

情緒注釈付き対話コーパスの誤り分析

Error analysis of emotion annotated dialogue corpus

古塩 貴行 徳久 雅人 村上 仁一 池原 悟
Takayuki Koshio Masato Tokuhisa Jin'ichi Murakami Satoru Ikehara

鳥取大学 工学部 知能情報工学科

Dept. of Information and Knowledge Engineering, Tottori University

We have been building emotion annotated dialogue corpus and rule-base for emotion recognition system in dialogues. While more annotaters are required in order to increase the corpus, what is important is the training for the annotation. This paper proposes a training program for the annotation. First, a novice annotater builds prototype corpus. Second, the emotion recognition system infers emotions, and lists differences between the corpus and the result of the inference. Third, an expert annotater checks them and discusses with the beginner. Finally the beginner repairs the corpus. In this paper, one annotater was trained and error ratio decreased from 40.2% to 3.2%. Then he annotated to 6 dialogues, which had been used to build the rule-base, and an expert annotater and the recognition system checked the tags. Since the error ratio of the trainee was 4.5% and the reduction ratio of the tags that the expert had to check was 67%, the effectiveness of the training program was confirmed.

1. はじめに

人間と計算機の対話における言語理解型的情绪推定システムを構築するために、大規模コーパスが必要となる。しかし、情緒注釈付き対話コーパスは体系が複雑なため、精度の高いコーパスの作成には熟練を要することが問題である。そこで本稿では、コーパス作成者の熟練度を向上させるためのコーパス作成の訓練手順を提案することを目的とする。

作成したコーパスを直接、熟練者が評価すれば、確実にコーパスの誤りが抽出できるが、手作業での評価はコストが高い。よって、情緒推定システムによる自動推論の推論過程と比較して、相違のある注釈について熟練者が評価する。そして、作成者がその評価を参考にコーパスの誤りを見直し、その傾向を分析する。

本稿では、30件の対話(323ターン)について、この手順を実施して、コーパス作成者の訓練を行う。そして、その後、別の6件の対話(58ターン)について注釈付与を行い、本提案手順の効果を評価する。

2. コーパス作成の訓練手順の提案

2.1 情緒注釈付き対話コーパス

(1) 情緒推定の方法

[徳久 98]は「喜び/悲しみ、好ましい/嫌だ、驚き、期待、恐れ、怒り」を基本情緒と定め、基本情緒の生起する原因の特徴を123種類に下位分類した。その生起特徴は、「生理、欲求、情緒、目標、プラン、予測、評価、記憶、認識、行動、言語」という心的状態が主要な構成要素となっている[徳久 01]。計算機が対話相手の情緒を推定する方法は、対話相手の目標やプラン、行動結果などの心的状態を対話文から理解し、情緒生起の原因の特徴とパターンマッチを行うことで実現する。この方法において、対話相手の心的状態は、必ずしも対話文に明示されるわけではないため、暗示された心的状態を対話の文脈から補完することが問題となる。

(2) コーパスのねらい

そこで、対話から心的状態を理解する過程を明確にするために、心的状態の注釈を付与した対話コーパスを作成する。コーパスには、コーパス作成者が、対話者の一方の立場に立ち、もう一方の対話者(相手)の心的状態を注釈として付与する。こうして作られたコーパスは、今後、計算機が対話者となると、相手の情緒を推論するための言語資源になる。

(3) 注釈

注釈には、心的状態を追加するための注釈、心的状態を棄却するための注釈、および、心的状態間の連鎖関係を表す注釈がある。

前者の注釈にはさらに、相手の心的状態の注釈、および、相手の信じる当方の心的状態の注釈の2つに分かれる。それぞれの表記は次のとおりである。

識別子, 心的状態名, 内容, 属性, 理由, 確信性

識別子, 信念-心的状態名, 内容, 属性, 理由, 確信性

中者の棄却のための注釈は次の表記となる。

close, 識別子, 理由

後者の関係を表す注釈は次の表記となる*1。

関係識別子, 関係名, 識別子組, 理由

注釈を構成する項目の記述内容を表1にまとめる。

対話で心的状態が明示されているかどうかは、注釈での「理由」の項目を参照することで区別できる。発話文から直接得られる心的状態の注釈には、理由の項目に発話行為の識別子を記入し、明示されていない心的状態の注釈には、理由の項目に他の心的状態の識別子を記入する。

