

時刻アノテーションから導出した時間順序関係による 時間情報解析

Fei Cheng

宮尾 祐介

国立情報学研究所 金融スマートデータ研究センター

{fei-cheng, yusuke}@nii.ac.jp

1 はじめに

自然言語テキストにおける時間情報の解析は、重要かつ挑戦的な課題である。時間情報解析の研究で最もよく用いられているコーパス TimeBank [5] では、イベント表現、時間表現、文書作成時刻 (Document Creation Time; DCT) (これらをまとめてメンションと呼ぶ) の間の相対的時間関係 (Temporal Link; TLINK) [7] がアノテーションされており、これを自動認識する手法がさかんに研究されている。例えば例 (1) では、*reported* で示されるイベントは、*Friday* の中で発生したことが読み取れるため、このペアに対して *is_included* という時間関係ラベルが付与される。

- (1) *The economy **created** jobs at a surprisingly robust pace in January, the government **reported** on Friday, evidence that America's economic stamina has **withstood** any disruption caused so far by the financial tumult in Asia.*

本研究では、文書中のメンションのペアに対して直接アノテーションを行うのではなく、各メンションに対して時刻をアノテーションし、時刻の前後関係を計算することでメンション間の時間順序 (Temporal Order; TORDER) を導出する手法を提案する (3 節)。提案手法で得られる TORDER は TLINK と類似の時間関係を与えるが、アノテーション作業がより容易であるという利点がある。TLINK はメンションペアに対してアノテーションを行うためメンション数の 2 乗の作業量が必要であるが、TORDER で必要な時刻アノテーションはメンション数に比例する作業量で済む。また、TLINK のアノテーションでは各ペアについて時間関係が読み取れる (顕著; salient) かどうかを逐一判断することが求められるが、提案手法ではこの判断は必要ない。一方、TORDER を導出するためには時刻アノテーションが必須であるが、ニュース記事のようにイベントの発生日時が明示されるテキストでは、この作業は比較的容易であると期待される。

4 節では、TORDER と TLINK を定性的および定量

的に比較する。TORDER と TLINK は類似しているが重要な差異があること、TORDER は TLINK では記述できない時間情報を記述できること、TORDER はラベル分布のバランスが取れていることを議論する。5 節では、イベント発生日を認識する日付認識タスク (Time Anchor Prediction)[6] において、TLINK あるいは TORDER アノテーションを用いて学習した時間関係認識器を比較し、提案手法の有効性を示す。

2 関連研究

時間情報解析では、TimeBank [5] が標準データとして広く利用されている。ただし、時間関係アノテーションについては、すべてのメンションペアを対象とするのではなく、作業者がアノテーションすべき「顕著なペア」と判断したものだけがラベル付けされている。続く研究では、アノテーションのカバレッジを改善する試みが行われている。Cassidy ら [2] は、同一文および隣接文中のすべてのメンションのペアについて TLINK のアノテーションを行い、TimeBank と比べて高密度なコーパス TimeBank-Dense を開発した。しかし、作業者が時間関係を読み取れない場合は *vague* ラベルを付与しており、時間関係が読み取れるかどうかの判断が必要であることに変わりはない。

Reimers ら [6] は、イベント表現に対して日付をアノテーションする手法を提案した。つまり、作業者は個々のイベントの発生日をテキストから読み取ることが求められる。単一の日に発生したイベント (単一日イベント) については、イベント発生日が 1 日に定まる (明確な) 場合は、その日付を「YYYY-MM-DD」の形式でアノテーションする。発生日が 1 日に定められない (不明確な) 場合は、できる限り範囲を絞り、タプル (*after*, *before*) としてアノテーションする。また、イベントが複数日にまたがる場合 (複数日イベント) は、その発生期間をタプル (*begin*, *end*) としてア

