

遷移型依存構造解析における擬似曖昧性の影響

進藤 裕之 永田 昌明

NTT コミュニケーション科学基礎研究所

{shindo.hiroyuki, nagata.masaaki}@lab.ntt.co.jp

1 はじめに

依存構造解析(係り受け解析)は、自然言語処理分野において重要な基本技術の一つであり、これまでに多くの研究が行われてきた [4, 5]。特に、シフト・リデュース法) [5, 3] に代表される遷移型依存構造解析は、比較的高速に動作し、大域的な文脈情報をモデルに取り入れることができるため広く用いられている。

遷移型依存構造解析では、解析器の入力である文から最終的な出力である依存構造を一連の状態遷移と捉え、入力状態から出力状態へ至る最適な行動系列を学習する。図 1 に、遷移型依存構造解析の状態遷移の例を示す。ここで、「状態」とは、部分的に依存構造が構築された文を指し、「行動」とは、まだ依存関係が付与されていない単語間に、新たに依存関係を構築することを指す。実際に解析を行う際には、一つの状態から複数の行動の可能性が存在するため、分類器を用いて最適な行動の一つを選択するという動作を繰り返して、最終的な解析結果を得る。

遷移型依存構造解析には、「単語を左から右の順番に読み込んで依存関係を決定する」といったように解析の方向を既定する方法 [5] と、順番をあらかじめ定めずに、解析モデルのスコアが高い順番に解析を行う方法 [1] とがある。前者は、正解の依存構造へ到達する経路は一つ、または少数しか存在しない。これを、「擬似曖昧性が低い」という。擬似曖昧性の低いモデルでは、学習やデコードに要する時間が短くて済むという利点があるが、文の後半の文脈を参照しなければ正しい依存関係が決定できない場合には、解析誤りを引き起こしやすい。そのため、従来の研究では、ビーム探索を援用して複数の状態の可能性(仮説)を同時に探索する方法によって精度を向上させることが常套手段となっ

ている。

一方、無方向性の解析手法は擬似曖昧性が高く、正解の状態へ到達する状態遷移系列が多数存在する。近年の研究では、擬似曖昧性の高いモデルでは、判断が困難な依存関係の決定を後回しにして容易な依存関係から先に決定できるため、決定的な解析では、擬似曖昧性の低いモデルよりも解析誤りを引き起こしにくいことがわかってきた。

このように、擬似曖昧性は遷移型依存構造解析の性能を左右する重要な要因の一つである。しかし、従来の無方向性解析に関する研究では、主に決定的探索を前提としており、複数の仮説を考慮する手法との比較や、そもそも擬似曖昧性の高いモデルとビーム探索を組み合わせることで精度向上の効果があるのかということ十分に明らかでない。さらに、双方の手法で用いられている統計モデルの素性も異なるため、公平な比較とはいえない。したがって、本研究では、擬似曖昧性の高い無方向性依存構造解析にビーム探索を導入するいくつかの手法を提案する。そして、擬似曖昧性の高いモデルにおいてもビーム探索が有効なのか、また、探索方向が既定されている他手法と比較しても優位なのかを比較・検証する。

2 無方向性依存構造解析におけるビーム探索

本節では、まず、無方向性依存構造解析の持つ擬似曖昧性について説明し、その次に無方向性依存構造解析におけるビーム探索について説明する。図 2 に、無方向性依存構造解析における擬似曖昧性の例を示す。図 2 に示されている二つの状態遷移系列は、最終的に同じ解析結果へ到達するが、状態遷移は異なっている。このように、無方向依存構造解析では、自由な順序で依存



