

重み付き意味ネットワークによる文章生成

日大生産工(院)

佐藤 裕介

日大生産工

松田 聖

1 はじめに

文を完全に生成することはとても困難なことである。だが、文の生成は早くから様々なアプリケーションシステムで実用的に用いられてきた。これは少しぐらい不自然な文であっても利用者がその内容を理解できていたからである。しかし計算機の発達により、より自然な文章を生成することが期待される。その中で、意味ネットワークからの文章生成が試みられている¹⁾。本研究では、意味ネットワークを拡張した重み付き意味ネットワークを使用した文章生成法を提案する。

2 重み付き意味ネットワーク

重み付き意味ネットワークとは、意味ネットワークの arc に対して重みと呼ばれる値を負荷した知識表現手法である²⁾。今までの重み付き意味ネットワークでは node に名詞、arc に動詞が割り当てられていた。しかし、動詞と名詞のみでは扱える知識に限界があり、文章生成をする上では力役不足である。そのため、本研究では node を体言 node と用言 node の 2 種類に分別する。これにより、動詞と名詞のみしか扱えなかった重み付き意味ネットワークが副詞や形容詞といった品詞も扱うことができる。また、arc として扱っていた動詞を node として扱うことで、文法を表現することや、文の構造を解析することが出来ると予想される。体言 node とは、物事を表わす node のことであり、名詞代名詞が割り当てられる。用言 node とは、動作様態を表わす node であり、動詞形容詞副詞等が割り当てられる。体言 node と用言 node がつなぐ、arc には体言 node 同士をつなぐ体言-体言 arc、体言 node と用言 node をつなぐ体言-用言 arc、用言 node 同士をつなぐ用言-用言 arc の三種類がある。

体言-体言 arc

体言-体言 arc とは、体言 node 同士をつなぐ arc である。体言-体言 arc に負荷された重みが表わすものは体言 node 同士の関連性の強さを表わしているといえる。また体言-体言 arc には重みとは別に、is-a, has-a, pert-of と言ったラベル付けがされる。

用言-用言 arc

動詞句のように用言同士が結合して始めて意味をなすような関係がある。また、副詞や助動詞のように用言を補佐する役割がある語が存在する。このように用言 node 同士を表現した arc は用言-用言 arc である。この arc に負荷された重みは文法規則を表わしているといえる。

体言-用言 arc

動作様態を表わす用言 node には、それに対してある種の意味役割をはたす体言 node が存在する。例えば、give(与える)にはいつ、どこで、誰が、何をなどの 5W1H のような意味役割を持った物事が存在する。このような意味役割を表わすものが体言-用言 arc である。体言-用言 arc には以下のような意味役割が存在する。

- agent:動作をする主体(動作主)
例) "咲く" "花"
- object:動作をされる対象
例) "植える" "花"
- a-object:属性を持つ対象
例) "赤い" "花"
- time:動作が行なわれる時間
例) "咲く" "春"
- place:動作が行なわれる場所
例) "咲く" "花壇"

以上の三種類の arc を主語となる体言 node から迎えることで、文を生成する。

Text Generation with Weighted Semantic Networks

Yusuke SATO[†] and Satoshi MATSUDA

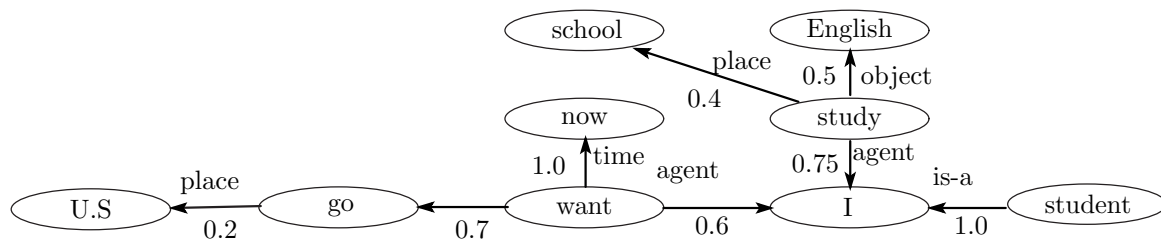


図 1: 重み付き意味ネットワーク

3 文章の生成

文章は、SVO、SVC などの基本 5 文型にそって生成される。そのため、重み付き意味ネットワークの中から S、V、O、C といった要素を抽出する。抽出された要素を、”3 単現の s” のような文法的な処理を施した上で、それを重み付き意味ネットワークを主語から辿るような順に出力することで文章を生成する。また、重み付き意味ネットワーク内を辿る場合、重みの大きい方を先に辿ることとする。

3.1 SVOC の抽出

主語 (S)

主語になりうる node は体言-用言 arc の *agentarc* の遷移先の node である。ここで、候補の node が複数ある場合は、候補 node の中で、連結される *agentarc* の本数の多い node 順に主語となる。連結される *agentarc* の本数が同じ場合は、それらに対する重みの平均の大きい方から順に主語となる。これは、主語が遷移する回数を減らし、知識の豊富な物から文章をつくることのできるためである。

動詞 (V)

動詞になりうる node は、先ほどの主語とは逆に *agentarc* の遷移元の node である。基本的には、ひとつの *agentarc* に対して S と V が一組存在するため、最低でも *agentarc* の数は文章が生成される。

目的語 (O)

目的語になりうる node は *objectarc* での遷移先 node となる。

補語 (C)

最後に C(補語) を決定する。補語になりうる node は *a-objectarc* の遷移先 node と、体言-体言 arc の *is-a* で接続された node である。*is-a* で結ばれた補語に対しては、主語との間に V がネットワーク上からは抽出することが出来ない。このため、*is-a* で結ばれた補語との V は be 動詞とする。

4 生成例考察

図 2 の重み付き意味ネットワークから、文章生成を行った例を以下に示す。

生成文

- I am student.
- I study English at school.
- I want to go the US.

生成された文章を表示する場合、文の順番も重要である。結果として得られた生成文の中で、1 文目の文章は、2 文目や 3 文目よりも後にくると不自然である。上記の結果は小規模の重み付き意味ネットワークからの生成例であった。しかし、本来の文章生成では、より大規模なネットワークからの文章生成をすることが必要になってくる。そのような大規模な重み付き意味ネットワークから node を確実に抽出できるか、またそのような大規模なネットワークをどのように作るのかということを考えていかなければならない。

5 おわりに

本論文では、重み付き意味ネットワークを用いた文章作成手法の枠組みを提案し、重み付き意味ネットワークによる文章作成が可能である事を示した。今後、重みの獲得、重みの更新等の知識をテキストより獲得する事により、本手法の検証を行なう。また、テキスト要約で抽出された文を一度分解し、本手法により文章生成を行うことにより、より理解しやすい要約文が作成できることが期待される。

参考文献

- [1] 長尾正太郎, 黒橋禎夫, 長尾真, 意味ネットワークからの文章生成, 情報処理学会, 自然言語処理研究会, 120-20, pp.133-140, 1997
- [2] 松田聖, 佐藤裕介. 重み付き意味ネットワークによる英文要約手法. 情報処理学会, テキストドキュメント研究会, 2004-DD-42 pp1-6, 2004