

話題遷移図の可視化と話題遷移に応じた関連情報提示による 議事録閲覧支援

白松 俊 駒谷 和範 尾形 哲也 高橋 徹 奥乃 博

京都大学 大学院情報学研究科

{siramatu,komatani,ogata,tall,okuno}@kuis.kyoto-u.ac.jp

1 はじめに

背景と問題意識 PI (Public Involvement; 行政への住民参画) は、住民公聴会等の公的討議を通じて住民の意見を集約し、行政の意思決定に反映させるプロセスである。PI においては、討議で展開された議論内容をいかに集約するか、その議論を住民や利害関係者の間でいかに共有するかが重要である。しかし、議事録は非常に長くなる傾向があるため、これらの作業には多大な労力を要する。

例えばそのような公的討議の一例として、淀川水系流域委員会が PDF 形式で公開している議事録 [1] がある。ここでの討議 1 回分の議事録は、文数にして数百文、ページ数にして数十ページに渡り、非常に長い。人手でまとめられた抄録も存在するが、収めきれない議論が存在するため十分な解決策とは言えない。

また、討議に参加しなかった住民にとって、議事録から議論内容を理解するのは困難を伴う作業である。例えば、既に参加者間で共有されている背景情報は省略されることが多いので、閲覧しながら関連情報を調べなければ発言内容を理解できない場合があり、多大な労力を要する。

目的と課題 本研究の目的は、議事録からの意見集約や議論共有を円滑にするため、長い議事録を読む労力を軽減する閲覧支援システムの開発である。そのためには、(1) 長い議事録の全体像を把握し、所望の議論を発見する労力の軽減と、(2) 閲覧中の箇所に関連する情報を調査し、背景を含めて議論内容を理解する労力の軽減が必要不可欠である。これら 2 つの課題を解決する議事録閲覧支援システムの開発を目指す。

本研究の位置付け 従来, ArguMed [2], Araucaria [3] 等の議論可視化研究では、発話同士をリンクで結び、局所的な議論のグラフ構造を可視化していた [4, 5]。それに対し、長い議事録の閲覧支援を目指す本研究では、局所的なグラフ構造ではなく議事録の全体像を

把握する労力を軽減する必要がある (上記課題 (1))。よって本研究では、従来の議論可視化研究とは異なる新しいアプローチを採用する。すなわち、議事録全体の「流れ」を表す話題遷移図の可視化によって課題 (1) を解決し、話題遷移に応じた関連情報の提示によって課題 (2) を解決する。

さらに、公的討議の議事録に議論構造を付与したコーパスを作成・分析する研究 [6] を促進するためにも、課題 (1), (2) の解決は研究者や分析者による議事録閲覧労力の軽減に繋がるので、有意義である。

2 SalienceGraph: 話題遷移図による 議事録閲覧支援

長い議事録の「流れ」を分かり易くし、所望の議論を発見する労力を軽減するため、われわれは図 1 のように文ごとに变化する語の「目立ち具合」を定量化

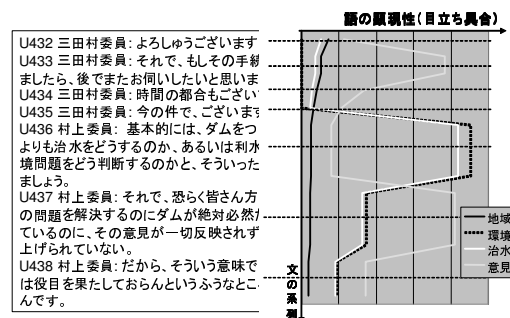


図 1 文ごとに動的に変化する語の顕現性 (salience)

