

感情を含む特徴変化情報付き対話コーパスの構築と それを用いた対話の自然さ推定

仲村 哲明[†]河原 大輔^{†‡}[†] 京都大学 大学院情報学研究科[‡] 科学技術振興機構 さきがけ

tnakamura@nlp.ist.i.kyoto-u.ac.jp

dk@i.kyoto-u.ac.jp

1 はじめに

近年、スマートスピーカーに代表される対話型インターフェースが普及しており、対話理解システムの重要性はますます増加すると予想される。一方で、我々が日常的に行う対話には感情表現が多分に含まれることは明らかであるが、計算機にこのような対話を理解させることは困難である。この一因として、発話と感情の関係に焦点を当てた大規模な対話コーパスがほとんど整備されていないことが挙げられる。

対話コーパスはいくつか開発されているが [4, 3, 5], 発話者の詳細な感情情報を記述しているものはほとんどない。文献 [2] のコーパスは感情に着目した数少ない対話コーパスの1つである。このコーパスでは発話に対して受け手の感情が付与されているが、人手で用意された感情に対応する明示的なキーワードリスト¹を用いて機械的なアノテーションを行っているため、処理対象が直接的な感情表現に限定されている。一方で、日常的な対話で使われる発話文には感情情報が明示的に出現するとは限らないため², 明示的な感情表現だけでなく暗黙的な感情表現も考慮した対話コーパスの整備は重要である。

我々は、対話における発話と感情の関係を理解するには、発話当事者（発話者、発話の受け手）の感情だけでなく、発話内容に登場する感情を持つ存在の感情も重要であると考えている（特に、感情的発話の理解や感情移入の理解など）。我々は、過去の研究 [6] 以降、事態文が表す事態の前後における文中の各項の様々な特徴変化（表 2）に関する知識ベース「日本語特徴変化知識ベース (JFCKB³)」の整備を継続している。本研究では JFCKB に基づいた対話コーパスを構築する。この対話コーパスでは、発話文（事態文）に対して、

文中の各項の特徴変化（JFCKB に準拠）および文の読み手/聞き手の感情変化が関連付けられている。このコーパスの有効性を検証するために、本研究では対話の自然さの判断タスクを機械学習によって行う。

2 提案コーパス

日本語特徴変化知識ベース JFCKB には、例えば、「妻が子供を引っ叩く」という事態文中の「子供」に対して、痛みの増加、嫌悪感の増加などの特徴変化情報が付与されている。さらに、同事態文には、「事態文の読み手」に関する感情変化情報（嫌悪感の増加など）も付与されている。これらの特徴変化情報はクラウドソーシングによって取得されている。本研究では、JFCKB をベースにした対話コーパスを構築することで、発話と感情の関係に焦点を当てた大規模対話コーパスの整備を目指す。

2.1 JFCKB

JFCKB は事態文に関する 3 種類の情報（事態文、格（ガヲニ格）となる項、各特徴の変化確率）で構成されている（表 1 の左側 3 列）。各特徴の変化情報は変化挙動の 3 つ組（増える、減る、変化なし）の確率値（0 ~ 1）で表現されている。

JFCKB における知識は表 2 に示す 47 特徴が想定されている。これらの特徴は、感情に関する従来研究 [7], 日本語シソーラス [11], 感情分析研究 [13], 他の言語資源 [1]などを参考に、基本レベルカテゴリ [8, 9] に沿う粒度となるように設計されている。ただし、現状の JFCKB は、表 2 のうち、主として感情特徴と感覚特徴の調査が終了した段階である。

JFCKB における事態文は以下の手順によって生成した。(Step 1) 京都大学ウェブ文書リードコーパス⁴, 京都大学格フレーム⁵, 日本語版 Winograd Schema

