

セマンティックオーサリングに基づく合議における貢献度の評価

上前田 直樹[†]

橋田 浩一^{†,‡,§}

[†] 東京工業大学大学院総合理工学研究科知能システム科学専攻

[‡] 産業技術総合研究所情報技術研究部門

[§] 科学技術振興機構 CREST

要旨

合議におけるキーパーソンを見出すことは、議論の流れを把握したり議論を活性化したりする上で重要であるが、従来は議論の意味内容に立ち入って解析するのが困難だった。本稿では、セマンティックオーサリングに基づく電子会議において、修辞関係と対話関係によって明示された議論の意味的な構造を用いて各参加者の貢献度を評価しキーパーソンを同定する方法を提案する。実際の合議の記録に関してそのような意味構造を用いて貢献度を計算し、この方法の妥当性を示す。

Evaluation of Participants' Contribution in Semantic-Authoring-Based Discussion

Naoki Kamimaeda[†]

Kôiti Hasida^{†,‡,§}

[†] Interdisciplinary Graduate School of Science and Engineering Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology

[‡] National Institute of Advanced Industrial Science and Technology

[§] Japan Science And Technology Agency

Abstract

It is important to detect key persons in a discussion in order to grasp the discussion flow and activate the discussion, but it has been difficult to analyze the semantic content of discussions. In this paper we propose a method to evaluate the participants' contributions and identify key persons by analyzing semantic structures of discussions explicitly represented in terms of discourse and dialogue relations in semantic-authoring-based discussions. This method is verified by calculating participants' contributions in the records of actual discussions.

1 はじめに

我々が何らかの意思決定を行う際、グループによる意思決定を行うことが多い。近年では、インターネット上での Web チャットや掲示板を用いた議論も行われるようになってきている。このように我々がグループによる意思決定を好む理由として「グループで話し合いを行うことによって、個人では思いもつけない優れた知恵が創発されるという期待」と「皆で決めたことは民主的であるという考え」の2つが

挙げられる [1]。

このような合議においてキーパーソンを見つけることは、議論の流れを把握したり議論を活性化したりする上で重要である。キーパーソンを見つけるためには議論の意味内容に立ち入って解析することが望ましいが、意味内容に立ち入って解析することは困難であった。したがって、従来は参加者からの投票や被参照関係等の表面的な情報のみでキーパーソンを見つけようとしていた。

我々は、修辞関係と対話関係によって議論の構造

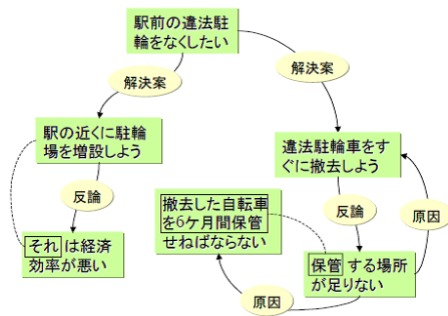


図 1: セマンティックオーサリングを用いた議論

を明示するセマンティックオーサリング [2] に基づく電子会議 (図 1 以下では共同 SA と呼ぶ) において、修辞関係と対話関係によって明示された議論の意味的な構造を用いて、各参加者の貢献度を評価しキーパーソンを同定する方法を提案する。対話関係や修辞関係といった中粒度のコンテンツ情報を利用することで議論の構造に基づいた解析が可能となり、従来は困難であった議論の意味内容に立ち入って解析を行うことが可能になる。こうすることで、表面的な情報のみでは見つけきれないキーパーソンを見つけることができると考えている。

以下では、2章で貢献度の定義、3章で貢献度の妥当性の評価とキーパーソンの同定を行う。

2 貢献度の定義

共同 SA を用いて行われた議論はグラフ型コンテンツとなる。以下の 4 つの特徴を持った発言を重要な発言と定義する。

1. 後続する多くの発言と談話・対話関係で結ばれている発言
2. 後の発言に与える影響力・情報量 (単語数) の多い発言
3. 後に続く発言のチェーンの長い発言
4. 重要リンクの端点になっている発言

議論をグラフ構造と捉え、発言の中身、対話関係と修辞関係、被参照関係に着目し、活性拡散のモデルを用いることでこれらの特徴を考慮した発言の重要度を定義することができる。そして、図 2 のような直感的に重要そうな発言の重要度を高くすることができる。グラフのノードは単文またはそれに準ず