図1 遷移型依存構造解析の状態遷移の例

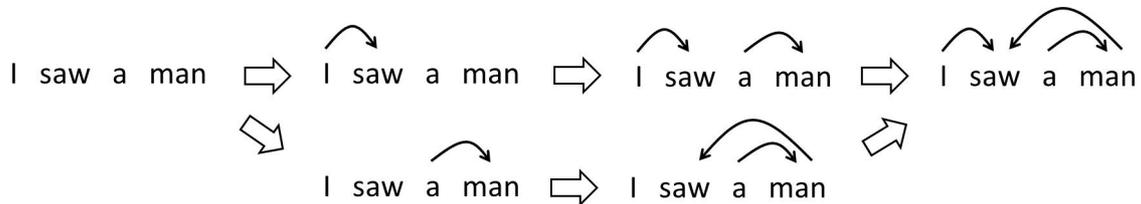


図2 無方向依存構造解析における擬似曖昧性の例

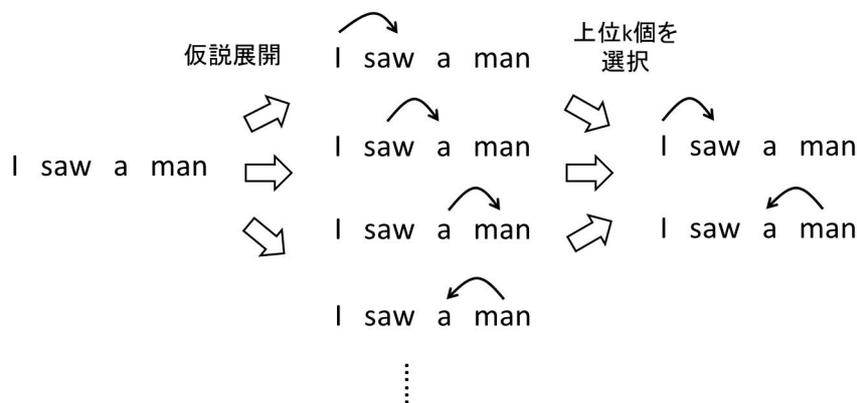


図3 ビーム探索の例

関係を構築できるため、同じ状態へ到達可能な状態遷移系列は他にも大量に存在する。これが擬似曖昧性の問題である。前述のように、擬似曖昧性の高い解析方法は探索空間が広大であり、最適な状態遷移が一意に決まらないという問題があるが、一方で、判断が困難な依存関係の決定を後回しにして、容易な依存関係から先に決定できるため、解析誤りを引き起こしにくいという利点がある。

正解の依存構造が付与された文データが与えられたとき、統計的手法を用いて正しい依存構造へ到達する最適な状態遷移系列を学習することができる。本稿では、最大エントロピーモデルを用いて最適な状態遷移系列を学習することを考える。最大エントロピーモデルは、各状態が与えられたときに、候補となる行動の中から最適な行動を選択するために用いられる。このとき、無方向性の手法では、正解の依存構造へ到達する状態遷移系列は複数存在するため、正解の依存構造が与えられていたとしても、そこから最適な状態遷移系列

を一意に決めることができない。そのため、学習の各反復において、正解の依存構造へ至る状態遷移系列の中で、現時点のモデルスコアが最も高い状態遷移系列を一つ決定する。そして、その最適な系列を正しい状態遷移系列と見なして、最大エントロピーモデルの素性重みを更新する。シフト・リデュース法のようにあらかじめ解析の順序が決まっている手法では、正解の状態遷移系列を事前に一意に決定することができるので、学習の各反復で最適な状態遷移系列が変わることはない。

ビーム探索は、複数の状態を仮説として保持しながら解析を行う方法である。図3にビーム探索の例を示す。図3において、入力文の次の状態は複数の可能性がある。ビーム探索では、候補となる各状態のスコアを計算し、そこから上位 k 個の状態のみを仮説として残す。各状態から n 通りの行動の可能性があるとするとき、次のステップでは、 k 個の状態からそれぞれ次の状態のスコアを計算し、 $k \times n$ 通りの全状態から、再び上



