

モダリティ表現認識・事象の事実性解析の同時学習

友利 涼^{1,a)} 村脇 有吾^{1,b)} 松吉 俊^{3,c)} 亀甲 博貴^{2,d)} 森 信介^{2,e)}

概要：モダリティ表現や事象の事実性などを正確に認識・解析することは、否定や推量などが含まれるテキストの言語理解や文生成を行ううえで重要である。本研究では、モダリティ表現認識器・事象の事実性解析器などをマルチタスク学習の枠組みを用いて同時学習する。これらのタスクは相互に関連しており、マルチタスク学習を用いてその関係性を自動的に学習することを目指す。実験では、同時学習による手法が単純な手法による精度を上回った。また、生コーパスを用いて事前学習することにより、さらなる精度向上を示した。

Joint Learning for Modality Expression Recognition and Event Factuality Analysis

1. はじめに

一般に、テキストには事象^{*1}に対する情報発信者の主観的な態度が多く含まれる。情報発信者の主観的な態度はモダリティと呼ばれ、それを示唆するテキスト中の表現をモダリティ表現という。事象とモダリティの例を以下に示す(下線は注目している事象)。

- (1) 太郎はお土産にモナカを買うかもしれない。
- (2) 綺麗な運河を観光したい。
- (3) 残念ながら、その列車では昼前に着かない。
- (4) 午後も晴れていれば自転車で街を走り回ります。

文(1)では、事象「太郎がお土産にモナカを買うこと」に対して、モダリティ表現「かもしれない」によって推量の態度が示されている。文(2)では、モダリティ表現「たい」に

よって希望が表明されている。文(3)においては、事象「その列車で昼前に着くこと」の成立をモダリティ表現「ない」により否定している。文(4)では、モダリティ表現「ば」により、仮定の出来事であることが表明されている。

テキストに存在するモダリティ表現を認識し、その意味ラベルを特定するタスクは、モダリティ表現認識と呼ばれる。このタスクの直接的な応用先は、上の例で示唆されるように、事象のモダリティを解析するタスクである [1]。その他の応用として、述語文節の正規化や平易化などがある [2]。

事象に対してそのモダリティを解析するタスクをモダリティ解析と呼ぶ。モダリティのうち、情報発信者による事象の成否判断のモダリティを(情報発信者による)事実性と呼び、モダリティ全体でなく、事実性のラベルのみを解析するタスクを事実性解析という [3]。事実性解析では、事象の成否判断を情報発信者が表明する確信度とともに解析する。モダリティ解析や事実性解析は、情報抽出 [4] やテキスト含意認識 [5]、テキストマイニング [6] などの応用に有用である。

上の例で示唆されるように、モダリティ表現認識と事実性解析を含むモダリティ解析は、互いに密接に関係したタスクである。先行研究では、モダリティ表現認識器とモダリティ表現の意味ラベルごとに作成した複雑なルールを用いて、パイプライン的に事実性解析を行う手法 [1] が提案されている。また、単純にそれぞれのタスクの認識器・解析器を独立に構築する手法もよく用いられる。しかし、各

¹ 京都大学 大学院情報学研究所
Graduate School of Informatics, Kyoto University
² 京都大学 学術情報メディアセンター
Academic Center for Computing and Media Studies, Kyoto University
³ 電気通信大学 大学院情報理工学研究所
Graduate School of Informatics and Engineering, The University of Electro-Communications
a) tomore.suzushi.72e@st.kyoto-u.ac.jp
b) murawaki@i.kyoto-u.ac.jp
c) matuyosi@uec.ac.jp
d) kameko@i.kyoto-u.ac.jp
e) forest@i.kyoto-u.ac.jp

*1 ここでいう事象とは、述語項構造によって表現される命題であり、行動や出来事のみでなく、述語項構造によって表現される状態や状況も含む。

タスクの認識器・解析器を独立に構築する手法では各タスク間の複雑な関係性を考慮できず、人手で複雑なルールを作成することはコストが高いため、コーパスからそれらの関係性を学習する手法が求められている。

そこで本研究では、複数のタスクを同時に解くマルチタスク学習手法 [7], [8] を用いて、モダリティ表現認識やモダリティ解析などを解く手法を提案する。提案手法はニューラルネットワークを用いており、各タスクの出力層は、文のエンコーダ部分を共有しながら学習する。文のエンコーダ部分を共有することで各タスクの関係性を暗黙的に学習することが期待できる。本研究では、モダリティ表現やモダリティがアノテーションされているゲーム解説コーパス [9] を用いる。このコーパスはアノテーション済みテキストが少量であるため、大量の生コーパスを用いてエンコーダを事前学習する場合との比較も行う。実験において、提案手法による精度が単純なベースラインを上回ったことを示す。

本論文は、以下のように構成される。第2節では、モダリティ表現やモダリティの関連研究、マルチタスク学習の関連研究について述べる。第3節では本研究で使用するコーパスの詳細と対象タスク、問題の定式化について述べる。第4節で提案手法を説明し、第5節で実験結果を述べる。最後に、第6節でまとめる。

