

ユーザ交流サイトにおける利用活性化のための 投稿データ分析

渡邊 侑弥^{1,a)} 静野 祥彦^{b)} 福田 裕大^{1,c)} 安部 裕大^{1,d)} 末永 高志^{1,e)} 生田目 崇^{2,f)}

概要: 近年では、交流サイトや Q&A サイトによるユーザ同士の交流が活発に行われている。しかし、同じようなサービスが複数あり、サイト運営者はユーザ利用を促進させることがサイト運営をする上でカギとなる。ユーザ同士の交流を活性化させるには、ユーザがどのようにサイトに対してどのようなニーズがあるのかを把握することが重要である。本研究では、妊婦や小さい子供を持つ女性を主なターゲットとした交流サイトでの実際の投稿データを用い、どのような交流が行われているのかを分析し、サイト活性化に向けた情報抽出を行う。

キーワード: Q&A データ, トピックモデル, 利用促進

Analysis of Post Data Analysis in Social Site to Increase Usage

YUYA WATANABE^{1,a)} YOSHIHIKO SHIZUNO^{b)} YUDAI FUKUDA^{1,c)} HIROYUKI ABE^{1,d)}
TAKASHI SUENAGA^{1,e)} TAKASHI NAMATAME^{2,f)}

Abstract: In recent years, users have been actively interacting with each other through exchange sites and Q&A sites. However, on the other hand, there are multiple similar services, and promoting user usage is the key to site management for the site manager. In order to activate interaction among users, it is important to understand each user has what kind of needs for the site. In this study, we analyze what kind of interaction is taking place using actual posted data on a social sites mainly targeting pregnant women and mothers who have small children. Then we extract information for site activation.

Keywords: Q&A data, Topic model, increase usage

1. はじめに

インターネットの普及は企業にとっての BtoC の情報・販売チャネルとしてだけでなく、CtoC、すなわち消費者間ネットワークの拡大に大きく寄与した。

消費者同士の情報の受発信としては、商品やサービスに

関する口コミが一般的であり消費者の購買行動に大きく影響を与えていることは周知の事実といえる。

これに対して、消費者同士が情報を受発信するようなコミュニティサービスも情報受発信に広く使われている。汎用コミュニティサイトや SNS も広く使われているが、特定の消費者セグメントや話題を対象としたコミュニティは同じニーズを持つ消費者間でよく使われている。こうしたサイトでは、コミュニケーション機能だけではなく、関連する様々な情報も提供されており、情報共有のワンストップサービスとしての価値は高い。

コミュニティに参加する理由としては、関連する情報の収集とともに、消費者自身の課題解決が挙げられる。インターネットコミュニティにおいては、スモールワールド性

¹ (株) NTT データ
NTT Data Corp. Toyosu, Koto, Tokyo 135-8671, Japan

² 中央大学
Chuo University, Kasuga, Bunkyo, Tokyo 112-8551, Japan

a) Yuya.I.Watanabe@nttdata.com

b) Yoshihiko.Shizuno@nttdta.com

c) Yudai.Fukuda@nttdata.com

d) Hiroyuki.Abe@nttdata.com

e) Takashi.Suenaga@nttdata.com

f) nama@kc.chuo-u.ac.jp

やネットワーク外部性といった概念も知られており、人や情報をいか多くに集め、良質な情報の共有ができるかが、コミュニティサイトの成否を分けるといっても過言ではない。そのため、こうしたサイトの管理者は、サイト活性化のために利用者の維持、利用の促進を図ることが求められる。

本論文では、妊娠や子育てをする母親を主なターゲットとした情報サイトを事例として取り上げる。妊娠や子育ては人生における大きなライフイベントであり、ライフステージの変化といえ、多くの親が様々な疑問や戸惑いを有しながら生活を送っている。東京都北区の調査 [1] によると、妊娠屋出産について困ったこと、困っていることについて 23.3% の回答者が「他の妊婦との交流の場がないこと」、16.4% が「妊娠・出産についての相談相手が身近にいないこと」を挙げており、出産費用を除くとコミュニティの欠如が上位の理由として挙げられている。また、ヤフー・データソリューションがヤフーの検索データを分析した「ビッグデータからみるストレス・悩み」によれば、子育てママは 20 時以降に子育てブログを見ることでストレス緩和をしている傾向がみられると報告している。

