

漸進的な結合モデルによる中国語の形態素・依存構造解析

羽鳥潤

松崎拓也

宮尾祐介

辻井潤一

東京大学大学院情報理工学系研究科
 {hatori,matuzaki}@is.s.u-tokyo.ac.jp

国立情報学研究所
 yusuke@nii.ac.jp

Microsoft Research Asia
 jtsujii@microsoft.com

1 緒論

単語間の区切りを持たない日本語・中国語等の言語において、単語分割・品詞解析はあらゆる自然言語処理タスクの礎となる非常に重要な問題である。これまでに、これら 2 つのタスクには強い相互作用がある事が知られており、日本語における Kudo et al. (2004)、中国語における Zhang and Clark (2010) 等多くの結合モデルが提案されてきた。例えば中国語に於いては、両タスクに対する結合モデルは各タスクの精度をそれぞれ 1% 以上向上させる事が報告されている (Zhang and Clark, 2008)。また、最近我々は、中国語の品詞・依存構造解析に対する漸進的結合モデルを提案し、これら 2 つのタスクの精度も結合モデルによって大きく向上する事を示した (Hatori et al., 2011)。しかし、このモデルは単語分割の問題を切り離して考えていたため、単語分割・品詞解析の間に大きな相互作用がある事を考えると、現実的な枠組みとは言えなかった。

そこで、本稿では単語分割・品詞・依存構造解析の 3 タスクに対する結合モデルを構築し、タスク間の大域的な相互作用を捉える事を目指す。単純な方法では計算量の爆発が不可避である事から、動的計画法を用いた漸進的構文解析手法 (Huang and Sagae, 2010) をベースとして、効率的な文字単位のデコーディング手法を提案する。このような結合モデルを構築する事で、下位タスクからのエラー伝播の軽減によって上位タスクの精度向上が期待されるだけでなく、長距離の統合的な情報の利用により単語分割・品詞解析の精度向上が見込まれる。特に統語的な情報は、外来語等の未知語の解析に有効であると考えられ、既存の単語分割システムにおけるエラーの大きな部分を占めていた未知語の処理精度が改善される事が期待される。

このような手法を構築する際には、大きく分けて 2 つの難しさがある。まず、ビーム探索をどのようにして効果的に行うかという問題である。単語分割と依存構造解析が同時に行われる際には、解析途中の状態が持つ依存弧の数は単語がどのように分割されたかに依存する為、終端状態に辿り付く迄のステップ数は一定にならず、単純な手法では状態間の同期をうまく取る事ができない。本稿では、どのように単語分割が行われても状態間の同期を適切に取る事ができる手法を提案する。また、素性セット自体は既存の手法 (Zhang and Clark, 2010; Huang and Sagae, 2010) を組み合わせただけのものになるが、これらを単一の漸進的枠組みの中に落とし込む事は容易ではない。単語分割・品詞・依存構造解析に関連した 3 種類の性質の異なった決定を扱う為に、各素性がいつ、どのような粒度の情報と組み合わせで用いられるかを細かく調整しなければならないほか、学習時に於いても、素性の種類に応じた学習率の調整が必要になった。このような点に注意してモデルを構築したところ、中国語ツリーバンク (CTB) を用いた実験に於いて、結合モデルの精度はパイプラインモデルと比べて各タスクの精度を有意に向上させる事が示された。

2 関連研究

中国語における単語分割・品詞解析の結合モデルは多くの研究例があるが、特に線形時間解析が可能な漸進的モデルを用いた Zhang and Clark (2010) は本研究の結合モ

デルを構築する上での大きな基礎となっている。彼らは、単語分割・品詞解析結合タスクに対して高速な文字単位のデコーディング手法により、従来手法よりも 10 倍程度高速でしかも遜色のない精度を持つモデルを提案した。また、漸進的な枠組みによる依存構造解析モデルという点では、中国語を対象とした品詞・依存構造解析結合モデル (Hatori et al., 2011) が直接の土台となっている。このモデルは、Huang and Sagae (2010) によって提案された動的計画法を用いた Shift-Reduce 法を拡張して、品詞解析を同時に行う事を可能にしている。