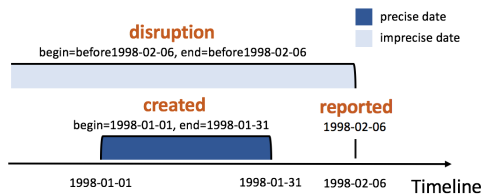


図 1: Reimers ら [6] の日付アノテーション

ノテーションする。ここで、*begin*, *end* は単一日のアノテーション記法を用いる。例 (1) の *created*, *reported*, *disruption* に対する日付アノテーションを図 1 に示す。

日本語については、Asahara ら [1] が現代日本語書き言葉均衡コーパス (BCCWJ) の新聞記事に対して TimeBank に準拠した基準でアノテーションを行った。坂口ら [8] は Reimers らの日付アノテーションの定義を拡張して、京都大学テキストコーパスに対する時間情報アノテーションを行った。本稿では、分析や評価に TimeBank-Dense を用いるが、提案手法はこれらの日本語データに対しても適用可能である。

3 TORDER: 時刻アノテーションから導出された時間順序関係

本研究では、2つの時刻アノテーション (例えば、日付アノテーション) から自動的に時間順序 (Temporal Order; TORDER) を計算し、イベント間、イベント-時間表現間、イベント-DCT 間の相対的時間関係として利用する手法を提案する。図 1 の例で、TORDER の導出方法を説明する。複数日イベントである *created* は、その終了日 (*end*=1998-01-31) が単一日イベント *reported* の日付 (1998-02-06) より以前であるため、順序関係 **BEFORE** が導出される¹。*reported* と *disruption* については、*disruption* の終了日が 1998-02-06 よりも前であることが示されているため、*reported* より以前であること (**BEFORE**) が導かれる。*created* と *disruption* の間は、*disruption* の開始・終了日が明確にはわからず、*created* との時間順序が定まらないため、**VAGUE** が導出される。このように、時刻の前後関係を計算することで、TORDER を導出することができる。

表 1 に単一日メンションのペア (S_1 と S_2) に対する TORDER の定義を示す。明確な日付どうしの前後関係は自明であるが、一方あるいは両方が不明確な日付の場合は、*after* あるいは *before* の値を比較し、一方のメンションがもう一方より確実に以前・以降であ

¹本稿では、TLINK のラベルを斜体・小文字で、TORDER のラベルをボールド体・大文字で表す。

TORDER	条件
明確な単一日 S_1 と S_2	
BEFORE	if $S_1 < S_2$
AFTER	if $S_1 > S_2$
SAME.DAY	if $S_1 = S_2$
明確な単一日 S_1 と不明確な単一日 S_2 (<i>after</i> ₂ , <i>before</i> ₂)	
BEFORE	if $S_1 < after_2$
AFTER	if $S_1 > before_2$
VAGUE	other cases
不明確な単一日 S_1 (<i>after</i> ₁ , <i>before</i> ₁) と S_2 (<i>after</i> ₂ , <i>before</i> ₂)	
BEFORE	if $before_1 < after_2$
AFTER	if $after_1 > before_2$
PVAGUE	if $before_1 = before_2$ and $after_1 = after_2$
VAGUE	other cases

表 1: 単一日メンション間の TORDER の定義

TORDER	条件
単一日 S_1 と複数日 M_2 (<i>begin</i> ₂ , <i>end</i> ₂)	
BEFORE	if S_1 BEFORE <i>begin</i> ₂
AFTER	if S_1 AFTER <i>end</i> ₂
IS_INCLUDED	if S_1 AFTER <i>begin</i> ₂ and S_1 BEFORE <i>end</i> ₂
VAGUE	other case
複数日 M_1 (<i>begin</i> ₁ , <i>end</i> ₁) と M_2 (<i>begin</i> ₂ , <i>end</i> ₂)	
BEFORE	if <i>end</i> ₁ BEFORE <i>begin</i> ₂
AFTER	if <i>begin</i> ₁ AFTER <i>end</i> ₂
IS_INCLUDED	if <i>begin</i> ₁ AFTER <i>begin</i> ₂ and <i>end</i> ₁ BEFORE <i>end</i> ₂
INCLUDES	if <i>begin</i> ₁ BEFORE <i>begin</i> ₂ and <i>end</i> ₁ AFTER <i>end</i> ₂
SAME.SPAN	if <i>begin</i> ₁ SAME.DAY <i>begin</i> ₂ and <i>end</i> ₁ SAME.DAY <i>end</i> ₂
PVAGUE	if <i>begin</i> ₁ PVAGUE <i>begin</i> ₂ and <i>end</i> ₁ PVAGUE <i>end</i> ₂
VAGUE	other cases