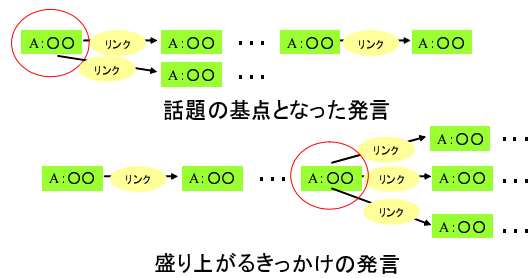


図 2: 良い意見

る分量の文章、リンクは対話関係と修辞関係である。

また、発言リンクの種類に着目すると、反対意見や解決案は議論を活性化の上で非常に重要であり、これらの意見を提示する人は非共有知識を持っている可能性が高いため、これらの意見は重要である。さらに、ある主張が成立するには何らかの根拠が必要であるから、お互いが理解しやすい議論を行うためにはただ単に主張を行うだけではなく、主張と共に論拠や根拠を述べるのが重要である [3]。また、お互いが理解しやすい議論を行うためには論拠や根拠だけでなく、具体例を挙げることや因果関係を述べることも重要であると考えられる。

これらを考慮するために対話関係と修辞関係のリンクに重要度を設け、重要度の高いリンクの端点になっている発言の重要度が高くなるようにした。また、リンクを張る際、自分の発言へのリンクと他人の発言へのリンクでは、リンクの種類 (関係) が同じでも重みは異なると考えられる。なぜなら、自分の発言に対して根拠や原因を明示することは自分の意見をわかりやすく伝えようとする上で普通の行為だが、他人の発言に対してこのようなことをするのはその人の発言を手助けしていることになり、この場合の貢献度はさらに高く設定する必要があると考えられるからである。したがって、リンクの種類によっては他人の発言に対して張る場合にはリンクの重みを増幅するようにした。共同 SA で用いるリンクの種類を図 3 に示す (我々が重要と考えている順に並べてある)。各種類の重要度は経験値を用いた。

発言の中身に関しては茶筌 [4] を用いて発言の形態素解析を行い、コミュニティ内での単語 t の総出現回数とある発言に含まれる単語 t がコミュニティ内で何回目に出現した単語 t かという出現した順番を元に発言に含まれる単語の影響力を計算し、その

重みが一定のリンク	重みが増幅されることのあるリンク
1. important dialogue acts • disapproval, solution 2. normal dialogue acts • response 3. summary • summary 4. anaphora • object 5. other dialogue acts • approval	1. derivative relations • evidence, conclusion, cause, result, et al. 2. expansive relations • example, restatement, et al. 3. multilateral relations • addition, similar, coordination, et al. 4. temporal relations • before, after, sometime, et al.
	リンクなし(重み:1.0)

図 3: リンクの種類

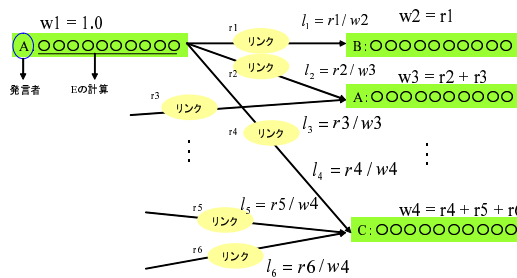


図 4: CW 値の計算

合計を発言の影響力とする。その際考慮する点としては以下が挙げられる。

1. 各発言は以前の発言の影響を受ける
2. 発言の影響力は時間とともに減衰する

これらを考慮して発言の影響力を以下のように定義する。Cはその発言に含まれる単語の集合である。

$$E = \sum_{t \in C} \log_e(\text{総出現回数}(t) - \text{出現した順番}(t) + e) \quad (1)$$

そして、発言の重み (CW: Comment Weight) は減衰定数 ρ 、(1) で求めた発言の影響力 E、リンクの重みの合計ベクトル R、接続行列 L を用いて、(2) のように定義する。ベクトル R の第 i 要素はノード i を端点とするリンクの重みの和である。また、接続行列 L は各列の合計が 1 になるようにリンクの重みに応じて正規化を行う (図 4)。

