

要約ソフトの現状 と将来展望

石崎 俊・岡本 潤



▶ 1 はじめに

電子化された文書が世の中に溢れ、様々な情報をホームページ等から入手する機会が多くなった。大量の関連情報の中から必要な部分だけを入手するために、自動要約技術などによって読み手の文書の量を制御し、短い時間で適確に内容を把握する必要性が高くなってきている。市販されている翻訳ソフトの中には、翻訳機能の他に要約機能を装備するものが増えてきている。それらはホームページを見るときに、ブラウザ上の内容の重要部分を要約・翻訳する機能を有する。ただし要約だけを扱う市販ソフトはなく、このように翻訳ソフトやワープロソフトなどに、オプションとして組み込まれている場合がほとんどである。指定された文書を読み込み、文の数、もしくは全体の文字数に対するパーセンテージで要約を行なう。現在、要約した結果の生成・表示は文書中で重要な文、パラグラフなどの単位で抽出・表示するものが主である。複数の文を接続詞などでつなげてまったく新しい要約文書を生成するものは少ないが、論旨の流れを考慮し、必要に応じて文の重要部分や骨格の抽出し、重要文をさらに小さい単位で表示するソフトも出てきている。

▶ 2 要約の方法

要約を行なうには、どの箇所が重要なかを判断する必要がある。要約機能を有するソフトでは文書中の文を一つの単位とし、なんらかの情報をもとに各文の重要度を計算し、読み手が必要とする文書の量にあわせて、重要な文を提示していくという方法をとっている。以下に、文の重要度を計算するための主な方法についてまとめる。

(1) 単語の重要度

要約対象の文書中に現われる単語の頻度を計算し、その順に重要語を定義する方法。頻度が多い語は、その文章の主題に大きく関係している語であるとする仮定のもとに、主に名詞の頻度を計算し、その語が多く含まれる文を重要文として抽出する方法がとられている。

(2) テキスト中での位置

たとえば新聞の記事は「見出し」「小見出し」「本文」などのような、ある程度決まった構造をもっている。また、内容のジャンルによってなんらかの構造上の特徴がある場合がある。新聞を要約するときはタイトルや、記事の前のほうに出ている語の重要度を上げ、文の重要度を計算する。Web上の文書も新聞に似た構造が多く、要約法も近いと考えられる。

(3) 手がかり表現

主張、結論、評価などの特別な語を含む文を重要文とする。英語の場合は「must」「should」などの助動詞、日本語の場合は「欲しい」「望ましい」「要するに」「まとめると」などが考えられる。このような文書の重要な記述部分を示す語を含む文や、その文に含まれる単語の重要度を他の単語より上げる方法がある。

(4) その他の基礎的な方法

- a. 接続詞、照応関係などから文間・単語間のつながりと、その関係の解析を行なって要約する方法
- b. 文書を意味ネットワーク化して、その上でコネクショニスト・モデルを用いて、接点の活性値の収束値を重要度として計算する要約方法
- c. スクリプトなどのトップダウンな知識を用いて、重要概念を予め定義しておいて効率的に要約する方法

▶ 3 要約の表示方法

現在の自動要約システムでは、人間の要約能力には遠く及ばないため、要約結果の表示方法に工夫をしたものが多い。たとえば下記のようなものがある。

- a. 対象とする文書全体を表示し重要箇所にアンダーラインを引くか、フォントの色を変えて目立たせるなどの方法
- b. 重要な文と文書全体とをフレームを分離して上下または左右に並べ、重要文をクリックすると他方のフレームの文書全体で重要文を含む箇所を表示する方法
- c. 重要文のみを表示する方法

一般に単語や重要文だけを並べるのでは理解しにくい。たとえば、重要文に照応詞が含まれていた場合、それが何を指しているのかわからない場合がある。その時は照応詞が指している内容に置き換える作業が必要になってくる。また、文単位の要約ではなく、単語や句レベルで要約する場合は、文生成の技術に関するさまざまな課題が残っている。