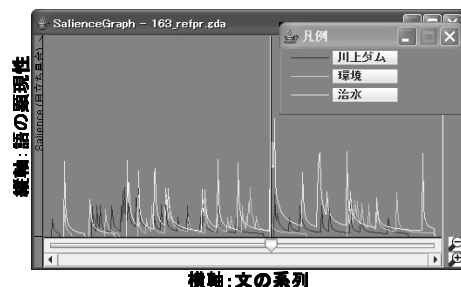


図 2 議事録中の話題遷移の可視化により、「流れ」の把握と所望の議論の発見を支援

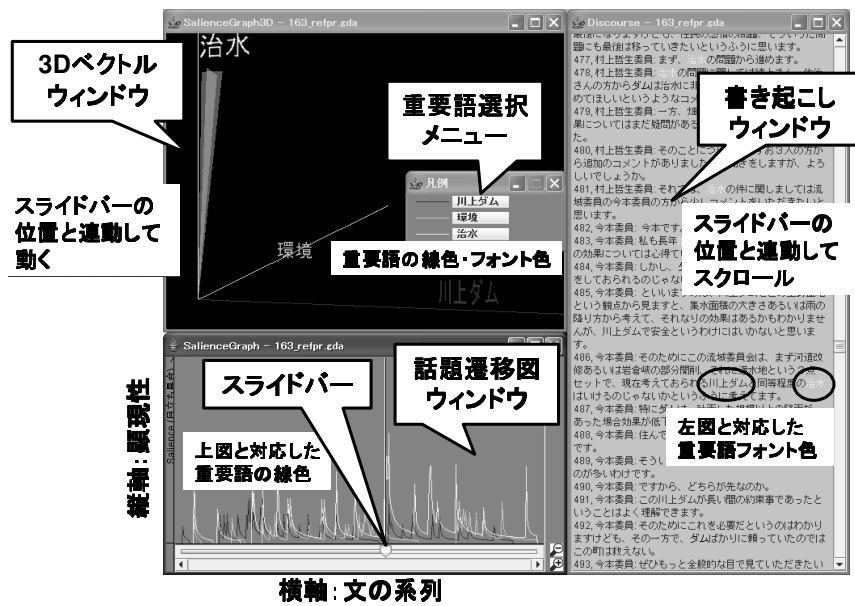


図3 SaliencyGraph: 話題遷移図を用いた議事録閲覧インタフェース

し、図2のように議事録中の話題遷移を可視化するアプローチを採用する。この図1横軸あるいは図2縦軸が表す量を、語の顕現性 (salience) と呼ぶ。では、この顕現性をどのように定量化すべきだろうか。

われわれは過去の研究 [7, 8] で、「注目されている単語ほど、次の発話で参照される確率が高い」という仮定に基づき、顕現性を以下の参照確率 (reference probability) で定義した。

$$\begin{aligned} & (\text{発話 } U_i \text{ 時点における単語 } w \text{ の顕現性}) \\ &= p(\exists w' \xrightarrow{\text{coref}} w \text{ in } U_{i+1} | \text{pre}(U_i)) \\ &= p(w | \text{pre}(U_i)) \end{aligned}$$

ただし $\text{pre}(U_i)$ は発話 U_i までの先行文脈 $[U_1, \dots, U_i]$ を表し、 $\exists w' \xrightarrow{\text{coref}} w \text{ in } U_{i+1}$ は単語 w の指示対象を参照する単語が後続発話 U_{i+1} 中に存在することを表す。 $p(w | \text{pre}(U_i))$ は参照確率の簡易表記である。

この参照確率による顕現性定義は、センタリング理論 (注意中心の連続性と代名詞化についての理論) [9] との整合性が実験により確認されている [7] だけでなく、コーパスに基づく統計的な計算が可能である [8]。この定義によって図1のような顕現性の文ごとの遷移を扱えるようになり、図2のような話題遷移図の可視化が可能になった。

現在、この話題遷移図を利用した議事録閲覧システムとして、図3に示すプロトタイプシステム SaliencyGraph を開発中である。SaliencyGraph で閲覧する

議事録に対しては、CaboCha [10] による係り受け解析と、先行文脈で出現した全名詞句 w の顕現性 $p(w | \text{pre}(U_i))$ の計算を、あらかじめ前処理として行う。言語構造を付与するための XML ボキャブラリである GDA (Global Document Annotation) [11] によりこれらの情報を議事録に自動タグ付けしておけば、SaliencyGraph で話題遷移図を可視化することができる。

SaliencyGraph は Shneiderman が提唱する *Visual Information-Seeking Mantra*、すなわち “Overview first, zoom and filter, then details on demand” [12] に基づいて設計されている。具体的には、以下のよう

Overview first ユーザは、左下の話題遷移図によって議事録全体の議題の流れを把握する。話題遷移図に現れる重要語は、重要語選択メニューからユーザが選択可能である。メニューには顕現性の総和と IDF (inverse document frequency) の積が上位の 15 語が列挙されるが、それ以外の語をメニュー最下部のテキストフィールドから入力することもできる。

Zoom and filter 話題遷移図ウィンドウ下部のスライダーを動かすと、右の書き起こしウィンドウが連動してスクロールする。例えば「治水」と「環境」についての議論を読みたい場合は、メニュー

から「治水」と「環境」を選択し、それぞれの顕現性が高い箇所へスライドバーを動かせば、右の書き起こしウィンドウに所望の箇所が表示される。この際、スライドバー周辺のみ話題遷移にズームすることもできる。