(4) 具体例

具体例として漫画「ちびまる子ちゃん」から作成したコーパスの一部を図1に示す。このコーパスでは、「まる子」の立場から「お母さん」の心的状態を注釈付けしている。お母さんの発話ターンでは、「お母さんの発話行為「プランの要求」 sa07.1から、お母さんの要求内容のプラン op01.1 が生じ、お母さんはまる子がプランを採用すると信じている pl01.1」ということが注釈付けされている。ここで、op01.1 は sa07.1から直接導かれた注釈であるので、明示された心的状態の注釈

連絡先: 古塩 貴行, 〒 680-8552 鳥取市湖山町南 4-101, 鳥取大学工学部, Tottori University, E-mail: kosio@ike.tottori-u.ac.jp

*1 本稿では、関係を表す注釈は、評価の対象外なので、具体例などからは省略する。

表 1: 注釈に記述する項目と内容

項目名	記述内容
識別子* (心的状態名)	ph(生理), ds(欲求), em(情緒), gl(目標), pl(プラン), op(要求プラン), pr(予測), ev(評価), mm(記憶), ac(行動), cg(認識), sa(発話行為)
内容 属性	プランの動作内容などを記述した簡潔な文 生理: [高い/中/低い], 欲求: [生起/解消], 情緒: [生起/解消], 目標: [設定/達成/解除], プラン: [候補/採用/不採用/終了], 要求プラン: [採用/不採用], 評価: [高い/低い], 行動: [実行中/終了]
理由 確信性	注釈を付与するために参照した注釈の識別子を記述 対話の現状で確認されている注釈には「c」, 予測した注釈には「p」を記述
関係識別子* 関係名	ar 生理-欲求, 生理-情緒, 欲求-目標, 情緒-目標, 目標-プラン, プラン-予測, 予測-評価, 評価-プラン, プラン-行動, 行動-認識 など
識別子組	識別子の 2 つ組

* 識別子および関係識別子の頭文字をここに示す。

である。pl01.1 は op01.1 から導かれた注釈であるので、暗示された心的状態の注釈である。まる子の発話ターンでは、お母さんは「『まる子がプランを受諾し sa08.1, プランを採用している pl01.2』と信じ、同意を得ることにより喜んでいる em01.1」ということが注釈付けされている。

4 お母さん:「でも、世話しなきゃならないのよ。」
sa07.1, 発話行為, プランの要求, [], c
op01.1, 要求プラン, プンチョウの世話をする, 採用, [sa07.1], c
pl01.1, 信念-プラン, プンチョウの世話をする, 採用, [op01.1], p
5 まる子:「そのなのまる子にまかせておいてよ。」
sa08.1, 信念-発話行為, プランの受諾, [], p
pl01.2, 信念-プラン, プンチョウの世話をする, 採用,
[op01.1, pl01.1, sa08.1], p
close, pl01.1, [pl01.2]
em01.1, 情緒, 同意による喜び, 生起, [op01.1, pl01.2], p

図 1: 情緒注釈付き対話コーパスの一部

2.2 コーパス作成の訓練手順

コーパス, 情緒推定システム, および, 作業員 (コーパス作成者, ルールベース作成者, 熟練者) の関係を図 2 に示す。

情緒推定システムは, 対話文の入力があると, 指定する立場から相手の情緒を推定する。ルールベースは, 心的状態, および, 情緒状態を推定するもので, 2 人の熟練者により構築したコーパスのうち, 異なり 6 対話 (のべ 12 対話, 注釈数 211 個) の注釈および 123 種類の情緒を全て推定できる。現在のルール数は 336 個である。なお, この 6 対話は, イソップ物語から引用しているので, 以後, イソップ対話集と呼ぶ。

コーパスと推定結果を比較した結果は, 通常ルールベースの強化に用いるが, 本稿では未熟なコーパス作成者の訓練に用いる。具体的には次の手順で行う。

(手順 1: コーパスの試作)

コーパス作成者が対話文を読み, 注釈を付与する。

(手順 2: 相違点の抽出)

