

隠れマルコフ木モデルを用いた単語アラインメント

近藤 修平^{1,a)} Kevin Duh^{1,b)} 松本 裕治^{1,c)}

概要：本稿では隠れマルコフ木モデルに基づく教師無し単語アラインメントモデルを提案する．提案法はアラインメント変数が目的言語側の依存構造木と同形の木構造を持つと仮定し，歪み確率を原言語側の依存構造木に基づいてモデルすることで，両言語側の統語構造を考慮する．日英データ上の単語アラインメントにおいて，提案法は IBM モデル 4 をアラインメント誤り率において 3 ポイント上回る性能を示した．また，提案法は事後確率の閾値に敏感なものの，適切な閾値のもとでは HMM アラインメントに近い性能を示した．

1. はじめに

自動単語アラインメントは統計的機械翻訳のパイプラインにおける最初のタスクである．通常，フレーズ等を単位とする翻訳モデルは単語対応済みの二言語コーパスから抽出され，単語アラインメントモデルに基づく単語対単語の翻訳確率も用いられる．単語アラインメントに最も広く用いられるモデルは IBM モデル 4 [1] および隠れマルコフモデル (HMM) [18] である．これらのモデルでは，ある単語のアラインメントはその直前の単語のアラインメントに依存するという仮定を置いている．このような仮定は英語とフランス語のように近い言語間のアラインメントに関しては適切であるが，日本語と英語のように離れた言語間では必ずしも成立しない．

近年になり，統語構造に基づく情報を単語アラインメントに取り入れるいくつかのモデルが提案されている．HMM アラインメントに対する拡張として，Lopez and Resnik [10] は原言語側の依存構造木に基づく並び替えモデルを，DeNero and Klein [5] は原言語側の句構造木に基づく並び替えモデルをそれぞれ提案した．一部の教師ありモデルは入力として受け取った構文木を素性の生成および探索のガイドとして用いた [15], [16]．Burkett ら [2] は単語対応の付いた構文木対から学習する構文解析と単語アラインメントの同時モデルを提案した．また，フレーズ対フレーズのアラインメントモデルとして，Nakazawa and Kurohashi [11] は，並列サンプリングによって学習する部分木アラインメントモデルを提案した．しかしながら，上述の手法の

いずれも，両言語側の構文に基づく情報を考慮した教師無し学習を効率的に行うものではない．

隠れマルコフ木 (HMT) モデル [3] は以上に述べた性質を満たすモデルである．HMT モデルは隠れ変数が木構造を持つことを仮定しており，これは単語間の依存関係の概念と相性がよい．また，ラベルなしデータから期待値最大化アルゴリズムによって学習することが可能であり，その時間計算量は HMM と同等である．

本稿では，HMT モデルに基づく単語アラインメントモデルを提案し，提案法は両言語側の依存構造木を考慮した教師無し学習を効率的に行えることを示す．また，日英データを対象とした場合のアラインメント誤り率 (AER) について，提案法と IBM モデル 4 および HMM 単語アラインメントモデルとの比較を行う．

2. IBM モデル 1 と HMM アラインメント

本節では IBM モデル 1 [1] と HMM アラインメント [18] について簡単に述べる．これらはどちらも以下のように書き表せる生成モデルである．

$$p(\mathbf{f}|\mathbf{e}) = \sum_{\mathbf{a}} p(\mathbf{a}, \mathbf{f}|\mathbf{e}) \quad (1)$$

$$p(\mathbf{a}, \mathbf{f}|\mathbf{e}) = \prod_{j=1}^J p_d(a_j|a_{j-}) p_t(f_j|e_{a_j}), \quad (2)$$

ここでは $\mathbf{e} = \{e_1, \dots, e_I\}$ は英語 (原言語) の文であり， $\mathbf{f} = \{f_1, \dots, f_J\}$ は外国語 (目的言語) の文である． $\mathbf{a} = \{a_1, \dots, a_J\}$ はアラインメントベクトルである． $a_j = i$ は目的言語側の j 番目の単語が原言語側の i 番目の単語に対応することを意味し， $a_j = 0$ は目的言語側の j 番目の単語が原言語側に対応する単語を持たないことを意味する． j_- は j 番目以前の目的言語側の単語で，原言語側に対応す

