

隣接発話らしさを利用した英語会話文完成問題の回答手法

Solving English dialogue completion problems using utterance adjacency

堂坂浩二^{*1}
Kohji Dohsaka

坂本祐磨^{*1}
Yuma Sakamoto

高瀬惇^{*1}
Jun Takase

^{*1} 秋田県立大学
Akita Prefectural University

This paper presents a solver for English dialogue completion problem using the utterance adjacency. The utterance adjacency shows the certainty that given two utterances appear adjacently in a dialogue. We use phrase pairs as features for utterance adjacency recognition. A phrase pair feature consists of a pair of n-grams with one n-gram taken from the first utterance and one n-gram taken from the second utterance. We trained an utterance adjacency recognizer from the situation dialogue corpus developed by NTT. The situation dialogue corpus contains 6,802 dialogues that have a similar style with those used for dialogue completion problems. The experimental result shows that the solver outperforms our previous solver that uses the flow of utterance intentions. In addition, we discuss how the performance of the solver was influenced by the extension of phrase pairs and by the choice of a dialogue corpus.

1. はじめに

我々は国立情報学研究所が主導する「ロボットは東大に入れるか」プロジェクト [Arai 2014] の英語問題に取り組み、センター試験および東大二次試験において高得点をとることを目指している。2014 年には我々が作成したシステムが「東ロボくん」として代ゼミセンター模試を受験し、全受験者の平均である 93.1 点 (200 点満点) を超える 95 点を達成した [東中 2015]。

筆者らは、なかでもセンター試験問題における会話文完成問題の解法を開発してきた。会話文完成問題とは、図 1 に示すような二人の話者の会話の空所に相応しい文を 4 つの選択肢から選び、会話文を完成させる問題である。この問題では空所 [27] に入る正解は選択肢(1)となる。2014 年の代ゼミセンター模試の際には、発話意図の流れの自然さと感情極性の流れの自然さを使った解法を開発した [東中 2015]。しかし、本論文で議論するように、この方法は数多くの試験問題で評価したところ、安定して正解率をあげることが難しく、特に発話意図の流れの自然さが正解率を下げる傾向をもつことが分かった。

そこで、本研究では、2014 年代ゼミセンター模試の際の解法の問題点を明らかにし、その問題点を解決する新しい解法として、発話意図の流れの自然さの代わりに、隣接発話らしさを使った会話文完成問題の解法を提案し、その評価結果を示す。

2. 発話意図を使った従来法

2.1 従来法の概略

2014 年の代ゼミセンター模試の際に開発した会話文完成問題の解法 [東中 2015] の概略を説明する。以下において、この方法を従来法と呼ぶ。従来法は、会話文完成問題の 4 つの選択肢の各場合について会話文の流れの自然さを推定し、最も自然な流れとなる選択肢を選ぶ。会話文の流れの自然さのスコアは、発話意図 (表明, 評価など) の流れの自然さのスコアと感情極性 (ポジティブかネガティブか) の流れの自然さのスコアの重み付き和として算出した。

発話意図の流れの自然さを認識するため、Switchboard

Taylor: Are you ready to leave for the baseball game?

Akira: Almost! My guitar lesson ran late and I got home ten minutes ago.

Taylor: Sara's on the phone. She's outside the stadium. What should I tell her?

Akira: [27] We'll never find her in the stadium.

Taylor: I'll also say we'll be there in 20 minutes. Is that OK?

選択肢: (1) Ask her to wait at Gate 11.

(2) I'm at a guitar lesson.

(3) Say we're already at the stadium.

(4) We'll meet her inside.

