

発言間の首尾一貫性を利用した会話文完成問題解法 Solving English dialogue completion problems using utterance coherence

高瀬 惇[†] 堂坂 浩二^{*}
Jun Takase Kohji Dohsaka

1. はじめに

近年、人工知能や自然言語処理技術の向上を目的として、人間が行う試験やコンテンツにチャレンジする研究プロジェクトが盛んに行われている。その例として人工知能により小説の創作を目指す「きまぐれ人工知能プロジェクト作家ですよ」[1]や、大学入学試験問題の解法の開発とそれを通じた人工知能の要素技術の発展を目指す「ロボットは東大に入れるか」プロジェクト[2]などがある。

筆者らは、「ロボットは東大に入れるか」プロジェクトの英語班として、センター試験英語問題の一つである会話文完成問題を解くプログラムの開発に取り組んできた。会話文完成問題とは、図 1 のように、2 人の話者による会話の中に空所があり、その空所に入る文を 4 つの選択肢の中から 1 つ選んで、会話文を完成させる問題である。この問題では選択肢(2)が正解となる。

我々は従来、隣接発話らしさと感情極性の流れの自然さの重み付き和で選択肢のスコアをつけ、スコアの一番高い選択肢を解答として選ぶ解法を提案してきた[3]。しかし、本論文で議論するように、この従来法では隣接している発言同士の関係しか捉えることができず、空所を含む発言とは隣接していない発言に解答のヒントがある問題においては正解を導くことが難しくなっている。

そこで、本論文では、従来法の問題点を解決する新しい解法として、隣接発話らしさと感情極性の流れの自然さに加えて、発言間の首尾一貫性を利用した会話文完成問題の解法を提案し、解法の正解率を評価する。

2. 隣接発話らしさを利用した従来法

2.1 従来法の概略

これまでに開発された隣接発話らしさを利用した会話文完成問題の解法 [3] 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、会話文完成問題の各選択肢における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。会話の流れの自然さは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算出する。

隣接発話らしさとは 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、2 つの発言”How are you?”と”I’m fine.”は会話の中で隣り合って現れやすいと考えられる。隣接発話らしさを特徴づける素性として、2 つの発言それぞれに含まれる単語 1-gram, 2-gram, 3-gram のペア [4] を用いた。この素性をフレーズペアと呼ぶ。隣接発話らしさの識別モデルは、NTT で作成された会話文完成問題と類似した形式の会話(6802 対話)を集めたシチュエーシ

David : I think I need to start exercising again. I didn't much all winter.
Ruth : I thought you said you go for a long walk every day.
David : I try to.
Ruth : Well, now that the weather is better, you have no excuse not to walk!
選択肢 : (1) Actually, I don't usually walk in the spring.
(2) But when it's cold and snowy, I get lazy.
(3) Exercising in the winter keeps me warm.
(4) In fact, I really like walking in the snow.

図 1 会話文完成問題の例

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 [5]を用い、Log-linear モデルを学習した。隣接発話らしさは、空所を含む発言とその直後の発言との間で隣接発話らしさを計算する。空所を含む発言が会話の最後の発言の場合は、空所を含む発言とその直前の発言と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 と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とは、連続する発言の間では感情極性(ポジティブかネガティブか)の確率分布の変化が小さいほど、自然な会話の流れであるという考え方に基づくスコアである。感情極性の流れの自然さを認識するために、感情極性コーパス [6] から SVM を用いて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 [5] を用い、

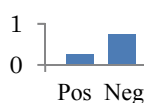
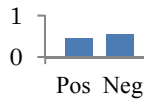
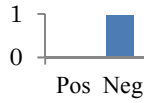
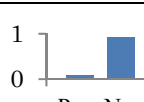
話者	発言	隣接発話らしさ	感情極性
David	I think I need to start exercising again. I didn't much all winter.		
Ruth	I thought you said you go for a long walk every day.		
David	I try to. [But when it's cold and snowy, I get lazy.]	フレーズペア (cold, walk)	
Ruth	Well, now that the weather is better, you have no excuse not to walk!	(snowy, weather)	

