

3C-4

一般化弁別ネットワークを用いた日本語解析

奥村 学, 秋葉友良, 伊藤克亘, 田中穂積

(東京工業大学工学部)

1 はじめに

弁別ネットワークは問題解決に広く用いられている表現形式であり、特に、自然言語処理においては、複数の語義を圧縮して表現するのに用いられている。弁別ネットワークには次のような利点がある。一つは、(複数の可能な解が存在する)曖昧性を表現する際に、選択肢として複数の候補を持つのではなく、(それらをすべて下位ノードとして包含する)一つの曖昧な表現(ノード)として持つことができることである。すると、ネットワークを下向きに一段ずつたどる過程は、制約プログラミングが目的とする解の増進的洗練過程と考えられる。二つ目は、線形探索と比較して、弁別ネットワーク上での探索アルゴリズムは効率がよいことである。これは、解の探索が葉ノードに位置するそれぞれの解に向けて根ノードからネットワークを下向きにたどる操作になり、この場合、枝に記述されている制約を索引として探索するので、探索空間が徐々に小さくなるからである。三つ目は、規則集合を圧縮して表現できることである。なぜなら、異なる規則中の同じ前提条件(制約)がネットワーク中の一つの枝としてまとめて表現できるからである。

弁別ネットワークにはこのように優れた特徴があるが、同時に二つの問題点がある。一つは、弁別ネットワークは予め決められた順序で制約が入力されないとたどることができない点である。制約は一般に決まった順序で得られるとは限らないので、ネットワークを下向きにたどるプロセスは、決められた順序で次に入力されるはずの制約が得られるのをしばしば待つ必要がある。二つ目の問題点はより深刻である。ネットワークを下向きにたどる際、次に入力されるはずの制約が得られない場合には、問題解決プロセスはもはやネットワークを下向きにたどることができず、デッドロックに陥ってしまう。

このような問題点を解決するため、任意の順序で制約が得られてもたどることができる、一般化弁別ネットワーク(Generalized Discrimination Network;GDN)を我々は提案している[4]。問題解決プロセスは、制約がどんな順序で得られても即座にGDNをたどることができる。

本稿では、GDNの日本語解析への応用について述べる。GDNを用いることにより、語義や係り受けの曖昧性解消、省略の補完処理がどのように行なわれるかについて述べる。

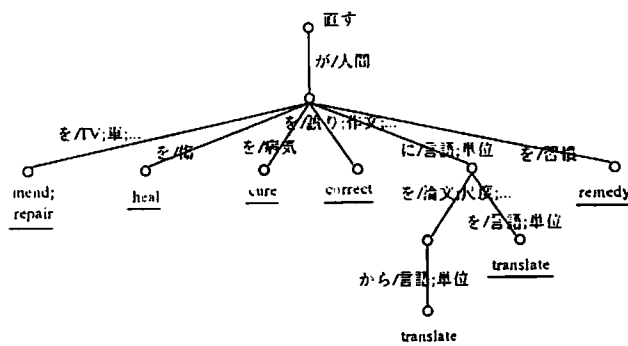


図1: 動詞「直す」の語義弁別ネットワーク(一部)

2 GDNを用いた日本語解析

図1に動詞「直す」の語義および格フレームを表現した弁別ネットワークの一部を示す。弁別ネットワークの枝には、動詞の取りうる表層格とその格を満たしうる名詞句に関する選択制限の組が記述される。ネットワークの葉ノードは曖昧性のない1つの語義を表す。葉ノードに付与されたラベル(下線を付したものはその語義を表すものとする。葉ノード以外のノードは、下向きにさらに枝をたどることで、複数のノード(語義)に到達しうることから、到達しうる語義をすべて含んだ曖昧な意味を表現していると考えられる。根ノードから葉ノードまでのパス中にある制約の集合が、その葉ノードの語義に対応する格フレームを表現する。

弁別ネットワークを用いた意味的曖昧性解消は、得られた情報が満足する制約の枝に沿って弁別ネットワークを根ノードから葉ノード(曖昧性のない語義)に向けて下向きに一段ずつたどる過程であると考えられる。この過程で、文中の回りの単語の情報(動詞が取る格要素の情報)から、異常な選択肢は排除され妥当な動詞の語義が弁別(選択)される。葉ノードに到達することが曖昧性が完全に解消されたことを意味する。

意味的曖昧性解消過程を、弁別ネットワークを下向きにたどる過程として実現する研究として、[2]などがある。これらの研究では、弁別ネットワークを意味的曖昧性解消に用いることの利点として、1節で挙げたようなものが述べられている。

弁別ネットワークの代わりにGDNを用いることにより、意味的曖昧性解消を増進的に行なうことができるといふ、新たな利点が付加される。文を解析する過程で得られる制約を利用して、できるかぎり早期の段階で曖昧性を段階的に解消していかなければ、曖昧性の数(探索空間の大きさ)は組合せ的に爆発してしまうことを考えると、増進的曖昧性解消手法

Japanese Analysis with a Generalized Discrimination Network  
OKUMURA Manabu, AKIBA Tomoyoshi, ITOU Katsunobu,  
TANAKA Hozumi  
Department of Computer Science, Tokyo Institute of Technology

