E-24

未知文の意味解析における類推と規則の融合解析手法
A Method for Hybrid Analysis Based on Analogy and Rules in Semantic Analysis for Unknown Sentences

渋木 英潔[†1] 荒木 健治[†2] 桃内 佳雄[†3] 栃内 香次[†4]
Hideyuki Shibuki Kenji Araki Yoshio Momouchi Koji Tochinai

1．はじめに

自然言語の解析において，限られた知識を用いて多様な言語表現を解析することは困難な問題の一つである．この問題を解決するために，比較的少数の規則を用いた規則的な解析と類推を用いた柔軟な解析を融合した類推規則融合法を提案する．融合解析に関する従来研究[1]は，複数の知識から得られる情報を利用して精度を向上させることを目的としたものであり，限られた知識から可能な限り多くの有用な情報を引き出すことを目的としたものではない．複数の知識を利用するという戦略は精度向上の観点からは有効であるが，複数の知識から情報が得られなかった場合に精度を向上させることはできない．さらに，いずれの知識からも情報が得られないような文，すなわち未知の文に対する解析が可能になるわけではない．本稿で提案する類推規則融合法は，人手で作成された規則などを用いずコーパス中の文を唯一の情報源として解析を行う．文を唯一の情報源として用いることで，単純に文を追加するだけでシステムの構築，改良が容易になる利点が得られる．コーパス中の文を利用して解析を行う従来研究には，規則性を確率で表現する統計ベースの解析，文を用例としてそのまま保持し解析時に類推する用例ベースの解析，文を一般化した規則を保持し解析時に規則をそのまま適用する学習型規則ベースの解析などが存在する．類推規則融合法は，用例ベースと学習型規則ベースを融合した手法であり，文を一般化した規則を保持し解析時に規則から類推することで解析を行う．文を唯一の情報源とし，有用な情報を引き出すことを目的として規則と類推を融合する点が本手法の新規性である．同一の文を情報源としながら，類推規則融合法の性能がそれぞれの解析を単独で用いた場合の性能よりも優れていることを3章の実験で確認する．

本稿では，二文節間の依存関係が明示された文に対して`AGENT'などの意味関係を割り当てる意味解析に類推規則融合法を適用する．我々は，以下の理由から意味解析を応用システムと分離して行うべきであると考えている．意味解析などの基礎技術は，様々な応用システムに共通した前処理として行われるため，分離することで汎用的な処理を行うことができ，保守管理が容易になる．応用システムにおける前処理として利用されることから，意味解析の結果が出力されない場合，それ以降の処理が全て失敗することを意味する．それゆえ，意味解析の結果は全ての場合に必ず出力されるべきであると考える．勿論，応用システムの中には被覆率よりも精度が重視されるものが存在する．しかしながら，意味解析結果を利用するかどうかの判断は応用システム側の問題である．意味解析側では，応用システム側でそのような判断ができるように，その出力結果に対する解析の信頼度，すなわち尤度を与えることができれば問題ないと考えられる．また，本手法は日本語を処理対象としている．

†1 北海学園大学ハイテク・リサーチ・センター
†2 北海道大学大学院工学研究科
†3 北海学園大学大学院工学研究科
†4 北海学園大学大学院経営学研究科

2．意味解析

意味解析の精度を実用的なレベルまで高めることは容易ではなく，そのための研究は本研究と別に行うべきである．本研究では，全ての文を正しく解析できるわけではないが，意味解析における従来研究を考慮し，以下の考え方に基づいて規則ベース，用例ベース，類推規則融合法の意味解析を行う．

