

学位論文 博士（工学）

機械学習を用いたコミュニティ QA サイト  
における品質向上に関する研究

2018 年度

慶應義塾大学大学院理工学研究科

島田 達朗

---

# 目次

<b>第1章 序論</b>	<b>1</b>
1.1 背景と目的	1
1.2 不適切な質問の抽出	2
1.3 回答率が低いと推察される質問の抽出	3
1.4 ユーザー属性の抽出	4
1.5 論文の構成	4
<b>第2章 関連研究</b>	<b>5</b>
2.1 コミュニティQAサイトの概要	5
2.2 質問ルーティング	7
2.3 質問の分類, タグ等の補足情報の付与	8
2.3.1 手作業による質問分類の先行研究	8
2.3.2 自然言語処理と機械学習を用いた質問分類の先行研究	9
2.3.3 質問文からの補足情報取得に関する先行研究	10
2.4 良質な回答に関する研究	11
2.5 ユーザーのプロファイル分類や解約予測に関する研究	11
<b>第3章 コミュニティQAサイトにおける質問テキストのみを用いた不適切な投稿の検閲</b>	<b>13</b>
3.1 はじめに	13
3.2 大規模コミュニティQAサイトにおける課題	15
3.3 質問テキストのみを用いた, 不適切な投稿の検閲手法	16
3.3.1 機械学習を用いた分類	17
3.3.2 提案手法	18
3.4 実験	20
3.4.1 データセット	20
3.4.2 実験手法	20
3.4.3 評価指標	20
3.4.4 目標とすべき指標数値の定義	21
3.4.5 実験結果	22
3.4.6 提案手法の有効性についての実システムでの検証	26

---

3.5	まとめと今後の課題	27
<b>第4章</b>	<b>コミュニティQA サイトにおける共感を求める質問の認識</b>	<b>29</b>
4.1	はじめに	29
4.2	共感の定義	31
4.3	質問分類タスクの定義	33
4.3.1	分類の定義	33
4.3.2	各分類の具体的な例	34
4.3.3	質問分類タスクにおける議論	34
4.4	分類手法	35
4.4.1	特徴語句による分類	35
4.4.2	機械学習を用いた分類	38
4.4.3	word2vec 用いた質問文拡張	38
4.5	実験	39
4.5.1	データセット	39
4.5.2	実験手法	40
4.5.3	目標とすべき指標数値の定義	41
4.5.4	実験結果	41
4.6	おわりに	48
<b>第5章</b>	<b>コミュニティQA サイトにおける質問テキストのみを用いたユーザー属性の抽出</b>	<b>49</b>
5.1	はじめに	49
5.2	ユーザー属性の活用と取得に関する課題	50
5.2.1	ユーザー属性の活用	50
5.2.2	ユーザー属性の取得に関する、コミュニティQA サイトの課題	50
5.3	質問文の特徴分析	52
5.3.1	目的	52
5.3.2	対象とするデータ	52
5.3.3	問題設定	52
5.3.4	特徴の分析方法	54

---

5.3.5	分析結果 . . . . .	54
5.4	実システムへの導入のための判別手法の提案 . . . . .	56
5.4.1	目標とすべき指標数値の定義 . . . . .	56
5.4.2	提案手法の概要 . . . . .	56
5.4.3	機械学習を用いた分類 . . . . .	58
5.5	実験 . . . . .	59
5.5.1	データセット . . . . .	59
5.5.2	実験手法 . . . . .	60
5.5.3	実験結果 . . . . .	62
5.6	終わりに . . . . .	65
<b>第 6 章</b>	<b>結論</b>	<b>67</b>
6.1	まとめ . . . . .	67
6.2	今後の課題 . . . . .	68
	<b>参考文献</b>	<b>70</b>
	<b>謝辞</b>	<b>80</b>

