

対話型意見収集システムの利用満足度調査結果の分析

大塚裕子¹ 伊藤裕美¹ 川野佐江子²

¹計量計画研究所 (IBS) ²立教大学大学院

1. 対話型意見収集システムへのニーズ

本稿では、開発した対話型意見収集システム（以下、対話型システムと呼ぶ）（丸元他 2008, 大塚他 2008）の今後の改良に向けた検証として実施したモニターによるシステム使用実験および満足度調査と、その結果の分析について報告する。

近年の市民参画型の政策立案および事業計画プロセスでは、情報伝達、意見把握などのコミュニケーション手法において、効果的かつ効率的なツールの開発が求められている。とくに意見把握では、参加主体である市民を中心とした利害関係者がより意見を述べやすい、また、行政担当者にとって意見を計画に反映しやすい手法を開発、提案する必要がある。しかし、そのためには、「賛成/反対」「A案がよい」といった立場表明の意見に留まらず、その背後にある計画への不安・懸念、不満など具体的な関心（インタレスト）や意見の理由を引き出す必要がある（矢嶋 2006）。

現状では、上述のようなインタレストを引き出す手法は、グループインタビューやワークショップなど対話の場の進行管理役であるファシリテーターによって行われる。しかし、計画の影響が広域になる場合、関与する市民も分散し多数であるため、すべての関係者と直接対話することが困難であること、対話の場へのアクセスが困難な市民がいることなどを踏まえると、一度に得られる意見の数と多様性には限界があるといえる。

一方、アンケート調査のように、コミュニケーションの場に関与する利害関係者の数が多数である場合には、具体的な不安や懸念、問題などを引き出すことは難しい。すなわち、現状の課題は、できるだけ多数の意見を得ることが可能であり、かつ、関係者間の対話が双方向であり、やりとりの頻度が多い手法の開発である。

2. 対話型意見収集システムの開発

上述の問題点に基づき、本研究では以下の方針で対話型システムを開発した（丸元他 2008）。

インターネットを介したシステムを利用し、参加の現場にアクセス困難なユーザー層を含め、多数多様なユーザーに意見表明の機会を作ること

上記 によって得られた意見に対し、機械学習手

法により自動的かつ即時的に意見分類を行うこと
上記 の分類に基づき、それに応じた対話設計を行い、ユーザーのインタレストを引き出すこと

現在、市民参画現場での意見収集ツールとして本研究を進めているが、これらの技術は企業のマーケティング戦略や組織運営などにも重要なツールと成り得る。本研究で開発している対話型システムは基本的に、ユーザーから入力された意見に対して、1) 意見の理由を尋ねる、2) 意見の詳細を尋ねる、という2種の発話を繰り返し、次のような対話を行う。県名など固有名詞の伏せ字および括弧による注記は筆者に拠る。

<対話事例 1> (U: ユーザー S: システム)

- 1 U: 空港周辺の環境へ配慮して欲しい
- 2 S: もう少し詳しく、Uさんの考えの理由についても、聞かせて下さい。(理由を引き出すための問い返し)
- 3 U: 空港の近くには海があるので、埋め立てなどはしないで欲しい
- 4 S: 一つずつ確認させてください。今、伺ったのは、おおまかに埋立規模に関するご意見ということでいいですか? (インタレストの推定)
- 5 U: はい
- 6 S: Uさんは、具体的には、埋立規模によって、どこに、どんな影響が出ることが気になりますか? (詳細情報を引き出すための問い返し)
- 7 U: 島周辺の海の汚染
- 8 S: これまでのご意見承りました。では、その他に、これまでのやり取りで十分述べていないご意見はありますか? なんでも気づいたことを教えてください。
- 9 U: ××は離島県なので、空港の機能を充実させて離島の方や観光客の方に気持ち良く利用してもらいたい
- 10 S: 一つずつ確認させてください。今、伺ったのは、おおまかに観光客に関するご意見ということでいいですか? (インタレストの推定)
- 11 U: 違います
- 12 S: 失礼しました。すみませんが、もう一度、少し簡単な言葉で教えてください。
- 13 U: 離発着の安全が確保できれば安心して利用できると思います