図 1. 会話文完成問題の例(1)

Dialog Act Corpus [Jurafsky 1997] から CRF を用いて会話中の発話意図列の識別器を学習した。素性は発言中の単語 1-gram, 2-gram を使い、クラスは 43 種類の発話意図を用いた。CRF の実装は CRFsuite [Okazaki 2007] を使った。この識別器により発話意図列の生起確率を推定し、選択肢ごとの発話意図列の生起確率の和が 1 になるように正規化したものをスコアとした。

感情極性の流れの自然さを認識するため、SVM を用いて感情極性コーパス [Pang 2004] から発言の感情極性の識別器を学習した。素性は文に含まれる単語 1-gram, 2-gram を用いた。SVM の実装は LIBLINEAR [Fan 2008] を使い、Log-linear モデルを学習した。この識別器を使って発言の感情極性がポジティブあるいはネガティブである確率を推定し、その確率に基づいて感情極性の流れの自然さのスコアを計算した。

図 1 の会話文完成問題の選択肢(1)の場合において、従来法が認識した発話意図と感情極性の流れの一部を図 2 に示す。従来法では発話意図と感情極性を文単位で認識する。発話意図としては、表明, Wh 疑問, YesNo 疑問といった発話意図が認識されている。この問題に関しては、4 つの選択肢のいずれの場合でもこの発話意図の流れとなった。発話意図列の生起確率に基づくスコアは選択肢の場合ごとに異なり、正解の選択肢(1)の場合が最も低いスコアとなった。感情極性に関しては、選択肢(1)の "Ask her to wait at Gate 11" の感情極性がポジティブに認識されているが、周囲の文の感情極性はネガティブと認識されているため、選択肢(1)が最も低い感情極性の流れの自然

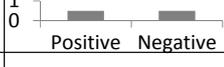
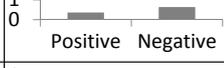
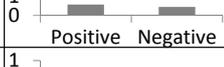
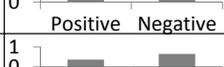
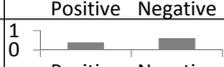
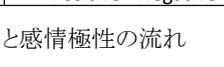
話者	発言	発話意図	感情極性
Taylor	Sara's on the phone.	表明	
Taylor	She's outside the stadium.	表明	
Taylor	What should I tell her?	Wh疑問	
Akira	Ask her to wait at Gate 11.	表明	
Akira	We'll never find her in the stadium.	表明	
Taylor	I'll also say we'll be there in 20 minutes.	表明	
Taylor	Is that OK?	YesNo疑問	

図 2. 従来法における発話意図と感情極性の流れ

さのスコアとなった。結果として、従来法は正解の選択肢(1)を選ぶことができなかった。

2.2 従来法の問題点

従来法の性能の評価結果を示し、従来法の問題点を明らかにする。評価の対象とした会話文完成問題は、大学入試センター試験の本試験過去問 51 問と追試験過去問 47 問 (1991～2015 年度の奇数年度および 2014 年度の本試験と追試験)、代ゼミセンター模試の 18 問 (2013 年第 1～4 回および 2014 年第 1 回)、ベネッセ模試 3 回分の 9 問 (2014 年 6 月, 9 月, 11 月)、および独自に収集したその他の問題 114 問の合計 239 問である。各試験問題は人手により XML フォーマットで入力された電子データで与えられた。

これらの問題を使って従来法の正解率を評価した結果を表 1 に示す。表 1 は、発話意図の流れと感情極性の流れの自然さのそれぞれを単独で使った場合の正解率と、2 つの流れの自然さを組み合わせて使った従来法の正解率を示している。従来法は、これまで報告した通り [松崎 2016], 代ゼミセンター模試に対する正解率は 0.44 であるが、他の模試を含めて 239 問で評価した場合の正解率は 0.21 となり、チャンスレートよりも低い結果となった。大学入試センター試験本試験に対する正解率も 0.25 と低かった。発話意図のみを使った場合と感情極性のみを使った場合を比べると、特に発話意図のみを使った場合の正解

表 1: 従来法の正解率

	発話意図のみ	感情極性のみ	従来法: 発話意図と感情極性
センター本試	0.24 (12/51)	0.24 (12/51)	0.25 (13/51)
センター追試	0.28 (13/47)	0.21 (10/47)	0.21 (10/47)
代ゼミ	0.39 (7/18)	0.44 (8/18)	0.44 (8/18)
ベネッセ	0.11 (1/9)	0.56 (5/9)	0.11 (1/9)
その他	0.15 (17/114)	0.25 (28/114)	0.15 (17/114)
合計	0.21 (50/239)	0.26 (63/239)	0.21 (49/239)

