

オンライン議論における テキスト意見分析によるパレートフロントの算出

The Calculation of the Pareto Front by Text Opinion Analysis in Online Discussion

佐藤 元紀^{*1} 仙石 晃久^{*2} 伊美 裕麻^{*1} 白松 俊^{*2} 藤田 桂英^{*3} 伊藤 孝行^{*1}
Motoki Sato Akihisa Sengoku Yuma Imi Shun Shiramatsu Katsuhide Fujita Takayuki Ito

^{*1}名古屋工業大学大学院 産業戦略工学専攻

Nagoya Institute of Technology, School of Techno-Business Administration

^{*2}名古屋工業大学大学院情報工学専攻

Nagoya Institute of Technology, Department of Computer Science

^{*3}東京農工大学大学院工学研究院

Tokyo University of Agriculture and Technology

We developed an open online workshop system called COLLAGREE that has facilitator support functions. We deployed it for an internet-based town meeting in Nagoya city, Japan, as a city project that lead by Nagoya city mayor and evaluated. Systems such as Twitter or Facebook that support users to share opinions and ideas have been used very widely. Although these systems enable us to obtain many opinions and ideas, there is a big burden for people to read all opinions. Therefore, we analyze the opinion and compute users' utility space of their opinions. Finally, we compute parato front between two users in online discussion.

1. はじめに

Web 上の議論に関する研究分野では、多様な視点を持った大規模な参加人数による意見集約が重要な研究課題となってきた。特に都市計画および公共事業等の分野では、オープンな環境で数百人レベルの多人数意見集約システムに期待が高まっている [1]。現状では地域ごとのタウンミーティングを行っているが、時間的及び地理的コストの制約のため、数人～数十人の少人数の意見のみしか集めることができない。そこで Web 技術の発達に伴い、時間的及び空間的に離れたユーザが議論する場を提供するシステムが普及している。しかし、既存のシステムでは意見を共有することは可能だが、意見を整理や集約する仕組みは実現されていないという問題がある。

本研究で提案するシステム COLLAGREE は、意見の発散、整理、および集約すべてに関する支援機能を提供する。特に集約に関して、実際のワークショップでも議論のリードを担う役割であるファシリテータとその支援機構の導入を行い、大規模な人数での意見集約を効果的に支援することを目指した [2]。大規模な人数で行うウェブの議論では、多様な意見が得られるが、ファシリテータが全ての意見を把握するのは負担が大きい。そこで本研究では、議論の可視化し、ファシリテータが議論全体を把握することを目指す。

議論の可視化のために、ゲーム理論での交渉空間の表現方法を用いる [3]。二者交渉問題の場合、一方の交渉者の効用を小さくしなければ、もう一方の交渉者の効用を高めることができないような状態をパレート最適であると呼ぶ。パレート最適な合意案候補を結んで得られる線をパレートフロントと呼ぶ。オンライン議論におけるユーザ間のパレートフロントを求めることができれば、議論の合意案として最適な意見候補を抽出でき、ファシリテータの負担を減らすことができる。また、ユーザの交渉空間を可視化し、議論のユーザ間の意見の違いを可視化できる。可視化のためにオンライン議論におけるユーザ

のテキスト意見を分析する。次に各ユーザの意見に対する効用値を推定し、交渉空間の可視化、及びパレートフロントの算出を行った。

本論文の構成を次に示す。2. 章では、先行研究での議論システムの大規模な意見集約を支援するための機能について議論する。3. 章では、テキスト意見からユーザの効用値の推定について述べる。4. 章では、名古屋市との大規模社会実験における意見分析について述べる。最後に 5. 章でまとめとする。

2. 大規模意見集約システム COLLAGREE

2.1 Web 上の議論システム

これまで、Web 上での議論の実現を目指し、いくつかの研究が行われている [4]。西本らの研究では、関連性と異質性を併せ持つ情報を抽出し、発想的指向活動を支援する門外漢モデルの開発により議論の進行支援を行っている [5]。専門分野を異とする門外漢を 1 人参加させ、議論に新たな視点をもたらすことで、意見の発散を支援している。発散支援に有用な議論の異質性の導入を行っているが、議論の集約には着目していない。より近い先行研究として、MIT CCI のプロジェクトがある [6][7]。ここでは、インターネットを使った大規模な議論や協議を支援し、大規模な意見共有を可能にするツールが構築されつつある。プロジェクトでは、大規模な意見の共有を目指して、参加者の意見を主張、賛成反対、および問題提起などに分類し、Argumentation Map と呼ばれる論理的構造（議論マップ）を明確に共有するシステムを構築しており、その評価実験が行われている。意見集約は完全に構造化した議論マップ上で行う必要があり、議論を主張・アイデア、賛成意見、反対意見、および問題提起で組み立てていく必要がある。

