

# BCCWJ-TimeBank を対象とした時間的順序関係の推定

吉川 克正

日本アイ・ビー・エム (株) 人間文化研究機構 国立国語研究所  
東京基礎研究所  
katsuy@jp.ibm.com

浅原 正幸

コーパス開発センター  
masayu-a@ninjal.ac.jp

飯田 龍

東京工業大学  
大学院情報理工学研究所  
ryu-i@cl.cs.titech.ac.jp

## 1 はじめに

近年盛んになりつつある時間情報解析は、事象・時間表現の抽出、事象・時間表現の属性同定、時間情報正規化、時間的順序関係推定など、いくつかの部分問題に分割される。中でも時間的順序関係推定は、その有用性から特に多くの研究が行われている問題である。時間的順序関係は因果関係と密接に関連しており、それを同定することは深い意味理解を行うために不可欠と考えられている。時間的順序関係推定は、TimeBank [8] の公開によって、機械学習による推定手法 [3, 7, 4] が提案されるようになり、その後、TempEval-1,2,3 [12, 13, 10] などの評価型ワークショップを中心に様々な手法が提案されてきた。これらの先行研究の多くは、コーパスが早くから構築されている英語、スペイン語、中国語などの主要な言語を対象としており、それら以外の言語を対象とした研究は少ない。日本語もこれまでは時間情報がアノテーションされたコーパスが存在しなかったため、日本語を対象とした時間情報解析の研究は少なかった。しかし近年、小西ら [14, 15] によって日本語時間情報タグ付きコーパス (BCCWJ-TimeBank) が構築されたことにより、機械学習による時間情報解析や、定量的な性能評価が可能になった。

図 1 に日本語時間的順序関係推定の例を示す。この例には、時間要素として 3 つの事象 (e1,e2,e3) と、1 つの時間表現 (t1) が含まれる。これらに加えて文書作成日時を表す DCT がある。図中の 7 つの矢印は時間要素間の時間的順序関係 (TLINK) を表す。これら時間的順序を表 1 にまとめる。時間的順序関係推定は分類問題として捉えることができ、そのための関係ラベルとして、BEFORE, OVERLAP, AFTER などが利用される。日本語での時間的順序関係推定は、島崎らが限定的なドメインにおける研究 [16] を行っているが、一般ドメインでの研究報告はなされていない。本稿では新聞記事文書を対象とした時間的順序関係推定を扱い、一般ドメインにおいてベンチマークとして利用可能な結果を報告する。

表 1: 図 1 中の時間的順序関係

タスク	順序関係		
<i>DCT</i>	DCT	AFTER	e1
<i>DCT</i>	DCT	BEFORE	e2
<i>DCT</i>	DCT	BEFORE	e3
<i>E2E</i>	e1	BEFORE	e2
<i>E2E</i>	e2	BEFORE	e3
<i>T2E</i>	t1	BEFORE	e2
<i>MAT</i>	e1	BEFORE	e3

## 2 BCCWJ-TimeBank

BCCWJ-TimeBank [14, 15] は、「現代日本語書き言葉均衡コーパス (BCCWJ) <sup>1</sup>」のコアデータ 54 文書に対し、TimeML フォーマット [9] を用いて時間情報を付与したコーパスである。このコーパスでは、下記の代表的な 3 つのタグによって時間情報を付与している。

まず 1 つ目は事象を表す *EVENT* である。BCCWJ-TimeBank では、動詞と形容動詞に限定してタグが付与されている。主な属性としては *class* があり、*STATE* や *REPORTING* などの 7 クラス <sup>2</sup> に分類される。*EVENT* はインスタンス化することで時間情報を表現し、その実体化タグ (*MAKEINSTANCE*) が 3839 個ある。

2 つ目は日付、時刻などの時間表現を表す *TIMEX3* である。*TIMEX3* の主な属性には *value* と *type* があり、それぞれ時間を正規化した値と表現の種類 (*date,time,duration,set*) を表す。BCCWJ-TimeBank には、54 の文書作成日時 (*DCT*) を含む、1198 個の *TIMEX3* がある [14]。

3 つ目は時間的順序関係を表す *TLINK* である。BCCWJ-TimeBank には表 2 に示す 4 種類の *TLINK* があり、関係を表現する属性、*relType* は、Allen の時間区間論理 [1] に基づく関係ラベルが付与されている <sup>3</sup>。

