情報処理学会論文誌 Vol. 50 No. 6 1622–1635 (June 2009)

日本語単語分割の分野適応のための部分的アノテーションを用いた条件付き確率場の学習

坪井祐太[†1] 森信介[†2] 鹿島久嗣[†1]
小田裕樹 松本裕治[†3]

本研究では文の一部にのみ単語分割情報を付与する部分的アノテーションに注目する．重要な部分や作業負荷の少ない部分にのみアノテーションをすることにより，新しい分野に対応するための学習データを効率的に作成できる．この部分的アノテーションを使用して条件付き確率場（CRF）を学習する方法を提案する．CRF は単語分割および自然言語処理の様々な問題でその有効性が示されている手法であるが，その学習には文全体へのアノテーションが必要であった．提案法は周辺尤度を目的関数にすることで部分的アノテーションを用いた CRF のパラメータ推定を可能にした．日本語単語分割器の分野適応実験において部分的アノテーションによって効果的に性能を向上させることが可能であったことを報告する．

Training Conditional Random Fields Using Partial Annotations for Domain Adaptation of Japanese Word Segmentation

Yuta Tsuboi,[†1] Shinsuke Mori,[†2] Hisashi Kashima,[†1]
Hiroki Oda and Yuji Matsumoto[†3]

In this paper, we address word-boundary annotations which are done only on part of sentences. By limiting our focus on crucial part of sentences, we can effectively create a training data for each new target domain by conducting such partial annotations. We propose a training algorithm for Conditional Random Fields (CRFs) using partial annotations. It is known that CRFs are well-suited to word segmentation tasks and many other sequence labeling problems in NLP. However, conventional CRF learning algorithms require fully annotated sentences. The objective function of the proposed method is a marginal likelihood function, so that the CRF model incorporates such partial annotations. Through experiments, we show our method effectively utilizes partial annotations on a domain adaptation task of Japanese word segmentation.

1. はじめに

日本語や中国語のように分かち書きされていない言語では単語分割は様々な言語解析の前提となる重要な問題である[⋆1]．単語境界は文脈を考慮して決める必要があり分割ルールの管理は複雑になるため，単語分割には統計的手法が活用されるようになっている．一方で，実際に統計的単語分割器を使用する際の課題の 1 つとして，学習データとは異分野の文書ではその性能が劣化してしまうという問題がある．この問題に対応するためには適応先分野に合わせて統計モデルを修正する必要があり，この作業は分野適応と呼ばれる．単純だが効果的に統計的単語分割器を分野適応する方法は，適応先分野の文に単語の境界をアノテーションした学習データを作成することである．しかし，現実の応用では新しい分野への適応時に人手によるアノテーション作業にかけられるコストは限定的であり，作業量は少ないことが望ましい．

そこで，性能向上効果の高い箇所や低コストで付与できる箇所に集中してアノテーションする方法を考える．2 章では具体例として単語リストおよび記号ルールによる部分的アノテーション法を紹介する．本論文での部分的アノテーションとは，文全体ではなく文の特定の単語や文字境界にのみ与えられている単語分割情報を指す．3 章では，部分的アノテーションの形式的な表現を定義し，部分的アノテーションを使用した教師付き単語分割の分野適応を定式化する．

統計的単語分割には N グラムモデルや隠れマルコフモデル（Hidden Markov Model; HMM）が使われてきたが，近年では識別モデルの条件付き確率場（Conditional Random Fields; CRF）[11] が適用され，その有効性が報告されている[10),15)]．識別モデルは入力変数の分布をモデル化する必要がなく，相関のある特徴を素性として扱うことができるため，広く自然言語処理で活用されるようになっている．また，CRF とこれまでの識別モデルに基づく分類器との大きな違いはラベル構造を出力するという点であり，各点の単語境界の決定に

†1 日本アイ・ビー・エム株式会社
IBM Japan, Ltd.
†2 京都大学
Kyoto University
†3 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

⋆1 日本語処理においては，単語分割と品詞付与を同時に行う形態素解析が一般的であるが，本研究では単語分割のみを対象とする．

© 2009 Information Processing Society of Japan

相互関係がある単語分割問題に適している．CRF は柔軟な素性が使えると同時に文全体の単語分割の整合性を考慮した単語分割器を学習できる．4.1 節では CRF の概要を説明する．

一方，CRF では文全体へのアノテーションが学習データとして必要になり，既存の CRF の学習アルゴリズムでは文の一部にのみアノテーションされたデータについては考慮されていなかった．4.2 節では，アノテーションされていない部分を和で消去した周辺尤度を目的関数とすることで，部分的アノテーションを用いた CRF 学習を可能にする手法を提案する．また，4.3 節で部分的アノテーションを用いたその他の学習方法を検討し，提案法との比較を行う．

5 章では会話例文から医療マニュアルへの分野適応実験により日本語単語分割における提案法の有効性を検証する．単語リストを活用して付与した部分的アノテーションからの学習では，文全体にアノテーションするのに比べて数%の作業量で性能向上が確認された．また，分割ルールによる部分アノテーションからの学習では，人手によるアノテーションなしに単語分割の性能を向上させることができる可能性が示された．6 章で関連研究を紹介し，最後に 7 章では結論と今後の課題について述べる．

2. 部分的アノテーション

本章では，例として，文の一部の単語分割のみをアノテーションする効率的な作業を 2 種類示す．

日本語や中国語のように分かち書きされていない言語では，単語境界を求めることは単純ではない．**図 1** は文字列「切り傷やすり傷」の正しい単語境界を実線で示している．また，単語になりうる複数の部分文字列の境界は破線で示している．この例のように，単語分割候補は複数存在しうるため単なる辞書引きでは単語境界は決定できない．単語分割問題は周辺文脈を考慮したうえで適切に決定する必要があり人手による調整には限界があるため，

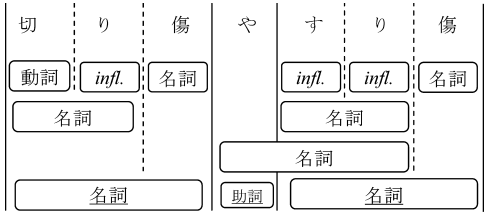


図 1　単語分割の曖昧性の例：*infl.* は動詞の活用語尾を示す
Fig. 1　An example of word boundary ambiguities: *infl.* stands for an inflectional suffix of a verb.

単語分割器の構築には統計モデルが活用されてきた．しかし，統計的手法では言葉の使われ方の違いなどによって学習データと異なる分野のデータでは性能劣化が発生することがある．特に，単語分割では元の分野では観測されなかった未知語の出現が分割誤りの主な原因となっている．たとえば図 1 で「すり傷」が未知語の場合には，文字列「切り傷やすり傷」は誤って「切り傷」「やすり」「傷」と分割されてしまう可能性が高い．

一方で，実際に分野適応を必要とする状況では適応先分野の専門用語辞書や製品名一覧が分野特有の単語リストとして使用可能なことが多い．Mori は適応先の分野単語リストが出現する文脈を評価する KWIC（KeyWord in Context）形式のユーザインタフェース（UI）を提案した[13]．分野単語リストのエントリ「すり傷」の UI での表示例を**図 2** に示す．アノテーション作業者は文字列「すり傷」が適応先テキストの各文脈で実際に単語として使用されている箇所に印を付ける．1 行目は 2 単語「こする」と「傷つく」の一部，2 行目は単語「こすり傷」の一部であるため，最後の 2 行にのみ正しい単語分割であることを示す印を付けている．この UI は文字列領域に対するアノテーションを YES/NO 入力に単純化することで作業を効率化している[⋆1]．

このような UI を使用することで重要な部分だけをアノテーションすることが効率的に行えるようになり，文全体へのアノテーションに比べて作業量の削減が期待できる．それだけでなく，部分的なアノテーションを許すことで作業者が自信のない部分は無理に判断する必要がないため，ノイズとなるアノテーションの追加を防止する効果も期待できる．現実の分野適応時には，分野知識はあるが言語学的知識のない複数人がアノテーション作業を行うことが多いため，この特徴は非常に重要である．

また，人手によるアノテーションだけでなく，事前知識に基づく決定的なルールによって

	が皮膚を強くこ	すり傷	ついてしまっ
	感染，角膜のこ	すり傷	，角膜潰瘍，
○	皮膚に切り傷や	すり傷	を負った場合
○	泥まみれの深い	すり傷	や，皮下深く

図 2　KWIC 形式アノテーション UI：○ は正しい単語分割として選択されたことを示す
Fig. 2　An example of KWIC style annotation UI: marked lines are identified as a correct segmentation.

⋆1 この UI は他の自然言語処理タスクにおいても有用な方法である．たとえば，係り受けアノテーションは単語間の構文的依存関係の有無を文脈を見ながら二値判断する作業の集積と考えることができる．

© 2009 Information Processing Society of Japan

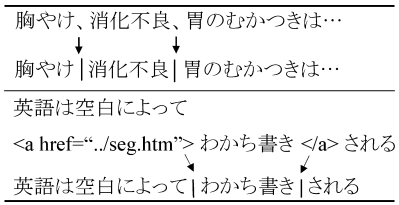


図 3　記号ルールによるアノテーション（実線）の例
Fig. 3　Examples of rule annotations by symbols: The solid lines denote annotated positions.

