1F3-1in

構造学習を用いたテキストからの地域イベント情報抽出

Local-area Event Extraction from Texts Based on Structured Output Learning

数原 良彦[*1] 鈴木 潤[*2] 鷲崎 誠司[*1]
Yoshihiko Suhara Jun Suzuki Seiji Susaki

[*1]日本電信電話株式会社 NTT サービスエボリューション研究所
NTT Service Evolution Laboratories, NTT Corporation

[*2]日本電信電話株式会社 NTT コミュニケーション科学基礎研究所
NTT Communication Science Laboratories, NTT Corporation

We aim to extract local-area event information from unstructured documents such as blogs. In this paper, we assume that local-area event consists of three items: event name, location, and event date. We propose a two-stage algorithm to tackle this problem; it first extracts named entities as candidates of local-area event from a given text by using a preliminarily trained named entity recognition tool. Then, it estimate the correct combination from candidates by using the structured output learning framework that we propose in this paper.

1. はじめに

スマートフォンの普及に伴い，ユーザの位置情報やユーザの訪問先をクエリとして利用する地域情報検索がより一般的になっている．たとえば現在位置近くで食事をしたいユーザに対して店舗を提示するグルメ検索や，ユーザが参加可能なイベント情報を提示する地域イベント情報サービスなどがある．地域イベント情報サービスとしては，GO 近所[*1]や，地域情報サイト ZAQ[*2]がある．これらのサービスを利用することでユーザは，週末の予定を立てたり，外出先でイベントに参加するなど，実世界行動の付加価値向上に活かすことができる．

このようなサービスでは，人手によって地域イベント情報を格納しているため，高品質なイベント情報を提供可能ではあるが，任意の地域において地域イベントを網羅することが困難である．たとえば地域に多数存在するサークル活動によるイベント告知など，地域掲示板に掲載されるような地域イベント情報はウェブ上にも数多く存在する．これらの地域イベント情報を抽出することができれば，人手によって構築された地域イベント情報と組み合わせることで，地域イベント情報サービスに更なる付加価値を与えることができる．

そこで我々は，イベントサイトでは扱われない地域イベント情報に着目し，これらの地域イベント情報の自動抽出を試みる．たとえば地方自治体などで定期的に更新される地域イベント情報の場合，フォーマットが定められているため，あらかじめ抽出ルールを作成することによって，半自動抽出が可能である．我々は，これらのページにも記載されない，ウェブ上で 1 文書のみ記述されている地域イベント告知から正確に地域イベント情報を抽出することを目指す．本稿ではウェブ情報源のうちブログ記事に着目し，ブログ記事からの地域イベント情報抽出に取り組む．

本稿では，地域イベント情報を (A) イベント名， (B) 開催場所， (C) 開催日時の 3 つ組情報と定義する．ここでイベント名，開催場所，開催日時を本稿では**カテゴリ**と呼ぶ．開催場所を緯度経度にマップすることができれば，たとえば地図インタフェースを用いてイベント情報をピン立てすることが可能となり，地域イベント情報提供サービスにそのまま利用が可能となる．そこで，開催場所については緯度経度情報を持つ住所表現または地名表現を候補とする．開催日時については日時表現を絶対時刻に変換することができれば，たとえばカレンダーインタフェースのようなものを用いて，目的の時期に開催される地域イベント情報の閲覧が可能となる．そのため，開催日時も同様に絶対時刻を持つものとする．

テキストからのイベント情報を自動抽出するアプローチとして，情報抽出のためのパターンのシードを作成し，パターンを拡張しながら複数文書を解析することで必要な情報を取り出すブートストラップアプローチがある [Xu 06]．しかしながら，地域イベント情報が記載されているのは一文書である，

