

音声認識された議事録からの議論マップ自動生成

Auto-Generation of Discussion-Map from Speech-Recognized Minute

趙雲超*¹ 松村真宏*² 谷内田正彦*³
Yunchao Zhao Masahiro Matsumura Masahiko Yachida

*¹京都大学大学院情報学研究科 *²大阪大学大学院経済学研究科
Graduate School of Informatics, Kyoto University Graduate School of Economics, Osaka University

*³大阪大学大学院基礎工学研究科
Graduate School of Engineering Science, Osaka University

An automatically generated minute by speech-recognition makes it difficult to reuse as it is because it contains many erroneous words. In this paper, we first specify correctly recognized words from a minute by making use of both multiple recognize candidates and time-series topic co-occurrence. Based on the specified words, we propose a method to visualize the flow of discussion as "discussion-map". Finally, we present our discoveries on experimental evaluations.

1. はじめに

会議やミーティングで議論が逸脱したり、議論の流れが掴めなくなってしまうことがよくある。その原因は様々考えられる。まず会議というのは、会議の進行をコントロールするモデレータや議長の技量に成果が左右されやすい。よってその技量に長けていない人が会議の進行役に当たれば、議論がなかなか進まない可能性が高い。また議論中には人の意見を聞かない人やただ話す要領を得ない人がいたり、複数の論点が悪く交錯したりして混乱を引き起こす。そのような状況において、議論の流れやテーマを常に把握しておくのは困難である。そこで、議論中にテーマの移り変わりやつながりを視覚情報として提示できれば議論の支援に役立つ。

本研究では音声認識された議事録からテーマキーワードを取り出し、それらを使って議論の流れや概要をグラフ化して表示することを目的とする。文脈の中で周りの多くの語とつながりを持ち、かつよく登場する語を「テーマキーワード」と呼び、議論の流れや概要をグラフ化したものを「議論マップ」と呼ぶことにする。また議論の中で展開されるテーマの表現方法は様々なものが考えられるが、本研究では名詞の集合で表せると仮定する。

1.1 関連研究

現在、音声認識を利用した様々な研究がなされている。例えば、携帯電話マニュアルの音声検索システム [1] では、ユーザーは電話音声認識サーバに電話をかけて、音声で携帯電話端末の操作に関する質問を行う。サーバは音声認識結果を用いて電子化された端末操作マニュアルを検索し、得られた結果候補を Web ページとして電話端末上に表示する。ユーザーは端末画面上で複数の検索結果候補から、目的の情報を選んで参照することができる。このシステムでは入力音声を認識する際、音響モデルと言語モデルを用いて最適な単語列を探索している。

講演録作成を目的とした講演音声自動要約の研究 [2] では講演音声の認識結果を自動的に文単位に分割した後、あらかじめ与えられた要約率に従い、文の抽出することで要約を行う。文の抽出を行う際、話題単語を多く含むかつ他の多くの意味的類似した文を重要文とみなし抽出する。

議論支援の分野でも多くの研究がなされている。例えば、電子会議システム TTS [3] では発言録の tf-idf を計算し、特徴語が大きく変化する点を話題の推移点として分割する。つぎに、各ブロックにおける要約を、キーワードの重み付けによる重要文抽出法により行うという手法を採用している。

議論構造の可視化により議論の論点の理解を促進させる研究 [4] では議事録から話題の単位 (セグメント) を同定し、さらに同定したセグメント間の関連を調べることににより、議論構造を構造化マップとして可視化するシステムを提案している。

プロジェクトを利用した議論の可視化の研究 [5] [6] ではコラジェクタ TM により議論の可視化を行い議論支援を行っている。コラジェクタ TM とは「プロジェクトを用いたコラボレーション」という意味の造語であり、PowerPoint などによって論旨・論点・論脈を可視化、共有することにより「目に見える議論」の実現を目指している。

1.2 研究目的

音声認識を用いた議論支援においては生の音声情報を解析する必要がある。しかし、一般的に音声認識の精度は低い。音声認識エンジンは入力音声に対して複数の変換候補を持っているが、第 1 変換候補のみを出力として提示するからである。しかし第 1 変換候補ではない変換候補に正解が含まれている場合がある。そこで本研究では複数の変換候補を用いることによって認識の精度を高める。音声認識には音声認識エンジン Julius [7] を用いる。