¹ 単語 *afraid* には感情 *fear*, 単語 *happy* には感情 *joy*, など。

² 例えば、「妻に引っ叩かれたよ」という発話には、発話者の「驚き」、「怒り」、「嫌悪感」に関する何らかの情報が含まれていると考えられる。

³ Japanese Feature Change Knowledge Base

⁴ <http://nlp.ist.i.kyoto-u.ac.jp/EN/index.php?KWDLC>

⁵ <http://nlp.ist.i.kyoto-u.ac.jp/index.php?京都大学格フレーム>

| 事態文 | 格 (語) | 特徴変化 (変化挙動の確率) | トリガ発話 | 応答 (自然さの確率) |
|-----------|--------------|--------------------------------------|-------------|-------------------------------------|
| 妻が子供を引っ叩く | ガ格 (妻) | 喜び (+, -, UNC) = (0, 1, 0) | 妻が子供を引っ叩いたよ | ひどいね (o, x, UNK) = (1, 0, 0) |
| | ヲ格 (子供) | 怒り (+, -, UNC) = (1, 0, 0) | | なんで? (o, x, UNK) = (1, 0, 0) |
| | 二格 (NULL) | N/A | | 暴力はいけないよ (o, x, UNK) = (1, 0, 0) |
| | 読み手 | 嫌悪感 (+, -, UNC) = (0.99, 0, 0.01) | | : |
| | | : | | |

表 1: 提案する対話コーパス (JDCFC) の例. 表の左側 3 列は, 我々が整備中の日本語特徴変化知識ベース JFCKB と同じ情報である. 各特徴には事態文の発生によって特徴がどのように変化するか確率値が付与されている. 各応答にはトリガ発話に対する応答としての自然さの確率値が付与されている. 特徴変化列の記号「+」, 「-」, 「UNC」は, それぞれ, 「増える」, 「減る」, 「変化しない」を表す. 応答列の記号「o」, 「x」, 「UNK」は, それぞれ, 「自然」, 「不自然」, 「わからない」を表す.

| 大分類 | 小分類 | 特徴 |
|-----|-----|--|
| 物理 | 形 | 長さ, 大きさ, 広さ, 太さ, 厚さ |
| | 色 | 赤さ, 橙色さ, 黄色さ, 緑色さ, 青さ, 紫色さ, 茶色さ, 白さ, 黒さ, 明るさ |
| | 感触 | 温度, かたさ, 粗さ, 粘り気 |
| | におい | 良さ, 悪さ |
| | 音 | 静かさ |
| | 味 | 甘み, 酸味, 苦み, 辛(から)み, 渋み |
| | 密度 | 粗密さ |
| | 数量 | 多さ |
| 心理 | 感情 | 喜び, 信頼感, 驚き, 嫌悪感, 恐れ, 悲しみ, 怒り, 期待感 |
| | 評価 | 極性 |
| 感覚 | 感覚 | 痛み, 眠気, 疲れ |
| 関係 | 関係 | 接触, 働きかけ, 力の有無, 所有, 社会的関係 |
| | 位置 | 近さ |

表 2: JFCKB で想定されている特徴. 表中の太字は, これまでに JFCKB で調査が終了している特徴である.

Challenge データセット [12] から, それぞれ, 最頻出 200 他動詞, 最頻出 1,000 他動詞, 全動詞を抽出する. (Step 2) 抽出された動詞について, 京都大学格フレームを用いて動詞のガヲニ格に関する代表語 (各格に関する最頻出語) を機械的に組み合わせることで, その動詞の代表文を作成し, その代表文をその動詞の事態文とする. (Step 3) 自然な文であるかを評価するクラウドソーシングタスク (のべ 1,559 名が参加) を実施し, 不自然な文であると判断された文を除外する.

事態文の生成後, 事態文の各項の特徴変化および事態文の読み手の特徴変化を収集するクラウドソーシングタスクを実施した. 作業員には, 提示された事態文の指定された項に関する指定された特徴が事態の前後でどのように変化するかを回答するよう求めた (e.g., 「妻が子供を引っ叩く」の「子供」の怒りはどのように変化しますか?). このタスクには, のべ 33,683 名が参加し, 最終的に 9,073 文 (975 動詞 (異なり)), の

べ 19,052 項 (異なり数 4,882), 5,647 格フレーム (異なり)) に関する特徴変化情報が得られた.

本研究で実施した全てのクラウドソーシングタスクでは, 各選択肢が選ばれた確率を文献 [10] で提案されている手法によって求めた. この手法は, 単純な多数決とは異なり, 作業員の回答一致率に基づいて作業員の回答能力や問題の難しさを推定することで確率値を計算する手法である.