ブログ記事に記載された地域イベント情報の例を図 1 に示す．この例ではイベント情報が整形されておらず，記述された部分からイベント情報を抽出する必要がある．その他の例としては，たとえば「名称:*」「場所:*」「日時:*」のようにある程度整形されて記述されているイベント情報も存在する．このような類の表現であれば，ルールベースアプローチによって一定の精度でイベント情報の抽出が可能であるが，ブログ記事群には構造化された表現，自然文で記述された表現が混在しているため，ひとつの枠組みでふたつの状況に対応できることが好ましい．

ブログ記事からの地域イベント情報に着目すると，以下の 2 つの要件を満たす方法が望ましい．(1) 地域イベント告知は主催者のブログのみに記述されることが多いため，ひとつの文書から正確に抽出可能なこと．また，(2) イベント情報の文書中での構造化の程度に依存しない抽出が可能なこと．

これらの要件を満たすため，本研究では網羅性を重視する言語解析的アプローチと，正確性を重視する情報抽出的アプローチを組み合わせたイベント情報抽出フレームワークを提案する．具体的には，固有表現抽出を用いて網羅的に候補を抽出する処理と，抽出された候補の中から，適切な組み合わせを選択する処理の二段階の処理により，テキストから地域イベント情報の抽出を実現する．

本研究の貢献は以下のとおりである．

- 抽出対象の多くが非構造化テキスト一文書のみに含まれる状況において高精度なイベント情報抽出を実現するた

連絡先: suhara.yoshihiko@lab.ntt.co.jp

*1 http://go-kinjo.jp/

*2 http://zaq.ne.jp/event/

タイトル: 民家の甲子園
本文: 本日 12 月 30 日より平成 25 年 1 月 27 日まで香南市夜須町にある道の駅やす、ヤ・シィーパークにおいて「民家の甲子園」のパネル展を開催します

図 1: ブログ記事に記載された地域イベント情報の例

表 1: イベント情報の独自性分析

当該ページのみ記述あり	#	他ページに記述あり	#
お店イベント告知	12	イベント告知転載	6
個人主催イベント告知	8	イベント紹介	3
サークル活動の告知	8	試合結果掲載	2
セミナーの参加募集	8		
ライブ告知	3		

め，固有表現抽出によってカテゴリ候補を網羅的に抽出し，抽出された候補の中から適切な組み合わせを選択する枠組みを提案する．

- カテゴリ候補の中から組み合わせを選択する問題を構造予測問題として定式化し，構造学習の枠組みで予測モデルを構築する方法を提案する．また，予測誤りに対してコストを設定し，各カテゴリの誤りに応じて損失を増やすことでモデルの予測精度向上させる方法も合わせて提案する．

本稿の構成は以下の通りである．2 章で地域イベント情報が記述される文書数について予備実験を行い，3 章で提案手法の詳細とアルゴリズムを述べる．4 章で評価実験について述べたのちに 5 章で関連研究を述べ，6 章でまとめる．

2. 予備実験

ブログ記事に記載された地域イベント情報がウェブ上に一意なものであるかを検証をするため，予備実験を行った．評価実験で用いたデータセットに記述されたイベント情報の中から，ランダムに選択した地域イベント情報 50 件について，ウェブ検索を利用して他のウェブメディアに記載されているか調査を行った．その結果，50 件中 39 件 (78%) の地域イベント情報が当該ブログ記事のみに記述されていたことを確認した．この結果より，ブログ記事に含まれる地域イベント情報の約 8 割については他メディアでは代替が困難であることがいえる．

また，予備実験に用いた 50 件についてどのような種類の地域イベント情報であるか分析を行った．分析結果の内訳を表 1 に示す．

3. 構造学習を用いた地域イベント情報抽出

提案手法においては，3 つ組の情報を抽出するために，(1) **カテゴリ候補の抽出**，(2) **カテゴリ組の選択**の 2 段階の処理を行う．以下，それぞれの処理について述べる．

3.1 カテゴリ候補の抽出

カテゴリ候補抽出処理においては，イベント名，開催場所，開催日時それぞれについて，カテゴリ候補の抽出を行う．以下，イベント名，開催場所，開催日時の順に候補の抽出方法を述べる．