▶ 4 要約・翻訳に関するアンケート調査

慶應義塾大学湘南藤沢キャンパスにおいて、平成12年度春学期に「自然言語論1」を履修している学生330名を対象に要約・翻訳などに関するアンケート調査を行なった。このアンケートはパソコンを日常的に使用し、Webの検索やメール交換、レポート作成にパソコンとネットワークを駆使している学生を対象としている。

コンピュータによる要約・翻訳の使用と使用ソフト、結果に関する学生の評価をまとめると表1から表4のようになる。要約ソフトの経験者は約1割、翻訳ソフトの場合は約3割と差があるが、満足度は前者が約15%、後者が約5%と逆になっている。しかし、

表1 要約システム使用経験の有無

コンピュータによる要約システムを使用した経験がある		
あ る	27人	10.1%
な い	239人	89.9%

表2 要約機能に対する満足度（複数回答あり）

評 価	満 足	普 通	不満足
合 計	4人	9人	16人
比 率	13.8%	31.0%	55.2%

表3 翻訳システム使用経験の有無

コンピュータによる翻訳システムを使用した経験がある		
あ る	75人	28.3%
な い	190人	71.7%

表4 翻訳結果に対する満足度（複数回答あり）

評 価	満 足	普 通	不満足
合 計	4人	31人	47人
比 率	4.9%	37.8%	57.3%



いずれにしても不満足が半数以上で評価は厳しいものになっている。

なお、要約の使用システムは、MicrosoftのWord、ジャスシステムの一太郎、AppleのSherlockなどである。一方、翻訳の方の使用システムはカテナのこりゃ英和！、Logo Vista EtoJ、ソニーの瞬間翻訳！ロボワード、日本IBMの翻訳の王様、富士通の翻訳サーフィンVer.3.0、L&HジャパンのE-JBANK、メディアヴィジョンのDr. Surf、フリーソフトの瞬訳名人パピロンなどである。

▶ 5 要約システムの紹介

市販されているワープロソフト、翻訳ソフトの中には要約機能がついているソフトも多い。なかでもワープロソフトのほとんどは要約機能を有している。ここでは今回調査した翻訳ソフトの特徴について考察を行なう。各ソフトでの要約の処理法自体は外部からは見にくいので、例文の要約結果に沿っていくつかの手法の評価を行なう。

- (1) 要約ソフト1は日本語、英語の文書ともにメニューバーの「ツール」から「要約の作成」を選択することにより要約文を生成する。要約の長さは文の数、語の数、または文書全体に対するパーセンテージで指定する。表示方法は「重要文のバックグラウンドの色を強調する」、「重要文だけ抽出する」、「文書の先頭に重要文を挿入」などの方法がある。単語の出現頻出をもとに重要度を計算し、比較的本文の初めの方にあっ

て重要語を含む文を重要文とする傾向が見られる。

- (2) 要約ソフト 2 は日本語、英語の文書ともにメニューバーの「ツール」から「文書要約」を選択することによって要約文を生成する。要約の長さは文書全体に対するパーセンテージで指定する。表示方法は「重要文のバックグラウンドの色を強調する」、「重要文だけを抽出する」という方法がある。また、オプションで「文字飾り」、「章、節、項や目次など」、「固有名詞」を重要とみなすかどうかをユーザーが決定できる。単語の出現頻出をもとに重要度を計算し、比較的長い文章を重要文とする傾向が見られる。また「文字飾り」、「章、節、項や目次など」、「固有名詞」で選んだ内容に著しく左右される。
- (3) 要約ソフト 3 は日本語、英語の文書ともに要約が可能であるが、インターネットにつなぎホームページを見ているときだけ使用できる。重要文は「Abstract」として文書の先頭に挿入する。また、ポップアップメニューで「要約」を選択すると別ウインドウに要約結果を表示することもできる。単語の出現頻出をもとに重要度を計算しており、接続関係についても考慮した要約を行なっている。
- (4) 要約ソフト 4 は、英語、日本語の文書とも要約可能である。要約を生成する際に文書中での位置や陳述部分の重視、頻出単語の重視の切り替えを行なうパラメータがある。談話の結束性を破らないように制約を定め、またフレーズ間の依存関係を判定するルールを設けている。要約結果は、目次、要約部分から原文の該当部分にハイパーリンクを張り表示することで、興味を引いた要約部分の詳細な内容を原文から探し出すことができる。
- (5) 要約ソフト 5 は、英語の文書のみで要約が可能である。「重要文抽出」で新しいブラウザを開き、重要度の高い文を示す。重要文の量は5%刻みにユーザーが選択することが出来る。重要文と原文は同時に提示され、重要文が原文のどこにあたるか一目で見わかるようになっている。(ここでは実験例を表示していない)