上記の対話例に見られるように、本研究で対象とした市民参画プロセスの現場では、空港計画に関する多様なインタレストを獲得することが目的となる。したがって、対話番号のように相手の関心事を推定

(推定インタレスト)し、対話番号4のように、それが妥当であるかどうかをユーザーに確定(確定インタレスト)してもらうことにより、インタレストを獲得する。さらに対話番号6のように、その詳細情報を聞き出す。インタレストは、空港の安全、サービス、環境問題などの意見の内容項目25件と、不安・懸念、不満など意図感情項目11件の組み合わせから計画関係者との協議の上作成した124件をリスト化している。リストにないインタレストが得られた場合には追加可能なシステムとなっている(リストの詳細については(丸元他2008)を参照のこと)。また、現状のシステムでは意図感情の判定をSVMによる機械学習で行い、内容の判定を正規表現によるパターンマッチングで行っている。

3. モニターによるシステムの使用実験と満足度調査

2007年8月29日から10月28日に実施されたある自治体の空港計画の意見公募(本調査)の、試行的かつ補助的調査として、開発した対話型システムによる意見収集を行った。意見公募期間の後半10月18日から意見公募終了時後3月末まで自治体のホームページからのアクセスを継続した。しかし、この期間に得られた意見は36件と少数であったため、モニターによる使用実験を追試し、満足度調査を併せて行った。

3.1 使用実験の方法と結果

ユーザーの居住地属性だけは、対象の自治体に限定し、年齢、性別など他の属性などは偏らないようサンプリングされた119名の被験者を対象としてweb上でのシステム使用実験と調査を行った。

実験と調査は、1)対話型システムの趣旨文教示、2)空港整備計画の調査内容(参考資料のPDF)表示、3)調査内容の理解度テスト、4)対話型システムの実施、5)満足度調査、という手順で行った。実験および調査の実施に先立ち、対話型システム使用の際の導入趣旨として、本調査実験が「対話型システムの使いやすさを調べる」という目的であることを明示した。これは、自治体の本調査である意見公募の一環と認識される誤解を防ぐためである。また、必要な最低限の知識を得るために、参考資料の報告書を読んでもらい、その内容を理解するテストのステップを設けている。理解度テストのwebページから、対話型システムにリンクを貼り、被験者が一連の調査実験として実施できるようにした。

実験結果の概要は表1のとおりである。この結果は、自治体の本調査の、システムによる補助的意見収集の結果(大塚他2008)と同様の傾向を示す。平

均ターン数、平均対話時間については、両結果ともに類似した値が得られた。しかし、最多ターン数および最長対話時間については(大塚他2008)のほうが数値は高かった。表1のとおり、モニター調査では、最多ターンが9、最長対話時間が34分36秒となっている。モニター調査は被験者実験であるのに対し、本調査の補助的意見収集時のユーザーは、自治体のホームページを辿ってシステムを使用しているため、空港計画に対して関心が高いことの表れとみなせる。平均値が類似しているのに対し、最多・最長値が異なっている要因はこのように考えられる。

表1 モニター調査の実験結果

評価項目	評価値
アクセス数(ログ数)	119
総対話時間	19時間40分59秒
平均対話時間	9分55秒
最長対話時間	34分36秒
最短対話時間	54秒
総ターン数	665
平均ターン数	5.59
最多ターン数	9
最小ターン数	1
推定インタレスト(システムが獲得したインタレスト)	93
確定インタレスト(ユーザーによって確認されたインタレスト)	56