も部分的に単語境界が付与された文を生成可能である．確実な単語境界・非境界を自動的にアノテーションする単純なルール例として，省略可能な特殊記号の出現箇所での分割を取り上げる．日本語では，読点などいくつかの記号は補助的に使われるため文から削除しても文法的には正しい文となる．また，HTML や XML などのマークアップ言語では，タグ記号を文から削除しても文としては正しい文である．**図 3** に読点（「、」）と HTML のリンクタグを削除した文に対する部分的アノテーション例を示す．これらの記号を文から削除し記号出現箇所を単語境界としてアノテーションすることで，部分的に分割されたテキストを自動生成することができる．このように，文の一部であればルールによって人手をかけずにアノテーションすることも可能である．

以上の例のように，全体の単語分割をアノテーションした文に比べて，部分的にアノテーションされた適応先の文は比較的低コストで得ることができる．

3. 問題定義

本章では，部分的アノテーションを用いた日本語単語分割の分野適応問題を定式化する．

本研究では，2 つの隣接文字の境界を入力単位とし，単語分割を文字境界列に対応する単語境界・非境界ラベル列を出力する問題として扱う[⋆1]．入力文字境界列 $\boldsymbol{x} = (x_1, x_2, \cdots, x_T) \in \boldsymbol{X}$ を文字境界の前後の文字列などを表す変数 $x_t \in X$ [⋆2]の列，ラベル列 $\boldsymbol{y} = (y_1, y_2, \cdots, y_T) \in \boldsymbol{Y}$ を単語境界の有無を表すラベル変数 $y_t \in Y$ の列とする．ただし，ラベル集合は $Y = \{\bigcirc, \times\}$ とし，$\bigcirc$ は単語境界，$\times$ は非単語境界のラベルを示すものとする．以上の定義より，単語分割は写像 $\boldsymbol{X} \to \boldsymbol{Y}$ によって表すことができる．次に，$\boldsymbol{y}$ の一部だけが与えられたデータを表現するために，$\boldsymbol{L} = (L_1, L_2, \cdots, L_T)$ を入力 $\boldsymbol{x}$ の各点 t がとりうるラベル変数の値集合 $L_t \in 2^Y - \{\emptyset\}$ の列とする．たとえば 2 章の KWIC 形式の UI で「すり傷」を文字列

$$怪_{x_1}我_{x_2}は\underbrace{_{x_3}す_{x_4}り_{x_5}傷_{x_6}}_{\text{単語アノテーション}}だ_{x_7}.$$

にアノテーションしたとすると，部分的アノテーションの表現は

$$\boldsymbol{L} = (\{\bigcirc, \times\}, \{\bigcirc, \times\}, \underbrace{\{\bigcirc\}, \{\times\}, \{\times\}, \{\bigcirc\}}_{\text{部分的アノテーション}}, \{\bigcirc, \times\})$$

となる．ただし，1 単語をアノテーションすることで L_3 と L_6 には単語境界，L_4 と L_5 には非単語境界が付与されている．なお，これまで CRF で扱われてきた文全体がアノテーションされたデータは，$t = 1, 2, \cdots, T$ のすべてにおいて要素サイズ $|L_t| = 1$ である $\boldsymbol{L}$ で表現される．よって，$\boldsymbol{L}$ が付与されたコーパスは完全にアノテーションされたコーパスを自然に一般化したものである．

最後に単語分割の分野適応を定義する．

- $D^S = \{(\boldsymbol{x}^{(n)}, \boldsymbol{L}^{(n)})\}_{n=1}^{N}$ を適応元の N 文の学習データ，
- $D^T = \{(\boldsymbol{x}^{(m)}, \boldsymbol{L}^{(m)})\}_{m=N+1}^{N+M}$ を M 文の部分的にアノテーションされた適用先の学習データとする．

部分的なアノテーションを使用した分野適応の目的は，D^S だけを学習データとして用いる場合に比べて D^S と D^T の両方を用いることで適応先分野の単語分割性能を向上させることである[⋆3]．

4. 部分的アノテーションを使用した条件付き確率場の学習

本章では，部分的アノテーションを活用した CRF の学習方法を提案する．

4.1 条件付き確率場

最初に単語分割器として本研究で使用する CRF を説明する．単語分割においての CRF の利点をまとめると，1) N グラムモデルに比べて様々な素性を取り込むことが可能である

⋆1 Peng ら[15] は文字の前が単語境界か否かを各文字に付与する問題として定式化した．しかし彼らの単語分割問題の定式化では，単語境界があることが自明である文の最初の文字にもラベルを付与することになり冗長である．また，Kudo ら[10] は辞書を前提に辞書マッチ結果（形態素ラティス）に対するラベル付与問題として問題を定式化した．しかし，彼らの手法では辞書に存在しない語に対応するためには文字列から単語候補を生成するルールが必要であるが，未知の単語候補の生成ルールは自明ではなく，分野適応で重要な未知語の処理には課題がある．

⋆2 x_t は前後の文字列のように点 t からの相対的な位置を区別した観測情報を含むだけでなく，文末の句読点などのように位置とは独立な観測情報も含めることができる．

⋆3 D^T のみを使用することも考えられる．しかし，D^T が少量で十分な性能を得られない場合には D^S と D^T を合わせて学習データとすることが多い[7]．

© 2009 Information Processing Society of Japan

と同時に，2) ラベル間の相互関係を学習できる点にある．

CRF 以前の N グラムの生成モデルによる単語分割では入力文字列と単語分割の同時分布 $P(\boldsymbol{x}, \boldsymbol{y})$ をモデル化していた．しかし，$P(\boldsymbol{x}, \boldsymbol{y}) = P(\boldsymbol{y}|\boldsymbol{x})P(\boldsymbol{x})$ であるから，同時分布を推定するには単語分割の予測に必要な条件付き分布 $P(\boldsymbol{y}|\boldsymbol{x})$ だけでなく入力文字列の分布 $P(\boldsymbol{x})$ も推定する必要がある．そのため，文字列自身と周辺文字列やその文字種などの相関のある素性を使用すると $P(\boldsymbol{x})$ として素性どうしの同時分布を考える必要がありその推定は非常に困難になる．結果的に，N グラムモデルでは個々の独立性が仮定できる簡素な素性のみを使うことになり素性設計の柔軟性に乏しかった．

一方，識別モデルである CRF は $P(\boldsymbol{y}|\boldsymbol{x})$ を直接モデル化し，素性の分布 $P(\boldsymbol{x})$ は推定する必要がない．そのため，素性の独立性を仮定せずに自由に素性を使用することによって複雑な言語現象に対応できる．たとえば，日本語では隣り合う文字の文字種が異なるとその間の文字境界は単語境界になりやすいが，一方で例外も多く存在する．CRF は相関のある「文字境界の前後の文字」と「文字境界の前後の文字種」を素性として同時に使用することができ，このような例外を含む事象に対応した単語分割器を設計することが可能になる．

また，CRF 以前の識別モデルの分類器は，各点 t ごとに x_t から y_t を独立に予測するモデルであった[⋆1]．しかし，2 章の図 1 で示したように，ある文字境界の単語分割 y_t の決定には前後の文字境界の単語分割 y_{t-1} や y_{t+1} の決定がおおいに関連している．CRF は入力列とラベル列の組 $\boldsymbol{x}$，$\boldsymbol{y}$ を学習データとしており，$\boldsymbol{y}$ に含まれる隣接するラベルの組 y_{t-1}，y_t などを素性として使用することで前後の単語分割の依存関係を表現することを可能にした．

次に CRF を定式化する．$\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{X} \times \boldsymbol{Y} \to \Re^d$ を入力列 $\boldsymbol{x}$ とラベル列 $\boldsymbol{y}$ の組から d 次元の任意の素性ベクトルへの写像，$\boldsymbol{\theta} \in \Re^d$ をモデルのパラメータベクトルとする．CRF は $\boldsymbol{x}$ が与えられたときの $\boldsymbol{y}$ の条件付き確率を次式でモデル化する．

$$P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{e^{\boldsymbol{\theta}\cdot\boldsymbol{\Phi}(\boldsymbol{x},\boldsymbol{y})}}{Z_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{Y}}}, \tag{1}$$

ただし，ベクトル $\boldsymbol{v}$ と $\boldsymbol{w}$ の内積を $\boldsymbol{v}\cdot\boldsymbol{w}$ とし，分母は確率（和が 1）にするための正規化項

$$Z_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{S}} = \sum_{\boldsymbol{y}\in\boldsymbol{S}} e^{\boldsymbol{\theta}\cdot\boldsymbol{\Phi}(\boldsymbol{x},\boldsymbol{y})}.$$

である（$\boldsymbol{S}$ は任意のラベル列集合）．パラメータベクトルは学習データから最尤推定値

$$\hat{\boldsymbol{\theta}} = \operatorname*{argmax}_{\boldsymbol{\theta}} \sum_{n=1}^{N} \ln P_{\boldsymbol{\theta}}(\boldsymbol{y}^{(n)}|\boldsymbol{x}^{(n)})$$

を求めることで決まる．なお，$\hat{\boldsymbol{\theta}}$ は次式の尤度関数の偏微分を用いて勾配法によって計算できる[21]：

$$\sum_{n=1}^{N}\left(\boldsymbol{\Phi}(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)}) - \sum_{\boldsymbol{y}\in\boldsymbol{Y}} P_{\boldsymbol{\theta}}(\boldsymbol{y}^{(n)}|\boldsymbol{x}^{(n)})\boldsymbol{\Phi}(\boldsymbol{x}^{(n)}, \boldsymbol{y}^{(n)})\right). \tag{2}$$

また，$\hat{\boldsymbol{\theta}}$ が与えられたもとでは，ラベル列は $\hat{\boldsymbol{y}} = \operatorname{argmax}_{\boldsymbol{y}\in\boldsymbol{Y}} P_{\hat{\boldsymbol{\theta}}}(\boldsymbol{y}|\boldsymbol{x})$ で予測する．式 (1) にはあらゆるラベル列 $\boldsymbol{Y}$ の和を含むが，マルコフ性を仮定すると動的計画法による効率的な計算方法が知られている[11]．また，動的計画法が適用できない問題においても *Loopy Belief Propagation* など近似的な計算法を用いて適用することができる[24]．

