

感情表出傾向の分析による対話破綻検出

松本 和幸 任 福継 吉田 稔 北 研二

徳島大学 理工学部

{matumoto, ren, mino, kita}@is.tokushima-u.ac.jp

1 はじめに

近年,人工知能技術による雑談対話システムは,柔軟な応答文生成が可能になってきている.しかし,それらのシステムを使用していると,しばしば対話破綻が起こる.対話破綻には,さまざまなタイプが存在する.対話破綻検出チャレンジにおいて提案された手法の多くは,対話破綻している基準として,対話文の意味や発話行為,話題の変化等に注目して検出するものであった¹.人間は,話題などの食い違いによる破綻以上に,感情の変化への対応の不十分さから違和感を感じ取っていることが多いと考えられる.たとえば,以下のような対話があったとする(S:システム発話,U:ユーザ発話).

S:「その洋服どこで買ったの?」

U:「 デパートで買ったのよ.お気に入りなの。」

S:「それ、安っぽいね。」

現実でも起こり得る対話の流れではあるが,「お気に入りの洋服」に対して,システム側が否定するような発言をすることにより,対話が破綻してしまうこともある.文脈的・話題的には破綻していないが,円滑な対話を進めるうえで,できれば避けたい流れといえる.ユーザ(人間)との円滑かつ親密な対話をする雑談対話を成立させるためには,従来の話題に着目するだけでなく,感情の変化に対応し,ユーザの発言に対して,同調,同情,共感することが必要不可欠である.

本研究では,ユーザの感情の変化に合わせてシステム自身の感情を表出させることができないために起こる対話破綻について,発話文から推測可能な感情傾向の比較により検出する手法を提案する.具体的には,発話文の分散表現ベクトルをもとに感情が表出されている強度を推定するモデルを用い,直前の2発話における推定結果との類似度の計算を行うことで,そのパターンを特徴量として対話破綻のラベルを推定する.評価実験では,提案手法を用いた対話破綻ラベルの推定をおこない,文

の分散表現ベクトルの類似度パターンによる手法との比較をおこなう.

2 関連研究

2.1 対話破綻検出

本研究では,Project Next NLP 対話タスクで収集された,雑談対話コーパス [1] の開発データを実験用データセットとして用いる.このコーパスは,雑談対話システムとユーザ(人間)との対話を記録したデータに対し,複数の作業者が対話破綻ラベルおよびコメントのアノテーションをおこなったものである.このデータには,3種類の対話破綻ラベル(O, , x)が付与されているが,各破綻の種類の詳細なカテゴリについては,未付与である.

東中ら [2] の研究では,対話破綻の種類を類型化している.本研究で着目している感情理解の欠如による対話破綻は,この対話破綻の類型の大分類「環境」の「社会性欠如」や「一般常識欠如」に該当するものと考えられる.しかし,ここで想定されているのは,感情のように感覚的なものが主ではないと考えられる.

2.2 発話文からの感情推定

Renら [3] が,ベースとなる感情コーパスから各種特徴量を抽出し,感情ラベル推定器を構築することにより,感情コーパスを半自動構築する手法を提案している.しかし,各感情ラベルの推定精度は高くなく,人手によるラベル付けの作業支援への応用にとどまっている.この原因としては,コーパスを自動構築するためには,典型的な用例文を大量に収集する必要があるが,実際に収集されているコーパスは典型的な用例ばかりではないといった点と,基本コーパスへのラベル付けの精度(信頼度)を保証できるほどの作業者数を集めることが困難な点あげられる.本研究では,人手により作成した感情ラベル付きコーパスをもとに,従来研究よりも

¹<https://sites.google.com/site/dialoguebreakdownndetection2>

表 1: 分類ラベルと付与数

分類ラベル	感情の例	付与数
A-1	喜び, 安心, 感動, 希望, etc.	24,044
A-2	愛, 尊敬, 好き, etc.	7,124
B-1	驚き, 仰天, 呆然, etc.	2,350
C-1	怒り, 苦しみ, 嫌悪, 軽蔑, etc.	13,037
D-1	悲しみ, 哀れみ, 罪悪感, etc.	8,201
D-2	不安, 恐怖, etc.	7,816

汎用性を高めるため, 文の分散表現ベクトルを素性として学習させた感情推定モデルを構築し, 感情ラベルの推定に用いる.