話者	発言	発話意図
A	Thank you.	感謝
B	You're welcome.	会話終了

図 3. 会話例に対する発話意図ラベル付与

率が低いことが分かる。

発話意図の流れの自然さに基づく方法の正解率が低いことの原因として、第一に、発話意図列の識別器の学習コーパスとして使った Switchboard Dialog Act Corpus (SWDA コーパス) [Jurafsky 1997] の発話意図ラベルの粒度が粗いことがあげられる。図 3 は、本研究で学習した識別器を使って、会話例に発話意図ラベルを付与した結果を示す。SWDA コーパスの発話意図ラベルには、感謝に呼応する応答を直接捉える発話意図ラベルが用意されていないため、第 2 文は会話終了という粗い発話意図ラベルが付与されており、感謝とそれに呼応する応答の流れを捉えることが難しいことが分かる。同様の問題として、SWDA コーパスでは Wh 疑問という発話意図ラベルが用意されているが、When, Where, What, Who, Why, How 疑問それぞれに対応する発話意図ラベルは用意されておらず、細かい粒度で Wh 疑問文と応答の流れを捉えることが難しい。

第二の理由として、SWDA コーパスは電話雑談を集めたものであり、必ずしも会話文完成問題とは会話の文体が合っていないことがある。例えば、SWDA コーパスでは、一方の話者が話し続け、もう一方の話者があいづち応答を行うような例が頻出するが、会話文完成問題にはそういった会話はほとんど現れない。

3. 隣接発話らしさを使った提案法

3.1 隣接発話らしさの識別

従来法の問題点を解決するため、発話意図の流れの自然さの代わりに、隣接発話らしさを使った解法を提案する。隣接発話らしさとは、与えられた 2 つの発言が会話の中で隣り合って現れる確からしさを表す。例えば、図 3 に示した 2 つの発言 "Thank you" と "You're welcome" は会話の中で隣り合って現れやすいと考えられる。このような隣接発話らしさを認識することができれば、会話文完成問題の選択肢の良さを評価することができる。

隣接発話らしさを特徴づける素性として、二つの発言それぞれに含まれる単語 1-gram, 2-gram, 3-gram のペア [Jamison 2014] を用いた。この素性をフレーズペアと呼ぶ。隣接発話らしさをフレーズペアという具体的な言語表現で特徴づけることにより、従来法で用いた発話意図列よりも、会話の流れの自然さを捉えることが容易になることが期待される。

また、隣接発話らしさの識別モデルを学習するために、NTT で作成されたシチュエーション対話コーパスを使った。シチュエーション対話コーパスは、会話の場面 (自宅, 職場, ホテル, レストラン, 同窓会など) と話題 (スポーツ, 健康, 気候, ショッピング, 自由話題など) を指定した上で、作業者に会話を作成してもらったものである。作業者 3401 人がそれぞれ 10 発話から成る対話を 2 つ作成し、合計 6802 対話が収集された。このように収集されたシチュエーション対話コーパスでは、会話文完成問題で対象とする会話と類似した文体の会話を集めることができた。従来法で使用した SWDA コーパスと比較して、正解率向上につながることを期待できる。

隣接発話らしさの識別モデルをシチュエーション対話コーパスから SVM を用いて学習した。素性はカイ 2 乗検定を使って上位 20,000 個のフレーズペアを選んだ。SVM の実装は LIBLINEAR [Fan 2008] を使い、Log-linear モデルを学習した。

学習に使った正・負データの作成方法を説明する。シチュエーション対話コーパスにおいて、ある作業者が以下のように会話1と2を作成したとする。ここでは U, V は発言を示す。

