

2-gramに基づいたQ&Aサイトにおける適切な回答者の選出手法の評価

Evaluation of method to select appropriate qualified respondents at Q&A Sites based on 2-gram

横山 友也^{1*}

Yuya Yokoyama^{1*}

¹東京都立産業技術大学院大学 Advanced Institute of Industrial Technology

*Corresponding author: Yuya Yokoyama, yokoyama-yuya@aiit.ac.jp

Abstract For the purpose of resolving mismatches between the questioners and respondents at Question and Answer (Q&A) sites, factor scores were estimated using feature values of statements based on the scores of nine factors obtained through impression evaluation experiment. A method to select respondents who can appropriately answer a question was then proposed. As a result of analysis, the proposed method successfully selected potential respondents that were more than approximately average when it came to appropriately answering a given question. This approach, however, was considerably dependent on the syntactic information extracted through morphological analysis. Therefore, N-gram has been applied to the methodology as an alternative method of morphological analysis. In applying N-gram, the estimation of factor scores and detecting of the respondent who can appropriately answer a question have been realized so far. Hence, this paper aims to investigate respondents who can properly answer a newly posted question based on 2-gram. An experiment is carried out to assess three sets of 100 answer statements extracted in accordance with the Euclidean distances based on 2-gram. The analysis result has shown that the proposed method utilizing 2-gram can choose potential respondents that were more than approximately average when it came to answering appropriately. It has also been implied that the proposed method using 2-gram would also be applicable.

Keywords factor score; impression; N-gram, Q&A site; respondent introduction

1 はじめに

近年、質問回答サイトの利用者が増加の一途をたどっている。質問回答サイトとは、例えばYahoo!知恵袋[1]のように、オンライン上で質問文や回答文を人手で投稿できるシステムである。これらのサイトは、多種多様で膨大な知識を含む巨大なデータベースとして知られている。質問回答サイトの基本的な構造は下記の通りである。質問者が質問文を投稿すると、回答の内容や精度に関わらず他のユーザが回答する。投稿された複数の回答文のうち、最も適切で満足している回答文を「ベストアンサー」(BA)と質問者が主観的に選択し、謝礼として回答者にポイント等の報酬を付与する。

しかし、利用者ならびに投稿回答文数が激増していくに伴って、回答者の興味や専門性に合致した回答文を発見することが困難になりつつある。従って、質問文が投稿されても、その質問文が必ずしも適切な回答者の目に留まり回答されるわけではない。一方で、前述の通り、質問回答サイトは社会の知恵袋となりつつあるものの、必ずしも正しい回答が蓄積されているわけではなく、誤答も多く発見される。適切な回答が得られないことで損害を被る状況が多発しており、社会的に大きな問題となってきた。このような理由から、質問文に対して適切に回答できる利用者を求めるることは、適切な回答を質問者に返すとともに正しい知識を蓄積するという点で非常に重要である。また、罵倒・暴言・誹謗中傷ならびに公序良俗に反する回答文も、読者や閲覧者に不快感を与えるものとなっている。

従って、適切な回答文を蓄積することを目的として、適切な回答を高く期待できる回答者を要求することは非常に望ましい状況となっている。前述の問題の解消を目的として、テキストの特徴量を用いた質問回答サイトの先行研究が多数行われてきている[2-4]。しかし、先行研究ではユーザの文体の傾向が考慮されているとは言い難い。また、これらの手法において、質問者に適切な回答者を引き合わせる手法が確立されているとは言い難い。従って、文章から想起される印象を調べるために、本研究では質問者に適切な回答者を引き合わせることを目的とする。本研究の発展により、さらに適切な回答文を収集す

ることや社会に有用な質問回答サイトが可能となり、ひいては社会活動の生産的な促進に貢献できると考えられる。

筆者は、本研究の初步として、Yahoo!知恵袋に投稿された4大ジャンル（オークション、パソコン、恋愛相談・人間関係、政治・社会問題）の質問回答文12組計60個の文章に対して印象評価実験を行った[5]。その結果に対して因子分析を施したところ、文章内容を表す9因子が得られた。また、全ての質問回答文の因子得点を推定可能にすることを目的として、文章の特徴量から文章の因子得点の推定を重回帰分析により試みたところ、全般的に良好な推定精度が得られた[6]。さらに、新規質問文に適切な回答を施すことが可能な回答者を探索する可能性を検証した。検証の結果、質問回答文間の距離と距離上位の出現回数は、質問文に適切に回答できるユーザの選択に役立てる可能性を示した[7]。この観察に基づき、出現回数と距離に基づくスコアと距離上位の出現回数に応じて、適切な回答者を決定する手法を提案した。提案手法を、スコアの平均値に基づく手法ならびに距離に基づく手法と、適合率と再現率で比較評価を行った[8, 9]。評価の結果、提案手法は、他手法よりも良好に回答者を推薦できることを示し、その精度は質問回答文のジャンルを考慮することによって向上することを示した[9]。