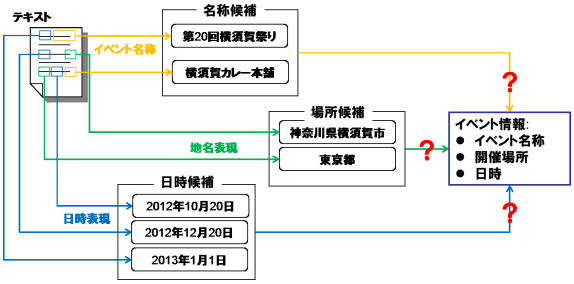

図 2: 固有表現抽出を用いた候補の抽出

イベント名候補の抽出には，固有表現抽出器の結果を用いる．本稿では固有表現抽出として，文献 [松尾 08] の方法を用いる．この固有表現抽出器を用いることで，従来の固有表現タイプに加えて，緯度経度に対応した地名表現と，絶対時刻に対応した日時表現も抽出できる．開催場所候補の抽出には，[松尾 08] の方法で緯度経度情報が付与された地名表現，および住所表現を対象とする．開催日時の抽出には同じく [松尾 08] の方法で日時表現と判定されたものを対象とする．具体的には廣嶋らの方法 [廣嶋 09] が利用されている．これにより，たとえば「明日」といった相対的な日時表現についても絶対日時に変換することが可能である．

以上の処理を利用することで，非構造化テキストからイベント情報の組を選択することができる．テキストから抽出した各カテゴリ候補の例を図 2 に示す．

3.2 構造学習に基づくカテゴリ組の選択

教師あり機械学習において，予測ラベル $\mathbf{y}$ が構造を持つ問題を扱う枠組みを構造学習と呼ぶ．構造学習においては，事例 $\mathbf{x}$ とラベル構造 $\mathbf{y}$ に対して素性関数ベクトル $\Phi(\mathbf{x},\mathbf{y})$ を定義する．そして，全ての可能なラベル構造 $\mathcal{Y}$ の中からモデルパラメータによって最大の値を返すラベル構造 $\hat{\mathbf{y}} = \mathrm{argmax}_{\mathbf{y}\in\mathcal{Y}} \mathbf{w}^T\Phi(\mathbf{x},\mathbf{y})$ を予測ラベル構造として出力する．モデルパラメータの学習には，構造化パーセプトロンや条件付き確率場などを用いる．ここで素性関数ベクトルは M 個の素性関数から構成される M 次元ベクトル $\Phi(\mathbf{x},\mathbf{y}) = (\phi_1(\mathbf{x},\mathbf{y}),\phi_2(\mathbf{x},\mathbf{y}),\ldots,\phi_M(\mathbf{x},\mathbf{y}))^T$ とする．素性関数の具体的な設計については後述する．

本稿においては，ラベル構造をイベント名，開催場所，開催日時から成る $\mathbf{y} = (y_{name}, y_{loc}, y_{time})^T$ と解釈し，さきほどの処理で抽出した候補の組み合わせ集合をラベル構造集合とし，候補の中から正しいイベント組を選択する構造予測問題として定式化する．すなわち，$\mathcal{Y} = \mathcal{Y}_{name} \times \mathcal{Y}_{loc} \times \mathcal{Y}_{time}$ である．ここで，それぞれのラベル集合は，イベント名 $y_{name} \in \mathcal{Y}_{name}$，開催場所 $y_{loc} \in \mathcal{Y}_{loc}$，開催日時 $y_{time} \in \mathcal{Y}_{time}$ に対応する．たとえば図 2 の例によると，全ての可能な組み合わせ集合 $\mathcal{Y}$ は $2 \cdot 2 \cdot 3 = 12$ 通りである．