また MIT CCI は、地球温暖化問題に焦点を当てて、解決プランを協議するシステムとして The Climate CoLab[8] というシステムを構築している。本システムでも、Argumentation Map を利用して意見の整理を行っている。さらに発散に向けた主となる機能として、Model-based planning を用いている。本機能は、地球温暖化に関する取り組み案を形式的に入力することで、その案が反映された世界を予想した簡単なシミュレーション結果を提供する機能である。例えば、各国の二酸化炭素

連絡先: 佐藤元紀, 名古屋工業大学大学院 産業戦略工学専攻, 〒466-8555 愛知県名古屋市昭和区御器所町, TEL: 052-745-7968, FAX: 052-735-7407, E-mail: sato.motoki@itolab.nitech.ac.jp



図 1: システムインターフェイス - 議論画面

排出量の変化を入力すると、温暖化の進行経過を確認できる。実際に自身のアイデアの実行をイメージでき、意見の発散を促している。最終的に、いくつか出た具体案に対して電子投票を行うことで、最終案の決定を行う。

本システムでは、参加者の自由な議論を重要視し、掲示板に近い形とした。なぜなら、自由な発言から斬新なアイデアの発想や問題提起ができると経験的に考えられるからである。自由な投稿による意見の発散を行い、議論の整理や集約といったプロセスを支援するための支援機能をいくつか整備する。また、議論集約の主な機能としてファシリテータの導入を試みる。

2.2 システムの実装

本システムの実装について説明する。本システムは Web アプリケーションであり、サーバサイドは Ruby on Rails により開発した。COLLAGREE は、複数のテーマについて自由に意見を投稿できる、一般的なインターネット掲示板のようなシステムをベースとしている。トップページには、議論が行われているテーマがサムネイルで表示される。トップページからテーマを選択することで、各テーマの議論画面に遷移する。本システムの議論画面を図 2 に示す。以下が、議論プロセスを支援するために実装した支援機能である。図 2 の各番号は各機能を示している。①賛成/反対の自動判定機能、②キーワード提示機能、③ファシリテーションフレーズの簡易投稿機能、④投稿並び替えおよび絞り込み機能、⑤論点タグ付加機能、⑥行動履歴（アクティビティ）機能、リマインダメール機能。①、②、および③が主なファシリテータ支援機構である。

【①賛成/反対の自動判定】返信操作に実装しており、投稿に対する各返信の意見を明確にすることができる。投稿の内容を分析し、自動的に賛成/反対を判別する。投稿が入力されると、リアルタイムに内容の賛成/反対度合いを計算し、スケールバーが移動する。判定結果に誤りがあるとユーザが判断した場合、手動で訂正することが可能である。

【②キーワード提示機能】議論内で注目されていると考えられるキーワードを抽出し、タグクラウド形式で表示する(図 2 の②)。キーワード抽出のため、テーマ内の全投稿の名詞を取り出し、TFIDF 法を用いてスコアリングしている。ユーザは、現在の議論が何に注目しているか把握することが可能になる。ファシリテータは、特にどの論点に焦点を当てていくかを検討する指標となる。さらに、単純な出現数のみではなく、「キーワード (+p, -n)」のように表示する。p, および n は上で示

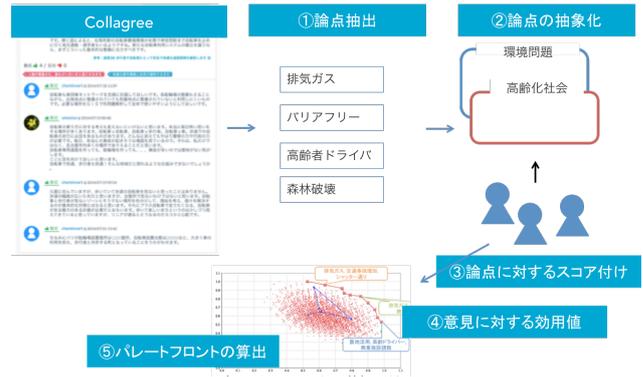


図 2: パレートフロント算出までの概要

した賛成/反対の自動判定機能によりそれぞれ、賛成/反対に判定された投稿中のキーワードの出現数である。

【③ファシリテーションフレーズの簡易投稿】ファシリテータがファシリテーションを用意に行うことができるように、ファシリテーションフレーズを用意した。ファシリテーションフレーズとは、ワークショップにおいて頻繁に意見されるフレーズである。公共政策に関するワークショップでの主催経験と共著者が所属する社会工学専攻の知見を活かし、十分に議論を重ねた上で選択した。ファシリテーションフレーズはユースケースごとに整理されており、議論状況に応じて適切に選択することが可能である。