図 2 従来法が認識した隣接発話らしさと感情極性の流れ

[†] 秋田県立大学電子情報システム学科

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であるかネガティブであるかの確率を推定し、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 2 は、図 1 の会話文完成問題の選択肢(2)の場合において、従来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示したものである。空所を含む発言中の単語“cold”と空所を含む発言の直後の発言中の単語“walk”の間のフレーズペアが抽出されている。感情極性として各発言の感情極性(Positive, Negative)の確率分布が示されている。この問題の正解は選択肢(2)となっているが、感情極性のスコアは選択肢(1)が一番大きくなり、隣接発話らしさのスコアは選択肢(4)が一番大きくなっている。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、選択肢(4)が最高スコアとなった。結果として従来法は正解の選択肢(2)を選ぶことができなかった。

2.2 従来法の問題点

図 1 の問題に従来法を適用したときの結果から従来法の問題点を明らかにする。図 1 の問題の正解は選択肢(2)である。この問題は、一番目の発言と各選択肢との間に、運動と季節に関する話題について述べているという共通点がある。人がこの問題を解く際、各選択肢において空所を含む発言と一番目の発言を比べて、解答として選択肢(2)を選ぶと考えられる。選択肢(1)は春の話をしており、一番目の発言と季節が異なるため矛盾している。選択肢(2)は冬に運動をしないという点で矛盾がなく自然な流れとなっている。選択肢(3)は冬に運動をするという点で矛盾している。選択肢(4)は冬に歩くことが好きという点で矛盾している。

以上のように、この問題では空所を含む発言と一番目の発言との間の関係を捉える必要があるが、隣接発話らしさでは隣り合った発言同士の関係しか捉えることができないため、一番目の発言と空所を含む発言との間の関係を捉えることができない。また、感情極性の流れの自然さにおいても隣り合った発言との間で感情極性の流れが自然であるかどうかを計算しているため、一番目の発言と空所を含む発言の間の関係を捉えることができない。そのため、隣接発話らしさと感情極性の流れの自然さのスコアを利用する従来法では、図 1 のような離れた発言同士の関係を捉える必要がある問題において正解を導くことが難しい。

3. 発言間の首尾一貫性を利用した解法

3.1 提案法の概略

従来法の問題点を解決するため、隣接発話らしさと感情極性の流れの自然さに加えて、発言間の首尾一貫性を利用した解法を提案する。提案法では会話文完成問題の 4 つの選択肢の各場合について、発言間の首尾一貫性のスコアと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から会話文の流れの自然さを推定し、最もスコアの高い選択肢を自然な流れとなる解答として選ぶ。

発言間の首尾一貫性とは、2 つの発言の流れが自然であり、矛盾していないことを意味する。ただし、2 つの発言は会話中で隣接している必要はない。

発言間の首尾一貫性のスコアは各選択肢に関して以下の 2 つの処理を行い計算する。

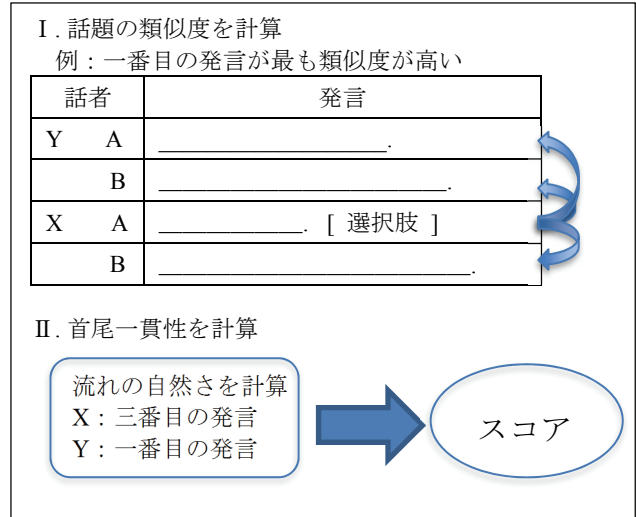


図 3 発言間の首尾一貫性の計算の流れ

I. 空所を含む発言を X とする。X と他の発言との間で話題の類似度を計算し、最も類似度が高い発言を Y とする。

II. 発言 X と発言 Y との間で発言間の流れの自然さを計算し、発言間の首尾一貫性のスコアとする。

図 3 は空所を含む発言と話題の類似度が最も高い発言として一番目の発言が選ばれた場合の発言間の首尾一貫性の計算の流れを示したものである。

空所を含む発言と話題が似ている発言を選択し、空所を含む発言と選択した発言の間で発言間の流れの自然さを計算することで、図 3 のように隣接していない発言同士の関係も捉えることができるのではないかと期待される。