## 3 分類学習による日本語時間的順序関係推定

この節では提案する日本語時間的順序関係推定法について述べる。本研究では、時間的順序関係推定を複数

<sup>1</sup>[http://www.ninjal.ac.jp/corpus\\_center/bccwj/](http://www.ninjal.ac.jp/corpus_center/bccwj/)

<sup>2</sup>参考文献 [15] p.666 3.2 節参照

<sup>3</sup>BCCWJ-TimeBank では 13 ラベルに加え *subevent* と “*vague*” を含めており、計 17 ラベルを利用している

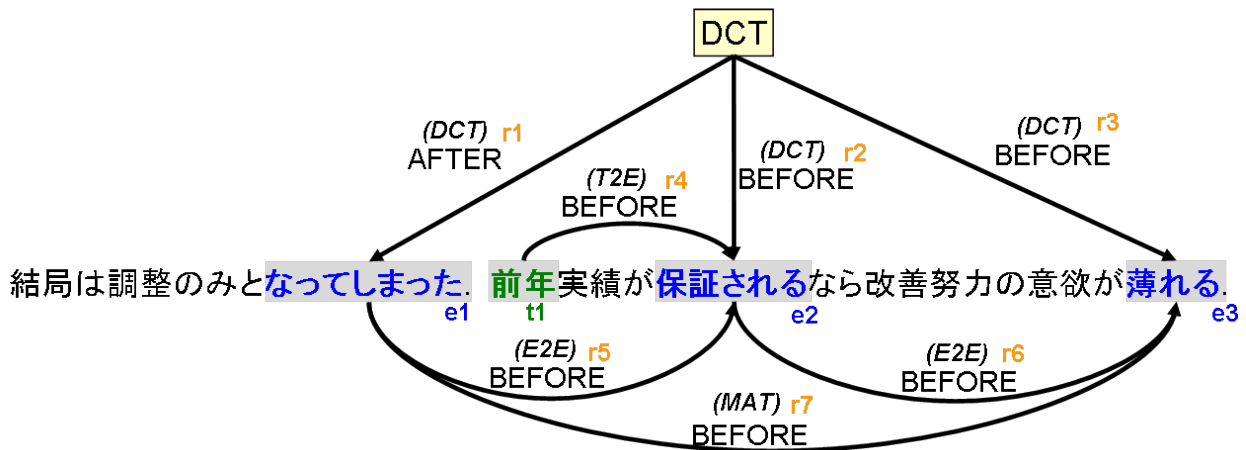


図 1: 日本語時間情報解析

表 2: 4 種類の TLINK

種類	対象となる関係
DCT	文書作成日時と事象間
T2E	同一文内の時間表現と事象間
E2E	隣接する事象間
MAT	隣接 2 文の主述語間

クラス分類問題として捉え、容易に再現可能な推定手法を提案する。分類学習器としては SVM [11] を利用する。SVM は多くの時間的順序関係推定研究でも利用されており [7, 5, 2], 再現が容易なことに加え、性能が高いことが知られている。

本研究では前節で示した 4 種類の TLINK の推定問題を部分タスクとして定義し、4 つのタスクで利用する素性をまとめて表 3 に示す。この表にある “Y” の意味は、対応するタスクでその素性が利用されることを示している。また、“e1-e2” や “e-t” が示すのは、2 つの時間要素の組み合わせ素性になるということである。例えば、Event-form の “e1-e2” であれば、2 つの事象が持つ表層形を組み合わせ素性とする (e.g. 行く-食べる)。表 3 に示した素性は素性名の接頭辞で次の 3 種類に分類される。事象の属性値及び語彙的な素性 (EVENT-), 時間表現の属性値及び語彙的な素性 (TIME-), 事象または時間表現間の統語的な素性 (DEP-) である。

表 3 では、BCCWJ-TimeBank が持つ属性値の素性とそれ以外の語彙的/統語的な素性を区別するため、前者については素性の種類名の直後を大文字にしてある。例えば、“Event-Class” は BCCWJ-TimeBank の属性値素性であり、“Event-form” は語彙的素性である。尚、事象や時間表現が複数単語からなる場合、本研究では全ての単語の語彙的素性を利用している。

事象の素性 (EVENT-) は、Class だけが属性値であり、残りは表層形 (form)、品詞 (pos\_Fine/Coarse)、表層形と品詞の組み合わせ (form-pos)、主辞 (lemma)、活用 (inflect)、活用形 (type) のように主要な語彙的素性を利用する。また対応する文節が、“ている”、“