表 2: 複数日メンションに対する TORDER の定義

る場合に **BEFORE/AFTER** が与えられる。それ以外の場合は時間の前後関係は決定できないが、同一の期間を参照している (*before* どうし、*after* どうしが同一日である) 場合は、まったく時間関係がわからない場合 (**VAGUE**) と区別するため、特別な関係ラベル **PVAGUE** (Partially Vague) を付与する。表 2 に単一日メンションと複数日メンション間、および複数日メンション間の TORDER の定義を示す。これらは、単一日に対する TORDER を用いて定義される。

TimeBank 等の既存データでは主に日付のアノテーションが行われており、時刻の最小単位を 1 日とすることが一般的である。本稿では、1 日を粒度とした日付アノテーション [6] を用いて分析および実験を行う。ただし、本手法はイベント・時間表現・DCT について同じ粒度でアノテーションが行われていれば適用可能であるため、1 日より大きいあるいは小さい単位の時刻アノテーションに対しても適用することができる。

4 TORDER と TLINK の比較

4.1 定性的比較

本研究の TORDER は、TLINK で定義されている時間関係と類似しており、多くの場合対応関係がある。**BEFORE** と **AFTER** は、あるメンションが時間軸上で

	<i>b</i>	<i>a</i>	<i>s</i>	<i>i</i>	<i>ii</i>	<i>v</i>
BEFORE	1486	24	0	22	26	542
AFTER	19	1242	5	26	66	503
SAME_DAY	155	93	83	164	343	647
SAME_SPAN	4	0	9	5	6	42
INCLUDES	104	61	2	225	25	372
IS_INCLUDED	56	71	1	25	214	333
PVAGUE	91	40	41	23	36	336
VAGUE	331	261	33	145	136	1464

表 3: TORDER と TLINK の対応関係。TLINK は以下の略記を用いる。*b*: before, *a*: after, *s*: simultaneous, *i*: includes, *ii*: is_included, *v*: vague.

別のメンションより前または後に発生することを表しており、TLINK の *before*, *after* にほぼ対応している。INCLUDES と IS_INCLUDED は、複数日メンションに別のメンションの時刻が包含されていることを示しており、TLINK における *includes* と *is_included* より厳密な定義が与えられている。SAME_DAY と SAME_SPAN は、最小単位（本稿では 1 日）の中で 2 つのメンションが同時に発生していることを示している。これは TLINK の *simultaneous* に類似しているものの、以下で議論するように興味深い差異がある。VAGUE と PVAGUE は時間順序が決定できないことを示しており、TimeBank においてアノテーションされていなかった「顕著でない」関係、あるいは TimeBank-Dense における *vague* に相当する。

TORDER は、TLINK では記述されない時間情報を記述することができる。例 (2) では、3 つのイベント *rose*, *pushing*, *leaving* の間の相対的時間関係は判断できず、TLINK では時間関係が記述できない。しかし、これらのイベントは同一日に起きたことが明白であるため、時刻アノテーションは同一日を指しており、したがって TORDER は SAME_DAY となる。

- (2) Stocks *rose*, *pushing* the Dow Jones industrial average up 72.24 points, to 8,189.49, *leaving* the index within 70 points of its record.