このように、子育てにおいて、他人との接点、および問題解決や情報検索において情報サイトや Q&A サイトが重要な役割を持っていることがうかがえる。こうしたコミュニティサイトの活性化は利用する消費者にとって価値あるものであるとともに、運営する企業にとっての安定したサイト経営においても、どのようにサイトの活性化を図るかを議論することは重要な課題といえる。

そこで本論文では、ある子育てに関する情報サイトに投稿された Q&A データの分析を通して、コミュニケーション活性化に関する分析と提案を行う。

2. 関連研究

本節では、本論文の研究に関連する既存の研究について整理する。

島田ら [3] は妊娠・出産・子育てをする女性の Q&A サイトを対象にサイト内での共感を得られるような質問項目がどのような内容であるのかについて分析している。この研究においては、回答数が少ないような質問の回答率を上げることを目的として、質問の分類を行い、質問文をどのように表現すればよいのかについて考察している。具体的には word2vec による単語の分散表現から、類似単語を質問文に追加して質問文の拡張を行っている。このほかにも、一連の研究として、ガイドラインに違反した投稿の検出方法や、ライフステージが変わる出産予定日の予測モデルなどを提案し、同サイトの価値向上を目指した方法論を示している。

他の話題を主なターゲットとしたものとしては、ネットコミュニティにおける質問者と回答者の間での情報のやり

取り上の問題として、互いのことがよくわからないこと、および、相手の態度や発言の表現が気に入らないといった理由から知識共有バリアが働くことを挙げている [6], [7]。こうした問題を解決するために、尹ら [8] は、質問者と回答者の間を仲介するエージェントツールを提案し、仲介システムの介在で、やり取りが活性化することを実証している。

このほかにも、ネットワーク分析 [9], [10] や、コンテンツとコミュニケーションの進み方などに着目した研究 [11] などもある。

ただし、本研究で対象とする妊婦・子育てのインターネットコミュニティに関しての研究は十分ではない。特に、妊娠・子育て時期はライフステージが短期間に大きく変わり、親にも十分な子育て経験がなく、核家族化が進んでいる社会においては、親自身が主体的に情報収集にあたることも多い。したがって、一つのコミュニティサイト上であっても、様々なトピックが混在していることが想定され、今後のサイト成長のためにはより詳細な分析とサービス拡大に対する知見が求められると考える。

3. データ概要

本論文では、経営科学系研究部会連合協議会主催、令和 3 年度データ解析コンペティション、妊娠・出産・子育てに関する情報サイト「ママリ」を運営するコネヒト株式会社から提供された、Q&A データを用いる。提供データは 2019 年 1 月から 2021 年 7 月までの同サイト上での質問と検索キーワード、並びにユーザ情報であり、ユニークユーザ数はおよそ 300 万人、質問件数は 540 万件でその回答はおよそ 50 万件である。本研究では使用しなかったが検索はおよそ 1 億 3 千万件であった。投稿された質問とその回答、並びに回答間の親子関係についても ID によって把握でき、どのユーザが質問や回答を投稿したかについても把握できる。また、ユーザごとの子供の月齢などについての情報も含まれる。

3.1 基礎集計と分析方針

対象データの集計から、質問に大して回答がなされているものは全体の 90% と割合は高いものの、回答があったうちおよそ 90% は 10 件未満であり、短期間にその話題は止まってしまっている。反面で、逆に数か月といった長期にわたって回答屋やり取りが続くような質問もあり、長い期間盛り上がるような質問もあることが分かる。

ここで、回答が多く投稿されるような質問を目視で確認したところ、数か月間にわたって同じような内容の回答が続いたり、会話のような形でトピックが続いたりなど、インターネット上の文章でのやりとりとはいえ、ユーザ同士の深い交流がなされていることが見て取れる。また、回答数が多い質問と少ない質問で、ユーザを比較すると、ユーザの休眠状態、すなわちアクティブユーザであるかどうかにか