¹ 奈良先端科学技術大学院大学 情報科学研究科

a) shuhei-k@is.naist.jp

b) kevinduh@is.naist.jp

c) matsu@is.naist.jp

る単語を持っていた最後の単語のインデックスを指す。

どちらのモデルにおいても $p_t(f_j|e_{a_j})$ は単語の翻訳確率であり、条件付き確率分布として表現できる。歪み確率については、どちらのモデルでも $p(a_j = 0|a_{j-} = i') = p_0$ は NULL 確率である。 $p(a_j = i|a_{j-} = i')$ はモデル 1 では均一であり、HMM では $c(i - i')$ に比例する。DeNero らは統語構造を考慮した歪みモデルを提案している [5]。このモデルでは、歪み確率が i 番目の単語と i' 番目の単語の距離ではなく、両者の句構造木上での経路に依存する。

これらのモデルはいずれも期待値最大化アルゴリズムによって効率的に学習可能である。実用上は、原言語側から目的言語側、およびその逆方向の両方向のモデルを学習し、両者の積や和を取るかあるいはヒューリスティックを用いて対称化することが一般的である。

3. 隠れマルコフ木モデル

Crouse ら [3] によって提案された隠れマルコフ木 (HMT) モデルは、画像分割やデノイズング、情報生命科学などの広い分野に応用されているが、自然言語処理の分野においてはあまり注目されていない。著者らの知る限りにおいては、HMT モデルを依存構造木ベースの機械翻訳システムにおけるトランスファーに用いた Žabokrtský and Popel [20] の研究が唯一の例外である。

HMT モデルは観測変数木 $X = \{x_1, \dots, x_N\}$ および隠れ変数木 $S = \{s_1, \dots, s_N\}$ からなり、両者は同形である。

同モデルのパラメータは

- 隠れ状態の事前分布 $P(s_1 = j)$
- 遷移確率 $P(s_t = j | s_{\rho(t)} = i)$
- 出力確率 $P(x_t = h | s_t = j)$

であり、ここでは $\rho()$ は隠れ状態ノードのインデックスをその親ノードのインデックスへと写像する関数である。これらのパラメータは期待値最大化アルゴリズムによって学習可能である。

Crouse ら [3] が提案した「上向き下向き」アルゴリズムは、HMM における前向き後ろ向きアルゴリズムに対応するものであり、期待値の計算に利用できるが、同時確率の分解に基づいているため、アンダーフローに陥る問題がある。

Durand ら [6] が提案した上向き下向きアルゴリズムの変種は事後確率の分解に基づいたものであり、アンダーフローを起こさない。この変種について、単語アラインメントの文脈において次節で説明する。

4. 隠れマルコフ木モデルを用いた単語アラインメント

本節では HMT モデルに基づく単語アラインメントを提案する。目的言語側の依存構造木 $f = \{f_1, \dots, f_J\}$ と原言語側の依存構造木 $e = \{e_1, \dots, e_I\}$ に対し、HMT 単語アライ

メントモデルは以下のように分解できる。

$$p(f|e) = \sum_a p(a, f|e) \quad (3)$$

$$p(a, f|e) = \prod_{j=1}^J p_d(a_j|a_{j-}) p_t(f_j|e_{a_j}) \quad (4)$$

これらの式は HMM アラインメントの式と同様であるが、 e, f および a が線形ではなく木構造を持つ点、および j_- が目的言語側の j 番目の単語の、原言語側に対応する単語を持つ最小祖先のインデックスを指す点が異なる。また、アラインメント変数 a_1, \dots, a_J からなる木 a が目的言語側の依存構造木 $f = \{f_1, \dots, f_J\}$ と同形である点は注意を要する。

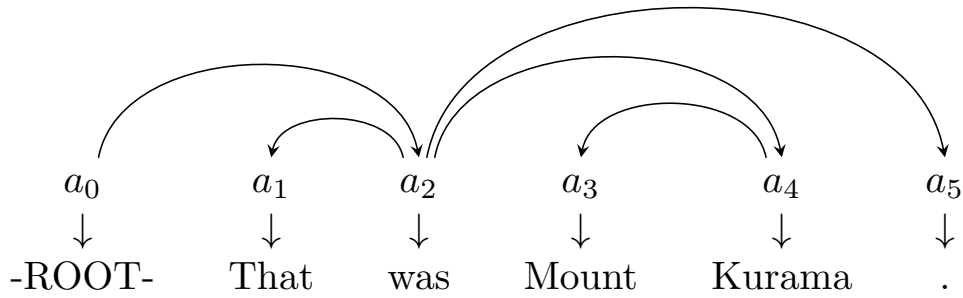
図 1 は目的言語側の依存構造木とアラインメント木、および原言語の依存構造木の例である。今後の説明では以下の表記法を用いるが、これは Durand ら [6] の表記法を単語アラインメントの文脈に合わせて修正したものである。