[3] は曖昧性解消のために望ましい方法であると考えられる。

しかし、増進的に得られる制約の順序はあらかじめ定められた通りであるとは限らず、また、次に入力されるはずの制約が得られない場合もあるため、解析過程で弁別ネットワークを上から下に順々にたどれる保証は一般にない。GDN は、増進的曖昧性解消手法を阻害している弁別ネットワークのこれらの問題点を解決し、解析過程で増進的に得られる制約を用いて、ネットワークを下向きにたどることを実現する。日本語は語順が比較的自由であるため、あらかじめ定められた順序からの逸脱の度合い、すなわち制約が得られる順序の非決定性がかなり大きい。さらに、日本語では、文脈から容易に推測可能な句が省略されることが多い。したがって、日本語解析で GDN の果たす役割は重要であると考えられる。

次に、GDN によりどのように省略の補完処理を行なうかについて述べる。GDN を用いた省略補完処理は、以下に述べるように、省略されている句の位置を検出したり、選択制限を用いて補う句の候補の妥当性を検査したりすることは容易に行なえるが、補う句の候補を発見したり、それらの間の優先性を文脈から決定したりするには他の機構が必要である。本稿では、「省略されている句に近いものを優先する」といった素朴なヒューリスティクスを仮定する。

省略された句の位置は GDN では 2 つの方法で検出される。if 節中の制約は得られていないが、得られるはずであることから、その位置を示していることになる。また、文全体の解析が終了した時点で、葉ノードに到達していない場合、到達しているノードと葉ノードの間にある制約は省略されているものとする。なぜなら、動詞の意味は本来曖昧性がない方が望ましいと考えられるからである。この場合、省略の補完処理を行なうことにより、語義の曖昧性をさらに解消することが可能になる。補う句の候補に、省略された句の位置の選択制限を適用することで、補う句の意味的妥当性が検査される。

最後に、GDN を用いてどのように増進的曖昧性解消と省略補完処理が実行されるかを例を用いて示す。次の文を考える。

太郎が日本語で論文を書き、  
英語に花子が直した。

例文中の「直す」の語義は図 1 に示したように曖昧である。また、制約の得られた順序は、[に/英語、が/花子] であり、ネットワークの本来の順序と異なる。さらに、「直す」の対象格の要素が省略されている。

最初の文が解析された時点で、省略された句を補う句の候補として、[論文、日本語、太郎] が得られる。ここでは、前にある要素の方が優先される。二番目の文は、次のように解析される。

「英語」は言語であり、「に」格の選択制限を満足する。よって、ノード 1 に到達できる。次に、「花子」が「が」格の選択制限を満足し、その結果、ノード 1 の到達可能性が無条件になる。「直す」の語義の曖昧性は、7 個から 2 個に減少するが依然曖昧である。文の解析が終了したが、到達したノードが葉ノードではないので、省略補完処理を試みる。そして、省略された句の位置の候補として、{を/言語; 単位}、あるいは {を/論文; スピーチ; 属性、から/言語; 単位} を得る。補う句の候補として上のように得られたものから、「論文」、「日本語」がそれぞれ省略された句の候補である「を」、「から」格

の選択制限を満足するので、省略補完処理が成功し、「直す」の語義は 1 つに決定される。二番目の文は最終的に、「花子が(太郎が書いた)論文を日本語から英語に翻訳した」を意味していると解釈される。

さらに、GDN を用いることで、係り受けの曖昧性も増進的に解消することが可能になる。GDN を用いて増進的に係り受けの曖昧性を解消する手法の詳細は、[1] に譲る。ここでは、GDN を用いた係り受けの曖昧性解消の利点として、異常な係り受けを(従来より)早期に排除できることについてのみ述べる。弁別ネットワークでは、上述したように、一つの格フレーム中で共起する名詞句の並びを根ノードから葉ノードまでのパス中の枝として表現している。このため、係り受けの解析の際に、今得られた名詞句がそれまでに得られた名詞句と同じ動詞に係るとして解析する場合は、(過去に得られた名詞句の情報を用いて)現在到達しているノードから、今得られた名詞句の情報を用いてさらにネットワークをたどることに対応する。したがって、現在到達しているノードから、今得られた名詞句の情報を用いてネットワークをさらにたどることができない場合、それまでに得られた名詞句と今得られた名詞句を同じ動詞に係るとする係り受けの候補は排除することができる。このような場合、係り受けの曖昧性が部分的に解消され、それらが別々の動詞に係るとする係り受けの候補のみが残ることになる。ここで、この解析(到達しているノードからネットワークをさらにたどること)は、動詞が出てくる以前に行なえることに注意して頂きたい。係り受けが名詞句と動詞の関係であることから、係り受けの曖昧性解消は従来動詞が現れて初めて実行されることが多い。それに対し、GDN を用いることで、名詞句が次々に現れた過程で動詞が現れるのを待つことなく、増進的に係り受けの曖昧性解消を行なうことができる。

## 参考文献

- [1] 秋葉友良, 伊藤克亘, 奥村学, 田中穂積. 増進的曖昧性解消モデルに基づいた統合的日本語解析. 「自然言語処理における統合」シンポジウム論文集, 1991. (予定).
- [2] S.L. Lytinen. Are vague words ambiguous? In S.L. Small, G.W. Cottrell, and M.K. Tanenhaus, editors, *Lexical Ambiguity Resolution: Perspectives from Psycholinguistics, Neuropsychology, and Artificial Intelligence*, pp. 109-128, Morgan Kaufmann Publishers, 1988.
- [3] C. S. Mellish. *Computer Interpretation of Natural Language Descriptions*. Ellis Horwood, 1985.
- [4] M. Okumura and T. Hozumi. Towards incremental disambiguation with a generalized discrimination network. In *Proc. of the 8th National Conference on Artificial Intelligence*, pp. 990-995, 1990.