

決定木に基づく対話行為のタグ付け支援

白井 清昭, 洪 陽杓

北陸先端科学技術大学院大学 情報科学研究科

本論文では、対話行為が付与された対話コーパスの構築を目的とし、対話行為のタグ付け作業を支援するツールについて述べる。このツールは、まず、対話を発話単位に自動的に分割する。さらに、作業者が発話単位を容易に変更できるインターフェースを提供する。次に、対話行為を推定する決定木を学習し、これを用いて発話に付与すべき対話行為の候補と信頼度のリストを提示する。実験の結果、決定木による対話行為推定の正解率は74.6%であった。最後に、対話行為を付与する発話に対して、その類似発話と過去のタグ付け事例を提示する。実験により、全体の81.7%の発話に対して対話行為タグ付けの参考となる適切な類似例文を提示できることがわかった。

Support for Dialog Act Annotation using a Decision Tree

Kiyoaki SHIRAI, Yoshaku KO

School of Information Science, Japan Advanced Institute of Science and Technology

This paper describes the tool to annotate a dialog corpus with dialog act tags. First, the tool divides a dialog into utterance units. It also provides a user interface to change boundaries of utterance units easily. Next, it provides a list of dialog act candidates with their reliability score for each utterance based on a decision tree trained from a corpus. The accuracy of dialog acts estimated by the trained decision tree is 74.6%. Finally, it indicates examples of tagging, i.e. similar utterances and their previously annotated dialog acts, for each utterance. Experimental results shows that examples presented by the tool are appropriate for 81.7% of utterances.

1 はじめに

近年、自然言語処理の様々な分野でコーパスを利用した機械学習に基づく手法が成功を収めている。対話研究の分野でもコーパスを利用した様々な研究が報告されている。しかし、新聞記事などのテキストコーパスと比べて、対話コーパスは十分な量のデータが整備されているとは言えない。そのため、対話コーパスの整備は重要な研究課題のひとつである。本研究では、談話タグの一つである対話行為をタグ付けしたコーパスの構築を目的とする。対話行為を自動的に決定する手法も数多く提案されてはいるが [3, 4, 7, 8, 9], その精度は十分高いわけではなく、対話行為タグ付けコーパスを自動的に構築することは困難である。したがって、自動的に決定された対話行為を修正するなど、人手による作業が必要不可欠となる。本研究は、人間による対話行為のタグ付け作業を支援する

様々な手法を提案し、それらを実装した対話行為タグ付けコーパス作成支援ツールを作成する。

2 提案システム

2.1 概要

対話行為タグ付け作業の労力を減らすため、主にマウス操作によってタグを付与する支援ツールを作成した。ツールを作成する際、タグ付けを行う対象となる対話コーパスとして交通経路案内タスクのコーパスを使用した。また、対話行為の定義として、荒木らによって提案された22個の対話行為 [1] を用いた。支援ツールのスナップショットを図1に示す。このツールの主な機能は以下の通りである。

1. 発話単位への分割

対話を発話単位に自動的に区切る。また、発話単位を簡単に変更することのできるインターフェー

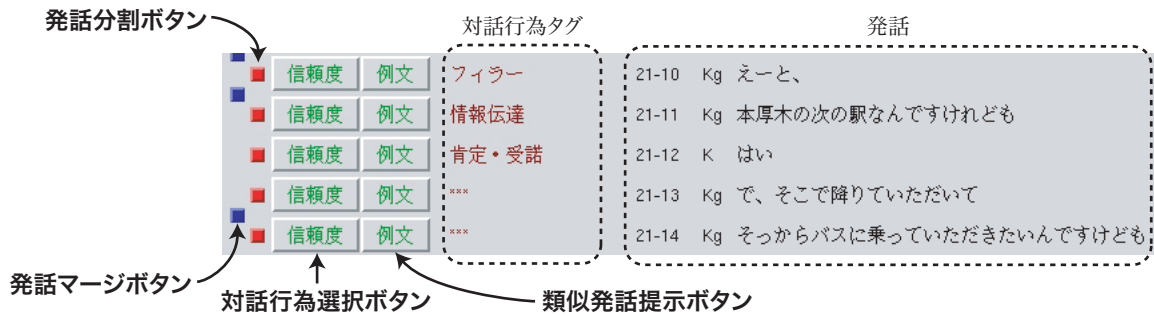


図 1: 支援ツールのスナップショット

スを提供する。図 1 の「発話分割ボタン」を押すと、発話が別ウィンドウで表示される。作業者はそのウィンドウ内で発話の区切りをマウスで指定することにより発話を分割できる。一方、「発話マージボタン」を押すと、そのボタンの前後の発話がマージされる。