---

## 目 次

2.1	コミュニティQA サイトにおけるユーザーの役割毎のアクション フローチャート . . . . .	6
3.1	辞書拡充のステップ . . . . .	18
3.2	機械学習と目視での投稿内容確認を組み合わせた提案手法 . . . . .	19
3.3	投稿のクラス分類の予測結果 (Prediction Outcome) と実際のク ラス (Actual Value) との混同行列 . . . . .	21
3.4	ipadic を利用して MeCab による形態素解析を実行した結果 . . . . .	23
3.5	mecab-ipadic-NEologd を利用して MeCab による形態素解析を実 行した結果 . . . . .	24
4.1	ママリの画面図 . . . . .	32
4.2	教師データ中の共感以外を求める質問サンプル数を変化させたとき の, テストデータにおける recall . . . . .	43
4.3	MeCab のシステム辞書における評価値の比較 . . . . .	44
4.4	各分類手法における recall (LR: Logistic Regression, GNB: Gaus- sian Naive Bayes, KN: k-Nearest Neighbors, RF: Random Forest, AB: AdaBoost, DT: Decision Tree) . . . . .	45
4.5	各分類手法における precision (LR: Logistic Regression, GNB: Gaus- sian Naive Bayes, KN: k-Nearest Neighbors, RF: Random Forest, AB: AdaBoost, DT: Decision Tree) . . . . .	45
4.6	各分類手法における F1 (LR: Logistic Regression, GNB: Gaus- sian Naive Bayes, KN: k-Nearest Neighbors, RF: Random Forest, AB:AdaBoost, DT:Decision Tree) . . . . .	46
5.1	妊娠週数に合わせたコンテンツやレコメンデーションの例 . . . . .	51
5.2	< 数値 > [w] 週 という特徴語句を持つ質問から出産日を起算日と した妊娠経過週を抜き出し予測した正解率と対象ユーザー率 (プ ロット上部の数値は出産日を起算日とした妊娠経過週) . . . . .	55
5.3	< 数値 > [w] 週 と < 数値 > [w] 週 < 数値 > [d] 日 それぞれの特 特徴語句を持つ質問から出産日を起算日とした妊娠経過週を抜き出 し予測した正解率と対象ユーザー率 (プロット上の数値は出産日 を起算日とした妊娠経過週) . . . . .	57

---

5.4	提案手法のフローチャート . . . . .	58
5.5	probability の閾値の導出方法 . . . . .	61
5.6	学習器と単語ベクトルの生成方法の組み合わせによる，出産日を 起算日とした妊娠経過週が 40 週時点の正解率と対象ユーザー率の 比較 (RF: RandomForest, W: word2vec) . . . . .	63
5.7	提案手法と < 数値 > [w 週]< 数値 > [d 日] の特徴語句を持つ質 問から出産日を起算日とした妊娠経過週を抜き出し予測した正解 率と対象ユーザー率の比較 . . . . .	64

---

## 表 目 次

3.1	学習器毎の各指標の比較 . . . . .	22
3.2	システム辞書毎の各指標の比較 . . . . .	22
3.3	辞書拡充前後の各指標の比較 . . . . .	24
3.4	サンプリングしたデータに対する判別結果 . . . . .	26
3.5	サンプリングしたデータに対する, 閾値調整後の判別結果 . . . . .	26
4.1	共感を求める質問と共感以外を求める質問それぞれ 300 件に対する未回答の数 . . . . .	30
4.2	共感を求める質問と共感以外を求める質問それぞれ 300 件に対する回答者数 (人) と回答者属性の割合 (%) . . . . .	31
4.3	特徴語句リスト . . . . .	36
4.3	特徴語句リスト (続き) . . . . .	37
4.4	4 段階の評価によるラベル付け . . . . .	40
4.5	教師データ中のクラス毎のサンプル数比を 1:1 にした際の, テストデータにおける共感を求める質問の分類結果 ( <b>B</b> : ベースライン, <b>M</b> -学習器 (システム辞書): 4.4.2 の手法, <b>W</b> (コーパス): word2vec 用いた質問文拡張) . . . . .	42
4.6	「つわり」の拡張単語 . . . . .	47

# 第1章 序論

本章では本研究における背景と目的を述べ、それを達成する上での3つの課題について触れる。また、それらの課題の解決方法について本論文における解決方法を対応づけて述べる。

## 1.1 背景と目的

コミュニティQAサイトの主な運営目標は、投稿された質問に最短時間で最も適切な回答を提供することである。しかし、ユーザー数の増加と毎日投稿される多様で膨大な数の新しい質問と回答により、コミュニティQAサイトは、日に日にサービス（質問、回答、ユーザーの最適な出会い）の提供が難しくなる。質問、回答、ユーザーの出会いを支援する、サポートシステムがなければ、結果として質問者は、満足な回答を許容可能な時間内に得ることはできず、コミュニティQAサイトは、主な運営目標を達成することが困難となる。

上記の問題に対して、満足のいく答えを提供する可能性が、最も高い潜在的な回答者に質問を表示、または回答を促す手法として、コミュニティQAの研究分野においては「質問ルーティング」という考え方が研究されてきた [22]。質問ルーティングの実現には、以下の3つのステップが必要である。

1. ユーザーの分類、ユーザーの能力・知識・関心等の属性情報の認識
2. 質問の分類、タグ等の補足情報の付与
3. 上記の情報を利用し、質問とユーザーのマッチング

3つ目のステップを実行し、質問ルーティングを実現するには、ステップ1もしくはステップ2（またはその両方）のステップの中で適切な情報の認識、抽出、付与を行う必要がある。しかし近年、スマートデバイスの普及によるサービス提供方法の変化により、コミュニティQAサイトによってはステップ1が利用しづらい状況になっている。その理由は、スマートデバイスに保持したIDのみで多くのサービスの利用が可能になったからである。これは、コミュニティQAサイ



