

ワークショップにおける討議内容の数值化と視覚化の試み

A Study on Digitalization and Visualization of the Discussion at Workshop*

丸石浩一**・佐々木邦明***

By Kohichi MARUISHI**・Kuniaki SASAKI***

1. はじめに

まちづくりに市民の多様な意見を吸い上げ、市民参画を深めることを目的として、まちづくりワークショップの開催が近年盛んにおこなわれている。ワークショップを行うこと自体が目的でなく、市民の参画の成果として、多様な意見を反映しながら、一つの方向性を出すことがワークショップをまちづくりに活用するための一つの鍵である。そのため単に意見を出し合うということだけでなく、一定の方向性を出すためには議論の集約が重要である。一般的には全体をいくつかのグループに分け、それぞれのグループ内での自由な議論を、KJ法などの手法を用いて、自由討議をまとめていく方法がとられている。

また、まちづくり等のワークショップへの参加は参加者の意識変化をもたらすことが知られている。特に近年のモビリティマネジメントに代表される意識変化を意図した交通政策においては、意識変化誘発のための有効な手法であるが、その意識変化はもちろん議論の内容に依存する。そのため意識変化の因果関係を知るためには、議論の内容の質的な表現だけでなく、議論の内容のある程度数值化することが望まれる。

これらを考え合わせると、より良いワークショップ開催のためには、議論がどのように進みどのように集約されたのかについての数値的・視覚的なデータがあることが望ましい。そこで、本研究はテキストマイニング手法であるtf*idf値およびSOMを用いて、ワークショップにおける議論を単語レベルの頻度や、文脈的な近接性などの指標として数値的に表現することを目的としている。またその結果を用いて、連続的なワークショップの議論の間の関連性の分析に適用し、議論がどのような影響を与えているのかについての適用例を示す。

*キーワード：交通まちづくり，ワークショップ，テキストマイニング

**正員，修士(工学)，東日本高速道路株式会社 関東支社所沢管理事務所 (〒359-0012 所沢市大字坂之下761-1)

***正員，博士(工学)，山梨大学大学院医学工学総合研究部 (山梨県甲府市武田4-3-11, TEL055-220-8671,

e-mail: sasaki@yamanashi.ac.jp)

2. 議論の数值化法

(1) 単語の頻度による数值化

まずはじめに、ワークショップ内の議論の内容において、そのグループの討議内容をよく示している「特徴的な単語」を探索し、その単語が各班の討議においてどのような位置づけであるかを分析することとした。グループの討議内容の中で特徴的な単語を抽出するために、tf*idf 値という指標を用い、この値を「特徴度」として比較する。tf*idf 値とは Harman が提案した tf (ターム頻度: Term Frequency) と idf (文章頻度の逆数: Inverse Document Frequency) の積であり、比較的簡素な計算ながら文章中における単語を表すことができる重み付けの手法である。

tf 値とは、ある単語 t_i が文章 D_j 中にたくさん現れるなら、その単語はその文章をよく表す単語であるという考え方に基づいて算出される値であり、式 (1) のようにあらわすことができる¹⁾。

$$tf_j^i = \text{freq}(i, j) \quad (1)$$

ただし、 $\text{freq}(i, j)$: 文章 D_j における単語 t_i の出現頻度

この値は、文章の構成によって異なるため、底 2 の対数をとったうえで、単語の種類数によって正規化することで比較可能な数値になる。正規化した際の tf 値を算出する式を、式 (2) に示す。

$$tf_j^i = \frac{\log_2(\text{freq}(i, j) + 1)}{\log_2(k_j)} \quad (2)$$

ただし、 k_j : 文章 D_j に出現する単語の種類数

また、idf 値は複数の文章を考慮した重み付けの数値である。複数の文章が存在するとき、各文章に広く用いられる単語は比較的高い頻度で現れる。しかし、特定の文章だけに集中的に存在する単語 t_i があつた場合、tf 値で表現すると低い値を取ってしまう。この場合、特定の文章を代表する単語が t_i ではうまく表現できない。そこで、式 (3) に示すように特定の単語が用いられている文章数を単語の評価値として用いる。

$$df_j^i = \text{Dfreq}(i) \quad (3)$$

ただし、 $\text{Dfreq}(i)$: 単語 t_i が出現する文章の数

tf 値同様に、文章群の文章数を用いて正規化を行い、2 を底とする対数をとることで、比較可能な数値、idf と

する。idf 値は式(4)のようになる。

$$\text{idf}_j^i = \log_2 \frac{N}{\text{Dfreq}(i)} + 1 \quad (4)$$

ただし、N：文章群中の文章の総数

tf*idf 値は、tf 値と idf 値の性質それぞれを併せ持ち、大きければ大きいほど「単語 t_i が属する文章群の中で、特徴的な単語である」ということを表す数値になる。tf*idf 値を式(5)に示す。