2. 関連研究

2.1 モダリティ表現

英語における主要なモダリティ表現は、助動詞である“must”や“may”などである。文献 [10] の TimeML において、これらのモダリティ表現は、事象を表す <MAKEINSTANCE> タグの @modality 属性に記述される。

日本語において、モダリティ表現の多くは機能語や複合辞からなる機能表現である。機能表現辞書 [11] からモダリティ表現を認識する手法 [2], [12] や、機能表現がアノテーションされたコーパスと条件付き確率場 (CRF) [13] を用いてモダリティ表現を認識する手法 [1] が提案されている。

「予測する」や「希望する」、「中止する」、「崩れる」のように、本動詞であるが、別の事象に対するモダリティ表現と解釈すべき語も存在する [14]。本研究で用いるゲーム解説コーパス [9] では、このような機能表現以外のモダリティ表現もアノテーションされている（例えば、表1における「崩れ」）。

2.2 モダリティ

情報発信者が表明する態度や成否判断などの、事象に対する総合的な情報はモダリティと呼ばれる [15]。本研究で用いるゲーム解説コーパスでは、事象らしさの情報や事象に対する情報発信者の態度が事象クラスとして、成否判断情報が事実性としてアノテーションされている。

英語では、事象とその時制などがアノテーションされた TimeML が提案されている [10]。文献 [16] では、TimeML の体系の上に、事象の成立に対する態度表明者の確信度とその成否極性がアノテーションされている。文献 [17] ではモダリティ表現などを用いて、事象の入れ子の外側から順に事実性を解析している。ニューラルネットワークモデルによる事実性解析 [18], [19] や文書単位の事実性解析モデル [20] など提案されている。

日本語においては、現代日本語書き言葉均衡コーパス (BCCWJ) [21] のコアデータのうち、新聞、書籍、Yahoo!知恵袋、白書のレジスターに対して、事象をラベル付与の対象とし、モダリティをアノテーションした研究がある [15]。文献 [15] では、タグ項目間の依存関係を考慮できる CRF である GRaphical Models in Mallet (GRMM) [22] を用いてモダリティ解析器を構築している。文献 [23] では、複数の事象間の依存関係を考慮できる Factorial CRF を用いてモダリティと時間ごとの事象の成否極性を解析する手法が提案されている。文献 [1] では、否定の機能表現なら事実性の成否極性を反転、意志なら確信度を減少させるなど、認識した機能表現の意味ラベルごとにルールを記述し、事実性解析を行っている。

2.3 系列ラベリング問題とマルチタスク学習

本研究では、それぞれのタスクを系列ラベリング問題として解く (第3節)。近年、系列ラベリングのタスクはニューラルネットワークベースの手法 [24], [25] によって解かれることが多い。これらの手法は単語や文のベクトルを生コーパスから事前学習することによって高い精度を実現している。系列ラベリング問題をマルチタスク学習によって解く手法もいくつか提案 [8] されており、品詞推定やチャンキングなどのタスクで成果を挙げている。文献 [26] では、いくつかデータセットの組み合わせのマルチタスク学習で実験・比較し、限られた組み合わせのときのみ精度が向上していることが示されている。

3. ゲーム解説コーパスと問題設定

3.1 ゲーム解説コーパス

ゲーム解説コーパス [9] は、将棋の対局について対局者以外の将棋熟練者が作成した解説文から構成されている*2。解説文のほとんどは局面に対するコメントであり、推量や予測が多く含まれる。これらの解説文の一部に対し、モダリティ表現や事象クラス、事実性が異なる層でアノテーションされている。本論文でいう事象クラスとは、モダリティから事実性を除いた、事象らしさの情報や情報発信者の態度情報である。2.2節で述べた文献 [15] のモダリティ

*2 本研究では使用していないが、コーパスには解説文が参照している局面の情報も含まれる。

表 1 ゲーム解説コーパスにおけるアノテーション例

単語	先手	は	美濃	囲い	が	崩れ	て	い	る	の	で	飛車	交換	は	後手	の	得	に	な	り	そう	だ	。	
モダリティ表現	O	O	O	O	O	ME _n -B	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
事象クラス	O	O	O	EVe	O	EVe	O	EVf	O	O	O	O	O	EV _i	O	O	O	O	O	O	O	O	O	O
事実性	O	O	O	FNc	O	FPc	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O	O
固有表現	Tu-B	O	Ca-B	Ca-I	O	Ao-B	O	O	O	O	O	O	Mn-B	Mn-I	O	Tu-B	O	Ee-B	O	Ao-B	O	O	O	O

ラベルの分布と比較^{*3}すると、ゲーム解説コーパスの事象クラスのラベル分布は新聞や書籍に対するものに近く、事実性のラベル分布は新聞やYahoo!知恵袋に対するものに近い。また、このコーパスは将棋解説文に特有の固有表現もアノテーションされており [27]、いくつかの固有表現インスタンスはその周辺の事象と密接な関係がある。