3. パレートフロントの算出

ユーザのテキスト意見から、ユーザ自身の効用値を算出し、パレートフロントを算出する概要を示す。本研究では、「ユーザ i が重要視する論点に言及する意見は、ユーザ i にとって効用値が高い」という仮定を置き効用値の算出を行った。

まず、テキスト意見から論点抽出を行う。次に抽出した論点に対して抽象化として、クラスタリングを行う。ここで、クラスタリングを行ったクラスタ集合を論点集合と呼ぶことにする。次に、各論点集合ごとにユーザの重要度を推定する。最後に論点集合の重要度から各意見に対してユーザの効用値を算出する。

各ステップについて詳しく述べる。

【論点抽出】まず論点抽出について述べる。青木ら [9] の手法を参考にした。名詞や名詞の連続の名詞句のパターンに一致する位置に出現する単語を論点として抽出する。論点の抽出には MeCab^{*1} を用いた。

抽出時の詳細な条件を以下に示す。

- 数詞は抽出しない。
- 平仮名のみ単語は抽出しない。
- 接頭詞は名詞として扱う。
- 1文字の単語は抽出しない。

【論点のクラスタリング】論点に対して抽象化としてクラスタリングを行う。クラスタリング後のクラスタを論点集合と呼ぶことにする。論点のクラスタリングを行うために word2vec^{*2}

*1 <https://code.google.com/p/mecab/>

*2 <https://code.google.com/p/word2vec/>

を用いて論点語の単語ベクトルを学習した。word2vec の Skip-gram モデル [10] で日本語 Wikipedia のコーパスから学習した単語ベクトルを用いた。クラスタリングアルゴリズムとして k-means アルゴリズムを用いた。クラスタリング数は、論点語数/3 を k-means のクラスタ数 k とした。

【ユーザの重要度の算出】ユーザの各論点に対する重要度を求めるために共起ネットワーク分析 [11] を用いた。共起ネットワークとは、テキスト内で共起した単語をノードとし、共起した単語ごとにエッジを張ったネットワークである。

ネットワーク分析として、PageRank アルゴリズム [12] を用いた。まず、各ユーザの投稿したテキスト意見からそれぞれ共起ネットワークを生成する。次に、各共起ネットワークごとに PageRank アルゴリズムによりノードの重要度を計算した。

各ノード（各論点）の重要度を元に、論点集合に対する重要度を計算していく。論点集合に対する重要度として、論点集合内の論点語の重要度の総和を計算し、論点集合に対する重要度とした。ユーザ i の論点集合 N に対する重要度を $I_{i,N}$ 、ユーザ i の論点語 n の重要度を $V_{i,n}$ とする。 $I_{i,N}$ を式 1 に示す。

$$I_{i,N} = \sum_{n \subseteq N} V_{i,n} \quad (1)$$

【意見に対する効用値の算出】意見に対する効用値の算出について述べる。意見内の論点語を先に述べた手法で抽出する。各論点語がどの論点集合に属しているかを求め、論点集合に対する重要度の平均を意見に対する効用値とした。ユーザ i の意見 o に対する効用値を $U_{i,o}$ 、ユーザ i の論点集合 N に対する重要度を $I_{i,N}$ 、意見 o に含まれる論点語の集合を $issues$ 、論点語 n を引数として与えるとどの論点集合 N に属しているかを返す関数を f とする。 $U_{i,o}$ を式 2 に示す。

$$U_{i,o} = \frac{\sum_{n \subseteq issues} I_{i,f(n)}}{|issues|} \quad (2)$$

各ユーザの効用値は 0.0 ~ 1.0 に正規化し、正規化後の効用値を元にパレートフロントを算出した。

4. 大規模議論のパレートフロント算出実験

4.1 実験内容

名古屋市との共催のもと、本システムを用いた大規模社会実験を行った。社会実験の議論内容からテキスト分析をし、パレートフロントの算出を行った。

【議論実験設定】共催：名古屋市役所，参加者数：266 人，実施期間：2013 年 11 月 19 日（火）午後 12 時 ~ 12 月 3 日（火）午後 12 時，議論テーマ：名古屋市次期総合計画に関する 4 題，ファシリテータ：専門家 9 名