3 提案手法

提案手法は, システム応答文とユーザ発話文のそれぞれから推定された感情の強度ベクトル間の類似度を求め, それらの組合せを類似度パターン特徴として用いることにより, 対話破綻ラベルの推定モデル(対話破綻分析器)を構築する. 以下, 各ステップについて説明する.

3.1 感情推定

感情推定の前処理として, 文またはフレーズ・単語をベクトル化するためのベクトル生成モデルを構築する必要がある. これには, Mikolov[4] が提案した paragraph2vec の実装の 1 つである, sentence2vec を用いる.

sentence2vec は, 文のベクトル化を, word2vec の単語ベクトルの手法を拡張して実現したものであり, 意味的に類似または関連する文を, 互いに類似したベクトルに変換できる. 本研究では, ベクトルの次元数を 500 とし, 文脈窓の幅を 10 と設定し, 学習をおこなった. このベクトル生成モデルの生成には, 3,403,658 ツイートをランダムに収集したものを日本語形態素解析器 MeCab を使用して形態素単位に分ち書きしたものをを用いた.

感情推定モデルを構築するため, 口語文のみを収集した発話文コーパス約 38,799 文と, 感情表現(単語やフレーズ)に感情の種類を付与した感情表現辞書(エントリ数: 19,529)を用いる. これらには, フィッシャー[5]の提案する感情の系統図をもとに, 表 1 に示す 6 種類のラベルを付与している. 発話文や感情表現によっては, 複数のラベルが付与されるものもあり, それらには優先順位として重みづけを行っている.

以上の方法により感情ラベル付きデータをベクトル化したものを Deep Neural Network により機械学習し, 感情推定モデルを構築する. ネットワークの種類は, 隠

表 2: ネットワークの構成とパラメータ

層	ユニット数	活性化関数	dropout 率
input	500	tanh	0.5
hidden1	128	tanh	0.5
hidden2	256	tanh	0.5
hidden3	512	tanh	0.5
hidden4	256	tanh	0.5
output	6	softmax	-

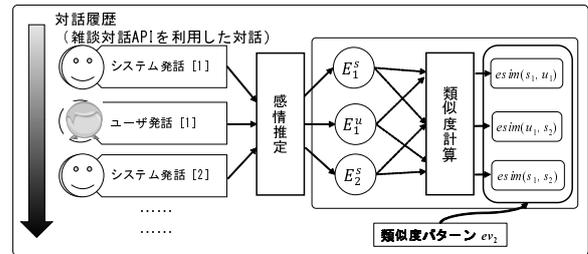


図 1: 感情の類似度パターン抽出過程

れ層 4 層の順伝播型で全結合とし, 各層のユニット数および活性化関数などのパラメータを, 表 2 に示す.

3.2 類似度パターン特徴の抽出

図 1 に, 発話における感情推定と, その結果得られた感情の類似度パターンの抽出過程を示す. 一般に, 感情理解の欠如を主要因としたシステムの応答発話における対話破綻は, 直前のシステム発話と, 直前のユーザ発話がそれぞれ表出している感情の違いが影響して起きると考えられる. そのため, システム応答発話ごとに, 以下のような特徴量を抽出する.

- 直前のユーザ発話文の感情 (E_{t-1}^u) と, 現在のシステム応答発話文の感情 (E_t^s) の類似度 $esim_{t-1,t}^{u,s}$
- 直前のシステム応答発話文の感情 (E_{t-1}^s) と, 現在のシステム応答発話文の感情 (E_t^s) の類似度 $esim_{t-1,t}^{s,s}$
- E_{t-1}^u と E_{t-1}^s との類似度 $esim_{t-1,t-1}^{u,s}$

3.3 対話破綻ラベルの推定手法

本研究では, 対話破綻分析器の学習に, 正準相関分析と k 近傍法を用いた. 以下, それぞれの手法について説明する.

