x_1, x_2, \dots, x_{i-1}$ 、後文脈を示す発話系列 s は $s = s_1, s_2, \dots, s_{n-(i+1)} = x_{i+1}, x_{i+2}, \dots, x_n$ となる。また、隣接する発話の距離を 1 としたとき、語彙的結束性の有無を判定する発話間の距離の最大値を指定するパラメータを d とする。実験では $d = 3$ とした。このとき、発話 x_a と発話 x_b の語彙的結束性を判定する関数 $\phi(x_a, x_b)$ と、 d によって発話間の距離を制限するための関数 $dist(a, b)$ を

$$\phi(x_a, x_b) = \begin{cases} 1 & \text{if } x_a \text{ と } x_b \text{ に語彙的結束性がある} \\ & \text{and} \\ & x_a \text{ と } x_b \text{ の少なくとも一方が人間の発話} \\ 0 & \text{otherwise} \end{cases}$$

$$dist(a, b) = \begin{cases} 1 & \text{if } |a - b| \leq d \\ 0 & \text{otherwise} \end{cases}$$

と定義すると、条件 (a) ~ (c) はそれぞれ以下の (2) ~ (4) のように書ける。

$$\sum_j \phi(x_i, p_j) dist(i, j) \geq 1 \quad (2)$$

$$\sum_j \phi(x_i, s_j) dist(i, j) \geq 1 \quad (3)$$

$$\sum_{j,k} \phi(p_j, s_k) dist(j, k) \geq 1 \quad (4)$$

図 1 の例を用いて説明する。図の「(4) 対話エージェントの発話」が評価対象発話とすると、それより前の発話系列が前文脈 (図中の (1) ~ (3) およびそれ以前の発話系列)、後の発話系列が後文脈 (図中の (5) ~ (7) およびそれ以後の発話系列) となる。 $d = 3$ のとき、条件

表 2 対話エージェント評価法の評価実験に用いた対話

対話の種類	対話数	発話数	評価値 (人手で評価)
(1) 人間同士の対話	20	693	0.99
(2) ELIZA 型 KELDIC と人間との対話	20	1200	0.54
(3) インタビュー型 KELDIC と 人間との対話	20	862	0.66

(a) の式 (2) において $dist(a, b) = 1$ となり, $\phi(x_a, x_b)$ の値が条件判定の結果に影響を与える発話のペアは $\langle(1),(4)\rangle, \langle(2),(4)\rangle, \langle(3),(4)\rangle$ の 3 つとなる. 同じく, 式 (3) では $\langle(4),(5)\rangle, \langle(4),(6)\rangle, \langle(4),(7)\rangle$, 式 (4) では $\langle(2),(5)\rangle, \langle(3),(5)\rangle, \langle(3),(6)\rangle$ が条件判定の結果に影響を与える.

3. 共起情報を用いた対話エージェント評価実験

3.1 概要

第 2.3 節で説明した語彙的結束性の判定条件に基づき, 周囲の文脈と語彙的結束性を有する発話がどの程度存在するかにより対話エージェントの性能を評価する.

実験に使用した対話は表 2 の通りであり, 合計 60 対話であり, 発話数にすると合計 2695 発話である. 対話は人間同士の対話 20 対話 (表 2 の (1)), 人間と対話エージェントとの対話 40 対話 (表 2 の (2)(3)) より構成されており, 以下の制約を課した.

- 30 分のテキスト対話とする
- 1 対 1 で交互に発話する
- 話題は制限しない
- 顔文字や方言は使用しない

人間同士の対話は, 互いに面識のない被験者により行われた. また, 対話エージェントと人間の対話データを収集する際, 人間は同じ対話エージェントと 1 度しか対話は行わないようにした. 対話エージェントには, 我々が考案した対話エージェントの KELDIC (Ken's Laboratory Dialogue Computer)¹⁾ のうち, ELIZA⁹⁾ を模して作られた ELIZA 型 KELDIC と, プロのインタビューを模したインタビュー型 KELDIC¹⁰⁾ を用いた. ELIZA 型 KELDIC はあらかじめ用意された, どのような発話に対し, どのような発話を生成するかという発話生成規則の中から, 対話相手の発話と一致する規則を探し, その規則に応じた発話を返す