表 3: 利用する素性

素性名	DCT	T2E	E2E	MAT
EVENT-Class	Y	Y	e1-e2	e1-e2
EVENT-form	Y	Y	e1-e2	e1-e2
EVENT-pos_Fine	Y	Y	e1-e2	e1-e2
EVENT-pos_Coarse	Y	Y	e1-e2	e1-e2
EVENT-form-pos	Y	Y	e1-e2	e1-e2
EVENT-lemma	Y	Y	e1-e2	e1-e2
EVENT-inflect	Y	Y	e1-e2	e1-e2
EVENT-type	Y	Y	e1-e2	e1-e2
EVENT-teiru	Y	Y	e1-e2	e1-e2
EVENT-teita	Y	Y	e1-e2	e1-e2
EVENT-ru	Y	Y	e1-e2	e1-e2
EVENT-ta	Y	Y	e1-e2	e1-e2
EVENT-sent_head	Y	Y	e1-e2	e1-e2
EVENT-sent_end	Y	Y	e1-e2	e1-e2
EVENT-NearestTime	Y			e1-e2
TIME-Type	Y	Y		
TIME-Value	Y	Y		
TIME-VFS		Y		
TIME-FuncInDoc	Y	Y		
TIME-form		Y		
TIME-pos_Fine		Y		
TIME-pos_Coarse		Y		
TIME-dct order		Y		
DEP-dependency		Y	Y	
DEP-path		Y	Y	
DEP-posPath		Y	Y	
DEP-simplePath		Y	Y	
DEP-word	Y	e-t	e1-e2	e1-e2
DEP-pos_Coarse	Y	e-t	e1-e2	e1-e2

ていた”, “る”, “た”, などの特徴的な語尾を持つ場合 (teiru, teita, ru, ta) や、対象の事象が文頭や文末に存在する場合 (sent\_head, sent\_end) には、それぞれ 2 値の素性を与えるものとする。“Nearesttime” は事象に最近傍の時間表現が持つ属性値の素性である。

時間表現の素性 (TIME-) には、TIMEX3 で定義される Type, Value, valueFromSurface(VFS), functionInDocument(FuncInDoc) の属性値を利用する。また DCT を除く時間表現では、表層形 (form)、品詞 (pos\_Fine/Coarse) の語彙的素性と、DCT との順序関

係を3値 (BEFORE, OVERLAP, AFTER) で定義した素性 (dct order)<sup>4</sup>を利用する。

統語的素性 (DEP-) としては、直接係り受け関係がある場合にはその関係ラベル (dependency) を、さらにその拡張として依存構造木上で最短のパスを、係り受けラベル (path), 品詞列 (posPath), 上下関係のみ (simplePath) の3種類で素性とする。また、係り先の単語の表層形 (form) と品詞 (pos) も素性とする。

## 4 実験と結果

この節では実験設定及び実験結果について述べる。

### 4.1 実験設定

実験データには2節で概説した BCCWJ-TimeBank を利用する。実験では英語での推定結果との比較を容易にするため、BCCWJ-TimeBank が持つ17種類の関係ラベルを TempEval-2 [12] と同等の6ラベルに簡略化して扱う。これには機械的な変換を用いており、それに用いた変換表を表4に示す。実験データは表5のように分割し、5分割交差検定を行った。4つの部分タスク (DCT, T2E, E2E, MAT) に対して個別の分類器を構築し、各モデルは対象となる1つの TLINK に対してただ1つの関係ラベルを出力する。分類学習には LIBLINEAR<sup>5</sup> を利用した。素性の抽出には、BCCWJ に人手で付与されている単語・文節区切り、品詞、そして文節ベースの係り受け情報を利用した。また、ベースラインとして最頻出のラベルを全て選択する手法を採用し、以下ではその結果を比較する。

### 4.2 実験結果

まず4つのタスクの実験結果をまとめて表6に示す。各列は左から順に、精度 (P), 再現率 (R), F 値 (F), そしてタスクの最頻ラベルとそれが占める割合である。

まず、タスク DCT において提案モデルは 75.6% の F 値であった。他の言語での推定結果でも、事象と文書作成日時間の関係が最も推定し易いことは知られており、日本語でも同様の結果となった。タスク DCT が他の3タスクに比べて性能が高い理由の一つはラベル分布の偏りにあり、AFTER が 68.3% を占める。つまり、新聞記事で記述されている事象は、既に起こった出来事についてのものが多いのである (文書作成日時の方が後になる)。本研究の提案モデルでは、この最頻ラベルによるベースラインと比較して 7.3% の性能改善が得られた。