試作したコーパスと情緒推定システムの推論結果を比較して, システムに推定されなかった注釈を相違点として抽出する。

(手順 3: 相違点の評価)

相違点について, コーパス作成者に非があるかシステムに非があるのかを熟練者の指導の下で判定する。

(手順 4: 誤りの分析)

手順 3 で「コーパス作成者に誤りがある」と評価された注釈についてその傾向を分析する。

(手順 5: コーパスの校正)

手順 4 の考察結果に従い, 注釈を修正する。

本手順でのコーパスの修正作業は, 熟練者が直接行うものではない。情緒推定の過程には個人差が元々存在するので, コーパス作成者の推定過程を熟練者が単純に否定してはいけぬ。修正のねらいは, コーパス作成者の推定過程における「短絡的な部分*2」をなくすこと, すなわち, 暗示された心的状態を書き尽くすことである。

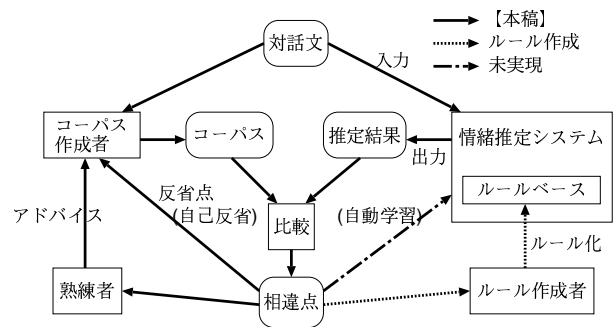


図 2: 提案するコーパス作成の流れ

3. 訓練手順の実施

3.1 対象

対象は, 漫画「ちびまる子ちゃん」に記述された対話文とした。理由は, 子供でも理解できる内容であること, 日常的な出来事が描かれていること, 平凡な喋り方の文であること, が挙げられる。こうした対話の中でも, 勧誘の場面を中心に, 話題が途中で変わらないこと, 5 ターン以上のやりとりがあること, 台詞に単文が多いこと, という条件で 30 対話 (323 ターン) を収集した。以後, まる子対話集と呼ぶ。

3.2 コーパスの試作と相違点の抽出

まる子対話集に注釈を付与した結果を, 表 2 にまとめる。1 ターンあたり平均 2.1 個の発話行為が含まれている。本稿では, 明示されていない注釈を調べるので, 発話行為の注釈およびそれに直接結びつく注釈を除外すると, 343 個の注釈が, 相違点を調べる対象となる。

情緒推定システムによる推定結果と上述の 343 個の注釈の相違を調べる。ここで, システムは, コーパスと異なるタイミングで心的状態を推定することがある。そこで, 注釈とシステムの推定を一致とみなす条件を次のとおり定め, 満たされない注釈を相違点とする。

- 心的状態名, 内容, 属性, 確信性が一致すること

*2 「短絡的な部分」の定義は, 心的状態の活性化連鎖の関係により定義する。たとえば, 目標設定に後続してプランの立案を行う, プランの立案に後続して実行状況の予測を行う, 予測状況に対して評価を行う, 評価に応じてプランを選択する, などである。対話においてこの関係が成り立たないことがあるので, 手順 3 および 4 の時点で熟練者との議論を要する。

表 2: コーパスの試作結果 (校正前)

ターン数	発話行為 注釈数	明示的な 心的状態 の注釈数	非明示的な 心的状態 の注釈数	総注釈数
323	663	436	343	1,442

- システムの推定が、注釈の出現と同一ターン、または、注釈の出現より早いターンであること

こうして判定した結果、224 個の注釈が相違点として抽出された。

3.3 熟練者による評価と誤り傾向の分析

熟練者の指導の下で、コーパス作成者は、相違点における誤り原因がコーパス作成者であるか、情緒推定システムであるかを判別し、対策を立てた。

相違原因の判別結果を表 3 に、コーパス作成者の誤りの傾向を表 4 に示す。表 3 のその他とは、誤りの原因がどちらにあるのか判別できなかった場合である。

表 4 より、コーパス作成者による誤りのうち上位 3 つが全体の 9 割以上占めていた。コーパスの誤りの主な原因は、コーパス作成者に心的状態の連鎖関係の知識が不足していることである。