$$CW = E \cdot R \cdot (I - \rho \cdot L)^{-1} \quad (2)$$

この式を用いて CW 値を計算し、各発言の重要度とする。

そして、各参加者の貢献度 (CD: Contribution Degree) は、(3) のように各参加者の発言の CW 値を合計することによって計算する。C_n は n 番参加者の発言の集合である。

$$CD(n) = \sum_{i \in C_n} CW_i \quad (3)$$

3 検証

3.1 検証方法

既存の電子会議室のデータに人手で対話関係と修辞関係を付けたデータと実際に共同 SA を用いた議論データを用いて本手法による貢献度の妥当性を検証し、それを元に電子会議室のデータでキーパーソンの検出を行う。電子会議室のデータには「たかじんのそこまで言って委員会大会議室 [5]」のデータ (1 トピック) を用いた。議論を構造化した例を図 5,6 に示す。

まず、2 節では電子会議室のデータと共同 SA を用いた議論データを用いて、本手法による貢献度の妥当性の検証を行う。また、電子会議室の議論と共同 SA を用いた議論では議論の構造そのものが異なる可能性があるため、両者を比較し、考察を行う。

3 節では中粒度のコンテンツ情報を用いた手法 (本手法) と発言間の被参照関係のみを用いた手法 (従来手法) を用いてキーパーソンの抽出を行い、比較することで、中粒度のコンテンツ情報を用いる有効性を示す。そして、キーパーソンの特徴を考察する。

3.2 貢献度の妥当性

3.2.1 本手法の妥当性

本手法の妥当性を示すために、本手法と対話関係と修辞関係を用いない場合の重要度との比較を行った。各発言の重要度を元にランキングを作成し、上位 20%、20% ~ 40%、40% ~ 60%、60% ~ 80%、80% ~ 100% のそれぞれの範囲にどのような関係を持った発言がどれくらいの割合で存在するかを比較した。

まず、電子会議室の議論データを用いた結果を図 7 に示す。図 7 より、対話関係と修辞関係を用いることで、重要な発言の重要度ほど上位に来ているこ

とがわかる。例えば、合議において我々が最も重要と考えている「反対意見・解決案」に着目してみると、対話関係と修辞関係を用いた本手法では上位に来ているものが多いことがわかる。一方、対話関係と修辞関係を用いていない手法では、かなり低い重要度のものが存在する。発言に重要なリンクを使用しているからといって、その発言に誰も反応を示していなければその発言の重要度はそこまで高くないため一概には言えないが、筆者らが議論のデータを

確認した結果、今回使用したデータでは反対意見や解決案に対して何らかの反応が示されており、その発言がきっかけで議論が盛り上がっていた。その他の関係についても対話関係と修辞関係を用いた場合のほうがより適切に重み付けができていた。

続いて実際に共同 SA を用いて学生 3 人で「SA の問題点」という議題で議論を行い、そのデータに本手法を適用した結果を図 8 に示す。「反対意見・解決案」に着目してみると、本手法を用いたほうが上