タスク T2E では 55.7% の F 値となっている。タスク T2E の最頻ラベルは OVERLAP であり、同一文内の時間表現と事象は同じ時区間を指すことが多いのが分か

<sup>4</sup>value の大小からルールによって算出したもの

<sup>5</sup><http://www.csie.ntu.edu.tw/~cjlin/liblinear/>

表 4: 17 から 6 ラベルへの変換表

元ラベル	変換後
after	AFTER
met-by	AFTER
overlapped-by	OVERLAP-OR-AFTER
finishes	OVERLAP-OR-AFTER
during	OVERLAP
started-by	OVERLAP
equal	OVERLAP
starts	BEFORE-OR-OVERLAP
contains	OVERLAP
finished-by	OVERLAP
overlaps	BEFORE-OR-OVERLAP
meets	BEFORE
before	BEFORE
is_included	OVERLAP
identity	OVERLAP
includes	OVERLAP
vague	VAGUE

表 5: 実験データセット

文書数	11	11	11	11	10
DCT	585	578	657	507	546
T2E	350	270	219	298	332
E2E	410	376	377	343	356
MAT	159	168	171	142	136

表 6: 各タスクの結果

	P	R	F	最頻ラベル (占有率)
DCT	75.6	75.6	75.6	AFTER (68.3)
T2E	55.7	55.7	55.7	OVERLAP (50.4)
E2E	59.9	59.9	59.9	BEFORE (43.2)
MAT	50.0	50.0	50.0	BEFORE (39.3)

る。このベースラインと比較して、提案モデルは 5.3% の性能改善を達成した。これは他のタスクの改善と比較すると小さいものの、統計的には有意である。

タスク E2E では 59.9% の F 値を達成している。このタスクの最頻ラベルは BEFORE になり、隣り合う事象同士は時系列順につながり易いと解釈できる。提案モデルの結果は、ベースラインと比較して 16.9% の性能向上で、4タスク中で最も高い数値となった。

タスク MAT の結果は 50.0% の F 値であり、4タスク中で最も低い数値になっている。ただし、このタスクは他の言語においても最も困難とされており、その理由は対象となる2つの事象が別の文にあることで、統語的な素性の多くが解析に有効に働かなくなるからである。また最頻ラベルはタスク E2E と同様に BEFORE であるが、その占有率は 39.3% と高くない。提案モデルの性能改善は 10.7% となった。

### 4.3 英語時間的順序関係推定との比較

今回定義した4つの部分タスクは、TempEval-1,2 のタスクと定義が近い。TempEval-1 のタスク A, B, C は、それぞれタスク T2E, DCT, MAT に相当し、TempEval-2 のタスク C, D, E, F は、タスク T2E, DCT, MAT, E2E に対応する。ただし、TempEval-2 のタスク F は、同一文

内の係り受け関係にある2事象間の関係推定であり、厳密な意味では日本語のタスク E2E とは一致しないことを補足しておく。

表7<sup>6</sup>に、TempEval-1,2で報告された英語時間的順序関係推定結果の平均を参考値として示す。タスク T2E では、本研究の結果が TempEval の結果よりもやや低い水準にあることが分かる。この原因が言語の特徴に起因するものかどうかは自明ではないため、今後の調査が必要である。

タスク DCT, E2E, MAT では本研究と TempEval の結果が同程度の水準にあることが分かる。しかしながら、TempEval のデータには、事象の持つ属性値として時制 (*tense*) とアスペクト (*aspect*) があり、それが文書作成日時-事象間及び、2事象間の関係を推定する上で大きな役割を果たしている。一方、BCCWJ-TimeBank には時制やアスペクトの情報が付与されていないため、推定結果が相対的に低くなっていると考えられる。また、隣接2文の事象間の関係推定を扱うタスク MAT が、TempEval の結果をやや下回るのに対し、文内の事象間関係を多く扱うタスク E2E では差が無いことから、時制・アスペクト情報の有無が影響していることが分かる。

## 5 おわりに

本研究では、BCCWJ-TimeBank を利用して日本語の時間的順序関係推定器を構築し、今後のベンチマークとなり得る実験結果を示した。またこの実験結果が英語における同様の研究結果 (TempEval) と比較しても遜色の無いことも確認できた。今後の展望として、言語に依存・非依存の特徴を切り分けて推定器を構築することにより、言語横断的な時間情報解析手法を確立したいと考えている。