▶ 6 要約システムの実行例

例文として文部省のホームページにある「文部省のあらまし」の中の「国際交流・協力」の文書 (<http://www.monbu.go.jp/aramashi/2000jpn/09/09.htm>) を使用した(図1)。ここでは文書内に用いられている画像とキャプション、文書の末尾にあるリンク先への項目は省略した。

この例文を上記要約ソフト 1 から 4 で要約した結果を図2 から図5 に示す。

図1 要約比較に用いた原文

国際交流・協力	
<p>文部省では、学生、教員、研究者、青少年、芸術家、スポーツ競技者等の交流の推進や国際会議の開催等を通じて、国際交流・国際協力を図っています。</p>	
<p>教育関係</p>	
<p>留学生交流を推進するため、外国人留学生の受入れ体制の整備を積極的に進めています。</p> <p>また、国際理解を増進するため、初等中等教育教員の海外派遣や社会教育指導者の海外派遣などを行うほか、外国語教育の改善に資するため、外国人教師等の受入れや外国語教員の海外派遣・研修事業も行っています。</p> <p>ユネスコを通じた国際協力として、アジア太平洋地域諸国において、識字教育普及のための事業や教育革新促進のための事業に協力しているほか、平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。</p> <p>さらにOECD及びAPECの教育事業に対して協力を行っており、加盟各国間の教育政策研究、情報交換を行うとともに、OECD/JAPANセミナー等の国際会議を開催しています。</p> <p>このほか、国立大学等との連携をとりながら、開発途上国における人づくりへの支援を行っています。</p> <p>また、平成12年4月には初の試みとして我が国でG8教育大臣会合を開催するなど、国際的な意見交換の場を積極的に設けるよう努めています。</p>	
<p>学術研究</p>	
<p>在外研究員、国際研究集会等への派遣、外国人研究員等の招致、日本学術振興会を通じての研究者交流事業などにより、研究者の交流を進めています。</p> <p>二国間あるいは多国間の国際共同研究の推進、南極地域観測事業の推進、ユネスコ事業及びOECDへの協力などを通じて学術交流を進めるとともに、国際社会への貢献に努めています。</p> <p>また、国連大学については、我が国の学界との連携・協力により研究・研修等の事業活動を推進するため、支援・協力を行っています。</p> <p>このほか、国立大学等と連携をとりながら、開発途上国に対する人づくり支援を行っています。</p>	
<p>文化・スポーツ関係</p>	
<p>優れた舞台芸術公演の海外フェスティバル等への参加・我が国で開催される国際フェスティバルへの支援や、若手芸術家の在外研修、国民文化祭等を通じた文化団体の招へい・派遣、国宝・重要文化財等の古美術品の海外展の開催、文化財の保存修復に関する国際共同研究等の実施及びユネスコによる世界遺産保護への協力などを推進し、また、スポーツ国際交流も積極的に進めています。</p>	