2. 対話行為タグの候補の提示

図 1 の「対話行為選択ボタン」を押すと、対話行為の一覧が表示される。この際、タグ付けする発話に応じて各対話行為の信頼度 (スコア) を計算し、その高い順に作業者に提示する。作業者は、表示された対話行為のリストの中から正しいものをマウスで選択する。

3. 類似発話の提示

図 1 の「類似発話提示ボタン」を押すと、タグ付けする発話に似ている発話と、それに対する過去の対話行為のタグ付け結果を表示する。作業者は、対話行為の選択に迷ったとき、過去のタグ付け結果を参照することができる。

以下、上記 1.~3. の処理の詳細について述べる。

2.2 発話単位への分割

対話行為タグ付けの最初の作業は、対話を発話単位に区切ることである。本研究では、堂坂ら [2] が提案した発話単位の定義に準じて対話を発話単位に分割し、個々の発話単位に対して対話行為を付与する。発話単位への自動分割は以下のように行う。

- 話者情報を事前に入力し、話者交替があったとき自動的に発話単位に分割する。
- 節の境界で発話単位を分割する。節の境界を検出するために、「けれども」「ので」や句点のような節の境界を示唆するキーワードを 15 種類用意した。
- フィラーやあいづちの前後で発話単位を分割する。フィラーやあいづちを検出するために、「えーっと」

や「はい」のようなキーワードを 40 種類用意した。

- 二人の人が同時に発話したとき、話者毎に発話単位を分割する。

予備実験の結果、本手法の発話単位区切りの精度は 97.9%、再現率は 90.9%であった。また、発話単位への自動分割は常に正しいわけではないので、2.1 項で述べたようにマウス操作によって発話単位の修正を行うインターフェースを実装した。

2.3 対話行為タグの候補の提示

本項では、作業者に対話行為タグの候補を提示する手法について述べる。まず、少量のタグ付きコーパス (訓練コーパス) が存在すると仮定し、発話の対話行為を推定する決定木を学習する。次に、対話行為を決めたい発話 u を決定木に入力し、 u が到達する決定木の葉 l を求める。訓練コーパスにおける発話のうち、 l に分類される発話の集合を U_l とする。そして、 U_l の各発話に付与されている対話行為の集合 D_l を作業者に提示する対話行為の候補とする。また、 D_l における対話行為の確率分布 $P(d|D_l)$ を求め、これを対話行為 d の信頼度とし、この順に対話行為を並べて作業者に提示する。また、支援ツールでは、決定木による分類が正しくない場合も考えられるので、 D_l 以外の対話行為も信頼度 0 として対話行為候補のリストの最下位に提示する。

2.3.1 決定木の学習

決定木の学習には C4.5 [6] を用いた。学習に用いた素性は以下の通りである。

- 前/後の発話との話者交替
発話の話者が前/後の発話の話者と同じか否か。
- 話者の役割
対話行為を決める発話の話者がシステム側かユーザ側か。本研究で対象とする対話はタスク指向型の対話であり、一人が情報を提供するシステムの

表 1: 類似発話の提示

[示唆]
S: あ, S*: ではそちらの駅の方へ, 来ていただいて, U: はい
U: えーと, S*: 折れていただいて, U: はい
[情報伝達]
S: で, 武蔵野住宅前だとー, バス停降りてー, S*: しばらく, 歩いて行って, U: はい
U: ほう. S*: の実験で来ましたというふうに言えば, 通してくれますので, U: はい

せるが, これらの発話に異なる対話行為が付与されているとき, それらは作業者が判断に迷う対話行為であると考えられる. そこで, 発話を決定木で分類したときに到達する葉を l としたとき, l に分類される過去のタグ付け事例 (発話) に付与されている対話行為²を求め, これらを類似発話を提示する対話行為とする. また, 2.3.2 と同様に, 状況に応じて 4 種類の決定木を使い分ける.