図4 状態を併合するビーム探索の例

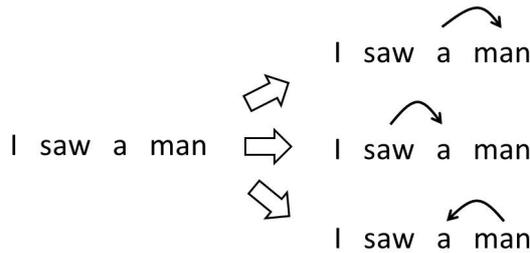


図5 排他的な仮説のみを考慮したビーム探索の例

位 k 個の状態のみを抽出する．最終的に k 通りの依存構造結果が得られるので，その中で最もスコアの高い依存構造を出力する．

無方向性依存構造解析は擬似曖昧性が高いので，単純なビーム探索では複数の仮説が同じ状態へ到達してしまう可能性が高いと予想される．したがって，本稿では，ビーム探索を用いても擬似曖昧性の問題を緩和できる以下の二つの手法を提案する．一つ目は，可能性のある全ての仮説を考慮するという点では通常のビーム探索と同じだが，同じ状態に到達してしまった場合はスコアの低い状態のみを残して，スコアの低い状態は削除するという方法である．これにより，状態遷移系列が異なる二つの同じ状態は一つに併合されるため，複数の同じ状態を仮説として考慮する必要がない．図4に例を示す．二つ目は，ビーム探索において，排他的仮説のみを考慮する方法である．排他的仮説とは，同じ状態へ到達不可能な仮説集合のことを指す．例えば，状態 A と B があったときに， A と B が排他的であるとは， A から到達可能な全ての状態と， B からの到達可能な全ての状態との積集合は空集合となることである．二つの状態が排他的であるかどうかは，各状態の保持している依存関係が互いに矛盾していないかをチェックすることで判定できる．図5に例を示す．

3 実験

擬似曖昧性の問題がビーム探索にどの程度影響を与えるかを評価するために，評価実験を行った．使用したデータは，英語の PennTreebank データを Yamada ら [6] の変換ルールを使用して依存構造へ変換したものである．学習データは，セクション2から21の中から10000文を無作為に抽出した．テストデータは，セクション23を用いた．最大エントロピーモデルの素性の重みの学習は，構造化パーセプトロンと max violation 学習法 [2] を用いた．無方向性解析の素性テンプレートは，Huang ら [3] のシフト・リデュース法の素性テンプレートと，Goldberg ら [1] の無方向性の素性テンプレートを組み合わせたものを使用した．ただし，Goldberg らの使用している素性テンプレートに含まれる PP attachment 素性は除いた．

実験結果を表1に示す．まず，決定的な解析（ビーム幅1）の場合，既存研究の素性を用いたシフト・リデュース法は，無方向性解析法と比較して解析誤りが多く，精度が低いことがわかる．これは，シフト・リデュース法では左から順番に解析を行うため，右の文脈情報が上手く利用できていないために解析を誤りやすいと考えられる．

一方，擬似曖昧性の高い無方向性解析は，シフト・リデュース法よりも精度が高い．これは，前述のように擬似曖昧性の高さを上手く利用して，判断が容易な依存関係から決定していく状態遷移系列を学習できたことを示唆している．これらの結果は，無方向性の既存研究で示されている結果と概ね一致する．

ビーム幅を増やしたときには，両者は異なる振る舞いを示した．まず，シフト・リデュース法は，ビーム幅を増やすと解析精度が大きく向上する．一方，無方向性解析では，単純なビーム探索を適用した場合，ビーム幅を増やしても精度はほとんど向上しなかった．これは，擬似曖昧性のために，仮説が同じ状態へ到達することが多かったためだと考えられる．排他的仮説のみを考慮する方法では，ビーム幅を増やすと解析精度は若干向上した．また，同一の状態を併合する方法は，さらに解析精度が向上することがわかった．排他的仮説のみを考慮した場合，ビーム探索による擬似曖昧性の問題は解消されるが，モデルのスコアがあまり高くない状態を仮説に含めてしまう可能性が相対的に高くなり，状態の併合に比べて精度が向上しなかったと考え

手法 / ビーム幅	1	2	4	8
シフト・リデュース ([3] の素性)	87.7	88.9	89.6	89.7
無方向	89.7	89.9	89.9	89.9
無方向 + 排他的	-	90.0	90.1	90.1
無方向 + 状態併合	-	90.2	90.2	90.3
シフト・リデュース (無方向と同じ素性)	89.2	90.0	90.3	90.6