また文脈を利用して議論支援のためのテーマキーワードを抽出する。議論支援を行うためにはテーマキーワードを抽出してから、それをいかに提示できるかが重要である。ただ順番に並べるだけではキーワード同士のつながりや議論の流れまでは把握できない。そこで本研究では抽出したテーマキーワード (ノード) 同士を矢印 (エッジ) でつなぎグラフ化することによって、議論の流れや概要を分かりやすく提示する。

音声認識テキストから直接テーマキーワードを抽出するとノイズワードがたくさん含まれているので、テーマキーワードを正確に抽出するためには、音声認識テキストの中からその候補となり得るものを最初に取り出す必要がある。よって本研究の流れは、まず音声認識テキストからテーマキーワード候補を取り出し、次に候補の中からテーマキーワードを取り出す。そして最後に議論マップを生成する。研究の流れを図 1 に示す。

連絡先: 趙雲超, 京都大学大学院情報学研究科,
京都府宇治市五ヶ庄, Tel : 0774-38-3018,
E-mail : uncho@sunflower.kuicr.kyouto-u.ac.jp

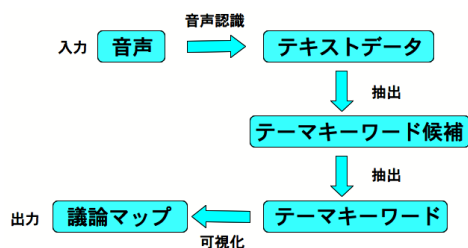


図 1: 処理の流れ

2. テーマキーワードの抽出

以下では、音声認識テキストの中からテーマキーワードを抽出するための具体的な手法を提案する。

2.1 テーマキーワード候補の抽出

まず座談会の資料を読み上げて音声認識させ、テキスト化する。次にその音声認識テキストに対して形態素解析を行い、その中から名詞のみを取り出す。形態素解析には形態素解析エンジン MeCab0.81 [8] を用いる。

形態素解析を行った後、得られた名詞集合からノイズワードを除去する。ここでノイズワードとは議論の流れに影響を与えない語を指す。本研究では一文字、語幹、非自立の語を除去する。

そして最後にテーマキーワード候補を抽出する。一般に音声認識エンジンは多数の認識候補を持っており、その内の第 1 認識候補のみを認識結果として出力している。しかし第 N 認識候補に正解が含まれている場合がある。また前後の文脈で複数回登場する語はテーマキーワードの可能性が高い。よって

1. 音声認識結果の第 N 候補まで利用する。
2. 前後の文脈を利用する。

の 2 点を考慮に入れて抽出した語をテーマキーワード候補とする。実際にテーマキーワード候補を抽出する際、議論中の局所的なテーマの盛り上がりを見るために、図 2 の解析の範囲であるウィンドを設定する。次にウィンド内で一定回数 (しきい値) 以上登場する語をテーマキーワード候補として抽出する。そして「ウィンドウサイズ w」「語のしきい値 t」として、最も抽出精度の良い w と t を実験により探索する。求まった値を用いてテーマキーワード候補の抽出を行う。

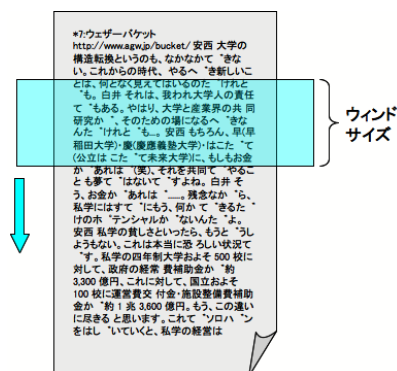


図 2: テーマキーワード候補の抽出

2.2 テーマキーワードの抽出

次に抽出したテーマキーワード候補の中からテーマキーワードを抽出する。

まず文脈中でのテーマの移り変わりを把握するために、抽出したテーマキーワード候補を座談会 [9] の発話中に出現した順に並べる。次に近くに登場するテーマ同士のつながりと流れを把握するために、N 番目の発話中に登場するノードから N+1, N+2, ..., N+n 番目^{*1}の発話中に登場するノードにそれぞれエッジを張りグラフを作成する。