2. 提示する発話の決定

1. で決めた対話行為について, 類似度の大きい N 個の発話を提示する. 二発話間の類似度は以下のように定義する.

(a) 決定木における共通上位ノードの深さ

決定木において近い位置に分類される発話に高い類似度を与える.

(b) 単語ベクトル間の類似度

発話に含まれる単語に対して式 (2) のような単語ベクトルを作成する.

$$\vec{v} = (x_1, x_2, \dots, x_n) \quad (2)$$

x_i は単語 w_i の重みであり, w_i が助詞, 助動詞などの付属語のときは 1, それ以外は 2 とする. 発話間の類似度は式 (2) の単語ベクトルのコサイン類似度とする.

発話の順位付けは, まず類似度 (a) にしたがって行う. (a) の類似度が同じときは類似度 (b) にしたがう. (a), (b) の類似度がともに同じときは, ランダムに順位付けする.

²これはツールが信頼度付きでユーザーに提示する対話行為タグの候補 D_l に等しい. 詳細は 2.3 項を参照.

表 2: 決定木の評価

決定木	上位一位	全て
$T_{0,0}$	68.1%	83.5%
$T_{0,1}$	68.1%	83.9%
$T_{1,0}$	74.6%	86.0%
$T_{1,1}$	72.8%	85.3%
ベースライン	24.0%	

3 評価実験

本節では提案手法の評価実験について述べる. 以下, 3.1 項では対話行為タグを推定する決定木 (2.3 項) の評価を行う. 3.2 項では, 類似発話の提示 (2.4 項) がどれだけ作業者に有効であるかを評価する.

3.1 決定木の評価

まず, 2.3.1 で述べた素性を用いて, 対話行為を推定する決定木を学習した. 訓練データとして, 正しい対話行為タグが付与された 1,224 発話からなる対話コーパスを用いた. 2.3.2 で述べたように, 前後の発話の対話行為を素性として用いるか否かによって 4 種類の決定木を学習した. 次に, 訓練データとは異なる 279 発話からなるコーパスをテストデータとし, 決定木が出力する対話行為タグの評価を行った. 結果を表 2 に示す. 表 2 において, 「上位一位」は決定木が出力する上位一位の対話行為が正解である割合を, 「全て」は決定木が出力する全ての対話行為の候補の中に正解が含まれる割合を表わす. 「ベースライン」は訓練データにおける最頻出対話行為 ([肯定・受諾]) を常に選択したときの正解率である. なお, 決定木が出力する対話行為の候補数の平均は 3.05 であった.

表 2 から, 4 種類の決定木の中では $T_{1,0}$ の正解率が一番高いことがわかる. $T_{1,0}$ の上位一位の対話行為の正解率は 74.6% である. これは, 例えば Reithingerらの手法 [7] による 74.7% に匹敵する値である. 対話行為タグの種類, 訓練データ量, 言語などの実験条件が異なるため単純に比較することはできないが, 学習した決定木による対話行為タグの自動推定は現在の機械学習手法によって得られる高い水準に達しているといえる. また, 本研究では対話行為タグ付け作業の支援を目的としているため, 必ずしも上位一位の対話行為が正しくなくてもよい. 決定木が出力する対話行為タグの中に正解が含まれていればよいと考え, 83~86% の発話について有効な支援が行えるとみなせる. また, 前の発話の対話行為を素

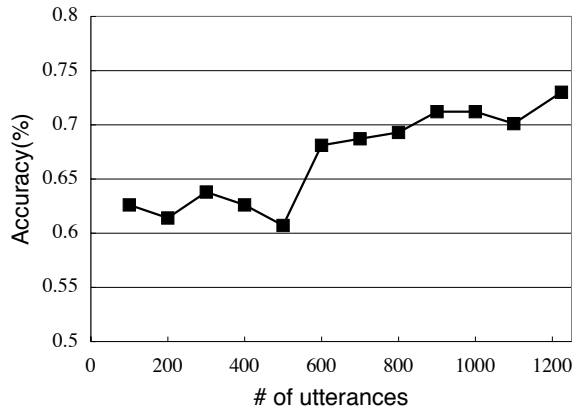


図 2: 決定木 $T_{1,0}$ の訓練データ量と正解率

性として使う決定木と使わない決定木では大きな差があるが、後の発話の対話行為については正解率の差はほとんどない。したがって、今回の実験では、後の発話の対話行為は前の発話の対話行為ほど有効な素性ではないことがわかった。