本研究で用いるゲーム解説コーパスにおけるテキストとそのアノテーションの例を表 1 に示す。このコーパスは、超短単位^{*4}を分割基準としており、本論文ではこの基準に従って分割された文字列を単語と呼ぶ。各タスクについてのアノテーションは単語ごとに行われている。

以下では、それぞれのアノテーションについて簡単に説明する。

3.1.1 モダリティ表現

ゲーム解説コーパスではモダリティ表現に、事実性関連や時間関連の 8 種類の意味ラベルのうち 1 つが付与されている。ラベルの一覧を表 5 に示す。第 2.1 節で前述したように、モダリティ表現の多くは機能表現であるが機能表現以外のモダリティ表現もアノテーションされている。アノテーションコスト削減のため、モダリティ表現が対象とする事象との対応はアノテーションされていない。

モダリティ表現は複数の単語から構成されることもあるため、BIO2 ラベル体系に従ってアノテーションされている。主ラベルに続く B はあるモダリティ表現の最初の単語、I は同種のモダリティ表現の継続、O はモダリティ表現以外の単語を表す。BIO2 ラベル体系では、モダリティ表現の層において合計で 17 (= 8 × 2 + 1) のラベルがある。

3.1.2 事象クラス

ゲーム解説コーパスでは、事象らしさ関連や態度関連の 8 種類の事象クラスが定義されている。ラベルの一覧を表 6 に示す。事象クラスのラベルは、事象を構成する核の述語 1 単語のみに付与されており、それ以外の単語には O ラベルが付与される。例えば、表 1 の事象「飛車交換が後手の得になること」の事象クラスは EV_e であり、そのラベルは単語「な」に対し付与されている。

3.1.3 事実性

事象の成否判断とその確信度として 6 つのラベルが定義されている。ラベルの一覧を表 7 に示す。事実性のラベルは、事象クラスが EV_e の事象に対して付与されており、その他については O ラベルが付与されている。

3.1.4 固有表現

将棋解説分野に特有の固有表現タイプが 21 種類定義されており [27]、BIO2 ラベル体系を用いてアノテーションされている。

3.2 問題設定

本研究では、入力単語列 $x_1, x_2, \dots, x_n, \dots, x_N$ に対し、各タスク m ($\in M$) のラベル列 $y_m = y_{m,1}, y_{m,2}, \dots, y_{m,n}, \dots, y_{m,N}$ を推定する問題設定を考える。ここで M を対象とするタスク全体の集合とする。本研究の対象とするタスクは、モダリティ表現認識と事象クラス解析、事実性解析、固有表現認識である。

4. 提案手法

図 1 に提案手法の概要を示す。第 3 節で述べたような複数の関連するタスクを高性能に解くために、マルチタスク学習を用いる。マルチタスク学習はそれぞれのタスクのために独立のモデルを構築するのではなく、エンコーダ部分のパラメータを共有しながらそれぞれのタスクの出力層を構築することで各タスクの関係性を暗黙的に学習できると考える。

4.1 マルチタスク学習

入力単語列 $x_1, x_2, \dots, x_n, \dots, x_N$ をエンコーダに入力することで各単語ごとのベクトル表現 $h_1, h_2, \dots, h_n, \dots, h_N$ を得る。これを各タスクの出力層への入力とする。本研究では、エンコーダとして両方向の long short-term memory (LSTM) [28] を組み合わせた BiLSTM と BERT [29] の 2 つを用いる (第 4.2 節)。

それぞれのタスク m に対し、出力層として CRF [13] をエンコーダの上に追加する。各 CRF はエンコーダから得られるベクトル列を受け取り、以下の式のように、 n 番目の単語ごとに出力スコアを計算する。