3 手法

3.1 漸進的な単語分割・品詞・依存構造結合モデル

本章では、Hatori et al. (2011) による品詞・依存構造解析モデルをベースとして、単語分割・品詞・依存構造解析タスクに対する結合モデルを提案する。Zhang and Clark (2010) を参考にして、新たな操作として APPEND 操作を加えると共に、SHIFT(t) 操作が文字単位の操作になっている。実際に用いられる操作は以下の通り：

- A (APPEND): キューの先頭文字をスタック最上位の単語に追加する。
- SH(t) (SHIFT(t)): キューの先頭文字を品詞タグ t を持つ新たな単語としてスタック最上位に移動する。
- RL/RR (REDUCE-LEFT/RIGHT): スタック最上位にある 2 つの部分木 (s_1, s_0) を、木 $s_1 \wedge s_0$ / $s_1 \vee s_0$ に統合する。

CTB では 33 個の品詞タグが定義されている事から、提案手法は合計で 36 個の操作を考慮する事になるが、実際には訓練データ中に出現した品詞タグ・限られた数の単語と関連づけられた品詞タグの情報等を用いる事で、多くの解析候補を削る事が出来る。訓練には平均化パーセプトロンを早期更新と共に用いた。

3.2 状態の整列

通常、漸進的な手法においてビーム探索を用いる際には、各状態はそれぞれに一意に割り振られたステップ数に従って整列され、同じステップにある状態に対して枝刈りが行われる。そのため、ビーム探索が効果的に働くためには、同じステップ数を持つ状態は全て比較可能であり、また、全ての終端状態が同じステップ数に属している事が望ましい。しかし、単語分割と依存構造解析を同時に行う時には、このような状態の整列を行う事が単純ではない。

まず、単語分割・品詞解析結合モデル (Zhang and Clark, 2010) が行っているように、SHIFT された文字の数で状態を揃える方法が考えられるが、この場合には、RL/RR 操作をステップ数を増やさずに行える事から、デコーダは依存弧が多い状態を優先的にビームに残してしまう傾向があり、RL/RR が早めに選ばれるようにバイアスがかかってしまう。また、代わりに品詞・依存構造解析結合モデル (Hatori et al., 2011) が行っているように適用された操作の数で状態を揃える事も可能であるが、この場合には単語分割が細かく行われるほど終端状態に辿り付くまでに必要な操作数が多くなってしまふ。従って、どちらの方法を用いてもビーム探索を効果的に行う事が出来ず、パイプライ

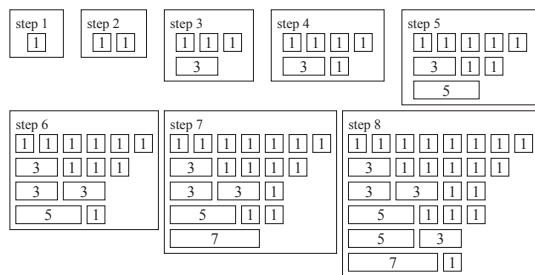


図 1: 状態間の整列の例

ン手法に精度の面で及ばない事が分かった¹。

このような問題を解決するために、我々は依存弧 (dependency arc) の数に基づいて状態を整列させる手法を提案する。まず、単語間の依存弧に加えて、長さ M の単語は $M-1$ 個の文字間依存弧を持っていると仮定する。例えば、一つの単語を \square で表すことにすると、3 文字からなる単語は $\square A \square B \square C \square$ のように 2 つの文字間依存弧を持つ。このように定義された文字間依存弧を用いて、shift された文字数と (単語間・文字間) 依存弧の総数を用いて状態を整列すると、以下の全ての性質を満たす事が出来る。

- (1) M 個の連続する文字に対する部分木は全て同じステップ数 ($2M-1$) を持つ。
- (2) 全ての終端状態のステップ数は $2N$ で等しい (N は文長、根への依存弧を含む)。
- (3) 全ての操作はステップ数を 1 以上進める。