大きく差があることが分かった。

このような状況から、本サービスの課題として、質問に対する回答率は高いものの、多く回答が続くような質問が少ないことが挙げられる。また、情報サイトとしての役割を考えると、解決したい疑問について質問を行うことでの確かな回答が得られ、その回答に対してさらに多くの回答が投稿されるといったネットワーク外部性を獲得することによって、様々な情報が集まるようなサイトになることが望まれる。

こうした役割を果たすためにはどのような利用をされているのか、また今後どのようにサイト利用を活性化すべきかについての方策が必要となる。そこで、本論文では、一つの質問に対する回答数を上記課題解決のための指標として定義し、分析を通じて本サービスの理想像に近づくようなサービス拡大アイデアを検討する。

3.2 サービス拡大のアイデア

サービス拡大の目的として、質問ごとの回答数を増やすことで、ユーザ同士のつながりを深めたいという課題を解決することを挙げる。

前節の基礎集計をもとに著者間のディスカッションを経て、本論文では、サービス拡大のアイデア創出のために以下の3つの分析の方向性を考えた。

- (1) 過去の回答数が伸びている質問の中で入力となる質問に近いものをリコメンド
- (2) 質問文を入力として回答数の予測
- (3) 回答数の多い過去の質問の分析

そこで、(1)については、内容が近い過去の質問を探すためにクラスタリングを実施し、大量の質問の特徴づけを行う。また(2)については、回答数を予測するために因果関係を分析する。そして予測される回答数と実際の回答数の比較を行う。これらの結果を元に(3)について考察する。

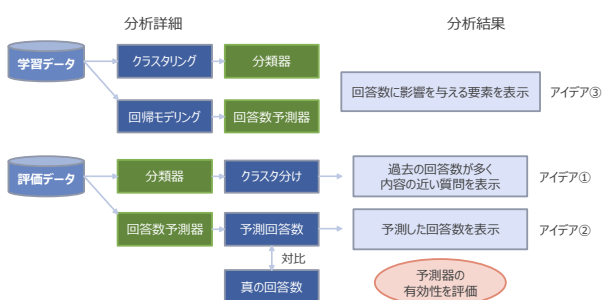


図 1 分析フレームワーク

4. 分析と結果

以下では、前節のアイデアのうち(1)と(2)に関する分析を行う。次節において、(1)と(2)についての分析と結果からの考察を行い、さらにこれらの分析結果を踏まえて

(3)についてのアイデアをまとめる。なお、分析にあたっては回答数が20件以上の質問からランダムに10,000件をサンプリングしている。

4.1 アイデア(1): クラスタの作成

アイデア(1)においては、過去の質問の中で回答数が多く、かつ、入力となる質問と内容が近い質問をレコメンドするための分析を行う。

まず、質問文の文章に含まれる単語に関してエンコーディングを行い、その結果を用いたクラスタリングを行う。エンコーディングは512次元のBERT (Bidirectional Encoder Representations from Transformers) [12]を用いる。

ただし、BERTは多くの文章によって事前学習が必要であり、辞書によって単語のエンコーディング結果が異なる。BERTの事前学習について様々な研究組織から自由に使えるものが提議されているが[13]、本研究では著者の属する企業内の研究グループが事前学習したもの[14]を用いる。

そして、その結果について、質問に関する階層型クラスタ分析によるクラスタ形成を行う。凝集方法はWard法を用いており、クラスタ数はクラスタ内の凝集度とクラスタ間の乖離度を評価できるシルエット係数が最大になる場合を採用した。

結果から、49クラスタモデルを採用した。

なお、提供データにもカテゴリラベルが付与されているが、このカテゴリとは異なるクラスタリングが行われている。定性的な評価ではあるが、本研究の結果はより具体的な課題にフォーカスして細分化されている様子が見取れた。提供データのカテゴリラベルは「その他」を除くと17カテゴリであり、そのうち「子育てグッズ」が全体の1/3と大変大きいカテゴリになっている。これと比較して、本分析の結果得られたクラスタサイズは概ね均等に近く、より適切に細分化されていることが特徴といえる。表1にクラスタリング結果の例を示す。なお、表中のfeature wordsは各単語のtf-idf値の上位10までの単語を列挙したものである。また、クラスタリングの様子を図2に示す。ここではエンコーディングされた512次元のデータについて各質問を対象として主成分分析を行い、2次元平面で表現したものを示している。