3.2 発言間の首尾一貫性の計算

発言間の首尾一貫性の計算方法について詳しく説明する。発言間の話題の類似度は、word2vec モデルを利用して各発言のベクトルを計算し、二つの発言間のベクトルの類似度を計算することで求められる。発言間のベクトルの類似度の計算法として 2 つの方法を用いた。1 つ目の方法を W2V 法と呼ぶ。W2V 法では、文に含まれる単語のベクトルに単語の重要度によって重みを付与した上で、発言内の単語の平均ベクトルを計算して、それを発言のベクトルと定義する。その後、二つの発言のベクトルのコサイン類似度を計算し、計算結果を発言間の話題の類似度とする。word2vec のモデルは 30 億単語から成る Google News コーパス から学習した 200 次元の既存モデル [7] を利用した。単語の重要度としては wikipedia の英語テキストから計算した tf-idf 値を用いた。

2 つ目の方法を WMD 法と呼ぶ。WMD 法では、文間の距離を計算するために Word Mover's Distance [8]を用いた。Word Mover's Distance では、word2vec モデルを使って、2 つの文中でベクトルの距離が近い単語同士を対応付け、発言内における単語間のベクトルの距離の総和を計算し、計算結果を発言間のベクトルの距離とする。発言間の話題の類似度は以下の式で求める。

$$\text{類似度} = \frac{1}{1 + \text{ベクトルの距離}}$$

WMD の実装にはプログラム [9] を用いた。Word2vec のモデルは W2V 法と同じモデルを利用した。

発言間の流れの自然さの計算には感情極性の流れの自然さの計算を用いた。二つの発言の感情極性の確率分布を P, Q としたとき感情極性の流れの自然さは JS ダイバージェンスを用いて以下の式で計算する。

$$D_{JS}(P||Q) = \frac{1}{2} \left(\sum_x P(x) \log \frac{P(x)}{R(x)} + \sum_x Q(x) \log \frac{Q(x)}{R(x)} \right)$$

$$R(x) = (P(x) + Q(x))/2$$

感情極性の流れの自然さ = $1 - D_{JS}$

この計算から求められた感情極性の流れの自然さを発言間の首尾一貫性のスコアとする。感情を認識するための認識器は従来法と同様の感情極性認識器を用いる。

4. 提案法の評価

評価のため、大学入試センター試験の本試験及び追試験の過去問、代ゼミセンター模試、ベネッセ模試、独自に収集したその他の問題を合わせて、合計 241 問の試験問題を使った。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として与えられた。241 問のうち 163 問を開発データセット、78 問をテストデータセットとした。開発データセットを使って最も正解率が高くなるような最適なパラメータを決めた後、テストデータセットでの正解率を評価した。

隣接発話らしさ、感情極性の流れの自然さ、発言間の首尾一貫性のスコアのそれぞれを単独で用いたときの開発データセットでの正解率とテストデータセットでの正解率を表 1 に示す。表 1 では開発データで最も正解率が高くなるように素性の数を調整している。表 1 を見ると隣接発話らしさのみを用いたときの正解率は 0.37 となっているが、感情極性の流れのみを用いたときの正解率は 0.27、首尾一貫性のみを用いたときの正解率は W2V 法では 0.30、WMD 法では 0.27 となっており、チャンスレートと大きな差がなく、低い正解率であることが分かる。このことから感情極性の流れの自然さのスコアと発言間の首尾一貫性のスコア単独では解ける問題に限られているのではないかと考えられる。