ここまで形態素解析を講じて抽出した構文情報に重きを置いた手法である。他方、構文解析のもう一種の手法であるN-gramを講じた手法も試みてきている。これまで、構文情報の代わりにN-gramを文章の特徴量として、質問回答文の因子得点の推定を試みた[10-12]。推定の結果、N-gramも構文情報と同様に因子得点の推定に利用可能であることを示している。また、日本語の場合と同様に、質問回答文間の距離と距離上位の出現回数は、質問文に適切に回答できるユーザの選択に役立てる可能性も示している[13]。

これらの観察に基づいて、本稿では、構文情報の場合と同様に、N-gramに基づいた場合においても適切な回答者を推薦することが可能であることを提案する[14]。この手法の初步として、2-gram(N=2)の場合で分析を行った。構文情報の場合と同様に、提案手法を、スコアの平均値に基づく手法ならびに距離に基づく手法と、適合率と再現率で実験的に比較評価を行う。

分析の結果、提案手法は他手法よりも良好な結果が得られている。以上より、N-gramに基づいた提案手法も、平均程度以上の回答者を推薦することが可能であることを示している[14]。

本稿の構成は次の通りである。第2節では、これまでの研究として、形態素解析に基づいた構文情報を用いた場合の適切な回答者の推薦について説明する。第3節では、第2節で示した手法の一部をN-gramに適用した場合の分析結果について説明する。第4節では、N-gramに基づいた場合の適切な回答者の推薦について詳述する。最後に、第5節で本稿をまとめた。なお、本稿は著者が国際会議 SERA2024において発表した内容である[14]。

2 形態素解析に基づいた構文情報を用いた適切な回答者の推薦

2.1 文章の性質を表す因子の抽出

本研究の初步として、質問回答文から得られる印象を抽出するための印象評価実験を実施した。評価素材として、2005年9月にYahoo!知恵袋[1]に投稿された12組60件の質問回答文を使用した。実験結果に対して因子分析を施したところ、文章に関する因子が9個得られた[5]。因子とは、複数の印象語により説明された文章の性質を意味する。的確性、不快性、独創性、容易性、執拗性、曖昧性、感動性、努力性、熱烈性の9因子が得られた[2]。なお、因子名は、各因子に対応する印象語を包括的に表現していると著者達が判断した名称を付与している。因子に対応する印象語を表1に示す。また、文章の特徴を表現するのに使用する因子得点も得られている[5]。

表1 9因子と対応する印象語[5]

因子	印象語				
第1因子（的確性）	説得力がある	流暢な	重要な	美しい	好ましい
	真実味がある	巧みな	清々しい	妥当な	充実した
	素晴らしい	的確な	丁寧な		
第2因子（不快性）	非常識な	憤慨した	不快な	残念な	不当な
	幻滅した	呆れる	怖い		
第3因子（独創性）	独創的な	予想外な	特殊な	斬新な	不思議な
第4因子（容易性）	易しい	明瞭な	難しい		
第5因子（執拗性）	細かい	しつこい	長い		
第6因子（曖昧性）	曖昧な	不充分な			
第7因子（感動性）	心温まる	感動的な			
第8因子（努力性）	涙ぐましい				
第9因子（熱烈性）	熱い	力強い			

2.2 文章の特徴量を用いた因子得点の推定

2.1節で得られた因子得点は、実験に用いた質問回答文60件から得られたものだけである。そこで、任意の文章の因子得点の推定も可能とするために、文章の特徴量に対し重回帰分析を講じた。ここで、構文解析の一種である形態素解析を通して抽出した文章の特徴量を用いて分析を行った[6]。分析に使用した文章の特徴量77個(g1~g77)を表2に示す。以下、各特徴量を簡潔に説明する[6]。

- 構文情報 (g1 - g36)：文章の数や長さ、名詞や動詞等の品詞の数や割合、といった構文情報を抽出した。感嘆符や疑問符等の具体的な記号も特徴量として使用している[6]。なお、g18のTTRはType Token Ratioの略記で、「ある文章における

語彙の豊富さを示す指標」であり、文章の総語数に対する語彙数の比率を表している[6]。

- 単語心像性 (g37 - g38)：単語から喚起されるイメージが、どの程度思い浮かべやすいかを示す主観的特性である[6]。
- 文末表現 (g39 - g64)：「ぞ」「だ」「よ」「ね」「か」「な」「し」「です」「ます」「たい」「ない」を使用している[6]。
- 単語親密度 (g65 - g71)：単語にどの程度なじみがあるかを表す指標である[6]。
- 表記妥当性 (g72 - g77)：単語表記の妥当性を表す[6]。

表2 文章の特徴量 [6]

(a) 構文情報

g	特徴量	g	特徴量
g1	助動詞（語彙数）	g19	全角記号 (%)
g2	接頭詞	g20	英数字 (%)
g3	記号（語彙数）	g21	全角英数字 (%)
g4	文数	g22	名詞 (%)
g5	文の長さ平均（字数）	g23	形容詞 (%)
g6	カタカナ（語数）	g24	副詞 (%)
g7	全角記号（語数）	g25	連体詞 (%)
g8	全角英数字（語数）	g26	接続詞 (%)
g9	形容詞（語数）	g27	感動詞 (%)
g10	副詞（語数）	g28	「！」の数
g11	連体詞（語数）	g29	「？」の数
g12	接続詞（語数）	g30	句点の数
g13	感動詞（語数）	g31	読点の数
g14	ひらがな (%)	g32	中点の数
g15	漢字 (%)	g33	3点リーダの数
g16	カタカナ (%)	g34	鍵括弧の数
g17	記号 (%)	g35	括弧の数
g18	TTR	g36	「/」の数

