

## 二分依存関係に基づく深層格推測手法 A Method for Deep Case Analysis Based on Two-forked Dependency

渋木 英潔<sup>†</sup>  
Hideyuki Shibuki

荒木 健治<sup>††</sup>  
Kenji Araki

桃内 佳雄<sup>††</sup>  
Yoshio Momouchi

柄内 香次<sup>††</sup>  
Koji Tochinal

### 1. まえがき

本稿では、深層格推測における学習効率の向上を目的として、従来手法における問題点の考察と新たな知識表現の検討を行う。

自動要約や対話処理などにおいて、深層格を利用することで精度を向上させる試みが従来から行われている。それゆえ、深層格を推測するための知識を自動的に獲得する手法に関する研究も数多く行われている [1, 2, 3]。我々のグループでは、個々のユーザに適応した知識をユーザが提示した事例から帰納的に学習する研究を行っており、深層格を対象とした研究を文献 [1] などで行っている。文献 [1] では、構文構造から意味構造へ変換するための規則を実例から帰納的に学習し、獲得された変換規則を用いて深層格の推測を行う。しかしながら、獲得される規則は可変長であり、規則の一般化は同じ長さの規則間でしか行われなため、規則の適用範囲や一般化の速度といった効率の面で再考の余地がある。

深層格の推測において、名詞 - 助詞 - 動詞となる一組の依存関係を基に行うことは、動詞の多義性などによる影響を受けやすい。したがって、文全体を考慮した知識を用いることが望ましいが、そのような知識は適用時の被覆率の点で問題が生じる傾向がある。動詞の多義性を解消する方法の一つとして直前の名詞を含む文節を参照する方法が知られている [4]。それゆえ、本稿では、深層格推測の知識表現を二つの名詞と一つの動詞を対象とした二分依存関係とすることで多義性と被覆率の両方の解決を試みる。

2 節では、問題点を解決するために二分依存関係の考え方を提案し、その考えに基づいた規則表現を提案する。3 節では、二分依存関係に基づいた規則の学習と適用について述べる。4 節は今後の予定である。

### 2. 二分依存関係

最初に、構文構造と意味構造を対応付ける依存関係の例を図 1 に示す。例となる箇所は二重下線を引いた「1941年、ナチスドイツはバルバロッサ作戦を發動して」の部分である。この箇所は三つの依存関係から構成されており、「發動」に対して、「1941年」、「ナチスドイツ」、「バルバロッサ作戦」がそれぞれ係っている。これに対応する機能語を抽出して規則の条件部を作成すると、「 $\times \times$ は  $\times \times$ をして」となり、この規則の適用により、 $\times \times$ の深層格が [time]、 $\times \times$ が [agent]、 $\times \times$ が [object] であると推測することが可能となる。その後、学習事例

1941年、ナチスドイツはバルバロッサ作戦を發動してソ連領内に電撃侵襲した。

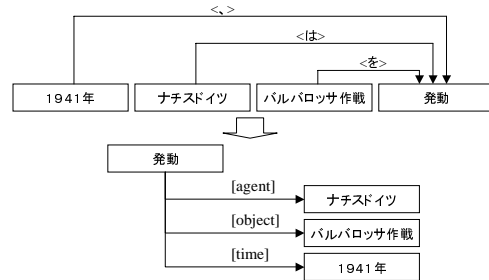


図 1: 依存関係の全体例

の中から規則の条件部と一致する事例を収集し、 $\times \times$ などの部分に対応する内容語をクラスタリングすることで一般性を高め、規則の適用範囲を拡大する。

しかしながら、この規則の条件部と一致する文の頻度は一般にそれほど大きくはないと考えられ、規則の一般化を行うために十分な量の事例を確保することが困難であると考えられる。この問題を解決するために、条件部を構成する依存関係を分解することで規則の適用範囲を拡大することができる。このとき、条件部を「 $\times \times$ は  $\times \times$ をして」とのように一組の依存関係に限定することで適用範囲を大幅に拡大できるが、そのような規則は「水は蒸発して」のように [agent] が相応しくない依存関係に対しても誤適用される危険性が高まる。文献 [4] では直前の格要素を用いて動詞の多義性を解消する手法を提案しており、本手法においてもこの考え方を採用した。条件部を構成する依存関係の中で、最も内側に位置する依存関係を必須として、それ以外の依存関係から任意の一つを組み合わせた形式を二分依存関係と定義する。また、前者の依存関係を必須条件部、後者を任意条件部と定義する。図 1 の依存関係を、二分依存関係に基づいて分解した例を図 2 に示す。どちらの二分依存関係においても必須条件部は「バルバロッサ作戦を發動して」であり、任意条件部はそれぞれ「ナチスドイツは」と「1941年」になる。

3. 規則の学習と適用

二分依存関係の考え方に基いて、ユーザから与えられた事例を分解し規則として学習する。しかしながら、「本を読む」のように一組の依存関係しか存在しない場合、二分依存関係として規則化することが困難である。本手法では、一組の依存関係しか明示されていない事例からは規則として学習しない。

### 3. 規則の学習と適用

次に規則同士を比較し、内容語の類似度を表現するためのネットワークを作成する。ネットワークのノードは

<sup>†</sup>北海道大学ハイテク・リサーチ・センター, High-Tech Research Center, Hokkai-Gakuen University

<sup>††</sup>北海道大学大学院情報科学研究科, Graduate School of Information Science and Technology, Hokkaido University

<sup>‡</sup>北海道大学工学部, Faculty of Engineering, Hokkai-Gakuen University

<sup>†††</sup>北海道大学経営学部, Faculty of Business Administration, Hokkai-Gakuen University