表 3: 注釈が推定されなかった原因とその件数 (校正前)

原因	注釈数
コーパス作成者が誤って付与した	138 個 (61.6%)
システムの不備で推定されなかった	58 個 (25.9%)
その他	28 個 (12.5%)
合計	224 個

表 4: コーパス作成者の誤りの傾向 (一部重複有り)

傾向	件数
項目「理由」の関係が不適切	60 件 (41.4%)
心的状態の連鎖関係を短絡	46 件 (31.7%)
信念の主の判断が不適切	25 件 (17.2%)
項目「内容・属性」が不適切	8 件 (5.6%)
漫画の絵や喋り方から推定	3 件 (2.1%)
発話行為の解釈誤りの影響	2 件 (1.4%)
情緒推定と無関係な注釈の付与	1 件 (0.7%)
合計	145 件

3.4 コーパスの校正

コーパスを校正する様子を図 3、図 4 を参照しながら説明する。

図 3 は校正前のコーパスである。誤りの箇所は、プラン pl01.2 の推定に必要な注釈の不足に起因している。つまり、プランを採用している状態 pl01.1 から、不採用の状態 pl01.2 を推定するには、「プラン」に対する「評価」が関わらないといけないが、「(カード売りの)おっちゃん帰った」という認識から短絡してコーパス作成者が推定している。

そこで、以下の手順で、図 4 のように校正する。

- (1) 初期状態の「プラン pl01.1」とターン 6 の「認識 cg02.1」から推定される「評価 ev02.1」を付与*3
- (2) (1) の「評価 ev02.1」の付与に伴い、理由部の「認識 cg02.1」を「評価 ev02.1」に変更

```

初期状態
pl01.1, プラン, 魔法カードを買いに行く, 採用, [], c
:
6 まる子:「おっちゃん帰ったみたいね。」
sa09.1, 発話行為, 認識の伝達, [], c
cg02.1, 認識, おっちゃんは帰った, [], [sa09.1], c
pl01.2, プラン, 魔法カードを買いに行く, 不採用,
[p101.1, cg02.1], c
close, pl01.1, [pl01.2]
    
```

図 3: 校正前のコーパス

```

初期状態
pl01.1, プラン, 魔法カードを買いに行く, 採用, [], c
:
6 まる子:「おっちゃん帰ったみたいね。」
sa09.1, 発話行為, 認識の伝達, [], c
cg02.1, 認識, おっちゃんは帰った, [], [sa09.1], c
ev02.1, 評価, 可能性, 低い, [pl01.1, cg02.1], c (1)
pl01.2, プラン, 魔法カードを買いに行く, 不採用,
[p101.1, ev02.1 (2)], c
close, pl01.1, [pl01.2]
    
```

図 4: 校正後のコーパス

以上のようにして、30 対話について校正を行った。結果を表 5 にまとめる。

表 5: コーパスの試作結果 (校正後)

ターン数	発話行為 注釈数	明示的な 心的状態 の注釈数	非明示的な 心的状態 の注釈数	総注釈数
323	665	451	374	1,490

表 2 と表 5 を比べると、発話行為が 2 個増え、明示的な心的状態が 15 個増え、非明示的な心的状態が 31 個増えた。そして、総注釈数は 48 個増えた。

3.5 校正後の相違点の評価

校正したコーパスを、再び情緒推定システムの推定結果と比較した。そして、相違点の抽出と評価を行った。

校正後のコーパスの誤り分析の結果を表 6 に示す。

表 3 と表 6 の比較より、「コーパス作成者の誤り」は 138 件から 12 件に減った。このことから適当な校正であったといえる。校正できなかった 12 件には「推定に必要な注釈が不足している」傾向があったが、必要な注釈が対話に明示されておらず、また、心的状態の活性連鎖の関係からも推定できなかったため、校正することができなかった。

一方、推定システムの不備によるものが 58 件から 190 件に増えた。これは、対話文から推論するには、世界知識およびルールが不足しているためである。また、一個の相違が、その後続く注釈の連鎖的な推定相違へと繋がったことも大きな要

