

依存ネットワークをもとにしたパラグラフ要約

山 脇 拓[†] 中 野 滋 徳[†]
足 立 顕^{††} 牧 野 武 則[†]

依存ネットワークは、語の間の主従関係に基づいた文の依存構造を表現する。名詞間の照応関係や接続詞による事象間関係によって、パラグラフ内の文は関係付けられ、1つあるいは複数の依存構造を形成する。要約文は、パラグラフの依存ネットワークで、事象ツリーのルートとなる述語から、そのネットワークをトラバースすることで形成されるとする。本論文では、ルートとなる述語の発見と、ネットワークをトラバース際の制約を制御することで、妥当な要約文が生成されるかを議論する。

Paragraph Summization based on Dependency Network

TAKU YAMAWAKI[†] SHIGENORI NAKANO[†] AKIRA ADACHI^{††}
and TAKENORI MAKINO[†]

This paper aims at proposing a summarization strategy. Dependency network represents a dependency structure with governor-dependent relationship. Sentences in a paragraph are related each other according to the anaphora between nouns and the event relationship between verbs by conjunctions. The paragraph is described by one dependency network or more. Summarized sentence is generated by traversing the dependency network from a rooted predicate of event tree in the network. We discuss how to find out the rooted predicate and to give constraints in traversing the network.

1. はじめに

要約に対する言語学的アプローチは、Schank¹⁾の概念依存モデルでその可能性が示されて以来、多くの研究がなされてきたが、その潜在的な能力にもかかわらず、技術的な困難さが指摘されている。その問題点は、表層の構文構造から文章全体の関係を把握する有効な方法が見出されていないこと、格構造などの意味表示からも有用な要約のための手がかりが見つからないことによる。要約の研究で、常に隘路となるのは、その結果の評価である。つまり、評価者によって要約の結果は依存するために、コンセンサスが得られる有効な評価法がない。このことは、「要約」の定義が

明確でないことによる。

まず、要約の定義を行う。与えられた制約のもとに抽出されるコンパクトな文章を要約文ということにする。すなわち、与える制約をどのように定義するかが、具体的な要約の定義である。

本論文では、要約の対象を、主題が比較的明確なパラグラフとする。パラグラフ内の文は何らかの関係を持ち、文脈を維持している。その関係を、照応関係、事象間関係によって与えて、パラグラフを1つあるいは複数の結合された依存ネットワークで表現する。次にパラグラフ依存ネットワークの事象間依存関係から、そのパラグラフにおけるルートとなる述語を発見し、その述語からパラグラフ依存ネットワークを、与えられた制約に基づきトラバースすることで、要約文を生成する。

本論文は依存ネットワークをもとにしたパラグ

[†] 東邦大学 大学院理学研究科 情報科学専攻
Department of Information Sciences Toho University
^{††} 富士通 情報メディアソリューション本部
Division of Media Solution, Fujitsu Co. Limited

ラフ要約を目的とする。依存ネットワークは、語と語の直接の関係を表層表現により表示し、ネットワークの構造を示したものであり、依存ネットワークを使った文生成²⁾、言い換えの研究³⁾⁴⁾がなされている。

本論文の構成は、まず、2節でパラグラフの構造化について述べ、3節で要約文生成方法について述べる。4節で要約文生成の例を、5節でパラグラフ要約の考察を述べ、最後に6節で本論文のまとめを述べる。

2. 依存ネットワークによるパラグラフの構造化

パラグラフはいくつかの文により構成されており、それぞれの文は表記上は独立しているが、何らかの関係で結ばれており、ひとつのパラグラフにまとめられている。これらの文の間関係を導入することで文章の構造化を行い、パラグラフ依存ネットワークで表現し、パラグラフにおけるルートとなる述語を発見する方法を提案する。

2.1 照応関係

● 代名詞・指示詞

代名詞には「彼」「彼女」「君」といったものが一般的で、指示詞には「こそあど」といった「これ」「それ」「あの」「どこ」といったものが代表的である。代名詞や指示詞といったものは、文と文をつなぐ機能の一つになる。これらは前述の文の名詞を指し示し、文と文の関係を与えている。

「太郎はサッカー選手だ。彼は世界的なスーパースターだ。」では、先行詞は「太郎」照応詞は「彼」となる。このパラグラフでは「彼」と「太郎」が照応関係にあることがわかる。