逆に、TORDER において時間情報が失われるケースは、不明確な時刻アノテーションに起因する。例 (1) で *withstood* と *disruption* の関係は TLINK では *after* が付与されている。一方、日付アノテーションは両方とも (begin=before 1998-02-06, end=before 1998-02-06) である。両イベントとも発生日が明確にはわからないため、TORDER では PVAGUE が付与される。

テキストのドメインや応用アプリケーションによって、TLINK あるいは TORDER が有利な場合がある。TLINK は、時刻があまり明示されないテキスト（小説や物語など）においても時間関係をアノテーションすることができる。しかし、TLINK のアノテーション作業はコストが大きく、またアノテーションすべき「顕

著な」ペアを同定する作業はアノテータにとって負担が大きい。TORDER は「顕著な」ペアの判断が必要なく、またメンション数に比例する作業量となるため、作業コストが大幅に小さいと期待される。しかし、イベントの時刻が同定できるようなテキスト（ニュース記事など）にしか適用できず、また時刻の最小単位より小さい粒度の時間関係は記述できない。

4.2 定量的比較

TimeBank-Dense のアノテーション対象と同一の 10,007 メンションペアに対して TORDER を導出し、TORDER と TLINK のラベルの頻度や対応関係を分析した。TORDER を導出するための時刻アノテーションとして Reimers ら [6] の日付アノテーション、および TimeBank の時間表現・DCT アノテーションを用いた。

表 3 は、TimeBank-Dense の TLINK と、提案手法で導出した TORDER のラベルごとの頻度を示す。TLINK は *vague* の割合が高く、疎なアノテーションであることがわかる。一方、TORDER はよりバランスの取れた頻度分布を示しており、TLINK より多くのメンションペアに対して時間順序を記述できることが示された。特に、*simultaneous* は非常に少ないが SAME_DAY は他のラベルと同程度観察されることが特筆される。これは、次節で詳述する日付認識タスクにおいて重要な時間情報を与えるものと考えられる。

5 日付認識タスクにおける評価

5.1 実験設定

本節では、イベントの発生日を認識する日付認識タスク (time anchor prediction) において提案手法の有効性を検証する。Reimers らのタスク設定 [6] にならい、単一日イベントを評価対象とする。認識手法は 2 段階からなり、第 1 段階では入力テキストに対し時間関係認識器を適用する。次に、各イベントについて認識された時間関係を集め、それらから導かれる時間情報の中で最も明確な日付を出力する。本稿では、Cheng ら [4] の双方向再帰型ニューラルネットワークモデル (Bi-LSTM) を TLINK あるいは TORDER で学習した時間関係認識器を適用し、日付認識タスクの精度向上を評価する。評価指標は、日付表現が厳密に一致する完全一致精度と、日付表現の一部が一致する部分一致精度（例えば、明確な単一日と、不明確な単一日の *before* が一致したものを正解とする）を用いた。

イベントの種類	利用した時間関係	TORDER		TLINK	
		Exact	Partial	Exact	Partial
明確な日付	DCT	0.586	0.866	0.387	0.570
	時間表現	0.384	0.555	0.216	0.288
	すべて	0.660	0.870	0.444	0.611
不明確な日付	DCT	0.351	0.631	0.234	0.395
	時間表現	0.074	0.217	0.051	0.133
	すべて	0.299	0.642	0.252	0.429
すべて	DCT	0.482	0.762	0.319	0.493
	時間表現	0.259	0.419	0.149	0.255
	すべて	0.501	0.769	0.360	0.530

表 4: TORDER および TLINK を用いた日付認識システムの精度。Exact, Partial は、それぞれ完全一致精度、部分一致精度を示す。

	Exact	Partial
CAEVO	0.442	0.553
Bi-LSTM TLINK	0.437	0.550
Bi-LSTM TORDER	0.586	0.811