4.2　周辺尤度による条件付き確率場の学習

オリジナルの CRF では，学習データにはラベル列 $\boldsymbol{y}$ が必要になるため，$\boldsymbol{y}$ の一部だけが与えられた $\boldsymbol{L}$ から直接学習することができない．そこで，$\boldsymbol{L}$ に適合するあらゆるラベル列の集合を $\boldsymbol{Y_L} = \{\boldsymbol{y} \mid y_t \in L_t \text{ for } 1 \le t \le T\}$ とし，$\boldsymbol{x}$ が与えられたときの $\boldsymbol{Y_L}$ の条件付き確率

$$P_{\boldsymbol{\theta}}(\boldsymbol{Y_L}|\boldsymbol{x}) = \sum_{\boldsymbol{y}\in\boldsymbol{Y_L}} P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}), \tag{3}$$

の尤度を考える．式 (3) はアノテーションが与えられていない未知の部分のラベル変数列を和で消去した周辺確率である．この周辺確率を使用することで部分的アノテーションの尤度を CRF のモデルで表現することが可能になる．パラメータ $\boldsymbol{\theta}$ の学習データ D での最尤推定値は次式の対数周辺尤度を最大化することで得られる：

$$\begin{aligned}\mathrm{LL}(\boldsymbol{\theta}; D) &= \sum_{(\boldsymbol{x},\boldsymbol{L})\in D} \ln P_{\boldsymbol{\theta}}(\boldsymbol{Y_L}|\boldsymbol{x}) \\ &= \sum_{(\boldsymbol{x},\boldsymbol{L})\in D} \sum_{\boldsymbol{y}\in\boldsymbol{Y_L}} \ln P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) \\ &= \sum_{(\boldsymbol{x},\boldsymbol{L})\in D} \left(\ln Z_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{Y_L}} - \ln Z_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{Y}}\right).\end{aligned} \tag{4}$$

この目的関数を使うことで CRF のモデルを変えずに部分的アノテーションの付いた D^T を使用して CRF のパラメータを学習することができる．

⋆1 CRF 以前にもラベル間の関係を識別モデルに取り入れた手法として，最大エントロピーマルコフモデル（Maximum Entropy Markov Model; MEMM）[25] が提案されていた．MEMM は前の単語分割 $y_1, \cdots, y_{t-2}, y_{t-1}$ が決定されたものとして y_t を予測するモデルである．しかし，学習データで曖昧性の少ないラベル遷移が優先されるという問題や，文を前から解析するモデルと後から解析するモデルでは結果が異なるという問題が知られている[10]．

© 2009 Information Processing Society of Japan

表 1　条件付き分布の違いの例：$\boldsymbol{L} = (\{\bigcirc, \times\}, \{\times\})$
Table 1　Two types of example distributions: $\boldsymbol{L} = (\{\bigcirc, \times\}, \{\times\})$.

$\boldsymbol{x}$	$\boldsymbol{y}$	$P_{\boldsymbol{\theta}}(\boldsymbol{y}\|\boldsymbol{x})$	$P_{\boldsymbol{\theta}}(\boldsymbol{y}\|\boldsymbol{Y}_{\boldsymbol{L}}, \boldsymbol{x})$
(a, b)	$(\bigcirc, \bigcirc)$	0.2	
	$(\times, \bigcirc)$	0.4	
	$(\bigcirc, \times)$	0.3	0.75
	$(\times, \times)$	0.1	0.25

表 2　素性の期待値の違いと勾配方向の例
Table 2　Two types of example feature expectations and gradients.

x	y	$\mathbb{E}_{\boldsymbol{y}\sim P_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y})}[\phi_{x,y}(\boldsymbol{x},\boldsymbol{y})]$	$\mathbb{E}_{\boldsymbol{y}\sim P_{\boldsymbol{\theta}}(\boldsymbol{y}\|\boldsymbol{Y}_{\boldsymbol{L}},\boldsymbol{x})}[\phi_{x,y}(\boldsymbol{x},\boldsymbol{y})]$	$\frac{\partial \mathrm{LL}(\boldsymbol{\theta};D)}{\partial \theta_{x,y}}$
a	$\bigcirc$	0.5	0.75	0.25
a	$\times$	0.5	0.25	-0.25
b	$\bigcirc$	0.6	0	-0.6
b	$\times$	0.4	1	0.6

残念ながら，式 (4) はパラメータ $\boldsymbol{\theta}$ に関して（上に）凸[⋆1]ではなく，効率的に最尤推定値を求めることは困難である．しかし，式 (4) の局所解は通常の CRF と同様に勾配法によって逐次的に計算することができる[⋆2]．勾配法に必要な目的関数の偏微分は次式で与えられる：

$$\frac{\partial \mathrm{LL}(\boldsymbol{\theta}; D)}{\partial \boldsymbol{\theta}} = \sum_{(\boldsymbol{x},\boldsymbol{L})\in D} \left(\sum_{\boldsymbol{y}\in \boldsymbol{Y}_{\boldsymbol{L}}} P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y}_{\boldsymbol{L}}, \boldsymbol{x}) \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y}) - \sum_{\boldsymbol{y}\in \boldsymbol{Y}} P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x}) \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y}) \right), \quad (5)$$

ただし，

$$P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y}_{\boldsymbol{L}}, \boldsymbol{x}) = \frac{e^{\boldsymbol{\theta}\cdot\boldsymbol{\Phi}(\boldsymbol{x},\boldsymbol{y})}}{Z_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{Y}_{\boldsymbol{L}}}} \quad (6)$$

は $\boldsymbol{L}$ に適合するラベル列 $\boldsymbol{Y}_{\boldsymbol{L}}$ のみで正規化した条件付き分布である．

式 (5) は局所解で 0 となるため，提案法では $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y}_{\boldsymbol{L}}, \boldsymbol{x})$ と $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})$ のもとでの素性の期待値 $\mathbb{E}_{\boldsymbol{y}\sim P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y}_{\boldsymbol{L}},\boldsymbol{x})}[\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})]$ と $\mathbb{E}_{\boldsymbol{y}\sim P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}[\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})]$ が等しくなるパラメータ $\boldsymbol{\theta}$ を選択しているといえる．式 (5) 右辺をオリジナルの CRF の偏微分（式 (2)）と比較すると，式 (2) では学習データを表す素性ベクトル $\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})$ そのものに $\mathbb{E}_{\boldsymbol{y}\sim P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}[\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})]$ が近づく向きに $\boldsymbol{\theta}$ を学習していく．一方，提案法の式 (5) では，$\mathbb{E}_{\boldsymbol{y}\sim P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y}_{\boldsymbol{L}},\boldsymbol{x})}[\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})]$ を学習データを表す素性ベクトルと考えて $\boldsymbol{\theta}$ を学習しているといえる．

直感的な理解のために，$X = \{a, b\}$ とし，部分的アノテーションの付与されたデータ $\boldsymbol{x} = (a, b)$，$\boldsymbol{L} = (\{\bigcirc, \times\}, \{\times\})$ を例として考える．また，あるパラメータ $\boldsymbol{\theta}$ での $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})$ が**表 1** の 3 列目に示す値であったとする．すると $\boldsymbol{L}$ が与えられたときの $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y}_{\boldsymbol{L}}, \boldsymbol{x})$ は 4 列目のようになる．また，仮に文中の各点の X と Y の組の数を合計した $\phi_{x,y}(\boldsymbol{x}, \boldsymbol{y}) = \sum_t [x_t = x \wedge y_t = y]$ を素性として考えると，各分布のもとでの素性の期待値と偏微分の値は**表 2** に示すようになる．ただし，$\theta_{x,y}$ は $\phi_{x,y}(\boldsymbol{x}, \boldsymbol{y})$ に対応するパラメータである．この例から，部分的アノテーションが与えられた点 $t = 2$ に隣接する点 $t = 1$ の素性（$\phi_{a,\bigcirc}(\boldsymbol{x}, \boldsymbol{y})$，$\phi_{a,\times}(\boldsymbol{x}, \boldsymbol{y})$）の期待値にも情報が伝播し，そのパラメータが変化することが分かる．

式 (4)，(5) は $\boldsymbol{Y}_{\boldsymbol{L}}$ および $\boldsymbol{Y}$ での和を含んでいるが，その数 $|\boldsymbol{Y}_{\boldsymbol{L}}| = |L_1| \times |L_2| \times \cdots \times |L_T|$ および $|\boldsymbol{Y}| = Y^T$ は指数的であり明示的に列挙することは困難である．しかし，マルコフ性を仮定すると式 (4)，(5) は既存の CRF の学習で使用される動的計画法を少し修正することで多項式時間で計算することが可能である．このアルゴリズムの詳細は付録 A.1 で説明する．

なお，CRF の学習では最尤推定による過学習を防ぐために，パラメータの事前分布 $P(\boldsymbol{\theta})$ を目的関数に含めて $\boldsymbol{\theta}$ の事後確率を最大化する推定法が用いられる．5 章の実験では平均 0，分散 σ^2 の正規分布を事前分布とした．また，分野適応の手法では目的関数における D^S と D^T のバランスを調整するパラメータを導入することが多い[7]．本研究では D^T の対数尤度に対する重み ω を調整パラメータとする．これら 2 点を考慮した提案法の目的関数は次式となる：

$$\mathrm{LL}(\boldsymbol{\theta}; D^S) + \omega \, \mathrm{LL}(\boldsymbol{\theta}; D^T) - \frac{||\boldsymbol{\theta}||^2}{2\sigma^2}.$$

また，その偏微分は次式で与えられる．

$$\frac{\partial \mathrm{LL}(\boldsymbol{\theta}; D^S)}{\partial \boldsymbol{\theta}} + \omega \frac{\partial \mathrm{LL}(\boldsymbol{\theta}; D^T)}{\partial \boldsymbol{\theta}} - \frac{\boldsymbol{\theta}}{\sigma^2},$$