- $\rho(j)$ は目的言語側の j 番目の単語の親のインデックスを指す。
- $c(j)$ は目的言語側の j 番目の単語の子のインデックスの集合を指す。
- $\bar{F}_j = \bar{f}_j$ は目的言語側の j 番目の単語を根とする部分木を指す。
- $\bar{F}_{c(j)} = \bar{f}_{c(j)}$ は目的言語側の j 番目の単語の子を根とする部分木の集合を指す。
- $\bar{F}_{j'}$ が \bar{F}_j の部分木である場合、 $\bar{F}_{j \setminus j'}$ は j 番目の単語を根とする部分木から j' 番目の単語を根とする部分木を取り除いたものを指す。
- $\bar{F}_{0 \setminus c(j)}$ は目的言語側の依存構造木から j 番目の単語の子を根とする全ての部分木を取り除いたものを指す。

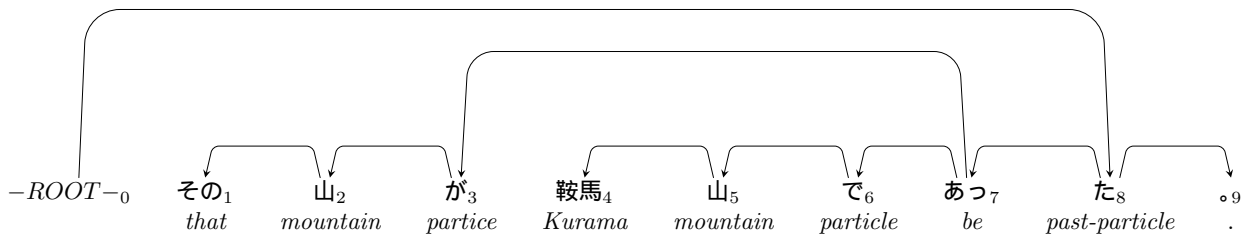
モデルパラメータのうち、3 節で述べた隠れ状態の事前分布は両側の依存構造木に仮定の ROOT ノードを置き、目的言語側の ROOT ノードが原言語側の ROOT ノードに対応すること、および目的言語側の他の単語が原言語側の ROOT ノードに対応しないことを強制することで表現できる。発生確率に対応する単語の翻訳確率 $p_t(f_j|e_{a_j})$ に関しては IBM モデル 1 および HMM アラインメントと同様である。

遷移確率に対応する歪み確率 $p_d(a_j = i|a_{j-} = i')$ は原言語側の i 番目の単語と i' 番目の単語の依存構造木上での距離 $d(i, i')$ に依存し、 $c(d(i, i'))$ によってモデルする。

提案法では、非負の距離の対 ($up, down$) によって $d(i, i')$ を表すが、 up は i 番目の単語から両単語の最小共通祖先 (lca) までの距離であり、 $down$ は i' 番目の単語から lca までの距離である。例えば図 1b では、 $d(0, 2) = (0, 4)$ 、 $d(2, 5) = (2, 2)$ および $d(4, 7) = (3, 0)$ である。実験では一定のウィンドウ幅 w を設けてそれ以上の距離は切り捨て、 $d(i, i') = (\min(up, w), \min(down, w))$ とし、 $c(d(i, i'))$ は



(a) 目的言語側および隠れ変数木



(b) 原言語側の木

図 1: 隠れマルコフ木アラインメントモデル

$(w + 1) \times (w + 1)$ の 2 次元行列に保存する .

期待値最大化アルゴリズムの期待値ステップには事後確率に基づく上向き下向きアルゴリズムの変種 [6] を用いることができる。以下に事後確率に基づく上向き下向きアルゴリズムについて、依存木対依存木の単語アラインメントの文脈で簡単に説明する。詳細に関しては原論文を参照されたい。

事後確率に基づく上向き下向きアルゴリズムでは、まず周辺確率

$$P(a_j = i) = \sum_{i'} P(a_{\rho(j)} = i') p(a_j = i | a_{\rho(j)} = i')$$

を目的言語側の各ノードの各状態について求める。ここで j 番目の単語が原言語側に対応する単語を持たない場合

$$p(a_j = i | a_{\rho(j)} = i') = p_0$$

であり、原言語側に対応する単語を持つ場合は

$$p(a_j = i | a_{\rho(j)} = i') = (1 - p_0) \cdot \frac{c(d(i', i))}{\sum_{i'' \neq i'} c(d(i', i''))}$$