対話戦略をとっている. インタビュー型 KELDIC は ELIZA 型 KELDIC を改良した対話エージェントであり, バラエティ番組の司会など, 対話のプロであるインタビューを模した対話エージェントである. インタビュー型 KELDIC は, ELIZA 型 KELDIC のような発話生成規則を用いて発話を生成することに加え, 対話の前に対話相手の情報をもとに大量の発話集 (スクリプト集) を用意して対話を行うインタビュー形式の対話エージェントであり, 情報を聞き出すための対話ではなく, 対話を盛り上げるための対話を目指している.

全ての発話には評価情報が人手によって付与されている. 評価は大学生 3 人によって個別に行われ, 各発話について「意味的に正しい」もしくは「意味的に正しくない」のどちらかを第 1 節で述べた発話の意味的な正しさの定義に従って判定した. 対話エージェントと人間の対話における 3 人の評価の kappa 統計量の平均は 0.61 であり, 比較的高い一致が確認された. 最終的に評価情報として付加されるものは 3 人の多数決によって決定し, 表 2 にその結果が示されている. 人手による評価値は, 対話エージェントによる全発話において, 意味的に正しいと判定された発話の割合である. 対話エージェントの評価は, 人手による評価値と同じように, 提案手法によって語彙的結束性を有すると判定された発話の割合によって行う.

3.2 実験結果

まず, 対話ごとに評価値を計算した場合の結果を示す. 提案手法による評価と人手による評価の比較が図 2 である. グラフ上の各点は 1 つの対話に対応している. 図 2 からわかるように, 人手により高く評価された対話ほど, 提案手法によっても高い値を示している. 相関係数は 0.84 であり, 人間同士の対話を除いた場合でも相関係数は 0.50 であった. 提案手法による評価と人手による評価との間に相関が確認されたことから, 本評価法は妥当であるといえる.

次に, 対話エージェントの比較評価の結果を表 3 に示す. 表 3 の精度と再現率は, $A =$ 評価法によって語彙的結束性を有すると判定された対話数, $B =$ 人手によって意味的に正しいと判定された発話数, $C =$ 評価法によって語彙的結束性を有すると判定された発話数, $D =$ 人手によって意味的に正しいと判定された発話数, としたとき,

$$\text{精度} = \frac{A}{B} \quad \text{再現率} = \frac{A}{C}$$

で計算された値である. 表 3 から明らかなように, 評価値は高い方から順に人間同士の対話, インタビュー型 KELDIC と人間の対話, ELIZA 型 KELDIC と人間の対話の順になっており, 表 2 で示した人手による評価と同様の傾向を示していることから, 本評価法が

表 3 対話エージェント評価法の評価実験結果

対話の種類	評価値	精度	再現率
(1) 人間同士の対話	0.85	1.00	0.85
(2) ELIZA 型 KELDIC と人間との対話	0.25	0.66	0.30
(3) インタビュー型 KELDIC と人間との対話	0.47	0.78	0.52

表 4 人手による評価

評価者	精度	再現率
A	0.88	0.77
B	0.75	0.91
C	0.92	0.86
平均	0.85	0.84

妥当であるといえる。

ELIZA 型 KELDIC は、発話生成規則の適用が失敗し、意味不明な発話を生成したり、人間の発話が発話生成規則に合致しない場合にランダムで発話を生成しており、その発話が意味的に正しくないことが多かった。その際、それに続く人間の発話は「何のことですか?」や「どういう意味ですか?」のような前後の文脈との間に共起語対を含まない発話となる場合が多かったため、最も低い評価値となった。インタビュー型 KELDIC は、発話として質問文を比較的多く生成するため、質問して人間が答え、さらに質問して人間が答える、という状況が生まれることがあった。そのような状況になった際には発話間に語彙的結束性を有する場合が多かった。そのため、インタビュー型 KELDIC は ELIZA 型 KELDIC と比べて高い評価値となったと思われる。しかし、生成する質問がそれまでの話の流れと合致していない場合や、ELIZA 型 KELDIC と同じように生成規則の適用に失敗し、生成された発話が意味不明である場合もあったため、人間ほど高い評価値とはならなかった。