## 1.2. 不適切な質問の抽出

---

トの利用時に email やパスワードの登録不要で、サービスの利用が開始できることを意味する。

こうして登録の敷居が下がったことにより、コミュニティQA サイトにおいて、入退会を繰り返すことで今までよりも容易にユーザー履歴を消すことが可能になった。従って、ユーザー履歴が十分に蓄積された状態でユーザーの分類、ユーザーの能力・知識・関心等の属性情報の認識、抽出、付与、活用を行うことよりも、質問テキストのみでユーザー属性を認識、抽出、付与、活用を行う必要がある。

本論文ではそうした背景の中で、重要性と課題解決の可能性を踏まえて以下の3つの課題に対して解決手法を提案することで、質問テキストのみでより良い質問ルーティングを実現することを目的とする。

1. 「不適切な質問」の抽出
2. 「回答率が低いと推察される質問」の抽出
3. 「ユーザー属性」の抽出

本論文においては、これらの課題を解決するべく、対象とするコミュニティQA サイトとして月間数百万の投稿が存在する大規模コミュニティQA サイト「ママリ [1]」を想定し、研究に取り組んだ。本論文における「商用データ」はママリのデータを指す。以下、本論文において取り組む課題と解決方法を対応づけて述べる。

## 1.2 不適切な質問の抽出

質問ルーティングの目標の1つとして、回答の質の向上は当然重要であるが、まず回答されないことにはコミュニティQA サイトとして成り立たないため、回答率の向上が非常に重要である。

活発なコミュニティQA サイトにおいては、本来のコミュニティQA サイトの利用用途にそぐわない投稿を行うユーザーが現れる（「荒らし」と呼ぶ）。過去の投稿が閲覧可能なコミュニティQA サイトにとって、ガイドラインに違反したコンテンツの投稿は「不適切な質問」であり、コミュニティQA サイトの目的を阻害してしまう。なぜなら、ユーザーは少しでもコミュニティが荒れた雰囲気を感じると投稿を控えるようになるからである。つまり、質問ルーティングにおいてコミュニティを荒らす投稿は誰にも送ってはいけない質問である。

### 1.3. 回答率が低いと推察される質問の抽出

---

「不適切な質問」を削除するためには、人による目視が必要である。しかしながら、人による目視での投稿の検閲はコストが高いという課題がある。そこで人による目視での確認の手法を残しつつ、可能な限り機械学習を用いて「適切な質問（通常の質問）」か「不適切な質問」かを分類することにより、人的コストを抑える形でこの問題を解決できると考えた。

大規模商用データを用いて、質問の不適切さを評価した後に、その評価値を元に、3個のクラス（不適切な投稿、人による判断が必要な投稿、適切な投稿）に分類する方法を提案し、人による介入を1/10以下に低減した上で、precisionを大きく低下させることなく、不適切な質問をほぼ100%除去する手法を提案する。この手法を用いることで課題1の解決を目指す。

### 1.3 回答率が低いと推察される質問の抽出

コミュニティQAサイトには、ユーザーがコミュニティに対して質問を投稿した後、十分な時間が経過しても回答が与えられず、未回答のままになってしまう質問が存在している。しかし、ユーザーは回答を求めてコミュニティに対して質問を行うため、回答が与えられないことには、ユーザーの望みを叶えることは困難である。

そこで、未回答の質問を調査したところ、その中には疑問等に対する答えを求めるのではなく、むしろ質問の具体的な答えそのものよりも、自分の思いへの共感を求める質問が存在していることが明らかとなった。自分の思いへの共感を求める質問に対して、商用データにおいて統計を取ると、共感を求める質問に対して回答する人は、そうでない質問に対して回答する人より有意に少ないという課題があることがわかった。さらに、共感を求める質問は全体の約2割を占めることがわかった。従って、共感を求める質問に対しても適切な回答を与えることができれば、ユーザーのコミュニティQAサイトに対する満足度が向上すると考えた。

共感を求める質問に対して、回答率の高いユーザー層に回答してもらえよう、質問をそうしたユーザー層に振り分けることによって、共感を求める質問であっても、回答される可能性が高まるものと考えた。そこで本論文においては、機械学習を用いて「共感を求める質問」か「共感以外を求める質問」かを分類する手法を提案する。この手法を用いることで課題2の解決を目指す。