ただし，σ と ω は問題に合わせて決定する必要があるハイパーパラメータである．

4.3 議　　論

本節では，部分的アノテーション $\boldsymbol{L}$ を用いて分野適応するための他の手法を検討し提案法と比較する．

$\boldsymbol{L}$ を用いて CRF を学習する別の方法としては，アノテーションの付与されていない点

⋆1 $\mathrm{LL}(\boldsymbol{\theta}; D)$ は 2 つの凸関数の差になっており，関数全体としては凸関数とはならない[5]．
⋆2 $\boldsymbol{\theta}$ の初期値によって局所解は異なる．5 章の実験では適応元の CRF のパラメータを初期値として使用した．

© 2009 Information Processing Society of Japan

$\{t \mid |L_t| > 1\}$ の y_t を適応元データで学習した CRF で予測し学習データとする方法が考えられる．適応元データで学習した CRF のパラメータベクトルを $\tilde{\boldsymbol{\theta}}$ とすると，部分的アノテーション $\boldsymbol{L}$ に適合するラベル列 $Y_{\boldsymbol{L}}$ から最も確率の高いラベル列 $\hat{\boldsymbol{y}} = \mathrm{argmax}_{\boldsymbol{y} \in \boldsymbol{Y_L}} P_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{y}|\boldsymbol{x})$ は動的計画法により効率的に計算可能である[6]．$\hat{\boldsymbol{y}}$ は $\boldsymbol{L}$ と適合するラベル列の 1 つであるため，生成された学習データには部分的アノテーションが反映されているといえる．また，この予測結果を使って学習データ $(\boldsymbol{x}, \hat{\boldsymbol{y}})$ とすれば，既存の CRF と同じ学習アルゴリズムを使うことができる（以降，**予測列 CRF** と呼ぶ）．しかし，確率が 1 番高いラベル列とそれ以外のラベル列とで $P_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{y}|\boldsymbol{x})$ の差が小さい場合にも，2 番目以降の候補は学習データとして考慮されない．このような場合，$\tilde{\boldsymbol{\theta}}$ のわずかな違いによって $\hat{\boldsymbol{y}}$ が変化してしまう．そのため，予測 $\hat{\boldsymbol{y}}$ を正解データとして学習すると $\tilde{\boldsymbol{\theta}}$ の推定誤差に敏感になり，予測列 CRF の学習結果のバラつきが大きくなってしまう可能性がある．一方，提案法は予測ラベル列の分布 $P_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{Y_L}, \boldsymbol{x})$ のもとでの $\boldsymbol{\phi}(\boldsymbol{x}, \boldsymbol{y})$ の期待値を学習データとしているため，$\hat{\boldsymbol{y}}$ のみを学習ラベル列とする予測列 CRF と比較して学習結果が安定すると期待できる．

また，CRF を使わずに，各文字境界 t に対する y_t を独立に学習・予測する分類器を使用する方法も考えられる（以降，**点分類器**と呼ぶ）．点分類器では部分的アノテーションが付いた点 $\{t \mid |L_t| = 1\}$ のみを学習データとすればよいため，部分的アノテーションに対応するために特別な考慮は必要ない．また，点分類器に識別モデルを使用することで柔軟に素性を使用することができる．しかし，4.2 節で述べたように点分類器では前後の単語分割を考慮しないため，完全にアノテーションが与えられている問題では CRF の方が良い性能を示すことが知られている[⋆1]．部分的アノテーションを扱う問題でも点分類器に対する CRF の優位性が提案法により保たれることが期待される．

5. 実　　験

本章では実データによって日本語単語分割の分野適応における提案法の有効性を検証した結果を示す．部分的アノテーションの付与には 2 章で説明した単語リストと分割ルールを用いた．最初に共通の実験設定を 5.1 節で説明し，5.2 節で単語リスト，5.3 節でルールを使用して部分的アノテーションを付与した実験結果を示す．

⋆1 CRF では，前後の単語分割を考慮することで，前方・後方の単語分割決定に寄与している遠距離の素性を間接的に参照していると解釈することもできる．点分類器でも明示的に遠距離の素性も使用することで CRF と同等の性能を示す可能性はあるが，遠距離の素性も使用すると素性数が膨大になるという問題がある．

表 3　データ統計
Table 3　Data statistics.

	分野	使用目的	分割済	文数	単語数
A	適応元	学習データ	○	11,700	145,925
B	適応元	検証データ	○	1,300	16,348
C	適応先	検証データ・部分的アノテーション	○	1,000	29,216
D	適応先	部分的アノテーション	×	53,834	N/A

5.1 実 験 設 定

本実験では適応元のデータとして日常会話辞書[9] の例文を用いた．また，適応先のデータとして医療マニュアル[4] の文を用いた[⋆2]．適応元の文は人手によりすべて単語に分割し，適応元の学習用データと性能検証用データに分けて使用した（**表 3** の A と B）．なお，データ A は文全体が完全にアノテーションされた学習データとして扱った．また，人手により適応先の 1,000 文を単語分割した（表 3 の C）．このデータ C は適応先の性能検証用データおよび，5.2 節での単語リストによる部分的アノテーション付きデータとして使用した．さらに 5.3 節では分割ルールを適用するために単語分割されていない適応先の約 5 万文を使用した（表 3 の D）．

性能評価には単語の分かち書きの再現率（R）と精度（P）の調和平均値 $F = 2RP/(R+P)$ を用いた：

$$R = \frac{\text{正解単語数}}{\text{全単語数}} \times 100, \quad P = \frac{\text{正解単語数}}{\text{システムの出力単語数}} \times 100.$$

本実験では 1 次のマルコフモデル CRF を実装した[⋆3]．文字境界を表す 2 値素性として，その境界周辺の**文字 N グラム**および**文字種 N グラム**（$N = 1, 2, 3$）を使用した．なお，文字種 N グラム素性は文字種が異なる文字境界では単語境界になりやすいという事前知識を

⋆2 単語分割の基準が各コーパスごとに異なるため，本研究では一般に公開されているコーパスではなく共通の基準で単語分割された 2 種類のコーパスを用いた．

⋆3 固有表現抽出などの分割問題ではセミマルコフモデル CRF[19] が使われることが多い．ただし，セミマルコフ CRF は解析時の計算量が $O(kT|Y|^2)$ であり，マルコフモデル CRF（$O(T|Y|^2)$）に比べて解析速度が劣る．なお，k は最大の単語長である．
本実験で使用した分野辞書エントリの最大長は 28 文字，また実験データに含まれる単語の最大長は 12 文字と比較的長い単語（主に専門用語）を含んでいた．計算量に対する最大単語長 k の影響が大きくなるため，本実験では解析速度を重視しマルコフモデル CRF を採用した．

© 2009 Information Processing Society of Japan

表 4　文字境界を表す素性：c の下付き添字は文字境界からの相対距離を示す
Table 4　Features for a character boundary: The subscript of c stands for the relative distance from the character boundary.

素性種別	参照文字位置
文字 N グラム	$c_{-1}, c_{+1},$
文字種 N グラム	$c_{-2}c_{-1}, c_{-1}c_{+1}, c_{+1}c_{+2},$
辞書 N グラム	$c_{-2}c_{-1}c_{+1}, c_{-1}c_{+1}c_{+2}$
辞書語開始	c_{+1} から文末まで
辞書語終了	c_{-1} から文頭まで

反映したものである[⋆1]．また，辞書の単語も素性に反映した．辞書素性の 1 つは，上記の N グラムが辞書に存在するかを表す**辞書 N グラム**素性[15] である．もう 1 つはある文字境界で開始または終了する辞書エントリが存在するかを表す素性（**辞書語開始**，**辞書語終了**）である．対象コーパスの単語は短単位であるため，汎用的な辞書として *unidic*[⋆2]（活用形を展開して約 28.1 万エントリ）を使用した．また，専門用語辞書としては電子カルテ用標準病名マスタ（*JSDCM*）[⋆3]（約 2.3 万エントリ）を使用した．使用した素性を**表 4** にまとめる．素性の例として，文字列「やすり | 傷を」の「り」「傷」の間の文字境界を表す素性を以下に示す．値を持つ文字 N グラム素性は { り |，| 傷，すり |，り | 傷，| 傷を，すり | 傷，り | 傷を } である．また，値を持つ文字種 N グラム素性は { H|，|K，HH|，H|K，|KH，HH|K，H|KH } となる．ただし，「|」は注目する文字境界の場所を示す補助記号であり，文字種のひらがなを H，漢字を K とする．さらに，この例で辞書に「やすり」「傷」の 2 語が存在するとすると，1 文字後の文字（| 傷）の辞書 N グラム素性（c_{+1}），辞書語開始素性（| 傷）および辞書語終了素性（やすり |）が値を持つ．なお，パラメータ数を減らすために検証・テスト用でないデータ A と D で高頻度な素性のみを使用した[⋆4]．最終的な入力素性数は約 30 万である．

部分的アノテーションを追加する前の予備実験として，適応元のデータ A で学習した CRF（以降 **CRF$^{\mathbf{S}}$**）の適応先データでの性能を調査した．この際，CRF$^{\mathrm{S}}$ のハイパーパラメータ σ は適応元の開発用データ B を使用して最適な値を選択した．**表 5** に予備実験の結果をまとめる．適応元のデータのみで学習した CRF は適応元での性能 $F = 96.92$ に比べて，適応先での性能が $F = 92.30$[⋆5] にとどまり分野適応の必要性が確認された．

表 5　予備実験結果（分野適応なし）
Table 5　The word segmentation performance without domain adaptation.

