

物語の自動要約

The Automatic Summary of a Tale

村上 聡^{*1} 上之蘭 和宏^{*1} 榎津 秀次^{*2} 古宮 誠一^{*1}

Satoshi Murakami Kazuhiro Uenosono Hidezi Enokizu Seiichi Komiya

^{*1} 芝浦工業大学大学院

Graduate School of Engineering, Shibaura Institute of Technology

^{*2} 芝浦工業大学

Shibaura Institute of Technology

Abstract Text information cannot become huge and the information needed truly cannot be extracted exactly in recent years. Therefore, the text automatic summary which helps to grasp required information quickly is increasing importance increasingly as basic technology which supports man's intellectual activity. This paper proposes the automatic summary of a tale using the list of proposition and the propositional network. The technique consists of formation of a proposition list of a text, proposition network construction by the connection rule, and important proposition derivation by connection pattern matching.

1. はじめに

現在、インターネット等を通じて大量の情報が溢れており、大量の情報の中から必要な情報を正確に抽出することが必要である。したがって、必要な情報を迅速に把握することを助けるテキスト自動要約は、人間の知的活動を支援する基礎技術としてますます重要性を増している。

自動要約の目標は、情報のソースを受け取り、そこから内容を抽出し、最も重要な内容をユーザーに、簡約した形で、かつ、ユーザーやアプリケーションの要求に応じた形で提示するものである。日本でも、これまで多くの優れた自動要約の研究が行われてきた。その中で、現在最も自動要約で行われているものは重要文抽出である。

本研究では、拡張連結ルールによって命題ネットワークを構築し、物語テキストにおける重要文に含まれる命題がどのような結合パターンをもつかどうかを検討する。その結果から、物語の自動要約を試みる。

2. テキスト理解

テキスト理解は、テキストを心的表現に変換する一連の過程によって実現されると考えられている。Kintschら(1978)は、この過程を逐語的処理によるマイクロ構造、すなわち、テキストベースの構築、マイクロ構造の骨子を表すマクロ構造の構築、状況モデルの形成によって説明している。テキストベースは、文から引き出される意味の連結によって構築されるものであり、命題ネットワークとして表すことができる。本研究では、Kintschらのテキスト理解に基づいて研究を行った。

2.1 命題の連結

文から引き出される意味がどのように連結されているかについては、独特な視点からの主張がテキスト理解の研究においてなされてきた。Kintschは、共通する概念を持つ命題が連結されると主張している。また、van den Brock(1990)は、因果連鎖ネットワークによってテキストの心的表現を表し、因果関係による命題の連結を提案している。さらに、構文構造において同一の木に含まれる概念を便宜的に独立させ、詳細な記述ほど想起しやすいという知見を補償するように命題に連結することでテキストの心的表現を表している研究もある。そして、物語文法に見ら

れるようにジャンルによる文章構造を反映した命題連結の重要性を主張する研究も存在する。

2.2 拡張連結ルール

本研究では、命題ネットワークを構築するための連結ルールに共通指示結合、因果結合、構文結合、展開結合、そして、それらの連結ルールのどれも適用できないが隣接する文から引き出される命題を連結する NULL 結合を含めた。これらの連結ルールは、解説文のみに適用されるものであるが、会話文のみに適用する会話結合、会話相互結合を連結ルールに加えた。物語テキスト中の会話文は、ある話題に基づく登場人物の相互交換的な発話を反映していることが多い。したがって、連続する会話文の範囲を明らかにするとともに発話の相互差異を考慮した連結ルールを案出した。

3. 命題リスト

命題リストとは、形態素解析を行なった文章から Fillmore(1975)の格文法に基づいて深層格の付与を行なったものである。命題とは、概念を表すもので、言語の表層的な表現にとられない表現形式である。命題は、一つの述語(主に用言概念)と一つ以上の項(主に名詞概念)から構成される。項にはそれぞれ深層格が割り当てられる。

3.1 命題リスト作成

物語テキストを構成するそれぞれの文に対して、形態素解析、構文解析、そして、表層格による意味解析を実施した(もとの文の意味を変えずに解析可能となるように修正した。また、感嘆詞、擬音語などは除かれている。)。そして、意味解析の結果を Kintsch の表記法に基づき、連結ルールの適用に必要な情報を付加した命題リストへと変換した。

