

# ファセット学習モデルを用いた公的討議のプロトコル分析\*

## Protocol Analysis of a Public Deliberation Using Facet Learning Model\*

鄭 蝦榮\*\*・羽鳥 剛史\*\*\*・小林 潔司\*\*\*・白松 俊\*\*\*\*

By Hayeong JEONG\*\*・Tsuyoshi HATORI\*\*\*・Kiyoshi KOBAYASHI\*\*\*・Shun SHIRAMATSU\*\*\*\*

### 1. はじめに

公共事業による社会資本整備は、限りある大切な資源を有効に活用し、国民の社会厚生を改善する役割を担う。現在社会の複雑性が増大する中、公共事業がその社会的要請を果たす上で、地域住民、企業、有識者等、事業に関わる様々な関係主体による公的討議の重要性が指摘されている。実際に、多くの公共事業において、パブリックインボルブメント(PI)が導入されているが、その中で関係主体間の公的討議が重要な役割を担っている。このような公的討議の目的は多様であるが、専門家や一般住民による討議を通じて、公共事業に関わる相互理解を深めるとともに、当該事業の社会的適切性に関わる判断情報を得ることが期待されている。

しかしながら、多様な利害関心を有する討議参加者は、公共事業に対して様々な期待や信念を有しており、参加者間で事業に対する認識が異なる場合が少なくない。公共事業を巡る認識の不一致は、円滑なコミュニケーションを阻害する要因であり、このような認識の不一致が存在する時、討議参加者間で社会的コンフリクトが生じる可能性が存在する。また、討議参加者がどのような関心や期待を有しており、どのような認識の不一致が存在しているかについての明確な理解が得られないまま、討議が終了することも懸念される。

公共事業に関わる認識の不一致問題に適切に対処する上では、公的討議における参加者の認識の共有化プロセスや学習メカニズムについて理解し、公的討議を適切に

進めていくことが重要であるが、そのためには、討議参加者の発言内容や参加者間の認識の不一致についての確に把握しておくことがまずもって重要である。

以上の問題意識に基づき、本研究では、公共事業を対象とする公的討議に対してプロトコル分析を実施し、討議参加者の認識構造を把握するための方法論を提案する。この目的の下、統計的言語学習モデルを用いて、討議参加者の発言をいくつかの概念カテゴリー(ファセット)に類型化し、それを基にして、認識の不一致や意見対立状況を明らかにすることを試みる。その際、公的討議のテキストデータに自然言語処理とパターン認識処理を施し、ファセット分類を行う。

### 2. 本研究の基本的な考え方

#### (1) 公的討議(Public Deliberation)の概念

公共事業を対象とした公的討議として、審議会、説明会、ワークショップ等、多様な実施形態が検討されているが、公的討議そのものを一義的に定義することは難しい。欧米で「討議」の概念を持つ用語として、「Debate」「Dispute」「Discussion」「Deliberation」が挙げられるが、ここで、Debateは「誰かとの口論や弁明」、Disputeは「論争」、Discussion「説明、観察、公平な決定」、Deliberationは「(i) 熟考・考慮、(ii) 自由化・開放」を意味するものと考えられている。このように公的討議は多義的な概念であるが、公共事業に関わる意思決定においては、その中でも、特に「Deliberation」が重要な概念として位置づけられている。Joseph M.<sup>1)</sup>は、deliberationについて「公共政策の利点について論証すること」と定義し、そのプロセスの重要な要素として情報(information)、主張(argument)、説得(persuasion)を挙げている。J. Elster<sup>2)</sup>は、熟議の民主制において、「討議参加者による合理的なかつ公正な価値の提示と主張に基づく意思決定」が求められると論じ、そのプロセスの要素としては、交渉(bargaining)と主張(argument)を挙げている。GutmannとThompson<sup>3)</sup>によれば、Deliberationの核心アイデアは、討議参加者が倫理的な観点から同意できないときは、参加者が受け入れられる決定に至るまで論証を続けることにあり、そのプロセスの要素につい

---

\*キーワード: 計画手法論、市民参加、プロトコル分析

\*\*学生員、工修、京都大学都市社会学研究科

(京都市西京区京都大学桂、

TEL 075-383-3224、FAX 075-383-3224)