分野	検証データ	F
適応元	B	96.92
適応先	C	92.30

以降の実験では，部分的アノテーションから学習する手法として，提案法以外に 4.3 節で説明した 1) D^T を CRF$^{\mathrm{S}}$ で予測した単語分割を CRF の学習データとした予測列 CRF と，2) 文字境界を 1 つの学習データと考えて独立に予測する点分類器を実装した．点分類器には条件付き分布 $P(y_t|x_t)$ をモデル化するロジスティック回帰（最大エントロピー分類器）を用いた．また，入力文字境界を表す素性は CRF と同じ素性を使用し，パラメータ推定には過学習を防ぐために CRF と同様に事後確率最大化推定を用いた．適応元データ A で学習した点分類器（**点分類器$^{\mathbf{S}}$**）の適応先データ C での性能は $F = 91.29$ であった．ただし，点分類器のハイパーパラメータ σ も CRF と同様にデータ B でチューニングした．この結果より，分野適応前の CRF$^{\mathrm{S}}$ は 点分類器$^{\mathrm{S}}$ より性能面で勝っているといえる．

5.2　単語リストによるアノテーション

本節では，単語リスト中の単語の適応先データでの出現箇所にアノテーションを付与して作成した部分的アノテーションを追加学習データとしたときの提案法の有効性を検証する．

本実験では適応先データ C を，C1）部分的アノテーション用および学習用 500 文と，C2）評価対象 500 文，に分け 2 分割交差検証を行った．単語リストとしては，病名辞書 JSDCM を使用した．JSDCM 中の病名が適応先データ C1 に実際に出現した異なり語数は各分割での平均で 224 語であり，また出現数は各分割において約 1,000 カ所であった．

5.2.1　部分的アノテーション数を変化させたときの性能

最初に，単語アノテーション数を 100 カ所から 1,000 カ所に変化させて性能を評価した．アノテーションする箇所には優先順位付けを行った．まず，単語リストの各単語ごとに重要度の高い 1 カ所のみをアノテーションした．重要度にはラベルのエントロピー

⋆1 文字種にはひらがな，カタカナ，漢字，英字，アラビア数字，記号を用いた．
⋆2 Ver. 1.3.5; http://www.tokuteicorpus.jp/dist/
⋆3 Ver. 2.63; http://www2.medis.or.jp/stdcd/byomei/
⋆4 具体的には，C_A と C_D をそれぞれデータ A と D における素性の出現頻度としたとき，単語分割のあるデータ A を重視して式 $C_A + 0.5C_D \geq 2$ を満たす素性のみを用いた．
⋆5 CRF$^{\mathrm{S}}$ から辞書素性を除いた CRF の性能は $F = 92.23$ であり，分野辞書 JSDCM の追加のみでは適応先での十分な性能向上は得られなかった．

© 2009 Information Processing Society of Japan

$H(\boldsymbol{y}_t^s) = -\sum_{\boldsymbol{y}_t^s \in \boldsymbol{Y}_t^s} P_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{y}_t^s|\boldsymbol{x}) \ln P_{\tilde{\boldsymbol{\theta}}}(\boldsymbol{y}_t^s|\boldsymbol{x})$ を用いた[2]．ただし，$\tilde{\boldsymbol{\theta}}$ は $\mathrm{CRF}^{\mathrm{S}}$ のパラメータ，$\boldsymbol{y}_t^s = (y_t, y_{t+1}, \cdots, y_s) \in \boldsymbol{Y}_t^s$ は $\boldsymbol{y}$ の t から s までの部分列である．直感的には，ある単語が出現した複数の文脈の中で，適応元モデル $\mathrm{CRF}^{\mathrm{S}}$ にとって判断が難しい文脈を優先的にアノテーションしていることになる[⋆1]．単語リストのすべての単語に対して 1 カ所アノテーションした後は，可能ならば各単語ごとにもう 1 カ所アノテーションを増やし，これを繰り返した．なお，1,000 カ所のアノテーションであってもデータ C1 の全出現単語数の約 7%程度であり，全単語をアノテーションするのに比べて非常に少ない作業量である．

適応先のデータは限られておりハイパーパラメータ調整用に適応先データが使えない場合を想定し，σ は $\mathrm{CRF}^{\mathrm{S}}$ または 点分類器$^{\mathrm{S}}$ と同じ値を使用し，$\omega = 1$[⋆2]と設定した．

図 4 (a) は単語アノテーション数を変化させたときの単語分割性能を示している．単語アノテーション数 = 0 の点が $\mathrm{CRF}^{\mathrm{S}}$ および 点分類器$^{\mathrm{S}}$ の性能である．なお，提案法の目的関数は局所解が存在するため，$\mathrm{CRF}^{\mathrm{S}}$ のパラメータ $\tilde{\boldsymbol{\theta}}$ を初期値として最適化を行った結果（**初期値=適応元 CRF**）を示す．どの手法も部分的なアノテーションを追加することで $\mathrm{CRF}^{\mathrm{S}}$ に比べて性能を向上させることができた．しかし，予測列 CRF はアノテーション数の変化に対して性能向上は安定的ではなかった．これは，4.3 節で述べた予測列 CRF の課題点と合致した結果となっている．一方，提案法は点分類器に対する優位性を保ったまま部分的アノテーションを用いて適切に学習できていることが示された．なお，Wilcoxon 符号順位検定（有意水準 5%）によって提案法とその他の手法との性能差は有意であることを確認した．また，単語アノテーション数=100〜200 の間で特に性能向上率が高かった．単語リスト中の単語がデータ C1 に実際に出現した異なり語数は 224 であるから，この間は単語リスト中のすべての単語を 1 回だけアノテーションしている過程である．よって，適応先の文に出現する単語が分野単語リストにより多く含まれている場合にはさらなる性能向上が期待できる．

5.2.2　アノテーション箇所の優先順位付けの効果

次に，アノテーション箇所の優先順位付けの効果を調査した．そのために，1) 優先順位付けした場合（**ラベルエントロピー順**）と，2) 各単語が出現した順にアノテーションを付与した場合（**出現順**）との提案法の単語分割性能を比較した．なお，ハイパーパラメータやパラメータの初期値は $\mathrm{CRF}^{\mathrm{S}}$ と同じ値を用いた．

単語アノテーション数を変化させたときの単語分割の性能を図 4 (b) に示す．アノテーションした単語数が 100 個から 500 個の間で，ラベルエントロピーによって優先順位を付けることで性能が良くなることが確認された．Wilcoxon 符号順位検定（有意水準 5%）で手法間の性能差は有意であり，実験結果は提案法とラベル箇所の優先順位付けを組み合わせることで少ないアノテーション数で効果的な学習が可能であることを示している．

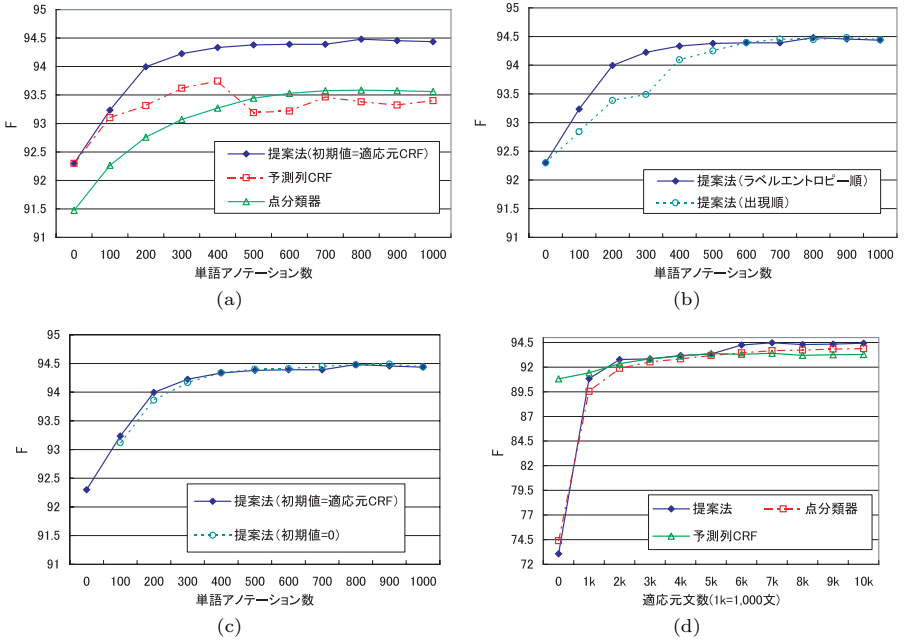


図 4　交差検証での平均性能：(a) 適応先部分的アノテーション数変化時（2 試行平均）；(b) 優先順位付け有無比較（2 試行平均）；(c) パラメータ初期値比較（2 試行平均）；(d) 適応元データ数変化時（6 試行平均）

Fig. 4　Average performances by cross-validation: (a) varying the number of word annotations over 2 trials; (b) with and without prioritization over 2 trials; (c) with parameter initialized as zero and $\mathrm{CRF}^{\mathrm{S}}$ over 2 trials; (d) varining the number of source domain data over 6 trials.