3.2 満足度調査の方法と結果

対話型システムの満足度調査については、1)問い返しの自然さ、2)システムの印象、3)システムの特徴、4)公共事業計画への導入、5)拡張整備についての知識の5カテゴリについて、表2のような印象評価文を用意し、それに対して、「非常にそう思う」「少しそう思う」「どちらともいえない」「あまりそう思わない」「まったくそう思わない」の5段階で評価する調査を行った。表2の設問カテゴリごとの集計結果を図1～3に示す。本稿では紙幅の関係上、1)問い返しの自然さ、2)システムの印象、3)システムの特徴の結果についてのみ取り上げる。印象評価文に対して「非常にそう思う」「少しそう思う」など肯定的な捉え方をしている回答をPos(青網がけ部:棒グラフ左)、「どちらともいえない」のように答えを保留しているものをNT(赤縦線部:棒グラフ中)、「あまりそう思わない」「まったくそう思わない」のように否定的な捉え方をNeg(薄黄色部:棒グラフ右)として示した。設問のうち、Q1-2,3,5,6、Q2-8,9、Q4-7,8のように、評価文が否定的であるものは、回答を反転させて集計している。

問い返しの自然さについて、システムの発話の長さや発話の理解のしやすさ等には比較的高評価であるが、単語・表現のわかりやすさ、応答の妥当さ、やりとりへの違和感については低評価である。単語・表現については、対話実例2のように「時代の

変化への対応」など表現のわかりにくさや「ハブ空港」など専門用語に要因があると考えられる。これらの応答表現については土木計画系の専門家と協議し、誤解なく伝わる表現を作成したが、システムの改良に向けてはユーザーにとってのわかりやすさを追求すべきだろう。

システムの印象については、思ったとおりに発言できた、考えが整理できた、のように、評価の主体がユーザー自身にあるものは比較的高い評価となったが、傾聴感や伝わっている感じ、尊重されている感じのようにシステムを評価する指標については低評価となっている。この結果は、ユーザーの意見表明の支援としては妥当だったものの、システム自体の対話能力としては十分でなかったことを示唆する。

システムの特徴については、本システムの開発課題である「できるだけ多数の意見を得ることが可能であり、かつ、関係者間の対話が双方向であり、や

表2 満足度調査のカテゴリと評価文

設問カテゴリ	設問内容
問い返しの自然さ	1 - . システムとの対話はスムーズにできた
	1 - . システムが話す単語や表現がわかりにくかった
	1 - . システムが話す文が長すぎた
	1 - . システムが話す内容は理解できた
	1 - . システムが自分の発言とは関係ないことを聞き返した
	1 - . システムに馬鹿にされているように感じた
	1 - . システムとのやり取りに関して違和感があった
システムの印象	2 - . 話を真剣に聞いてもらっていると感じた
	2 - . 自分の思ったとおりに発言できた
	2 - . 自分の言いたいことが伝わっていると感じた
	2 - . 自分の考えを整理することが出来た
	2 - . 自分の考えを深めることが出来た
	2 - . 自分の意見が尊重されているように感じた
	2 - . 自分の意見が公正に扱われていると感じた
	2 - . 自分の意見が誘導されていると感じた
システムの特徴	3 - . 人となくシステムが答えること自体に違和感を感じた
	3 - . 時間と場所を選ばず意見を言える
	3 - . 発言機会が何度もある
	3 - . 発言を促してもらえる
	3 - . 考えが変わっても後から訂正できる
	3 - . 後から気付いたことを追加できる
	3 - . 建前の意見でなくて、本音がいえる
	3 - . 通常のアンケートと比べて親しみやすい
公共事業計画への導入	4 - . 通常のアンケートと比べて考えながら意見が言える
	4 - . 市民の意見をより詳しく聞こうとしている
	4 - . 計画に市民の考え方をなるべく取り入れようとしている
	4 - . よく考えた上での意見が取り入れられるようになる
	4 - . 市民が計画にもっと関心を持つようになる
	4 - . 市民の意向を無視することが少なくなる
	4 - . 計画の進め方がもっと公正なものになる
	4 - . 計画主体の期待する方向に意見が誘導されることになる
拡張整備についての知識	5 - . 人が対応しない機械に任せるとは失礼だ
	5 - . 公共事業にはこのようなアンケートを是非とり入れるべきだ
	5 - . 拡張整備について、もともと興味があった
	5 - . 拡張整備については、以前よりよく知っていた
	5 - . このアンケートで拡張整備について興味がわいた
	5 - . このアンケートで拡張整備についてよく知ることができた
	5 - . 自分はもともと好奇心が旺盛なほうだ
	5 - . 自分はもともと社会問題について関心が高い