3.3.1 正準相関分析による推定

感情表出傾向のベクトル類似度パターン ev_i に対応する対話破綻ラベルの分布 dv_i の推定は, ある実数値ベク

トル ev_i から、別の実数値ベクトル dv_i を推定する問題に当てはめることができる。本研究では、前述の推定をおこなう多変量解析の手法として正準相関分析を用いる。正準相関分析は、複数の変量間の相関係数（正準相関係数）を求める手法であり、画像とテキスト、音楽と生体情報など、異種メディア間の相関を求めて情報推薦する手法に用いられることもある [6]。

3.3.2 k 近傍法による推定

事例数が少ない場合に、正準相関分析のような手法ではうまくいかない場合がある。類似度パターンが類似する事例との比較をコサイン類似度によりおこない、類似度が大きい順に k 個の事例の対話破綻ラベルの分布の重み付き平均を計算することで対話破綻ラベルの分布を推定する。

4 実験

提案手法の有効性を示すため、対話破綻検出チャレンジ 2 用のデータセットとして公開されている 3 種類の雑談対話システム (DCM, DIT, IRS) による対話データ (各 50 対話の開発データ) をそれぞれ別々のデータセットとして用いて評価実験をおこなった。今回は評価データは用いず、開発データのみで対話単位での交差検証をおこなう。k 近傍法の学習パラメータとして、類似度の閾値 th_{sim} と、 k の値について、 $k = (1, 2, 3, \dots, 100)$, $th_{sim} = (0.1)$ の組合せを試行し、最適値を求めた。提案手法とベースライン手法 (文の分散表現ベクトルの類似度パターンを素性とした手法) とを、以下の評価指標により比較する。

- 正解ラベルの分布と対話破綻分析器が出力したラベル分布の間の
 - コサイン類似度 (cos)
 - 相関係数 (cc)
 - JS Divergence (JSD)
 - Mean Squared Error (MSE)
- 対話破綻応答文の検出率 (破綻文を破綻文として推定できた割合)

コサイン類似度、相関係数および検出率は高いほど性能がよく、JSD と MSE は低いほど性能が良いことを示す。

表 3: 正準相関分析による実験結果

指標	DCM		DIT		IRS	
	ev	sv	ev	sv	ev	sv
cos	0.419	0.237	0.566	0.269	0.471	0.287
cc	-0.028	0.050	0.087	-0.013	0.062	-0.007
JSD	0.542	1.079	0.335	0.968	0.539	0.814
MSE	1.095	6.924	0.668	4.095	1.554	2.826
検出率	0.230	0.351	0.333	0.485	0.560	0.372

表 4: k 近傍法による最大検出率と k の最適値

手法	データセット	k の最適値	検出率
sv	DCM	66	0.569
	DIT	76	1.000
	IRS	23	0.648
ev	DCM	26	0.553
	DIT	84	1.000
	IRS	88	0.832

4.1 実験結果と考察

正準相関分析による実験結果を表 3 に示す。表中、提案手法を ev、ベースライン手法を sv で示している。

この結果をみると、各指標、各データセットにおいて、ほとんどの場合において提案手法がベースラインを上回るという結果になった。また、検出率に関しては、データセットが IRS の場合のみ提案手法がベースラインを上回った。

つぎに、k 近傍法による実験結果を、表 4 に示す。破綻ラベル (x) の検出率 (再現率) は、提案手法とベースラインとであまり差が出ない。また、データセット DIT を用いた実験では、検出率 1.0 という値が出ている。類似度パターンに差異が出にくく、結果として、データセットのラベルの分布に従い、ほとんど破綻ラベルとして出力してしまったことが考えられる。

図 2, 3, 4 は、それぞれのデータセットにおける k 近傍法による各ラベルの検出率の k の値による推移を示したものである。‘O’, ‘X’, ‘T’ は、それぞれ、‘○’, ‘x’, ‘’ を表す。データセットによって異なる傾向が見てとれるが、比較的安定した検出率である DCM の結果においても、k の値が 30 を超えるあたりから出カラベルに偏りが出てしまい、検出率がばらつくという結果となっている。

5 おわりに

本研究では、対話破綻の一因として考えられる「感情理解の欠如」が感情表出傾向の差異に表れると仮定し、直前発話との感情推定結果の類似度パターンを特徴量

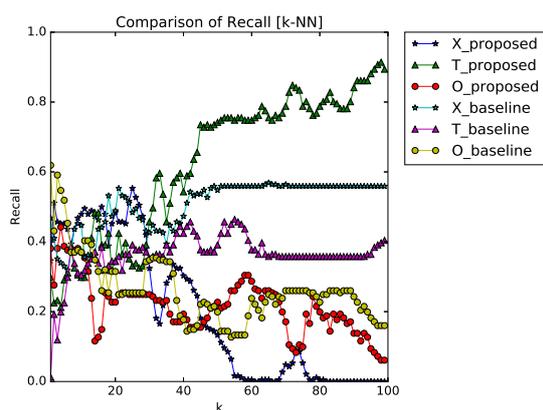


図 2: k 近傍法による各ラベルの検出率の比較 (DCM)