\*\*\*正員、助教、東京工業大学大学院理工学研究科

(東京都目黒区大岡山2-12-1、

TEL 03-5734-2590、FAX 03-5734-2590)

\*\*\*正員、教授、京都大学都市社会学研究科

(京都市西京区京都大学桂、

TEL 075-383-3224、FAX 075-383-3224)

\*\*\*\*非会員、工修、京都大学大学院情報学研究科

(京都市左京区吉田本町

TEL 075-753-4952、FAX 075-753-5977)

ては、互惠性 (reciprocity)、公開性(publicity)、説明責任(accountability)、基本的な自由(basic liberty)、基本的な機会(basic opportunity)、公平な機会(fair opportunity)がDeliberationを構成すると主張されている。

これらの既往研究を踏まえれば、Deliberationの中心概念は、討議参加者が考慮する合理的、公的な価値について自由な論証を通じて決定していくプロセスであると捉えることが出来る。その主要な関心事項として、発言内容についての論証及びその論証プロセス、公的な関心(価値)、協調の形態などが挙げられる。そして、理想的な公的討議(Ideal Deliberation)とは、これらの要件の成立可否を判断し、欠陥している要素を同定し、補完していくものと考えられる。以上の考え方に基づき、本研究では、公的討議の分析枠組みとして、討議参加者の発言の(i)公共性(公的な価値について述べているか否か)、(ii)協調性(協調的な態度であるか否か)、(iii)論証性(理由を提示しているか、単なる主張か)、(iv)賛否(同義の余地があるか否か)を取り上げ、これら四つの要件に基づいて、討議参加者の発言を分析する。

## (2) 公的討議のプロトコル分析

本研究では、上記の考え方に基づき討議参加者の発言内容を分析する上で、公的討議のプロトコル分析を実施する。プロトコル分析は、発話思考法によって得られたデータを分析するための認知科学の方法論として、Ericsson and Simonによって提唱された。ここで、プロトコルとは広義に「人の発話行為によって得られた言語的なデータ」を表すものと解釈されている。プロトコル分析の特徴は、発話者の認知過程を通じて生起する現象として、プロトコルデータを捉える点にある。すなわち、プロトコル分析は、単に発話内容だけではなく、発話をもたらした発話者の認知過程を明らかにすることを目的とする。特に、公共事業を対象とする公的討議においては、多様な価値観や認識を有する参加者の間で広範な文脈の下で議論が展開される。そのため、個々の発言自体をそれぞれ独立に切り出して分析するのではなく、発言とそれが生起する様々な文脈との関わりの中で、討議参加者の認識構造を明らかにすることが重要である。本研究で提案するプロトコル分析は、討議参加者の発言の類型化を通して、参加者間の認識の不一致を明らかにする点に特徴がある。

## (3) ファセット理論に基づく発言の類型化

公的討議における参加者の認識を発言内容から類型化し、討議参加者の認識構造や認識の不一致を検討することを試みる。その際、先行研究に基づき<sup>5)</sup>、討議参加者の発言が幾つかのファセット(概念カテゴリー)の要素の組み合わせで構成されるものとして、個々の発言の空間

構造を規定する。本研究では、2(1)の理想的討議の要件に基づいて、表-1に示す四つのファセットを設定し、参加者の発言をこれらのファセットから構成される枠組みの中で規定する。

表-1 ファセットの分類

ファセット a) 公的利害関心、私的利害関心
ファセット b) 協調的表現、非協調的表現
ファセット c) 説明的表現、主張的表現
ファセット d) 賛成表明意見、反対表明意見

ここで、ファセット a) は発言の公共性を表し、その参加者が公的価値に基づいて述べているか否かを基準として発言を分類する。ファセット b) は、発言の協調性を表し、参加者が協調的な態度を示しているか否かに基づいて発言を分類する。ファセット c) は、発言の論証性を表し、参加者が自分の主張の理由を述べているか否かに基づいて発言を分類する。ファセット d) は、発言の賛否を表し、参加者が賛成意見を述べているのか反対意見を述べているのかに基づいて分類する。例えば、「明らかなのは、ダムができると川が死にます」という発言については、ファセット a) は公的利害(関心対象の公共性)、ファセット b) は、非協調的表現(参加の態度)、ファセット c) は主張的表現(理由の不説明)、ファセット d) は反対表明意見(利害の方向)と類型化することが出来る。このような発言の類似化を通じて、認識の不一致を定量的に把握するとともに、参加者の認識の変化や参加者間の相互作用による影響を時系列的に観察することが可能である。