このように定義された指標を用いると、各状態は図 1 に示されるように整列される事になる。この方法を用いると、ステップ数を 1 増やす操作 (SH(t) と RL/RR) と 2 増やす操作 (for A) が混在するが、2 つのビームを用いる事でデコーディングが可能である。従って、1 文を処理するのに必要な計算量は、 $O(B(T+3) \cdot 2N)$ (B はビーム幅、 T は品詞数、 N は文長である) となり、単語分割・品詞解析結合モデル (Zhang and Clark, 2010) と同じオーダーを持つ事が分かる。

3.3 素性

素性セットとしては、実験の項でベースライン手法として用いる最先端の単語分割・品詞解析結合モデル (Zhang and Clark, 2010) からの素性と、依存構造解析器 (Huang and Sagae, 2010) からの素性を組み合わせて利用した。しかし、前述の通り、単一の漸進的モデルで 3 種類の決定 (単語分割・品詞・統語構造解析) を行う為、各素性をどのようなタイミングで発火させ、どのような粒度の情報と組み合わせるかを調整する事が、効果的な結合解析を行う為に不可欠となる。紙幅の関係で詳細な記述は省くが、各素性をどのように組み込むかについて、概略を示す。

まず、Zhang and Clark (2010) からの素性のうち、部分文字列に対するスコアを表す素性は、A と SH(t) にのみ用いられるが、新たに割り当てられる品詞情報を含む素性は、SH(t) にのみ用いられる。また、SH(t) 操作によって境界が確定する単語の情報と関連付けられていた素性は、本手法では SH(t)・RL・RR の各操作に対して用いられるが、RL・RR 操作に対してはスタック最上位の単語の境界を初めて確定させるような状況に対してのみ適用される。これは、2 回以上連続して RL/RR 操作が適用された際に、2 回目以降の操作には単語の区切りを確定させるような効果がない為である。更に、このような状況に於いて、単語境界の確定を表す一般化された操作ラベル (SH') を素性と組み

¹例えば、適用された操作数で状態を描いた場合には、パイプライン手法と比べて単語分割精度が 0.2-0.3% 程度低かった。

	訓練			開発			テスト		
	文	語		文	語	未知	文	語	未知
CTB-5d	16k	438k	804	21k	1.2k	1.9k	50k	3.1k	
CTB-5j	18k	494k	352	6.8k	553	348	8.0k	278	
CTB-5c	15k	423k	-	-	-	-	-	-	-
CTB-6	23k	641k	2.1k	60k	3.3k	2.8k	82k	4.6k	
CTB-7	31k	718k	10k	237k	13k	10k	245k	13k	

表 1: データセットの統計情報 (文数・単語数)

合わせて用いる事で、これらの操作に対する共通の振る舞いを学習できるようにした。

次に、Huang and Sagae (2010) からの素性は、ほぼそのまま用いられる。しかし、キュー内の単語の語形や品詞タグと関連付けられた素性 (先読み素性) はそのままでは漸進的なモデルに組み込むことが出来ないため、Hatori et al. (2011) によって提案された遅延素性を用いて組み込む必要がある。我々は、これによってモデル全体の精度が更に向上する事を確認したが、遅延素性の利用は計算時間の大きな増加を伴う為、本稿の実験では用いていない。

3.3.1 辞書素性

中国語の単語分割においては辞書に含まれている単語の情報が強力な情報として働く事が知られており、辞書情報を用いる事で統語的な情報の寄与をより現実的に即した状況で調べられると考えられる。そこで、辞書素性として、境界を確定しようとしている単語の語形が辞書に含まれるかどうかを表す二値素性、及びそれと品詞タグとの組み合わせを辞書素性として定義した。これらの素性はベースライン手法に於いても同時に用いられた。

3.4 素性重みの学習率の調整

予備実験の結果、3 タスクに対する結合モデルを構築する際には、素性の種類に応じて重みの学習率を調整する必要がある事が分かった。学習率の比による寄与を調べるため、まず、状態 ψ に対して操作 a が適用された場合の素性ベクトル $\vec{\phi}$ 及びスコア Φ を以下のように定義する。

$$\Phi(\psi, a) = \vec{\lambda} \cdot \vec{\phi}(\psi, a) = \vec{\lambda} \cdot \left\{ \vec{\phi}_{st}(\psi, a) + \sigma_p \vec{\phi}_p(\psi, a) \right\},$$

