

相談者の期待の種類を考慮する恋愛相談対話システムの検討

Love Counseling Dialogue System Considering User Expectations

中下咲帆¹ 藤後英哲¹ 菊池英明¹ 藤倉将平²

NAKASHITA Sakiho¹, TOGO Eitetsu¹, KIKUCHI Hideaki¹ and FUJIKURA Shohei²

¹ 早稲田大学

¹ WASEDA University

² 株式会社サイシキ

² Saishiki, Inc.

Abstract: 本研究では恋愛相談対話システムの構築に向けて、応答分類モデルを機械学習により構築し、精度評価を行った。モデルの学習には、情緒的支援と情報提供のラベルが付与された Empathetic-Dialogues (共感的対話データ) と Yahoo!知恵袋の恋愛相談データを使用した。2名の評価者によって付与されたラベルの一致率はカッパ係数が 0.49 で、適度な一致 (Moderate) となった。また、作成した全てのラベル付きデータを学習した応答分類モデルの正解率は 0.68, F1 スコアは 0.67 となった。

1. はじめに

1.1. 背景

2022年時点では、日常生活のサポートとして悩みの相談相手や聞き役として活躍する対話システムが増加している[1][2][3]。また、悩みを人間よりもチャットボットに相談したいと考える人が八割以上に達したという調査結果[4]もあり、相談対話システムが必要とされている。

人間同士の相談場面を分析した先行研究[5]では、相談者の期待が情緒的支援と情報収集に集約されることが示されている。これを踏まえ、従来の先行研究では、期待される応答の一部を定型表現で返答する試みが多く行われてきた[6][7][8]。しかし、定型表現では応答の多様性が低く、相談者の期待を十分に満たすことは難しい。

1.2. 目的

本研究では、情緒的支援と情報収集への期待を考慮し、多様な応答を返せる恋愛相談システムの構築を目的とする。ただし、今回は目的の第一段階として、恋愛相談システムに含まれる応答分類モデルの構築のみを行う。情緒的支援と情報収集への期待を考慮した対話の例を図1と図2に示す。



図1. 情緒的支援への期待を考慮した対話例

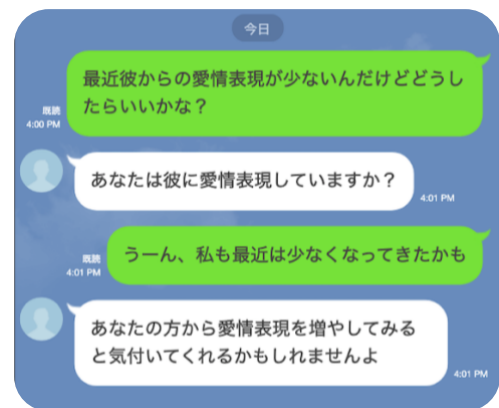


図2. 情報収集への期待を考慮した対話例

2. 提案手法

2.1. 概要

本章では、1章に示した目的を達成するための手法について述べる。はじめに、本研究におけるシステムのフローを図3に示す。

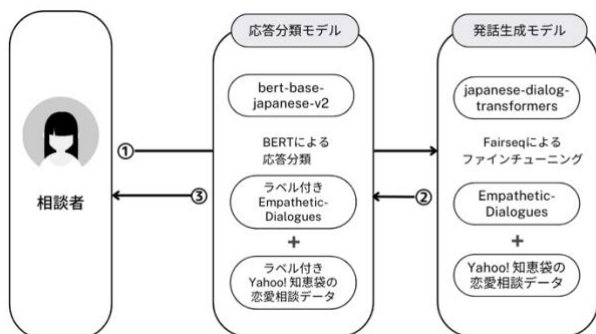


図3. システムのフロー

提案システムではまず、入力として相談者からの悩み文が発話生成モデルに渡される（図3①）。続いて、発話生成モデルでは入力に対する応答候補が出力され、BERTで構築した応答分類モデルに渡される（図3②）。最後に、応答分類モデルで相談者の期待に対応したカテゴリの応答文を選択し、選択された応答が相談者に返される（図3③）。