### 1.4 ユーザー属性の抽出

多くのオンラインサイトにはユーザー属性の登録機能があり、コミュニティQAサイトでも自分のユーザー属性を入力できるサイトが多い。ユーザーとしては、ユーザー属性から質問に対しての必要な情報を想起させやすくすることや、回答者としてのスタンスを明示することに利用ができるため、能動的にユーザー属性の入力を行うユーザーも一定数存在している。サービス提供側としてはユーザーが入力をしたユーザー属性を用いることにより、ユーザー自身が求める情報のマッチングが可能になるため、より良い質問ルーティングを実現することも可能である。

しかし、ユーザー自身がユーザー属性を入力する場合、入力に手間がかかり、ユーザーに負担がかかってしまう。また、スマートデバイスに保持したIDのみで多くのサービスの利用が可能になったために、入退会が容易になった。この影響により、今までよりも、登録したユーザー属性を削除することが容易になったため、ユーザー属性を行動履歴から得られないという課題が発生している。

そこで、より良い質問ルーティングを実現するために、ユーザー属性の入力を必要とせずに、ユーザーが投稿した質問から機械学習を用いて、ユーザー属性を予測する手法を提案する。この手法を用いることで、課題3の解決を目指す。

### 1.5 論文の構成

本論文の構成は以下のとおりである。第2章では、本研究の関連研究として、コミュニティQAサイトにおける質問分類や、分析の概要と本研究で用いる機械学習の手法について述べる。第3章では、適切な質問か不適切な質問かを分類するための、投稿検閲システムの構築および不適切な投稿の検閲フローについて具体的に述べる。第4章では、回答率の低いと推測される質問の抽出を行うために、共感を求める質問とそれ以外の質問の定義を述べ、その上で機械学習を用いて、共感を求める質問と共感以外を求める質問かを分類する手法について述べる。第5章では、ユーザーが投稿した質問から、機械学習を用いてユーザー属性を予測する方法について述べる。最後に第6章では、本論文のまとめと、今後の課題および展望について述べる。

## 第2章 関連研究

本章では、コミュニティQAサイトにおける基本事項について述べた後に、コミュニティQAサイトでこれまで取り組まれてきた関連研究について述べ、それらの手法、評価指標についても言及する。

### 2.1 コミュニティQAサイトの概要

「QAサイト」と呼ばれるサイトの中には様々なWebサイトが含まれる。例えば、医療や法律などのドメインについて特定の知識を持ったエキスパートがユーザーからの質問に回答するサイトや、人間ではなく機械が自動的に回答するサイトなどである。本論文で扱うコミュニティQAサイトは、ユーザー同士による質問と回答を特徴とするコミュニティQAサイトを指す。コミュニティQAサイトにおけるユーザーの役割毎のアクションフローチャートを図2.1に示す。

図2.1の質問者は、質問を作成し、コミュニティに対して投稿するユーザーである。投稿の際にはカテゴリやタグといった補助情報をつけて投稿するのが一般的である。その質問に対して回答を行うユーザーが回答者である。コミュニティ内の不特定多数のユーザーは自由に質問を閲覧することができ、また1つの質問に対して複数のユーザーが、回答することができる。それらの回答に対して質問者が満足した場合、質問に対する回答を終了させることができる（「閉じる」と呼ぶ）。これは、自分が投稿した質問に対して、これ以上の質問の回答を受け付けることを、止めることである。また、コミュニティQAサイトによっては、この際に最も有用であると判断した回答を選択し、感謝していることを示すマークを付ける。こうしてやり取りされた質問や回答はQA Databaseに蓄積される。そうして蓄積されたQA Databaseに存在している過去の質問や回答に対して検索を行い、自分に必要な情報を探すユーザーも存在しており、図2.1内では閲覧者として表現している。このようなコミュニティQAサイトは現在のインターネットにおける重要なリソースになっており、従来の情報検索システムでは得ることの難しい主観的文脈を含んだ質問と回答、そして会話のような体験を得ることができる。