構造学習においてはラベル構造にも依存した素性関数を設計することが可能であるため，たとえばイベント名，開催場所，開催日時の記述が文書中の近くに出現するといった特徴を素性関数として用いることが可能である．本稿で用いた素性関数を以下に示す．以下でカテゴリ候補と記述している部分は，それぞれ { イベント名，開催場所，開催日時 } のカテゴリ候補を表す．

- カテゴリ候補を構成する形態素のユニグラム
- 3 つのカテゴリ候補が出現する文の差
- カテゴリ候補の直前 5 形態素のユニグラム
- カテゴリ候補の直後 5 形態素のユニグラム

- カテゴリ候補がブログの記事タイトルに含まれるか

なお，形態素については全ての品詞を用いる．カテゴリ候補を構成する形態素のユニグラムを利用することにより，たとえばイベント名においては「第 x 回」「イベント」「セミナー」といったイベント名を構成する形態素表現が選択に役立つと期待する．また，カテゴリ候補の直前 5 形態素のユニグラムを利用することで「場所:」「日時:」といった部分的に構造化されたカテゴリ情報の特徴を捉えることを期待する．

3.2.1 誤りコストの考慮

提案手法では，3 つ組のカテゴリ情報の構造予測問題としているため，個々のカテゴリ (イベント名，開催場所，開催日時) に対する予測が正解であるか判別することが可能である．そのため，3 つ組としては誤りの組み合わせにおいても，1 カテゴリ誤り，2 カテゴリ誤り，3 カテゴリつ全て誤りが存在する．しかしながら，構造学習を愚直に適用するだけでは，これらの誤りは等しく誤りとして扱われる．今回の問題設定においては上述のとおり，カテゴリ毎に正解を規定することが可能であるため，各カテゴリにおける誤りに対して個別に損失を与えることにより，より選択精度の高いモデル構築を実現する．

イベント名誤り，開催場所誤り，日時誤りに対する誤りコストをそれぞれ $c_{name}, c_{loc}, c_{time}$ とする．これらの誤りコストはハイパーパラメータとして与えられるものとする．

正解ラベル構造 $\mathbf{y}$ と予測ラベル構造 $\hat{\mathbf{y}}$ とした場合，正解ラベル構造のカテゴリ候補と予測ラベル構造のカテゴリ候補が一致している場合に 0 を返すような関数を，

$$\epsilon(y_{category}, \hat{y}_{category}) = \begin{cases} 1 & (y_{category} \neq \hat{y}_{category}) \\ 0 & (\text{otherwise}) \end{cases}$$

として定義する．簡単のため $\epsilon(y_{category}, \hat{y}_{category})$ を $\epsilon_{category}$ と表記する．これを用いて誤りに対するコスト関数を以下の通り定義する:

$$\rho(\mathbf{y}, \hat{\mathbf{y}}) = \epsilon_{name} \cdot c_{name} + \epsilon_{loc} \cdot c_{loc} + \epsilon_{time} \cdot c_{time}. \quad (1)$$

たとえば $c_{name} = c_{loc} = c_{time}$ と設定することにより，誤りカテゴリ数に応じたペナルティを設定することが可能である．

3.2.2 アルゴリズム

本稿では構造学習の学習アルゴリズムとして，オンラインの識別学習手法である Passive-Aggressive (PA) [Crammer 06] を用いる．PA は t 回目の更新における重みベクトル $\mathbf{w}_t$ の更新を $\mathbf{w}_{t+1} = \text{argmin}_{\mathbf{w} \in \mathbb{R}^n} \frac{1}{2}||\mathbf{w} - \mathbf{w}_t||_2^2$ s.t. $\ell_t = 0$ という最適化問題として定式化する．ここで n は重みベクトル $\mathbf{w}$ の次元数を表す．ここで ℓ_t は現在の重みベクトルにおける事例の損失を表し，$\ell_t = 0$ ならば何もせず，$\ell_t > 0$ ならば，更新の大きさが最小になるように，$\mathbf{w}$ を更新する．この最適化問題は，ラグランジュの未定乗数法を用いて解くことにより，閉じた解で $\mathbf{w}_{t+1}$ を求めることができる．このように 1 回の更新においては，更新量を閉じた解で求めることができるのが PA の利点のひとつである．また，誤りを許容するためにスラック変数 ξ を導入した場合も，同様に閉じた解で重みベクトルの更新が可能である．正則化項を $C\xi$ と設定した手法は PA-I と呼び [Crammer 06]，よりロバストな学習が可能であることが知られている．本稿ではこれより PA-I を用いる．