*3 まる子はカードを買おうとしていたが、それを売っていたおじさんは既にいなくなっていたため、カードを手に入れる可能性は低いという連鎖

因であった。これら 190 個の注釈は、今後のルールベースの強化に役立つ。

表 6: 注釈が推定されなかった原因とその件数 (校正後)

原因	注釈数	
コーパス作成者が誤って付与した	12 個	(5.1%)
システムの不備で推定されなかった	190 個	(80.5%)
その他	34 個	(14.4%)
合計	236 個	

4. 訓練手順の評価実験

4.1 実験 1: 熟練度の評価

イソップ対話集 (6 対話, 58 ターン) を用いて, 3 章で訓練したコーパス作成者を評価する。

6 対話についての注釈数を表 7 に示す。

熟練者が判定したところ, 67 個のうち 3 個の注釈は誤りであった。誤り率は 4.5% (3/67) である。3 章で用いたまる子対話集の場合, 初期の誤り率が 40.2% (138/343) であったことと比較すると, 飛躍的な熟練度の向上が伺える。よって, 提案した手順は有効だといえる。

表 7: イソップ対話集についてのコーパス作成結果

ターン数	発話行為 注釈数	明示的な 心的状態 の注釈数	非明示的な 心的状態 の注釈数	総注釈数
58	87	50	67	204

4.2 実験 2: 自動判定の精度と効果

情緒推定システムを用いてイソップ対話集の情緒推定を行い, 4.1 節で作成したコーパスとの相違点に, どれだけコーパス作成者の誤りが含まれているか数えることで, 自動判定の精度を検討する。なお, ルールベースの構築にイソップ対話集を用いたので (2.2 節), システムにとってこの対話集への情緒推定は, クローズドテストとなる。

4.1 節より抽出目標の注釈数は 3 個である。相違点として 22 個が挙げられ, その中に目標の注釈が全て含まれていた。したがって自動判定による誤り抽出の精度は, 再現率 100% (3/3), 適合率 13.6% (3/22) である。

非明示的な心的状態 67 個のうち, 22 個が相違点として抽出されたので, その差の 45 箇所は熟練者が評価する必要がなくなった。つまり, 人手による判定作業のコストが 67%削減できたといえる。

5. おわりに

注釈の体系が複雑な情緒注釈付き対話コーパスを高い精度で作成するには熟練を要するため, 付与作業のできる者が限られていることが問題であった。そこで, 本稿ではコーパス作成者の熟練度向上のための訓練手順を提案した。

そして, 30 件の対話について, この手順を実施して, コーパス作成者の訓練を行った。その後, 提案手順を評価するため, 別の 6 件の対話について注釈付与を行い, コーパス作成者の熟練度の評価, および, 自動判定の精度・コスト効果の評価実験を行った。そして, 以下の結果を得た。

- 1 熟練者でない者がコーパスを作成する際, 心的状態の連鎖関係の知識不足に注意する必要がある。
- 2 提案した手順により, コーパス作成者の誤り率を 40.2% から 3.2%まで低減でき, コーパスの精度を向上させることができた。
- 3 訓練後, 新たな対話文において, コーパス作成者の誤り率が 4.5%であることから, 熟練度の向上が確認できた。
- 4 情緒推定システムにとってクローズドテストとなる対話文においては, システムによる相違点の自動抽出により, 人手での誤り判定作業のコストが 67%削減できた。

以上より, 本稿で提案したコーパス作成手順の有効性を示すことができた。また, 訓練手順自体は特別なものではないが, 明確で決定的な解答のない情緒注釈を付与するタスク, すなわち, 非明示的な心的状態を書き尽くすことを目標とするタスクを対象とした場合においても, 訓練の効果を確認することができた。

謝辞

ルールベースの強化を行った有田真康君 (平成 14 年度 鳥取大学工学部卒業生) に感謝します。

参考文献

- [徳久 98] 徳久雅人, 岡田直之: パターン理解的手法に基づく知能エージェントの情緒生起, 情報処理学会論文誌, Vol.39, No.8, pp.2440-2451, (1998) .
- [徳久 01] 徳久雅人, 中野育恵, 山下智之, 岡田直之: 情緒を加味した深いタスク指向の対話理解のためのルールベースの構築, 信学技報, TL2001-25, pp.21-28, (2001) .