Details on demand このようにしてユーザは所望の議論を発見し、読むことができる。スライドバーを動かした際、左上の3Dベクトルの動きは、スライドバー付近の話題遷移を把握する助けとなる。

さらに、現バージョンではまだ未実装であるが、Details on demand フェーズではスライドバー位置の話題に関連する新聞記事や別の議事録を提示することで、ユーザが議論の背景を理解するヒントを与える。そのような関連情報提示機能を実現するためには、スライドバー位置の文における語の顕現性を利用して関連情報を検索すればよい。以下では、その検索手法を述べる。

3 話題遷移に応じた関連情報提示による議事録閲覧支援

スライドバー位置の文 U_i における顕現性から成るベクトル $[p(w_1|\text{pre}(U_i)), \dots, p(w_n|\text{pre}(U_i))]$ は、 U_i 時点までの話題遷移を反映したクエリベクトルとして用いることができる。なぜなら、 U_i における顕現性 $p(w|\text{pre}(U_i))$ は、そこまでの先行文脈 $\text{pre}(U_i)$ を反映した値として計算されるからである。

つまり、各文における顕現性ベクトルを用いて近傍探索をすれば、話題遷移に応じた関連情報を検索できる。

ただし、議事録に含まれる語彙数は数万を超えるため、語彙数の次元を持つ顕現性ベクトルの近傍探索には膨大な計算量が必要となる。そこで本研究では、顕現性ベクトルを pLSA (probabilistic Latent Semantic Analysis) [13] により次元圧縮した後、ANN (Approximate Nearest Neighbor) [14] を用いた近傍探索により、スライドバー位置の話題に関連する情報を検索した。その手順を以下に示す。

1. 議事録と、その背景を理解するヒントとしての検索対象 (新聞記事や資料) に対し、係り受け構造と各文における語の顕現性の値を GDA 形式で付与する。
2. 議事録中の全文における顕現性の総和と IDF の

... そこで、今いろんな問題が起きていると思うんですが、ああいう中で私どもはその 38 年間背負ってきた中で泣く泣くふるさとを離れて 8 年たちます。その間小さな農村が、ダムに賛成、反対でいろいろやって 2 つに、隣の人と口もきかないような現状があったわけなんです。それは、客観的にみれば、ああ、どこのダムの対応でもそういうことはあるやないかと皆さん方はおっしゃるかも知れません。しかし、現実にはその中にいる人間にとっては国の事業が我々住民に対して与えた影響というのはこれほど大きいんです。

閲覧対象の議事録 (太字がスライドバー位置の文と仮定)

... 本当に立派な代替案ですばつと別のところに河川をつくって、そんなもん田んぼから何から全部つぶさないと、できるわけがないやん。そんなことしたら、ダムが幾つでもできるではないか。実は、そういう観点の中で、これは流域委員会の意見を重視して住民参加、住民参加とこのパンフレットには書いてあるけども、私は実際の本当の意見というのは聞いておられないような気がしてならない。皆さん必死ではないですか。余呉町なんて立てた計画が全部狂ってしまうやないか。活性化の問題もある。

最近傍の候補 (太字の文の顕現性ベクトルが最近傍)

図 4 顕現性ベクトルに基づいて提示できる関連情報の例

積が大きい上位 n 語を抽出する。

3. 議事録および検索対象の各文における n 次元顕現性ベクトルを、pLSA により次元圧縮する。
4. ユーザがスライドバーで指定した位置の文における圧縮済みの顕現性ベクトルに対して、近傍のベクトルを持つ文を、ヒントとなる関連情報の候補とする。近傍ベクトルを持つ候補文とその先行文脈をタブウィンドウで表示し、ユーザは上位候補 5 件のいずれかをタブで選択して読む。

本稿では上記の手法の妥当性を検証するため、淀川水系流域委員会 [1] の 226 議事録 (176,971 文) を対象として関連情報の検索実験を行った。語彙数 n 、すなわち圧縮前の顕現性ベクトルの次元数 n は 30,000 とした。また、pLSA により圧縮した後の次元数は 20 とした。議論の背景を理解するヒントとしての検索対象には、閲覧中の議事録を除く 225 議事録 (約 176,000 文)、Yahoo! 新聞記事横断検索 [15] で検索した「淀川水系」を含む 774 記事 (9,989 文)、Google [16] で検索した「淀川水系」を含む 2,277 ページ (106,094 文)、「淀川水系」を含む Wikipedia [17] の 228 記事 (5,000 文) を用いた。

