

制限付き疑似ベイジアンネットを用いた組合わせ範疇文法パーザ

高橋 直人

一杉 裕志

国立研究開発法人 産業技術総合研究所 人工知能研究センター

{naoto.takahashi, y-ichisugi}@aist.go.jp

1 はじめに

失語症の症例研究や、fMRI を使った観察などから、ヒトの文法処理能力には大脳皮質の一部(いわゆる言語野)が強く関与していることは間違いない [5]。また近年の計算論的神経科学においては、大脳皮質がベイジアンネット的計算を行っているという仮説が注目を集めている [6, 2]。ヒトの文法処理能力が大脳皮質において実現されており、かつ大脳皮質がベイジアンネット的計算を行っているのならば、ベイジアンネットを用いてヒトの文法処理能力を再現することが少なくとも原理的には可能であるはずである。

筆者らはこの考えに基づき、通常のベイジアンネットや、制限付き疑似ベイジアンネット [3, 4, 9] を用いて、文脈自由文法や日本語係り受けのパーザの実装を試みてきた [8, 10]。

本稿では、制限付き疑似ベイジアンネットを用いた組合わせ範疇文法パーザに関して新たに報告する。

2 制限付き疑似ベイジアンネット

言語能力などの高次認知機能をベイジアンネット等の機械学習技術を用いてモデル化する際には、たとえ小規模なプロトタイプであっても、ハイパーパラメタの調整など非本質的な作業を伴う点が問題となる。また実装上の問題や局所解・過適合の問題などが関係してくるため、意図した結果が得られない場合に、モデルが本質的に間違っているのか、それとも単なる調整不足なのかの切り分けが難しい。

そこで筆者らは、ベイジアンネットを用いてヒトの高次認知機能をモデル化する際のプロトタイプングツールとして、制限付き疑似ベイジアンネットを提案した [3, 4, 9]。制限付き疑似ベイジアンネットは、確率値の 0 と非 0 のみを区別するようにベイジアンネットを簡略化した上で、さらに条件付確率表に制限を加えることでパラメタ数の爆発を抑えたものである。制

限付き疑似ベイジアンネットには、信号の流れを制御するためのゲートと呼ばれる機構が備わっており、これによって論理回路的感覚で生成モデルを設定することが可能となっている。

制限付き疑似ベイジアンネットは、条件付確率表の値をデータから学習することはできないので、設計者が手で条件付確率表を記述する必要がある。その代わりに設計者は問題を解く上で非本質的な作業から解放され、問題の定性的な面・本質的な部分に専念することが可能となる。

3 組合わせ範疇文法

組合わせ範疇文法 [7, 1] は、古典的範疇文法に関数合成規則を取り入れたものである。組合わせ範疇文法の弱生成能力は文脈自由文法と文脈依存文法の間位置し、自然言語の文法記述に適していると考えられている。

通常の句構造文法では、個々の統語範疇に対して固有の非終端記号を割り当てる。たとえば文には S 、名詞句には NP 、動詞には V が割り当てられることが多い。

これに対し組合せ範疇文法では、基底範疇と演算子を用いて統語範疇を表現する。たとえば英語の動詞句は、その左側の名詞句 (NP) と組合わされて文 (S) となる。そのため S と NP を基底範疇とした場合、組合わせ範疇文法では動詞句を $S \setminus NP$ と表現する ($X \setminus Y$ は「左側の Y と組み合わせられて X となる統語範疇」を意味する。) さらに動詞とは、その右側の名詞句 (NP) と組合わされて動詞句 ($S \setminus NP$) となるものであるから、 $(S \setminus NP) / NP$ と表現できる (X / Y は「右側の Y と組み合わせられて X となる統語範疇」を意味する。)

通常の範疇表現

X/Y

$(X/Y)\backslash Z$

$X/(Y\backslash Z)$

固定長前置表現

/	Y	X	.	.
---	---	---	---	---

\	Z	/	Y	X
---	---	---	---	---

/	\	Z	Y	X
---	---	---	---	---

図 1: 組み合わせ範疇文法における一般的な範疇表現 (左) と、ベイジアンネットのノードによる固定長前置表現 (右) の対応。ピリオドはフィラーを表す。

4 実装

本節では、組み合わせ規則のうち、順/逆関数適用規則をベイジアンネットを実装する方法を説明する。

4.1 ベイジアンネットによる統語範疇表現

組み合わせ範疇文法における統語範疇表現は、理論上任意の長さが想定可能であるが、人間の情報処理能力の制約からここでは有限であると仮定する。今回は 1 範疇を 5 ノードの固定長で表現することとする。各ノードは、値として基底範疇あるいは演算子をとる。