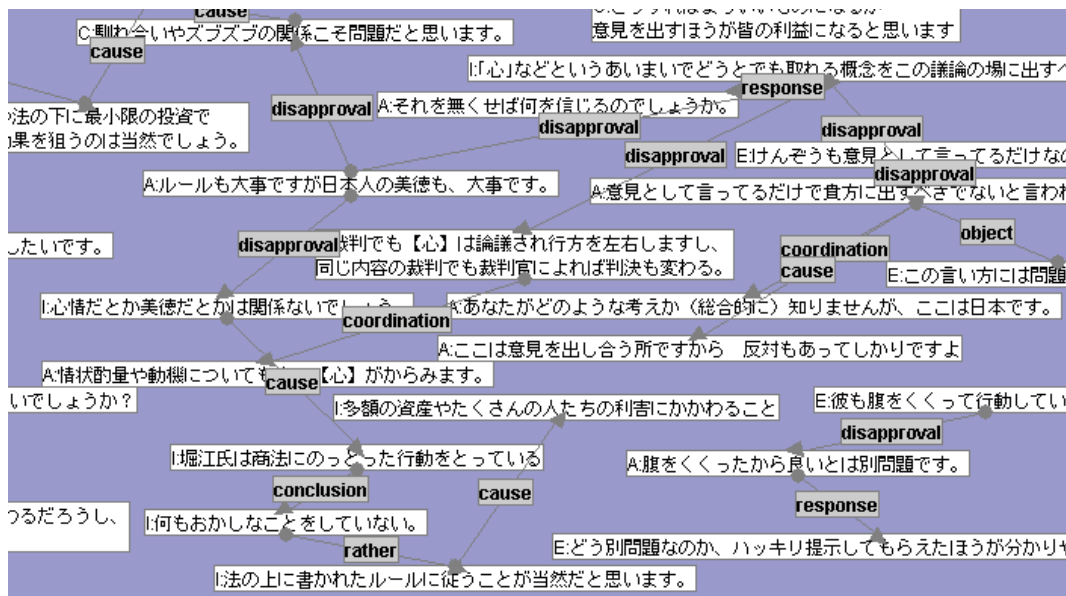


図 5: 議論の例 1

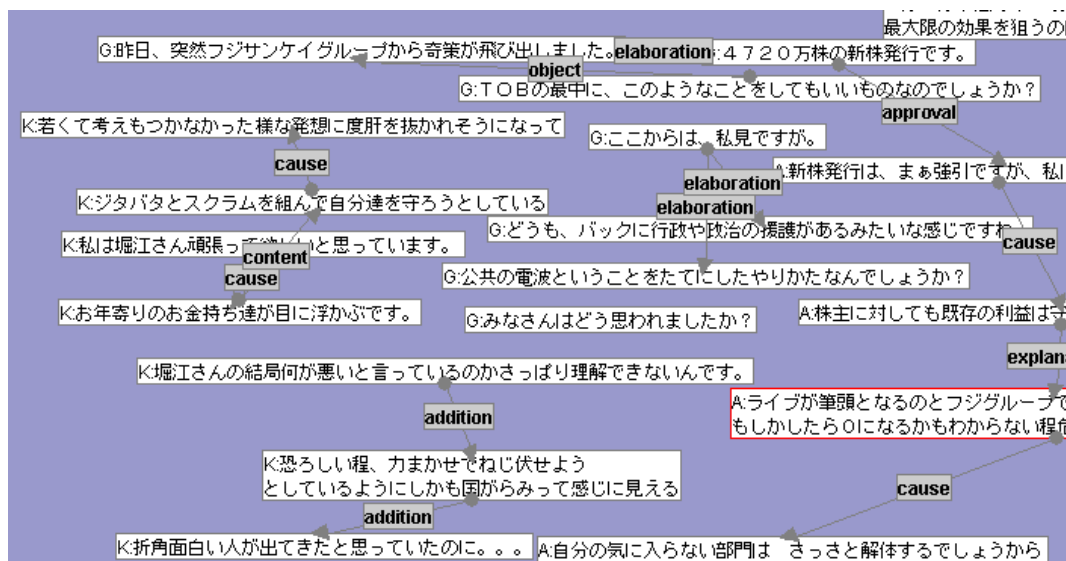


図 6: 議論の例 2

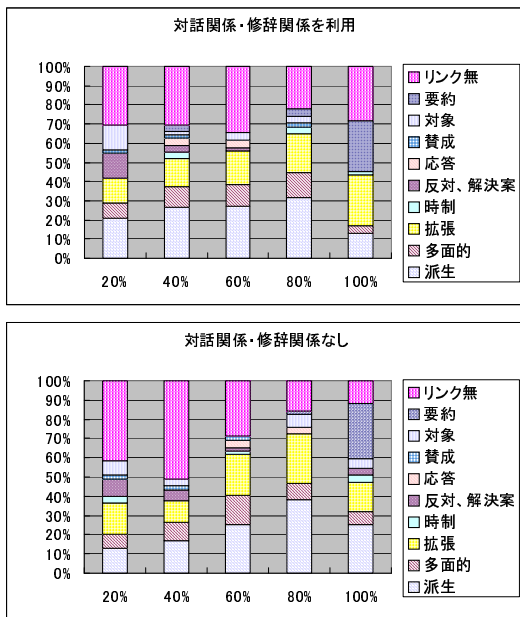


図 7: (電子会議室) 対話・修辞関係有と無の比較

位に来ているものが多いことがわかる。実際にデータを見てみると、上位に来ている「反対意見・解決案」は議論の中心的な発言となっていた。逆に同じ反対意見でも「試作段階だからしょうがない」のように発言内容が短く、その後の議論が活性化する要因となっていなかった発言の重要度は低かった。その他の関係に関してもその発言以後、議論が盛り上がっている発言や発言の論拠や根拠を述べている発言は全体的に重要度が高くなっていた。