規則ベースの解析では，一般に選択制限の考え方に基づいた規則を用いて行われる[1]．例えば，係り元の内容語，係り先の内容語，機能語に対する選択制限がそれぞれ，[place]，[move]，「…に～」や「…へ～」を包含するクラスタ C_1 である場合，その意味関係は`GOAL'であるという規則があるとする．「学校に行く」という文を解析する場合，「学校」，「行く」，「に」はそれぞれ[place]，[move]，$C_1$ の一種であることから，この規則を適用することができ，意味関係を`GOAL'と決定することができる．ある単語がある概念に属しているかどうかの判断はシソーラスに基づいて行う．紙面の制約により詳述は文献[2]で述べているが，規則の獲得は文中の単語をシソーラスを用いて一般化することにより行われる．また，尤度の求め方に関しても文献[2]で述べている．

用例ベースの解析手法[3]では，一般に入力文と類似した用例の解析結果を出力するという原理に基づいて行われる．例えば，「学校に行く」という文を解析する場合において，「学校に行く」と類似した用例として「病院へ向かう」という文が存在すると仮定する．このとき，「病院」の意味関係が「向かう」の`GOAL'であることが判明しているならば，「学校」の意味関係も「行く」の`GOAL'であると推測できる．二つの単語 w_1，w_2 の類似度 sim は文献[1]と同じ式(1)に従って計算される．

$$sim(w_1, w_2) = 2 \times depth(cw)/(depth(w_1)+depth(w_2)). \quad (1)$$

ここで，$depth(w_1)$はシソーラスにおける深さでルートから w_1 までのノードの数で表される．cw は w_1 と w_2 の共通上位概念の中で最も深い位置に存在するノードである．

類推規則融合解析では，最初に規則ベースの解析と同様に入力文に適用可能な規則を探す．もしも，適用可能な規則が存在しなかった場合，用例ではなく，規則を利用して類推を行う．ここが用例ベースの解析における類推と異なる点である．規則からの類推は，規則の選択制限を文とみなすことで行う．例えば，上の例であげた規則と用例が存在する状況で「彼に渡す」という文を解析する場合を仮定すると，規則を「[place]C_1[move]」という

表1　実験結果

	被覆率　(度数)	精度　(度数)
規則ベース	3.8% (　75)	29.3% (　22)
用例ベース	100.0% (2,219)	28.8% (639)
類推規則融合法	100.0% (2,219)	30.4% (675)

並びの文とみなし，「彼」と[place]，「渡す」と[move]，「に」と C_1 の類似度を基に規則との類似度を計算する．入力文の単語も規則の概念も同じシソーラス上に存在するので，式(1)を用いて類似度を計算することができる．仮に，入力文中の単語が未知語であるならば，その単語をシソーラスのルートとみなして類似度を計算する．これは未知語に対する特定の概念の影響を避けるためであり，ルートが最も一般化された概念を表すノードであるからである．

3. 実験

我々は前章で述べた解析手法を用いて，EDR コーパスを対象にした実験を行った．それぞれの実験には同じトレーニングデータとテストデータを用い，各データは以下のように作成した．まず，EDR コーパスからランダムに 1,000 文を選択した．その半分の文をトレーニングデータとし，残りの文をテストデータとした．トレーニングデータとテストデータは互いに独立しており，学習や類推を用いることなしに字面上で直接一致する依存関係は一つも存在しなかった．トレーニングデータに含まれる依存関係の数は 4,121 で，テストデータは 4,361 であった．また，獲得された規則の数は 250 であった．

評価基準として，被覆率 cov と精度 acc を使用し，それぞれ以下の式で計算される．

$$cov = acceptable_d / total_d \times 100. \quad (2)$$

$$acc = correct_d / acceptable_d \times 100. \quad (3)$$

ここで，$total_d$ は全ての依存関係の数，$acceptable_d$ は解析できた依存関係の数，$correct_d$ は EDR コーパスの概念関係子と解析結果が一致した依存関係の数である．複数の解析結果が存在する場合には，尤度が一位の結果を用いて評価した．EDR コーパスの構文解析結果の中には，意味解析結果に現れない依存関係が存在する．そのような依存関係の概念関係子は不明なので，評価の段階でそれらを無視することとした．また，本手法では依存関係にある文節は全て後方依存と仮定して解析を行ったため，前方依存である依存関係も評価対象から除外した．その結果，評価した依存関係の数は 2,219 であった．

