

言い換えによる自由記述アンケート回答の要求意図判定基準の作成および検証

乾裕子<sup>†‡</sup> 内山将夫<sup>†</sup> 井佐原均<sup>†‡</sup>

<sup>†</sup>通信総合研究所 <sup>‡</sup>神戸大学大学院

1. はじめに

「要求を知る」すなわち、何が求められているのかを認識するということが社会のあらゆる局面で必要とされている。現状の問題点を探り、改善するという問題解決には「要求を知る」ことが重要だからである。近年、パソコンなどの電子媒体やインターネットの普及に伴い多くの要求が電子化テキストに蓄積されるようになった。これらの電子化テキストから「要求を知る」こと、すなわち要求を取り出すことは、要素技術として研究開発が試みられてきている。たとえば、情報抽出やテキストマイニングなどの技術を用いた、コールセンターへの問い合わせ分析システムの開発 [14] や F A Q 生成支援システムの開発 [11, 12, 15]、また、要求を取り出す仕組みとして提案された、ソフトウェアシステム開発における要求獲得会議での要求獲得法 [4] などである。しかし、「要求を知る」ことは、様々な「要求者」の意図を知ることであり、単にテキストに文字通り表現されている要求表現を抽出することではない。この目的に対しては、現状の工学的な技術としての意図認識および意図理解は十分でない。

以上を踏まえ、本研究では要求が蓄積されたテキストとしての自由回答を対象とし、回答テキスト中の要求意図を抽出し分類する研究を進めている [6, 7]。とくに、意図抽出のための基礎研究として、抽出の際に必要な要求意図の特定方法に関する基準や規則に着目している [8, 9]。本稿では、要求意図判定基準を提案し、この基準の妥当性を検証するために客観性・再現性・有効性の三つの観点から行った実験の方法と結果について述べる。

2. 言い換えを用いた要求意図判定基準の作成

調査者はアンケートなどの手段を通して回答者の要求を知ろうとし、回答者は回答することで調査者に要求を伝えようとする。したがって、回答テキストに記された表現から回答者の要求を取り出す方法を確立することは重要である。

まず「要求」の表現について演繹的に考察する。回答文を読むと、日本語母語話者であれば殆どの人が「要求」であると判断できる「～してほしい」「～てください」「～くれ」「～もらいたい」のような「直接依頼表現」(直接要求表現)

[10] がある。これらの直接要求表現は「要求を知る」ための手がかりになっている。この認知過程は図1における「言語直感的に要求と判断できるか否か」の判定に相当する。これを、図2に示す要求意図判定基準の第一段階とみなし、回答が「直接要求表現を含むか否か」という判定基準に相当すると位置付ける。

次に、要求や依頼を示す直接的な表現がない場合、すなわち図1における非要求の回答の中にも要求であると判断できる場合について考える。たとえば「交通量の激しい道路の歩道にはガードレールをつけるべきだ」「環境に配慮した道路づくりが重要である」といった回答は、調査者・回答者・状況を踏まえた語用論的解釈から「歩道にガードレールをつけてほしい」「環境に配慮してほしい」といった要求とみなすことができる。しかし、図1の第二の判定に示すとおり、この解釈方法を可能にするのは暗黙の「何らかの」基準である。このような基準を用いて得られる図1の要求②は判定者に依存して判定結果が変わる可能性があるため、要求を含む回答を取り出す際に再現性のある結果が得られない。そこで、われわれは、この暗黙の基準を要求意図判定基準として顕現化することを試みた。この顕現化が、図2に示した第二の「回答を『てほしい』を含む文に言い換えられるか否か」という判定である。回答文を典型的な要求表現である「てほしい」を含む文に言い換えることができれば、それは要求文であるといえる。この判定基準によって得られるのが図2における要求意図表現を含む要求であり、この要求を含む回答は明確な基準に基づいて取り出されるため再現性が保証される。