このようなコミュニティQAサイトの主な運営目標は、投稿された質問に最短

## 2.1. コミュニティQAサイトの概要

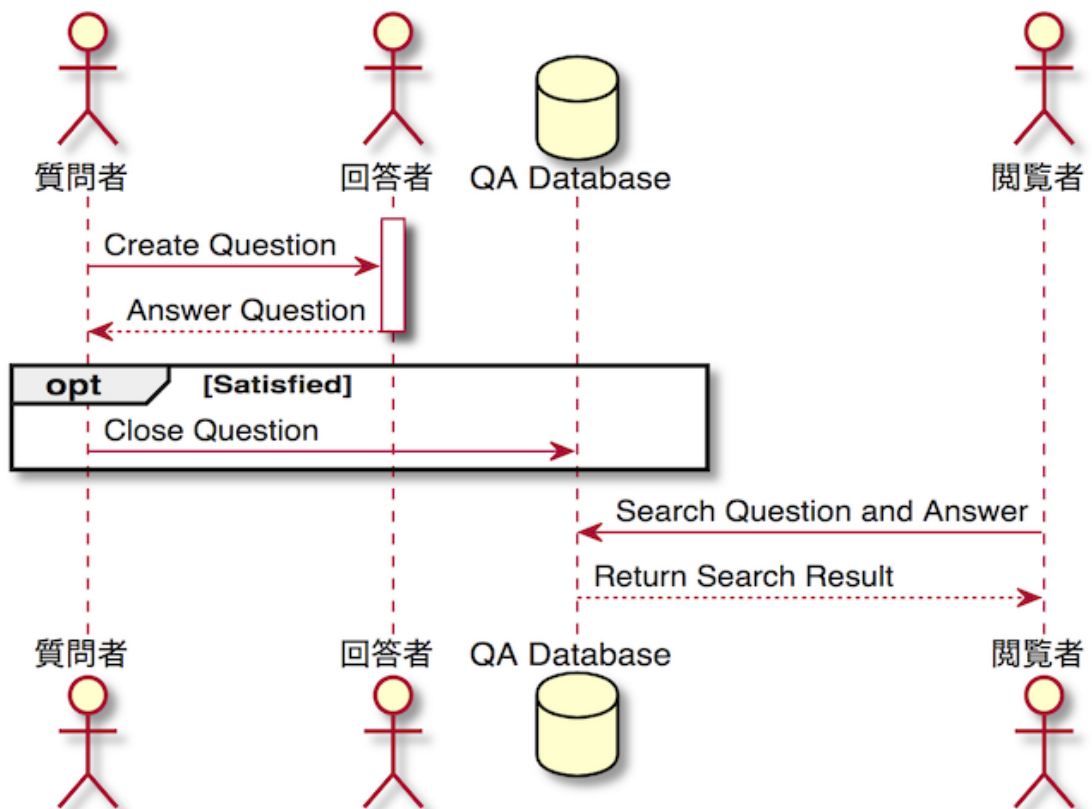


図 2.1: コミュニティQAサイトにおけるユーザーの役割毎のアクションフローチャート

時間で最も適切な回答を提供することである。しかし、ユーザー数の増加と毎日投稿される多様で膨大な数の新しい質問とそれに対応した回答により、コミュニティQA サイトは、日に日に質問、回答、ユーザーの最適な出会いを提供するのが難しくなる。質問、回答、ユーザーの最適な出会いを支援するシステムのサポートがなければ、結果として、質問者は許容可能な時間内に満足のいく回答を得ることは困難である。従って、コミュニティQA サイトの主な運営目標を達成することが困難となる。この問題を解消するため、質問と回答のプロセスをサポートすることを目的として、多くのアプローチが提案され、これまで質問、回答、ユーザーに関するケーススタディやデータ分析が数多く行われてきた [21]。

## 2.2 質問ルーティング

コミュニティQA サイトの主な運営目標は、投稿された質問に最短時間で最も適切な回答を提供することであるが、回答がなされない質問が存在するという問題や、質問者が満足する回答が得られないという問題が存在している。このような問題の発生は、ユーザーと質問数の増加に比例して多くなる。

Zhou ら [23] はメジャーなコミュニティQA サイトの1つである Yahoo Answers [3] において Yahoo Answers の 26 個の第 1 階層の各カテゴリから、ランダムに 140 件の質問をサンプリングし、合計 3,640 件の質問について、質問が解決したかどうかのステータスを時系列で確認した。その結果、これらの 3,640 件の質問は、1 日後、434 件 (11.95 %) のみが解決され、2 日後、726 件 (19.95 %) が解決されたことがわかった。この調査結果は、投稿された質問の多くが、短時間で解決されることは、困難であることを示している。同様の問題はその他の研究 [24, 25] でも確認されており、様々なコミュニティQA サイトにおいて大きな課題の1つとなっている。

質問ルーティングは、満足のいく答えを提供する可能性が、最も高い潜在的な回答者に質問を表示、または回答を促すことによって、この問題の解決を図る手法である。質問ルーティングの実現には、以下の3つのステップが必要である。

1. ユーザーの分類, ユーザーの能力・知識・関心等の属性情報の認識
2. 質問の分類, タグ等の補足情報の付与
3. 上記の情報を利用した, 質問とユーザーのマッチング

3つのステップを実行し、適切なマッチングを実現するにはステップ1, ステップ2で適切な情報を認識, 抽出, 付与を行う必要がある。本論文では質問の分類

と情報抽出を主に扱うため, 2.3節で「質問の分類, タグ等の補足情報の付与」についての関連研究を述べる. また, 2.4節では「良質な回答」についての関連研究を述べる. 最後に, 2.5節で「ユーザーのプロファイル分類や解約予測」についての関連研究を述べる.