図2 要約ソフト1

10%	<p style="text-align: center;">学術研究</p> <p>在外研究員、国際研究集会等への派遣、外国人研究員等の招致、日本学術振興会を通じての研究者交流事業などにより、研究者の交流を進めています。</p>
15%	<p>文部省では、学生、教員、研究者、青少年、芸術家、スポーツ競技者等の交流の推進や国際会議の開催等を通じて、国際交流・国際協力を図っています。</p> <p style="text-align: center;">学術研究</p> <p>在外研究員、国際研究集会等への派遣、外国人研究員等の招致、日本学術振興会を通じての研究者交流事業などにより、研究者の交流を進めています。</p>
25%	<p>文部省では、学生、教員、研究者、青少年、芸術家、スポーツ競技者等の交流の推進や国際会議の開催等を通じて、国際交流・国際協力を図っています。</p> <p style="text-align: center;">学術研究</p> <p>在外研究員、国際研究集会等への派遣、外国人研究員等の招致、日本学術振興会を通じての研究者交流事業などにより、研究者の交流を進めています。</p> <p>二国間あるいは多国間の国際共同研究の推進、南極地域観測事業の推進、ユネスコ事業及びOECDへの協力などを通じて学術交流を進めるとともに、国際社会への貢献に努めています。</p>

図3 要約ソフト2

10%	ユネスコを通じた国際協力として、アジア太平洋地域諸国において、識字教育普及のための事業や教育革新促進のための事業に協力しているほか、平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。
15%	文部省では、学生、教員、研究者、青少年、芸術家、スポーツ競技者等の交流の推進や国際会議の開催等を通じて、国際交流・国際協力を図っています。 ユネスコを通じた国際協力として、アジア太平洋地域諸国において、識字教育普及のための事業や教育革新促進のための事業に協力しているほか、平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。
25%	文部省では、学生、教員、研究者、青少年、芸術家、スポーツ競技者等の交流の推進や国際会議の開催等を通じて、国際交流・国際協力を図っています。 ユネスコを通じた国際協力として、アジア太平洋地域諸国において、識字教育普及のための事業や教育革新促進のための事業に協力しているほか、平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。 優れた舞台芸術公演の海外フェスティバル等への参加・我が国で開催される国際フェスティバルへの支援や、若手芸術家の在外研修、国民文化祭等を通じた文化団体の招へい・派遣、国宝・重要文化財等の古美術品の海外展の開催、文化財の保存修復に関する国際共同研究等の実施及びユネスコによる世界遺産保護への協力などを推進し、また、スポーツ国際交流も積極的に進めています。

図4 要約ソフト3

要約率固定	教育関係 留学生交流を推進するため、外国人留学生の受入れ体制の整備を積極的に進めています。また、国際理解を増進するため、初等中等教育教員の海外派遣や社会教育指導者の海外派遣などを行うほか、外国語教育の改善に資するため、外国人教師等の受入れや外国語教員の海外派遣・研修事業も行っています。ユネスコを通じた国際協力として、アジア太平洋地域諸国において、識字教育普及のための事業や教育革新促進のための事業に協力しているほか、平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。さらにまた、平成12年4月には初の試みとして我が国でG8教育大臣会合を開催するなど、国際的な意見交換の場を積極的に設けるよう努めています。
-------	--

図5 要約ソフト4

要約率 高	文部省では、国際交流・国際協力を図っています。 教育関係 国際理解を増進するため、初等中等教育教員の海外派遣や社会教育指導者の海外派遣などを行うほか、外国語教育の改善に資するため、外国人教師等の受入れや外国語教員の海外派遣・研修事業も行っています。平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。 学術研究 文化・スポーツ関係
要約率 中	文部省では、国際交流・国際協力を図っています。 教育関係 外国人留学生の受入れ体制の整備を積極的に進めています。また、国際理解を増進するため、初等中等教育教員の海外派遣や社会教育指導者の海外派遣などを行うほか、外国語教育の改善に資するため、外国人教師等の受入れや外国語教員の海外派遣・研修事業も行っています。平成11年度からは、世界の高校生、大学生、教員等を対象とする交流事業を開始しました。 学術研究 国連大学については、支援・協力を行っています。 文化・スポーツ関係 優れた舞台芸術公演の海外フェスティバル等への参加・我が国で開催される国際フェスティバルへの支援や、若手芸術家の在外研修、国民文化祭等を通じた文化団体の招へい・派遣、国宝・重要文化財等の古美術品の海外展の開催、文化財の保存修復に関する国際共同研究等の実施及びユネスコによる世界遺産保護への協力などを推進し、また、スポーツ国際交流も積極的に進めています。