$$o_{m,n} = W_m h_n + b_m$$

ここで $o_{m,n}$ の次元数はタスク m でのラベル数である。出力スコアとラベルの遷移スコアを用いてラベル列の出現確率を計算する。

*3 ただしラベル体系が異なる部分もあるので単純に比較できるわけではない。

*4 文献 [21] の短単位に対して活用語尾を分割したもの。

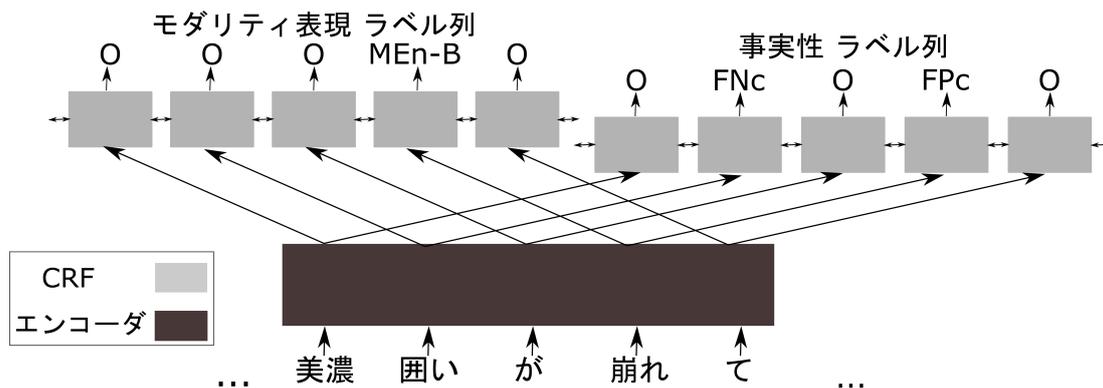


図 1 マルチタスク学習を用いたモダリティ表現認識と事象クラス解析、事実性解析、固有表現認識

$$p(\mathbf{y}_m | \mathbf{o}_m, \mathbf{T}_m) = \frac{\prod_{n=1}^{N+1} \exp(\mathbf{o}_{m,n}^{y_{m,n}} + \mathbf{T}_m^{y_{m,n-1}, y_{m,n}})}{\sum_{\mathbf{y}'_m \in \mathcal{Y}_m} \prod_{n=1}^{N+1} \exp(\mathbf{o}_{m,n}^{y'_{m,n}} + \mathbf{T}_m^{y'_{m,n-1}, y'_{m,n}})}$$

ここで $\mathbf{o}_{m,n}^{y_{m,n}}$ は $\mathbf{o}_{m,n}$ におけるラベル $y_{m,n}$ のスコア、 $\mathbf{T}_m^{y_{m,n-1}, y_{m,n}}$ はラベル $y_{m,n-1}$ から $y_{m,n}$ への遷移スコアを表し、 $n=0$ と $n=N+1$ のときの $y_{m,n}$ はそれぞれ文頭と文末の特別なラベル BOS と EOS とする。

タスク m の目的関数は、学習データの集合 D_m を用いて

$$NLL_m = - \sum_{D_m} \log p(\mathbf{y}_m | \mathbf{o}_m, \mathbf{T}_m)$$

と表される。最終的な目的関数は

$$NLL = \sum_{m \in M} \alpha_m NLL_m$$

で表され、この式の値を最小化することでモデルの学習を行う。ここで $\alpha_m \geq 0$, $\sum_{m \in M} \alpha_m = 1$ である。単純に、タスクごとの目的関数の和を目的関数とするといくつかのタスクの目的関数が収束しない可能性があるため、パラメータ更新のたびに α_m を MGDA アルゴリズム [30] を用いて調整する*5。

4.2 エンコーダ

本研究では、それぞれのタスクの入力に用いる単語ごとのベクトルを得るために 2 種類のエンコーダを別々に用いる。

BiLSTM はリカレントニューラルネットワークの拡張で、入力系列の長距離の依存関係を考慮できる。まず、入力の各単語 x_n をベクトル \mathbf{e}_n に変換し、順方向の LSTM への

*5 予備実験では MGDA アルゴリズムを用いた手法はタスクごとに重みづけしない目的関数を用いた場合よりもわずかながら精度が高かった。

入力とする。ここで n 番目の順方向の LSTM の出力は、順方向の LSTM の直前の出力を用いて以下の式で表される。

$$\vec{\mathbf{h}}_n = \text{LSTM}_f(\mathbf{e}_n, \vec{\mathbf{h}}_{n-1})$$

同様に、 $n+1$ 番目の逆方向の LSTM の出力を用いて n 番目の逆方向の LSTM の出力 $\overleftarrow{\mathbf{h}}_n$ を得る。その後、以下の式のようにそれぞれの方向の LSTM の出力を連結し、BiLSTM の出力を得る。

$$\mathbf{h}_n = [\vec{\mathbf{h}}_n : \overleftarrow{\mathbf{h}}_n]$$

BERT [29] は近年提案されたエンコーダであり、大量の生コーパスから事前学習する。入力単語列のいくつかの単語をマスクし、それを予測することで生コーパスから文や単語のベクトルを事前学習する。その後、タスクごとに出力層を追加しファインチューニングすることで様々なタスクで高精度を達成している。BERT をマルチタスク学習のエンコーダとして用いることで、少量のアノテーションデータからでも高い精度のモデルを構築できることが期待できる。

5. 実験

5.1 実験設定

表 2 に実験で用いたコーパスの諸元を示す。事前学習には日本語の Wikipedia データを用い、学習や評価にはゲーム解説コーパス [9], [27] を用いた。学習や評価の際、入力にはゲーム解説コーパスにアノテーションされている正解単語列を用いた。事前学習用のコーパスは単語列に分割されていないため、KyTea*6 [31] を用いて単語列を自動推定した。ハイパーパラメータは開発データを用いて調整した。

5.2 比較手法

ベースラインとしてスパースな素性 (sparse) を入力とする CRF モデルをタスクごとに構築した。素性として注目している単語の周辺の単語 n -gram と品詞 n -gram

*6 <http://www.phontron.com/kytea/index-ja.html>

表 2 コーパス諸元

	文数	単語数	固有表現数	モダリティ表現数	事象数	事実性のラベル総数
事前学習	18,697,554	602,326,608	-	-	-	-
学習	1,325	23,370	6,658	1,014	3,231	1,982
開発	247	4,073	1,596	196	600	390
テスト	469	6,744	2,103	413	1,093	719