図 2 は、訓練データ量を変化させたときの決定木 $T_{1,0}$ の正解率を示したグラフである。この結果から、訓練データ量を増やせば増やすほど正解率が向上することがわかる。また、今回の実験で用いた訓練コーパスの量は、1,224 発話であるが、この付近でも正解率は上昇する傾向にあるため、より多くの訓練データを用意することによって正解率の更なる向上が期待できる。例えば、対話行為タグ付け作業が完了した発話を訓練コーパスに逐次加え、決定木を再学習することによって、タグ付け作業の途中の段階でもツールが提示する対話行為の候補の正確性を増すことも可能であろう。

決定木 $T_{1,0}$ が出力する上位一位の対話行為の正解率を対話行為毎に調べたのが表 3 である。ただし、 O_{train} , O_{test} はそれぞれ訓練データ、テストデータにおける対話行為タグの出現頻度である。表 3 から、訓練データにおける出現頻度の高い対話行為ほど決定木による推定精度が高い傾向があることがわかる。

3.2 類似発話の評価

前項で決定木の評価に用いた 279 発話について、2.4 項の手法にしたがって類似発話を提示し、それが正しい対話行為の判定の参考になるかどうかを調べた。提示すべき過去のタグ付け事例として、決定木の学習に用いた 1,224 発話を用いた。決定木は $T_{1,0}$ を用いた。結果を表 4 に示す。

表 4 の (A) は類似する発話が 1 つも見つからない

表 3: 対話行為別の正解率 ($T_{1,0}$)

対話行為	O_{train}	O_{test}	正解率
肯定・受諾	312	67	94.0%
情報伝達	231	45	91.1%
フィラー	219	75	85.3%
了解	89	17	82.4%
あいさつ	25	8	62.5%
未知情報応答	16	7	57.1%
示唆	39	13	46.2%
その他の働き掛け	30	8	37.5%
確認	120	13	30.8%
未知情報要求	17	8	25.0%
依頼	23	5	20.0%
真偽情報要求	27	7	14.3%
約束・申し出	5	3	0%

表 4: 類似発話の評価

	発話数
(A) 類似発話なし	5 (1.8%)
(B) 正解の対話行為なし	36 (12.9%)
(C) 参考になる発話がない	10 (3.6%)
(D) 参考になる発話がある	228 (81.7%)

場合を表わす。また、(B) は提示した類似発話の中に正解の対話行為が存在しない場合を表わす。言い換えれば、ある発話について決定木を辿ったとき、(A) は到達した葉の中に分類される過去のタグ付け事例が 1 つもないことを、(B) は過去のタグ付け事例の中に正解の対話行為を持つ発話がないことを示している。いずれの場合も、正解となる対話行為に関して類似発話を表示できていないので、作業者にとって有効ではない。本研究では、類似発話を提示する対話行為の候補は決定木によって決めているため、正解の対話行為に関する類似発話を提示できるかどうかは決定木の正解率に依存する。したがって、(A),(B) の割合が低くなるように改善することは、対話行為を推定する決定木の正解率を向上させることと深く関係する。

一方、表 4 の (C) は、正解となる対話行為について類似発話を提示しているが、その対話行為が正しいかどうかの参考にならない場合を表わす。これは、提示した類似発話が元の発話とあまり似ていないため、そのタグ付け結果を参考にしてよいかかわらな

い場合である。この割合を低くするためには、まずタグ付け事例の量を増やすことが必要である。参照する事例の数が増えれば増えるほど、類似した発話が見つかる可能性は高くなる。また、発話間の類似度の定義を改良することも必要であろう。例えば、式(2)の単語ベクトルを作成する際、文末の単語の重みを重くしたり、前後の対話行為をベクトルに加えたりすることによって、類似する発話の順位が上位になる可能性がある。

表4の(D)は対話行為タグ付けの参考となる発話が提示できた場合を表わす。したがって、約8割の発話に対して、タグ付けの参考となる類似発話が提示できたことがわかる。「あいさつ」や「フィラー」などのようにほとんど迷いなく対話行為を決めることができる場合も含まれてはいるが、約8割の発話が(D)に該当することから、本研究で提案する類似発話の提示が対話行為タグ付け作業の有効な支援になりうるといえる。