また議論の中では、あるテーマを中心に議論が展開されていくのでテーマとなるキーワードはその付近に登場する多くの語と関連している。また議論の中で特定の語とペアでよく登場する語は互いに関係のある可能性が高い。そこで本研究では次の 2 つの条件を同時に満たす語をテーマキーワードとして抽出する。

条件 1 : 多くの語と関連を持っている語 (エッジ数の多い語)

条件 2 : 特定の語と関連の強い語 (共起回数の多い語)

条件 1 は大局的な視点から見たときに重要である語で、条件 2 は局所的な視点から見たときに重要である語である。テーマキーワード候補で生成したグラフの中では上記の条件 1 はエッジが多いノードを示し、条件 2 は共起回数が多い (上位 30) ノードのことを表す。よってテーマキーワード候補で生成したグラフの中からエッジの数が多くかつ共起回数が多い (共起回数の閾値より高い) ノード抽出すればそれがテーマキーワードとなる。

3. 議論マップの生成

続いて議論マップの生成方法について述べる。可視化の際 Graphviz [10] というグラフにより視覚化を行うツールを用いる。グラフによる視覚化は構造的な情報を抽象的なグラフネットワークとして表現する 1 つの手法である [10]。

議論マップの生成手順は、まず文脈中でのテーマの移り変わりを把握するために、抽出したテーマキーワード候補を座談会の発話中に出現した順に並べる。次に近くに登場するテーマ同士のつながりと流れを把握するために、N 番目の発話中に登場するノードから N+1, N+2, N+3 番目の発話中に登場するノードにそれぞれエッジを張りグラフを作成する。

またテーマキーワード候補で生成したグラフの中で、エッジの数が多くかつ共起回数が多いノードをテーマキーワードとして抽出したので、テーマキーワードの条件に満たないノードをグラフの中から削除する。

その結果残ったグラフは議論の大局的なテーマの流れをよく表していて、かつ局所的なテーマのつながりや移り変わりもよく表しているの、これを議論マップとする。生成した議論マップの例を図 3 に示す。また次に図 3 のノードを文脈に登場した順に上から下へ並び替えた議論マップの例を図 4 に示す。

4. 実験

4.1 テーマキーワード候補評価実験

音声認識テキストの中から抽出されたテーマキーワード候補が最も精度が高くなるようなウィンドサイズと語のしきい値を求めるために評価実験を行った。評価に用いるのは Precision (適合率), Recall (再現率), F-measure (F 値) の 3 つの値である。Precision と Recall はトレードオフの関係にあるので、実験

*1 予備実験の結果より本論文では n=3 とした。

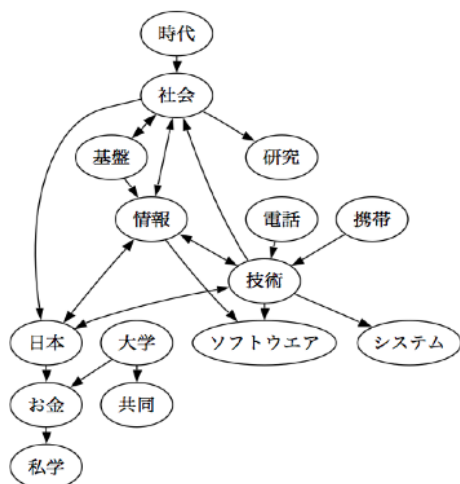


図 3: 議論マップの例 1

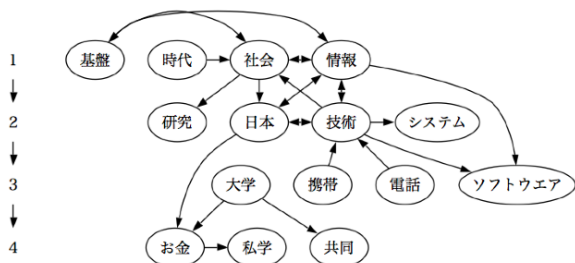


図 4: 議論マップ例 2

では F-measure が最大となるものを求める。本実験において Precision, Recall, F-measure はそれぞれ次のように定義する。

$$\text{Precision} = \frac{\text{正しく認識した名詞の種類数}}{\text{認識した名詞の種類数}} \quad (1)$$

$$\text{Recall} = \frac{\text{正しく認識した名詞種類数}}{\text{座談会中名詞の種類数}} \quad (2)$$

$$\text{F-measure} = \frac{4\text{Precision} \times \text{Recall}}{3\text{Precision} + \text{Recall}} \quad (3)$$

