

# 反復語句・必須格・文間深層格を考慮した要約システム ABISYS

野口 貴<sup>†</sup> 韓 東力<sup>‡</sup> 原田 実<sup>‡</sup>

<sup>†</sup>青山学院大学大学院 理工学研究科 経営工学専攻

<sup>‡</sup>青山学院大学 理工学部 情報テクノロジー学科

## 1. はじめに

原田研究室では昨年度、入力文章から意味解析システム SAGE<sup>[1]</sup>によって意味解析した格フレーム群から要約文を生成する ABISYS<sup>[2]</sup>を開発した。従来の ABISYS では、反復語句を重要語句とし、必須格を補うことで要約を作成していた。本研究では、図 1 に示すように、重要語句の選定用に従来の反復得点に、文間の深層格情報、位置情報、意見語情報、主題・焦点情報など複数の視点からの得点を加えて、これらからマハラノビスの汎距離を用いて体言に対する重要語得点の総合化を行い、そこから正しい日本語として必要な語句を追加して要約文を生成する手法を提案する。

## 2. 重要語得点要素

本研究では、体言が文章の内容のポイント（要点）を示すものと考え、体言を重要語候補とする。重要度合いの得点は反復得点、文脈得点、位置得点、見解得点、主題・焦点得点の 5 種類とし、それらの得点を総合化することで体言に対する重要語得点とする。以下にそれぞれの得点の求め方を示す。

### 2.1. 反復得点

ここで言う反復語句とは、「文章中の異なる文に 2 度以上出現する同一語句・同意語句・同カテゴリ語句」のことである。同一語句とは見出しが一致する語句、同意語句とは見出しは異なるが概念 ID(語意)が等しい語句、同カテゴリ語句とは語と語の概念距離が 1 である語句を言う。

### 2.2. 文間深層格

文間深層格とは、隣接する 2 文間の意味的な関係を定義したもので 21 種類の格がある。この文間深層格と要約の関係性を調査したところ、話の転換を示す inter-conversion 等一定の格を持つ文が要約に残りやすく、詳細化を示す inter-detail 等の一定の格を持つ文が残りにくいことが分かった。この調査を基に文を構成する体言に対して加点・減点する。

## Summarizing System ABISYS based on repeated words, indispensable case and deep case between sentences.

Takashi Noguchi<sup>†</sup>, Dongli Han<sup>‡</sup> and Minoru Harada<sup>‡</sup>

<sup>†</sup>Graduate School of Industrial and Systems Engineering, Aoyama Gakuin University.

<sup>‡</sup>Department of Integrated Information Technology, Faculty of Science and Engineering, Aoyama Gakuin University.

## 2.3. 位置得点

一般に新聞記事では冒頭文が重要であるとされる。また、最終文にはテーマに関する今後の展開などが書かれる場合が多い。論説文においても冒頭文および最終文は筆者のまとめが書かれることが多いとされている。本手法では、第 1 文目と最終文を重要視し、それらの文を構成する体言に対して加点する。

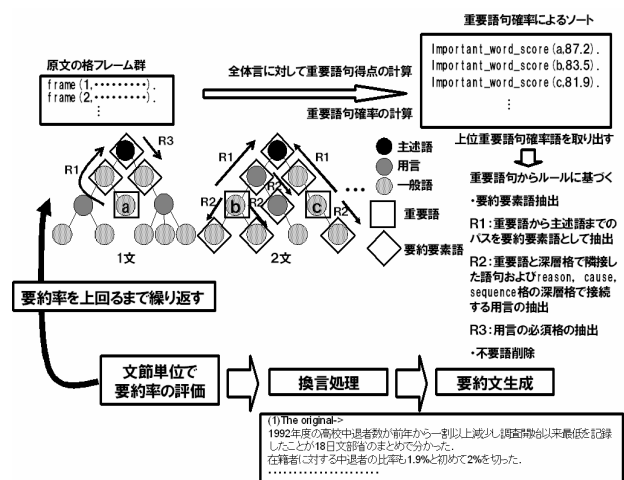


図 1 ABISYSによる要約手順

## 2.4. 意見語得点

論説文においては、筆者の主張、意見、希望がもっとも重視されるものである。そこで意見語を特定するための意見語句(「思う」など)を調査し、その語句と深層関係がある語、さらにその語とも深層関係がある語句に対して意見得点として加点する。これは意見語句から 2 つ以内の深層関係において筆者の意見が集中しているためである。

## 2.5. 主題・焦点得点

要約というものは、「原文の主題や結論に焦点を、自分の言葉でまとめること」であり、主題・焦点を特定できれば、原文の趣旨を読み取った要約を作成することが出来る。そこで本研究では、次のようなルールで簡易的に主題・焦点を決定し、それぞれの語に対して加点する。(1) 名詞 A+助詞「は」「も」「が」「を」「に」「で」の形を探す。(2) この名詞 A のうちで他の語句に対して「agent」「object」「a-object」「goal」である名詞 B を特定する。(3) この名詞 B を主題・焦点語とし加点する。