隣接発話らしさ、感情極性の流れの自然さ、発言間の首尾一貫性の重み付き和をスコアとした提案法と従来法を用いたときの開発データセットでの正解率とテストデータセットでの正解率を表 2 に示す。表 2 を見ると、従来法の正解率は 0.40 となっており、表 1 の隣接発話らしさのスコアと感情極性のスコアを単独で用いたときの正解率より高くなっている。このことからスコアを組み合わせることにより、正解率が向上することが分かる。提案法では、隣接発話らしさと発言間の首尾一貫性を組み合わせると、W2V 法と WMD 法のいずれも正解率が 0.38 となっており、従来法よりも低く、隣接発話らしさのスコア単独での正解率との間で大きな変化がなかった。しかし、提案法において、隣接発話らしさ、感情極性の流れの自然さ、発言間の首尾一貫性の 3 つの重み付き和のスコアを使った場合、W2V 法を用いたときのテストデータでの正解率が従来法よりも高い 0.44 となっており、WMD 法を用いたときのテストデータの正解率も 0.42 となった。結果として提案法では従来法と比較して正解率の向上が見られた。このことから発言間の首尾一貫性と隣接発話らしさと感情極性の流れの自然さ

表 1 各スコア単独の評価結果

解法	開発データ	テストデータ
隣接発話のみ	0.36 (59/163)	0.37 (29/78)
感情極性のみ	0.28 (46/163)	0.27 (21/78)
首尾一貫性 (W2V 法)のみ	0.26 (42/163)	0.30 (23/78)
首尾一貫性 (WMD 法)のみ	0.23 (37/163)	0.27 (21/78)

表 2 従来法と提案法の評価結果

解法	開発データ	テストデータ
従来法 ・隣接発話らしさ ・感情極性の流れ	0.37 (60/163)	0.40 (31/78)
提案法 ・隣接発話らしさ ・首尾一貫性(W2V 法)	0.37 (60/163)	0.38 (30/78)
提案法 ・隣接発話らしさ ・首尾一貫性(WMD 法)	0.34 (56/163)	0.38 (30/78)
提案法 ・隣接発話らしさ ・感情極性の流れ ・首尾一貫性(W2V 法)	0.36 (58/163)	0.44 (34/78)
提案法 ・隣接発話らしさ ・感情極性の流れ ・首尾一貫性(WMD 法)	0.34 (55/163)	0.42 (33/78)

を組み合わせることによって、正解率を上げることができないのではないかと考えられる。

5. 考察

提案法で解けなかった問題例を示し、その原因について考察する。

5.1 首尾一貫性の判別方法

提案法では、感情極性の流れの自然さを使って発言間の首尾一貫性を判別していたが、季節などの時間の違いや未来・過去などの時制の違いなどから首尾一貫性を判別すべき問題があることが分かった。図 1 の問題の正解は選択肢 (2) である。この問題に発言間の首尾一貫性を適用したところ、選択肢 (1) と (2) を空所に当てはめた場合、空所を含む発言と話題の類似度が最も高い発言として、問題の一番目の発言が選ばれた。しかし、一番目の発言と選択肢 (1)、選択肢 (2) はいずれも感情極性が Negative となっており、感情極性の流れの自然さを計算したところ、正解である選択肢 (2) よりも選択肢 (1) の方が高いスコアになってしまった。人がこの問題を解く場合、一番目の発言では冬の話をしており、選択肢 (1) では春の話をしていることから一番目の発言と選択肢 (1) の内容は季節が一致しておらず首尾一貫していないと考え、選択肢 (1) を答えとして選ばない。このことから発言間の首尾一貫性を用いて様々な問題を解くためには感情極性の流れの自然さの以外にも時間の違いや時制の違

Waiter : So, are you ready to order?
 Customer : No, not really. What do you recommend?
 Waiter : The shrimp pasta is very good!
 Customer : Really?
 選択肢: (1) Eventually, I'll get over that.
 (2) I guess I'll go on with that.
 (3) I'll have that then.
 (4) OK, so then I'll be that.