## 3. ファセット学習モデルを用いた発言類型化手法

### (1) パターン認識によるファセット自動分類の概要

ファセットを機械的に分類するためにパターン認識という統計機械学習手法を用いる。パターン認識とは、文字や映像や音声などのような入力データから、その特徴を取り出し、標準的なパターンに基づいてクラスを判別させる技術である。近代よく使われているパターン認識の統計機械学習手法として、NN(Neural Network)とSVM(Support Vector Machine)がある。NNは、人間の脳の構造をまねて作った情報処理機構として記憶→連想→認識という段階を経てパターン認識をさせる手法である。NN手法については過剰学習の問題、収束の遅さ、局所最適解への収束、モデルの選択の恣意性などのいくつかの問題が指摘されている。一方、SVMはVapnikらが提案した手法であり、最適超平面法とKernel法でパターン認識を行う。最適超平面法により、学習データと識別境界の間隔(マージン)を取り扱い、その中で最小距離を持つマージンを最大化にする戦略に基づく識別境界の位置を

決定する明確な基準を持つ。さらに、Kernel 法との組み合わせにより、非線形のモデル空間を仮定することや、複数の素性の組み合わせを考慮した学習が可能である。したがって、学習データの次元数に依存しない極めて高い汎化能力を持つ。未学習データについては、識別超平面を基準として、異なるクラスを識別できる<sup>6)7)</sup>という優位性を有していることから、本研究ではSVMを用いてファセットを機械的に分類する。

一般に、SVMは2値分類器である。本研究では、表-1のように一つのファセットの要素を二つに設定して、要素i(●)、要素ii(○)、要素が欠陥している(●)、三つに分類する。これは、2値分類を2回実施して、図-1のような二つの境界超平面を求めることで、ファセットの要素を三つに分類することができる。

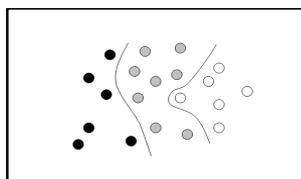


図-1 ファセット分類の仕組み

(2) SVMを用いたファセット分類方法

ファセットの自動分類は、図-2のように行われる。



図-2 ファセット自動分類の手法

第1段階では、まず、どのようにデータを分類するかを設定する。ここでは、表-1のファセット分類を学習モデルとして用いる。次に、学習モデルに従って、調査者が各発言に対してファセット要素の値を付与し、その結果から学習データを取得する。この際、精度が高い標準パターンを得るために、二人以上の観察者が作業を行い、一致したパターンを標準パターンとする。

第2段階では、調査者による学習データを用いて機械学習を行う。まず、学習器を使うために自然言語処理を施し、自動的にラベリングする。具体的には、Global Document Annotation (GDA)によりメタ言語を得て、n次元の空間Rを仮定し、空間R上に各発言を表すベクトル $u = \{t_1, t_2, \dots, t_n\}$ を配置する。ここで、空間上の各座標は、それぞれ分析対象となるデータ内の各単語を表している。したがって、次元nは全単語数を表している。そして、発言ベクトルの要素について、単語iがその発言内に登場する場合、 $t_i=1$ 、そうでなければ、 $t_i=0$ と設定する。発言ベクトルuに、その発言に該当するファセットの要素の値 $y_i (\in \{-1, 0, 1\})$ を加えることで、学

習データの集合 $S = \{(u_1, y_1), \dots, (u_j, y_j)\}$ が得られる。ここで、 $y_i$ は、そのファセットの要素が要素iであれば1、要素iiであれば-1、要素欠陥であれば0とする。集合Sに対して空間上の位置(標準パターン)を学習させる際にSVMを用いて実行し、候補となる識別超平面を求め、データを学習させる。最後に、調査者による分類とSVMによる分類との一致率を確認することから、機械学習の精度(Accuracy)の良さを評価し、学習結果の信頼性を図る。精度を測る手法としては、K-fold Cross Validation法を用いる。これは、既存のサンプルの一部を評価用に指標構成のためのサンプルからK単位で切り離し、既存のサンプルの中で指標構成と評価を行う手法である。本研究ではKを10と設定した。