⋆1 リスト中の単語の出現箇所の平均エントロピーが 0.32 であったのに対し，リスト外の単語の出現箇所では平均 0.2 であったことから，分野辞書を用いてエントロピーの点で重要度の高い箇所を選択することが可能であったといえる．
一方，単語リストに限定せずに任意の部分文字列をアノテーション箇所の候補とすることも可能である．しかし，文の一部よりも文全体をアノテーションしたほうが好ましいことからも明らかであるように，部分文字列は長いほどその重要度は大きくなるがその分作業負荷も高くなる[2]．そのため任意の部分文字列の優先順位付けには作業負荷も考慮した重要度指標に課題が残る．

⋆2 文献 7) では，$\omega > \frac{N}{M}$ と設定することで適応先の性能が良くなることを示している．しかし，そのためには ω を調整するためのアノテーション済み適応先データが必要になる．

© 2009 Information Processing Society of Japan

5.2.3　パラメータ初期値への依存度

また，提案した目的関数に局所解があることによる単語分割性能への影響を調べた．図 4 (c) にパラメータの値をすべて 0 に初期化して学習した CRF（**初期値=0**）と，CRF^{S} のパラメータを初期値とした CRF（**初期値=適応元 CRF**）の結果を示す．有意水準 5%の Wilcoxon 符号順位検定では初期値による性能差は認められず，提案法におけるパラメータの初期値の影響は小さいことが確かめられた．

5.2.4　適応元文数を変化させたときの性能

最後に，適応元の学習データ数を変化させたときの性能を示す．部分的アノテーション数を固定（1,000 単語）し，適応元データの文数を変化（0〜10,000）させ，提案法・点学習器・予測列 CRF を学習した結果を図 4 (d) に示す．ただし，適応元データはデータ A から部分集合をランダムサンプリングし，サンプリング試行 3 回および適応先データの 2 分割交差検証の計 6 試行の平均性能である．なお，提案法のパラメータは適応元文数が 0 のときは 0 に初期化し，1,000〜10,000 ではそれぞれの文数で学習した適応元 CRF のパラメータを初期値とした．また，適応元文数が 0 のときは予測列 CRF は CRF^{S} による適応先の単語分割結果を使用して学習している．そのため厳密には提案法・点学習器より予測列 CRF は有利な条件であることに注意してほしい．

提案法は適応元データが少ない場合において予測列 CRF・点学習器より性能が劣る[*1]が，それ以上の文数では他の手法を上回った．提案した部分的アノテーションからの学習方法は数千文以上の適応元の完全なアノテーションと適応先の部分的アノテーションを併用したときに有効であることが本実験によって示された．なお，Wilcoxon 符号順位検定（有意水準 5%）で提案法と他の手法の性能差は有意であった．

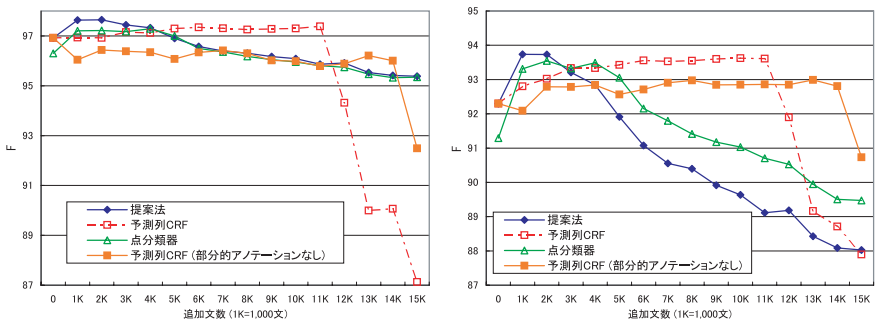
5.3　分割ルールによるアノテーション

本実験では，適応先データに 2 章で述べた記号による分割ルールを適用した．ルールにより部分的にアノテーションされた学習データを自動生成し，これを用いて学習した単語分割器の性能を検証した．

学習データとしては，適応先の未分割データ D に対して分割ルールを適用し部分的アノテーションが付与された文を生成した．なお，分割ルールの記号は多くの場合省略可能である**表 6** の 11 種を使用した．分割ルールで生成したデータのうち 15,000 文を学習データとし，学習データ追加による性能変化を調べた．15,000 文の選択はランダムサンプリングであり，サンプリング 3 回での平均性能を評価した．

表 6　分割ルールに使用した記号
Table 6　Symbols for segmentation rule.

、	。	(	)	”	「	」	／	：	；	・



(a) 適応元データ B で評価　　(b) 適応先データ C で評価

図 5　部分的アノテーションされた文数を変化させたときの性能（3 試行の平均）
Fig. 5　Performances varying the number of partially annotated sentences (averaged over 3 trials).

分割ルールによって付与された単語境界アノテーションは人手による確認は行われない．そこで，この分割ルールによる部分的アノテーション自身の有効性についても検証するために，ルール適用前のデータ D の単語分割を CRF^{S} で予測し学習データとした CRF（**予測列 CRF（部分的アノテーションなし）**）も比較対象とした．また，前節の実験と同様，$\omega = 1$ とし，σ は CRF^{S} または 点分類器^{S} と同じ値を用い，提案法の初期値を $\tilde{\boldsymbol{\theta}}$ として最適化を行った．

5.3.1　部分的アノテーション数を変化させたときの性能

図 5 は，部分的にアノテーションされた文数を $M = 1{,}000$〜$15{,}000$ と変えたときに，各手法の性能を適応元データ B（図 5 (a)）と適応先データ C（図 5 (b)）で評価した結果を示す．予測列 CRF（部分的アノテーションなし）の適応元データ B での性能がつねに CRF^{S} を下回ったのを除けば，どの手法も特定の文数を追加した時点では CRF^{S} を上回る性能を示した．性能のピーク値を比較すると提案法が 1,000 文を追加した時点で最も良い性能を示し

[*1] この原因としては，4.2 節で示したように提案法は部分的アノテーションが与えられた箇所以外の y_t の確率分布を最適化中のパラメータから推定する点があげられる．部分的アノテーションのみが与えられている場合には最適化中のパラメータは信頼性が低く，誤った y_t の確率分布を正解データとして学習することで性能が劣化したと考えられる．

© 2009 Information Processing Society of Japan

た．CRF^{S} と比較してパラメータが増加した素性には「か | しこ」・「シン |」・「| 血小」・「腹痛 |」（いずれも $y_t = ◯$）などがあった．このことから「しこり」・「エピルビシン」・「血小板」・「腹痛」などの医療分野特有の語の単語境界を学習できていることが分かる．一方，分割ルールによる部分的アノテーションを使用していない予測列 CRF（部分的アノテーションなし）のピーク性能が最も低かった．この結果より，ルールによって自動生成した部分的アノテーションによる性能向上の可能性が示された．またデータ B で良い性能であった D^T の文数はデータ C でも他の追加文数と比べて良い性能を示し，適応元の性能と適応先の性能の間には正の相関関係があることが観察された．

5.3.2　適応元・適応先データのバランス調整による性能改善

しかし，D^T の文数を増やすことで最終的にはどの手法も分野適応前の CRF^{S} の性能を下回る結果となった．提案法の誤りを分析したところ，D^T の文数が増えると単語を短く分割してしまうエラーが CRF^{S} に比べて増加していた．実験で使用したルールでは，単語境界のアノテーションは与えられるが単語を構成する文字列内の非単語境界のアノテーションがまったく与えられない．その結果，ある文字境界は単語境界になりやすいという情報が学習データ中に相対的に多くなり，細かく単語分割する傾向になったと考えられる[⋆1]．

そこで，学習データにおける偏りを補正するために適応元データを用いて重み ω を 1 より小さい値に調整した．**図 6** に ω を調整したときの適応先データ C での単語分割性能を示す．ただし，適応先には十分な評価データがないことを想定し，各追加文数での ω の決定には適応元データ B で性能が最も良くなる値を $\omega = 0.05$〜1 の間から選択した．重み付けの調整により部分的アノテーションを自動付与した文数にかかわらず CRF^{S} よりも適応先での性能を向上させることができた[⋆2]．その中でも，提案法はつねに他の手法に比べて良い性能を示した．提案法と他の手法との性能差は Wilcoxon 符号順位検定（有意水準 5%）により有意であった．

ただし，ここでは本実験データにおける適応元と適応先の性能との相関関係を利用して，適応元データで重み ω の選択を行った．適応元と適応先の間で性能の相関がつねに正である保証はないことから，重み ω の決定方法には実用上の課題が残る．

⋆1 たとえば，CRF^{S} では正しく分割されていたフレーズ「人体/の/しくみ」（「/」は単語境界）は，「人/体/の/し/くみ」として誤って分割されてしまっていた．「人体」「しくみ」が単語として D^S に出現しない一方で，D^T に多く観測された素性「| 体」・「し |」・「K|K」・「H|H」（いずれも $y_t = ◯$）のパラメータ値が大きくなったため，このような分割誤りの原因となったと考えられる．

⋆2 一方で，適応元データ B の性能を参照して行った σ の調整では，D^T の文数 M を変えたときの適応先での性能を安定させることはできなかった．

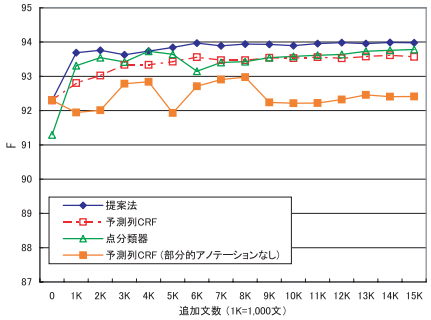


図 6　部分的にアノテーションした文に対する重み ω を調整したときの適応先での性能（3 試行の平均）
Fig. 6　Performance improvements in the targe domain by tuning ω for partially annotated sentences (averaged over 3 trials).