表 4 は、参考として本評価法の結果の代わりに、人手による評価を行った A, B, C の 3 人の評価者の評価結果 (人間同士の対話を除く) を与え、精度と再現率を計算したものである。人手による評価の場合、精度・再現率ともに 0.8~0.9 程度の値となることがわかる。ここから本評価法の精度は比較的高い値が得られたといえる。一方、本評価法の再現率に関しては、表 4 の人手による評価の再現率と比べ、低い値を示した。特に ELIZA 型 KELDIC と人間の対話においてより低い値となった。実際の対話を見ると、ELIZA 型 KELDIC と人間の対話では同じ話題が長く続いているようなことはほとんど無く、ELIZA 型 KELDIC による不適切な発話によって頻りに話題が中断されていた。したがって、正しい発話であっても共起語対が周辺に存在せず、このような低い再現率になったと考えられる。ここまでの精度・再現率の分析から、発話が周囲の文脈と語彙的結束性を有することは、発話が意味的に正しいことの十分条件であるが、必要条件ではないことがわかる。ただし、本研究では全発話について意味的な正しさを厳密に判定することを目的とせず、対話エージェントがどの

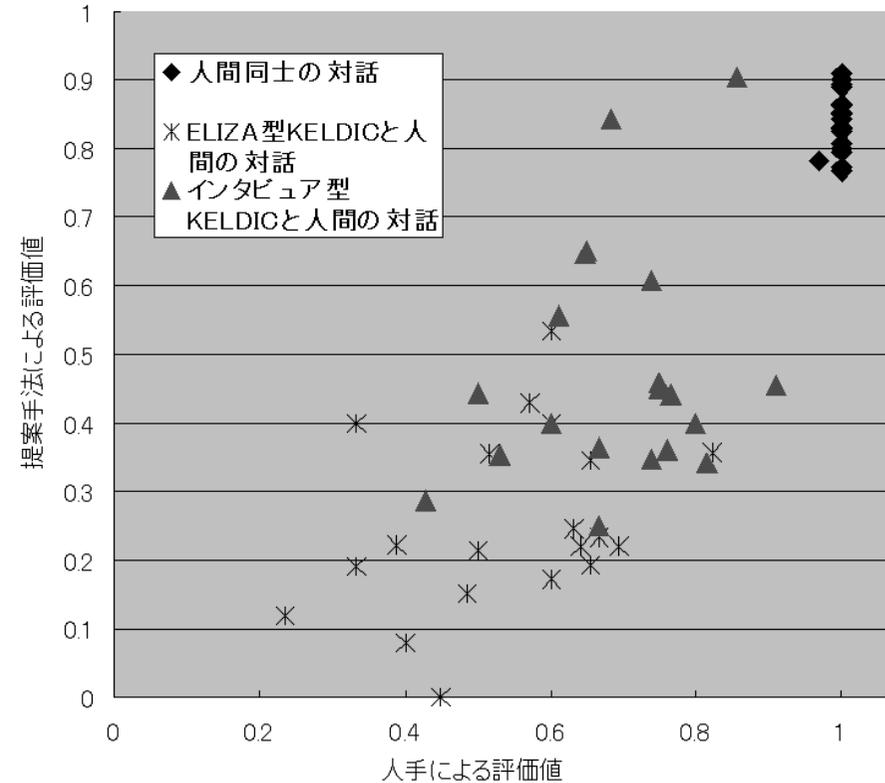


図 2 評価法と人手による評価との比較

程度意味的に正しい発話を生成可能かという傾向を評価することが目的であるため、人手による評価に比べて再現率が低いことは、精度が比較的高い値が得られている点を考慮すると、それほど大きな問題とはならない。

表 5 と表 6 に実際の評価例を示した。表 5 は結束語彙が発話間に存在し、かつ人手による評価も「正しい」で一致している。一方、表 6 は正しく評価できなかった例である。表 6 の発話番号 4 と 6 の発話は、人手による評価が「正しくない」であるにもかかわらず、語彙的結束性があると判定されている。特に発話番号 6 の発話では、質問・時間という意味的に