$$\omega_j^i = \text{tf}_j^i \times \text{idf}_j^i \quad (5)$$

(2) ベクトル空間法による文章群の類似度表現

tf*idf 値をもとに文章群をベクトル的に表記することができる。つまり、各単語に対して一連の文章は、各単語数の次元を持つベクトル空間として表現できるといふことである。文章 D_j をベクトル空間で表現したものを式(6)に表す。

$$\vec{D}_j = (\omega_j^1 + \omega_j^2 + \omega_j^3 + \dots + \omega_j^n) \quad (6)$$

単語の数が n 個であった場合、ベクトル式も n 次元で表現することが出来る。このようにベクトル化することによって、2つのベクトルの類似度をベクトルのなす角によって表現することが出来る。ベクトルのなす角は、ベクトルの内積の式を変形することによって計算でき、式(7)となる。これを本研究での文章群の類似度指標とする。このように本研究で用いる類似度は最大値1を取り、最小値0となる指標である。

$$\cos \theta = \frac{\vec{x} \cdot \vec{y}}{|\vec{x}| |\vec{y}|} \quad (7)$$

3. 富士吉田公共交通まちづくりワークショップ

(1) 富士吉田市の現況

本研究で分析対象としたのは平成19年に富士吉田市で行われたワークショップである。対象地域である山梨県富士吉田市は、人口5万3千人の山梨県郡内地域の中心都市である。平成19年現在、富士吉田市の高齢化率21.3%であり、人口は1990年をピークに減少に転じている。主要産業としては富士山や娯楽施設を中心とした観光業、および古くからの織物業があげられる。近年では「吉田のうどん」を観光の目玉の一つとして宣伝しており、市内に60軒近くの店舗が存在している。交通の現況としては、他の地方都市と同様にモータリゼーションが広く浸透し、公共交通の分担率は非常に低い。鉄道は富士急行富士吉田線が市内を縦貫し、富士吉田駅および市民病院を中心として、富士急行バスおよび富士吉田市のコミュニティバスが市域の大部分をカバーしている。

(2) まちづくり検討委員会

今後人口減少期を迎える富士吉田市においては、まちづくりに投下できる資本は限られたものとなり、今後はより効率的な行政運営が求められている。そこで既存の市街地である中心市街地を再び活性化させることにより、新規に市のコアを形成する場合や、郊外に分散した商業施設などに分散して投資を行うよりも効率的な市の運営化を図ることができる。また、高齢化率の上昇は、近い将来に自動車が運転できなくなる交通弱者の増加といった形で現れてくるため、交通弱者のモビリティを確保する公共交通の拡充が欠かせない。これらの問題は相互に関連しており、まちづくりと公共交通を市民と協働で一体的に考える必要性が生じていた。そこで国土交通省関東運輸局・富士吉田市・交通事業者および市民中心とした「富士吉田まちづくりと一体となった公共交通サービスのあり方検討委員会（以下検討委員会）」を平成19年に組織した。この検討委員会では、広く市民の意見を求め、検討委員会での討議の参考とするために四日幅広い市民が参画可能なワーキングを実施することとなった。

(3) 市民参画ワーキング

市民参画ワーキング(以下WS)は、各回それぞれテーマを設定した上で3回実施された。ワークショップ開始に先立ち、市民からWS参加者を募り市内外から53名(市民参加者31名・検討委員会関係者22名)の参加者が集まった。

各回のWS開催概要は以下の通りである。各会では参加者を2~4班に分け、それぞれ、テーマに沿った内容を付箋に記入し、それを各自が模造紙に張り付けながら紹介し、KJ法を用いて類似のものを集めることによって議論の集約を行い、各班の発表を通じて相互理解を深めた。

①第1回

テーマ：富士吉田市の魅力と課題を考えてみよう

日時：2007年12月1日

参加者：31名を4班に分ける

②第2回

テーマ：色々な立場から、富士吉田市がこうなったらいいな、を考えてみよう

日時：2007年12月15日

参加者：27名を4班に分ける

③第3回

テーマ：提案をまとめてみよう

日時：2008年2月16日

以上の3回のWSの議論を班ごとに録音し、それを文章としてテキスト化したのち、各文章の特徴を得るために、

名詞・形容詞を取り出して各文章の構成要素とした。ただし、第1回のワークショップではB班のデータが欠落しており、第2回のワークショップではA班とD班のデータが欠落している。第3回についてはデータが存在しない。また、付箋で貼られた内容と議論との関連性の深さを分析するため、発表用紙にある付箋の内容も同じくテキスト化し、それを分析データとして用いた。