2.2. 応答分類モデル

応答分類モデルの学習時に使用する教師データは、情緒的支援の応答が多くみられると想定される Empathetic Dialogues [9]と、情報提供の応答が多くみられると想定される Yahoo!知恵袋[10]の恋愛相談データを使用した。表1と表2にそれぞれのデータ例を示す。

表1. Empathetic Dialogues のデータ例

話者	対話
A	ここだけの話にしてね。明日ね、同じクラスの男子に告白するの！
B	すごい、勇気あるねー。いい感じに進んでるってこと？

表2. Yahoo! 知恵袋恋愛相談のデータ例

話者	対話
A	元カノと復縁したいです。この気持ちはいつ伝えるべきでしょうか。
B	思ったことはすぐ伝える方が後悔しないですよ。

さらに、これらの教師データには情緒的支援・情報提供・その他のデータが混在しているため、それぞれラベリングを行った。表3にラベル付きデータの例を示す。ラベルは、情緒的支援であれば1、情報提供であれば2、その他であれば3とする。複数のカテゴリに該当するデータがある場合は、複数回答を許している。

表3. ラベル付きデータの例

ラベル	話者	対話
	A	ここだけの話にしてね。明日ね、同じクラスの男子に告白するの！
1	B	すごい、勇気あるねー。いい感じに進んでるってこと？
	A	元カノと復縁したいです。この気持ちはいつ伝えるべきでしょうか。
2	B	思ったことはすぐ伝える方が後悔しないですよ。
	A	友達がさ、親友だった子に彼氏奪われたって泣きながら電話かけてきたんだよね。ひどいと思わない？
1,2	B	そうだね。でもその子は、友情を天秤にかけてでも奪いたいほど好きだったんだろうね。

各データに対して2名の評価者がラベリングを行い、評価者数はのべ60名となった。作成したラベル付きデータのカテゴリ毎のデータ数を調べた結果、以下の表4と表5に示す割合となった。

表4. ラベリング結果（1回目）

データ	情緒的支援	情報提供	情緒的支援かつ情報提供
Empathetic-Dialogues	428	60	108
Yahoo!知恵袋恋愛相談データ	17	296	91

表 5. ラベリング結果 (2 回目)

データ	情緒的 支援	情報 提供	情緒的 支援かつ 情報提供
Empathetic-Dialogues	492	55	39
Yahoo!知恵袋 恋愛相談データ	39	302	58

続いて、上のラベル付きデータを使用し応答分類モデルの学習を行なった。以下では使用した学習済みモデルと学習時の設定について説明する。

(1) 使用した学習済みモデル

学習には、BERT の事前学習済みモデルの出力層に分類用のレイヤーを追加したネットワーク構成である BertForSequenceClassification [11] を使用した。BertForSequenceClassification に前処理済みの学習データを入力し、応答分類モデルを構築する。BERT の事前学習済みモデルは多数存在するが、今回は日本語の文章分類を目的としているため、2022 年時点で高精度な日本語事前学習モデルである bert-base-japanese-v2 [12] を使用した。

(2) 学習時の設定

学習時にはハイパーパラメータとして最適化手法、バッチサイズ、エポック数を指定する必要がある。今回は最適化手法を AdamW(Adam Weight Decay fix) とした。また、バッチサイズとエポック数は学習データによって調整した。使用した学習データの種類は 3 種類である。一つ目は作成した学習データ全て、二つ目は 2 名の評価者間でラベルが完全一致した学習データ、三つ目は 2 名の評価者間で単一のラベルのみが完全一致した学習データである。

3. 人間同士での分類一致率

分類タスクの難易度を分析するため、評価者間でのラベルの一致の程度についてカッパ係数[13]を算出したところ 0.49 となり、[14]の基準では適度な一致(Moderate)となった。また 1/5 のデータに情緒的支援と情報提供の両方のラベルが付与された。この結果から、今回の分類タスクの難易度は人間にとって中程度であると解釈した。