りとり頻度が多い手法の開発」を実現できたことを示す評価が得られたと考えられる。しかし、対話でありながら、親しみやすさへの評価が得られなかったことは、上述のように対話能力の低さに問題があるといえる。

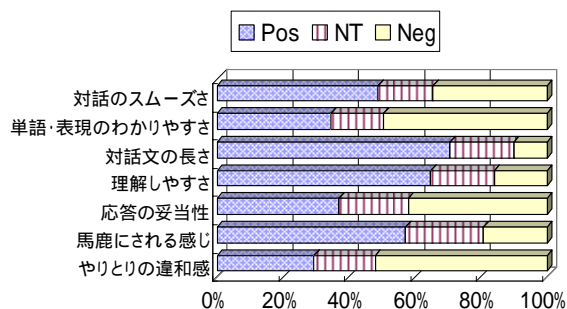


図1 問い返しの自然さ

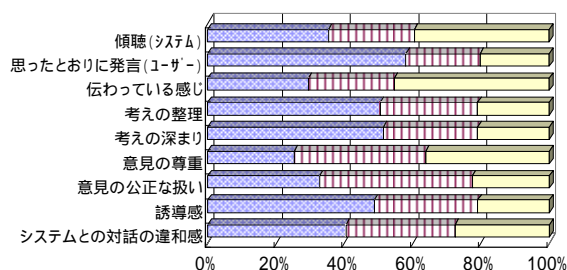


図2 システムの印象

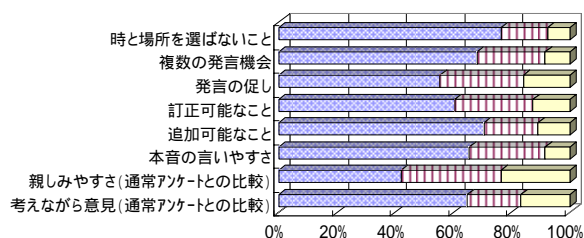


図3 システムの特徴

4. 満足度の違いに基づく対話の分析

実用化を目指した意見収集の対話型システムでは、1) 得たい情報(本研究ではインタレスト)が得られたかという調査主体にとっての評価と、2) 対話のしやすさというシステムユーザーにとっての評価の2点、コンテンツおよびプロセスに関する評価によって、システムの改良が検討される必要がある(大塚他 2008)。この2点を踏まえ、ここでは「システムとの対話のスムーズさ」(Q1-)に着目する。まず、満足度調査のユーザーごとの回答相関をとったところ、問い返しの自然さについては0.65以上の比較的高い相関を示す設問は、Q1- と (0.69)、Q1- と Q2- (0.67)、Q1- と Q3- (0.65)であっ

表3 満足度の違いに着目した対話履歴の分析から得られた特徴要因

			合計							平均				
			ユーザー数	ターン数	推定インタレスト獲得数	確認回数(システム)	確認への否定回数(ユーザー)	確定インタレスト獲得数	明確な否定がないためインタレストが確定された回数	ターン数	推定インタレスト獲得率 (獲得数/ユーザー数)	確定インタレスト獲得率(獲得数/ユーザー数)	確定インタレスト獲得率(推定数比)	最初のインタレスト推定までのターン数
Q1-	Q1-	Pos	32	179	28	23	0	23	4	5.59	0.88	0.72	0.82	2.22
		Neg	39	222	30	27	8	19	4	5.69	0.77	0.49	0.63	2.40
	Q2-	Pos	29	159	19	18	1	17	1	5.48	0.66	0.59	0.89	2.13
		Neg	37	197	24	23	7	16	2	5.32	0.65	0.43	0.67	2.13
	Q3-	Pos	36	199	27	24	1	23	3	5.53	0.75	0.64	0.85	2.22
		Neg	23	120	15	13	6	7	2	5.22	0.65	0.30	0.47	2.20