表 1 に結果を示す．括弧内の数字は依存関係の数である．この結果は類推規則融合解析が他の二つの解析よりも有効にトレーニングデータを活用したことを示しており，同じトレーニングデータから異なる結果が導き出されたことが分かる．

規則ベースの 3.8%という被覆率に対し，100%という被覆率から，類推規則融合解析と用例ベースの解析が，類推を用いて全ての文を柔軟に解析するという目的を達成していることが分かる．

精度において，類推規則融合解析の結果(30.4%)と規則ベースの解析結果(29.3%)が，用例ベースの解析結果(28.8%)よりも優れていることが示された．この理由として，規則を用いた解析が，類推による解析と異なり，決定的に意味関係を解析できたことが挙げられる．用例中の単語が規則中の概念に一般化されるとき，その概念に用例中の単語以外の単語が含まれる可能性がある．その含まれる単語が正しいならば，類推規則融合解析と規則ベースの解析は，その単語を含む文に対しても正しく決定的に解析することができる．しかしながら，用例ベースの解析では，必ずしもその規則を獲得する元となった用例が最も類似した用例として選ばれることを保証しない．何故ならば，類推の際には，その単語だけではなく，入力文中の他の単語による影響も受けるからである．この違いが類推規則融合解析との精度の差(1.6%)と規則ベースの解析との精度の差(0.5%)となって現れたと考えられる．

類推規則融合解析の精度が規則ベースの解析の精度よりも優れていた理由について考察する．どちらの解析においても規則を直接適用できた依存関係の数は等しいので，類推規則融合解析は類推を用いて 653 の依存関係を正しく解析したことになる．この数は，用例ベースの解析結果である 639 よりも多い．これは，未知の文に対する類推規則融合解析の類推が用例ベースの解析による類推よりも優れていたことを示している．未知語の場合，その単語の概念はシソーラスのルートとみなされ，類推規則融合解析においてはルートと規則中の概念を用いて類推を行う．規則中の概念が多くの単語を含むほど，その概念はルートとの距離が近くなり類似度も高くなる．それゆえ，類推規則融合解析は多くの単語を含むような概念をもつ規則を用いて未知の文を解析することになる．そして，その結果，具体的な用例から類推した結果よりも精度が向上した．このことは，未知の文を解析する際に，具体的な用例を用いて類推するよりも適度に一般化された規則を用いて類推する方が有効であることを示している．

類推規則融合解析では，部分係り受け解析などで主張されているような被覆率とのトレードオフにすることなく精度を向上させることができ，限られた文を用いて未知の文を解析する状況において有効であると考える．

4. おわりに

限られた文を用いて未知の文を有効に解析するために，規則ベースと用例ベースを融合した類推規則融合法について提案した．EDR コーパスに対して，類推規則融合解析，規則ベースの解析，用例ベースの解析を用いて二文節間の意味関係の割り当て実験を行った結果，類推規則融合解析が最も高い値を示した．これにより，類推規則融合解析が限られた文を用いて未知の文を解析する状況において有効に働くことを実験により確認した．

謝辞　本研究の一部は，北海学園大学ハイテク・リサーチ・センター研究費による補助のもとに行なわれた．

参考文献

[1] Kurohashi, S. and Sakai, Y.: Semantic Analysis of Japanese Noun Phrases : A New Approach to Dictionary-Based Understanding, *Proc. 37th ACL*, pp.481-488 (1999).

[2] 渋木英潔，荒木健治，栃内香次: 帰納的学習を用いた括弧つきコーパスからの意味解析規則の自動獲得手法の性能評価，信学技報，NLC2001-32，(2001).

[3] Al-Adhaileh, M.H. and Kong, T.E.: A flexible Example-based parser based on the SSTC, *Proc. 36th ACL*, pp.687-693, (1998).