ここで $\vec{\phi}_{st}$ が Zhang and Clark (2010) からの素性に、 $\vec{\phi}_p$ が Huang and Sagae (2010) からの素性に対応している。素性ベクトルを 2 つの要素に分け両者の比を調整する事で、例えば σ_p が 1 より小さな値に設定された場合には、構文解析素性に対する重みの更新幅は学習の初期段階において小さく保たれる。このようにして、統語素性の学習率を相対的に制御する事が可能になる。

4 実験

4.1 実験設定

評価用データとして、Chinese Penn Treebank ver. 5.1・6.0・7.0 を用いた (以下 CTB-5・CTB-6・CTB-7)。それぞれのデータは、先行研究と同一のデータ分割に従って訓練・開発・テスト用に分けた。CTB-5 については、Huang and Sagae (2010) と同一の分割を行ったものを CTB-5d、Zhang and Clark (2010) と同一の分割を行ったものを CTB-5j と表記する。また、CTB-5 から CTB-5d と CTB-5j の開発・テストデータとの共通部分を除外した上で、交差テスト用に 5 等分したものを CTB-5c-1 ~ CTB-5c-5 と表記する。CTB-6 は付属ドキュメントに従って分割し、CTB-7 は Wang et al. (2011) と同一の分割を行った。各データのサイズを表 1 に示す。辞書として、HowNet Chinese Word List² (91,015 語) および中国語

²http://www.keenage.com/html/e_index.html

Wikipedia³ (2011-10-26 版、709,352 語) を用いた。これらの辞書は表層語形のみを含み、頻度や品詞などの情報は含まない。

単語分割および品詞解析の評価尺度としては、通例どおり単語単位の適合率、再現率および F1 スコアを用いた。依存構造解析の評価の際は、係り元および係り先の 2 つの単語がともに正しく分割されているときのみ正しい依存構造関係として数え、ラベルなしの係り先識別精度 (句読点は除く) を評価値として用いた。統計的有意性の検定には McNemar 検定を用いた ($\dagger: p < 0.05, \ddagger: p < 0.01$)。

4.2 ベースライン手法・提案手法

実験では、以下のベースライン・提案手法を比較した。

- **SegTag**: Zhang and Clark (2010) の単語分割・品詞解析結合モデルの再実装。ビーム幅として、元論文で最高精度を示した 16 を用いた。
- **Dep'**: Huang and Sagae (2010) の依存構造解析器の再実装。Hatori et al. (2011) で用いられている実装と等しい。
- **Dep**: Dep' から先読み素性を除いたもの。
- **TagDep**: Hatori et al. (2011) の品詞・依存構造解析結合モデルから、先読み素性を除いたもの⁴。
- **SegTag+Dep/SegTag+Dep'**: SegTag と Dep (あるいは Dep') のパイプラインによる連結。
- **SegTag+TagDep**: SegTag と TagDep のパイプラインによる連結。ただし、SegTag の出力のうち、単語分割の部分だけを TagDep の入力として用い、品詞解析の評価には TagDep からの出力を用いる。
- **SegTagDep**: 単語分割・品詞・依存構造解析の 3 タスク結合モデル (提案手法)。

上記全てのモデルにおいて、単語分割と品詞解析の素性 (Zhang and Clark, 2008; Zhang and Clark, 2010) 及び依存構造解析部分の素性 (Huang and Sagae, 2010) は同一のものを用いている。よって、結果の差異は純粋に解析モデルの違いによるものであり、手法間の比較によって結合モデルによる改善を評価する事ができる。

4.3 開発データによる分析

各モデルには、依存構造解析の素性の重み σ_p ・ビーム幅・訓練の際の繰り返し回数など、いくつかの調整可能なパラメータがある。これら全てのパラメータは CTB-5c による交差検定を用いて決定したが、CTB-5j・CTB-6・CTB-7 を用いた実験では、訓練の繰り返し回数を開発データを用いて決定した。