まず括弧を消去するために、演算子は前置記法とする (中置記法による処理も別途検討中である。) 演算子の直後には元来演算子の右側に書かれていた統語範疇を記述し、続けて演算子の左側に書かれていた統語範疇を記述する。5 個 1 組のノードは左詰めで使用し、使われないノードには値として明示的にフィラーを入れる (図 1)。

4.2 順関数適用規則

順関数適用規則を実現する制限付き疑似ベイジアンネットを図 2 に示す。また、この場合に非 0 の確率で生じる値の組み合わせを表 1 に示す。

左統語範疇の先頭ノード (すなわち最上位の演算子) が / であり、かつそれに続く 3 ノードの値が右統語範疇の先頭 3 ノードとマッチした場合 (図 3) は、左統語範疇の最終ノードと関数適用後の統語範疇の先頭ノードを結ぶゲートが開き、両ノードの値が等しくなる。同時に関数適用後の統語範疇のうち、先頭ノード以外のノードはフィラーで埋められる。

また左統語範疇の先頭ノードが \ であり、かつそれに続く 1 ノードのみが右統語範疇の先頭ノードとマッチした場合 (図 4) は、左統語範疇の残り 3 ノードと

判定器

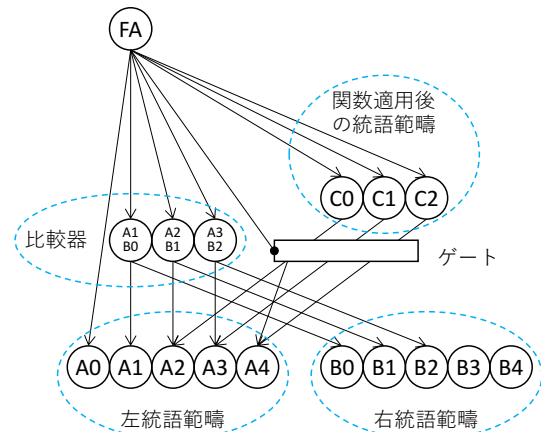


図 2: 順関数適用規則を実現する制限付き疑似ベイジアンネットの全体図。統語範疇を表現する各ノードは、基底範疇あるいは演算子を値としてとる。比較器は、2 個の子ノードの値が等しい場合はそれらと同じ値をとり、そうでない場合は否定記号としてフィラーと同じ値をとる。判定器は比較器の値に応じて、左統語範疇と関数適用後の統語範疇の間のゲートを閉鎖する。関数適用規則が適用された場合は統語範疇の表現が必ず短くなるので、この図では最初の 3 ノードのみを示している。

関数適用後の統語範疇のノードの値が等しくなるようにゲートが動作する。

4.3 逆関数適用規則・統語範疇数の拡大

順関数適用規則と同様の方法で逆関数適用規則を実現する制限付き疑似ベイジアンネットを図 5 に示す。順関数適用規則の左右を入れ替えて考えればよいので、説明は省略する。

図 2 と図 5 を重ね合わせたネットワークを作成することで、2 個の統語範疇間の関数適用可能性を判定することができる。そのように重ね合わせたネットワークを図 6 の上のように略記することになると、重ね合わせたネットワーク同士を合成することで同図下のようになり、3 個の統語範疇間の関数適用可能性を判定できるようになる。

3 個の統語範疇間で関数適用が可能であるためには、F と G の少なくとも一方で関数適用が成功しなければならず、そのためには D と E の少なくとも一方で関数適用が成功している必要がある。実際に制限付き疑似ベイジアンネットを用いて図 6 下のネットワークを作成したところ、関数適用可能なすべての入力パターン

FA	A0	C0	C1	C2	A1B0	A2B1	A3B2	C0A2	C0A4	C1A3	C2A4
C0	/	X	.	.	/,\	X	X	切	入	切	切
C1	/	/,\,X	X,.	X,.	X	.	-	入	切	入	入

表 1: 図 2 において非 0 の確率で生じる値の組み合わせ. X は任意の基底範疇 (列ごとに違っていてもよい), ピリオドは明示的なフィラー, あるいは「マッチせず」を意味する否定記号, アンダースコアは任意の値を意味する. 最後の 4 列はゲートによるノード間の連結・非連結を示している. FA の値は各行を区別できるものであれば何でもよい.