会話1 : $U_1, U_2, U_3, U_4, U_5, U_6, U_7, U_8, U_9, U_{10}$

会話2 : $V_1, V_2, V_3, V_4, V_5, V_6, V_7, V_8, V_9, V_{10}$

このとき、正データは次のように一つの会話の中で隣り合う発言のペアから作成した。

正データ1 : $(U_1-U_2), (U_2-U_3), \dots, (U_9-U_{10})$

正データ2 : $(V_1-V_2), (V_2-V_3), \dots, (V_9-V_{10})$

負データは、以下の U_1 と V_2 , V_1 と U_2 のように作成した。

負データ1 : $(U_1-V_2), (U_2-V_3), \dots, (U_9-V_{10})$

負データ2 : $(V_1-U_2), (V_2-U_3), \dots, (V_9-U_{10})$

3.2 提案法の概略

提案法は、会話文完成問題の4つの選択肢の各場合について、隣接発話らしさのスコアと感情極性の流れの自然さのスコアの重み付き和として会話文の流れの自然さを推定し、最も自然な流れとなる選択肢を選ぶ。

提案法では、話者の発言を単位として隣接発話らしさと感情極性を認識する。隣接発話らしさは、空所を含む発言が会話の最後の発言でないなら、空所を含む発言とその直後の発言の間で隣接発話らしさを計算し、さもなければ、空所を含む発言とその直前の発言の間で隣接発話らしさを計算する。選択肢ごとの隣接発話らしさの和が1になるように正規化したものをスコアとした。感情極性の流れの自然さのスコアは従来法と同様に計算した。

図1の会話完成文問題の選択肢(1)の場合において、提案法が認識した隣接発話らしさと感情極性の流れの一部を図4に示す。空所を含む発言とその直後の発言の間のフレーズペアの一部が示されている。例えば、空所を含む発言中の単語"ask"と空所直後の発言中の単語"ok"の間のフレーズペアが抽出されている。この問題の場合、選択肢(1)の場合の隣接発話らしさのスコアが最も高くなった。感情極性のスコアは、従来法と同様に、選択肢(1)が最も低いスコアとなったが、2つのスコアの重み付き和として計算される最終スコアでは、正解の選択肢(1)が最高スコアとなった。

3.3 素性の拡張

隣接発話らしさを捉える特徴量として、単語自体だけでなく単語のカテゴリまで考慮してフレーズペアを拡張することによって、典型的な隣接発話を捉えやすくなる可能性があると考えた。そこで、感情極性リスト[Hu 2004]と固有表現認識[Manning 2014]を用いてフレーズペアを拡張することを試みた。

感情極性リスト[Hu 2004]は、製品レビューデータから収集したポジティブ単語 2006語、ネガティブ単語 4783語の合計 6789語から構成されている。このリストを使って、会話中のポジティブあるいはネガティブな感情を表す単語を感情極性カテゴリ#positive#, #negative#に置き換えた。例えば、次の会話で単語"fine"がポジティブ単語リストに含まれているとする。

A: How are you ?

B: I am fine.

このとき会話を次のように書き換えた。

A: How are you ?

B: I am #positive#.

書き換え後の会話で感情極性カテゴリを含む n-gram のフレーズペアを抽出することにより、フレーズペアを拡張した。これ

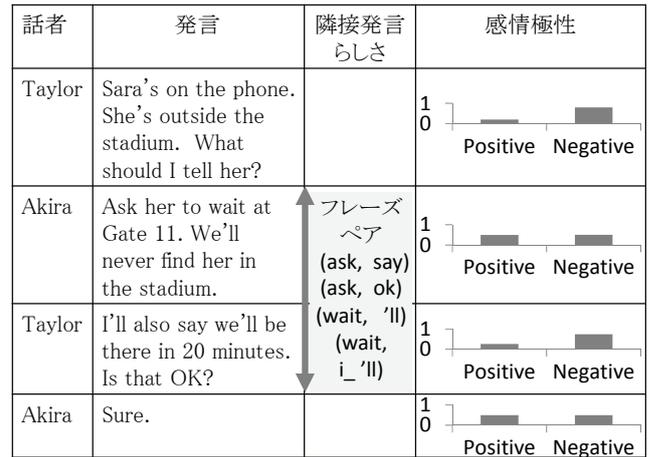


図4. 提案法における隣接発話らしさと感情極性の流れ

により、感情や意見のやり取りにおいて起こる会話の流れを捉えやすくなるのが期待できる。

Stanford Core NLPの固有表現認識 [Manning 2014] を使って、文中の人物、場所、組織、時間、日付、割合、金額、その他の固有表現を認識し、それぞれ固有表現カテゴリ#person#, #location#, #organization#, #time#, #date#, #percent#, #money#, #misc.#に置き換えた。例えば、次の会話で、"New York"が場所の固有表現として認識されたとする。

A: Where are you ?