4. 事例適用

淀川流域整備計画における「住民との意見交換会」の一つの議事録から総1393発言を得て調査者かつ機械による学習を行い、表-2のような結果が得られた。全体の精度は60%ほどであるが、要素によってはデータの貧弱さから精度が低い場合があり、今後より精度の高い結果を得るためには、データの量および質を補完することが必要である。

表-2 学習結果とその精度

総 発言 1393	Facet①		Facet②		Facet③		Facet④		
	公私利害		肯定/否定		説明/主張		賛成/反対		
	数	%	数	%	数	%	数	%	
調 査 者	●	498	35.8	401	28.8	276	19.8	331	23.8
	○	112	08.0	229	16.4	403	28.9	113	08.1
	●	938	67.3	918	65.9	869	62.4	1104	79.3
S V M	●	599	43.0	443	31.8	173	12.4	486	34.9
	○	47	3.4	175	12.6	417	29.9	81	5.8
	●	747	53.6	775	55.6	803	57.6	826	59.3
精度	894	64.2	859	61.7	819	58.8	836	60.0	
再 現 率 Recall	334	67.1	221	55.1	69	25.0	177	53.5	
	6	05.4	67	29.2	194	48.1	10	08.8	
	554	70.8	571	74.8	556	77.9	649	68.4	
適 合 率 Precisi on		55.8		49.9		40.0		36.4	
		12.8		38.2		46.5		12.3	
		74.2		73.7		69.2		78.6	
F値		60.9		52.4		30.7		43.3	
		07.5		33.2		47.3		10.3	
		72.4		74.3		73.3		73.1	

表-3は、司会者(S10,S11,S12)を除く参加者の認識構造を表したものである。ファセット①については、全参加者が公的利害を認識しているが、他のファセットについては認識の不一致が存在することが分かる。

表-3. 学習結果による各参加者の認識構造

		Facet①		Facet②		Facet③		Facet④	
		公私利害		協調/非協調		説明/主張		賛成/反対	
S1	人 機	78/18	●	70/24	●	52/53	●	81/1	●
		80/3	●	56/18	●	21/51	○	61/8	●
S2	人	24/17	●	20/28	○	18/33	○	27/0	●

	機	42/7	●	24/18	●	9/30	○	31/6	●
S3	人	93/34	●	54/62	○	34/95	○	79/18	●
	機	104/7	●	34/95	○	18/101	○	105/6	●
S4	人	45/4	●	24/22	●	19/30	○	30/16	●
	機	48/0	●	27/11	●	14/23	○	40/4	●
S5	人	30/31	●	55/17	●	41/34	●	34/6	●
	機	38/12	●	42/10	●	15/36	○	40/0	●
S6	人	16/2	●	14/4	●	7/11	○	0/0	●
	機	25/1	●	17/3	●	8/14	○	13/6	●
S7	人	17/0	●	14/3	●	3/4	○	8/7	●
	機	16/0	●	13/4	●	7/9	○	14/6	●
S8	人	67/2	●	48/25	●	27/49	○	5/44	○
	機	51/3	●	45/11	●	13/35	○	32/12	●
S9	人	37/2	●	28/11	●	26/16	●	9/8	●
	機	71/1	●	53/16	●	23/33	○	48/10	●
S13	人	26/2	●	29/2	●	13/23	○	17/8	●
	機	28/0	●	23/4	●	5/21	○	20/7	●
S14	人	14/0	●	14/0	●	5/9	○	5/0	●
	機	14/0	●	14/2	●	4/11	○	9/4	●

以下の図-3、4は、S3(賛成市民)、S8(反対専門家)の認識の変化を時系列に表したものである。S3の場合は、横軸を基準として公的利害(青い線)、非協調的(ピンクの線)、主張的(黄色い線)、賛成(緑色の線)の認識を持つ傾向が見られる。