● 同一名詞

同一名詞が、パラグラフ内で複数出現するケースがある。それらの名詞は互いに同一と考えられ、文と文の関係を与える。

「地球温暖化をもたらす大気中の二酸化炭素を削減するのに大きな役割を果たすのが森林だ。樹木は二酸化炭素を吸収する。その森林はアマゾンで広大な地に広がっている。」この文では、「二酸化炭素」が1文と2文に、「森林」が1文と3文に現れている。これらはそれぞれ文と文の間につながりを示している。

● 名詞部分一致

同パラグラフ内に現われる同じ名詞を含む語句（名詞が後方部分一致している語句）の関係は非常に大きい。このように、名詞部分一致したものはネットワーク内で、文と文をつなぐ要素として扱う。

● その他の照応

その他の照応として、「同～」といった前文に現れる名詞を指し示すものも照応関係をとることができる。このとき、前文の名詞を指定するためには、意味を考慮する必要がある。

2.2 接続詞による事象間関係

接続詞は事柄と事柄の関係を明らかにするために、話し手の態度を述べる語詞であり、常にそれに先立つ叙述を必要として、これを受けているものである。

● 接続詞の機能

接続詞は文と文の論理関係を担う形式であるが、文や語句を一定の意味関係・論理関係のもとに結びつけて、文脈を展開していく上で、接続詞や接続助詞などが重要な役割になっているとされている。文の羅列でも意味が通じるが、その場合受け手に論理構成を委ねる形になり、受け手が正確に論理構成ができない場合は誤解が生じてしまう。接続詞や接続助詞を使用することで、受け手に誤解されずに論理構成を明確にすることができる。

接続詞の機能⁵⁾として、4つの分類に分け代表的な接続詞について検証した。これを、表1に示す。

文と文のつながりについては接続詞を挟む前後の文は、ほとんどの場合、後文に重みがある。しかし、「補足の接続詞」は後文が前文を説明しているため前文に重みがある。

● 動詞と動詞の関係

パラグラフにおける、文と文のつながりにおいて、動詞の関係が非常に重要である。

「EUは1993年1月に市場統合を実現させた。そして、今回の通貨統合で各国は自国通貨を放棄し、2002年には国籍は違っても約3億人が同じ通貨を使用する。」このパラグラフにおいて、動詞の関係を検証すると、「実現させた」というのは、「そして」という対等の接続詞にあたり、後文に重要度がある。したがっ

表1 接続詞の機能

接続詞の種類	接続詞の機能	例
対等の接続詞	様々な意味合いで語句を並べ立てる	そして・また・さらに・しかも・つまり・すなわち など
承前の接続詞	前文を前提ないし先攻条件として、後文に結びつける	すると・それから・ならば・でなければ・だから・しかし・でも・ところが など
転換の接続詞	それまでの文脈や話題を断ち切り、新しい話題を切り出す。	とにかく・ともあれ・さて・ところで・それで・では など
補足の接続	前文の内容を補足する内容を後文に述べる。	なぜなら・というのは・だって・なお・ちなみに・たとえば など

て「実現させた」という動詞は「使用する」という動詞に依存する。

2.3 構造化の例

「EUは1993年に市場統合させた。今回の通貨統合で各国は自国通貨を放棄し、国籍は違っても、2002年には約3億人が同じ通貨を使用する。ユーロランドを機に欧州人意識が強まるだろう。」を構造化した結果を図1に示す

3. 要約文生成方法

前章で構造化したパラグラフネットワークで、その構造を示唆する関係は事象間関係である。事象間関係には、因果関係、時間関係、逆説、話題変化などがあり、有向アークの向きが、パラグラフに置ける論理展開の向きを表している。ここで、事象間関係のルート（親を持たないノード）を、そのパラグラフにおける要約の述語とする。これを、パラグラフの主述語と呼ぶ。パラグラフは単一、あるいは複数の帰結ノードを持つ。一つの帰結ノードから、アークの向きが逆であるアークをたどって到達できる事象列が、そのパラグラフの要約事象列とする。要約事象列を、そのパラグラフの中心となる事象列とみなし、その事象列から要約文を生成する。パラグラフの文を、関係付けて形成される。