B: I am in New York.

このとき会話を次のように書き換えた。

A: Where are you ?

B: I am in #location#.

書き換え後の会話で固有表現カテゴリを含む n-gram のフレーズペアを抽出することにより、フレーズペアを拡張した。これにより、where の後には#location#が現れやすいといった会話の流れを捉えやすくなるのが期待できる。

感情極性リストと固有表現認識のそれぞれを使ってフレーズペアを拡張した後、カイ2乗検定を使って上位 20,000個のフレーズペアを選択した。

4. 提案法の評価

2.2節で述べた会話文完成問題 239問を使って提案法の正解率を評価した結果を表2に示す。表2は、隣接発話らしさと感情極性の流れの自然さのそれぞれを単独で使った場合の正解率と、2つを組み合わせ使った提案法の正解率を示している。

会話文完成問題 239問に対する正解率は、従来法が 0.21に対して提案法が 0.39であり、正解率が改善したことが分かる。提案法は、大学入試センター試験、代ゼミセンター模試、ベネッセ模試のいずれの場合も従来法よりも正解率が改善した。隣接発話らしさのみを使った場合と感情極性の流れの自然さのみを使った場合の正解率を比べると、感情極性の流れの自然さもわずかに正解率向上に貢献しているが、隣接発話らしさの方の貢献が大きいことが分かる。発話意図の流れの自然さの代わりに隣接発話らしさを使うことにより、より安定して会話文完成問題を解くことができるようになった。

なお、感情極性の流れのみを使った場合の正解率が表1に示した従来法と異なるのは、従来法では文単位に感情極性を認識するのに対して、提案法では話者の発言を単位として感情極性を認識しているためである。

表 2. 提案法の正解率

	隣接発話のみ	感情極性のみ	提案法: 隣接発話と感情極性
センター 本試	0.37 (19/51)	0.27 (14/51)	0.37 (19/51)
センター 追試	0.30 (14/47)	0.21 (10/47)	0.28 (13/47)
代ゼミ	0.50 (9/18)	0.39 (7/18)	0.50 (9/18)
ベネッセ	0.67 (6/9)	0.33 (3/9)	0.78 (7/9)
その他	0.38 (43/114)	0.30 (34/114)	0.39 (45/114)
合計	0.38 (91/239)	0.28 (68/239)	0.39 (93/239)

表 3. 素性拡張と学習コーパス変更の影響

感情極性リスト によって素性を 拡張した場合	固有表現認識 によって素性を 拡張した場合	隣接発話らしさ の学習コーパス に SWDA コーパス を使った場合
0.20 (47/239)	0.38 (90/239)	0.23 (56/239)

次に、表 3 に感情極性リストと固有表現認識を使って素性を拡張した場合の正解率と、隣接発話らしさの学習コーパスに SWDA コーパスを使った場合の正解率を示す。感情極性リストによって素性を拡張した場合は、正解率が格段に下がった。感情極性カテゴリを含むフレーズペアを導入した結果、識別に有効であった元のフレーズペアが素性として選択されなくなってしまうためと考えられる。固有表現認識によって素性を拡張した場合は正解率にほとんど変化は見られなかった。"where" と "#location#" のフレーズペアなど有用なフレーズペアが抽出されていたが、正解率向上に結びつかなかったのは学習に用いたデータが少なかった可能性もある。隣接発話らしさの学習コーパスに SWDA コーパスを使った場合は正解率が下がった。SWDA コーパスは、シチュエーション対話コーパスに比べて大量の会話データを含むが、会話文完成問題の会話とは文体が合わない会話も多く含むため、成果率を下げた可能性がある。