6. 関 連 研 究

近年の単語分割研究の中心的な課題は未知語の処理である．未知語処理研究の主なアプローチは，1) 未知語を扱う統計モデルの工夫と，2) 語彙の獲得である[14]．N グラム生成モデルにおいては単語連接の統計モデルとは別に未知語のモデルも推定する必要があった．しかし，未知語モデルで使用していた文字種などが素性として取り入れることが可能になったため，識別モデルでは未知語モデルを特別に扱う必要はない．一方，語彙獲得は識別モデルにおいても重要な課題である．提案法では分野単語リストを仮定したが，分野単語リストは実際の文書に出現する単語をカバーするには十分でないことが多い．文献 15) はコーパス中の辞書に存在しない文字列から CRF の確率モデルを使用して新語を検出する方法を提案している．ただし，文献 15) では誤りを含む新語候補の追加によって CRF の単語分割性能が低下する場合があることも報告している．新語候補の出現文脈に人手で部分的アノテーションを付与し，提案法で CRF を学習することによって，確実な性能向上が期待できる．

構文解析の文脈では，Pereira ら[16] の研究で部分的に構成素領域がアノテーションされた構文木から確率的文脈自由文法を獲得する方法が提案されている．文献 16) は木構造出力における部分的アノテーションからの生成モデルの学習であり，提案法を応用することで構文解析における部分的アノテーションからの CRF の学習が可能になると期待できる．

5.2 節の実験では，アノテーション箇所の優先順位付けを CRF を使用して行ったが，これは能動学習として研究されている手法である．しかし，アノテーションのために部分構造

© 2009 Information Processing Society of Japan

を選択する能動学習の先行研究[1),3),18),20)]では，学習器に N グラムモデル，HMM，点分類器を仮定していた．提案法によって CRF と部分構造の能動学習を組み合わせることが可能になったといえる．

また，部分的アノテーションを扱った研究としては，Culotta ら[6)] は CRF を活用したインタラクティブなアノテーションシステムの研究において，付与された部分的なアノテーションを満たすラベル列を修正し作業量を減らす方法を提案している．この手法では，最終的には文全体のラベルが修正されることを仮定しており，CRF は通常どおり学習される．作業者がすべてのラベルを修正できない場合には部分的アノテーションが残るため，提案法を組み合わせることが有効である．

アノテーションがあるデータとアノテーションがまったくないデータを使用して学習する半教師付き学習の枠組みで CRF を学習する手法[8),12)] が提案されている．半教師付き学習は適応先データに限りがある分野適応で有望な手法ではあるが，構造データの一部にアノテーションがある場合は考慮されていなかった．部分的アノテーションのある適応先データでは提案法を使用し，アノテーションがまったくないデータでは半教師付き学習を用いることで適応先データを有効活用できる．

また，我々の研究とは独立に，文献 26) は部分的にラベル付けされた画像から CRF を学習する手法を提案している．目的関数は式 (4) と同じであるが，彼らの問題（場面分割）では目的関数と偏微分の評価が多項式時間で行えないため，*Loopy Belief Propagation* を用いて近似的に計算を行っている．

7. 結論と今後の課題

本研究では，日本語の単語分割の分野適応時に学習データ作成作業を削減するために，部分的アノテーションを使用した条件付き確率場のパラメータ推定法を提案した．計算機実験により，提案法がアノテーション作業を減らすとともに単語分割性能を向上させることを確かめた．単語分割器の分野適応を容易にすることによって，単語分割器を基盤技術として使用したテキストマイニングや機械翻訳などの分野適応作業の効率向上も期待できる．

5.3 節の実験で明らかになったように，部分的アノテーションを使用した学習ではアノテーション箇所に偏りがある場合に性能悪化が発生する可能性がある．本研究の実験では適応先データへの重みを適切に調整することで問題を回避できることを示したが，一般の分野適応タスクにおける重みの選択方法には課題が残った．分野適応において部分的アノテーションを広く活用するには，適応先の正解データなしに重みの決定を可能にする手法の開発が課題である．本研究では適応元と適応先データのバランスを重みで調整したが，付与されるラベルの偏りを補正するための重みを用いる手法なども重要な検討事項である．また，記号を削除するルールで部分アノテーションを自動的に付与した文の中には，結果として日本語として不自然な表現も生成されてしまうことも確認された．これらの不自然な事例の学習への影響を排除するなど，ノイズを含む部分的アノテーションから頑健に学習する手法も重要であると考える．一方で，単語分割を自動的に付与するルールを利用するためには部分的な非単語分割アノテーションを与えるルールの検討も今後の課題である．

また，部分的アノテーションを用いて CRF 以外の構造学習器を学習する手法の検討も興味深い課題である．本研究では，条件付き確率モデルを正則化項付きの最尤推定により学習したが，たとえば正しい出力と最も尤度の高い不正解出力との対数尤度比の最大化による分類器を学習する方法も考えられる．対数尤度比：

$$r_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = \ln \frac{\boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}{\max_{\bar{\boldsymbol{y}}\neq\boldsymbol{y}} \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}$$

に対し，対数尤度比が 1 より小さい事例のみを対象とした $\max(1 - r_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}), 0)$ を損失関数（ヒンジ損失）とした最小化問題を考えると多クラスのサポートベクタマシン（SVM）を導くことができる[17)]．文献 22) の欠損値の扱いと同様に，部分的アノテーションが与えられたデータの対数尤度比は次式のように定めることができる．

$$r_{\boldsymbol{\theta}}(\boldsymbol{x},\boldsymbol{y}) = \ln \frac{\sum_{\boldsymbol{y}\in\boldsymbol{Y_L}} \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}{\max_{\bar{\boldsymbol{y}}\notin\boldsymbol{Y_L}} \boldsymbol{p}_{\boldsymbol{\theta}}(\boldsymbol{y}|\boldsymbol{x})}.$$

すると，部分的アノテーションを含むデータを用いた SVM の学習は次の制約付き最小化問題を解くことに相当する．

$$\min \sum_{n=1}^{N+M} \xi^{(n)} + \frac{||\boldsymbol{\theta}||^2}{2\sigma^2}$$

制約条件：$\ln Z_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{Y_L}} - \max_{\boldsymbol{y}\notin\boldsymbol{Y}_{\boldsymbol{L}^{(n)}}} \boldsymbol{\theta}\cdot\boldsymbol{\Phi}(\boldsymbol{x}^{(n)},\boldsymbol{y}) \geq 1 - \xi^{(n)}$ かつ，$\xi^{(n)} \geq 0$.

さらに，本研究では凸関数の差の形式の目的関数 (4) を勾配法を用いて最適化し局所解を求めたが，文献 22) では確率モデルに限らず凸関数の差となる目的関数を Concave-Convex Procedure（CCCP）と呼ばれる最適化法により局所解を求め，分類器を学習する方法を提案している．CCCP と A.1 節の計算方法と組み合わせることで，式 (4) の最適化にも適用可能である．部分的アノテーションを用いた SVM の学習には，CCCP の使用も検討の余地があるだろう．

© 2009 Information Processing Society of Japan

謝辞　有益なコメントをいただいた査読者の皆様に感謝を申し上げる．

参 考 文 献

1) Anderson, B. and Moore, A.: Active Learning for Hidden Markov Models: Objective Functions and Algorithms, *Proc. 22nd International Conference on Machine Learning*, pp.9–16 (2005).
2) Anderson, B., Siddiqi, S. and Moore, A.: Sequence Selection for Active Learning, Technical Report CMU-IR-TR-06-16, Carnegie Mellon University (2006).
3) Argamon-Engelson, S. and Dagan, I.: Committee-Based Sample Selection for Probabilistic Classifiers, *Journal of Artificial Intelligence Research*, Vol.11, pp.335–360 (1999).
4) Beers, M.H.：メルクマニュアル医学百科—最新家庭版，日経 BP 社 (2004).
5) Boyd, S. and Vandenberghe, L.: *Convex Optimization*, chapter 3.2.1, Cambridge University Press (2004).
6) Culotta, A., Kristjansson, T., McCallum, A. and Viola, P.: Corrective Feedback and Persistent Learning for Information Extraction, *Artificial Intelligence Journal*, Vol.170, pp.1101–1122 (2006).
7) Jiang, J. and Zhai, C.: Instance Weighting for Domain Adaptation in NLP, *Proc. Annual Meeting of the Association of Computational Linguistics*, Association for Computational Linguistics, pp.264–271 (2007).
8) Jiao, F., Wang, S., Lee, C.-H., Greiner, R. and Schuurmans, D.: Semi-Supervised Conditional Random Fields for Improved Sequence Segmentation and Labeling, *Proc. Annual Meeting of the Association of Computational Linguistics*, pp.209–216 (2006).
9) Keene, D.，羽鳥博愛，伊良部祥子，山田晴子（編）：会話作文英語表現辞典，朝日出版社 (1992).
10) Kudo, T., Yamamoto, K. and Matsumoto, Y.: Applying Conditional Random Fields to Japanese Morphological Analysis, *Proc. Empirical Methods in Natural Language Processing* (2004).
11) Lafferty, J., McCallum, A. and Pereira, F.: Conditional Random Fields: Probabilistic models for segmenting and labeling sequence data, *Proc. 18th International Conference on Machine Learning* (2001).
12) Mann, G.S. and McCallum, A.: Efficient Computation of Entropy Gradient for Semi-Supervised Conditional Random Fields, *Proc. Human Language Technologies Conference — Annual Conference of the North American Chapter of the Association for Computational Linguistics* (*NAACL-HLT*), pp.109–112 (2007).
13) Mori, S.: Language Model Adaptation with a Word List and a Raw Corpus, *Proc. 9th International Conference on Spoken Language Processing* (2006).
14) Nagata, M.: A part of speech estimation method for Japanese unknown words using a statistical model of morphology and context, *Proc. Annual Meeting of the Association of Computational Linguistics*, pp.277–284 (1999).
15) Peng, F., Feng, F. and McCallum, A.: Chinese Segmentation and New Word Detection using Conditional Random Fields, *Proc. International Conference on Computational Linguistics* (2004).
16) Pereira, F.C.N. and Schabes, Y.: Inside-Outside Reestimation from Partially Bracketed Corpora, *Proc. Annual Meeting of the Association of Computational Linguistics*, pp.128–135 (1992).
17) Rätsch, G. and Smola, A.J.: Adapting Codes and Embeddings for Polychotomies, *Advances in Neural Information Processing Systems* (2002).
18) Roth, D. and Small, K.: Margin-based Active Learning for Structured Output Spaces, *Proc. European Conference on Machine Learning*, pp.413–424, Springer (2006).
19) Sarawagi, S. and Cohen, W.W.: Semi-Markov Conditional Random Fields for Information Extraction, *Advances in Neural Information Processing Systems* (2005).
20) Scheffer, T. and Wrobel, S.: Active learning of partially hidden Markov models, *Proc. ECML/PKDD Workshop on Instance Selection* (2001).
21) Sha, F. and Pereira, F.: Shallow Parsing with Conditional Random Fields, *Proc. Human Language Technology — NAACL*, Edmonton, Canada (2003).
22) Smola, A.J., Vishwanathan, S. and Hofmann, T.: Kernel methods for missing variables, *Proc. 10th International Workshop on Artificial Intelligence and Statistics* (2005).
23) Sutton, C. and McCallum, A.: An Introduction to Conditional Random Fields for Relational Learning, *Introduction to Statistical Relational Learning*, Getoor, L. and Taskar, B. (Eds.), MIT Press (2006).
24) Sutton, C., Rohanimanesh, K. and McCallum, A.: Dynamic Conditional Random Fields: Factorized Probabilistic Models for Labeling and Segmenting Sequence Data, *Proc. International Conference on Machine Learning* (2004).
25) Uchimoto, K., Sekine, S. and Isahara, H.: The unknown word problem: A morphological analysis of Japanese using maximum entropy aided by a dictionary, *Proc. Conference on Empirical Methods in Natural Language Processing*, pp.91–99 (2001).
26) Verbeek, J. and Triggs, B.: Scene Segmentation with CRFs Learned from Partially Labeled Images, *Advances in Neural Information Processing Systems* (2007).