である。歪み確率が線形距離に依存する場合は違い、依存構造木上では j 番目の単語からの距離が等しい単語が複数存在しうするため、 $p(a_j = i | a_{\rho(j)} = i')$ は人為的に正規化しなくてはならない点に注意が必要である。

次に上向き再帰で $P(a_j = i | \bar{\mathbf{F}}_j = \bar{\mathbf{f}}_j)$ を計算する。まず、各葉ノードにおいて

$$\begin{aligned} \beta_j(i) &= P(a_j = i | F_j = f_j) \\ &= \frac{P(f_j | e_i) P(a_j = i)}{N_j} \end{aligned}$$

によって初期化を行う。ここで

$$N_j = P(F_j = f_j) = \sum_i P(F_j | e_i) P(a_j = i)$$

である。次に、葉から根に向かって以下のように再帰的に計算を行う。

$$\begin{aligned} \beta_j(i) &= P(a_j = i | \bar{\mathbf{F}}_j = \bar{\mathbf{f}}_j) \\ &= \frac{\{\prod_{j' \in \mathbf{c}(j)} \beta_{j,j'}(i)\} P(f_j | e_i) P(a_j = i)}{N_j} \end{aligned}$$

ここで

$$\begin{aligned} N_j &= \frac{P(\bar{\mathbf{F}}_j = \bar{\mathbf{f}}_j)}{\prod_{j' \in \mathbf{c}(j)} P(\bar{\mathbf{F}}_{j'} = \bar{\mathbf{f}}_{j'})} \\ &= \sum_i \{ \prod_{j' \in \mathbf{c}(j)} \beta_{j,j'}(i) \} P(f_j | e_i) P(a_j = i) \end{aligned}$$

$$\begin{aligned} \beta_{\rho(j),j}(i) &= \frac{P(\bar{\mathbf{F}}_j = \bar{\mathbf{f}}_j | a_{\rho(j)} = i)}{P(\bar{\mathbf{F}}_j = \bar{\mathbf{f}}_j)} \\ &= \sum_{i'} \frac{\beta_j(i') P(a_j = i' | a_{\rho(j)} = i)}{P(a_j = i')} \end{aligned}$$

である。

上向き再帰の完了後、下向き再帰で $P(a_j = i | \bar{\mathbf{F}}_0 = \bar{\mathbf{f}}_0)$ を計算する。まず根ノードで

$$\xi_0(i) = P(a_0 = i | \bar{\mathbf{F}}_0 = \bar{\mathbf{f}}_0)$$

による初期化を行う．次に根から葉に向かって各ノードの各状態に対して

$$\begin{aligned} \xi_j(i) &= P(a_j = i | \bar{\mathbf{F}}_0 = \bar{\mathbf{f}}_0) \\ &= \frac{\beta_j(i)}{P(a_j = i)} \sum_{i'} \frac{P(a_j = i | a_{\rho(j)} = i') \xi_{\rho(j)}(i')}{\beta_{\rho(j),j}(i')} \end{aligned}$$

を計算する．

歪み確率の推定に用いる条件付き確率

$$\begin{aligned} P(a_j = i, a_{\rho(j)} = i' | \bar{\mathbf{F}}_0 = \bar{\mathbf{f}}_0) \\ = \frac{\beta_j(i) P(a_j = i | a_{\rho(j)} = i') \xi_{\rho(j)}(i')}{P(a_j = i) \beta_{\rho(j),j}(i')} \end{aligned}$$

は下向き再帰中に計算可能である．

最大化ステップにおいて，単語の翻訳確率は IBM モデルおよび HMM アラインメントと同様に

$$P(f|e) = \frac{c(f, e)}{c(e)}$$

で更新できる．しかし， $c(i, i')$ を保存する行列は確率分布ではないため，歪みモデルの更新方法は自明ではない．提案法では，最尤推定の近似として，期待値ステップにおいて計算した $c(d(i, i'))$ の値を訓練データ中の距離 $d(i, i')$ を持つ遷移の回数で割った上で，行列の各要素を

$$c(d(i, i')) = \frac{c(d(i, i'))}{\sum_{i=0}^w \sum_{i'=0}^w c(d(i, i'))}$$

で正規化する．

5. 実験

本節では提案法の隠れマルコフモデルに基づく単語アラインメントモデルの性能を公開されている英日データ上で評価し，IBM モデル 4 [1]，距離に基づく HMM アラインメント (D-HMM) [9]，[18] および統語構造を考慮した HMM アラインメント (S-HMM) [5] と比較する．