## 2.3 質問の分類, タグ等の補足情報の付与

### 2.3.1 手作業による質問分類の先行研究

質問の分類に関する研究では, アンケートを用いて収集した質問を分析し, 手作業で様々な質問タイプに分類する方法が提案されている.

三浦ら [28] は知識共有コミュニティにおける参加動機を分析する過程で, 投稿される質問には「正解あり」と「正解なし」の二つのタイプが存在していることを指摘した. 「正解あり」タイプの場合は, 質問者が決まった答えや正解を求める, つまり信頼性が高い知識や正しい情報を求める質問タイプである. 「正解なし」タイプの場合は, 決まった答えを求めるのではなく, 質問者の悩みに対して他の人に経験や意見を求める相談のような質問タイプである.

栗山ら [8] は Yahoo!知恵袋 [2] において, 500 件の質問を対象にして, 手作業で質問のタイプを次の3つに分けた. 1つ目はサーチエンジンや図書館のレファレンス・サービスを利用して解答を探すことが可能な, 「情報検索型」質問で, 事実, 真偽, 定義・記述, 方法・手段, 原因・理由, 効果・結果を尋ねる質問である. 2つ目は客観的な正解はなく, 特定の個人あるいは集団に対してアンケート調査を行うことで回答を得るような「社会調査型」の質問で, 助言, 意見, 施行, 推薦, 経験を尋ねる質問である. 3つ目は情報検索やアンケート調査によって客観的あるいは主観的な回答を得ることが目的ではなく, 質問者が自分の主張に対する反響・反応を求めた質問であり, こうした記述表現を持つ「非質問型」で, 主張もしくは記述としては, 何が書かれているのか分析者には理解できなかったものである. 栗山らはユーザーの質問投稿時において, 質問のトピックによるカテゴリ選択の支援だけでなく, これらの質問のタイプを提示することは利用者の支援になるという考えを示した.

渡邊ら [26] は質問者が回答に期待する内容と属性に着目し, 質問者が回答に期待する内容として事実, 根拠, 経験, 提案, 意見という5種類の質問タイプに分類している. 質問に対する唯一の正解が存在するような内容の質問では, 正解となる事実を含む回答を要求する場合と, 事実に加え, その根拠をも要求する場合とがあり, 前者の質問タイプを「事実」とし, 後者を「根拠」とした. 一方, 唯一の正解が存在しないような内容の質問では, 回答者の主観に基づく回答が要求

される。この場合、回答者自身の経験に基づく回答や、回答者による提案を含む回答などが要求されると考えられる。前者の質問タイプを「経験」、後者を「提案」とし、これらに分類されない場合の質問タイプを「意見」として定義した。

### 2.3.2 自然言語処理と機械学習を用いた質問分類の先行研究

前項のような手作業による質問の分類の関連研究は、質問の自動分類のための手法を示唆する。コミュニティQAサイトにおいて、質問や回答のテキストデータを対象として、特徴を抽出した上で、自然言語処理と機械学習を用いて、質問の分類を行っている研究には、様々な研究が存在している [9, 10, 29, 30, 31, 32, 33, 34, 35, 36, 27, 37]。

例えば、Quら [9] は、質問と回答のテキストデータを用いて、Yahoo Answers における質問を SVM [19], Naive Bayes, Maximum Entropy を用いてカテゴリ分類を行っている。Aikawaら [10] は、Yahoo!知恵袋の質問において、質問テキストから抽出した特徴量のみで、subjective な回答を求める質問か objective な回答を求める質問かの分類を、SVM と Naive Bayes を用いて行った。

Zhouら [11] は、質問や回答のテキストデータから抽出された特徴量に加えて、like (ユーザーが有益と感じたときに回答に対してされるアクション), vote (ユーザーが最も良いとされる回答に対して投票を行なうアクション), source (回答の参照元), poll and survey (投票機能を用いた質問), answer number (回答数) といった要素から、自動的に subjective な回答を求める質問かどうかの分類を行った。

質問自体が適切か、不適切かどうかを分類する研究に関しては、小林らの研究が挙げられる。小林ら [36] は、Yahoo!知恵袋のデータセットを用いて、人手で行っている不適切な投稿の削除の自動化を試みた、正例あるいは負例の片方だけに確かなコーパスが存在し、残りが不確かなコーパスであるという条件下で、SVM を用いて質問が適切か、不適切かどうかを分類することに取り組んだ。

渡邊ら [27] は、前述の質問者が回答に期待する内容と属性に着目し、質問者が回答に期待する内容として事実、根拠、経験、提案、意見という、5種類の質問タイプに分類するため、機械学習を用いた手法についても取り組んでいる。渡邊らは質問が投稿された時点で質問を分類することを目指すため、質問文から得られる特徴量だけを機械学習の素性として分類を行った。