226 議事録中の 176,971 文それぞれに対し、本手法

で上位 5 件の関連情報候補を決定した。その一例を、図 4 に示す。図上部が閲覧中の議事録、その中の太字の文がスライドバー位置の文であると仮定した。これに対して ANN で近傍探索を行った結果、図下部の太字で表された文の顕現性ベクトルが最近傍であった。この例では、閲覧中の箇所では「ダム計画が現地住民に与えた影響の大きさ」について述べられている。これに対して検索された箇所では、そのような影響の一例である「余呉町では立てた計画が狂ってしまう」という事例とともに「住民の意見を聞いていないのでは」という問題提示がされており、議論の背景を理解するヒントとなり得る関連情報である。ただし、この例では 1 位以外の候補 (2 位から 5 位まで) は議論の背景を理解するヒントとして適切ではなかった。このことから、上位 5 候補からユーザが簡単に候補を選択できるインタフェースの重要性が示唆される。

4 おわりに

本稿では、長い議事録における議題の「流れ」を分かり易くし、所望の議論を発見する労力を軽減するため、話題遷移図の可視化手法と、それに基づく議事録閲覧インタフェース SaliencyGraph を開発した。SaliencyGraph は、*Visual Information-Seeking Mantra* に基づいて議事録閲覧支援できるように設計した。また、議論の背景を理解するヒントとしての関連情報を提示するため、語の顕現性ベクトルをクエリとして用いた関連情報検索手法を開発した。これを用いて淀川水系流域委員会の議事録を対象とした関連情報の検索実験を行い、提示できる関連情報の例を示した。今後は、本稿で示した関連情報提示機能を SaliencyGraph に追加する。さらに、所望の議論を発見する労力と省略された背景情報を把握する労力とを、本稿で示した手法により軽減できることを確かめるために、被験者実験を行う予定である。

謝辞 淀川水系流域委員会の議事録の存在をお教え頂き、議論に応じて下さった鄭蝦榮さん、小林潔司先生に感謝します。また、参照確率による顕現性定義や GDA について多大なご指導を頂いた橋田浩一先生に感謝します。本研究は科研費補助金 (特別研究員奨励費) の支援を受けました。

参考文献

- [1] 淀川水系流域委員会. 議事録一覧. http://www.yodoriver.org/doc_list/gijiroku.html.
- [2] B. Verheij. Artificial argument assistants for defensible argumentation. *Artificial Intelligence*, Vol. 150, No. 1-2, pp. 291-324, 2003.
- [3] C.A. Reed et al. Araucaria: Software for argument analysis, diagramming and representation. *International Journal of AI Tools*, Vol. 13, No. 4, pp. 961-980, 2004.
- [4] P. Kirschner et al. *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making*. Springer, 2003.
- [5] S.W. van den Braak et al. A critical review of argument visualization tools: Do users become better reasoners? In *Workshop Notes of the ECAI-2006 Workshop on CMNA*, pp. 67-75, 2006.
- [6] H. Jeong et al. Discourse analysis of public debates: A corpus-based approach. In *Proc. of IEEE SMC 2007*, pp. 1782-1793, 2007.
- [7] S. Shiramatsu et al. A Game-Theoretic Model of Referential Coherence and Its Empirical Verification Using Large Japanese and English Corpora. *ACM-TSLP*, Vol. 5, No. 3, Article 6, 2008.
- [8] S. Shiramatsu et al. SaliencyGraph: Visualizing Saliency Dynamics of Written Discourse by Using Reference Probability and PLSA. In *Proc. of PRICAI-08*, pp. 890-902. Springer-Verlag, 2008.
- [9] B. Grosz et al. Centering: A Framework for Modeling the Local Coherence of Discourse. *Computational Linguistics*, Vol. 21, No. 2, pp. 203-226, 1995.
- [10] T. Kudo et al. Japanese dependency analysis using cascaded chunking. In *Proc. of CoNLL-2002, COLING 2002 Post-Conference Workshops*, pp. 1-7, 2002.
- [11] 橋田浩一. 大域文書修飾 Global Document Annotation (GDA). <http://i-content.org/gda/>.
- [12] B. Shneiderman. *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Pearson Addison Wesley, 1998.
- [13] T. Hofmann. Probabilistic latent semantic analysis. In *Proc. of UAI 1999*, pp. 289-296, 1999.
- [14] S. Arya et al. Approximate Nearest Neighbor Queries in Fixed Dimensions. In *Proc. of SODA '93*, pp. 271-280, 1993.
- [15] Yahoo! 新聞記事横断検索. <http://gsearch.news.yahoo.co.jp/gs?ty=g>.
- [16] Google 日本語版. <http://www.google.co.jp/>.
- [17] Wikipedia 日本語版. <http://ja.wikipedia.org/>.