参加者数は、システムに登録を行った人数であり、すべて無償で登録している。266 人が発言可能な状態であるタウンミーティングの実現は簡単ではなく、本システムの利点である。さらに、本実験では日本ファシリテーション協会（FAJ）の協力のもと、ファシリテータの専門家 9 名が参加し、各テーマの議論で支援機能を用いたファシリテーションを行った。ファシリテータの投稿を除外すると全意見の投稿数は 108 件であった。

【評価実験設定】議論テキストから各参加ユーザの効用値を推定し、二者間のユーザのパレートフロントを求めた。発言数が多かった 5 人のユーザについて実験対象とし、 ${}_5C_2 = 10$ 組についてアンケートを実施し評価を行った。

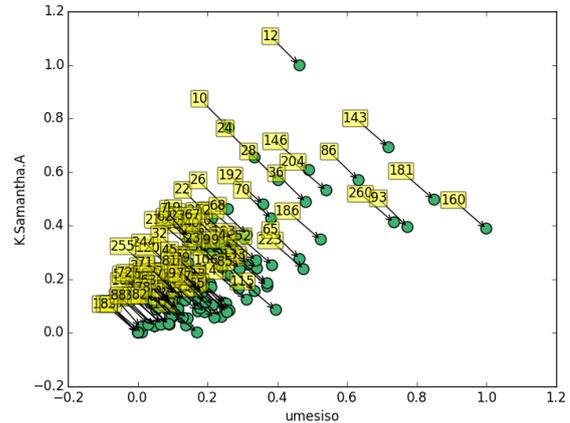


図 3: 二者間の交渉空間

参加者：4 人，評価日時：2015 年 3 月 23 日（月）

評価実験を実施し 4 人にアンケートを実施した。まず、評価実験参加者に発言数が多かった 5 人のユーザの全てのテキスト意見を読んでもらう。その上で以下の 2 つの質問を実施した。

- Q1. 本研究で推定した効用値を見ていただきユーザの効用値として適切かどうかを答えてもらう。
- Q2. 二者間のパレートフロントの意見が二者間の合意案として適切かどうかを答えてもらう。

4.2 実験結果

発言数の多い上位 5 人のユーザ「umessiso さん」, 「Nayabashi5 さん」, 「それいゆさん」, 「K.Samantha.A さん」, 及び「かつやんさん」について効用値を推定し、二者間のパレートフロントを求めた。

「umessiso さん」と「K.Samantha.A さん」の二者間のユーザの各意見に対する効用値を推定し、図 3 に示す。緑の丸が各意見を示し、黄色の四角が意見 id を示している。この場合、パレートフロントは「12」, 「143」, 「160」, 「181」と求めた。パレートフロントの意見について表 4.2 に示す。パレートフロントの算出によって 108 件から 4 件に絞ることができた。

文書 id	投稿したユーザ	意見内容
12	K.Samantha.A	水辺の利用について
143	それいゆ	ごみの分別
160	umessiso	高齢者の集まりの場について
181	umessiso	自然とコミュニケーション

実験結果として、「Q1. 各ユーザの効用値の推定として適切だと思いますか？」の回答結果を図 4 に示し、「Q2. 各パレートフロントは適切に表現できていると思いますか？」の回答結果を図 5 に示す。

図 4 の結果より「umessiso さん」の「効用値の推定として適切だと思いますか？」という質問に対して、3 人のユーザが「そう思う」と答えおり、「K.Samantha.A さん」も 1 人が「非常にそう思う」2 人が「そう思う」と回答している。一方、「それいゆさん」について、1 人が「そう思う」3 人が「どちらでもない」と回答しているため、「umessiso さん」の推定結果と比較し適切な効用値が推定できていないと考えられる。

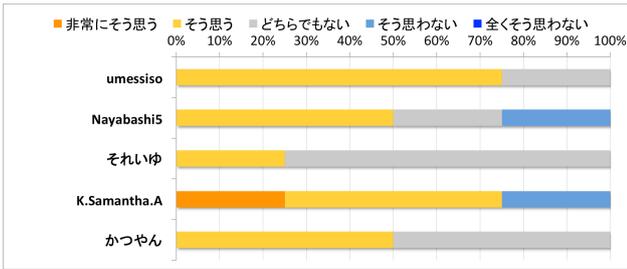


図 4: Q1. 各ユーザの効用値の推定として適切だと思いますか?

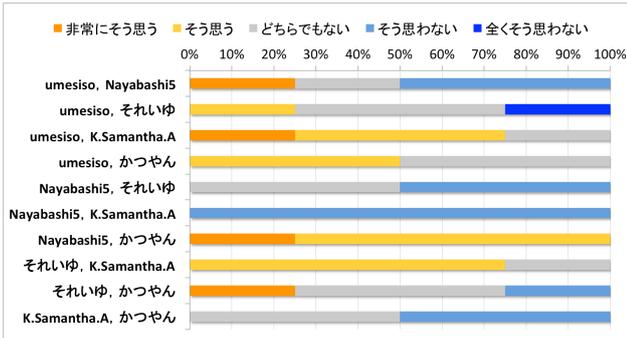


図 5: Q2. 各パレートフロントは適切に表現できていると思いますか?