実験には京都フリー翻訳タスク (KFTT) [12] version 1.3 を用いる．表 1 にコーパスの統計情報を示す．データセットのデフォルトの設定とは微妙に数値が異なるが，これは以下に述べる前処理方法の違いによるものである．

5.1 前処理

英語側データは Stanford Tokenizer^{*1} を用いてトークナイズし，Berkeley Parser^{*2}[14] で句構造解析を行った．この句構造木は Berkeley Aligner の統語構造に基づく歪みモデルに用いるとともに，依存構造木に変換して提案法の歪みモデルに用いた^{*3}．また，訓練データ (約 33 万文) 中の

^{*1} <http://nlp.stanford.edu/software/>

^{*2} 解析モデルの訓練には Ontonotes[7] の Wall Street Journal 部分を用いた．

^{*3} 変換には Stanford のツール [4] を用いた．

7 文に対しては Berkeley Parser が解析に失敗したため，両言語側から取り除いた．

日本語側データに関しては，まずデータセットのデフォルトによる分かち書き後の文中の機能語部分を，データセットの作者が公開しているスクリプト^{*4}を用いて連結した．次に MeCab^{*5} version 0.996 を用いて再度形態素解析を行い，CaboCha^{*6} version 0.66 を用いて依存構造解析を行った．CaboCha の出力は CoNLL 形式で行い，括弧や句読点などの記号が親ノードになっている箇所は修正した．このような手続きを経たのは，データセットのデフォルトの分かち書きと解析器にとって扱いやすい分かち書きの間でバランスを取るためである．

また，前処理の違いが原因でデフォルトの正解アラインメントを用いることが不可能なため，データセットの作者が公開しているスクリプト^{*7}を用いて，新たな分かち書きに合うように正解アラインメントを修正した．

5.2 学習

提案法のモデルは同時学習した [9] IBM モデル 1 の翻訳パラメータ (5 イテレーション) を用いて初期化し，依存構造に基づく二方向のモデルを独立に各 5 イテレーション学習した．学習に要した時間は 3.3 GHz の Xeon CPU 上で合計で約 4 時間であった．

IBM モデル 4 の学習には GIZA++ [13] と Moses [8] のスクリプトを用いた．

D-HMM および S-HMM アラインメントモデルの学習には Berkeley Aligner を用いた．同時学習した IBM モデル 1 (5 イテレーション) で初期化し，二方向の HMM アラインメントを独立に 5 イテレーション学習した．IBM モデル 1 の同時学習による初期化が有効だったのに対し，HMM アラインメントの同時学習はこのデータ上での性能には悪影響を与えた．

5.3 結果

提案法および HMM アラインメントのデコーディングには *soft union* による *posterior thresholding* および *competitive thresholding* [5]，[9] を，IBM モデル 4 には *grow-diag-final-and* ヒューリスティックを用いた．

表 2 および図 2 に結果を示した．提案法については *competitive thresholding* を用いた場合，HMM アラインメントについては *competitive thresholding* を用いない場合の方がよい性能を示したため，各デコーディング基準同士と各モデルの最も性能の良かった基準同士に関して適合率 / 再現

^{*4} <https://github.com/neubig/util-scripts/blob/master/combine-predicate.pl>

^{*5} <http://code.google.com/p/mecab/>

^{*6} <http://code.google.com/p/cabocho/>

^{*7} <https://github.com/neubig/util-scripts/blob/master/adjust-alignments.pl>

	Sentences	English Tokens	Japanese Tokens
Train	329,974	5,912,543	5,893,334
Dev	1,166	24,354	26,068
Tune	1,235	30,839	33,180
Test	1,160	26,730	27,693

表 1: KFTT の統計データ.

	Precision	Recall	AER
HMT (Proposed)	71.77	55.23	37.58
IBM Model 4	60.58	57.71	40.89
D-HMM	69.59	56.15	37.85
S-HMM	71.60	56.14	37.07

表 2: 各モデル毎の最高のアラインメント誤り率 (AER)

率とアラインメント誤り率 (AER) の曲線を比較した.

表 2 に示したように, 提案法の最高性能は IBM モデル 4 よりも AER で 3 ポイント低く, D-HMM と S-HMM の中間に位置した. これは提案法の歪みモデルが S-HMM に比べて単純なものであることを考えれば自然な結果であり, 提案法の歪みモデルをより洗練されたもので置き換えることにより性能の向上につながることを期待できる.