図 4 場面の知識を利用する問題

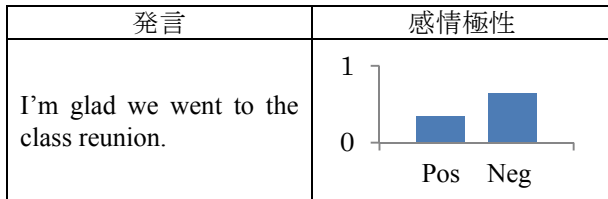


図 5 感情極性の認識結果

いなどを利用して首尾一貫性を判別する方法を開発する必要があると考えられる。

5.2 場面の知識を使う問題

常識や場面の知識を使って解答する問題など、発言間の首尾一貫性のスコアを適用すべきでない問題があることが分かった。図 4 は場面の知識を用いて解答すべき問題である。図 4 は客がウェイターにお薦めの料理を尋ね、薦められた料理を注文するという内容になっている。人がこの問題を解く場合、レストランでの場面の知識を利用して解答として「料理の注文」をする選択肢(3)を選ぶと考えられる。この問題に発言間の首尾一貫性を適用したところ、すべての選択肢において二番目の発言が空所を含む発言との話題の類似性が高い発言として選ばれた。各選択肢において空所を含む発言との首尾一貫性を計算したところ、選択肢(1)のスコアが一番高くなり、答えとして選択肢(1)を選択してしまった。このことから発言間の首尾一貫性は場面の知識や常識などを利用して解答を決める問題に適しておらず、図 4 のような問題には発言間の首尾一貫性のスコアを適用すべきではないのではないかと考えられる。

5.3 感情極性の認識精度

感情極性の認識が適切でない場合があることが分かった。図 5 に感情極性を適切に認識できなかった発言と、認識した感情極性を示す。図 5 の発言には「glad」という単語があるためポジティブな内容の発言であると考えられるが、感情極性の認識結果はポジティブが 0.35、ネガティブが 0.65 となり、ネガティブな発言として認識されていた。このことから現在の感情極性の認識器では感情極性の認識が適切でない場合があることが考えられる。

6. おわりに

本稿では、センター試験の英語会話文完成問題において、隣り合っていない発言同士の関係を捉えることができないという従来法の問題点を解決するために、隣接発話らしさと感情極性の流れの自然さに加えて、発言間の首尾一貫性を利用した解法を提案した。提案法を評価した結果、発言間の首尾一貫性のスコアを単独で用いた場合は正解率が低

くなることが分かった。このことから発言間の首尾一貫性で解くことができる問題が限られているのではないかと考えられる。従来法と提案法を比較したところ、隣接発話らしさと感情極性の流れの自然さと発言間の首尾一貫性の重み付き和のスコアを利用した場合において、従来法よりも高い正解率を得ることができた。このことから発言間の首尾一貫性を用いることにより、会話文完成問題を解く解法の性能を向上させることができるのではないかと考えられる。

今後の課題として、まず感情極性の流れの自然さ以外にも発言間の首尾一貫性を判別する方法を開発する必要がある。次に、発言間の首尾一貫性では解ける問題が限られているため、発言間の首尾一貫性のスコアを適用できる問題のパターンを見つけることが必要である。また、感情極性の認識器の精度を上げるため、学習データの数を増やすことが課題としてあげられる。

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。試験問題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

- [1] 松原仁, ショートショートの新入らしさに関する考察, 2014 年度人工知能学会大会, IC2-OS-14a-1 (2014).
- [2] 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, Vol.23, No.1, pp.119-159 (2016).
- [3] 堂坂浩二, 坂本祐磨, 高瀬惇, 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, 2016 年度人工知能学会大会, 1K3-4 (2016).
- [4] Emily K. Jamison and Iryna Gurevych, Adjacency Pair Recognition in Wikipedia Discussions using Lexical Pairs, Proc. PACLIC, pp.479-488 (2014).
- [5] 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, Vol.9, pp.1871-1874 (2008).
- [6] Bo Pang and Lillian Lee, A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts, Proc. ACL'04 (2004).
- [7] word2vec-GoogleNews-vectors, <<https://github.com/mnihaltz/word2vec-GoogleNews-vectors>>, 参照日時 2017/02/08 12:00.
- [8] Matt J. Kusner, Yu Sun, Nicholas I. Kolkin and Kilian Q. Weinberger From Word Embeddings To Document Distances, Proc. ICML, pp.3-4 (2015).
- [9] Vlad Niculae, Word mover's distance classification in Python, <<http://nbviewer.jupyter.org/github/vene/vene.github.io/blob/pelican/content/blog/word-movers-distance-in-python.ipynb>>, 参照日時 2017/03/10 12:00.