表 5 正しく評価できた例

発話者	共起語対 (一部)	人手による評価	発話
C (対話エージェント)	休日・買い物 (後文脈)	正しい	あなたの休日の過ごし方って何ですか？
D (人)	-	-	レポートをやったり、 買い物にいったり、 それからバイトですね。
C	買い物・食料 (後文脈)	正しい	おお、それからバイトなんですね。 どんな買い物をするんですか？
D	-	-	いろいろですね、 食料から装飾品まで。
C	装飾品 (前文脈) ・アクセサリ (後文脈)	正しい	なるほどね、いろいろなんですね。 装飾品ってどんな感じですか？
D	-	-	服とか、アクセサリとかですね。

表 6 正しく評価できなかった発話を含む例

発話番号	発話者	共起語対 (一部)	人手による評価	発話
1	F (人)	-	-	見に行った映画が つまらなかったのは ショックでした。
2	E (対話エージェント)	ジャンル ・映画 (前文脈)	正しい	あなたの好きな映画の ジャンルって何ですか？
3	F	-	-	アクション、コメディ等が 好きです。
4	E	コメディ ・映画 (前文脈)	正しくない	コメディ等はどうか？
5	F	-	-	どういう意味ですか？
6	E	質問 ・時間 (後文脈)	正しくない	難しい質問ですね。 話は変わりますが、 何をしますのですか？
7	F	-	-	時間があつたら のんびりしたいです。

関係の薄い語が共起語対として獲得されてしまったために誤判定された。コーパスから自動で獲得した共起語対にはこのようなノイズがいくつか混入しており、同じように誤判定された例がいくつか見られた。共起語対を使うことにより、このような誤判定の問題はあるが、本評価法の精度が比較的高い値であったことから、誤判定は許容できるものと思われる。

4. おわりに

大規模コーパスから獲得した共起情報を用いて非タスク指向型対話エージェントを対話の「意味」の観点から定量的に評価する方法を提案した。本手法では、対話の中で語彙的結束性を有する部分は意味的に正しい対話が行われていると仮定し、まず、大規模コーパスから語彙的結束性を与える共起語対を獲得した。次に共起語対を発話間を含むかどうかによって対話の評価を行った。実験の結果、人手によって性能がよいと評価された対話エージェントほど評価値が高くなることを確認した。また、人手による評価と提案手法による評価値との相関関係を確認した。

今後の課題としては、対話の楽しさや盛り上がりを定量的に評価することが挙げられる。人間を楽しませることを目的とした対話エージェントを設計する際には本研究で評価したような応答が意味的に正しいことに加え、話が盛り上がり、楽しいものである必要がある。

参 考 文 献

- 磯村直樹, 鳥海不二夫, 石井健一郎. HMM による非タスク指向型対話システムの評価. 電子情報通信学会論文誌 D, Vol.92, No.4, pp. 542-551, 2009.
- M.A. Walker, D.J. Litman, C.A. Kamm, and A.Abella. PARADISE: A Framework for Evaluating Spoken Dialogue Agents. *Proceedings of the 35th Annual Meeting of the Association of Computational Linguistics*, pp. 271-280, 1997.
- R.Soricut and E.Brill. A Unified Framework for Automatic Evaluation using N-gram Co-Occurrence Statistics. *Proceedings of ACL*, pp. 613-620, 2004.
- D.Jurafsky and J.H. Martin. *Speech and language processing: an introduction to natural language processing, computational linguistics, and speech recognition*. MIT Press, 2000.
- 石塚満. 自然言語テキストの共通的概念記述. 人工知能学会論文誌, Vol.21, No.6, 2006.
- P.Grice. *Studies in the way of words*. Harvard University Press.
- 工藤拓, 賀沢秀人. Web 日本語 N グラム第 1 版. 言語資源協会発行.
- T.Dunning. Accurate methods for the statistics of surprise and coincidence. *Computational linguistics*, Vol.19, No.1, pp. 61-74, 1993.
- 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.
- 岡田譲二, 鳥海不二夫, 石井健一郎. インタビューを模した対話エージェントのための質問文自動生成. *Proceedings of JAWS 2007*, 2007.