解析の範囲を示すウィンドサイズを 1 から 30 行、語のしきい値を 1 から 20 回の間で色々変えながら F-measure が最も高くなるウィンドサイズと語のしきい値の組み合わせを求める。

ウィンドサイズが 12、語のしきい値が 3 のときに F 値が最も高くなる。そのときの Precision, Recall, F-measure の値は以下の結果になった。

$$\text{Precision} = 0.6804$$

$$\text{Recall} = 0.8843$$

$$\text{F-measure} = 0.8226$$

よってウィンドサイズが 12、語のしきい値が 3 を用いてテーマキーワード候補の抽出を行う。

音声認識の際、第 1 認識候補のみを用いた場合、テーマキーワード候補の精度は以下になる。

$$\text{Precision} = 0.6594$$

$$\text{Recall} = 0.8961$$

$$\text{F-measure} = 0.8223$$

実験の結果、第 5 認識候補までを用いてテーマキーワード候補を抽出すると F 値が 0.8226 であるのに対し、第 1 認識候補のみで抽出したときは F 値が 0.8223 であるので、あまり向上が見られなかった。これは、ウェブに公開されていた編集済みの議事録を音声認識させて実験したために冗長な部分が削除されていた可能性があり、それが実験結果に影響を及ぼした可能性がある。しかし抽出したテーマキーワードの候補の種類は異なるので、後の議論マップの生成に影響を与えらる。

4.2 議論マップ評価実験

テーマキーワードの抽出手法が異なれば、その結果生成される議論マップも異なる。従って、どの手法を用いて作成した議論マップが議論の支援に役立つのかを評価するために評価実験を行った。

具体的な実験方法と条件について述べる。

評価実験はアンケート調査により行った。アンケートの協力者は学生 10 人である。手順としてはまず座談会の資料を読んでもらい、次に生成した議論マップを見てもらい、アンケートに従って質問 Q3 ~ Q7 (Q1 は回答者の名前、Q2 は回答者への指示) に答えてもらう。その際に用いる議論マップは手法 1 ~ 8 の 8 つの手法により生成されたものである。

抽出したテーマキーワード候補をグラフ化したときノードのエッジ数は他のノードとどれだけ多く関連しているかを示している。共起回数は特定の語とどれだけ深く関連しているかを示している。その中でエッジ数が多くて共起回数が多いノード、つまり多くの語と関連があり、かつ特定の語と結びつきが強い語をテーマキーワードとして抽出しているの手法 1 ではテーマキーワードをそのまま使って議論マップを生成する。

手法 1: ノードのエッジ数とノードの共起回数により生成次にノードのエッジ数が生成される議論マップにどのような影響を与えるのかを見るため、手法 2 ではノードの共起回数のみを条件にして議論マップを生成する。

手法 2: ノードの共起回数により生成

またノードを文脈に登場した順に並び替えて生成した議論マップと、そうでない議論マップのどちらがより議論支援に適しているかを比較するために手法 1, 2 のノードをそれぞれ手法 3, 4 で並び替える。

手法 3: ノードのエッジ数とノードの共起回数により生成 + ノードを文脈に登場した順に並び替える

手法 4: ノードの共起回数により生成 + ノードを文脈に登場した順に並び替える

最後に第 5 変換候補まで使って生成した議論マップと第 1 変換候補のみを使って生成した議論マップのどちらがより議論支援に適しているかを比較するために手法 5 ~ 8 では第 1 変換候補を使って議論マップを生成する。

手法 5: ノードのエッジ数とノードの共起回数により生成

手法 6: ノードの共起回数により生成

手法 7: ノードのエッジ数とノードの共起回数により生成 + ノードを文脈に登場した順に並び替える

手法 8: ノードの共起回数により生成 + ノードを文脈に登場した順に並び替える

質問 Q3 では上記の 8 つの手法に対してそれぞれ 3 枚の議論

表 1: 順位の平均値

	手法 1	手法 2	手法 3	手法 4
Q3	5.2	4.45	3.95	4.1
Q4	3.2	5.55	3.95	6.75
Q5	3.9	4.65	4.6	4.5
Q6	4.4	3.7	4.75	4.8
Q7	3.4	3.4	4.1	4.1