## 3. 重要語句得点の総合化

従来の研究ではそれぞれの得点に重みをかけ、それぞれの要素得点の和を総合得点とし、重要語句を抽出するもの

がほとんどであった。しかし、それらの方法はすべて経験的に求めたものであり、統計的には根拠が乏しい。そこで本研究では、マハラノビスの汎距離を用いて確率的な重要語句決定方法を採用した。まず、5変数で定義された空間、つまり「反復得点」、「文脈得点」、「位置得点」、「意見得点」、「主題・焦点得点」の5つの得点を持つ空間において、あらかじめ大量の重要語句候補を正解、不正解に人手で判別しておく。ここで言う正解とはその文章の要約において抽出されるべき語句ということである。その上で新たな重要語句候補を解析する際、それが正解、不正解どちらの群に近いかをマハラノビスの汎距離を用いて判別し、正解重要語句確率という形で算出する。要約率に応じて、その確率の高い順に重要語句とする。

#### 4. 正しい日本語としての出力

抽出された重要語句から正しい日本語として要約を出力するための方法を述べる。本研究では次に説明する2つの方法を用いて重要語句から正しい日本語としての要約を得る。

##### 4.1. 必須格の抽出

他研究では、主に表層情報を用いて、用言に対し「ガ格」、「ヲ格」なら必須格とするなどとしている。しかし、表層情報だけでは必須格の抽出は不十分である。また、必須格をそれぞれの述語に対して決定している研究もあるが、その研究においても文の主たる述語に対してのみ必須格を補っているため、1文中に複数の述語があるときには、主たる述語以外の述語に適切な語を補うことができない。そのため、本研究では、EDR コーパス中の用言の各深層格  $x$  の共起辞書内における出現割合を統計処理し、平均値  $m_x$  を求め、用言  $v$  の格  $x$  の出現割合が  $m_x$  以上であれば、 $x$  を  $v$  の必須格とする。これで表層情報では抽出できない語を抽出することが可能となり、統計処理を行うことで、どの深層格がどの用言の必須格になりやすいかを客観的に決定することができる。

##### 4.2. 重要語句から要約要素語の抽出

重要語句からの要約要素語の抽出は図2に示すように次の4ステップで行う。(1)重要語句から主述語までのパスを要約要素語として抽出する。(2)重要語句と深層格で隣接した語句と「reason」「cause」「sequence」格で隣接する用言を抽出する。(3)重要語句と深層格で下向きに隣接した語句を抽出する。(4)抽出した要約要素語中に、「こと」「の」「という」が含まれていた場合、これらの語は何か他の語を伴って意味が生ずると考えられるので、さらにその which 格の語を抽出する。

#### 5. 文を縮める技術

意味的に重複するものを削除することにより、要約文を短くすることができる。補足語修飾節の削除、引用動詞の削除の2つを採用した。

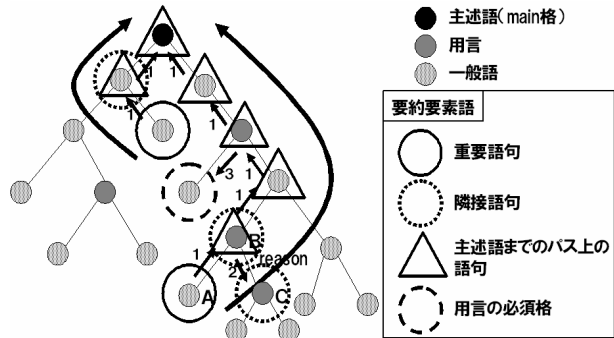


図2 要約要素語の抽出

#### 5.1. 補足語修飾節の削除

補足語修飾節の削除は次のステップで行う。(1)要約要素語から、被修飾名詞 A を探索する。(2)被修飾名詞 A が持つ補足語修飾節を表す which 格の先の語 B を探索する。(3)被修飾名詞 A が持つ修飾語を表す modifier 格の先の語 C を探索する。(4)語 B が被修飾名詞 A に対して、object 格、goal 格、place 格、agent 格を持ち、かつ、語 C が存在すれば、補足語修飾節を構成する語 (語 B を根に持つ部分木) をすべて削除する。

#### 5.2. 引用動詞の削除

引用動詞の削除は次の4ステップで行う。(1)上位概念が「考える」という概念を表す “30f878” や「思考する」という概念を表す “444dda” の語 A を探索する。(2)語 A が持つ、logical 格、timing 格、purpose 格の語 B を探索する。(3)語 B (語 B が構成語ならそれを含む複合語) の品詞が用言であり、かつ語 B の助詞が「と」「ように」「とか」ならば、語 B と深層格でつながる語 A を引用動詞と判断し、語 A を削除する。(4)語 A の引用節以外の深層格先の語を削除する。

#### 6. おわりに

本システムによる要約と、人手による要約を主観評価と文の修正の度合いに関して評価したところ、従来の手法に比べ、要約品質が向上した。よって本手法の有効性を示すことができたと言える。

#### 謝辞

本研究の一部は文部省科学研究費基盤研究 C 『日本語文章の常識を用いた意味理解・文脈理解システムの開発研究』の補助金を用いて行われました。

#### 参考文献

- [1] 原田実, 田淵和幸, 大野博之: “日本語意味解析システム SAGE の高速化・高精度化とコーパスによる精度評価”, 情報処理学会論文誌, Vol. 43, No. 9, pp. 2894-2902, (2002. 9).
- [2] 矢後友和, 原田実: “反復語句と深層格に基づく要約システム ABISYS の開発”, 情報処理学会自然言語処理研究会研報, 2003-NL-153, pp. 17-23 (2003. 1).