(b) 単語心像性

g	特徴量	g	特徴量
g37	単語心像性4点台（語数）	g38	単語心像性6.5以上7.0未満（語数）

(c) 文末表現

g	特徴量	g	特徴量
g39	か（語数）	g52	ぞ (%)
g40	な（語数）	g53	だ (%)
g41	し（語数）	g54	よ (%)
g42	たい（語数）	g55	ね (%)
g43	ない（語数）	g56	か (%)
g44	だ（文末語数）	g57	です (%)
g45	か（文末語数）	g58	ます (%)
g46	な（文末語数）	g59	ない (%)
g47	し（文末語数）	g60	か（文末%)
g48	です（文末語数）	g61	ですか（語数）
g49	ます（文末語数）	g62	ないです（語数）
g50	たい（文末語数）	g63	ますか（語数）
g51	ない（文末語数）	g64	ました（語数）

(d) 単語親密度

g	特徴量	g	特徴量
g65	単語親密度該当単語率	g69	単語親密度5.5以上6.0未満（語数）
g66	単語親密度6.5以上7.0未満（語彙数）	g70	単語親密度6点台（語数）
g67	単語親密度4点台（語数）	g71	単語親密度6.0以上6.5未満（語数）
g68	単語親密度5点台（語数）		

(e) 表記妥当性

g	特徴量	g	特徴量
g72	表記妥当性該当単語率	g75	表記妥当性4点台（語数）
g73	表記妥当性3点台（語数）	g76	表記妥当性4.0以上4.5未満（語数）
g74	表記妥当性3.5以上4.0未満（語数）	g77	表記妥当性5点台（語数）

印象評価実験で使用された60個の質問回答文に関して、表

2に示した計77個の説明変数に基づく281個の二次項（説明変数同士の積）を説明変数とし、表2で示した9因子の因子得点を目的変数として、重回帰分析を施した[6]。推定精度の良好性を示す重相関係数を表3に示す。全9因子とも値が0.9以上であるため、9因子とも推定精度が非常に良好であるといえる。

表3 重相関係数（構文情報）[6]

因子	重相関係数
第1因子（的確性）	0.989
第2因子（不快性）	1.000
第3因子（独創性）	0.999
第4因子（容易性）	1.000
第5因子（執拗性）	0.925
第6因子（曖昧性）	1.000
第7因子（感動性）	0.963
第8因子（努力性）	0.950
第9因子（感動性）	1.000

2.3 質問回答文の印象と適切度

新規質問文に適切な回答を施すことが可能な回答者を探査する可能性を検証した。ここでは、質問文の印象と類似している回答文の印象により、質問文に適切な回答を施すユーザを探査することが可能であるという前提で、質問文の因子得点と既に投稿された回答文の因子得点との距離を計算し利用する。具体的には、2.2節での重回帰分析で得られた重回帰式をもとに、2.1節での印象評価実験に使用していない任意の質問回答文の因子得点を算出する[7]。

質問回答文間の印象の差異と回答文の適切性の検証に使用した質問回答文について、質問文は、オークション、パソコン、恋愛相談・人間関係の各1件の文章である。回答文は、66,238件の回答文を使用している。質問文と66,238件の回答文の各ユーリッド距離は、式(1)を用いて算出する。

$$D = \sqrt{\sum_{k=1}^9 (Fac_{Q_k} - Fac_{A_k})^2} \quad (1)$$

ここで、 Fac_{Q_k} と Fac_{A_k} は、それぞれ質問文の因子得点、 k 番目の回答文の因子得点を示している。分析の結果、質問回答文間の距離は、適切な回答者を選択することに役立てることが示された[7]。また、実際の回答投稿者以外も、適切な回答者になる可能性も示されている[7]。

2.4 適切な回答者の選出手法と評価

2.4.1 目的

2.3節で説明した観察結果に基づいて、新規に投稿された質問文に対して適切な回答を施す可能性が高いと考えられる回答者を選出手法を提案する[8, 9]。提案手法では、以下の基準に基づいて適切な回答者になり得る回答者を選出手する。