提案手法のアルゴリズムを Algorithm 1 に示す．PA-I は訓練データ D とイテレーション回数 T，トレードオフパラメータ C を入力とする．イテレーション回数 T だけ以下の処理を繰り返す．ランダムに事例を選択し (ステップ 3)，選択された事例の中から損失が最大の予測ラベル構造 $\hat{\mathbf{y}}_t$ を選択する (ステップ 4)．ここで $\mathbf{y}_t$ は t 試行目に選択された事例の正解ラベル構造である．また，コスト $\rho(\mathbf{y}_t, \mathbf{y})$ は式 (1) を用いて算出する．選択された予測ラベル構造に基づき t 試行目における損失 ℓ_t を計算する (ステップ 5)．ステップ 5 で算出された損失 ℓ_t と，正解ラベル構造における素性ベクトルと予測ラベル構造における素性ベクトルの差のノルムに基づいて更新量 τ_t を算出し (ステップ 6)，重みの更新を行う (ステップ 7)．処理の最後に全ての試行における重みベクトルの平均を求める [Collins 02] (ステップ 9)．

Algorithm 1 Structured Output Learning with PA-I

Input: $\{(\mathbf{x}_n, \mathbf{y}_n)\}_{n=1}^N \in D$, T, C
Output: $\mathbf{w}^*$
1: $\mathbf{w}_0 \leftarrow \mathbf{0}$
2: **for** $t = 1$ to T **do**
3: Obtain random sample $(\mathbf{x}_t, \mathbf{y}_t)$ from D
4: $\hat{\mathbf{y}}_t = \underset{\mathbf{y} \in \mathcal{Y}}{\text{argmax}} \left\{ \mathbf{w}_t^T \Phi(\mathbf{x}_t, \mathbf{y}) - \mathbf{w}_t^T \Phi(\mathbf{x}_t, \mathbf{y}_t) + \rho(\mathbf{y}_t, \mathbf{y}) \right\}$
5: $\ell_t = \mathbf{w}_t^T \Phi(\mathbf{x}_t, \hat{\mathbf{y}}_t) - \mathbf{w}_t^T \Phi(\mathbf{x}_t, \mathbf{y}_t) + \rho(\mathbf{y}_t, \mathbf{y})$
6: $\tau_t = \max \left\{ \frac{\ell_t}{\|\Phi(\mathbf{x}_t, \mathbf{y}_t) - \Phi(\mathbf{x}_t, \hat{\mathbf{y}}_t)\|^2}, C \right\}$
7: $\mathbf{w}_{t+1} \leftarrow \mathbf{w}_t + \tau_t \left(\Phi(\mathbf{x}_t, \mathbf{y}_t) - \Phi(\mathbf{x}_t, \hat{\mathbf{y}}_t) \right)$
8: **end for**
9: $\mathbf{w}^* = \frac{1}{T} \sum_{t=1}^T \mathbf{w}_t$
10: **return** $\mathbf{w}^*$

組み合わせを選択するためには，現在のパラメータにおいて argmax 計算を行う必要があり，組み合わせ数に応じて計算コストが高くなる．たとえば，各ラベルを部分的に決定し，全ラベル構造の探索を枝刈りすることによって argmax を近似的に計算することにより高速化を行うことが可能ではあるが，本稿においては候補の数がそれほど大きくないため，全探索によって argmax 計算を行う．