2.2 JDCFC: 特徴変化情報付き日本語対話コーパス

本研究で提案する対話コーパス (Japanese Dialogue Corpus with Feature Changes: JDCFC) は JFCKB をベースに構築する. JDCFC は表 1 に示す事態文に関するレコードを単位として構成される. 各レコードは, 事態文, 事態文のガヲニ格, ガヲニ格および読み手の特徴変化情報, 対話でのトリガ発話としての事態文, トリガ発話に対する応答候補, トリガ発話に対する応答候補の自然さ (確率) から構成される. JDCFC で使用される特徴は JFCKB と同じ特徴である.

JDCFC におけるトリガ発話は JFCKB に基づいて作成される. トリガ発話となる事態文は, JFCKB に登録されている全ての事態文 (9,073 文) のうち 1 つ以上の特徴変化⁶が発生する事態文 (2,428 文) である.

JFCKB での事態文と JDCFC でのトリガ発話の違いは文末表現 (JFCKB では現在形, JDCFC では過去形 + 終助詞「よ」) である. これは, 対話における事態文の文末表現はこのような表現のほうが自然であると考えられるためである. 上記 2,428 文の事態文に対するトリガ発話の生成後, 2 つのクラウドソーシ

⁶ある特徴 f の増加または減少する確率が 0.75 以上である場合に, f の変化が生じるとした.

| | |
|--|-------------|
| 会話の相手が以下の発言をした時、あなたはどのような応答をしますか？(自然な会話になるように、あなたの応答を記入してください) | |
| 相手の発言 | 妻が子供を引っ叩いたよ |
| あなたの応答 | |

図 1: 対話の応答を収集するタスクの例.

| | |
|--|-------------|
| 以下に表示されている 2 人 (A さんと B さん) の会話は自然なやり取りですか？(A さんが話した後に、B さんが応答していると考えて下さい) | |
| A さん | 妻が子供を引っ叩いたよ |
| B さん | なんで？ |

図 2: 応答の自然さを判定するタスクの例.

グタスクを実施することで JDCFC を構築した.

第 1 のタスクではトリガ発話に対する応答を収集する. このタスクにはのべ 8,370 名の作業者が参加し, 作業者はトリガ発話を提示され適切な応答をするよう求められた (図 1). 各トリガ発話に関して, 収集された応答のうち重複するものを除外し, さらに, 極めて低品質で不適当な応答 (提示文のコピーアンドペーストや空欄) を除外した. その結果, 2,428 文のトリガ発話文に対して, のべ 23,196 応答が得られ, 1 トリガ発話あたりの異なり応答数は約 9.6 件であった.

第 2 のタスクではトリガ発話に対する応答の自然さを推定する. このタスクにはのべ 5,605 名の作業者が参加し, 作業者は提示された対話 (トリガ発話と 1 つの応答のペア) が自然かどうか回答するよう求められた. トリガ発話とペアになる応答は, 上述のトリガ発話に対応する複数の応答の中から 1 つが選ばれ, 全ての応答候補のペアを作成した. 結果的に, 23,196 件の対話ペアに対する自然さの回答が得られた. 与えられた対話ペアが自然であると判断された確率は, JFCKB の構築時と同様に, 文献 [10] の手法によって算出した. 表 1 に示した応答および自然さの確率は実際に得られた結果である. タスクに使用した 23,196 対話ペアのうち, 作業者が自然であると判断した確率が 0.8 以上となった対話ペアは 22,357 件であった. 本研究では, この 22,357 件の対話ペアを用いて対話コーパスの有用性を評価した.

3 評価実験

JDCFC は詳細な感情変化情報を付与した対話コーパスである. このコーパスの有用性を検証するために, このコーパスを学習データとして用いた機械学習器に対話の自然さを推定させる実験を行った. 本研究では, クラウドソーシングによって作業者が自然な対話ペアであると判断した確率が 0.8 以上の対話ペア (22,357 件) を自然な対話ペアであるとして評価実験に用い