▶ 7 要約システムの将来展望

図2から図4を見ると、同一の入力文に対して異なった要約結果を示している。これは、要約の方針や観点を変えたというよりも、システムの誤差というべきものであろう。これらのシステムは内容を理解した上で、重要な部分を要約しているわけではない。内容の理解そのものが、現在のコンピュータで難しいことが原因である。構文解析や意味解析の手法が不足しているだけでなく、人間の持っている常識や当該分野の知識などが不足しているのである。

このような知識の不足に対して、オントロジーの研究が進められている。これは概念辞書とも言われるもので、人間の持っている知識を構造化してコンピュータで扱えるように電子化したデータベースである。また、大規模な連想実験によってこのような人間の知識を抽出し電子化する試みもある。このような知識を用いた要約システムの出現が期待されている。

もう一つの現実的な手法として、最近では文書にタグをつけて、意味を構造化して表現する手法が進歩してきている。たとえば、GDA(Global Document Annotation)では、多言語間に共通の統語・意味等に関するXMLタグを文書につけている。タグとしては以下のようなものが考えられている。

1. 意味タグ：語(句)の意味
2. 統語タグ：係り受け・構成素構造、主辞の情報など
3. 指示タグ：代用表現の指示物、時間、場所など
4. 語用論タグ：文章の種類(広告、論説、手紙...)、談話参加者の社会的関係など

一般に要約には深い意味処理と多くの背景知識が必要である。しかし、これまでの研究の多くは表層的な手がかりや、文書の持つ何らかのスタイルに関するヒューリスティックを用いるものであった。GDAに基づく要約では、GDAタグを文書に付与することによって得られる、文の構成要素の重要度(活性値)を計算し要約を行なう。文書内のパラグラフ、文、句、語などの構成要素をノードとして、ネットワーク構造で表現する。また文書内の語、文間の意味的なつながり、照応関係は、対応する構成要素をリンクで結ぶことで表現する。文の構成要素(ノード)の活性値を活性拡散を用いて計算することで、活性値の高い要素を含む文を重要文とする。文より細かい単位で活性値を計算しているため、必須要素・照応を考慮しながら活性値の小さな部分を切り落とし、一文をさらに短くすることもできる。この研究はまだ進行中であり、実用システムにはなっていないが、近い将来に使用可能な性能の良いシステムとして実現が期待されている。

参考文献

- [1] 石崎俊, 文書要約からみた談話理解モデル, 「談話理解モデルとその応用」シンポジウム, 1988.
- [2] Koiti Hashida, Shun Ishizaki, Hitosi Isahara, A CONNECTIONIST APPROACH TO THE GENERATION OF ABSTRACTS.
- [3] K.Nagao, K.Hashida, Automatic Text Summarization Based on the Global Document Annotation, B-3215, COLING'98, 1998
- [4] Global Document Annotation, <http://www.etl.go.jp/etl/nl/gda/>
- [5] 中野潔, 「要約ソフトが突きつける 読むの意味」, 月刊言語, pp54-59, Vol.27, No.2, 1998.

- [6] 奥村 学, 難波 英嗣, “テキスト自動要約に関する研究動向”, 自然言語処理, 「テキスト要約のための言語処理」特集号, Vol. 6, No. 6, 1999.
- [7] 望月源, 要約機能付き市販ソフト, <http://galaga.jaist.ac.jp:8000/pub/research/summarization/software.html>
- [8] 野村直之, ConceptBaseの言語処理と新しいソリューション, 情報処理学会, NL-129-1, pp1-8, 1999.

(石崎 俊 慶應義塾大学政策メディア研究科教授)

(岡本 潤 慶應義塾大学政策メディア研究科博士後期課程)