1) 質問回答文間の因子得点の距離の昇順で上位100件の回答文内の回答者の出現回数

2) 複数の回答者が同じ回数出現した場合、の合計。ただし、 $Score_k$ は式(2)より求める。

$$Score_k = \frac{Dis_A_1}{Dis_A_k} \quad (1 \leq k \leq 100) \quad (2)$$

ここで、 Dis_A_k は距離の昇順で k 番目の回答文の距離を表す。したがって、あるユーザの回答を $A_{k_1}, A_{k_2}, \dots, A_{k_n}$ とすると、そのユーザが有するスコア S は式(3)により求まる[8, 9]。

$$S = \sum_{i=1}^n Score_{k_i} \quad (3)$$

2.4.2 客観的な適切度（FR）

適切な回答者を機械的な方法で決定することは不可能であるため、人手の協力を要する実験を実施する必要があった。ここで、適切な回答者を決定する手法は存在しており、ある回答文が質問文を解決するのに適切であるかどうかを被験者が決定する方法である。しかし、回答の内容によっては、一見によらず実際は適切でない場合もあるため、回答を適切であるかどうかを被験者が判定するのは容易ではない。そこで、データセット内の質問文に対する回答文を施したユーザが適切な回答を施してくれそうかどうかを実験的に獲得する。この適切度を“Fitness Rate (FR)”と定義する[8, 9]。ここで、ユーザに対するFRの平均値を“Averaged Fitness Rate (AFR)”と称する[8, 9]。AFRは平均値であるので、客観的な尺度であると考えられる。データセット内の範囲でユーザが回答文を複数施している場合、代表的なAFRを決定する必要がある。ここでは、最良のAFRをユーザの代表的なAFRとして定めて、これを“Best Averaged Fitness Rate (BAFR)”と称する。ユーザがデータセット内で1件しか回答文を施していない場合は、その回答文のAFRが自動的にBAFRとなる。

2.4.3 FRを用いた実験

適切な回答者を機械的な方法 14名（22~29歳の男性10名、女性4名）の被験者に対して印象評価実験を行った[8, 9]。評価素材は、質問文1件と質問文に非対応の回答文100件から構成されるセットを3組（オークション、パソコン、恋愛相談の各1組）用意している。100件の回答文の選出方法は、質問回答文間の距離が上位100件を抽出している。被験者には、まず質問文を読んでもらった後で、各回答文を読んでもらい、回答文を施した回答者が質問文に適切な回答を施すことが期待できる度合の評価として、FRを5段階（5：非常に期待できる、4：期待できる、3：どちらともいえない、2：あまり期待できない、1：全く期待できない）で付与してもらった。評価素材には同一回答者が施した回答文が複数あつたが、この旨は被験者に伝えていない。

各手法の妥当性を適合率と再現率により評価を行う。適合率と再現率は、式(4)ならびに(5)でそれぞれ算出する[8, 9]。

$$\text{適合率} = \frac{\text{検索適切回答者数}}{\text{検索回答者数}} \quad (4)$$

$$\text{再現率} = \frac{\text{検索適切回答者数}}{\text{適切回答者数}} \quad (5)$$

ここで、適切回答者とは、「BAFR が閾値n以上の回答文を含む回答者」と定義する。適合率と再現率は、閾値nを 2.0~4.0 の区間ににおいて 0.5 間隔で設定して求める。分析の結果、提案手法が他の 2 手法よりも良好な精度が得られている[8, 9]。また、平均程度以上の回答者を推薦することが可能であることを示した。さらに、質問回答文のジャンルを考慮することで向上することも示している[9]。

3 N-gram に基づく手法

3.1 目的

2 節で述べた分析手法は、主に形態素解析を通じて抽出された文章の特徴量を使用した手法である。しかし、表 2 に示したように、使用した文章の特徴量 77 種のうち 36 種が形態素解析に基づいた構文情報であることから、この手法は形態素解析に重きを置いた手法となっている。また、因子得点を推定するための重回帰式も複雑なモデルとなっている。一方で、形態素解析と同様に構文解析として知られている方法として N-gram がある。そこで、図 1 に示すように、形態素解析に基づいた構文情報を主に特徴量に使用した場合と同様の分析手法で、N-gram を特徴量として用いた場合に因子得点の推定が可能であるかどうかを検証する[10-12]。また、質問回答文の印象と適切度についても、構文情報と同様の傾向を確認している[13]。その上で、N-gram の場合に基づいた場合においても適切な回答者の選出手法と評価が可能であることを示す。

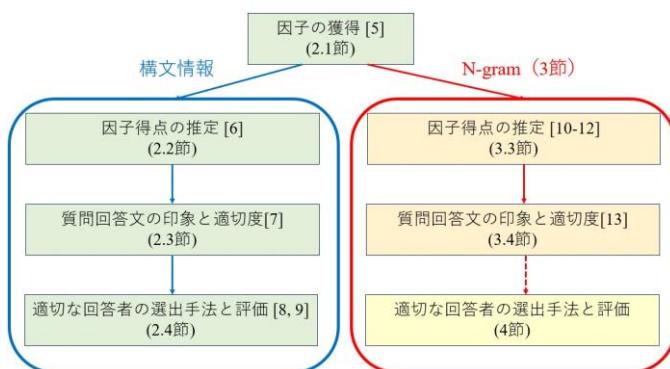


図 1 本研究手法のフローチャート

3.2 N-gram

N-gram は、テキスト内における N 個単位の文字や形態素、あるいは品詞の連続した形態である[15, 16]。N は 2 以上の任意の自然数が入るが、一般的には N=2 または N=3 が適用され、それぞれバイグラム (bigram)、トリグラム (trigram) と称される。たとえば、テキスト内における「今日は」を例にすると、文字 3-gram であれば「今 日 は」のような 3 文字の連鎖、単語 2-gram であれば「今日 は」のような 2 単語の連鎖、品詞 2-gram であれば「名詞 助詞」のような 2 品詞の連鎖をそれぞれ表している。このように、N-gram は隣接する要素を機械的に抽出する手法であり、様々な分野に適用されている[9]。文字 N-gram は、形態素解析を要することなく集計することが可能であり、形態素解析における分かち書きの誤り等の影響を受けずにテキストを分析することが可能である[10]。さらに、品詞 N-gram は、文章を品詞の単位に抽象化するため、