図 2 に、CTB-5c-1 に対する SegTagDep の精度と訓練繰り返し回数との関係を、構文解析素性の重みを変化させながら調べた結果を示す。図中、“Seg”・“Tag”・“Dep” はそれぞれ単語分割・品詞解析・依存構造解析の F1 スコアである。この実験では辞書は用いず、ビーム幅は 32 とした。この結果で興味深い点は、 σ_p を単純に 1 とした場合、解析精度が低い値に収束している事である。また、 σ_p を 0.2 とした場合、単語分割と品詞解析の精度は最適な設定である $\sigma_p = 0.5$ の場合の精度とほぼ等しくなるが、収束はより遅くなっている。これらの結果から、以下の実験では $\sigma_p = 0.5$ を用いた。ビーム幅としては、図 3 から分かるように、各タスクの精度が飽和に近づく 64 を用いた。

4.4 結果

この節では、提案手法とベースライン手法を比較した実験結果を示す。表 2-表 5 に、CTB-5c に対する単語分割・品詞解析・依存構造解析の F1 スコアを示す。括弧中

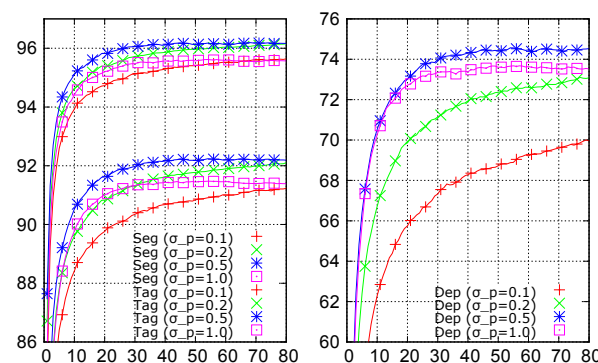


図 2: 依存構造解析の素性の重みごとの、訓練繰り返し回数と CTB-5c-1 に対する SegTagDep の F1 スコア (%) の関係

の数値は SegTag+Dep に対する精度の差である。辞書の使用の有無に依らず、SegTagDep による結果では品詞解析及び依存構造解析の精度が大きく改善している (品詞解析で 0.56–0.63%, 依存構造解析で 2.34–2.44% の改善)。SegTagDep による依存構造解析精度の改善は、先読み素性を用いたモデルである SegTag+Dep' と比較した場合でも有意であった。しかし、辞書を用いない場合、単語分割精度には有意な改善が見られなかった。これと対照的に、辞書を用いた場合、単語分割精度はベースラインである SegTag に比べ一貫して改善が見られた。全単語に対する改善は 0.1% 程度に留まるものの、未知語⁵のみに対して分割精度を評価した場合に 1% 以上の改善が見られた。辞書の使用による結果の違いは、結合モデルに於いては、単語分割部分の能力が十分でない場合、統語的な情報は単語分割に対してむしろノイズとして作用するが、基本的な単語分割部分の能力が十分である場合は、統語的な情報が単語分割に対して有効に働く事を示唆している。

部分的な結合モデルである SegTag+TagDep は、依存構造解析に関しては程々に良い精度を示している。辞書を用いた場合、SegTag+TagDep は SegTag+Dep に比べて依存構造解析精度が 2.02% 改善しており、これは SegTagDep と比較して 0.32% 低いだけである。しかし、SegTag と比べ、SegTag+TagDep の品詞解析精度にはほぼ改善が見られないのに対し、SegTagDep の品詞解析精度は辞書有/無それぞれの場合に 0.63%/0.56% と一貫して改善している。この品詞解析精度に関する差は、結合モデルによるエラー伝播の軽減と、統語的情報の品詞解析への寄与によるものと考えられる。更に、未知語に対する SegTag+TagDep の品詞解析精度は SegTag に比べて一貫して悪くなっており、このモデルにおいては、統語的情報は未知語に対する品詞予測にむしろ悪影響を与えているように見える⁶。これに対し、SegTagDep ではこのような統語的情報による品詞解析精度への悪影響は見られず、総合評価・未知語のみの評価の両方に於いて品詞解析精度が改善している。これは 3 タスクを結合した提案手法の有効性を示す結果と言える。