	手法 5	手法 6	手法 7	手法 8
Q3	5.25	5.3	3.85	3.9
Q4	1.75	5.25	3.1	6.45
Q5	3.85	3.75	4.95	5.8
Q6	4.85	4.25	5.05	4.2
Q7	5.1	5.1	5.4	5.4

マップを生成する。1 枚目は音声認識によりテキスト化されたデータ全体の 3 分の 1 までで生成した議論マップである。2 枚目はデータ全体の 3 分の 2 までで生成した議論マップで、3 枚目はデータ全体で生成した議論マップである。そして 8 つの手法中、1 枚目～3 枚目の流れが議論の流れをよく表している順に順位付けをしてもらった。

質問 Q4～Q7 では上記のような 8 つの方法で生成した議論マップを、議論の概要をよく表していると思う順に順位付けをもらった。ただし Q4 では議論マップを生成の際共起回数の上しい値を 3、Q5 では 4、Q6 では 5、Q7 では 6 にそれぞれ固定している。

次に実験の結果を示す。各手法に対する 10 人のアンケートの協力者が付けた順位の平均値を表 1 に示す。

4.3 考察

Q3 の結果、時系列で表示している手法 3、4、7、8 の議論マップの順位が高かった。手法 1、2、5、6 ではテーマキーワードは時系列で並んでいないので、テーマ間の流れが把握し難かった。よって議論中のテーマの流れを見るには時系列式の議論マップが見やすいことがわかった。

Q4 の結果、エッジの数を限定してテーマキーワードを抽出すると、その結果生成された議論マップの順位が高かった。テーマキーワード抽出のときエッジの数を限定しないと、生成された議論マップに孤立したノードが現れて見難かったことが手法 2、4、6、8 の順位の低さの原因であった。よってグラフ中のノードのエッジ数を考慮しないとノード同士のつながりが弱い議論マップが生成されると考察できる。

Q5、Q6、Q7 の結果、手法の違いによる議論マップ同士の差があまり見られなかったが、音声認識の第 1 認識候補のみを用いると、生成される議論マップ中に孤立するノードが出現する場面がある。よって第 1 認識候補のみを用いて議論マップを生成すると、テーマ間のつながりが反映されない場面があることがわかった。

5. おわりに

本研究では、議論支援を行うために音声認識された議事録からテーマキーワードを抽出し、議論マップを生成する手法について提案した。そのための手段として文脈中からテーマキーワードの候補を抽出しその精度について評価を行った。さらにテーマキーワードの候補からテーマキーワードを抽出する手法を提案した。そして最後に議論マップを生成しその評価を行った。今後は実際の議論の中でこのシステムを用いて評価を行っていく予定である。

参考文献

- [1] 磯谷亮輔, 畑崎香一郎, 服部浩明, 奥村明俊, 渡辺隆夫: 話し言葉認識技術とその応用, NEC 技報, Vol.58, No.5, (2005)
- [2] 堀智織, 古井貞熙: 講演録作成を目的とした講演音声自動要約, 日本音響学会 2001 年秋季講演論文集, 2-1-10, pp.67-68, (2001.10)
- [3] 古田一雄, 前原基芳, 高島亮祐, 中田圭一: 知的支援機能を備えた電子会議システム, 社会技術研究論文集, Vol.1, 299-306, (2003.10)
- [4] 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, Vol.15, No.5, pp.554-564 (2003.9)
- [5] 妹尾堅一郎: IT 環境におけるコラボレーション能力の開発: コラジェクタ TM による議論の可視化, PC カンファレンス論文集, CIEC, pp.126-127, (2001)
- [6] 桑畑幸博, 妹尾堅一郎: コラジェクタによるコンセプトワーク能力開発, PC カンファレンス論文集, CIEC, pp.276-277, (2002)
- [7] 大語彙連続音声認識システム Julius <http://www.sourceforge.jp/>
- [8] MeCab: Yet Another Part-of-Speech and Morphological Analyzer <http://www.chasen.org/~taku/software/mecab/>
- [9] 最先端科学をいかに産業へ移転するか「ユビキタス & IT」徹底討論 <http://www.sangakukan.jp/journal/main/200502/002-02/002-02.pdf>
- [10] Graph-Graph Visualization Software <http://www.graphviz.org>