文章の内容の影響に左右されることなく文章の構造を捉えることができる利点がある[16]。

2.1 節で使用した 60 件の質問回答文のうち、1 つの質問文に品詞 2-gram を適用した例を表 4 に示す。便宜として、表 4-(a) に例として示した質問文を QA04 と表す。表 4-(b) は、QA04 の品詞 2-gram の例と出現数について、出現数の降順で示している。出現数が 2 以上の品詞 2-gram に関しては、それぞれの一例のみを「例」の項に示している。

表 4 品詞 2-gram の例 [8-10]

(a) 文章の例 (QA04)

QA04	パソコン初心者です。デジカメで撮った画像をプリントアウトしたところ画像が暗いのですが、明るくする方法をご存知の方回答をお願いします。
------	--

(b) 2-gram の適用例ならびに出現数

2-gram	例	出現数
[名詞 - 助詞]	[画像 - を]	6
[名詞 - 名詞]	[パソコン - 初心者]	4
[助詞 - 名詞]	[の - 方]	4
[動詞 - 助動詞]	[する - ます]	3
[助動詞 - 記号]	[ます - 。]	2
[助動詞 - 名詞]	[た - ところ]	2
[名詞 - 助動詞]	[初心者 - です]	2
[名詞 - 動詞]	[お願い - する]	2
[記号 - 形容詞]	[、 - 明るい]	1
[記号 - 名詞]	[。 - デジカメ]	1
[形容詞 - 動詞]	[明るい - する]	1
[形容詞 - 名詞]	[暗い - の]	1
[助詞 - 記号]	[が - 、]	1
[助詞 - 形容詞]	[が - 暗い]	1
[助詞 - 動詞]	[で - 撮る]	1
[助動詞 - 助詞]	[です - が]	1
[動詞 - 名詞]	[する - 方法]	1

3.3 N-gram を用いた因子得点の推定

2.2 節で記述した手法と同様に、N-gram を用いた文章の特徴量を使用して因子得点の推定を試みた。2-gram (N=2) から 5-gram (N=5) までの場合について分析を行った。統計解析フリーソフ R[17]において、RMeCab のパッケージを用いて品詞 N-gram の特徴量を抽出した。60 件の質問回答文をもとに、全パターンの品詞 N-gram を生成し、それぞれの品詞 N-gram の合計出現数を算定する。その後、合計出現数の降順に品詞 N-gram を並べて、試験的に上位 17 件の品詞 N-gram を特徴量として使用した。N-gram の特徴量を Ngr_g1, Ngr_g2, ..., Ngr_g17 と表している。したがって、2-gram を例とすると、2gr_g1, 2gr_g2, ..., 2gr_g17 と表す。3-gram, 4-gram, 5-gram についても、2-gram と同様の表記に準じる。一例として、品詞 2-gram の特徴量を表 5 に示す。

表5 2-gram の特徴量 [8]

2-gram	特徴量
2gr_g1	[名詞-助詞]
2gr_g2	[助詞-動詞]
2gr_g3	[助詞-名詞]
2gr_g4	[名詞-名詞]
2gr_g5	[記号-名詞]
2gr_g6	[動詞-助動詞]
2gr_g7	[助詞-記号]
2gr_g8	[動詞-助詞]
2gr_g9	[助動詞-助詞]
2gr_g10	[名詞-助動詞]
2gr_g11	[助動詞-記号]
2gr_g12	[動詞-名詞]
2gr_g13	[名詞-動詞]
2gr_g14	[助動詞-名詞]
2gr_g15	[助動詞-助動詞]
2gr_g16	[記号-記号]
2gr_g17	[助詞-助詞]

重回帰分析に関して、2.2 節の場合と同様に、質問回答文 60 件の因子得点を目的変数に設定する。一方で、2.2 節での分析と異なる点として、形態素解析に基づいた構文情報 (36 種:g1-g36) に代わって、N-gram に基づいた特徴量 (17 種:Ngr_g1-Ngr_g17) を新規特徴量として説明変数に用いる[6-8]。他方、2.2 節で使用した単語心像性 (g37-g38)・文末表現 (g39-g64)・単語親密度 (g65-g71)・表記妥当性 (g72-g77) の計 41 種については引き続き説明変数として使用する。以上より、計 58 種の特徴量を説明変数として設定する。

重回帰分析の推定精度の指標を示す重相関係数を表 6 に示す。この結果より、どの N-gram においても重相関係数が 0.9 以上になっており、全般に良好な推定精度が得られている[6-8]。表 3 の形態素解析を用いた場合の結果と比較すると、形態素解析を用いた場合と同等程度の推定精度が得られている。