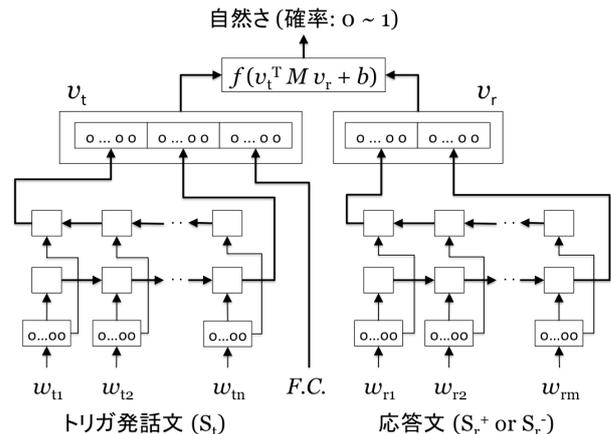


図 3: 提案する対話コーパスを評価するための LSTM モデル. 記号 $F.C.$ はトリガ発話の述語に関する特徴変化ベクトルを表す.

た. 実験には, 22,357 件の自然な対話ペアをベースに, 22,357 件の学習データセット (正例と負例のセット) を作成した. これらのセットの作成に関しては, トリガ発話 S_t に対する応答のうち, 自然な対話ペアであるとされた応答文を S_t に対する正例 S_t^+ として採用した. 一方, S_t とは異なるトリガ発話に対する応答のうち, S_t に対して得られた応答と重複しない応答をランダムに 1 つ選択し, それを S_t に対する負例 S_t^- として S_t^+ と組み合わせ, S_t に関する 1 つの学習データセットとした. 実験タスクには, 正例と負例のセットを与えた時にどちらが自然な応答であるかを選択するタスクを採用した.

3.1 モデル

機械学習器には文献 [5] で用いられている双方向 LSTM モデルを用いた (図 3). このモデルによる対話ペア (トリガ発話文と応答文) の自然さの推定手順は以下の通りである. (1) トリガ発話文と応答文に出現する word embeddings を LSTM の入力とする. ただし, トリガ発話文と応答文の単語の並びを, それぞれ, $w_{t1}, w_{t2}, \dots, w_{tn}$ および $w_{r1}, w_{r2}, \dots, w_{rm}$ とする. 本研究で使用した word embeddings は 128 次元のベクトルである. (2) 与えられた word embeddings からトリガ発話文と応答文の隠れ層ベクトルを双方向 LSTM によって計算する. 本研究では隠れ層ベクトルの次元数を 100 にした. (3) トリガ発話文の文ベクトル v_t は, 先の (2) によって最終的に出力された (w_{tn} に対して出力された) 前向き隠れ層ベクトル, 最終的に出力された (w_{t1} に対して出力された) 後向き隠れ層ベクトル, および, トリガ発話文の述語に関する特徴変化ベクトルを連結したベクトルである. 応答文の

文ベクトル v_r は、 v_t の場合と同様の方法で出力された前向き隠れ層ベクトルと後向き隠れ層ベクトルを連結したベクトルである。この評価では、特徴変化として用いる特徴を表2の8感情に限定したため、特徴変化ベクトルは96次元⁷の確率値のベクトルである。(4)最終的な出力 o (入力された対話ペアの自然さ) は式(1)により得られる。式(1)において、 f , W , b は、それぞれ、活性化関数、重み行列、バイアスを示す。

$$o = f(v_t^T W v_r + b) \quad (1)$$

本研究では、活性化関数にシグモイド関数を用い、パラメータ最適化手法には Adam を用いた。

3.2 実験設定

この評価では2つのモデル(ベースラインモデル, 提案モデル)を比較した。ベースラインモデルは図3のモデルから特徴変化情報を除外したモデルで、提案モデルは図3に示したモデルそのものである。ただし、図3の隠れ層は説明のために1層であるが、実際の両モデルの隠れ層は2層とした。

学習時には正例と負例に対してそれぞれ1と0を出力するように学習した。評価に関しては、正例と負例のセットを与えた時に正例に対する出力が負例に対する出力よりも大きければ正しい推定であるとした。

学習データ22,357セットのうち、80%, 10%, 10%を、それぞれ、訓練データ、開発データ、評価データに割り当てた。評価は訓練データ10エポックの学習後に行った。

3.3 結果と考察

評価結果を表3に示す。この表から、提案モデルはベースラインモデルよりも良好な結果であることがわかる。この結果は対話の自然さを推定する際に特徴変化情報が有効であることを示唆していると考えられる。