表 1 クラスタリング結果の一部

クラスタリング名	質問例	サイズ	特徴語
妊娠時間連	0 人目の妊娠がわかった時、何週ごろに病院行きましたか?	197	性別、妊娠、お腹、胎動、心拍、エコー、妊娠報告、確定、報告、感じる
保育園関連	共働きの方に質問です!保育園が休園したらどのような対応をしますか? 子供をどこかに預ける、自分が仕事休んで面倒をみる等、皆さんの対応を参考までに教えてください。	307	保育園、幼稚園、預ける、入園、行く、保育、入れる、通う、働く、4 月
食事関連	0 歳 1 ヶ月の息子がいます皆さん今日の晩御飯なんですか? 参考にさせてください ほぼ取り分けであげています	255	食べる、離乳食、あげる、作る、1 歳、ベビーフード、ご飯、パン、始める、朝ごはん
グッズ関連	チャイルドシートとベビーカーのおすすめを教えてください! こんなの買ったらいいよなどありますか? しばらくベビーカーはなくてもいいですよ?	133	抱っこ紐、ベビーカー、チャイルドシート、買う、オススメ、抱っこ、保育園、購入、自転車、持つ
母乳関連	生後 0 ヶ月半の完母です!母乳どのくらい飲めてれば良いですかね?	224	飲む、母乳、ミルク、完ミ、授乳、あげる、生後、完母、卒乳、おっぱい
離婚関連	旦那さん側から離婚を切り出されて、離婚された方いらっしゃいますか?	90	離婚、再婚、旦那さん、養育費、旦那、方位、理由、子供、調停、元カノ
睡眠関連	夜勤の旦那さんがいる方、旦那さん何時まで寝ますか?ギリギリまで起きないですか?	107	起きる、寝る、昼寝、旦那さん、旦那、皆さん、眠れる、あげる、起こす、休み

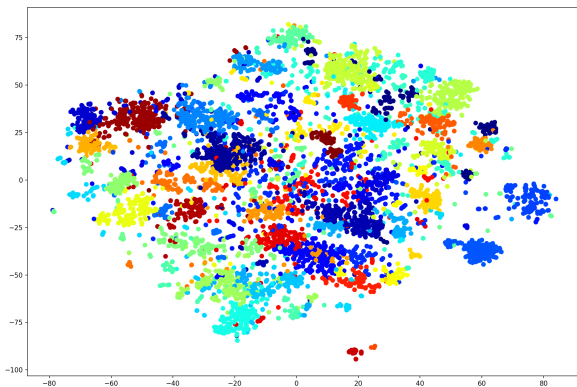


図 2 質問のクラスタリング

各クラスタの質問文について目視で確認したところ、多くは回答内容が似ているような質問がうまく選ばれているように見受けられる一方で、複数の質問要素が含まれるような質問については全く異なる質問が選択されているような場合もあった。

次に、クラスタについて考察する。得られたクラスタの特徴についてまとめる。サイズが大きかったクラスタを見たところ、話題が明確であり、「保育園、幼稚園に関する話題」「寝かしつけに関する話題」「妊娠に関する話題」「離婚に関する話題」「離乳食に関する話題」がトップ 5 であった。これらは経験者のコメントが役立つことが多いと思われることと、ライフイベントや最初の子供の時に判断に困ることが多い話題であり、コミュニティにおいても多くの人が情報発信をしている様子が見られる。

また、図 2 を見ると、2 次元平面において重なりはあるもののクラスタごとの質問の座標がおおよそ近い位置にあることが見て取れる。このことは、各クラスタの特徴語がほかの単語に比べて十分異なる特徴量を持っており、クラスタ形成に関係していることを表しているといえる。

各クラスタの tf-idf の上位 10 位までの単語が各クラスタ内の単語数に占める割合は、最小で 0.16%、最大で 19.55%、平均が 3.67% であり、単語数が多いことから考えると比較的トピックとなる単語が特定のクラスタに集まっている様子が見られる。