以上のとおり、本稿では回答から要求を適切に取り出すための手法として、典型的な要求表現への言い換えによる要求意図判定基準を提案する。

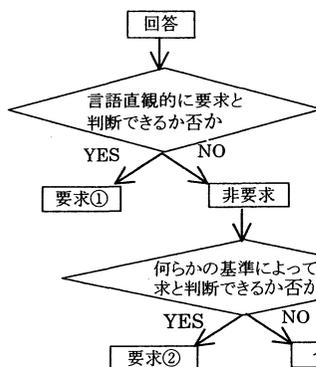


図1 表現としての「要求」を判断する階層

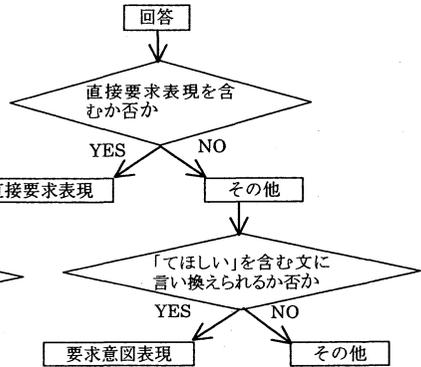


図2 要求意図判定基準

### 3. 同一判定者による検証

#### 3.1. 分析的検証と考察

要求意図判定基準をもとに、1) 回答文に「てほしい」「もらいたい」等の直接要求・依頼表現を含むか否か、2) 回答を「てほしい」で終わる文に言い換え可能か、を目標により判断、分類する。1) の判断には森田ら [13] の分類において要求・依頼に列挙された表現を対象とした。分析対象は、将来的な道路計画に市民の声を活かす目的で行われたアンケートの自由回答で、回答人数 35,674 人、回答数 (意見数) 113,316 件の大規模調査である [5]。分析データは、複数文から成っている回答についても一文ごとに分け、無作為に取り出した 3000 文を対象にした。それに対して 2 節で提案した要求意図判定基準を適用し、表 1 の結果を得た。

表1 要求意図判定基準の適用結果<sup>1</sup>

	直接要求表現	言い換え	3000 文中
①	含む	可	547
②	含む	不可	3
③	含まない	可	1190
④	含まない	不可	1252

表 1 の①は「～てほしい」「～てください」などの直接要求表現を含む文である。これらは、すべて「～てほしい」への言い換えが可能であり、3000 文中の約 2 割を占める。③の表現が 2 節の図 2 における要求意図表現に相当する。

要求意図表現には動詞や名詞など、従来、モダリティの表現とは認められていない品詞の語や、従来の助動詞および助動詞相当語句としてのモダリティ表現の中で、構文論的な意味分類においては「要求」とみなされていなかった様々な表現も含まれる [9]。従来の言語学的知見から得られたパタン分類の手法では、表 1 の①からのみ「要求」を取り出そうとしていた。しかし、提案手法を用いることにより、表 1 の③の要求意図表現を手がかりに「要求」を取り出せることがわかった。

#### 3.2. 機械学習による客観的検証

本節では、要求意図判定基準が機械学習を利用することにより自動化できることを示す。また、これにより 3.1 節の分析的検証が十分に客観的であることを示す。

##### 3.2.1. 機械学習による意図判定の概要

2 節で述べた意図判定法は、機械学習という観点からは、与えられた自由回答を、「てほしい」に書換え可能なクラス(「可」と呼ぶ)と不可能なクラス(「不可」と呼ぶ)に分類する、2 値分類問題として定式化できる。

##### 3.2.2. タグ付きデータ

機械学習のためのデータとしては、3001 の自由回答からなるタグ付きデータを 3.1 節の分析対象とは別に用意した。

このタグ付きデータにおいて、「可」の回答数は 1944 であり、「不可」の回答数は 1057 である。したがって、もし、全ての回答について「可」を選んだとすると、 $1944/(1944+1057) = 0.648$  の正解率 (= 正しく分類された回答の数/全回答の数) が得られる。この方法は、最頻のクラスを選ぶ方法であるので最頻法と呼ぶ。最頻法は、全ての機械学習法のベースラインとなる手法であるが、最頻法では、高精度な分類をすることはできない。