以上より、本手法のように対話関係と修辞関係を用いることで、より正確に発言の重要度を計算できることがわかる。また、本手法では他人の意見に何らかの反応を示し、自分の意見を言う際にきちんと論拠や根拠を述べている人ほど CW 平均が高かった。これは論拠や根拠を明示し、きちんと自分の意見を参加者に伝えようとしている人の発言ほど重みが高くなっていることを示している。このことより、本手法は的確に発言の重要度を計算できていると言える。

3.2.2 電子会議室の議論と共同 SA を用いた議論

電子会議室の議論と共同 SA を用いた議論では議論の構造そのものが異なる可能性があるため、電子会議室の議論と共同 SA を用いた議論を用いて二つ

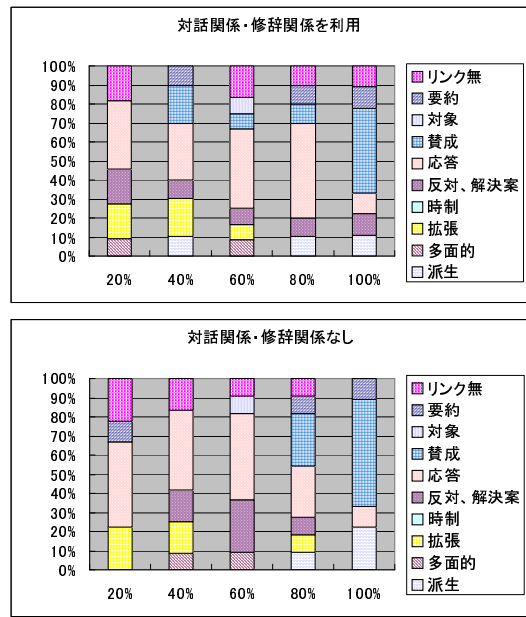


図 8: (共同 SA) 対話・修辞関係有と無の比較

の議論を比較する。2つの議論の使用リンクの傾向を図9に示す。図9からもわかるように、共同 SA を用いた議論では他人の発言に反応を示すためのリンクの使用が多かった。

従来の電子会議室の議論では、相手の意見を直感的に把握しづらいため、相手に何らかの反応を示すというよりも自分の意見を述べる傾向が強いと思われる。一方、共同 SA を用いた議論では、議論を構造化することで相手の発言を理解しやすくなり、相手の意見に対して何らかの反応を示したくなったのだと考えられる。このように、共同 SA を用いることで、従来の電子会議よりも相手の発言を理解しやすくなり、より多く相手に何らかの反応を示すようになるため、より議論を活性化させることができる可能性がある。

3.3 キーパーソンの発見

3.3.1 コンテンツ情報の有効性

本節では、電子会議室のデータを用いてキーパーソンの抽出を行う。そして、本手法と従来手法の比較を行い、中粒度のコンテンツ情報を用いることの有効性を示す。表1に本手法と従来手法の違いを示す。従来手法は発言の重みをその発言に含まれる単

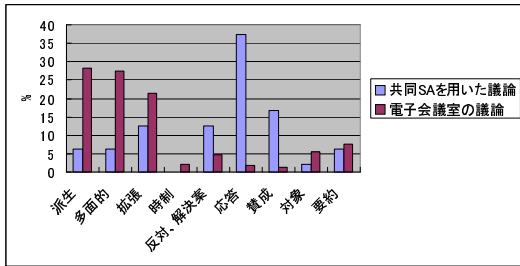


図 9: 共同 SA を用いた議論と電子会議室の議論

語数とし、被参照関係を用いて活性拡散を行う手法を用いた。

貢献度は基本的に発言数の多い人ほど高くなる傾向があるため、発言内容の情報量に極端な差がある場合には貢献度のみを基準としてキーパーソンを抽出するのは不適切である。発言数は少ないが、発言一つ一つの重みの高い参加者もキーパーソンであると考えられる。したがって、それぞれの手法について CD/CD の最大値 $\geq \Theta_1$, CW 平均/ CW 平均の最大値 $\geq \Theta_2$ の条件に当てはまる参加者をキーパーソンとして抜き出す。 Θ_1, Θ_2 は閾値である。ただし、閾値以上の参加者が 5 人に満たない時は、上位 5 位までを抜き出す。結果を表 2 に示す (太文字は閾値以上の参加者を表す)。