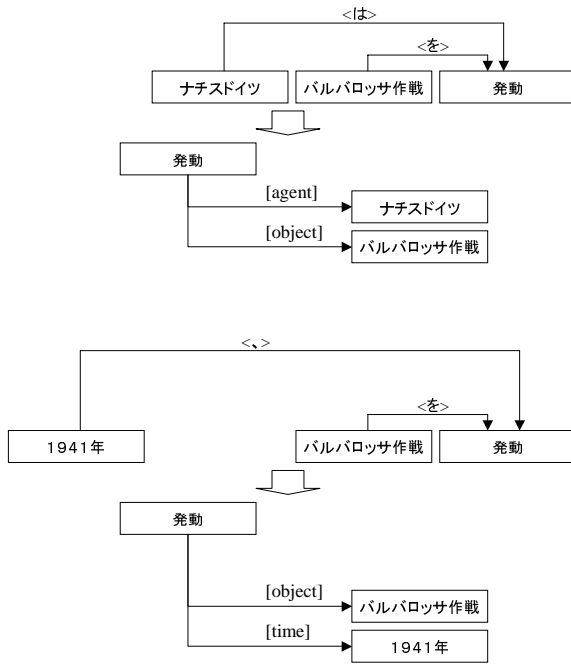


図 2: 二分依存関係への分解例

事例中に出現した単語であり，アークは連結される単語間の類似度を 0 から 1 の範囲で表す．二つの規則を比較して，必須条件部と任意条件部に対応する機能語がそれぞれ同一であり，対応する内容語が異なる場合に，その内容語の間にアークを張る．例えば，「ナチスドイツはバルバロッサ作戦を発動して」と「法務大臣は指揮権を発動して」からは，「ナチスドイツ」と「法務大臣」，「バルバロッサ作戦」と「指揮権」の間にそれぞれアークを張る．

張られたアークの類似度を決定する要因として，適用して推測される深層格と他の内容語の同一性を考慮する．深層格の同一性とは，条件部が同一ならば推測される深層格も同一となるはずであるという前提に基づくものであり，条件部の機能語が同一でありながら深層格が異なる場合，その原因は内容語の違いに帰着すると考えられる．したがって，深層格が同一の場合にはアークの類似度が強まる方向へ，異なる場合には弱る方向へ学習を行う．強まる方向を示す指標を  $ps$ ，弱まる方向の指標を  $ns$  と定義し，以下の式によりアークの類似度  $sim$  を計算する．

$$sim = \frac{ps + 1}{ns + 1} \quad (1)$$

分子と分母の+1 は 0.5 を基準値とするためのものである．深層格の同一性により類似度の変化の方向を求めるが，変化の程度に関しては他の内容語の同一性により求める．他の内容語の同一性も，深層格の同一性と同様の前提に基づくものである．二分依存関係には三つの内容語が含まれるが，全ての内容語が異なる場合よりも，一つの内容語のみが異なっている場合の方が，その違いが原因である可能性が高いと考えられる．したがって，同

一である内容語が多いほどアークの類似度の変化の度合いが大きくなるようにする．以上の考え方に基づき，変化の度合い  $ds$  を以下の式により計算する．

$$ds = \frac{mcw + 1}{3} \quad (2)$$

$mcw$  は同一である内容語の数であり，一つの内容語のみが異なる場合に  $ds$  は最大の 1 をとる． $ps$  および  $ns$  は両方とも  $ds$  の和で表現され，推測される深層格が同一となる規則から計算された  $ds$  の和を  $ps$  とし，深層格が異なる規則による  $ds$  の和を  $ns$  とする．

規則の適用は，条件部に対応する構文構造をもつ文に対して行われる．条件部に含まれる機能語が全て一致する場合，その規則を適用可能とする．ただし，任意条件部の構文構造が必須条件部の構文構造と連続している必要はない．また，一組の依存関係しか存在しない文に対しては，必須条件部と一致すれば適用可能であるとした．異なる深層格を推測する複数の規則が適用可能である場合，条件部に含まれる内容語の類似度に基づいた尤度計算を行い深層格を決定する．ある深層格  $d$  と推測する規則の中で適用可能な規則を抽出し，それぞれの規則に対して以下の式により尤度  $pl(d)$  を求める．

$$pl(d) = \sum_{r \in R} \prod_{w \in CW} scw(r, w) \quad (3)$$

$R$  は適用可能な規則の集合， $CW$  は規則の条件部に含まれる内容語であり， $scw$  は上述のネットワークを用いた内容語の類似度である．内容語の類似度は，対象となるノード間のアークを辿ることで計算される．目的となるノードに辿りつくまでの経路が  $i$  通り存在し，各経路におけるアークの数を  $na_i$ ，経路に含まれる各アークの類似度を  $sim_{i,j}$  として，以下の式で計算される．

$$scw = \arg \max_i \left( \frac{\prod_{j=1}^{na_i} sim_{i,j}}{na_i} \right) \quad (4)$$

#### 4. 今後の予定

発表では，文献 [1] と同じ EDR コーパスを用いた実験を行い，その結果を報告する予定である．

#### 参考文献

- [1] 峨家正樹, 荒木健治, 柘内香次: 帰納的学習を用いた言語に依存しない意味解析手法の評価, 情報処理学会研究報告, NL-146-12 (2001)
- [2] 大石亨, 松本裕治: 格パターン分析に基づく動詞の語彙知識獲得, 情報処理学会論文誌, Vol.36, No.11, pp.2597-2610 (1995).
- [3] 原田実, 水野高宏: EDR を用いた日本語意味解析システム SAGE, 人工知能学会論文誌, Vol.16, No.1 (2001)
- [4] 河原大輔, 黒橋禎夫: 用言と直前の格要素の組を単位とする格フレームの自動構築, 自然言語処理, Vol.9, No.1, pp.3-19 (2002).