4. 命題ネットワークの構築

得られた命題リストより、命題ネットワークを以下の ~ までの規則に従いそれぞれ順番に命題同士を連結する内部処理を行う。

展開結合(P): 動作主格の移動が確認される、時間の移動がある場合、命題番号1とリンクさせる。

構文的結合(G): 命題リストにリンク先命題番号が記されている場合、それに基づいてリンクを張る。

連絡先: 村上聡, 芝浦工業大学大学院, m103201@sic.shibaura-it.ac.jp

因果結合(C): 1文中に動詞が複数出てくる場合は、支配されている動詞の命題番号とメインの動詞の命題番号をリンクさせる。

会話結合(K): 文中に会話がくる場合はその文の前文の先頭にある命題番号をリンクさせる。

会話相互結合(R): 会話文が続く場合は、前の会話と後に続く会話のそれぞれ最初の命題番号同士をリンクさせる。

共通指示結合(A): 1文中の会話の中に同じ引数がある場合は、その引数同士をリンクさせる。

Null 結合(N): 上記の ~ にあてはまらない場合は、その命題番号の一つ前の命題番号とリンクさせる。

上記のルールによって構築されたものが命題ネットワークになる。表 1 は、構築された命題ネットワークを表したものである。

表 1. 命題ネットワーク

	2	3	4	5	6	7	8
文番	1	1	1	2	2	2	2
1	G	G	G				
2							
3							
4			N				
5				C			
6					G	A	
7							
8							

5. 重要命題導出

連結パターン照合による重要命題導出を行うために、重要文に含まれる命題が命題ネットワーク上でどのように結合されているかを明らかにする必要がある。

5.1 実験(1)と結果

10人の参加者に5つの物語テキストを読ませ、重要文と判断した部分を抜き出させる調査を実施した(表2)。

人間が重要文と判断し、抜き出した部分が命題ネットワークではどのようなパターンであるかを検討した結果、重要文には構文的結合、因果結合、共通指示結合によって多くの他の命題と連結している命題が含まれることが発見された。

表 2. 実験結果(1)

	物語の文の数	重要文と判断した文(平均)	圧縮率
つるのおんがえし	50	14	28%
かさじぞう	58	17	29%
ももたろう	53	17	32%
おむすびころりん	60	18	30%
いっすんぼうし	46	14	30%

5.2 実験(2)と結果

実験(1)の結果から、多くの他の命題と構文的結合、因果結合、引数結合によって連結している命題を含む文を重要文とするシステムを構築した。システムが重要文として出力した結果を表3に示す。

5つの物語テキストにおいて人間が重要文として判断した文とシステムが重要文として出力した文を比較した。その結果、1)人間が重要文として判断した文は平均16文であったが、重要文として出力した文は平均14文である、2)人間が重要文として判断した文と重要文として出力した文の一致率は約80%である。

予備的な調査ということを考えれば、この結果は、さらに改善を要するものの、物語テキストを自動要約するための土台として拡張連結ルールにより形成された命題ネットワークは有効であることを示唆している。このことは、テキスト理解において構築されると考えられているテキストベース、すなわち、命題ネットワークの姿を暗示してもいる。

表 3. 出力結果

	物語の文の数	重要文と出力した文(平均)	圧縮率
つるのおんがえし	50	14	28%
かさじぞう	58	12	21%
ももたろう	53	15	28%
おむすびころりん	60	15	25%
いっすんぼうし	46	13	28%

6. まとめ

本研究は、拡張連結ルールによって構築される命題ネットワークに基づく、物語テキストの自動要約を提案した。

今後の課題は、物語文法をより反映した連結ルールを加えることで、一致率が向上すると思われるため、拡張連結ルールを改善する予定である。

参考文献

- [Kintsch 1978 年] Kintsch,W and vanDijk,T: Recalling and summarizing, Language,40,pp98-116,1978.
- [Kintsch 1973 年] Kintsch,W and vanDijk,T: Reading rate and retention as a function of the number of propositions in the base structure of sentences, Cognitive psychology,5pp257-274,1973.
- [上之蘭 2004 年] 上之蘭和宏: 遺伝的プログラミングによる係り受け関係からの格フレーム辞書自動生成システム, 第159回 自然言語処理研究会 2004.
- [長尾 1996 年] 長尾真: 自然言語処理, 岩波書店, 1996.
- [森 2001 年] 森敏昭: おもしろ言語のラボラトリー, 北大路書房, 2001.
- [Inderjeet 2003 年] Inderjeet Mani: 自動要約, 共立出版, 2003.