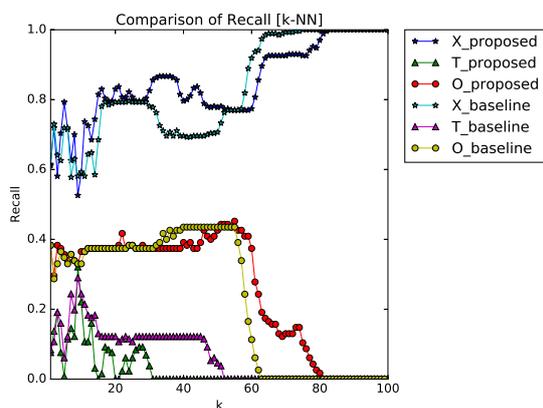


図 3: k 近傍法による各ラベルの検出率の比較 (DIT)

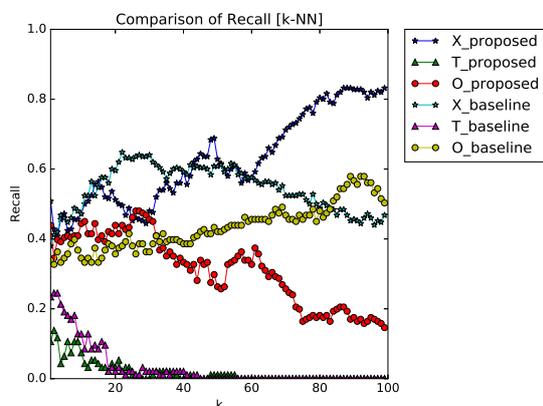


図 4: k 近傍法による各ラベルの検出率の比較 (IRS)

として抽出し、対話破綻ラベルの分布を推定する手法を提案した。評価実験において、文の分散表現ベクトルの類似度パターンを用いた場合と比較し、条件によっては評価指標が向上するという結果が得られた。また、対話破綻分析器により検出できた対話破綻応答文に対し、「感情理解の欠如」が生起していたかどうかは、今回は分析できなかったため、今後の課題としたい。

本稿で用いた感情推定モデルは、対話に特化したものではなかったため、今後は、人手による対話データへの感情ラベル付与などの作業もおこない、対話システム用の感情推定モデルを検討したい。さらに、対話破綻検出チャレンジで用いられたベースライン手法や、発話のトピック間類似度の対話破綻検出手法と組み合わせることにより、検出精度の向上を目指したい。

謝辞

本研究の一部は、科学研究費補助金（基盤研究（A）：15H01712）の補助を受けて行った。

参考文献

- [1] 東中 竜一郎, 船越 孝太郎: Project Next NLP 対話タスクにおける雑談対話データの収集と対話破綻アノテーション, 人工知能学会 言語・音声理解と対話処理研究会 第 72 回, pp.45-50, 2014-12-15.
- [2] 東中 竜一郎, 船越 孝太郎, 荒木 雅弘, 塚原 裕史, 小林 優佳, 水上 雅博: 雑談対話データの収集と対話破綻アノテーションおよびその分類, 自然言語処理, Vol.23, No.1, pp. 59-86, 2015.
- [3] Ren, F. and Matsumoto, K.: Semi-automatic Creation of Youth Slang Corpus and Its Application to Affective Computing, IEEE Transactions on Affective Computing, Vol.7, No.2, pp.176-189, 2015.
- [4] Mikolov, T.: Distributed Representations of Sentences and Documents, Proceedings of the 31 st International Conference on Machine Learning, JMLR: W&CP Vol.32, 2014.
- [5] Fischer, K. W., Shaver, P. and Carnchan, P.: A skill approach to emotional development: From basic- to subordinate-category emotions. In W. Damon (Ed.), Child development today and tomorrow (pp.107-136), 1989.
- [6] 赤穂 昭太郎: カーネル正準相関分析, 2000 年情報論的学習理論ワークショップ, 2000.