表6 重相関係数 (N-gram) [8-10]

因子	2-gram	3-gram	4-gram	5-gram
第1因子 的確性	0.989	0.993	0.999	0.998
第2因子 不快性	0.999	0.987	0.985	0.991
第3因子 独創性	0.981	0.998	0.971	0.976
第4因子 容易性	0.990	0.995	0.993	0.994
第5因子 執拗性	0.993	0.976	0.994	0.999
第6因子 曖昧性	0.998	0.994	0.983	0.986
第7因子 感動性	0.999	0.996	0.945	0.992
第8因子 努力性	0.995	0.968	0.988	1.00
第9因子 热烈性	0.995	0.998	0.973	0.954

3.4 2-gram を用いた質問回答文の印象と適切度

2.3 節で記した構文情報に基づいた手法と同様に、N-gram に

基づいた手法でも新規質問文に適切な回答を施すことが可能な回答者を探索する可能性を検証した[13]。ここでは、質問文の印象と類似している回答文の印象により、質問文に適切な回答を施すユーザを探索することが可能であるという前提で、2-gram を用いた場合の検証を行った。データセットは 2.3 節で示したものと同じである。分析の結果、構文解析の場合と同様に、N-gram の場合においても、質問回答文間の距離は、適切な回答者を選択することに役立てることが示された[13]。

4 2-gram に基づいた適切な回答者の選出手法と評価

4.1 手法の概要

3.4 節で示した観察に基づいて、構文情報の場合と同様に、N-gram の場合でも適切な回答者の選出手法が可能であるかどうかを検証する[14]。この手法の初步として、2-gram の場合で検証を行う。提案手法の選択に基づく基準は、2.4.1 節で示すように、回答文の出現回数ならびに質問回答文間の距離の順位に基づくスコア S である。

2.4.3 節で示したと同様の実験に被験者 11 名（男性 8 名、女性 3 名）が参加した。質問文を一読し、その後に質問文に非対応の回答文の集合に対する評価を 5 段階評価で実施した[14]。実験の評価素材として、質問文 1 件と非対応の回答文 100 件のセット 3 組を使用した。回答文 100 件の選定は、3.3 節で示した 2-gram に基づく距離から得られている。被験者に提示した回答文の順序については、無作為に並べた上で提示している。

4.2 t 検定

実験結果の性差の影響を調べるために、男性 8 名の回答と女性 3 名の回答に有意差があるかどうかを有意水準 1% で t 検定を行った[14]。検定の結果、p 値が 4.47×10^{-10} となり、p 値が 1% を大きく下回るため男女間に大きな性差があることを示している。したがって、以後の評価において、男性分の回答と女性分の回答は別々に取り扱った上で分析を実施する。

4.3 評価手順

構文情報に基づいた実験の場合と同様に、提案手法を他の 2 手法と比較検証する。回答文のスコアの平均値に基づく「スコア平均の手法」と、回答文との距離の最小値に基づく「距離最小の手法」の 2 手法である。提案手法を含めた 3 つの手法について、2.4.3 節において示した式(4)ならびに式(5)のように、適合率と再現率を用いて評価を行う[14]。

適合率と再現率を評価するにあたって、2.4.3 節で示したように、「回答者の BAFR が T_{FR} 以上の回答者」を適切回答者と定義する。適合率と再現率を $T_{FR} = 4.0, 3.5, 3.0, 2.5$ と設定した上で計算する。実例を用いながら適合率と再現率の計算方法について詳述する。実際の結果の一部として、 $T_{FR}=4.0$ の時にオーケションに対する男性 8 名分のスコアを提案手法で評価した場合を表 7 に示す。提案手法に基づいて、出現回数ならびにスコア S をソートする。「順位 (提案手法)」という項は提案手法に基づく順位を示しており、BAFR の項は各回答者の BAFR で示す。

表 7 実際の評価の例

(オークション、男性、提案手法、 $T_{FR} = 4.0$) [14]						
順位 (提案手法)	出現回数	S	BAFR	$T_{FR} = 4.0$		
				適切回答者	適合率	再現率
1	35	18.3	3.71	×	0.000	0.000
2	22	11.9	4.00	○	0.500	0.333
3	7	3.97	4.00	○	0.667	0.667
4	3	1.63	3.57	×	0.500	0.667
5	3	1.50	3.00	×	0.400	0.667
6	3	1.49	3.00	×	0.333	0.667
7	2	1.12	4.00	○	0.429	1.00

閾値 $T_{FR}=4.0$ の時、BAFRが4.0以上の回答者が適切回答者と判定され、その場合は「○」が「適切回答者」の項に付与され、BAFRが4.0未満の場合は「×」が付与される。再現率が1.00に達した場合、たとえ未検索の回答者が残っていたとしても、全ての適切回答者の検索が完了したことを表す。表7の例では、ユニークユーザ30名中、 $T_{FR}=4.0$ の時にBAFRが4.0以上であるために、提案手法の基準に基づいた順位2番目・3番目・7番目のユーザが全て適切回答者として判定される。これらの回答者は表7において網掛けを施している。この例では、検出回答者数は7、適切回答者数は3である。したがって、式(4)より、適合率は $3 \div 7 = 0.429$ となる。同様に、式(5)より、再現率は2番目・3番目・7番目においてそれぞれ適切回答者が検索される度に $1 \div 3 = 0.333$ 、 $2 \div 3 = 0.667$ 、 $3 \div 3 = 1.00$ となる。このようにして、他の場合についても適合率・再現率の計算を施している。