述語ノードは、関係する事態を表すために必ず必要となる格が定まっている。

格助詞は次の9つに分類される。

- 「ガ格」
動きや状態の主体を表す用法、状態の対象を表す用法。
- 「ヲ格」
動作や感情を向ける対象を表す用法、移動の

場所を表す用法、移動の起点を表す用法。

- 「ニ格」
人や物の存在場所、所有者、移送の着点、動作の相手・対象、状態の対象、原因、移動動作の目的、事態のときを表す用法
- 「カラ格」
移動の起点、受け取り相手の動作の相手、移動の起点としての動作の主体、時の起点、出来事の発端としての起点、判断の根拠
- 「デ格」
出来事・動作の場所、道具・手段、材料、原因、範囲、限度、基準、動作の主体を表す。
- 「ヘ格」
方向・目的地を表す用法
- 「マデ格」
移動の終わる場所、事態の終わる時を表す用法。
- 「ヨリ格」
比較の相手、時の起点を表す用法。

要約文生成において必須格の扱いに考慮する。

3.1 要約文生成の流れと制約

要約文生成についての流れと制約を示す。(1)～(4)は要約文生成の流れであり、(5)～(11)については要約文生成の制約とする。

- (1) まず、先頭の文の動詞を探す。その動詞の周辺に存在する名詞で、次の文と照応関係をとれる名詞を探す。あればその名詞にシフトし、シフトした名詞の周辺の動詞を探す。以後、シフトができなくなるまでそれを繰り返す。辿り着いた動詞をパラグラフの主述語とする。

(10) 提題に関する処理

提題化された名詞は、次の「は」が出てくるまで、パラグラフ中のすべての述語に関係を持つとする必要がある。よって、要約文を作成するにあたっては、必要に応じて主語を補う。

(11) 独立文に関する処理

パラグラフ内で独立している文は、要約に無関係な文と判断できるためネットワーク上から取り除く。

4. 実 験

ジャンルを問わず 50 のパラグラフについて要約文の生成を行った。その結果、41 のパラグラフについては要約文を生成することができた。要約文の生成が成功した例を以下に示す。

図 2 は「米大リーグマリナーズの佐々木投手 (35) は 20 日、都内のホテルで記者会見した。その席で、代理人を通じて同球団に退団の申し入れと日本球界復帰の希望を伝えたことを表明した。」をパラグラフ依存ネットワークで示した図である。このパラグラフ依存ネットワークをもとに要約文を生成する。パラグラフ依存ネットワークからの文生成は、文生成アルゴリズム²⁾を利用する

しかし、残りの 9 パラグラフにおいては要約文の生成を行うことができなかった。

図 2 では、提題の「は」が 1 回しか出てきていない。つまり、最初の文「米大リーグマリナーズの佐々木投手 (35) は」が全体に係っている状態である。制約より、パラグラフの主述語は「表明する」になり、このパラグラフの主述語を中心に、パラグラフ依存ネットワークをトラバースすると、「同球団に退団の申し入れと日本球界復帰の希望を伝えたことを表明した。」となるが、主語が抜けているので、この文に主語を補うと、「佐々木主浩投手 (35) が同球団に退団の申し入れと日本球界復帰の希望を伝えたことを表明した。」となる。ここで「米大リーグマリナーズの」と「同球団」は照応関係が与えられている。よって、

要約文：「佐々木主浩投手 (35) が米大リーグマ

リナーズに退団の申し入れと日本球界復帰の希望を伝えたことを表明した。」

が得られる。

また、本論文ではジャンルを問わずパラグラフ要約を行ったが、ジャンルごとに要約文生成の制約を決めることで、要約文がうまく取り出せるといえる。

5. 考 察

本実験において、要約文の生成が行えないものがあつた。その原因として

- 不明確な名詞を含むパラグラフの処理
- たどれるアークを複数持つパラグラフの処理
- パラグラフが補足文で構成されている場合の処理

次に制約に関して、次のことがわかつた。

- 副詞に関する処理
文中の時を表す副詞は省略ができると、本論文では述べたが、文章の構成によっては省略できない場合がある。例えば、
「7 日、8 日の 2 日間を使って PKO と生物資源の調査の規制についての国連会議が開かれ、生物資源探査の規制は 8 日に話し合われた。」というパラグラフでは、後半の文の「8 日」を省くと、前半の文の「7 日」、「8 日」のどちらを指しているかわからない。
- 提題に関する処理
提題化された名詞を含む文は、次の「は」が出てくるまでパラグラフ内のすべての述語に関係を持つ可能性があるため、生成するにあたっては最初に抜き出すことで、文生成がしやすくなることがわかつた。

6. ま と め

本論文では、従来の要約研究の語の頻度をもとにした統計的アプローチではなく、文の構造、依存文法の観点から要約文を生成する際に制約を設けることで、言語学的アプローチによる手法を用いて要約文生成を試みた。