表1 シフト・リデュース法と無方向法とのラベルなし構文解析精度の比較

ビーム幅	2	4	8
	0.22	0.92	2.58

表2 状態併合の起こった割合

構文解析精度 (ラベルなし)	
シフト・リデュース (本研究)	92.3
無方向 + 状態併合	92.0
シフト・リデュース ([3])	92.1

表3 標準データ (セクション2から21)での構文解析精度の比較

られる。同一の状態を併合する方法は、通常のビーム探索と同様に必ずスコアの上位 k 個を保持し、かつ、同じ状態を仮説としてもたないため、決定的解析よりも精度が大幅に向上した。以上の結果から、擬似曖昧性の高いモデルは、上手くビーム探索を導入することで、決定的解析でない状況においても擬似曖昧性の低いモデルと同等以上の解析精度を達成することができることが新たにわかった。ただし、他の言語やドメインのデータでは異なる結果となる可能性があるため、更なる検証が必要である。

次に、両者をより公平に比較するため、無方向性解析で使用した素性テンプレートを使ったときのシフト・リデュース法の結果を表1の下段に示す。その結果、シフト・リデュース法は既存研究の結果よりも大幅に精度が向上した。決定的解析では無方向性解析が優位だが、ビーム幅を増やした場合には、シフト・リデュース法のほうが精度が高い。これは、擬似曖昧性の高いモデルでは、同じ状態の仮説を併合する方法を適用しても、いずれ同じ状態となる似たような状態は事前に取り除くことができないため、ビーム探索の効果は限定されていると考えられる。

表2は、無方向性解析において、状態併合が起こった割合を表している。例えば、ビーム幅8では、各ステップにおいて平均2.58個の状態併合が起こっている。

最後に、セクション2から21の全てを学習データとして用いて構文解析精度の比較を行った。結果を表3に示す。ビーム幅は8に設定した。無方向性解析は、従来は決定的な解析において用いられてきたが、ビーム探索と組み合わせても他手法と同等の結果を実現できることが明らかとなった。また、シフト・リデュース法に関しては、無方向性解析の素性を加えることで、Huangらの手法よりも精度を向上させることができた。

4 おわりに

本稿では、無方向性の依存構造解析にビーム探索を導入するいくつかの方法を提案し、擬似曖昧性の高いモデルにおいても複数の仮説を同時に考慮する効果があるかを検証した。無方向性依存構造解析のように擬似曖昧性の高いモデルにおいては、単純にビーム探索を導入することの効果はほとんどないが、仮説が同じ状態へ到達しない工夫を行うことで、ビーム探索によって精度が向上することを確認した。また、シフト・リデュース法に、無方向性解析の素性を導入することで、更なる精度向上が実現できることを示した。今後は、他の言語やドメインのデータで擬似曖昧性の影響を検証する必要がある。

参考文献

- [1] Yoav Goldberg and Michael Elhadad. An efficient algorithm for easy-first non-directional dependency parsing. In *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, pages 742–750. Association for Computational Linguistics, 2010.
- [2] Liang Huang, Suphan Fayong, and Yang Guo. Structured perceptron with inexact search. In *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 142–151. Association for Computational Linguistics, 2012.
- [3] Liang Huang and Kenji Sagae. Dynamic programming for linear-time incremental parsing. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1077–1086. Association for Computational Linguistics, 2010.
- [4] Ryan McDonald, Fernando Pereira, Kiril Ribarov, and Jan Hajič. Non-projective dependency parsing using spanning tree algorithms. In *Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing*, pages 523–530. Association for Computational Linguistics, 2005.
- [5] Joakim Nivre. Algorithms for deterministic incremental dependency parsing. *Computational Linguistics*, 34(4):513–553, 2008.
- [6] Hiroyasu Yamada and Yuji Matsumoto. Statistical dependency analysis with support vector machines. In *Proceedings of IWPT*, volume 3, 2003.