4.4 実験結果

2-gramに基づいた適切な回答者の選出手法の評価について、紙面の都合のため、結果の一部のみを計算する。オークションに対する男性のスコアを提案手法で評価した時の適合率と再現率を図2に示す。横軸は再現率、縦軸は適合率をそれぞれ示している。 $T_{FR} = 3.5$ または $T_{FR} = 3.0$ の時、再現率が低いうちに複数の回答者を検出しているため、提案手法は上位の回答者を選択することが可能であるといえる[14]。この傾向は、図3に示す構文情報に基づいた場合と同様のものが得られている。

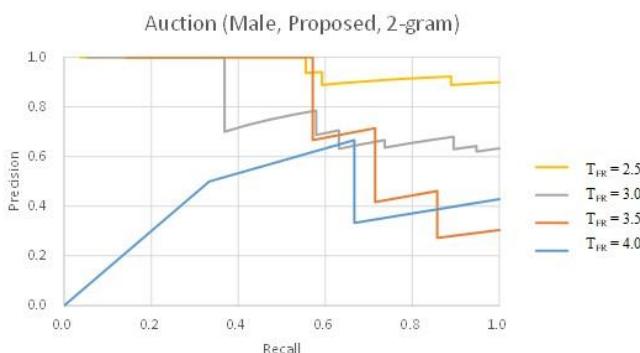


図 2 適合率と再現率
(オークション、男性、提案手法、2-gram) [14]



図 3 適合率と再現率
(オークション、男性、提案手法、構文情報) [9]

しかし、2-gramと構文情報どちらの場合においても、適切回答者数が不十分なため、 $T_{FR} = 4.0$ の挙動は不安定である。概して、構文情報の場合と同様に、2-gramに基づいた提案手法も、質問に適切な回答にあたって平均程度以上の回答者を選択することが可能であることがいえる[14]。図2以外の他の組み合わせ（男性・女性、オークション・パソコン・恋愛相談、提案手法・スコア平均の手法・距離最小の手法）においても、概ね同様の傾向が得られている。

4.5 考察

スコア平均の手法、距離最小の手法の3手法の比較を行う。紙面の都合上、男性の $T_{FR} = 3.0$ の時の結果を図4、5、6に示す。図2、3と同様に、横軸が再現率、縦軸が適合率をそれぞれ表している。総じて、提案手法が他の2手法よりも良好な結果が得られている[14]。但し、オークションにおいては、適合率が約0.35以上においては他の2手法よりも下回っている。一方で、構文情報の時は、オークションとパソコンが一番良好な結果を示していたが、恋愛相談の場合は距離最小の手法が提案手法を上回る結果を示していた。これらの傾向を確認するためには、2.1節で述べた印象評価実験を含めた追加実験を行う必要がある[14]。

次に、2-gramと構文情報の場合で提案手法を用いた時の適合率を表8に示す[14]。再現率の右側に、上から順に検索適切回答者数と適切回答者数がそれぞれ示されている。4.3節において表7で示した実際の評価例（オークション、男性、提案手法、 $T_{FR} = 4.0$ ）において、検出回答者数7を適切回答者数3で除算することで、適合率は0.429となり、4.3節での説明と符合することが確認できる。これらを比較すると、2-gramと構文情報は類似した数字が得られている。 $T_{FR} = 3.5$ ならびに $T_{FR} = 3.0$ の時、適切回答者数が高い順位にランク付けされるため、提案手法は適切な回答者を選択するのに実用的であるといえる[14]。 $T_{FR} = 2.5$ の時は、適合率が0.920にも到達するため、提案手法は平均程度以上の回答者の選択を回答することが可能であることがいえる[14]。総じて、N-gramは適切な回答者を選択するのに構文情報と同等の役割をなすといえる[14]。また、3.4節で述べたように、N-gramのNを2から5まで適用して因子得点の推定結果を示している。因子得点の推定は2-gramの時点では十分な推定結果が得られている[10-12]。したがって、3-gram以降に基づいた場合において適切な回答者を選

出する手法を行うには及ばない[14]。しかし、 $T_{FR} = 4.0$ の場合においては、適切回答者数が不十分なため、適合率の数値が不安定である。この課題を解消すべく、さらに被験者数を増やした上で追加実験を行う必要がある。

5 まとめ

本稿では、構文情報と同様の傾向が得られるかどうかを調べるために、N-gramに基づいた適切な回答者の選出手法について評価を行った。提案手法は、回答者の出現回数ならびに適切な回答者の順位に基づいて決定されたスコアを使用している。この手法について、提案手法をスコアの平均値に基づく手法や距離に基づく手法と比較しながら、適合率と再現率で実験的に比較評価を行った。分析の結果、提案手法は他手法よりも良好な結果が得られたため、N-gramに基づいた提案手法も、平均程度以上の回答者の推薦が可能であることを示した。