4 関連研究

対話行為タグ付きコーパスを作成する手法のひとつは、発話の対話行為を自動的に決定するシステムを構築することである。近年では、教師あり機械学習の手法を用いて対話行為を自動的に決定する研究がいくつか行われている。Reithingerらは、発話に含まれる単語のn-gramや前の発話の対話行為などを用いて、対話行為を決定する確率モデルを学習した[7]。彼らのモデルによる対話行為タグ付けの正解率は74.7%であった。一方、Samuelらは、対話行為の決定にTransformation-Based Learning(TBL)という機械学習手法を用いた[8]。Reithingerらと同じ対話コーパスを用いて実験を行い、彼らよりも少ない訓練データで71.2%の正解率が得られたと報告している。木村らは、日本語の対話コーパスを対象に、Samuelらと同じくTBLを用いて対話行為を決定するモデルを学習し、72.6%の正解率を得ている[4]。しかし、これらの手法の精度は自動的に対話行為タグ付けコーパスを構築できるほど高くはない。したがって、本研究のように人間の対話行為のタグ付け作業を支援する技術もまた重要である。

対話行為タグ付けを支援することを目的とした研究として駒谷らの研究がある[5]。彼らもまた、本研究と同じように、対話行為タグ付けを支援するGUIツールを作成している。駒谷らのツールでは、作業者に提示する対話行為の候補はルールベースの手法

によって決定している。これに対し、本研究では機械学習の手法を用いている。機械学習を用いる利点としては、対話行為推定の精度が高いこと、作業が進むにつれて利用可能な訓練データの数が増えることなどが挙げられる。訓練コーパスとして正しい対話行為がタグ付けされたコーパスが必要だが、コーパス作成作業の最初の段階で少量のタグ付きコーパスを用意することはそれほど大変ではないだろう。また、類似発話を提示する機能を実装している点も本研究の特徴のひとつである。

5 おわりに

本研究では、対話行為タグ付けコーパスの作成を支援するツールについて述べた。また、対話行為を推定する決定木と類似発話を提示する手法の評価を行い、いずれも対話行為タグ付けの支援に有効であることがわかった。今後の課題として、決定木の正解率を向上させることや、決定木以外の学習アルゴリズムの導入を検討することなどが挙げられる。また、対話行為を決定する際に、音声から得られる韻律などの情報は重要な手がかりとなる[3]ので、実際の発話の音声データを再生する機能も現在の支援ツールに組み込みたい。

参考文献

- [1] 荒木雅弘, 伊藤敏彦, 熊谷智子, 石崎雅人. 発話単位タグ標準化案の作成. 人工知能学会誌, Vol. 14, No. 2, pp. 251-260, 1999.
- [2] 堂坂浩二, 島津明. タスク指向型対話における漸次的発話生成モデル. 情報処理学会論文誌, Vol. 37, No. 12, pp. 2190-2200, 1996.
- [3] 細木光宏, 樽松明. 韻律を用いた自由発話対話の意図の判定. 人工知能学会研究会資料, SIG-SLUD-A003-5, pp. 25-30, 2001.
- [4] 木村幸彦, 新美康永, 荒木雅弘, 西本卓也. 機械学習を用いた談話タグ推定手法. 人工知能学会研究会資料, SIG-SLUD-A003-8, pp. 43-48, 2001.
- [5] 駒谷和範, 荒木雅弘, 堂下修司. 対話コーパスにおける発話単位タグの一推定手法. 人工知能学会誌, Vol. 14, No. 2, pp. 273-281, 1999.
- [6] J. Ross Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, 1993.
- [7] Norbert Reithinger and Martin Klesen. Dialogue act classification using language models. In *Proceedings of the EuroSpeech*, pp. 2235-2238, 1997.
- [8] Ken Samuel, Sandra Carberry, and K. Vijayshanker. Dialogue act tagging with transformation-based learning. In *COLING-ACL 98*, pp. 1150-1156, 1998.
- [9] 高野健治, 島津明. 談話構造を用いた対話行為の認識. 言語処理学会第9回年次大会, pp. 699-702, 2003.