最後に、図 3 に各モデルに対し解析精度と処理時間の関係を調べた結果を示す。SegTagDep は SegTag+Dep/TagDep と同等の精度を達成するために数倍の処理時間を必要とするものの、処理時間を度外視すれば更に精度の改善の余地がある事が分かる。

⁵ここでいう未知語は、訓練データ及び辞書に現れない単語

⁶同様の観察として、Hatori et al. (2011) でも、品詞解析と依存構造解析の結合モデルでは、統語的な役割が変わるようなタグ付け誤りが減少する一方、一般名詞と固有名詞のような統語的に類似のタグ間では誤りがやや増える事が報告されている。

³<http://zh.wikipedia.org/wiki>

⁴先読み素性を除いたことによる精度低下は CTB-5d の開発データで 0.26%。

	SegTag		SegTagDep	
	ALL	OOV	ALL	OOV
辞書無	96.22	72.24	96.19±0.07(-0.02 ±0.08)	72.24±0.42(0.00 ±0.48)
辞書有	96.82	78.32	96.90±0.08(+0.08[‡]±0.04)	79.38±0.75(+1.06[‡]±0.35)

表 2: CTB-5c に対する単語分割 F1 スコア

	SegTag +Dep	SegTag +Dep [‡]	SegTag+TagDep	SegTagDep
辞書無	72.58	72.94	74.60±0.22(+2.02 [‡] ±0.10)	74.92±0.12(+2.34[‡]±0.31)
辞書有	73.53	73.90	75.45±0.24(+1.92 [‡] ±0.06)	75.97±0.14(+2.44[‡]±0.26)

表 3: CTB-5c に対する依存構造解析 F1 スコア

システム	Seg	Tag
Kruengkrai '09	97.87	93.67
Zhang '10	97.78	93.67
Sun '11	98.17	94.02
Wang '11	98.11	94.18
SegTag	97.66	93.61
SegTagDep	97.73	94.46
SegTag(d)	98.18	94.08
SegTagDep(d)	98.26	94.64

表 4: CTB-5j に対する最終結果

	SegTag		SegTag+TagDep		SegTagDep	
	ALL	OOV	ALL	OOV	ALL	OOV
辞書無	91.74	59.82	91.86±0.14(+0.12 [‡] ±0.09)	58.89±1.22(-0.94 [‡] ±0.55)	92.30±0.08(+0.56[‡]±0.06)	61.03±0.80(+1.20[‡]±0.41)
辞書有	92.34	65.44	92.35±0.16(+0.01 ±0.07)	63.20±1.52(-2.24 [‡] ±0.68)	92.97±0.16(+0.63[‡]±0.10)	67.40±1.11(+1.96[‡]±0.23)

表 5: CTB-5c に対する品詞解析 F1 スコア (括弧内は SegTag に対する精度の差)

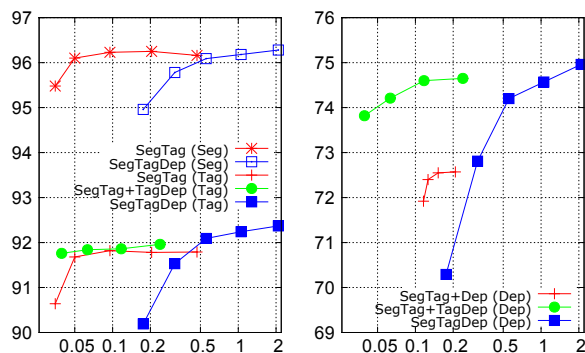


図 3: CTB-5c-1 に対する処理時間 (秒/文) と解析精度の関係 (各点はビーム幅 4・8・16・32・64) に対応。SegTag+Dep 及び SegTag+TagDep の一部として用いられる SegTag にはビーム幅 16 を用いた。)

4.5 先行研究との比較

表 4 及び表 6 に単語分割及び品詞解析に関する先行研究との比較結果を示す。“Kruengkrai+ '09”は Kruengkrai et al. (2009) によるラティスベースの手法、“Zhang '10”は Zhang and Clark (2010) による漸進的解析手法であり、これら 2 つのシステムでは CTB コーパス以外の辞書等の資源は用いられていない。“Sun+ '11”は Sun (2011) による CRF をベースとした、複数のモデルの組み合わせを用いた手法で、イデオムに関する外部辞書を用いている。“Wang+ '11”は Wang et al. (2011) による半教師付き学習による手法で、CTB に加え Chinese Gigaword Corpus をデータとして用いている。