今後の課題として、質問回答文の意味や内容を考慮する必要性がある。本手法で使用した特徴量の大半が日本語に基づいているため、本手法を他言語にも適用できるかどうかを検証する必要がある。また、本手法を他分野のデータセットにも汎用可能であるかどうかを検証することも、今後の課題に含まれる。

謝辞

本研究は、一部、科研費（26008587, 20K19933）の助成を受けて行われたものである。また、国立情報学研究所から提供を受けた、Yahoo! 知恵袋のデータを利用している。

参考文献

1. Yahoo! 知恵袋. [cited 19 Nov 2024]. Available: <http://chiebukuro.yahoo.co.jp/>.
2. E. U. Haq, T. Braud and P. Hui, "Community Matters More Than Anonymity. Analysis of User Interactions on the Quora Q&A Platform. Proc. of the International Conference Series on Advances in Social Network Analysis and Mining (ASONAM 2020). 2020: 94-98.
3. Wang, L., Zhang, L. and Jiang, J. IEA: An Answerer Recommendation Approach on Stack Overflow. *Science China Information Sciences*. Volume 62, 2019: 19 pages.
4. Anandhan, A., Ismail, M. A. and Shuib, L. Expert Recommendation through Tag Relationship in Community Question Answering. *Malaysian Journal of Computer Science*. 2022;35;3: 201-221.
5. 横山友也, 宝珍輝尚, 野宮浩揮, 佐藤哲司. 質問回答サイトの質問文と回答文の印象評価とベストアンサーの推定. 日本感性工学会論文誌. 2011;10;2: 221-230. Available: https://www.jstage.jst.go.jp/article/jjske/10/2/10_2_221/_pdf/-char/ja
6. Yokoyama Y., Hochin T., Nomiya H. Using Feature Values of Statements to Improve the Estimation Accuracy of Factor Scores of Impressions of Question and Answer Statements. *International Journal of Affective Engineering*. 2013;13;1: 19-26. Available: https://www.jstage.jst.go.jp/article/ijae/13/1/13_19/_pdf/-char/ja
7. Yokoyama Y., Hochin T., Nomiya H. Towards Detecting Appropriate Respondents to Questions Posted at Q&A Sites by Disregarding and Considering Categories of Answer Statements. *International Journal of Affective Engineering*. 2016;15;2: 167-

8. 横山友也, 宝珍輝尚, 野宮浩揮. 質問回答サイトにおける質問文への適切な回答者の選出法. 日本感性工学会論文誌. 2016;15;1: 21-29. Available: https://www.jstage.jst.go.jp/article/jjske/15/1/15_TJSKE-D-15-00033/_pdf/-char/ja/
9. Yokoyama Y., Hochin T., Nomiya H. Improvement of Obtaining Potential Appropriate Respondents to Questions at Q&A Sites by Considering Categories of Answer Statements. *International Journal of Affective Engineering*. 2017;16;2: 63-73. Available: <https://doi.org/10.5057/ijae.IJAE-D-16-00023>
10. Yokoyama Y., Hochin T., Nomiya H. Application of 2-gram and 3-gram to Obtain Factor Scores of Statements Posted at Q&A Sites. *International Journal of Networked and Distributed Computing*. 2022;10;1-2: 11-20. Available: <https://link.springer.com/article/10.1007/s44227-022-00005-2>
11. Yokoyama Y., Hochin T., Nomiya H. Using 4-gram to Obtain Factor Scores of Japanese Statements Posted at Q&A Sites. *Proceedings of the 13th International Congress on Advanced Applied Informatics (AAI 2022-Winter)*. 2022: 25-31. Available: <https://ieeexplore.ieee.org/document/10123522>
12. Yokoyama Y. Application of 5-gram to Obtain Factor Scores of Japanese Q&A Statements. *Proceedings of the 14th International Congress on Advanced Applied Informatics (AAI 2023)*. 2023: 69-75. Available: <https://ieeexplore.ieee.org/document/10371593>
13. Yokoyama Y. Impression and Suitability of Q&A Statements through Factor Scores Using 2-gram. *Proceedings of the 15th International Congress on Advanced Applied Informatics (AAI 2023-winter)*. 2023: 45-51. Available: [10.1109/IIAI-AAI-Winter61682.2023.00017](https://doi.org/10.1109/IIAI-AAI-Winter61682.2023.00017).
14. Yokoyama Y. Using 2-gram to Detect Potential Appropriate Respondents to Questions at Q&A Sites. *Proceedings of the 22nd IEEE/ACIS International Conference on Software Engineering, Management and Applications (SERA 2024)*. 2024: 304-309. Available: [10.1109/SERA61261.2024.10685602](https://doi.org/10.1109/SERA61261.2024.10685602).
15. 石田基広. R によるテキストマイニング入門 (第2版). 森北出版; 2017: 94.
16. 小林雄一郎. R によるやさしいテキストマイニング[活用事例編]. オーム社; 2018: 86-87.
17. The R Project for Statistical Computing. [cited 19 Nov 2024]. Available: <https://www.r-project.org>