表 3 sparse で用いた素性 (x_n は注目している単語、 pos_n はその単語の品詞)

$x_{n-2}, x_{n-1}, x_n, x_{n+1}, x_{n+2},$
$(x_{n-2}, x_{n-1}), (x_{n-1}, x_n), (x_n, x_{n+1}), (x_{n+1}, x_{n+2}),$
$(x_{n-2}, x_{n-1}, x_n), (x_{n-1}, x_n, x_{n+1}), (x_n, x_{n+1}, x_{n+2}),$
$pos_{n-2}, pos_{n-1}, pos_n, pos_{n+1}, pos_{n+2},$
$(pos_{n-2}, pos_{n-1}), (pos_{n-1}, pos_n),$
$(pos_n, pos_{n+1}), (pos_{n+1}, pos_{n+2}),$
$(pos_{n-2}, pos_{n-1}, pos_n), (pos_{n-1}, pos_n, pos_{n+1}),$
$(pos_n, pos_{n+1}, pos_{n+2})$

($1 \leq n \leq 3$) を用いた (表 3)。ゲーム解説コーパスには品詞情報は含まれていないため、品詞は KyTea で推定した。

エンコーダに BiLSTM と BERT を用いたベースラインとして、各タスクごとにエンコーダと出力層を学習し、パラメータの共有は行わないモデルを構築した。これをエンコーダのパラメータを共有する提案手法 (+multi) と比較する。

BiLSTM をエンコーダとして用いた際には、表 2 の事前学習コーパスと GloVe [32] を用いて単語のベクトルを事前学習し、エンコーダへの入力とした。BERT のエンコーダを事前学習する際には、KyTea で分割された単語列に対し、WordPiece [33] を用いてサブワードへと分割し、得られたサブワード列を入力して学習した。サブワードの語彙サイズは 32,000 とした。ファインチューニング時において、BERT のエンコーダはサブワードごとにベクトルを出力するが、単語を構成する最初のサブワードのベクトルを抽出し、そのベクトル列を出力層への入力とした。

勾配降下法として Adam [34] を用いた。その他のハイパーパラメータを付録の表 A.1, A.2 に示す。

5.3 結果と分析

評価指標には、以下の式で計算される適合率、再現率、F 値を用いた。

$$\text{適合率} = \frac{\text{正解数}}{\text{システムが出力したインスタンス数}}$$

$$\text{再現率} = \frac{\text{正解数}}{\text{コーパスに出現するインスタンス数}}$$

$$\text{F 値} = \frac{2 \times \text{適合率} \times \text{再現率}}{\text{適合率} + \text{再現率}}$$

事象クラスや事実性のラベルに関しては O ラベル以外に出現したラベルの数、モダリティ表現や固有表現に関しては、BIOES2 ラベル体系で B が付いているラベルの数をインスタンス数とした。

表 4 各タスクの精度 (太字は比較手法の中で最も高い精度を示し、下線は 2 番目に高い精度を示す)

	Model	適合率	再現率	F 値
モダリティ表現	sparse	0.811	0.550	0.655
	BiLSTM	0.763	0.593	0.668
	+multi	<u>0.810</u>	0.588	<u>0.682</u>
	BERT	0.719	<u>0.627</u>	0.670
	+multi	0.755	0.634	0.689
事象クラス	sparse	0.798	0.664	0.725
	BiLSTM	0.767	0.676	0.719
	+multi	0.805	0.719	0.760
	BERT	0.803	<u>0.745</u>	<u>0.773</u>
	+multi	<u>0.804</u>	0.752	0.777
事実性	sparse	0.803	0.675	0.733
	BiLSTM	0.796	0.746	0.770
	+multi	<u>0.823</u>	0.748	0.784
	BERT	0.812	0.791	<u>0.801</u>
	+multi	0.837	<u>0.775</u>	0.805
固有表現	sparse	0.881	0.708	0.785
	BiLSTM	0.870	0.760	0.811
	+multi	<u>0.879</u>	0.757	0.813
	BERT	0.865	<u>0.868</u>	<u>0.866</u>
	+multi	0.864	0.873	0.868

実験の結果を表 4 に示す。提案手法である BERT+multi がすべてのタスクにおいて最も高い F 値を達成した。モダリティ表現と固有表現の適合率を除き、全体的な傾向としてニューラルネットワークベースの手法がスパースな素性である sparse を大幅に上回った。BERT をエンコーダに用いたモデルは他の手法と比べ再現率が大幅に向上している。提案手法であるマルチタスク学習によるモデルでは、わずかながら一貫して精度の向上が確認できた。

タスクごとの詳細な結果として、BERT+multi のモデルでのモダリティ表現や事象クラス、事実性のラベルごとの精度をそれぞれ表 5, 6, 7 に示す。

表 5 において、MEp のものを除き、モダリティ表現の出現数は多くない。ゲーム解説コーパス全体における出現数上位 10 個のモダリティ表現を表 8 に示す。これらがモダリティ表現トークンの約 65% を占めており、モダリティ表現の出現数分布はロングテールになっている。学習データにおけるモダリティ表現の異なり数は 154 であるが、テストデータに対する文字列完全一致のカバレッジは約 36% のみであった。モダリティ表現認識の精度を上げるためには、このロングテール問題を考慮する必要がある。