システムが正しく推定できた一例としては評価セット(トリガ発話: 俺が使徒を倒したよ, 正例: それはすごいです, 負例: それはつらいですね)がある。このトリガ発話に関しては、JDCFC/JFCKBにおける文の読み手の特徴変化が、「怒りの減少」や「驚きの増加」と関連していた。この場合の正例は、このような読み手の感情変化を反映しているように見える。我々は、トリガ発話に対する応答が読み手/受け手の特徴変化に関連している場合に、特徴変化情報が対話の自然さの判断にうまく働くと推測している。

⁷各述語に関して、各特徴は3次元(増える, 減る, 変化なし), 項は4種類(ガ格, ヲ格, 二格, 読み手), 特徴数は8(8感情)であるので96次元となる。

| | ベースラインモデル | 提案モデル |
|----|-----------|-------|
| 精度 | 64.2% | 71.0% |

表3: 評価結果.

4 おわりに

本研究では、感情と発話の関係に焦点を当てた対話コーパスを構築した。このコーパスでは、発話文中の各項の感情変化とその発話文の読み手/聞き手の感情変化が関連付けられている。このコーパスを学習データとして用いた機械学習器に対話の自然さを推定させる実験を行った結果、コーパスの有用性を確認できた。

参考文献

- [1] Joshua K. Hartshorne, Claire Bonial, and Martha Palmer. The verbcorner project: Findings from phase 1 of crowd-sourcing a semantic decomposition of verbs. In *Proc. of ACL 2014*, pp. 397–402, 2014.
- [2] Takayuki Hasegawa, Nobuhiro Kaji, Naoki Yoshinaga, and Masashi Toyoda. Predicting and eliciting addressee’s emotion in online dialogue. In *Proc. of SCL2013*, pp. 964–972, 2013.
- [3] Ryuichiro Higashinaka, Kotaro Funakoshi, Yuka Kobayashi, and Michimasa Inaba. The dialogue breakdown detection challenge: Task description, datasets, and evaluation metrics. In *Proc. of LREC 2016*, pp. 3146–3150, 2016.
- [4] Seokhwan Kim, Luis Fernando D’Haro, Rafael E. Banchs, Jason D. Williams, Matthew Henderson, and Koichiro Yoshino. The fifth dialog state tracking challenge. In *Proc. of SLT 2016*, pp. 511–517, 2016.
- [5] Ryan Lowe, Nissan Pow, Iulian V. Serban, and Joelle Pineau. The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems. In *Proc. of the SIGDIAL 2015*, pp. 285–294, 2015.
- [6] Tetsuaki Nakamura and Daisuke Kawahara. Constructing a dictionary describing feature changes of arguments in event sentences. In *Proceedings of the 4th Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pp. 46–50, 2016.
- [7] Robert Plutchik. *A General Psychoevolutionary Theory of Emotion*, 1, pp. 3–33. Academic Press, 1980.
- [8] Eleanor Rosch, Carolyn B. Mervis, Wayne D. Gray, David M. Johnson, and Penny Boyes-Braem. Basic objects in natural categories. *Cognitive Psychology*, Vol. 8, pp. 382–439, 1976.
- [9] John R. Taylor. *Linguistic Categorization: Prototypes in Linguistic Theory*. Clarendon Press, 1995.
- [10] Jacob Whitehill, Paul L. Ruvolo, Jacob Bergsma Tingfan Wu, and Javier R. Movellan. Whose vote should count more: Optimal integration of labels from labelers of unknown expertise. In Y. Bengio, D. Schuurmans, J. D. Lafferty, C. K. I. Williams, and A. Culotta, editors, *Advances in Neural Information Processing Systems 22*, pp. 2035–2043. Curran Associates, Inc., 2009.
- [11] 池原悟. 日本語語彙大系. 岩波書店, 東京, 1997.
- [12] 柴田知秀, 小浜翔太郎, 黒橋禎夫. 日本語 Winograd Schema Challenge の構築と分析. NLP 2015 発表論文集, pp. 493–496, 2015.
- [13] 徳久良子, 乾健太郎, 松本裕治. Web から獲得した感情生起要因コーパスに基づく感情推定. 情報処理学会論文誌, Vol. 50, No. 4, pp. 1365–1374, 2009.