図 2a および 2d を詳しく見ると, 提案法は適合率 50 から 60% の範囲では HMM アライメントを若干上回っており, これは *posterior thresholding* の閾値では 0.15 から 0.35 の範囲に相当する. また, 図 2b および 2e からは, *competitive thresholding* を用いた場合提案法は適合率で 60 から 70 %, 閾値では 0.2 から 0.4 の範囲で良好な性能を示すことが分かる. さらに, どの場合でも提案法の性能曲線は HMM アラインメントに比べて尖っている.

このような振る舞い方の原因の一つとして, 提案法の歪みモデルが HMM アラインメントの歪みモデルに比べてより均一になりやすい傾向を持つことが考えられる. 例えば, 提案法では同じノードを親に持つノードへの親ノードからの歪み確率は全て等しくなるが, これは HMM アラインメントの歪みモデルでは一定の範囲内にあるノードが通常異なる歪み確率を持つのは対称的である. より均一に近い歪み確率の元ではアラインメントの事後分布もより拡散するものと思われ, これが閾値に対する敏感性につながっていると考えられる.

6. 関連研究

Lopez and Resnik [10] は依存構造木上の距離に基づく HMM アラインメントモデルを提案した. DeNero and Klein [5] は句構造木上の経路に基づく HMM アラインメントモデルを提案し, Tree-to-string トランスデューサのルール抽出における精度の向上を示した. どちらのモデルも隠れ変数 (アラインメント) の構造として鎖を仮定している点, および片方の言語においてのみ統語構造を考慮し

ている点において, 隠れ変数が木構造を持つと仮定するとともに両言語側の統語構造を考慮する提案法とは異なる.

Nakazawa and Kurohashi [11] は依存構造に基づくフレーズ対フレーズのアラインメントモデルを提案している. 一部の教師ありモデルでは, 統語構造を素性の生成や探索のガイドとして用いる [2], [15], [16].

Žabokrtský and Popel [20] はトランスファー方式の機械翻訳システムに隠れマルコフモデルを用いたが, パラメータ自体の学習は完了していることが前提とされていた.

7. 結論

本稿では隠れマルコフモデルに基づく単語アラインメントモデルを提案し, 提案法は両言語側の統語構造を考慮した教師なしアライメントの学習を効率的に行えることを示した. 日英データを用いた実験における提案法の性能は IBM モデル 4 を上回り, HMM アラインメントに近い水準であった.

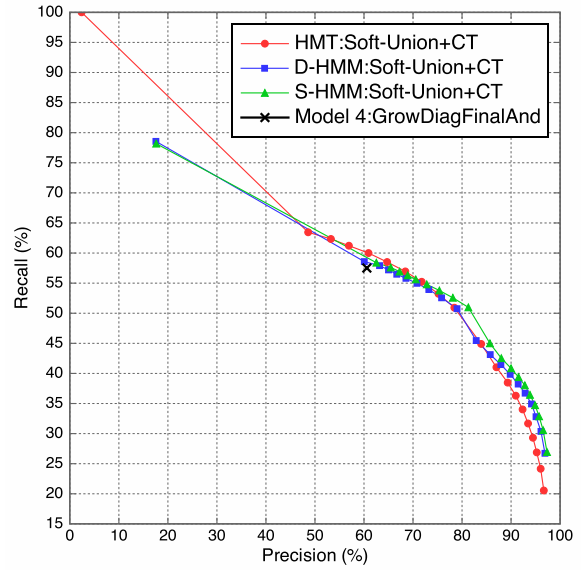
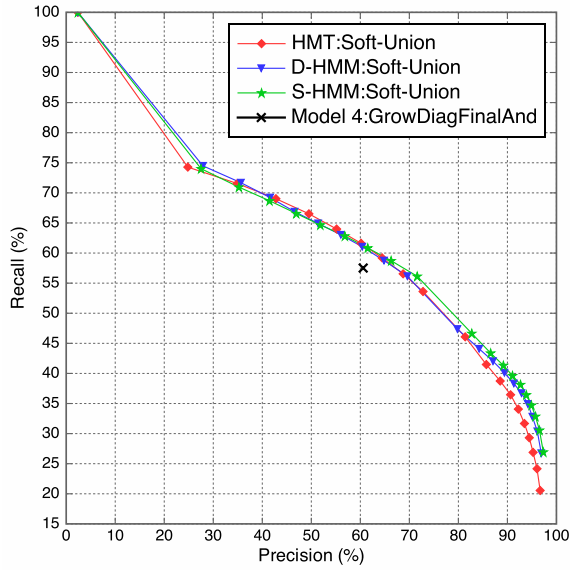
今後の最優先の課題としては, 提案法の歪みモデル部分を改善することがあげられる. DeNero and Klein[5] の統語構造を考慮した歪みモデルに近いものがその有力な候補である.