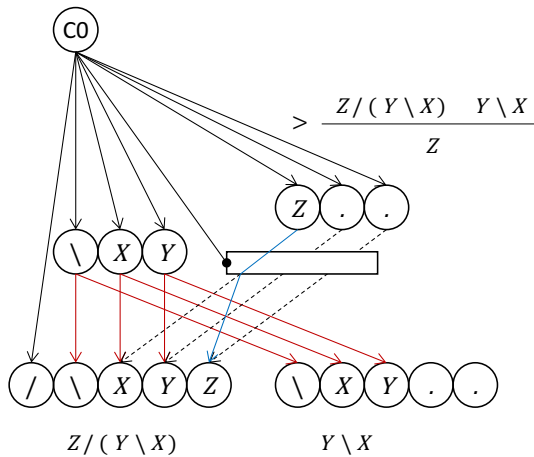
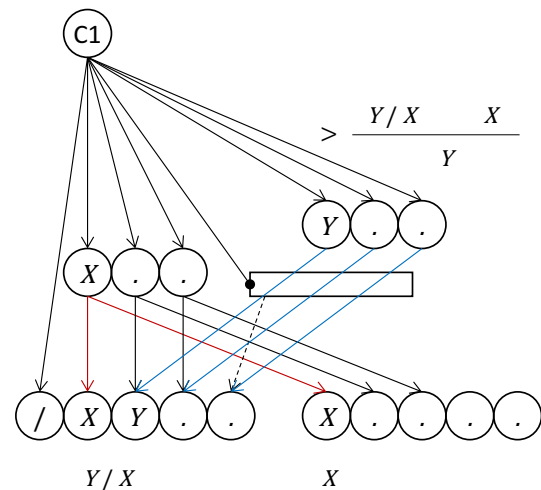


図 3: 順関数適用規則において左統語範疇の一部と右統語範疇の先頭 3 ノードがマッチした状態. 左統語範疇の最終ノードと関数適用後の統語範疇の先頭ノードが等しくなるようにゲートが動作する.



において関数適用が成功し, 適用不可能なパターンは解に含まれないことが確認できた. 4 個以上の統語範疇に対しても, 同様の実装が可能であると考えられる.

5 おわりに

本稿では, 制限付き疑似ベイジアンネットを用いた組み合わせ範疇文法パーザの実装可能性を検討した. 現段階では, 数種類ある組み合わせ規則のうち関数適用規則のみが実装できたに過ぎない. また受理可能な統語範疇の数も非常に小さく制限されている. 実用的なパーザに発展させるためには, 現在の実装方針の下で受理可能な統語範疇を増加させたときに, ネットワークのパラメータ数が指数関数的に増大しないことを確認する必要がある.

また, 組み合わせ範疇文法においては, 組み合わせ規則によって統語導出と意味合成が同時に行なわれるというエレガントな特徴があるので, その特徴を活かした意味表現方式を考案する必要もある.

制限付き疑似ベイジアンネットは確率が非 0 になる

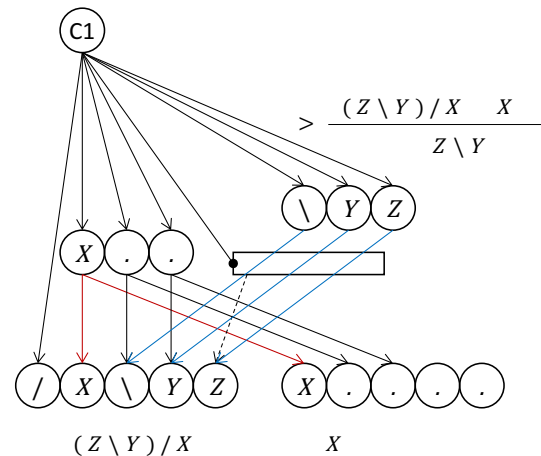


図 4: 順関数適用規則において左統語範疇の一部と右統語範疇の先頭 1 ノードのみがマッチした状態. 左統語範疇の最終 3 ノードと関数適用後の統語範疇が等しくなるようにゲートが動作する.