また、クラスタごとの返信数の中央値を求め、回答が伸びやすいカテゴリやクラスタを調べた。その結果、平均

的にはクラスタの中央値が 25 程度である中で、ママリサービスのサービスに関する話題のカテゴリが中央値 262 と非常に大きかった。このほかにもクラスタサイズが小さい (14 の質問が該当) クラスタの返信数が多かったが見受けられたが、このクラスタは大きな特徴はなく、分析に使用した特徴量にも明確な傾向はなかった。その他のクラスタは返信数の中央値に大きな差はなかった。

以上のように一部解釈が難しいクラスタがあるものの、全体的には質問に付与されているラベルよりも詳細な質問の分類ができたものと考えられる。

4.2 アイデア (2) : 回答数予測

次に、アイデア (2) のために、質問文を入力として回答数を予測する。

サンプリングした 10000 件についてホールドアウト法により 8,000 件を学習用に、残る 2,000 件を検証用とした。説明変数は前節で得られた BERT による 512 次元のベクトルであり、目的変数は各質問の回答数である。比較したモデルは、

- 線形回帰
- リッジ回帰
- ランダムフォレスト
- 勾配ブースティング
- 多層パーセプトロン

である。精度評価としては MAE (mean absolute error) を用いた。結果を表 2 にまとめる。

表 2 モデルの精度比較

モデル	MAE
線形回帰	4.385
リッジ回帰	4.186
ランダムフォレスト	4.340
勾配ブースティング	4.297
多層パーセプトロン	5.196

この結果から、正則化項考慮したリッジ回帰が最良の結果となった。誤算の分布を確認したところリッジ回帰と勾配ブースティングが似た形をしているが、勾配ブースティングの方が、誤差 5 あたりの頻度が大きくなっている。

また、様々な回帰モデルを用いて検証を行ったり、他にも入力としてユーザ属性情報の追加も検証を行ったが、本結果よりも精度の高い回帰モデルの作成には至らなかった。

説明変数が512次元と非常に高次元であることから、リッジ回帰における係数に大きな差はなかった。そこで、ランダムフォレストモデルによる変数重要度から、分岐に影響を与える上位の次元を抽出した。ランダムフォレストの変数重要度の平均は0.002であった。そこで上位である変数重要度が0.01を超える5つの次元に着目した。

そして、その次元の値が大きいもしくは小さい質問内容を確認したところ、

- (1) 値が大きい質問では全体に文字数が多く、相手に回答を求めるような表現が含まれていることが多い
- (2) 値が大きい質問は子供の成育内容を含んでいることが多く、逆に小さい質問では子供に直接関係しない質問が多い

という傾向があった。

ただし、全体で共通するような明確な特徴があったわけではなく、質問文の内容自体が回答数に必ずしも大きな影響を与えているわけではないということは考察される。

ただし、この予測結果を質問に際して参考として表示することで、回答を得やすい質問を検知することにつながり、コミュニティの活性化につながる質問抽出ができた可能性は示唆できると考える。

4.3 アイデア(3)：回答数が多い質問に関する分析

本節では、アイデア(3)のために、前節までに得られた知見を元に、回答数を増やす要因について考察する。

そのために、まず、質問投稿後の質問者のコミュニティ内での行動が回答数にどのように影響を与えているかを分析する。

図3は、質問に対する回答のうち、各質問者自身のコメントの割合の様子を示している。図3の各点は、1つの質問を示し、横軸は各質問において質問者がコメントの投稿を行っている割合を求めたものである。すなわち、0.0であれば、質問者は質問投稿以降、一度もコメントしておらず、0.5は質問に付いているコメントの半数が質問者によるものであり、1.0は質問に付いているコメントが全て質問者によるものである。なお、縦軸は回答数である。