本手法と従来手法を比較すると、両手法とも A,B,C,F,I をキーパーソンとして抜き出している。図 5 からわかるように今回使用した議論データは、A の意見と I 意見が衝突している関係になっており、この二人の発言内容が中心となって議論が盛り上がっていた。また、B,C,F は自分の意見を根拠を明確にしながら発言していた。中でも B は発言内容の情報量が多く、多くの意見を議論の場に述べていた。したがって、これらはキーパーソンとして適当である。また、本手法では D,E,H をキーパーソンとして抜き出している。E,H はどちらも CW 平均が高く、自分の意見をきちんと根拠を示しながら発言をしていた。また、D は発言内容の情報量が多く、今後の展開までも意見として述べていた。そして、図 5 からわかるように E は A と意見の衝突が見られ、I と同じく議論の盛り上がりの中心にいた。よって、これらもキーパーソンとして適当だと考えられる。しかし、本手法では G をキーパーソンとして抜き出していない。G はこの議題の提案者であり、

表 1: 本手法と従来手法

手法	コンテンツ粒度	被参照関係の利用	対話・修辭関係の利用
本手法	中粒度		
従来手法	粗粒度		x

表 2: キーパーソン

手法	キーパーソン	貢献度による抽出	CW 平均による抽出
本手法	A B C D E F H I	A B C D E	I C A F B E H
従来手法	A B C F G I	A B I C G	B I C A F

一見キーパーソンだと思われるが、実際に発言内容を見てみると議題の提案は行っているが、それに対する自分の意見をあまり述べておらず、その意見に対しても他の参加者から反応を示されていなかった (図 6)。よって、キーパーソンとしては不適当であると考えられる。

以上より、中粒度のコンテンツ情報を用いることで、従来の表面的な情報のみでは拾うことのできなかったキーパーソンを拾うことができると言える。また、表面的な情報だけではキーパーソンに見えてしまうが、実際にはキーパーソンとは言えない参加者を誤って抜き出すことはなかった。このように、中粒度のコンテンツ情報を用いることで、より議論の内容に基づいたキーパーソンの同定を行うことが可能であると言える。

3.3.2 キーパーソンの特徴

前節で抜き出したキーパーソンの特徴を調べる。まず、貢献度の高い人と貢献度の低い人の比較を行った (図 10)。

C,H,I は、派生関係リンクの使用が多く、意見を述べるときに論拠や根拠を明示していることが多いと考えられる。しかし、H は他の参加者の意見に対して何らかの反応を示すようなリンクの使用は少ないため、発言数のわりに貢献度が低くなっていると考えられる。また、I は CW 平均も高く、意見の構造としては参加者の中で最も良かった。しかし、発言数がかなり少ないため、貢献度は低くなってしまっている。このような参加者に意見を求めることで、より議論を活性化できるのではないかと考えられる。A に関しては、発言数も多く、他の参加者の意見に対して反応を示すようなリンクの使用も多