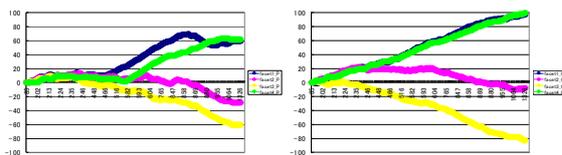


図-3 S3の認識変化

(左：人による分類、右：SVMによる分類) 一方、S8の場合は、公的利害、協調的、やや主張的、反対の認識を持つ傾向が見られる。

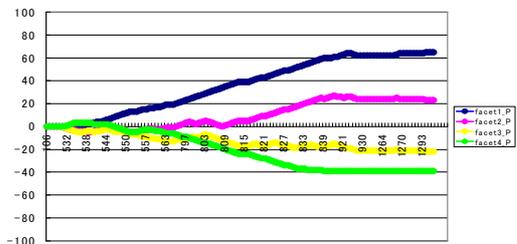


図-4 S8の認識変化

上図から、各々の参加者の微細な認識変化や参加者間の認識の類似性を明確に確認することが可能である。

最後に、討議参加者間の認識の類似性(cognitive similarity)と言語の類似性(semantic similarity)の関係を確認する。図-5は、先行研究<sup>8)</sup>で求められた言語の類似性に基づいて空間上に各参加者を載せて、その上に本研究で得られた認識の分類結果を色に分けて表したものである。ここで、公的利害は認識が一致していることから省略した。

図の真ん中にS4, S13が近く位置して、色づけも類似している。二人の間では言語と認識すべてが類似していると考えられる。一方、S4, S13とS5は近く位置しているが、色付けが異なる。S4, S13とS5の間では、言語は類

似しているが、認識は異なるものと考えられる。

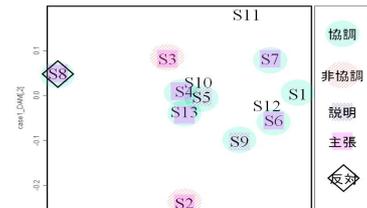


図-5 ファセット分類の仕組み

以上の分析を通じて、(i) 言語の類似性と認識の類似性の双方が高い場合は、参加者間で共通認識が生成されていること、(ii) 認識の類似性が高く、言語の類似性が低い場合は、討議参加者の属する社会的言語的環境は類似していないものの、共通の価値観や利害関心を持ち得ること、(iii) 言語の類似性が高く、認識の類似性が低い場合は、社会的言語的環境は類似しているものの、共通の価値観や利害関心を有していないこと、(iv) 言語の類似性と認識の類似性の双方とも低い場合は、コミュニケーションのための共通点が見え難い状況であることを推察することが可能であると考えられる。

## 5. おわりに

本研究では、統計的言語学習モデルを用いて、公共事業を対象とした公的討議に対してプロトコル分析を実施し、討議参加者の認識構造を把握する方法論について検討した。また、本手法を事例に適用し、認識の不一致を明確にする等その有効性を検討した。今後は、本分析手法を確立することによって、討議参加者の公共事業に関わる認識の共有化や学習プロセスについての知見を得ることが重要である。

## 参考文献

- 1) Bessette, M Joseph. : The Mild Voice of Reason - Deliberative Democracy & American National Government Chicago : University of Chicago Press, 1994.
- 2) Jon Elster. : Deliberative democracy, Cambridge University Press, 1998.
- 3) Amy Gutmann, Dennis Thompson. : Democracy and Disagreement, The Belknap Press of Harvard University Press, Cambridge, 1996.
- 4) 田窪行則 共著：談話と文脈、岩波書店、2004
- 5) 羽鳥 剛史, 川除隆広, 小林潔司, 夏目卓生, 藤崎英司：ファセット理論に基づく公的討論過程のプロトコル分析, 土木計画学論文集, No. 23, pp. 91-102, 2006
- 6) 柳本 豪一, 大松 繁：カーネル法を用いた関連フィードバックによる興味抽出, 電気学会論文誌C, Vol. 126, No. 3, 395-400, 2006
- 7) 工藤 拓, 松本 欲治: Support Vector Machineを用いたChunk同定, 自然言語処理, Vol. 9, No. 5 pp 3-2, 2002
- 8) Hyeong JEONG, Tsuyoshi HATORI, Kiyoshi KOBAYASHI, "Discourse Analysis of Public Debates: A Corpus-based Approach", IEEE, SMC 2007 (掲載決定).