表 6 を見ると、断定・推量を表す EVe の抽出精度は比較

表 5 BERT+multi によるモダリティ表現のラベルごとの認識結果

ラベル	説明	出現数	適合率	再現率	F 値
MEy	確実な肯定	21	0.333	0.048	0.083
MEa	肯定の可能性	65	0.559	0.292	0.384
ME0	可能性保留	31	0.548	0.548	0.548
MEm	否定の可能性	8	1.000	0.625	0.769
MEn	確実な否定	57	0.800	0.702	0.748
MEp	過去	167	0.806	0.868	0.836
MEf	未来	20	1.000	0.350	0.519
MEh	仮定	44	0.697	0.523	0.597

表 6 BERT+multi による事象クラスのラベルごとの認識結果

ラベル	説明	出現数	適合率	再現率	F 値
EVe	事象であり、事実性付与が必須	719	0.860	0.847	0.854
EVc	事象可能	113	0.833	0.575	0.681
EVf	機能語につき、事象ではない	110	0.745	0.636	0.686
EVi	仮定など、時間軸上にない概念を指示	127	0.545	0.528	0.536
EVs	比喩	1	0.000	0.000	0.000
EVa	希望、依頼、命令、勧誘	0	0.000	0.000	0.000
EVq	疑問、問いかけ	23	0.739	0.739	0.739
EVp	許可	0	0.000	0.000	0.000

的高いことが分かる。述語項構造解析の前処理として事象を特定するタスクは、事象クラスのうち、EVf・EVcと残り6つを区別し、後者を事象として採用するものである。EVf・EVcの解析精度はF値で約0.68であるため、これらをフィルタリングする上記タスクもこのような性能であると思われる。今回利用したゲーム解説コーパスのアノテーション部分は少量であったため、EVqを除くマイナーなラベルに関する評価は適切にできていない。マイナーなラベルに対する提案手法の効果を測るために、これらのラベルを含むアノテーション部分を増やす必要がある。

事実性解析において、断定を表すFPcはデフォルト値であることから、これ以外のラベルの評価が重要である。表7から、主にモダリティ表現「ない」により特徴付けられるFNcの精度はF値で約0.67であり、モダリティ表現「そうだ」や「ようだ」により特徴付けられるFPrの精度はF値で約0.44であったことが分かる。適合率が低い主な原因は、「～だと、… そうだ」や「～すれば、… そうだ」のような仮定節を持つ文のうち、「…」部分が長いものであると推測される。これらは、局所的にはFPrと同じような表層文字列をしているが、仮定節を持つため、事象クラスEViかつ事実性0の事例である。FNcに関しても同様のことが言える。今後は、このような長距離係り受け関係を適切に捉えることが必要であると思われる。一方、再現率が低い主な原因は、上述したモダリティ表現のロングテール問題であると思われる。

本研究では使用していないがゲーム解説コーパスには将棋局面のデータも含まれており、通常の教師あり学習では高精度で解くことが難しい問題の解法に役立つ可能性を持っている。ゲーム木探索の技術を用いることで、解説文

が参照している将棋局面の将来の局面なども予測することができる。例えば、過去の局面から現在までに選ばれなかった指し手を探索することで、EViの認識精度向上が期待できたり、事実性の確信度推定の高精度化が期待できる。

6. おわりに

本研究ではモダリティ表現と事象のモダリティの間の複雑な関係性を学習するために、マルチタスク学習手法を用いてモダリティ表現認識と事象クラス解析、事実性解析、固有表現認識の4つのタスクを同時学習した。また、学習に用いたアノテーションコーパスが少量のため、大量の生テキストを用いて事前学習を行った。実験したすべてのタスクにおいて、マルチタスク学習を用いた手法は一貫して精度が向上した。

提案手法はゲーム解説コーパス以外のコーパスや他のタスクにも適用可能である。今後の課題として、その他ドメインへの適用や、テキストが参照している世界情報*7を活用することがあげられる。また、可能世界意味論の枠組みを用いて、モダリティ表現を可能世界集合内に接地する手法 [35] において必要となる、モダリティ表現認識と固有表現認識に応用することが期待される。

謝辞

本研究の一部は、JSPS 科研費による基盤研究 (C) 18K11427 「実世界と可能世界が参照可能であるテキストの日本語モダリティ解析」、および、若手研究 19K20341 「音声対話による将棋の感想戦支援システムの構築」の助成を受けたものです。