4. 評価実験

提案手法の有効性を検証するために評価実験を行った．

4.1 データセット

クロールによって取得した日本語ブログ記事のうち，地域イベント情報を含むブログ記事 500 件に対してアノテータによってイベント名，開催場所，開催日時のタグづけを行った．アノテーションを行ったブログ記事の中からイベント名，開催場所，開催日時を含み，1 文書あたり 1 つのイベント情報が含まれる 309 件のデータを評価に用いた．

4.2 実験条件

カテゴリ情報の選択精度の検証を行うため，カテゴリ候補から，固有表現タイプによって組み合わせを判別する手法をベースラインとして用いた (Baseline)．提案手法は，コスト関数を利用しない PA-I (PAStruct)，$c_{name} = c_{loc} = c_{time} = 1$ と設定したコスト関数を用いる PA-I (PAStruct+cost) の二種類を用いた．提案手法については訓練データが必要なため，データセットを 5 分割し，4 つを訓練データ，1 つをテストデータとして用いる 5-Fold cross-validation によって評価を行った．PA-I に対するパラメータは予備実験によって決定し，イテレーション回数は 5,000，C パラメータは 0.1 とした．

評価指標として，正しく 3 つ組が抽出されている正解率，および各カテゴリの正解率を求めた．本実験に用いたデータセットは，全ての事例について正しい組み合わせが存在するため，本実験における正解率は適合率と同時に再現率も表す．

なお，正解判定方法としてイベント名については正解アノテーションの 2/3 以上文字列が一致することとした．開催場

表 2: 実験結果

Method	Overall	Name	Loc.	Time
Baseline	0.133	0.378	0.240	0.426
PAStruct	0.389	0.735	0.466	0.641
PAStruct+cost	**0.417**	**0.781**	**0.474**	**0.719**

所については厳密に正解アノテーションと一致するかに基づいて判定を行った．すなわち，横須賀市久里浜という正解に対して，横須賀市を選択する予測は誤りとする．開催日時については，日付単位の一致を正解と判定した．

4.3 結果と考察

評価結果を表 2 に示す．Overall が 3 つ組の正解率，Name, Loc., Time がそれぞれイベント名，開催場所，開催日時の正解率を示している．Baseline と PAStruct を比較すると，3 つ組抽出，全てのカテゴリについて高い正解率で正しい組み合わせを選択していることがわかる．また，PAStruct と PAStruct+cost を比較すると，コスト関数を利用することにより，更に高い正解率を達成することがわかる．したがって，提案手法を用いることでベースラインに比べて高精度にカテゴリ情報の組を選択可能である．また，コスト関数の利用により，3 つ組正解率と全てのカテゴリ選択正解率について高い値を示しているため，コスト関数の利用によって高精度なモデル構築が可能であるといえる．

しかしながら，PAStruct+cost において 3 つ組正解率は 0.417 と十分に高いとは言えない．カテゴリ別に結果を見ると，開催場所が他のカテゴリに比べて著しく低い値を示している．これは，本実験における開催場所について部分的な一致を誤りとみなす判定が原因だと考えられる．

5. 関連研究

情報抽出分野においてイベント情報抽出は古くから研究されており，MUC[*3] や ACE[*4]においてさまざまな手法が検討されてきた．これらの情報抽出タスクにおいては，たとえばオリンピック開催やノーベル賞受賞など，社会的な意味でのイベント情報を対象としている．これらのタスク向けに開発された方法では，複数の文書を解析し，テンプレートに埋まる情報を正確に抽出するという課題の解決を目指している．たとえば Xu ら [Xu 06] の方法では，あらかじめ与えられたルールのシードを元にブートストラップに基づいてルールを拡張することで，テンプレートを埋めるための情報抽出を実現している．