朱ら [37] は、質問タイプを分類することによって質問タイプによるベストアンサー、または良質な回答を推定する手法を提案している。



### 2.3.3 質問文からの補足情報取得に関する先行研究

コミュニティQAサイトにおける質問の話題は様々である。従って、ユーザーへ多くの異なる話題に関する質問を、カテゴリ化せずに蓄積・表示するべきではない。そこで一般的に質問は「タグ」や「カテゴリ」といった質問の補足となる情報を質問者が選択して付与するか、システムが自動的に付与する。これらは、回答率の低い質問を判別して、判別された質問をユーザーへ表示することに使われたり、同様の関心を持つユーザーを見つけるために使用される。また、回答する際や解決された質問のアーカイブを検索するとき、閲覧者にとって必要な質問へのナビゲーションをより簡単に提供する。

最も一般的な補足情報は、階層化したカテゴリである。Yahoo AnswersやYahoo!知恵袋でもこの方法は採用されている。カテゴリ情報の付与については、通常、カテゴリ階層中の最下層のカテゴリの情報を質問に付与する。しかし多くの質問は、より多くの情報と関連する可能性がある[38]。タグに階層を持たせたコミュニティQAサイトがある一方、Stack Overflow [5]のようにタグは階層関係のない単純な構造のみを持つサイトも存在している。多様なユーザーの増加により、コミュニティQAにおいて前述のタグや最下層のカテゴリ数はサービスの成長とともに多くなっていき、その結果、経験の少ない質問者にとっては適切なタグやカテゴリの選択は難しくなる。

このような問題は、既存または新しく投稿された質問に自動的にタグやカテゴリを割り当てる方法で解決が可能である。しかし、質問投稿直後に利用可能な情報(質問のテキスト中に含まれている情報)は、通常、情報量が非常に限られているため、あらかじめ定義されたカテゴリへの分類やタグの付与は難しい。例えば、Yahoo Answersには1,000以上のカテゴリが存在しており、最適なカテゴリを選ぶのは困難である。

質問テキストという限られた情報を用い、モデルを作るアプローチとして、これまでBoW (bag-of-words) モデル[38], translation-based language モデル[40]さらに、Wikipediaから得られた意味論的知識を用いた translation-based language モデル[39], カテゴリベースの言語モデル[41]が存在している。その他nグラム, 品詞 (POS) タグ付け, 構文木[42]などを用いたアプローチがある。また, adaptive probabilistic hypergraph [43]により質問の内容からだけでなく, 質問者とそのフォロワーの質疑応答の履歴からモデルを構築するというアプローチも存在している。

学習器としてはk-Nearest Neighbors, Maximum Entropy Classifier, SVM等が用いられている。Nieら[43]は意味論的に類似した質問を判別するためのprobabilistic hypergraph学習を提案し, その後, 得られたタグ候補をさらにフィルタ

リングするヒューリスティックなアプローチを提案した。

## 2.4 良質な回答に関する研究

質問に対する回答や、回答に対して、質問者が回答の中で最も良い回答として選ぶ「ベストアンサー」についても幅広く研究が行われてきた。

石川ら [44] は Yahoo!知恵袋を対象に、人間の判定者がベストアンサーに相応しい回答を選出し、その結果と「ベストアンサー」情報とを併せて、良質な回答の特徴分析を行った。また、その結果に基づき、良質な回答の自動予測システムを機械学習によって構築した。

Kim ら [7] は Yahoo Answers において、465 件の質問とその質問者からベストアンサーが与えられた回答について手作業で分類を行い、ベストアンサーの選択理由について考察を行っている。Kim らはそれらの質問を information (特定の事実の探索や現象の理解), suggestion (助言, 推薦, 実行可能な解決法の探索), opinion (他人の意見・感じ方の調査, 議論の開始), others (先の 3 つのタイプに入らないもの) に分類した結果、それらの割合はそれぞれ information が 35%, suggestion が 23%, opinion が 39% であった。また、ベストアンサーの選択理由の分布は質問のタイプによって違いがあることを示し、opinion タイプの質問に対しては、回答者の態度や感情面での支援など社会的・感情的な理由である socio-emotional の要素がベストアンサーの選択理由として大きな要素であるとしている。

## 2.5 ユーザーのプロファイル分類や解約予測に関する研究

Pal ら [60] はコミュニティ QA サイトの TurboTax Live Community [67] において、ユーザーがコミュニティに参加後に、良い回答を与えるアクティブなユーザーに発展するかどうかを予測する研究に取り組んだ。同様に、Movshovitz-Attias[62] らは、ユーザーの初期活動の違いによるコミュニティへの長期的な貢献についての分析を行った。