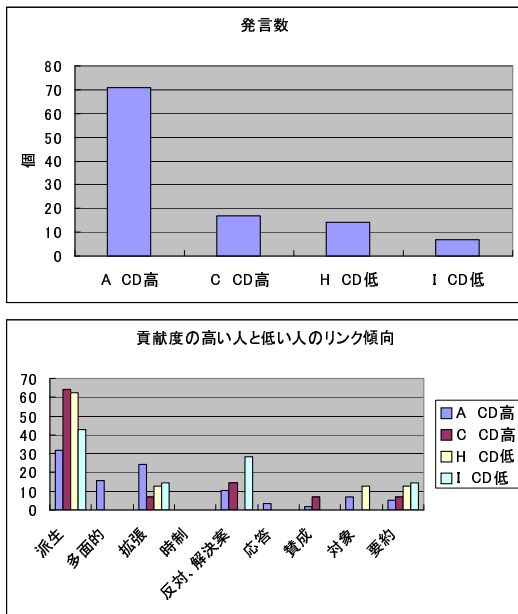


図 10: 貢献度の高い人と低い人の比較

く、その他のリンクも全般的に使用している。したがって、意見を述べる際にある程度論拠や根拠を明示していると考えられ、議論の中心的な役割を担っていると考えられる。

次に、CW 値の平均の高い人と低い人の使用リンクの傾向を比較した (図 11)。CW 平均の高い二人の特徴として、反対と解決案のリンクと派生関係のリンクをたくさん使用していることが挙げられる。これは、その人が議論において反対意見や解決案を述べ、その際に派生関係を用いて、論拠や根拠を明確化していると考えられる。実際にこの二人の発言を見ても、何らかの意見を言う際にはその原因・結果と言ったものを明示していた (図 12)。一方、CW 平均の低い 2 人は反対意見や解決案といった発言はなく、他人の意見に対して何らかの意見を言うというより自分自身の意見をただ述べているだけという傾向が強かった (図 6 の G と K)。合議の場では、自分の意見を主張するだけでなく、他人の意見に対して、何らかの反応を示すことが重要である。他人の意見に対して反応を示すことで、新しい考えが創発される可能性が生じ、議論が活性化するのである。

以上のように、貢献度は発言数と意見の構造によって決定されるが、貢献度が低くても一つ一つの意見の重みの高い参加者も存在するため、キーパー

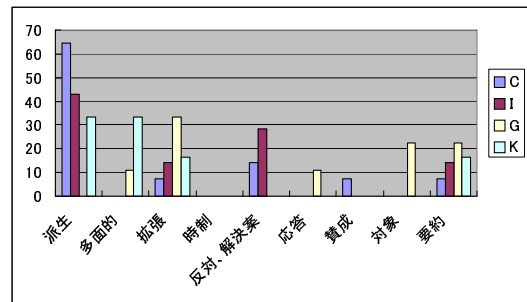


図 11: CW 平均の高い人と低い人の使用リンク傾向

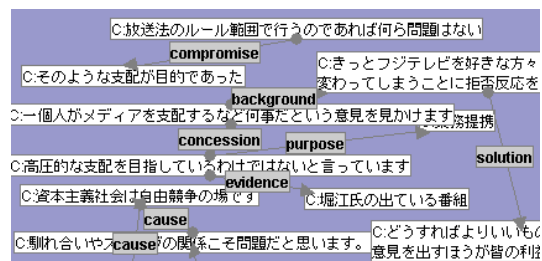


図 12: CW 平均の高い人の発言

ソンを抜き出す際に貢献度と意見の構造の両面から抜き出すことは妥当であると考えられる。また、CW 平均の高い参加者は発言を行う際に論拠や根拠を明確にしていると考えられ、このような参加者の発言を促進することで議論の場に良い意見がたくさん投入されるようになり、議論をより活性化することができるのではないかと考えられる。

4 関連研究

「議論における貢献度や参加者の役割を明示するため」や「マーケティング等に用いるため」に掲示板や電子会議室上での参加者の貢献度や発言の重要度を計算し、そのコミュニティ内でキーパーソンを見つける研究は数多くなされている。また、近年 Web 等から人と人のつながりを抽出する研究 [6] やある知識を持った人を抽出する KnowWho に関する研究 [7] もなされており、キーパーソンを発見する研究が活発に行われている。キーパーソンを見つける手法には「ユーザーからの投票によって重要度を計算するもの」「システムが何らかの基準に基づ

き自動で計算するもの」の大きく2種類に分けられる。

「ユーザーからの投票によって重要度を計算するもの」として、Yahoo!知恵袋 [8] の貢献度が挙げられる。しかし、このような手法では議論の内容とは関係のないところで参加者にコストがかかってしまい、タイムラグも発生してしまうため、システムが自動で計算することが望ましい。