図 5 の結果より「Nayabashi5, かつやん」の二者間のパレートフロントに関して“パレートフロントとして適切だと思いますか?”に対して、1 人が「非常にそう思う」3 人が「そう思う」と回答しており適切なパレートフロントが求められていると考えられる。一方、「Nayabashi5, K.Samantha.A」について、4 人が「そう思わない」と回答しており適切なパレートフロントが求められていないと考えられる。

5. まとめ

本研究では、大規模意見集約支援システム COLLAGREE におけるテキスト意見の分析を行い、ゲーム理論における交渉空間の表現方法を用いて議論の可視化を行い、パレートフロントの算出を行った。ユーザの効用値を推定するために、テキスト意見から論点抽出をし、論点集合のクラスタリングを行った。各ユーザの意見語の共起ネットワークによる重要度の算出を行い、各意見に対してユーザがどの程度重要視しているかを推定した。

評価実験として人手によるアンケートを実施し、効用値が適切に推定できているか、二者間のパレートフロントが適切かどうかを評価した。結果としてユーザの効用値として適切であると考えられる結果を得ることができたユーザもいたが、一方で二者間のパレートフロントを算出した際に全く適切でないと考えられる結果となった。そのため、評価者を増やし評価を再度行うことや、実際の発言者にアンケートを実施し評価するなど評価方法の検討が必要である。

今後の課題として、意見の集約に向けてパレートフロントだけではなく、パレートフロント付近の意見を抽出し議論を進める機能をシステム内に実装し、評価する必要がある。また、本研究ではまず二者間でのパレートフロントを算出したが、多人数でのパレートフロントを算出し大規模議論でのファシリテーション機能として実装する必要がある。

参考文献

- [1] 伊美裕麻, 伊藤孝行, 伊藤孝紀, 秀島英三. ファシリテータ支援機構に基づく大規模意見集約システム collagree の開発と評価 名古屋市次期総合計画のネット上のタウンミーティングでの社会実験. 情報処理学会第 76 回全国大会, Jan 2014.
- [2] Takayuki Ito, Yuma Imi, Motoki Sato, Takanori Ito, and Eizo Hideshima. Incentive mechanism for managing large-scale internet-based discussions on collagree. *Collective Intelligence 2015*, 2015.
- [3] 中山幹夫, 船木由喜彦, 武藤滋夫. 協力ゲーム理論. 協力ゲーム理論, 2008.
- [4] Takayuki Ito, Yuma Imi, Takanori Ito, and Eizo Hideshima. Collagree: A facilitator-mediated large-scale consensus support system. *Collective Intelligence 2014*, 2014.
- [5] 西本, 間瀬, 中津. グループによる発散的思考における自律的情報提供エージェントの影響. 人工知能学会会誌, Vol. vol.14, No. No.1, pp. 58-70, 1999.
- [6] Mark Klein. Achieving collective intelligence via largescale on-line argumentation. *CCI Working Paper 2007-001*, Vol. MIT Sloan School of Management 4647-07, , 2007.
- [7] L. Iandoli, M. Klein, and G. Zollo. Enabling on-line deliberation and collective decision-making through large-scale argumentation: A new approach to the design of an internet-based mass collaboration platform. *International Journal of Decision Support System Technology*, Vol. vol. 1, , 2009.
- [8] Joshua Introne, Robert Laubacher, Gary Olson, and Thomas Malone. The climate colab: Large scale model-based collaborative planning. *IEEE Collaboration Technologies and Systems (CTS)*, Vol. 2011 International Conference, , 2011.
- [9] 青木伸也, 湯本高行, 角谷和俊, 新居学, 高橋豊. 論点に対する極性に注目したニュース記事からの編集意図の抽出手法. 情報処理学会研究報告. データベース・システム研究会報告, Vol. 2009, No. 16, pp. 1-8, 2009.
- [10] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [11] 吉田知訓, 間瀬心博, 北村泰彦. 質問応答 web サイトからの関連語ネットワークの自動抽出. 電子情報通信学会技術研究報告. AI, 人工知能と知識処理, Vol. 108, No. 119, pp. 75-80, 2008.
- [12] Lawrence Page, Sergey Brin, Rajeev Motwani, and Winograd. The pagerank citation ranking: Bringing order to the web. 1998.