*7 将棋解説文が参照する将棋局面など。

表 7 BERT+multi による事実性のラベルごとの認識結果

ラベル	説明	出現数	適合率	再現率	F 値
FpC	肯定かつ確定	629	0.840	0.835	0.835
FpR	肯定かつ高い確信度	50	0.563	0.360	0.439
FpS	肯定かつ低い確信度	3	1.000	0.667	0.800
Fnc	否定かつ確定	27	0.667	0.667	0.667
Fnr	否定かつ高い確信度	9	0.250	0.111	0.154
Fns	否定かつ低い確信度	1	0.000	0.000	0.000

表 8 出現数上位 10 個のモダリティ表現 (「表現」列における括弧内文字は、意味を判別するための後続形態素)

表現	ラベル	出現回数	割合	累積割合
た	MEp	618	0.381	0.381
な (い)	MEen	170	0.105	0.486
そう (だ)	MEea	56	0.035	0.520
ば	MEh	47	0.029	0.549
か	MEo	46	0.028	0.577
よう (だ)	MEea	31	0.019	0.596
と	MEh	28	0.017	0.614
だろう	MEea	22	0.014	0.627
た (い)	MEf	20	0.021	0.640
ず	MEen	19	0.012	0.651

参考文献

- [1] Kamioka, Y., Narita, K., Mizuno, J., Kanno, M. and Inui, K.: Semantic Annotation of Japanese Functional Expressions and its Impact on Factuality Analysis, *Proceedings of The 9th Linguistic Annotation Workshop*, pp. 52–61 (2015).
- [2] Izumi, T., Imamura, K., Asami, T., Saito, K., Kikui, G. and Sato, S.: Normalizing Complex Functional Expressions in Japanese Predicates: Linguistically-Directed Rule-Based Paraphrasing and Its Application, Vol. 12, No. 3, pp. 11:1–11:20 (2013).
- [3] Saurí, R. and Pustejovsky, J.: Are You Sure That This Happened? Assessing the Factuality Degree of Events in Text, *Computational Linguistics*, Vol. 38, No. 2, pp. 261–299 (2012).
- [4] Lee, K., Artzi, Y., Choi, Y. and Zettlemoyer, L.: Event Detection and Factuality Assessment with Non-Expert Supervision, *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pp. 1643–1648 (2015).
- [5] de Marneffe, M., Manning, C. D. and Potts, C.: Did It Happen? The Pragmatic Complexity of Veridicality Assessment, *Computational Linguistics*, Vol. 38, No. 2, pp. 301–333 (2012).
- [6] 叶内 晨, 北川善彬, 荒牧英治, 岡崎直観, 小町 守: Web 情報からの罹患検出を対象とした事実性解析・主体解析の誤り分析, 自然言語処理, Vol. 22, No. 5, pp. 363–395 (2015).
- [7] Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K. and Kuksa, P.: Natural language processing (almost) from scratch, *Journal of Machine Learning Research*, Vol. 12, pp. 2493–2537 (2011).
- [8] Søgaard, A. and Goldberg, Y.: Deep multi-task learning with low level tasks supervised at lower layers, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, Association for Computational Linguistics, pp. 231–235 (2016).
- [9] Matsuyoshi, S., Kameko, H., Murawaki, Y. and Mori, S.: Annotating Modality Expressions and Event Factuality for a Japanese Chess Commentary Corpus, *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (chair)*, N. C. C., Choukri, K., Cieri, C., Declerck, T., Goggi, S., Hasida, K., Isahara, H., Maegaard, B., Mariani, J., Mazo, H., Moreno, A., Odijk, J., Piperidis, S. and Tokunaga, T., eds.) (2018).
- [10] Saurí, R.: TimeML Annotation Guidelines Version 1.2. 1, http://www.timeml.org/site/publications/timeMLdocs/annguide_1.2.1.pdf (2006).
- [11] 松吉 俊, 佐藤理史, 宇津呂武仁: 日本語機能表現辞書の編纂, 自然言語処理, Vol. 14, No. 5, pp. 123–146 (2007).
- [12] Suzuki, T., Abe, Y., Toyota, I., Utsuro, T., Matsuyoshi, S. and Tsuchiya, M.: Detecting Japanese Compound Functional Expressions using Canonical/Derivational Relation, *Proceedings of the Eighth International Conference on Language Resources and Evaluation* (2012).
- [13] Lafferty, J., McCallum, A. and Pereira, F.: Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data, *Proceedings of the 18th International Conference on Machine Learning*, pp. 282–289 (2001).
- [14] 江口 萌, 松吉 俊, 佐尾ちとせ, 乾健太郎, 松本裕治: モダリティ、真偽情報、価値情報を統合した拡張モダリティ解析, 言語処理学会第 16 回年次大会, pp. 852–855 (2010).
- [15] 松吉 俊, 江口 萌, 佐尾ちとせ, 村上浩司, 乾健太郎, 松本裕治: テキスト情報分析のための判断情報アノテーション, 電子情報通信学会論文誌. D, 情報・システム = The IEICE transactions on information and systems (Japanese edition), Vol. 93, No. 6, pp. 705–713 (2010).
- [16] Saurí, R. and Pustejovsky, J.: FactBank: a corpus annotated with event factuality, *Language Resources and Evaluation*, Vol. 43, pp. 227–268 (2009).
- [17] Saurí, R. and Pustejovsky, J.: Determining modality and factuality for text entailment, *International Conference on Semantic Computing*, IEEE, pp. 509–516 (2007).
- [18] Rudinger, R., White, A. S. and Van Durme, B.: Neural Models of Factuality, *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 731–744 (2018).
- [19] Qian, Z., Li, P., Zhang, Y., Zhou, G. and Zhu, Q.: Event factuality identification via generative adversarial networks with auxiliary classification, *Proceedings of the 27th International Joint Conference on Artificial Intelligence*, AAAI Press, pp. 4293–4300 (2018).
- [20] Qian, Z., Li, P., Zhu, Q. and Zhou, G.: Document-Level Event Factuality Identification via Adversarial Neural Network, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1*