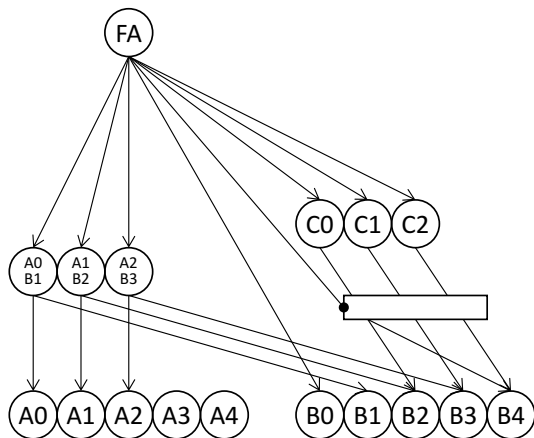


図 5: 逆関数適用規則を実現する制限付き疑似ベイジアンネットの全体図 .

組合せを全解探索で求めるが、真のベイジアンネットは一番もってもらいし組合せを高速に近似計算することができる。ヒトは文を解析する際に意識に上る形での全解探索は行っていない上、深い中央埋め込み文などの処理に失敗することから、やはり何らかの近似計算を用いることで実時間での解析を実現している可能性が高い。制限付き疑似ベイジアンネットによるプロトタイピングで方向性を定め、真のベイジアンネットを用いてヒトの言語能力をモデル化することが筆者らの目標である。

謝辞

この成果は、国立研究開発法人新エネルギー・産業技術総合開発機構 (NEDO) の委託業務の結果得られたものである。

参考文献

- [1] 戸次大介. 日本語文法の形式理論. くろしお出版, 2010.
- [2] 一杉裕志. 大脳皮質とベイジアンネット. 日本ロボット学会誌, Vol. 29, No. 3, pp. 412 – 415, 2011.
- [3] 一杉裕志. 疑似ベイジアンネットを用いた認知モデルのプロトタイピング手法の提案. 人工知能学会研究会資料 SIG-AGI-004-01, 2016.
- [4] 一杉裕志, 高橋直人, 尾崎竜史. 大脳皮質の計算論的モデルを用いた組み合わせ範疇文法パーザ実装の構想. 第 23 回言語処理学会年次大会, 2017.

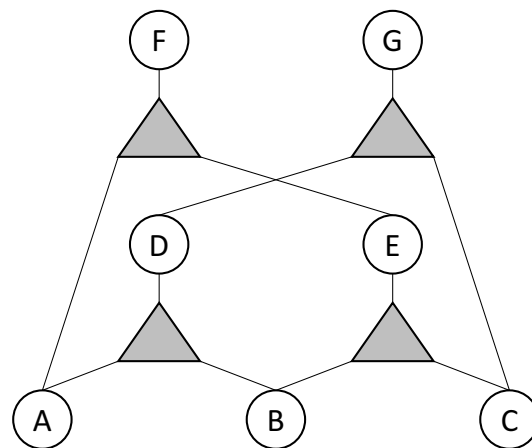
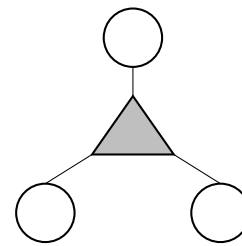


図 6: 図 2 と図 5 を重ね合わせた場合の略記法 (上) と、それを用いて 3 個の統語範疇間の関数適用可能性を判定するネットワーク (下) .

- [5] David Kemmerer. *Cognitive Neuroscience of Language*. Psychology Press, 2015.
- [6] Tai Sing Lee and David Mumford. Hierarchical bayesian inference in the visual cortex. *Journal of Optical Society of America A*, Vol. 20, No. 7, pp. 1434–1448, 2003.
- [7] Mark Steedman. *The Syntactic Process*. MIT Press, Cambridge, MA, USA, 2000.
- [8] 高橋直人. ベイジアンネットを用いた疑似日本語の係り受け解析. 第 30 回人工知能学会全国大会, 2016.
- [9] Naoto Takahashi and Yuuji Ichisugi. Restricted quasi bayesian networks as a prototyping tool for computational models of individual cortical areas. In *Proceedings of Machine Learning Research*, Vol. 73. PMLR, 2017.
- [10] 高橋直人, 一杉裕志. 制限付き疑似ベイジアンネットを用いた文脈自由文法解析の試み. 第 31 回人工知能学会全国大会, 2017.