##### 3.2.3. 利用した機械学習手法

高精度な分類のために、本節で用い、かつ、比較する機械学習法は最大エントロピー法(ME) [1] と Support Vector Machine(SVM) [2] である。これら手法の詳細は本節では省略するが、概要は以下の通りである。まず、回答  $X$  は数値ベクトル  $x=[x_1, x_2, \dots, x_n]$  により表現する。ここで、
$$x_i = \begin{cases} 1 & \text{素性 } i \text{ が回答 } x \text{ に含まれるとき} \\ 0 & \text{素性 } i \text{ が回答 } x \text{ に含まれないとき} \end{cases}$$
 である。

ME や SVM では、あらかじめ「可/不可」に分類された訓練データ(およびパラメタ調整用のデータ)を利用して、回答における素性値の 0/1 のパタンから「可/不可」への写像を学習し、それを利用してテストデータにおける回答を分類する。なお、本節では、ME の学習には maxent<sup>2</sup> を用い、SVM の学習には TinySVM<sup>3</sup> を用いた。SVM の学習においては、1 次の多項式カーネルを用いた<sup>4</sup>。

最後に、学習に利用した素性の集合について述べる。本節では 3 通りの素性集合を比較した。それらは、 $n=1, n=2, n=3$  と示すことにするが、 $n=1$  は任意の形態素(表記)の集合、 $n=2$  は  $n=1$  に任意の形態素の 2 連鎖を追加した集合、 $n=3$  は  $n=2$  に任意の形態素の 3 連鎖を追加した集合である。また、ある回答  $x$  に含まれる素性とは、 $x$  が形態素列により  $w_1, w_2, \dots, w_m$  で表現されているとすると  $w_1 = \langle b \rangle, w_m = \langle e \rangle$  という文頭と文末を示す特殊形態素として、 $n=1$  については  $\{w_i \mid 2 \leq i \leq m-1\}$  に含まれる形態素であり、 $n=2$  については  $n=1$  の場合に加えて  $\{w_i w_{i+1}, w_{i+2} \mid 1 \leq i \leq m-1\}$  を追加した集合に含まれる形態素(連鎖)であり、 $n=3$  については  $n=2$  の場合に加えて  $\{w_i w_{i+1} w_{i+2} \mid 1 \leq i \leq m-2\}$  を追加した集合に含まれる形態素(連鎖)である。なお、回答は茶釜<sup>5</sup>を利用して形態素に分割した。

##### 3.2.4. 実験手順と実験結果

実験の手順としては、10 分割交差検定(10-fold cross validation)を利用した。ME と SVM の全体の正解率  $P_{ME}, P_{SVM}$  は各分割での正解率の平均により定義する。また、ME と SVM の精度を統計的に比較するためには、ウェルチ(Welch)

<sup>2</sup> <http://www.crl.go.jp/jt/al32/members/mutiyama/software.html>

<sup>3</sup> <http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/>

<sup>4</sup> 2 次の多項式カーネルも用いたが、その精度は、1 次の場合よりも統計的に有意差がないかもしくは低かったため、ここでは割愛する。

<sup>5</sup> <http://chasen.aist-nara.ac.jp/>

<sup>1</sup> ①～④に含まれない 8 文については、前後文脈を見ないと判断に迷う例である。②は直接要求表現を含むが、引用文中に現れているため言い換えは不可であった。これは例外事例といえる。

検定を利用した。なお、MEと最頻法、およびSVMと最頻法との比較においては、最頻法の正解率は一定値0.648(3.2.2節参照)であるとして、MEとSVMとがその一定値よりも大きな正解率をとるかどうかをt-検定で比較した。これらの検定はいずれも片側検定とした。このようにして求めた正解率を表2に示す。表において、「n=1」「n=2」「n=3」の行はそれぞれの素性集合を用いた正解率であり、「ME」「SVM」「最頻法」の列はそれぞれの学習法による正解率である。