(*Long and Short Papers*), pp. 2799–2809 (2019).

[21] Maekawa, K., Yamazaki, M., Maruyama, T., Yamaguchi, M., Ogura, H., Kashino, W., Ogiso, T., Koiso, H. and Den, Y.: Design, Compilation, and Preliminary Analyses of Balanced Corpus of Contemporary Written Japanese, *Proceedings of the Seventh conference on International Language Resources and Evaluation*, Valletta, Malta, European Languages Resources Association (ELRA) (2010).

[22] Sutton, Charles: GRMM: GRaphical Models in Mallet., <http://www.globus.org/toolkit/> (2006).

[23] Inui, K., Abe, S., Hara, K., Morita, H., Sao, C., Eguchi, M., Sumida, A., Murakami, K. and Matsuyoshi, S.: Experience Mining: Building a Large-Scale Database of Personal Experiences and Opinions from Web Documents, *2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, Vol. 1, pp. 314–321 (2008).

[24] Ma, X. and Hovy, E.: End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Association for Computational Linguistics, pp. 1064–1074 (2016).

[25] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2019).

[26] Martínez Alonso, H. and Plank, B.: When is multitask learning effective? Semantic sequence prediction under varying data conditions, *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, Valencia, Spain, Association for Computational Linguistics, pp. 44–53 (2017).

[27] Mori, S., Richardson, J., Ushiku, A., Sasada, T., Kameko, H. and Tsuruoka, Y.: A Japanese Chess Commentary Corpus, *Proceedings of the tenth International Conference on Language Resources and Evaluation* (2016).

[28] Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, *Neural Computation*, Vol. 9, No. 8, pp. 1735–1780 (1997).

[29] Devlin, J., Chang, M.-W., Lee, K. and Toutanova, K.: BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding, *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, Minneapolis, Minnesota, Association for Computational Linguistics, pp. 4171–4186 (online), available from (<https://www.aclweb.org/anthology/N19-1423>) (2019).

[30] Sener, O. and Koltun, V.: Multi-task learning as multi-objective optimization, *Advances in Neural Information Processing Systems*, pp. 527–538 (2018).

[31] Neubig, G., Nakata, Y. and Mori, S.: Point-wise Prediction for Robust, Adaptable Japanese Morphological Analysis, *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, Portland, Oregon, USA, Association for Computational Linguistics, pp. 529–533 (online), available from (<http://www.aclweb.org/anthology/P11-2093>) (2011).

[32] Pennington, J., Socher, R. and Manning, C. D.: GloVe: Global Vectors for Word Representation, *Empirical Methods in Natural Language Processing*, pp. 1532–

1543 (2014).

[33] Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., Krikun, M., Cao, Y., Gao, Q., Macherey, K., Klingner, J., Shah, A., Johnson, M., Liu, X., Kaiser, L., Gouws, S., Kato, Y., Kudo, T., Kazawa, H., Stevens, K., Kurian, G., Patil, N., Wang, W., Young, C., Smith, J., Riesa, J., Rudnick, A., Vinyals, O., Corrado, G., Hughes, M. and Dean, J.: Google’s neural machine translation system: Bridging the gap between human and machine translation, *arXiv preprint arXiv:1609.08144* (2016).

[34] Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, *arXiv:1412.6980* (2014).

[35] 亀甲博貴, 松吉 俊, 村脇有吾, 森 信介: モンテカルロシミュレーションによる認知的モダリティ表現のグラウンディング手法の検討, 言語処理学会第 25 回年次大会, pp. 731–734 (2019).

付 録

表 A.1 BiLSTM に用いたハイパーパラメータ

単語ベクトルの次元数	100
BiLSTM の層数	1
LSTM の隠れ層の次元数	100
ドロップアウト率	0.25
学習率の初期値	0.01
ミニバッチサイズ	10
エポック数	100

表 A.2 BERT に用いたハイパーパラメータ
事前学習時

単語ベクトルの次元数	768
Transformer の層数	12
隠れ層の次元数	768
self-attention のヘッド数	12
ドロップアウト率	0.1
学習率の初期値	0.0001
ミニバッチサイズ	16
エポック数	15
ファインチューニング時	
ドロップアウト率	0.1
学習率の初期値	0.00002
ミニバッチサイズ	10
エポック数	20