## 参考文献

- [1] James Allen. Maintaining knowledge about temporal intervals. In *Communications of the ACM*, pp. 832–843, 1983.
- [2] Steven Bethard. Cleartk-timeml: A minimalist approach to tempeval 2013. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 10–14, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
- [3] Branimir Boguraev and Rie Kubota Ando. Timeml-compliant text analysis for temporal reasoning. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence*, pp. 997–1003, 2005.
- [4] Yuchang Cheng, Masayuki Asahara, and Yuji Matsumoto. Constructing a temporal relation tagged corpus of chinese based on dependency structure analysis. In *TIME '07: Proceedings of the 14th International Symposium on Temporal Representation and Reasoning*, pp. 59–69, Washington, DC, USA, 2007. IEEE Computer Society.

<sup>6</sup>表中の\*では、NCSU-joint [6] をはずれ値として省いている

表 7: TempEval-1,2 の結果平均

	T2E	DCT	MAT	E2E
TempEval-1 (En)	59.0	76.0	51.0	-
TempEval-2 (En)	61.0	78.0*	<b>53.0</b>	<b>60.0*</b>
本研究の結果 (Ja)	54.0	76.0	<b>50.0</b>	<b>60.0</b>

- [5] Yuchang Cheng, Masayuki Asahara, and Yuji Matsumoto. Naist.japan: Temporal relation identification using dependency parsed tree. In *Proceedings of the Fourth International Workshop on Semantic Evaluations (SemEval-2007)*, pp. 245–248, Prague, Czech Republic, June 2007. Association for Computational Linguistics.
- [6] Eun Ha, Alok Baikadi, Carlyle Licata, and James Lester. Ncsu: Modeling temporal relations with markov logic and lexical ontology. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 341–344, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
- [7] Inderjeet Mani, Marc Verhagen, Ben Wellner, Chong Min Lee, and James Pustejovsky. Machine learning of temporal relations. In *ACL-44: Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, pp. 753–760, Morristown, NJ, USA, 2006. Association for Computational Linguistics.
- [8] James Pustejovsky, Jose Castano, Robert Ingria, Reser Sauri, Robert Gaizauskas, Andrea Setzer, and Graham Katz. The timebank corpus. In *Proceedings of Corpus Linguistics 2003*, pp. 647–656, 2003.
- [9] James Pustejovsky, Jose Castano, Robert Ingria, Reser Sauri, Robert Gaizauskas, Andrea Setzer, and Graham Katz. Timeml: Robust specification of event and temporal expression in text. *IWCS-5, Fifth International Workshop on Computational Semantics.*, 2003.
- [10] Naushad UzZaman, Hector Llorens, Leon Derczynski, James Allen, Marc Verhagen, and James Pustejovsky. Semeval-2013 task 1: Tempeval-3: Evaluating time expressions, events, and temporal relations. In *Second Joint Conference on Lexical and Computational Semantics (\*SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013)*, pp. 1–9, Atlanta, Georgia, USA, June 2013. Association for Computational Linguistics.
- [11] Vladimir N. Vapnik. *The nature of statistical learning theory*. Springer-Verlag New York, Inc., New York, NY, USA, 1995.
- [12] Marc Verhagen, Robert Gaizaukas, Frank Schilder, Mark Hepple, Graham Katz, and James Pustejovsky. Semeval-2007 task 15: Tempeval temporal relation identification. In *Proceedings of the 4th International Workshop on SemEval-2007.*, pp. 75–80, 2007.
- [13] Marc Verhagen, Roser Sauri, Tommaso Caselli, and James Pustejovsky. Semeval-2010 task 13: Tempeval-2. In *Proceedings of the 5th International Workshop on Semantic Evaluation*, pp. 57–62, Uppsala, Sweden, July 2010. Association for Computational Linguistics.
- [14] 小西光, 浅原正幸, 前川喜久雄. 『現代日本語書き言葉均衡コーパス』に対する時間情報アノテーション. 自然言語処理, Vol. 20, No. 2, pp. 201–222, 2013.
- [15] 保田祥, 小西光, 浅原正幸, 今田水穂, 前川喜久雄. 『現代日本語書き言葉均衡コーパス』に対する時間表現・事象表現間の時間的順序関係アノテーション. 自然言語処理, Vol. 20, No. 5, pp. 657–682, 2013.
- [16] 島崎聡, 飯田龍, 徳永健伸. ドメイン知識を利用したレシピテキストにおけるイベント間の時間関係の解析. 情報処理学会自然言語処理研究会予稿集, 第 NL-211-1 巻, pp. 1–7, 2013.