4. 応答分類モデルの評価

応答分類モデルの精度評価を行うため、以下の 3 つの方法で 3 クラス分類の正解率と F1 スコアを算出した。表 5 にその結果を示す。

表 5. 学習データ毎の正解率と F1 スコア

学習データ	正解率	F1-Score
全てのラベル付きデータ	0.68	0.67
2 名のラベルが完全一致したラベル付きデータ	0.79	0.78
2名の単一ラベルが完全一致したラベル付きデータ	0.92	0.91

3 章で示した分類タスクの難易度を踏まえると、表 5 で示した正解率 0.92, F1 スコア 0.91 の精度は高い精度だと考えられる。

5. おわりに

本研究では、情緒的支援と情報収集の期待を考慮した恋愛相談システムの構築を目標とし、その前段階として期待に沿った応答を返すための応答分類モデルを構築した。また分類精度は 4 章に示した通りとなった。今後は、今回構築した応答分類モデルを使用し、1 章で示した目標である対話システム全体の構築を目指す。また完成した対話システムとの対話印象評価を通して、今回構築した各モデルの実運用上の精度を測定する必要がある。

謝辞

本実験の実施にあたり、ご協力いただきました参加者の皆様に厚くお礼申し上げます。

参考文献

- [1] Woebot, <https://woebothealth.com/> (最終閲覧日 2022/12/4)
- [2] KOKOROBO, <https://www.kokorobo.jp/> (最終閲覧日 2023/1/15)
- [3] 青山商事株式会社, https://www.y-aoyama.jp/campaign/chat_bot/snack.html (最終閲覧日 2023/1/15)
- [4] Oracle(2020), 日本の「職場における AI」調査: AI 利用は世界 11 カ国で最下位も, 87 %が不安やストレスを相談する相手としてロボット・AI を受け入れると回答, Oracle 日本・ニュースコネクト,

<https://www.oracle.com/jp/corporate/pressrelease/jp20201104.html> (最終閲覧日 2022/12/4)

- [5] Sachiko Takahashi(2013), A Consulting Process with Significant Others on Interpersonal Stress, Japanese Journal of Counseling Science, 46, pages 1-10.
- [6] 石田真也, 井上昂治, 中村静, 高梨克也, 河原達也 (2016), 傾聴対話システムのための発話を促す聞き手応答の生成, 人工知能学会研究会資料, SIG-SLUD-504-01.
- [7] 石田真也, 井上昂治, 中村静, 高梨克也, 河原達也 (2017), 傾聴対話システムにおける自分語りを含む多様な聞き手応答の生成, 情報処理学会第 79 回全国大会, 2. 239-2. 240.
- [8] Kazuki Isoshima, Masafumi Hagiwara(2020), An Automatic Consultation System Focusing on Sympathy and Advice —A Case of Relationship Counseling, 情報処理学会論文誌, Vol.62, No.1, 378-386.
- [9] Hannah Rashkin(2019), Towards Empathetic Open-domain Conversation Models: a New Benchmark and Dataset, In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, pages 5370-5381.
- [1 0] Yahoo! 知恵袋 (2022), 恋愛相談, Yahoo!JAPAN, <https://chiebukuro.yahoo.co.jp/category/2078675272/question/list> (最終閲覧日 2022/1/16)
- [1 1] Huggingface.BERT(2019), https://huggingface.co/transformers/v2.2.0/model_doc/bert.html (最終閲覧日 2022/12/4)
- [1 2] 東北大学乾・鈴木研究室 (2020), cl-tohoku/bert-base-japanese-v2. <https://huggingface.co/cl-tohoku/bert-base-japanese-v2> (最終閲覧日 2022/12/4)
- [1 3] Cohen, J.(1960), A coefficient of agreement for nominal scales, Educational and Psychological Measurement, 20(1):37-46, doi:10.1177/001316446002000104.
- [1 4] Landis JR, Koch GG(1977), The measurement of observer agreement for categorical data, Biometrics, 33(1):159-74.