「システムが何らかの基準に基づき自動で計算するもの」を以下に挙げる。小谷・岡本らは議論の活性化への参加者の貢献度を“好意的発言影響度”として表わし、各参加者の発言意欲を高めることを目的とした研究を行い、その有効性を示している [9]。好意的発言影響度は提案、賛成意見、反対意見などの発言意図を元に算出している。松村、石塚らは、ある人の発した語への興味が他の人に全般していくプロセスに着目し、テキストによるコミュニケーションの影響の普及を表すモデルを提案している [10]。コメントチェーン上を伝播していく語の割合に着目し、各々の媒体が他の媒体に及ぼす影響力である媒介影響度を算出している。友部・長尾らは議論構造をグラフ構造として捉え、重要発言からリンクがある発言や重要発言にリンクしている発言は重要であるという考えの下、活性拡散のモデルを用いて発言の重要度を算出している [11]。

我々の手法も友部・長尾らと同じように議論構造をグラフ構造と捉え、活性拡散のモデルを用いて発言の重要度を算出しているが、対話関係と修辞関係を利用している点で異なる。また共同SAでは、議論グラフのノードは単文もしくはそれに準ずる分量を想定しているため、松村・石塚らが指摘するような前の発言内容に含まれる単語が発言に含まれる可能性は低いと考えている。したがって、語の伝播の割合に着目するのは難しい。そこで、我々はコミュニティ内での単語 t の総出現回数とある発言に含まれる単語 t がコミュニティ内で何回目に出現した単語 t かという出現した順番を元に単語の影響力を定義した。

5 おわりに

本論文では、対話関係や修辞関係によって議論の構造を明示するセマンティックオーサリングに基づく電子会議において、各参加者の貢献度を評価する

方法を提案し、それを元にキーパーソンを見つけ、その妥当性を検証した。対話関係や修辞関係といった中粒度のコンテンツ情報を利用して発言の重要度と各参加者の貢献度を計算することで、より正確に貢献度を計算できることがわかった。また、それらを元にキーパーソンを見つけることで、従来手法では見つけることのできなかつたキーパーソンを見つけることができた。

合議における問題点の一つとして、「貢献度が参加者間で明らかでないため、個々の参加者の課題解決への動機付けが低下する」ということが挙げられる [1] が、今回提案した貢献度の計算手法を用いて参加者間に貢献度を明示することで、この問題は軽減できると思われる。今後は、対話関係や修辞関係といった中粒度のコンテンツ情報を用いることで議論の状態を同定し、議論の状態と各参加者の貢献度やタイプに応じて支援を行う方法について検討していきたいと考えている。

参考文献

- [1] 亀田達也:合議の知を求めて, 共立出版,1997.
- [2] Kôiti Hasida Semantic Authoring and Semantic Computing. In Koiti Hasida and Katsumi Nitta (eds.) New Frontiers in Artificial Intelligence: Joint Proceeding of the 17th and 18th Annual Conferences of the Japanese Society for Artificial Intelligence, Springer, 2005
- [3] 福沢一喜: 議論のレッスン, 日本放送出版協会, 2002
- [4] 茶筌, <http://chasen.naist.jp/hiki/ChaSen/>
- [5] たかじんのそこまで言って委員会大会議室, http://www.ytv.co.jp/takajin/bbs/takajin_bbs.cgi
- [6] 松尾豊, 友部博教, 橋田浩一, 中島秀之, 石塚満: Webの情報からの人間関係ネットワークの抽出, 人工知能学会論文誌, 20 巻 1 号 E, 2005
- [7] 岩倉友哉, 塚本浩司, 井形伸之, 津田宏: テキストからのメタデータ情報抽出と KnowWho 検索への応用, 人工知能学会全国大会 (第 18 回), 2004
- [8] Yahoo!知恵袋, <http://knowledge.yahoo.co.jp/>
- [9] 小谷哲郎, 関一也, 松居辰則, 岡本敏雄: 好意的発言影響度を取り入れた議論支援システムの開発, 人工知能学会論文誌, 19 巻 2 号 SP-4, 2004
- [10] 松村直宏, 大澤幸生, 石塚満: テキストによるコミュニケーションにおける影響の普及モデル, 人工知能学会論文誌, 17 巻 3 号 SP-B, 2002
- [11] 友部博教, 長尾確: ディスカッションマイニング: 議事録集合からの知識発見, 情報処理学会第 67 回全国大会, 2005