

対話型文書情報の自動構造化手法の提案

東京大学 学生会員 岸尾 雄太
東京大学 正会員 堀田 昌英

1. 本研究の背景・目的

組織における意思決定の歴史を記した議事録を解析し視覚化することは、決定の背後にある膨大な組織の知識をもとに現在直面している課題を解決するための支援に繋がる。しかしながら旧来、議事録を解析する作業は人手により行われており、膨大な作業労力と時間を要するものであった。

本研究では、この作業を自動化することにより、対象の議事録を即座に、また簡便に構造化する手法の提案を試みる。この手法では、分析対象の議事録の単語の重要度に着目することにより、議事録を自動的にツリー状に構造化する。

2節で議事録の視覚化の手法としてツリー構造化の利点について述べ、3節では構造化の手法について述べる。4節で解析結果と考察を述べる。

2. 議事録のツリー構造化

通常議事録は議事中の発言者名とその発言が列挙されているだけのものであり、その議論において何が主な論点となったのか、またどれだけの深さで議論が展開されたのか、といったことは議事録を長時間に亘って丁寧に読まなければ分かりえない。

堀田・神野(2001)の開発したCRANES等の研究はこのような性質を持った議事録をツリー構造に変換することにより、話題の転換点や議論の深さ等を効果的に視覚化することに着目した意思決定支援システムである。これにより、議事録からユーザーが必要な情報をより短時間で直感的に得ることが可能となった。

しかしながら、議事録をツリー構造化する作業は膨大な時間と手間を要する。数百回・数千回に亘る議事録のデータを必要とする際に、人手により議事録をツリー化する作業を行うことは現実的には難しい。

そこで本研究は、過去の議事録の有効活用を目的として、議事録を計算機により自動的にツリー構造化することを試みた。

3. 提案する議事録の自動構造化手法

はじめに議事録中に表記されている発言者名を検出し、その前後を発言の区切りとする。このときに発言者名等の発言内容以外のものを指し示すノイズを削除しておく。

次に解析対象の議事録の文章を形態素解析機「MeCab」を用いて形態素解析し、出現した全名詞を抽出する。ただし、「とき」「もの」などのように文章の意味を測るのに不必要な単語は除外しておく。

各発言の前後において、これにより得られた単語配列と、この単語群が単語の頻度に従ってランダムに生じたとする単語配列との尤度を求める。この値が小さいほど単語がランダムに生じたとは考えにくく、その前後では意味の繋がりが強いと考えられる。この値が極大となる位置を意味の結束性が極小になる位置、すなわち話題の変化位置であるとみなす。この話題の変化位置の検出手法は中川(未発表論文)の手法を参考にした。

これにより得られたいくつかの話題のかたまり(以後セグメントと呼ぶ)間の話題の類似度を計算することにより、セグメント間の関係性を分析し、セグメントをツリー構造に整理する。話題の類似度の計算には、

キーワード 政策論議, ツリー構造, 自動構造化, 自然言語処理

連絡先 〒113-8656 東京都文京区本郷7-3-1 大学院工学系研究科社会基盤学専攻 TEL03-5841-6143

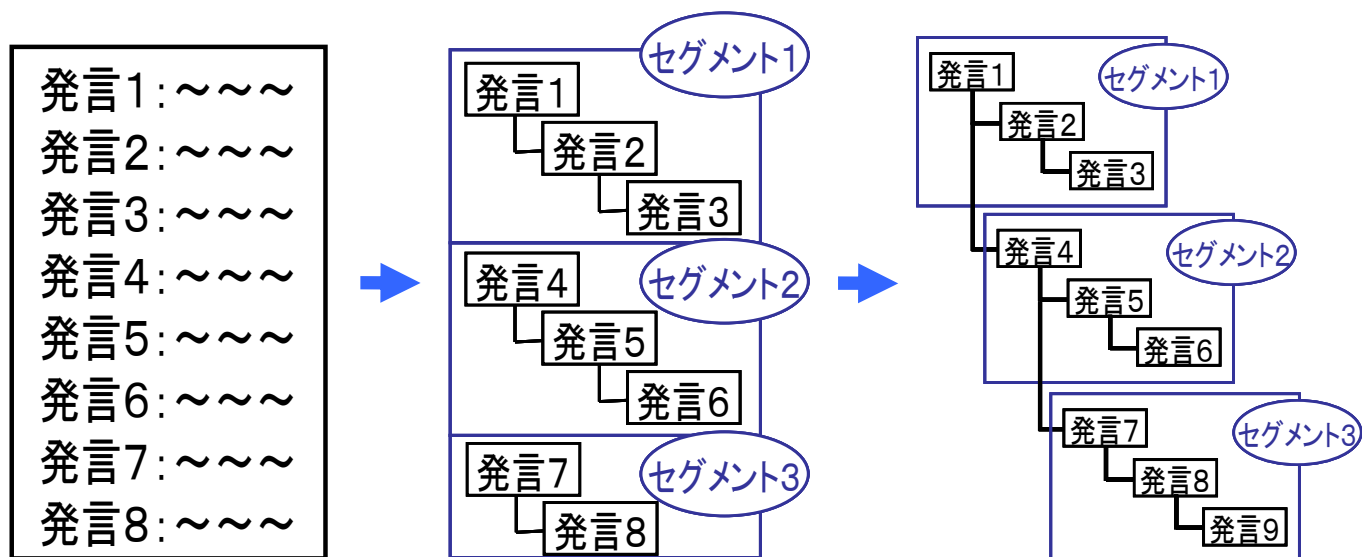


図 1-A 元の議事録

図 1-B セグメント分けされた議事録

図 1-C 議事録のツリー構造

tf-idf という単語のその文章における重要度を示す指標を単語ベクトルに重みとして付与し、ベクトル間の余弦をとることで計算を行った。

4. 結果・考察

解析には JICA 環境社会配慮ガイドライン改定委員会（第二回～第四回）の議事録を用いた。解析によって得られたツリーと、同じ議事録から人の手によって解釈的に作ったツリーの、各発言の親となる発言が一致しているかどうかを指標とし、それぞれの議事録について、一致した発言の数と全体の発言数の割合を求めたところ、第二回委員会の議事録では 68%，第二回委員会の議事録では 67%，第二回委員会の議事録では 71% という結果が得られた。このことから、精度に改善の余地があるとはいえ、実用に耐えうる精度であると考えられる。本研究手法，意思決定支援システム等への応用が期待される。

更に精度を上げるために今後の課題としては、本実験では実装されなかったが、シソーラス（類語辞典）の導入やコーパス（文章を解析するために用いた日本語における単語の使用頻度を分析するための文章群）の改善などが考えられる。

参考文献

- [1] 堀田昌英・榎戸輝揚・岩崎伸卓「多元的議論構造の可視化手法：社会技術としての政策論議支援」（社会技術研究論文集, 2003.10）
- [2] 堀田昌英・神野由紀, 2001, 「参画型パブリック・マネジメントのための情報基盤システム CRANES の開発」（土木学会論文集 2001, VI-52, 109-120）
- [3] 松村真宏・加藤優・大澤幸生・石塚満「議論構造の可視化による論点の発見と理解」（日本ファジィ学会誌, 2003）
- [4] 中川善典「単語配列の不自然さに注目したテキストセグメンテーション」（未発表論文）