た。そこで、これらの回答者について、各設問評価に対し、いずれもポジティブな評価(P)、いずれもネガティブな評価(N)を行った際のシステムとの対話履歴を分析した。その特徴についてまとめたのが表3である。この結果から、P/N の評価を分ける要因としては、確認への否定回数と確定インタレストの獲得率の影響が高いことが明らかになった。分析前われわれはシステムがインタレストを推定できるか否かが満足度に影響すると考えていた。インタレストを推定できない場合、システムは対話事例 1 の 6 のような、理由を尋ねる問い返しを際限なく繰り返すからである。しかし、表3に示すように、推定インタレストの獲得率の違いは、満足度の P/N の評価との関係性を見出しにくい。対話への満足度は、むしろ、推定されたインタレストが、ユーザーによって確定される過程、すなわち、システムが推定したインタレストを肯定するか否定するかという過程に起因することが結果に示されている。

以下に Q1- と、Q1- の評価がいずれもネガティブな対話、すなわち、対話がスムーズでなく、やりとりで違和感があったと評価された典型例を示す。この例では、推定インタレストに対するユーザーからの明確な否定が2回あり、確定インタレストは得られなかった。

<対話事例2>

- 1 U: 24 時間空港化の早期実施
- 2 S: 一つずつ確認させてください。今、伺ったのは、おおまかに時代の変化への対応に関するご意見ということでもいいですか？(インタレストの推定)
- 3 U: 違います
- 4 S: 失礼しました。すみませんが、もう一度、少し簡単な言葉で教えて下さい。
- 5 U: 早めに空港を 24 時間オープンさせる
- 6 S: これまでのご意見承りました。では、その他に、これまでのやり取りで十分述べていないご意見はありますか？なんでも気づいたことを教えてください。
- 7 U: 空港を拠点とした交通網の早期整備
- 8 S: 一つずつ確認させてください。今、伺ったのは、おおまかにハブ空港に関するご意見ということでもいいですか？
- 9 U: 違います

10 S: 失礼しました。すみませんが、もう一度、少し簡単な言葉で教えて下さい。

11 U: 空港への路線を早めに整備し直す

一方で、推定インタレストへの肯定はしていないにもかかわらず、満足度がポジティブな評価もある。

<対話事例3>

- 1 U: 沖縄県が観光依存の経済から脱却するためにも、貨物を中心とした那覇空港のハブ空港化は必要だと思います。
- 2 S: 一つずつ確認させてください。今、伺ったのは、おおまかに観光業に関するご意見ということでもいいですか？
- 3 U: 観光業もありますが、その他の業種にも影響のあることですね。
- 4 S: Uさんは、具体的には、××の観光業に、どんな問題があると考えていますか？
- 5 U: リピーターの確保やアジア以外の地域からの観光客の誘致、長期滞在型観光への移行の為の施策が必要と思います。(略)

現段階の分析では、1) 対話事例3のようにシステムの推定するインタレストの指す範囲が広いとユーザーの意見との違いが明確にならず許容される、2) 対話事例2のように具体性が高いとユーザーの意見との違いが焦点化されて許容されない、という傾向が見られている。ユーザーの満足度を高め、かつ対話を続けてもらい多様なインタレストを得るためには、インタレストの推定精度を高めると同時に、対話履歴の分析に基づく問い返し設計の工夫が必要である。また、対話事例3では推定されたインタレストのずれが、ユーザーの満足度の低さの要因にならなかったが、このような明確でない否定は、誤解が生じる要因になりやすいため、今後は、対話事例の2,3に示すような明確な/明確でない否定の前後の対話過程の分析も行い、得られた特徴や要因を対話設計の改良に活かしていく予定である。

参考文献:

- 大塚裕子, 乾孝司, 鈴木泰山, 伊藤裕美, 丸元聡子, 奥村学(2008) 対話型意見収集システムの評価方法の検討, 言語処理学会第14回年次大会発表論文集.
- 丸元聡子, 鈴木泰山, 大塚裕子, 伊藤裕美, 乾孝司, 奥村学(2008) 空港計画における対話型意見収集システムの実装と課題, 言語処理学会第14回年次大会発表論文集.
- 矢嶋宏光(2006) ワークショップ開催上の留意点, 全日本建設技術協会編『ワークショップ事例集--協働によるこれからの地域づくり』