4. tf*idf 値による各班の議論の数値化

(1) 各班の議論の概要

tf*idf 値とベクトル空間法による類似度の算出を行うにあたり、討議時における頻出語をみると、第1回WSでは、いずれの班においてもバスがもっとも話された単語となっていた。以下、公共交通に関する話題（公共交通・鉄道）、街作りに関する話題（まちづくり）、観光に関する話題（富士山・河口湖・観光・観光客など）、地場産業（うどん）、そしていくつかの形容詞となっている。第2回WSでは、バスの言及数はやや中位に位置している。B班では商店街の話題がもっとも話されており、次いで地場産業（地場産業・うどん）や観光（観光）などが話題に上る。C班は観光の話題がもっとも話されており、次いで交通に関する話題（鉄道・バス・自動車など）、商店街に関する話題などがあつた。

これらを tf*idf 値を用いて重みづけし、議論を特徴づける単語として抽出した結果が表-1である。

表-1 tf*idf 値による特徴語の抽出結果（上位3語）

	1a	1c	1d	2b	2c
1	景観	年寄り	自治会	無い	理想
2	補助	一般	遊休店舗	花火	食事
3	サンフーズ	集中	インターチェンジ	大会	便利

分析結果としては、「たくさん用いられた単語」より「その班でのみ語られた単語」が多く抽出される傾向が見られた。この結果を用いて、各班の議論の類似度を計算したところ表-2に示す結果になった。この結果をみると、各班の内容は同じWSにおいても類似度は低く、各班とも独立した内容が話されているように思われる。しかし、先ほど述べたように tf*idf 値は各班特有の単語が強調して抽出されるようになる性質があるため、このような結果が得られたと考えられる。続いて、班の討議内容と班の発表時に貼られていた付箋の間の類似度を計算した。これは、各班の議論の特徴と、最初に提示された付箋の内容が一致するのかを検証するために行った。また、この発表内容は他の班の構成員にも情報として伝わるため、複数回のWSを行う際には、この情報が全員に気わたることになる。発表の付箋内容

を同じように tf*idf 値を計算し、それが高いつまり、付箋を特徴づける単語を示したのが、表-3である。

表-2 空間ベクトル法による討議内容の類似度

	1a	1c	1d	2b	2c
1a	1.000				
1c	0.184	1.000			
1d	0.176	0.241	1.000		
2b	0.171	0.235	0.218	1.000	
2c	0.218	0.264	0.197	0.261	1.000

表-3 tf*idf 値による特徴語の抽出結果（上位3語）

	1a	1c	1d	2b	2c
1	時刻表	都内	回遊	本町通	施設
2	公共交通	施設	御師	イベント	観光
3	多い	買い物	富士五湖	生活	子供

この結果を表-1と比較すると、リストに出てくる単語が大きく変わっている。これはWSで話された内容と付箋として貼られたものが必ずしも一致しないということを示している。これは、付箋は議論の前に書いたものであるが、それを紹介する過程においては他人との意見交換が存在し、議論の内容はキーワードとして貼り付けられたものからどのように発展していくのかに依存する。その意味では、表-3をベースとして、表-1の内容がそこからの発展として議論されたと考えられる。また、その確認のために、WSの討議内容と付箋内容の類似度の比較を表-4に示した。

表-4 討議内容と付箋内容の類似度比較

		討議				
		1a	1c	1d	2b	2c
付箋	1a	0.387	0.122	0.117	0.128	0.131
	1c	0.131	0.307	0.144	0.146	0.183
	1d	0.126	0.178	0.301	0.157	0.172
	2b	0.146	0.135	0.163	0.297	0.187
	2c	0.133	0.196	0.134	0.168	0.344

※着色したセルは互いに対応する(討議→発表)

計算された結果をみると、全ての組み合わせで類似度が高いとはいえないが、同じ回で同じ班の討議と発表を比較すると、関係のない組み合わせと比べ高くなる。しかし、同じ班の異なる回の組み合わせにおいては、関係のない組み合わせとほぼ同様の値であった。