また, 提案法によるアラインメントが翻訳精度に及ぼす影響を確認することも重要な課題である. 中でも, 文から依存木へのモデル [17] や依存木から文へのモデル [19] など, 近年提案されている依存構造を取り入れた翻訳モデルとの組み合わせは特に興味深い.

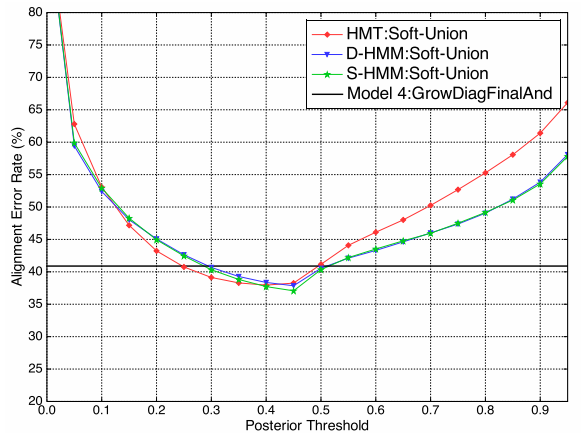
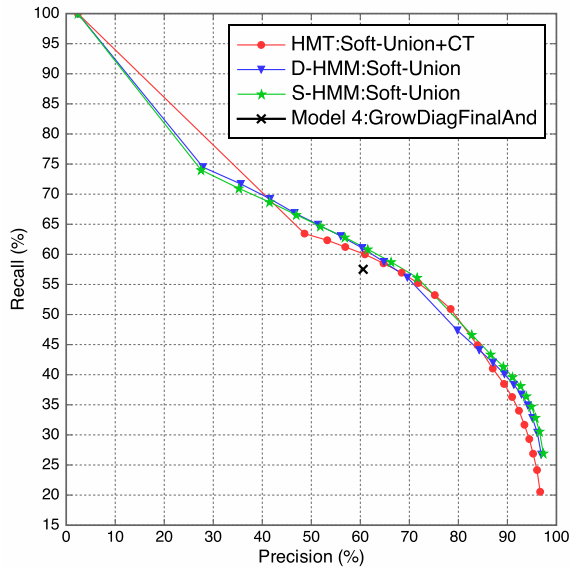
謝辞 京都フリー翻訳タスクの作成および公開に携わった方々に心より感謝いたします.

参考文献

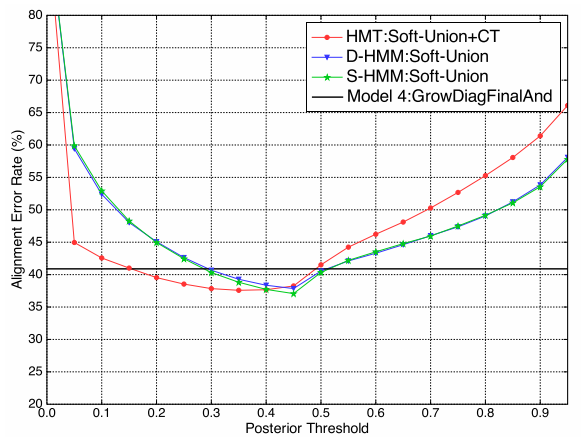
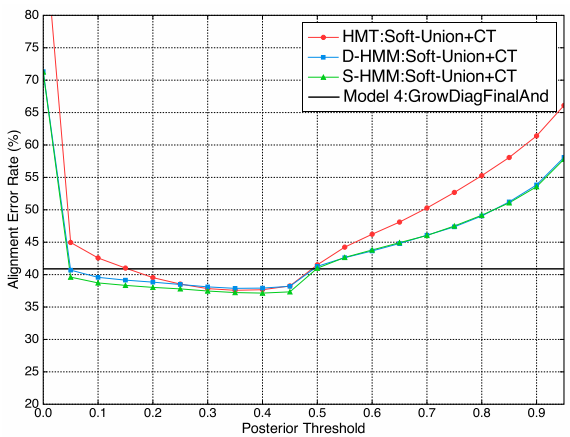
- [1] Brown, P. F., Della Pietra, V. J., Della Pietra, S. A. and Mercer, R. L.: The Mathematics of Statistical Machine Translation: Parameter Estimation, *Computational linguistics*, Vol. 19, No. 2, pp. 263–311 (1993).
- [2] Burkett, D., Blitzer, J. and Klein, D.: Joint Parsing and Alignment with Weakly Synchronized Grammars, *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pp. 127–135 (2010).
- [3] Crouse, M. S., Nowak, R. D. and Baraniuk, R. G.: Wavelet-Based Statistical Signal Processing Using Hidden Markov Models, *Signal Processing, IEEE Transactions on*, Vol. 46, No. 4, pp. 886–902 (1998).
- [4] de Marneffe, M.-C., MacCartney, B. and Manning,



(a) Competitive Thresholding を用いない場合の再現率 / 適合率曲線 . (b) Competitive Thresholding を用いた場合の再現率 / 適合率曲線 .