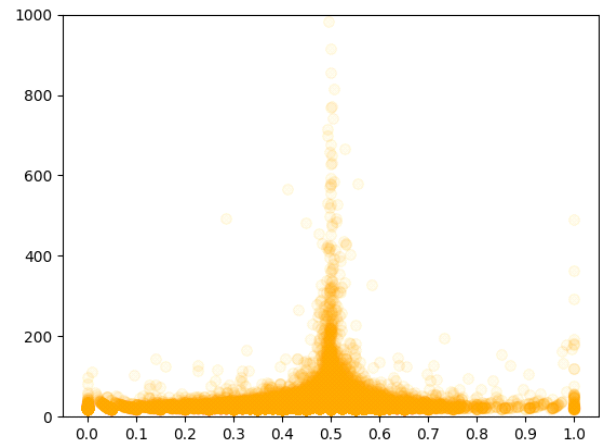


図3 返信率と回答数

これを見ると0.5のところで回答数が突出しており、半数が質問者である場合、すなわちやり取りを丁寧にしていると回答数が伸びやすい傾向にあることが示唆される。横軸について0.5から離れるほど全体として回答数が伸びない傾向にあることがわかる。

次に、質問投稿後の質問者の行動が回答数にどのような影響を与えているか分析するために、返信速度、すなわち質問者が回答に返信するまでの平均時間について考察する。

図4は質問投稿後の返信の速さと回答数についての関係を表したグラフである。

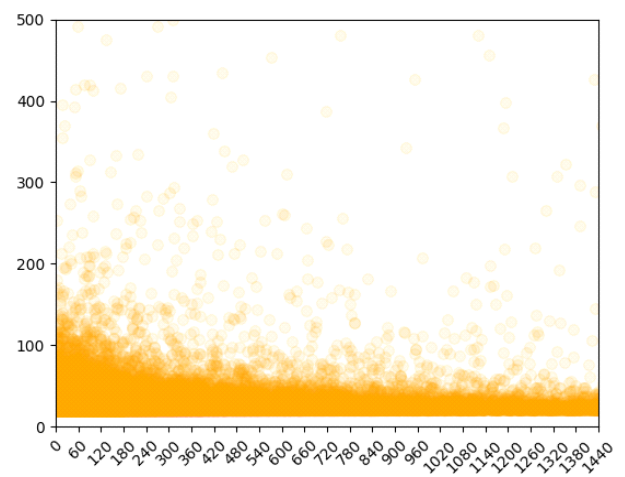


図4 返信速度と回答数(横軸は分)

この図より、回答数の上振れはあるものの、グラフの密度の高い部分を見ると、返信までの時間が長くなるとともに回答数は低位に推移しており、質問投稿後にできる限り早く返信することが、回答数のアップにつながることを示唆される。

このように、返信率や返信の速さが回答数の増加に寄与している様子が分かる。

以上より、対象としたQ&Aサイトにおけるコミュニティ活性化においては、(1)適切な質問クラスタに分割す

ることによって、類似した質問の参照をやすくするなど工夫をしてコミュニティ内の流動性を上げること、(2) 回答数が伸びそうな質問をあらかじめ検知することによって、その内容を学習したり、またその質問への流入を増やすなどしてコミュニティ利用促進をすること、(3) 質問者と回答者のやり取りを密にするような施策を講ずること、といったアイデアが考えられる。

5. まとめと今後の課題

本論文では、妊婦・乳幼児を持つ母親向けサイトに Q&A データを用い、コミュニティ活性化を促進する戦略立案のための分析を行った。

そのための 3 つのアイデアを示し、それぞれのアイデアに対する分析を行った。各分析の結果から、以下のような考察を

まず、アイデア (1)、すなわち過去の回答数が伸びている質問の中で入力となる質問に近いものをレコメンドするための分析では、質問内容が近く過去に回答数が伸びている質問を提示するレコメンドモデルを作成した。その結果、クラスタリングでは従来のカテゴリと比較して、より課題にフォーカスしたクラスタ形成をすることができた。

次に、アイデア (2)、質問文を入力として回答数を予測に対して、質問文を入力として回答数を予測する回帰モデルを作成した。回答数に影響すると考えられる質問文の部分的な定量的・定性的な特徴が抽出できたものの、その結果からは質問内容自体が回答数に大きな影響を与えていないということが分かった。

最後に、アイデア (3)、回答数が多い過去の質問を評価するために、質問投稿後の質問者の行動が回答数にどのような影響を与えているか分析を行った。その結果、質問者の返信率と返信速度が回答数を伸ばす大きな要因になっていると分かった。

