

チャートを用いたトップダウンによる文生成

4 A E - 9

小林伸嘉 牧野武則
東邦大学大学院理学研究科

1. はじめに

文生成は文解析と表裏の関係にあり、簡素で見通しのよいアルゴリズムが求められている。Martin Kay によって提案されたチャート法[1][2]は、文脈自由文法を受理するパーザとして知られ、頑強なことから実用システムにおいて使用されている。

文生成への適用は M.Kay 本人によって提案された[3]。この戦略にはボトムアップが用いられ、各語に相当するアークを独立に作成し、これらを結合していくことによって文の生成を行うというものであった。

しかし、文の大局的な生成を行う場合はトップダウンに生成する方が戦略を立案しやすい。

この論文ではチャート方を用いて文の開始記号からトップダウンに文を生成する手法を提案する。

2. 生成への応用

今回提案する生成プロセスでは開始記号、すなわち一文を表す一本のアークにその文におけるメインとなる動詞のアークを当てはめ、そこから適合しうるルールに沿ってノードを分割してアークを結んでいくことで、各語を繋いでいく。つまり、トップダウン型での生成を行う。

2.1 フラットな論理表現

生成における入力としては、Copestake らによって提案されたフラットな論理表現と呼ば

Top-down generation with chart.

Nobuyoshi Kobayashi

Takenori Makino

Department of Information Sciences,

Faculty of Science, Toho University

2-2-1 Miyama, Funabashi, Chiba, Japan 274

れるものに若干の記述変更を加える。この論理表現は一階述語論理に近い構造を持つ表現であり、基本として述語記号と、頂点と呼ばれる個体変数及び個体定数からなる。

ここで新たにメイン述語について次の例のような記述を行う。

(1) give1(Vmain, S | O1O2)

このとき(1)は文中における唯一の生成の基点である。この動詞を示す頂点を Vmain で表わし、これは当然一文を表わす論理表現において、複数の動詞にあってはならない。動詞を表わす頂点に加え、その前後に来る句へリンクされた頂点を与えることで一つの動詞を表わす。

2.2 生成プロセス

今回は簡単な例としてつぎのような論理表現の文生成の過程を述べる

(2) dog(d), cat(d), def(d), def(c), past(s),
see(s, d | c)

この例は“The dog saw the cat”の論理表現として与えている。この場合、動詞 see がメイン述語であり、Vmain に相当する頂点は s となる。

これらの論理表現から、該当しうる語を辞書から検索することで、文生成に使われる語を得る。今回の例ではこの検索結果が次のようになったものとする。

(3)

語	品詞分類	論理表現
dog	n(x)	dog(x)
cat	n(x)	cat(x)
the	det(x)	def(x)
saw	v(x, y, z)	see(x, y z), past(x)

これらの結果に対し、次のような文法規則を左から右へ適用していくことで文生成を行う。

- (4) $s(x) \rightarrow np(y) vp(x,y)$
- (5) $vp(x,y) \rightarrow v(x,y,z) np(z)$
- (6) $np(x) \rightarrow det(x) n(x)$

3. Top-down での生成過程

先にも述べたとおり、トップダウン生成では、文のメイン述語を補う動詞“saw”のアーキ、そしてその上に文全体を示すアーキを引いたものが初期状態である。つまりこの場合、図1に示すように $v(s,y,z)$ と $s(s)$ を示すアーキからのスタートとなる。

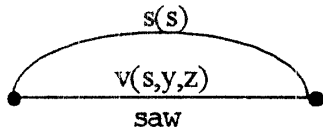


図1 初期状態のアーキ

- 実際の生成過程を見ていくと次のようになる。
- まず(4)の規則から、 $np(y)$ 及び $vp(s,y)$ が生成される。このうち、(5)の規則により $v(s,y,z)$ への到達可能なものは $vp(s,y)$ なので、 $vp(s,y)$ のアーキは $v(s,y,z)$ のアーキを覆うかたちとなり、この左側に $np(y)$ のアーキが引かれる。この時、(3)の一番下の論理表現より、 $v(s,y,z)$ の左側にくる頂点は y は d に相当する。
- さらに(6)の規則によって $np(d)$ は $det(d)$ と $n(d)$ に分割される。
- $v(s,d,z)$ の右側に関しては(5)の規則から $np(z)$ が生成され、 $v(s,d,z)$ の右側は c であり、 z は c に相当する。
- 同様に(6)の規則によって $np(c)$ は $det(c)$ と $n(c)$ に分けられる。
- このときの品詞の並びをみると、 $det(d)-n(d)-v(s,d,c)-det(c)-n(c)$ 。すなわち、以上で“The dog saw the cat.”の文生成がなされた。

以上のプロセスによって出来上がった全てのアーキを図2に表す。

なお、終了の判定はそれぞれの語に関連した論理形式が全て使われているかによって検証す

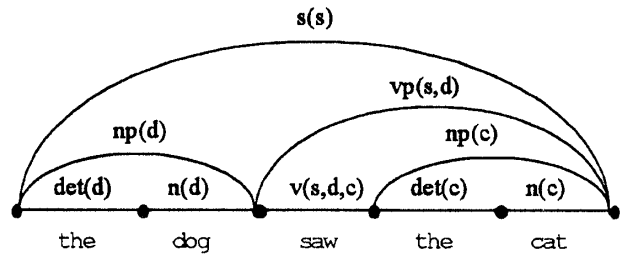


図2 top-down によって引かれたアーキ

る。これが重複したり、不足していた場合はさらに深くまで探索、生成を行うか、もしくは別の探索ルートでの生成を行う。

4. おわりに

論理形式からの文生成へのアプローチとしてトップダウンでのチャート生成の試みを行った。

従来のボトムアップによる生成では、初期状態において各語を示すアーキが独立した状態で存在し、そこからそれぞれが別々に接続を試みるため、生成までのプロセスが複雑になりがちであった。これに対し、トップダウン方式においては、メインの動詞を基点とした規則適用をおこなうため、文に展開する手順を明確に表わすことができる。

なお、今後の課題としては、この戦略を用いたシステムの構築と、サンプルを用いての検証と評価を行う。

参考文献

- 1) 牧野：図解自然言語処理、オーム社(1991)
- 2) 田中、辻井：自然言語理解、オーム社(1988)
- 3) Kay, M. (1996): Chart Generation, 34th Annual Meeting of the Association for Computational Linguistics.
- 4) Brill, E. and S,giorgio(1996): Efficient Transformation-based Parsing, 34th Annual Meeting of the Association for Computational Linguistics.