表2 正解率の比較

	ME	SVM	最頻法
n=1	0.892	0.887	0.648
n=2	0.912	0.909	0.648
n=3	0.913	0.915	0.648

まず、表から、どのnについても、MEとSVMの両手法とも最頻法よりも正解率が高いと言える。これは、統計的にも極めて有意である(有意水準 $1 \times 10^{-12} \sim 1 \times 10^{-14}$ )。次に、MEとSVMとを比較した場合には、いずれのnにおいても、MEとSVMとで正解率に有意差はない。したがって、学習法としてはどちらを用いても良い。また、n=1とn=2、n=1とn=3、n=2とn=3について、それぞれの最高正解率同士の比較をすると、n=2とn=3については有意差はないが、n=1とn=2、n=1とn=3には有意差(有意水準 $1 \times 10^{-3}$ )がある。そのため、素性としてはn=2かn=3を用いた方が良い。以上は相対的な正解率の比較である。絶対的な正解率について言えば、n=2,3においては、MEもSVMも91%前後の正解率である。この正解率は最頻法よりも十分に高いので、機械学習が有効であったことを示している。また、これは自動的に精度良く書き換え可能性を判定できることも示している。つまり、このことは、われわれの主張点である「判定基準の客観性」を十分に保証するものであると言える。

#### 4. 異なる判定者による検証

3節では分析的検証、およびその検証に基づいて作成したコーパスを対象として機械学習により客観的検証を行うことで、同一判定者による検証を行った。ここでは異なる判定者による、再現性および有効性の検証を行う。

##### 4.1. 再現性の検証 一複数の作業者による判定一

実験の被験者すなわち言い換え作業者は日本語を母語話者とする大学生3人で、工学部に在籍する20代男性である。3節で説明したアンケート回答の中から無作為に抽出した述べ24000文を対象に作業を行った。一回答が複数文から成る場合は、一文ごとに分けて対象文とした。なお、延べ24000文の内訳は、三人の作業者ABCが各自8000文を担当したが、AB、BC、ACとそれぞれ二人ずつ共通する4000文に対して作業したため、文の異なりは12000である。なお、3.1節の表1で示したように、直接要求表現は「てほしい」に言い換え可能であるため、本節では図2の第二段階のみの判定作業すなわち「てほしい」への言い換え判定だけを行う。「てほしい」を含む文への言い換

えに際しては作業仕様を作成し参照してもらった。

表3 判定による言い換え可否の事例数

		B			$\kappa$ 値
A		可	不可	計	
可		2372	970	3342	
不可		36	622	658	
計		2408	1592	4000	
		C			$\kappa$ 値
A		可	不可	計	
可		3123	264	3387	
不可		171	442	613	
計		3294	706	4000	
		C			$\kappa$ 値
B		可	不可	計	
可		2119	50	2169	
不可		934	897	1831	
計		3053	947	4000	

各作業者が言い換え作業を行い、表3の結果を得た。表中の「可」「不可」は作業者が言い換え可、言い換え不可とした事例を指す。表3の値から作業の一致率を計算するにあたり、偶然の一致を排除した一致率を統計的に計算する $\kappa$ 値<sup>6</sup>を比較する。 $\kappa$ 値は近年、言語処理の分野でも人が行う分類の一致率などに使われている[3]。計算の結果、 $\kappa$ 値は表3の最右列に示す値になった。一般に、 $\kappa$ 値は1に近いほど一致していることを表し、 $\kappa=1$ になった場合は完全な一致となる。 $\kappa$ が0.81~1.00の間にあればほぼ完全な一致、0.61~0.80の間にあれば実質的に一致、0.41~0.60の間ならば中等度の一致とみなされる。したがって、言い換え作業の結果、作業者ACの間では実質的な一致が、AB、BCの間では中程度の一致が見られた。