© 2009 Information Processing Society of Japan

付　　録

A.1　目的関数と偏微分の効率的計算法

マルコフ性を仮定することで，対数尤度（式 (4)）とその偏微分（式 (5)）を動的計画法により効率的に計算する方法を以下に示す．なお，以下の説明では 1 次のマルコフモデルを説明するが，2 次以上のマルコフモデルやセミマルコフモデル[19] にも容易に拡張可能である．

1 次のマルコフモデルでは，$(\boldsymbol{x}, \boldsymbol{y})$ の各点 t において，入力変数とラベル変数の組の素性 $\boldsymbol{f}(x_t, y_t) : X \times Y$ と隣接するラベル変数の組の素性 $\boldsymbol{g}(y_{t-1}, y_t) : Y \times Y$ [⋆1]を考え，その線形結合を $\boldsymbol{\phi}(x_t, y_{t-1}, y_t) = \boldsymbol{f}(x_t, y_t) + \boldsymbol{g}(y_{t-1}, y_t)$ と書く．すると，素性ベクトルは $\boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y}) = \sum_{t=1}^{T+1} \boldsymbol{\phi}(x_t, y_{t-1}, y_t)$ と分解できる．ただし，ラベル列の先頭と末尾を表す特別なラベルをそれぞれ S と E とするとき，$\boldsymbol{\phi}(x_t, y_{t-1}, y_t)$ は先頭 $t = 1$ で $\boldsymbol{\phi}(x_t, S, y_t)$，末尾 $t = T + 1$ で $\boldsymbol{g}(y_{t-1}, E)$ と定義する．

式 (4)，(5) の第 1，2 項を効率的に計算するポイントは，ラベル列ごとの再計算を避けるために，ある $\boldsymbol{\theta}$，$\boldsymbol{x}$，$\boldsymbol{L}$ に対して行列 $\alpha_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}[t, j]$，$\beta_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}[t, j]$ をあらかじめ計算することである．なお，$\boldsymbol{Y}$ の和の計算時は $\boldsymbol{L} = (Y, \cdots, Y)$ とすればよい．これはよく知られた *Forward-Backward* アルゴリズムを制約 $\boldsymbol{L}$ を満たすように拡張したものであり，α は $t = 1, 2, \cdots, T$，β は $t = T + 1, \cdots, 2, 1$ の順で以下のように計算する．

$$\alpha_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}[t, j] = \begin{cases} 0 & \text{if } j \notin L_t \\ \boldsymbol{\theta} \cdot \boldsymbol{\phi}(x_t, S, j) & \text{else if } t = 1 \\ \ln \sum_{i \in L_{t-1}} e^{\alpha[t-1,i] + \boldsymbol{\theta} \cdot \boldsymbol{\phi}(x_t, i, j)} & \text{else} \end{cases}$$

$$\beta_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}[t, j] = \begin{cases} 0 & \text{if } j \notin L_t \\ \boldsymbol{\theta} \cdot \boldsymbol{g}(j, E) & \text{else if } t = T + 1 \\ \ln \sum_{k \in L_{t+1}} e^{\boldsymbol{\theta} \cdot \boldsymbol{\phi}(x_t, j, k) + \beta[t+1,k]} & \text{else} \end{cases}$$

以降，α，β，Z の添え字 $\boldsymbol{\theta}$，$\boldsymbol{x}$，$\boldsymbol{L}$ は誤解のない範囲で省略する．行列 $\alpha[t, j]$（$\beta[t, j]$）は各点 s でラベル y_s が L_s に含まれる点 t までの前（後）からの部分ラベル列の中で，$y_t = j$ であるすべての部分ラベル列の指数の和の対数値[⋆2]を格納している．α，β の計算量は $O(T|Y|^2)$ である．

最後に，式 (4)，(5) を α，β を使って計算する方法を示す．式 (4) の Z の対数は α，β を使って次式で得られる：

$$\ln Z_{\boldsymbol{\theta},\boldsymbol{Y_L}} = \ln \sum_{j \in L_T} e^{\alpha_{\boldsymbol{\theta},\boldsymbol{L}}[T,j] + \boldsymbol{\theta} \cdot \boldsymbol{g}(j, E)},$$

また，式 (5) の第 1，2 項は同様に次式で計算できる．

$$\sum_{\boldsymbol{y} \in \boldsymbol{Y_L}} P_{\boldsymbol{\theta},\boldsymbol{L}}(\boldsymbol{y}|\boldsymbol{x}) \boldsymbol{\Phi}(\boldsymbol{x}, \boldsymbol{y})$$
$$= \sum_{t=1}^{T} \sum_{j \in L_t} \left(\gamma_{\boldsymbol{L}}(t, j) \boldsymbol{f}(x_t, j) + \sum_{i \in L_{t-1}} \varepsilon_{\boldsymbol{L}}(t, i, j) \boldsymbol{g}(i, j) \right) + \sum_{i \in L_T} \varepsilon_{\boldsymbol{L}}(T, i, E) \boldsymbol{g}(i, E)$$

ただし，$\gamma_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}$ および $\varepsilon_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}$ は次式の周辺確率である（上式で下付きの $\boldsymbol{\theta}$，$\boldsymbol{x}$ は省略した）．

$$\gamma_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}(t, j) = P_{\boldsymbol{\theta},\boldsymbol{L}}(y_t = j|\boldsymbol{x}) = e^{\alpha[t,j] + \beta[t,j] - \ln Z_{\boldsymbol{Y_L}}}$$

$$\varepsilon_{\boldsymbol{\theta},\boldsymbol{x},\boldsymbol{L}}(t, i, j) = P_{\boldsymbol{\theta},\boldsymbol{L}}(y_{t-1} = i, y_t = j|\boldsymbol{x}) = e^{\alpha[t-1,i] + \boldsymbol{\theta} \cdot \boldsymbol{\phi}(x_t, i, j) + \beta[t,j] - \ln Z_{\boldsymbol{Y_L}}}$$

以上の方法で，式 (4)，(5) は $O(T|Y|^2)$ で計算できる．

(平成 20 年 4 月 21 日受付)
(平成 21 年 3 月 6 日採録)

坪井　祐太（正会員）

2002 年奈良先端科学技術大学院大学情報科学研究科博士前期課程修了．同年日本アイ・ビー・エム（株）入社．同社基礎研究所にてテキストマイニングの研究開発に従事．2009 年奈良先端科学技術大学院大学情報科学研究科博士後期課程修了．工学博士．

⋆1 x_t を含めて $\boldsymbol{g}(x_t, y_{t-1}, y_t) : X \times Y \times Y$ としても以下の議論には影響しない．しかし，ラベル数の 2 乗に比例して $\boldsymbol{g}$ の素性数（$= |X||Y|^2$）が増加する．そのため，実際には x_t を含めて $\boldsymbol{g}$ を設計しないことが多く，表記を簡略にするために x_t は $\boldsymbol{g}$ の引数から省略した．

⋆2 数値計算時にはオーバフローを防ぐための工夫が必要になる．詳細は文献 23) の 1.4.6 項を参照．

© 2009 Information Processing Society of Japan

森　　信介（正会員）

1998 年京都大学大学院工学研究科電子通信工学専攻博士後期課程修了．同年日本アイ・ビー・エム（株）入社．2007 年 5 月より京都大学学術情報メディアセンター准教授．工学博士．1997 年本学会山下記念研究賞受賞．言語処理学会会員．

鹿島　久嗣（正会員）

1999 年京都大学大学院工学研究科応用システム科学専攻修士課程修了．同年日本アイ・ビー・エム（株）入社．同社基礎研究所にて，機械学習，データマイニング手法の開発と，バイオインフォマティクス，オートノミックコンピューティング，ビジネスインテリジェンス等への応用に従事．2007 年京都大学情報学研究科知能情報学専攻博士後期課程修了．情報学博士．

小田　裕樹（正会員）

1999 年徳島大学大学院工学研究科博士前期課程知能情報工学専攻修了．同年 NTT ソフトウェア（株）入社．言語処理・情報検索システム等の開発，コンサルティング業務に従事．確率・統計的自然言語処理およびその応用に興味を持つ．工学博士．言語処理学会会員．

松本　裕治（正会員）

1979 年京都大学大学院工学研究科修士課程情報工学専攻修了．同年電子技術総合研究所入所．1984〜1985 年英国インペリアルカレッジ客員研究員．1985〜1987 年（財）新世代コンピュータ技術開発機構に出向．京都大学助教授を経て，1993 年より奈良先端科学技術大学院大学教授．工学博士．専門は自然言語処理．人工知能学会，日本ソフトウェア科学会，言語処理学会，認知科学会，AAAI，ACL，ACM 各会員．

© 2009 Information Processing Society of Japan