(c) 各モデルに最適な基準を用いた場合の再現率 / 適合率曲線 . (d) Competitive Thresholding を用いない場合のアライメント誤り率 .



(e) Competitive Thresholding を用いた場合のアライメント誤り率 . (f) 各モデルに最適な基準を用いた場合のアライメント誤り率 .

図 2: モデルおよびデコーディング基準毎の再現率 / 適合率曲線とアライメント誤り率

- C. D.: Generating Typed Dependency Parses from Phrase Structure Parses, *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Vol. 6, pp. 449–454 (2006).
- [5] DeNero, J. and Klein, D.: Tailoring Word Alignments to Syntactic Machine Translation, *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, pp. 17–24 (2007).
- [6] Durand, J.-B., Gonçalves, P. and Guéron, Y.: Computational Methods for Hidden Markov Tree Models-An Application to Wavelet Trees, *Signal Processing, IEEE Transactions on*, Vol. 52, No. 9, pp. 2551–2560 (2004).
- [7] Hovy, E., Marcus, M., Palmer, M., Ramshaw, L. and Weischedel, R.: OntoNotes: The 90% Solution, *Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers*, pp. 57–60 (2006).
- [8] Koehn, P., Hoang, H., Birch, A., Callison-Burch, C., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, C., Zens, R., Dyer, C., Bojar, O., Constantin, A. and Herbst, E.: Moses: Open Source Toolkit for Statistical Machine Translation, *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics Companion Volume Proceedings of the Demo and Poster Sessions*, Prague, Czech Republic, Association for Computational Linguistics, pp. 177–180 (online), available from <http://www.aclweb.org/anthology/P07-2045> (2007).
- [9] Liang, P., Taskar, B. and Klein, D.: Alignment by Agreement, *Proceedings of the Human Language Technology Conference of the NAACL, Main Conference*, pp. 104–111 (2006).
- [10] Lopez, A. and Resnik, P.: Improved HMM Alignment Models for Languages with Scarce Resources, *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, pp. 83–86 (2005).
- [11] Nakazawa, T. and Kurohashi, S.: Bayesian Subtree Alignment Model based on Dependency Trees, *Proceedings of 5th International Joint Conference on Natural Language Processing*, pp. 794–802 (2011).
- [12] Neubig, G.: The Kyoto Free Translation Task, <http://www.phontron.com/kfft> (2011).
- [13] Och, F. J. and Ney, H.: A Systematic Comparison of Various Statistical Alignment Models, *Computational linguistics*, Vol. 29, No. 1, pp. 19–51 (2003).
- [14] Petrov, S., Barrett, L., Thibaux, R. and Klein, D.: Learning Accurate, Compact, and Interpretable Tree Annotation, *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pp. 433–440 (2006).
- [15] Riesa, J., Irvine, A. and Marcu, D.: Feature-Rich Language-Independent Syntax-Based Alignment for Statistical Machine Translation, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 497–507 (2011).
- [16] Riesa, J. and Marcu, D.: Hierarchical Search for Word Alignment, *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp. 157–166 (2010).
- [17] Shen, L., Xu, J. and Weischedel, R.: A New String-to-Dependency Machine Translation Algorithm with a Target Dependency Language Model, *Proceedings of ACL-08: HLT*, pp. 577–585 (2008).
- [18] Vogel, S., Ney, H. and Tillmann, C.: HMM-Based Word Alignment in Statistical Translation, *Proceedings of the 16th conference on Computational linguistics-Volume 2*, pp. 836–841 (1996).
- [19] Xie, J., Mi, H. and Liu, Q.: A Novel Dependency-to-String Model for Statistical Machine Translation, *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, pp. 216–226 (2011).
- [20] Žabokrtský, Z. and Popel, M.: Hidden Markov Tree Model in Dependency-based Machine Translation, *Proceedings of the ACL-IJCNLP 2009 Conference Short Papers*, pp. 145–148 (2009).