以上の結果は、本手法が同一判定者による分析・実験だけでなく、異なる判定者による実験からも意図判定基準として要求か否かを判定できること、すなわち、訓練された判定者以外にも要求を取り出すことが再現できることを示している。

言い換え不可で一致した文には、下記#11のように消極的に現状を受容する意図、#12のように積極的に受容する意図が見られた。また、#13は「理由は」のように、明示的に理由を述べる表現から始まる文である。これらは要求文一理由文という複数の文からなる回答の一部であり、要求の動機を示している。

#11 必要があれば料金の値上げもしょうがないと思う。

#12 車イスの人でも、楽にあちこち一人で買い物や、散歩できる町、道路ですばらしい。

#13 理由は全体的な発展が望めない。

<sup>6</sup> 作業者Xの言い換え可の割合:  $x(\alpha)$   
 作業者Xの言い換え不可の割合:  $x(\beta)$   
 作業者Yの言い換え可の割合:  $y(\alpha)$   
 作業者Yの言い換え不可の割合:  $y(\beta)$   
 作業者XYが共に言い換え可の割合:  $xy(\alpha)$   
 作業者XYが共に言い換え不可の割合:  $xy(\beta)$   
 両者の言い換え可が偶然に一致した割合 a:  $y(\alpha) \times x(\alpha)$   
 両者の言い換え不可が偶然に一致した割合 b:  $y(\beta) \times x(\beta)$   
 観察された一致率  $P_o$ :  $xy(\alpha) + xy(\beta)$   
 偶然の一致率  $P_e$ :  $a+b$   
 $\kappa$  値:  $(P_o - P_e) / (1 - P_e)$

以上のように、言い換え可の文は要求文を、言い換え不可の文は、現状の受容あるいは要求動機を示すということがわかった。

#### 4.2. 有効性の検証 ー基準を用いない意図判定ー

ここでは、提案した判定基準の有効性を検証するため、基準を用いずに回答テキストが要求を示すか否かの判定実験を行った。被験者は日本語を母語話者とする大学生 2 人で、一人は教育学部に在籍する 20 代男性、もう一人は文学部に在籍する 20 代女性である。4.1 節の実験で作業員 B C が対象とした回答テキスト 4000 文と同じものを使用した。作業員間で相談することなく、個別に判定を行った。なお、判定にあたって次の作業指示を提示した。

- ・例示する各回答文には文脈的つながりがないこと
- ・回答文が要求であると思う場合は 1、要求ではないと思う場合は 0 を直観的に判断すること
- ・無回答やその他の回答は無効になるので、必ずいずれかを選択すること

上記の実験により表 4 の結果が得られた。被験者 E に無回答が 2 件あったため、得られた回答数は 3998 件である。この検証は 4.1 節の実験で作業した B C と共通のデータを使用しているため、表 4 では結果を併記する。D E の作業の一致率は、B C より低く、かつ 4.1 節における再現性の実験結果 (表 3) のいずれの場合よりも低い 0.17 の値となっている。また、0.17 という値は、 $\kappa$  値の観点からは一致がないことを示している。このことから、主観的な判定による言い換えは基準がないため、結果が安定しないことがわかった。すなわち、言い換えを用いた判定基準の有効性が示された。

#### 5. おわりに

本稿では、自由記述アンケートの回答から「要求」を取り出すために、要求意図を判定する基準として、典型的な要求表現である「てほしい」への言い換えによる要求意図判定基準を作成した。この判定基準について分析的検証を行い、判定基準によって要求が取り出せること、かつ、従来よりも網羅性が高いことがわかった。また、分析にもとづいて作成されたコーパスを対象に行った機械学習による実験から、判定基準の適用結果が客観的であることがわかった。さらに、判定基準を用いた複数の作業員による言い換え実験における言い換え可否の一致率から判定基準の再現性の高さを示した。また、この結果と、判定基準を用いずに要求か否かを主観的に判断する実験の結果との比較から、判定基準の有効性を示すことができた。