以上より、本研究で取り組んだ分析によって得られたクラスタに関して、ユーザからの興味関心が高いものが抽出されたと考えられる。そして、新たなカテゴリとして設定することによって検索の利便性向上や、ユーザのより詳細な意見交換が見込める。また、一定の期間単位で注目を集めているトピックを動的にカテゴリとして設定することによって、人気な話題に対する投稿のさらなる活性化が見込め、ユーザが注目しやすい内容を抽出することが可能なため、公式投稿の内容検討に使用できる。

質問投稿後の質問者の行動における返信率と返信速度が回答数に影響を与えるという結果は有効活用できるのではないかと考えられる。ホームへの掲載や質問者に対する Tips や公式投稿として共有するなど、ユーザに周知を行うことによって投稿の活性化につながる可能性が示唆される。

今後の課題としては、レコメンドモデルの定量的評価方法について検討することや、今回の分析で得られたクラスタリングの活用方法検討等が挙げられる。また、今回は利用者の個人・子供の月齢などは考慮しなかったため、こうした個人属性を質問者と回答者間に考慮することで、より適切なマッチングを実現することが期待される。

参考文献

- [1] 東京都北区, 「子供子育て支援計画」, (2020) <https://www.city.kita.tokyo.jp/k-mirai/documents/kokokeikaku2020.pdf>
- [2] ヤフー・データソリューション, 「ビッグデータから見るストレス・悩み」, 202. <https://ds.yahoo.co.jp/report/20210520.html>
- [3] 島田達朗, 櫻井彰人: コミュニティサイトにおける共感を求める質問の認識, 知能と情報 (日本知能情報ファジィ学会誌), Vol.29, No.4, pp.611-618 (2017).
- [4] 島田達朗, 櫻井彰人: 日本最大級の女性向けコミュニティサイトにおける違反コンテンツ検閲システムの構築—安心して使い続けてもらうためのコミュニティを目指して—, デジタルプラクティス, Vol.8, No.1, pp.92-96 (2016).
- [5] 島田達朗, 櫻井彰人: 日本最大級の女性向けコミュニティサイトにおける出産予定日予測システムの構築, 情報システム学会誌, Vol.14, No.1, pp.1-12 (2018).
- [6] Hew, K.F. and Hara, N.: Knowledge sharing in online environments: A qualitative case study, *Journal of the Association for Information Science and Technology*, Vol.58, No.14, pp.2310-2324 (2007).
- [7] 尹浩, 山本景子, 倉本到, 辻野嘉宏: バーチャルエージェントに対する好意を用いた Q&A ウェブサイトにおける回答意欲向上効果, ヒューマンインタフェース学会論文誌, Vol.20, No.4, pp.413-416 (2018).
- [8] 尹浩, 山本景子, 倉本到, 辻野嘉宏: Q&A システムにおける回答意欲を向上させるためのバーチャルエージェントデザインツール, 情報処理学会論文誌, Vol.61, No.3, pp.618-627, 2020
- [9] Adamic, L.A. Zhang, J., Bakshy, E. and Ackerman, M.S.: Knowledge sharing and Yahoo Answers: Everyone knows something. *Proceedings of the 17th international conference on World Wide Web*, pp.665-674 2008.
- [10] 甲谷優, 川島晴美, 藤村考, 奥雅博: QA サイトにおける質問応答グラフの成長パターン分析, 日本データベース学会論文誌, Vol.7, No.3, pp.61-66 (2008).
- [11] Zhongbao, K. and Changshui, Z.: Reply networks on a bulletin board system. *Physical Review E*, Vol.67, No 3, p.36117 (2003).
- [12] Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V. and Zettlemoyer, L.: BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension, *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp.7871-7880 (2020).
- [13] HalloRusk: フリーで使える日本語の主な大規模言語モデルまとめ, Zenn エンジニアのための情報共有コミュニティ (2021). <https://zenn.dev/hellorusk/articles/ddee520a5e4318>
- [14] 和知徳磨: 汎用言語モデル「BERT」のビジネス実用化技術に迫る, NTT 技術ジャーナル, Vol.33, No.3, pp.76-81 (2021). <https://journal.ntt.co.jp/article/11275>