

複数の情報における要約技術を活用した知的システムに関する考察

馬 文鵬[†]皆月 昭則[‡]林 秀彦^{††}鳴門教育大学[†]釧路公立大学情報センター[‡]鳴門教育大学^{††}

1. はじめに

人間は情報を要約する場合、複数の情報を得てから、頭の中でキーワードや重要な部分を抽出し、一つの要約を自動的に生成できる。そのプロセスは、脳において不必要な情報や冗長な情報を除去し、重要な部分を抽出していると考えられる。これにより、人間は情報を比較・整理・合併・濃縮等のプロセスを通して要約を実現するモデルを構築できる(図1)。また、複数の情報に対して、その中の関連性を探し出すことは人間の思考の特徴の一つである。このような人間の思考プロセスに基づき、複数の種類の情報を要約することができれば、様々な分野への貢献が期待できる。本研究では、人間の思考プロセスをモデル化し、複数の情報における要約技術を活用した知的システムを提案する。

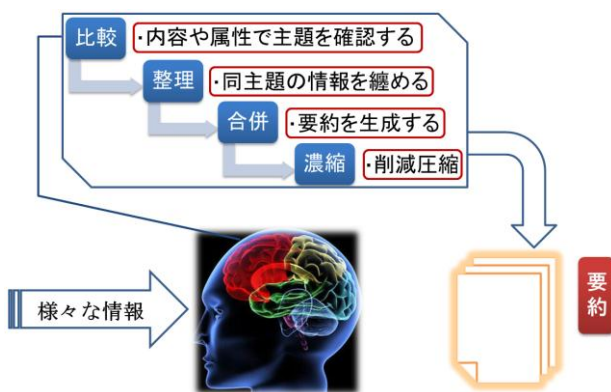


図1 人間の情報処理モデル

2. 要約技術の先行研究

要約技術には様々な手法があるので先行研究を概観する。テキスト自動要約手法としては、伝統的に、テキスト中の重要な文を抜き出す重要分抽出法が用いられてきた。研究が活発化するとともに研究の方向性も多様化してきた。以下のようないくつかの要約手法に関する研究が行われている。奥村はテキストの自動要約手法を以下の3つの

Research of Intelligent System on Multiple Information using Summarization Technology

[†]Bunhou MA・Naruto University of Education.

[‡]Akinori MINADUKI・Center for ITS, Kushiro Public University.

^{††}Hidehiko HAYASHI・Naruto University of Education.

観点からまとめている[1]。

- (1) 文から文中の重要箇所を抽出することによる要約手法。
- (2) 単一テキストではなく、複数テキストを対象にした要約手法。
- (3) ユーザに特化した要約を動的に作成する要約手法。

このようなテキストの自動要約手法に加えて、ネットで欲しい商品の評判を一言で知ることができるような「クチコミ要約」手法[2]、長時間の録画映像を要約して、数分で確認出来る「監視カメラ映像要約」手法[3]など、研究の方向性が多様化していることがわかる。しかしながら、人間と同じように思考し、複数の情報を要約できる技術の発展はこれからである。

3. 複数情報の要約

本論文では、複数情報の要約について述べる。複数情報の要約は、関連する多くの情報を比較・整理し・合併・濃縮を通して、一つの要約を実現するものである。なお、これらの比較、整理、合併、濃縮の4つの情報処理プロセスは、以後の説明のため便宜上設定しているが、必ずしも明確に分離して処理されるものではないと仮定している。以下に詳細を述べる。

情報の比較・整理：莫大な情報であっても、小分けに情報を分類できる。一つ一つの小さな情報から、各キーワードを通じて、分類された情報の内容を分析し、似ている情報の関連性を確立する。表1に基づいて「関連性」について説明する。表1に示す「分類」、「段落」、「キーワード」の各項目は以下のように説明できる。

表1 「分類」、「段落」、「キーワード」の関係

分類	段落	キーワード
分類1	段落1	山田、34歳、社員
	段落2	眼鏡掛け、ひげ
	段落3	洋服、ネクタイ、時計
関連性	山田の外見	

「キーワード」とは、テキスト情報であれば、表1に示すように、ある対象を簡潔に表す主題を表現する単語である。映像情報であれば、キーワードは単語に限るものでは

なく、例えば、ある映像シーンの主題を表すフレームを指す。

「段落」とは、テキスト情報であれば、文書の段落を指している。映像情報であれば、ある映像の一画面に区切られたシーンを指す。例えば防犯用監視カメラに記録された映像で説明すると、一日に記録された映像を時間単位で分割し、その中に含まれる人物を抽出したものを段落にすることができる。他にも、キーワードに相当する情報のまとめ方によって、段落になり得る部分は様々である。そのため、どのように段落に分けるかは課題の一つである。