最後に、提案法で解けなかった問題例を図 5 に示す。この問題では提案法は正解の選択肢(3)を選ぶことができなかった。選択肢(3)の場合の隣接発話らしさを捉えるためには、"make it" (間に合う)と時間の固有表現のフレーズペアを利用することが考えられるが、提案法では、固有表現認識で素性を拡張した場合でも、そのようなフレーズペアを抽出することはできなかった。

5. おわりに

本稿では、隣接発話らしさを使って会話文完成問題を解く解法を提案した。隣接発話らしさを認識するための素性として、フレーズペアと呼ぶ具体的な言語表現に近い素性を使ったこと、会話文完成問題で対象とする会話と類似した文体の会話を収集したシチュエーション対話コーパスを利用したことにより、発話意図の流れの自然さを使った従来法よりも正解率が向上した。

今後の課題としては、会話文完成問題にとって有用な大量の会話データを集める方法を開発すること、会話文特有の発言の流れ(質問-応答、申し出-受諾等)や、構文情報を利用して、会話の流れの自然さをモデル化することがある。

Rachel: Can you drive me to soccer practice today, Dad?
Sam: Yes. It's at six o'clock, right?
Rachel: No. Our coach wants us there at 5:30 today. [28]
Sam: No. I have a meeting until five o'clock.
Rachel: Please, Dad. My coach said it's important.
Sam: Sorry, Rachel. You should ask Mom.
選択肢: (1) Can I play soccer today?
(2) Can my coach drive me?
(3) Can you make it?
(4) Can you practice with me?

図 5. 会話文完成問題の例(2)

謝辞

本研究を推進するにあたって、大学入試センター試験問題のデータをご提供下さった独立行政法人大学入試センターおよび株式会社ジェイシー教育研究所に感謝いたします。実験データをご提供くださいました学校法人高宮学園、株式会社ベネッセコーポレーションに感謝いたします。また、NTT コミュニケーション科学基礎研究所にはシチュエーション対話コーパスをご提供いただきました。謹んで感謝の意を表します。

参考文献

- [Arai 2014] Noriko H. Arai and Takuya Matsuzaki: The impact of A.I. on education—Can a robot get into the University of Tokyo?, In Proc. ICCE, 1034-1042, 2014.
- [Fan 2008] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin: LIBLINEAR: A library for large linear classification, Journal of Machine Learning Research, 9, 1871-1874, 2008.
- [東中 2015] 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩: センター試験における英語問題の回答手法, 言語処理学会第 21 回年次大会, 2015.
- [Hu 2004] Mingqing Hu and Bing Liu: Mining and Summarizing Customer Reviews, Proc. KDD-2004, 2004.
- [Jamison 2014] Emily K. Jamison and Iryna Gurevych: Adjacency Pair Recognition in Wikipedia Discussions using Lexical Pairs, Proc. PACLIC2014, 479-488, 2014.
- [Jurafsky 1997] Daniel Jurafsky, Elizabeth Shriberg and Debra Biasca: Switchboard SWBD-DAMSL shallow discourse-function annotation coders manual, Draft 13, University of Colorado, Boulder, Institute of Cognitive Science, Technical Report, 97-02, 1997.
- [Manning 2014] Christopher D. Manning, Mihai Surdeanu, John Bauer, Jenny Finkel, Steven J. Bethard and David McClosky: The Stanford CoreNLP Natural Language Processing Toolkit, Proc. ACL'14: System Demonstrations, 55-60, 2014.
- [松崎 2016] 松崎拓也, 横野光, 宮尾祐介, 川添愛, 狩野芳伸, 加納隼人, 佐藤理史, 東中竜一郎, 杉山弘晃, 磯崎秀樹, 菊井玄一郎, 堂坂浩二, 平博順, 南泰浩, 新井紀子: 「ロボットは東大に入れるか」プロジェクト: 代ゼミセンター模試タスクにおけるエラーの分析, 自然言語処理, 23(1), 119-159, 2016.
- [Okazaki 2007] Naoaki Okazaki: CRF suite: a fast implementation of conditional random fields (CRFs), 2007.
- [Pang 2004] Bo Pang and Lillian Lee: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proc. ACL'04, 2004.