(2) tf*idf 値の安定性

ここで、tf*idf 値の安定性をみるために、双方のWSでの全文章中の頻出上位58単語に限定した tf*idf 値を

求めて類似度を計算した。その結果を表-5に示す。

この結果をみると、これまでの分析結果と比較して、その類似度が高いことがわかる。もちろん、ベクトルの次元が低くなったために類似度は高くなるのであるが、特に第1回WSのC班の議論の内容が第二回WSのB班およびC班の議論の内容と高い類似度を示している。また、同じように第1回WSのD班の議論は第2回WSのB班の議論と類似している。この傾向は表-2の結果と変化していないが、その特徴は明確に表れるようになった。この指標は本来相対的なものであり、比較において明確に特徴が表れることは望ましいため、単語数を絞ることは、議論の傾向を把握するためには望ましいと考えられる。

表-5 第1回WSと第2回WSの議論の類似度

tf*idf 行列		WS1/議論			WS2/議論	
		A	C	D	B	C
WS1/ 議論	A	1.000	0.657	0.663	0.625	0.640
	C		1.000	0.677	0.731	0.762
	D			1.000	0.753	0.704
WS2/ 議論	B				1.000	0.797
	C					1.000

(3) SOMを用いた単語のクラスタリング

本節では、ある単語が登場する文章においての利用される文脈が重要との認識より、単語の同時出現度に基づいて文脈を解析する手法を適用する。その考え方は、ある単語が他の単語とどの程度同時に使われているかを示すベクトルを定義して、同時に使われることの多い単語は近い位置に配置し、ほとんど同時に使われない単語は遠くに配置する、文脈の視覚化を適用する。けっかとして、単語の文脈に応じたクラスタリングを行うことになる。文脈ベクトルでは1つの単語を1つのベクトルで表すため、今回の計算のように多数単語間のクラスタリングでは、多次元(50次元~100次元)の計算を行う必要がある。そこで、本研究ではその単語の配置方法にSOM(自己組織化マップ, Self Organizing Maps)を利用した。紙面の関係上そのアルゴリズムの詳細は参考文献³⁾を参照してほしい。

実際に討議内容に対してSOMを適用した。単語は数多く存在するが、その例として「東京」というキーワードに対して、2つの異なる班を比較した。すると、班によって周囲に配置される単語が大きく異なっており、図-1左図ではアクセス手段に関する単語が多く、東京とのアクセス手段に重みを置いて話をしていたと推測され

る。図-1右図では富士吉田の観光資源に関する単語が多く、東京からの観光に重点を置いていたと考えられる。このようにSOMによって班によって同じ語でも使われ方や関連性が異なる点を図化して表すことが出来た。

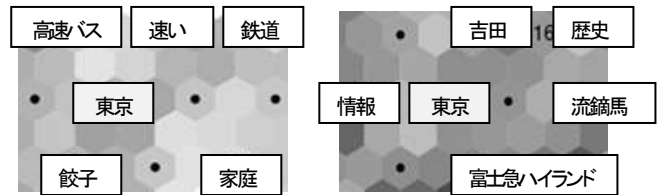


図-1 SOMによる単語のクラスタリング例

5. おわりに

本研究では、WSでの議論の内容を数値化することを目的として、富士吉田市で行われた公共交通まちづくりWSを対象として、tf*idf値を用いてその数値化を行った。また、SOMを用いて単語の類似度を視覚的に表現した。その結果、各班の議論の特徴を発表内容とは異なる視点から見つけることができた。そのうえで、類似度を定義することにより、議論の内容が必ずしも発表の普選に現れる内容とは一致しないことを明らかにした。また、異なる班の議論の類似度を定義でき、発表による意見の共有が他の班にどのような影響を与えているかを数値で示すことができた。また、SOMを用いた結果からは、単語の文脈的な位置づけの違いをみることができた。

本研究の最終的な目的は議論の内容と意識変化の分析である。今回得られた結果の意識指標に対する適用が今後の課題である。

謝辞

本研究は関東運輸局企画観光部交通企画課、富士吉田市まちづくりと一体となった公共交通サービスのあり方検討委員会および市民参画ワーキングに参加していただいた皆様に研究の機会と貴重なデータを提供していただいた。ここに記して感謝の意を表す。

参考文献

- 1) D Harman: An experimental study of factors important in document ranking. In Proceedings of ACM conference on Research and development in information retrieval, pp. 186-193, Pisa, Italy, September 1986.
- 2) T. Landauer and S. Dumais: A solution to plato's problem: The latent semantic analysis theory of the acquisition, induction, and representation of knowledge. Psychological Review, Vol. 104, No. 2, pp. 211-240, 1997.
- 3) 田中雅博, 古河靖之, 谷野哲三: 自己組織化マップを利用したクラスタリング, 電子情報通信学会論文集, Vol. J79-DII, No. 2, pp. 301-304, 1996.