提案法及び部分的な結合モデルを辞書と共に用いた場合の精度 (‘(d)’ マーク) は他の先行研究とほぼ同等であり、特に SegTagDep は辞書以外の追加資源を用いず、単語分割と品詞解析に対し最高精度を示している。特に、品詞解析精度については、統語情報の利用により、これまでの最高精度に比べ 0.4% 以上の改善を達成している。

5 結論

本稿では、中国語の単語分割・品詞・依存構造解析の 3 タスクに対する初の結合モデルを提案した。漸進的な解析モデルの利用と効率的な文字レベルのデコーディングにより現実的な計算時間で 3 タスクを統合する事に成功し、最先端の手法を組み合わせたパイプライン手法と比べて全てのタスクで有意な精度向上を記録した。特に、品詞・依存構造解析の精度は大きく向上し、それぞれ 0.6%・2.4% の

モデル	CTB-6 テスト			CTB-7 テスト		
	Seg	Tag	Dep	Seg	Tag	Dep
Kruengkrai '09	95.50	90.50	-	95.40	89.86	-
Wang '11	95.79	91.12	-	95.65	90.46	-
SegTag+Dep	95.46	90.64	72.57	95.49	90.11	71.25
SegTagDep	95.45	91.27	74.88	95.42	90.62	73.58
(差)	-0.01	+0.63 [‡]	+2.31 [‡]	-0.07	+0.51 [‡]	+2.33 [‡]
SegTag+Dep(d)	96.13	91.38	73.62	95.98	90.68	72.06
SegTagDep(d)	96.18	91.95	75.76	96.07	91.28	74.58
(差)	+0.05	+0.57 [‡]	+2.14 [‡]	+0.09 [‡]	+0.60 [‡]	+2.52 [‡]

表 6: CTB-6 及び CTB-7 に対する最終結果

精度向上を記録している。単語分割に対しては、全体の精度向上幅は 0.1% 程度に留まったが、未知語のみに対する評価では 1% 以上の精度向上を記録している。

今後は、確率的なモデルによる探索の枝刈りを行う事で、更にデコーディングを高速に行う事が必要であると考えられる。特に、ビーム幅を大きくした際には更なる精度向上が見込まれる事から、より効率的なデコーディングを開発する事により統語的な依存関係の寄与をより明確に示せるのではないかと期待される。

参考文献

- Jun Hatori, Takuya Matsuzaki, Yusuke Miyao, and Jun'ichi Tsujii. 2011. Incremental joint POS tagging and dependency parsing in Chinese. In *IJCNLP-2011*.
- Liang Huang and Kenji Sagae. 2010. Dynamic programming for linear-time incremental parsing. In *ACL-2010*.
- Canasai Kruengkrai, Kiyotaka Uchimoto, Jun'ichi Kazama, Yiu Wang, Kentaro Torisawa, and Hitoshi Isahara. 2009. An error-driven word-character hybrid model for joint Chinese word segmentation and POS tagging. In *ACL-2009*.
- Taku Kudo, Kaoru Yamamoto, and Yuji Matsumoto. 2004. Applying conditional random fields to Japanese morphological analysis. In *EMNLP-2004*.
- Weiwei Sun. 2011. A stacked sub-word model for joint Chinese word segmentation and part-of-speech tagging. In *ACL-HLT-2011*.
- Yiu Wang, Jun'ichi Kazama, Yoshimasa Tsuruoka, Wenliang Chen, Yujie Zhang, and Kentaro Torisawa. 2011. Improving Chinese word segmentation and POS tagging with semi-supervised methods using large auto-analyzed data. In *IJCNLP-2011*.
- Yue Zhang and Stephen Clark. 2008. Joint word segmentation and POS tagging using a single perceptron. In *ACL-HLT-2008*.
- Yue Zhang and Stephen Clark. 2010. A fast decoder for joint word segmentation and POS-tagging using a single discriminative model. In *EMNLP-2010*.