本稿では割愛したが、検証結果の考察から 4.1 節の言い換え作業における言い換え可否の不一致は作業仕様において要求の動機と内容を明確に区別していなかったことが原因であるとわかった。今後は「ので／ため／から」等の理由節に言い換えられる節および文を「要求動機」とし、「ので」を付与して要求動機に言い換えられるか否かとい

表 4 有効性の実験結果

D	E			B	C		
	要求	非要求	計		可	不可	計
要求	562	1880	2442	可	2119	50	2169
非要求	39	1517	1556	不可	934	897	1831
計	601	3397	3998	計	3053	947	4000
DEの $\kappa$ 値	0.17			BCの $\kappa$ 値	0.49		

う判定基準を作成・検証する予定である。そのうえで、要求動機・要求内容・要求意図といった要求の要素の区別を適切に行い、「要求を知る」ための抽出技術の向上を目指すと考えている。

**謝辞:** 研究データとして道路審議会基本政策部会「21世紀の道を考える委員会」が実施されたボイス・レポートについて研究利用を快諾してくださった(財)国土技術研究センター調査第二部の前田様、川原様のご厚意に深謝いたします。

#### 参考文献:

- [1] Berger, A.L., Della Pietra, S.A. and Della Pietra, V.J. (1996) A maximum entropy approach to natural language processing. *Computational Linguistics*, Vol. 22, No. 1, pp39-71.
- [2] Cristianini, N. and Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines*. Cambridge University Press.
- [3] Carletta, J. (1995) Assessing agreement on classification tasks: the kappa statistic. *Computational linguistics*, Vol. 30, No. 11
- [4] 土井昇一・蓬萊尚幸・渡部勇・片山佳則・園部正幸 (2003) 要求獲得会議を分析することによるユーザ指向要求獲得法. *情報処理学会誌*, vol. 44 No. 1, pp48-58.
- [5] 道路審議会基本政策部会 (1996) ボイス・レポート. 21世紀の道を考える委員会 建設省道路局 建設省都市局.
- [6] 乾裕子・内元清貴・井佐原均 (1998) モダリティ分析に基づく自由回答アンケートの分類. *言語処理学会第 4 回年次大会発表論文集*, pp540-543.
- [7] Inui, H., Murata, M., Uchimoto, K. and Isahara, H. (2001) Classification of Open-Ended Questionnaires based on Surface Information in Sentence Structure. *NLPRS2001*, pp315-322.
- [8] 乾裕子・高梨克也・井佐原均 (2001) 自由記述型アンケート回答を対象にした要求意図特定スキーマの提案. *言語処理学会第 8 回年次大会発表論文集*, pp104-107.
- [9] 乾裕子・井佐原均 (2002) 拡張モダリティの提案 - 自由回答から回答者の意図を判定するために -. *信学技報*, Vol. 102 No. 414, NLC2002-43, pp31-36.
- [10] 国立国語研究所 (1960) 「話しことばの文型」, 秀英出版.
- [11] 松井くにお・田中穂積 (2002) 単純質問文から蓄積された質問応答へのナビゲーション. *信学技報*, Vol. 102 No. 414, NLC2002-40, pp13-18.
- [12] 松澤裕史 (2002) 自然言語処理技術と構造化パターンマイニングを用いた F A Q 生成支援システム. *FIT 情報科学技術フォーラム*, pp69-70.
- [13] 森田良行・松木正恵 (1989) 「日本語表現文型」, アルク.
- [14] 那須川哲哉 (2001) コールセンターにおけるテキストマイニング. *人工知能学会誌*, Vol. 16, No. 2, pp219-225.
- [15] 柳瀬隆史・丸元聡子・難波功・落谷亮 (2002) 文末の述語表現を利用した質問テキストの解析. *言語処理学会第 8 回年次大会発表論文集*, pp647-650.