一方でこのアプローチでは，一文書から正確に情報を抽出するということが困難な場合が多い．そのような場合には，固有表現抽出のようにテキストの中から正確にイベント情報を構成するカテゴリ情報を抽出する．しかしながら，このような言語解析的アプローチには，大量のアノテーションが必要になるため，テキストからのイベント情報抽出をチャンキング問題として定式化することは難しいと考えられる．

はじめに述べたように我々の提案手法は，網羅性を目指す言語解析的アプローチと，適合性を目指す情報抽出的アプローチのハイブリッド手法と捉えることができる．また，地域イベント情報抽出のために既に開発された手法の結果を利用することができるため，ドメインが変化した場合においても適切な候補抽出手法を選択することにより，提案フレームワークが利用可能である．

*3 http://www.itl.nist.gov/iaui/894.02/related_projects/muc/

*4 http://www.itl.nist.gov/iad/mig/tests/ace/

6. おわりに

本稿では，ブログ記事のような構造化されていないテキスト情報から地域イベント情報を抽出するために，イベント名，開催場所，開催日時に対する候補を複数抽出し，抽出された候補の中から適切な組み合わせを選択するという二段階の処理によってイベント情報を抽出する枠組みを提案した．特に組み合わせ選択において，ラベルが構造を持つ構造学習の枠組みで予測モデルを構築することで，ラベル構造に依存する特徴を利用することを可能とし，高精度な組み合わせ選択を実現する．評価実験を通じて，抽出された候補の中から固有表現タイプで組み合わせを決定するベースライン手法に比べて，構造学習を用いた提案手法によって，候補の中から高精度にイベント情報組を選択することが可能であることを検証した．

我々は，カテゴリ候補抽出方法の検討，複数イベント情報が記載された文書からのイベント情報抽出を主たる課題として考えている．提案手法においては，カテゴリ候補抽出によって適切な候補が選択されているかということが最終的なイベント抽出精度に直接影響を与える．評価実験においてはこの部分の性能評価を行っていない．イベント情報抽出の再現率を考慮すると，カテゴリ候補抽出のフェーズでより多くの候補を抽出するように設定する方が望ましい．一方で適合率を考慮すると，カテゴリ候補抽出のフェーズである程度候補を絞る，またはこの段階でフィルタリングを行うという方法も考えられる．本稿における提案手法は，1 文書 1 イベントという仮定を置いているが，実際の文書には複数イベント情報が記載されていることがある．我々の提案手法の枠組みを適用するためには，たとえば文書をイベント情報が含まれる単位に分割して提案手法を適用する方法が考えられる．しかし，たとえば複数のイベント情報が同じ開催場所や同じ開催日時を共有している場合には，単純な領域分割では提案手法を適用することができない．このように複数イベント情報が含まれる文書からのイベント情報抽出が今後の課題である．

参考文献

[Collins 02] Collins, M.: Discriminative training methods for hidden Markov models: theory and experiments with perceptron algorithms, in *Proc. EMNLP '02*, pp. 1–8 (2002)

[Crammer 06] Crammer, K., Dekel, O., Keshet, J., Shalev-Shwartz, S., and Singer, Y.: Online Passive-Aggressive Algorithm, *Mach. Learn.*, Vol. 7, pp. 551–585 (2006)

[Xu 06] Xu, F., Uszkoreit, H., and Li, H.: Automatic event and relation detection with seeds of varying complexity, in *Proc. the AAAI workshop event extraction and synthesis*, pp. 12–17 (2006)

[松尾 08] 松尾義博, 小林のぞみ, 平野徹, 高橋いづみ：Web2.0 時代の名寄せを実現する固有表現グラウンディング技術, NTT 技術ジャーナル, Vol. 20, No. 6, pp. 16–19 (2008)

[廣嶋 09] 廣嶋伸章, 戸田浩之, 松浦由美子, 片岡良治：ブログ記事からの特定の時間帯を表す時間表現の抽出, 第 16 回言語処理学会年次大会 (2009)