「分類」とは、複数の段落がある場合、その中で関連している段落を集めて、まとめたものである。例えば、テキスト情報では、各段落を集めて、そのなかの主題の関連性に基づき、まとめたものを指している。映像情報においても、テキスト情報の段落と同様に、主題の関連性を用いて、まとめたものを「分類」とすることができる。ここで、関連性とは、人間の思考プロセスにおける特徴的な機能を指している。

情報の合併・濃縮：

ある対象には、あらゆる観点があり得る。例えば、テキスト情報、映像情報ともに、ある観点によって比較・整理された情報は、もちろん、共通する受取り方もあるが、別の観点をもつ受け手によって、受取り方（印象・感想）が異なることもある。そのため、ある対象には複数の属性があり、その中に属性間の関連性を見出すことも可能である。従って、属性（キーワード）を基にして単一要約（要素）を多量に生成することができる。これらの単一要約の集合を要約集合と呼び、その中の関連性がある情報をまとめ、同じ内容を含む部分あるいは冗長な情報を除去し、全面的な要約を生成する。この要約に対して、削減と圧縮を行い、主題から外れる情報や冗長な情報を除去する。さらに、不要な部分（動詞連体修飾節）を除去し、文字数を減らすことで、内容を圧縮する（図2）。

このように、図2に示す複数の情報における要約手法では、人間と同じように情報の比較・整理、冗長情報の合併・濃縮を行い、一つの最終要約を生成している。

4. 人間の思考プロセスに基づく知的システム

本論文では、複数情報の要約技術によって、人間の思考プロセスと同じように、複数の情報から関連性を把握し、一つの要約を生成する知的システムを構想している。そのシステムは、図2のように比較・整理、冗長情報の合併・濃縮のプロセスを通して実現されるシステムである。例えば、ネットで商品を買うとき、ユーザは好きな色、規格、

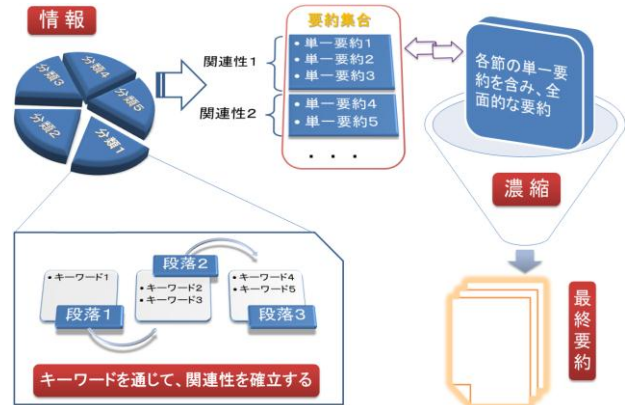


図2 複数情報の要約システムの概念図

ブランド、使用後の評価、コストパフォーマンスなどに関心があるとする。このとき提案する知的システムは、大量の情報を分類・比較し、合併・濃縮のプロセスを経て、ユーザが知りたい情報の要約を生成する。これらの要約された情報を参考にすることで、ユーザは複数の情報に対する可読性を高め、大幅な時間の削減が可能となる。

5. まとめ

本研究では、これまでの要約技術を概観し、従来のテキスト情報の要約のみならず、クチコミ、映像などの要約技術の多様な方向性を示した。そして、人間と同じように思考し、複数の情報を要約できる技術を実現する知的システムを提案した。提案システムは、人間の思考プロセスに基づくように概念設計した。これを応用して、複数の情報に対する可読性を高め、大幅な時間の削減が可能となることを示唆した。今後、より人間の一般的な思考プロセスを詳細に分析し、脳と同じように複数の情報の関連性を探す方法を組み込む知的システムを開発してゆく予定である。

参考文献

- [1]奥村学, テキスト自動要約, 情報処理, Vol.45 No.6 pp.574-579(2004)
- [2]goo ラボ, <http://youyaku.labs.goo.ne.jp/>
- [3]セコム IS 研究所, <http://www.secom.co.jp/isl/research/ps/report02/>
- [4]酒井浩之, 増山繁, ユーザの要約要求を反映するためにユーザとのインタラクションを導入した複数文書要約システム, 知能と情報 (日本知能情報ファジィ学会誌), Vol.18, No.2, pp.265-279(2006)
- [5]池田諭史, 牧野恵, 山本和英, 濃縮還元型文要約モデルの検討(談話・要約), 情報処理学会 研究報告 2006-NL-174(13), pp.71-76(2006)