表 5: 既存の TLINK 認識器との比較

5.2 実験結果

表 4 に、TORDER および TLINK 認識器を用いた日付認識システムの精度を示す。ここで示す実験結果は 6 分割交差検定で得られた精度の平均である。イベントの種類は、明確な日付のイベント、不明確な日付のイベント、あるいは全イベントを評価対象とした結果を示す。利用した時間関係は、第 2 段階のアルゴリズムがイベント-DCT ペアのみ、イベント-時間表現ペアのみ、あるいは両方を入力とした場合の精度を示す。

TORDER は、TLINK と比較して 3 つすべての設定（イベント-DCT ペアのみ、イベント-時間表現ペアのみ、またはイベント-DCT + イベント-時間表現）において、大幅な精度向上を示している。特に、イベント-DCT 間の TORDER を利用した時に高い精度を達成している。これは、ニュース記事で報告されるイベントは DCT と同じ日に発生した場合が多いため、TORDER の SAME_DAY によってイベントの日付が正しく認識できたためと考えられる。イベント-時間表現を利用した場合にはさらに精度向上が見られ、特に明確な日付のイベントについては完全一致精度が 7.4 ポイント向上している。システム全体の精度（最下段）では、TORDER は TLINK より完全一致精度で 14.1 ポイント、部分一致精度で 23.9 ポイントの大幅な精度向上を示しており、提案手法の有効性が実証された。

5.3 既存の TLINK 認識器との比較

最後に、既存の TLINK 認識器 CAEVO [3] を用いた場合との比較を行う。表 5 に、時間関係認識器と

して CAEVO を用いた場合、Cheng ら [4] の手法を用いた TLINK 認識器 (Bi-LSTM TLINK) と TORDER 認識器 (Bi-LSTM TORDER) を用いた場合の結果を示す。この実験は先行研究のデータ分割設定を用いた²。CAEVO と Bi-LSTM TLINK はほぼ同等の精度を示すが、TORDER 認識器を用いることで他の 2 つのシステムよりも完全一致精度で約 14 ポイント、部分一致精度で約 26 ポイント高い精度を示した。

6 おわりに

本稿では、時刻アノテーションから時間順序 TORDER を計算することで時間関係を自動的に得る手法を提案した。TORDER は既存の時間関係 TLINK と類似しており、機械学習に基づく時間関係認識をそのまま適用できる。しかし、同じメンションペアに対して異なる関係ラベルが付与されることがあり、その原因やコーパス中の頻度について分析を与えた。評価実験では、日付認識タスクにおいて TORDER 認識器を用いることで大幅な精度向上が示された。

謝辞 大阪大学 GI センターの李明先生には、論文の日本語作成にご協力いただき、深く感謝いたします。

参考文献

- [1] M. Asahara, S. Kato, H. Konishi, M. Imada, and K. Maekawa. BCCWJ-Timebank: Temporal and event information annotation on Japanese text. In *IJCLCLP 2014*.
- [2] T. Cassidy, B. McDowell, N. Chambers, and S. Bethard. An annotation framework for dense event ordering. In *ACL 2014*.
- [3] N. Chambers, T. Cassidy, B. McDowell, and S. Bethard. Dense event ordering with a multi-pass architecture. *TACL 2014*.
- [4] F. Cheng and Y. Miyao. Classifying temporal relations by bidirectional LSTM over dependency paths. In *ACL 2017*.
- [5] J. Pustejovsky, P. Hanks, R. Sauri, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro, et al. The timebank corpus. In *Corpus Linguistics 2003*.
- [6] N. Reimers, N. Dehghani, and I. Gurevych. Temporal anchoring of events for the timebank corpus. In *ACL 2016*.
- [7] A. Setzer. *Temporal information in newswire articles: an annotation scheme and corpus study*. PhD thesis, University of Sheffield, 2002.
- [8] 坂口智洋, 河原大輔, 黒橋禎夫. 京都大学テキストコーパスに対する網羅的な時間情報アノテーション. 情報処理学会研究報告 2017.

²<https://github.com/nchambers/caevo/blob/master/src/main/java/caevo/Evaluate.java>