Pal, Harper ら [61] は専門知識を持つユーザーを特定するために、彼らが回答する質問を選択する際にバイアスがあることに着目した。専門知識を持つユーザーは回答が与えられていない、もしくは質が低い少数の回答のみが与えられている質問を選択し、回答をすることでコミュニティへ貢献するという特徴を利用した。

## 2.5. ユーザーのプロファイル分類や解約予測に関する研究

---

Song ら [63] は Quora[4] のデータを用いて、コミュニティにとって大事なユーザーを判別することに取り組んだ。具体的には、コミュニティの中の秩序を維持し、いくつかの分野において他のユーザーに知識を構築するように促すようなユーザーである。Furtado ら [64] も同様に、質問数などの過去の行動履歴から、ユーザーの分類に取り組んだ。

Dror ら [65] ならびに Pudipeddi ら [66] はユーザーの解約予測を行った。これらの研究ではユーザーの行動履歴を利用して、ユーザーの分類や解約予測を行っている。しかし、前述の通りスマートデバイスの普及によりユーザーの行動履歴を用いずに、即時的にユーザーの属性を抽出する方法が必要とされている。

# 第3章 コミュニティQAサイトにおける質問テキストのみを用いた不適切な投稿の検閲

本章では、適切な質問か不適切な質問かを分類するための、コンテンツ検閲システムの構築および不適切なコンテンツの検閲フローについて具体的に述べる。

## 3.1 はじめに

コミュニティQAサイトのような活発に質問と回答が行われる環境下においては、「荒らし」と呼ばれるユーザーが現れる。「荒らし」とは本来のコミュニティQAサイトの利用用途にそぐわない投稿を行うユーザーである。「荒らし」からの投稿には様々なものがある。例えば、不適切な商売への勧誘、サイトの炎上につながるような様々な表現、根拠のない誹謗中傷、いじめ等があるが、様々なタイプのものが現れるため、予め論理的に規定するのは困難である。

コミュニティQAサイトが小規模な時期においては、ユーザーの違反報告や、サイト運営側での目視による監視によって対処することは可能であるが、利用者の増大とともに、目視での運営は多大なコストがかかってしまう。そうした、ガイドラインに違反したコンテンツの投稿が、ユーザーの目に触れることはコミュニティにとってクリティカルな問題となる。なぜなら、ユーザーは少しでもコミュニティが荒れた雰囲気を感じると投稿を控えるようになるからである。以下はコミュニティQAサイトにユーザーから送られてきた違反報告の文章の一例である。

「正直、せっかく投稿したのにそれを攻撃されるなら、もう投稿したく無いしアプリに足も運びたく無くなります。たまたま否定的な意見の人が集まると、いじめのような状態になっていて、怖いです。」

上記のように感じるユーザーを減らすため不適切な投稿に対応することはコミュニティ運営上、非常に重要な意味を持つことがわかる。

先行研究では、このような問題に対して様々な手法で問題解決に取り組んできた。例えば Kayes ら [45] は Yahoo Answers のデータを用いて悪意のあるユーザーについて分析し、学習器と 29 種類の特徴量を用いて悪意のあるユーザーと悪意のないユーザーに分類し、83%という高い精度を出している。しかし、Kayes らが用いていた回答やフォロー、フォロワーといったユーザー同士の関係をはじめとした 29 種類の特徴量を収集し、利用できる形にすることは実応用においては可能であったにしても、高いコストがかかる。さらに、このような特徴量を利用することができない場合も存在している。

近年スマートデバイスの爆発的な普及により、インターネットへのアクセスが簡単になったことから、コミュニティQA サイトも気軽に利用できるようになった。スマートデバイスには特定の ID を保持することが可能である。そして、その ID のみでユーザーの認証が可能になったため、コミュニティQA サイトの中には email やパスワードが不要でニックネームのみでサービスの利用が開始できるサイトも多い。登録の敷居が下がったことにより、入退会を繰り返す、今までのユーザー履歴を消すことが容易になったため、「荒らし」も活動しやすい環境になった。このような状況下では、1人のユーザーに対して、質問投稿時に十分な過去のデータが存在しないために、回答やユーザー間の関係性（フォロワー）等の特徴量は利用することができず、悪意のあるユーザーとして判別するまでの十分なデータを集めることは困難である。

こうした場合の対策としては、質問テキストのみで、その質問が不適切な質問であるかどうかを判別し、コミュニティを荒らす投稿を防ぐことが考えられる。

コミュニティQA サイトにおいて、不適切な投稿かどうか質問をテキストのみを用いて分類する研究に関しては、小林ら [36] の研究が挙げられる。小林らは Yahoo!知恵袋において、人手で行っている不適切な投稿の削除を、機械学習の SVM を用いて自動的に行うことを試みた。コミュニティQA サイト以外ではスパムメールフィルター [47] やコンテンツベースで検出 [46] する取り組みも行われている。しかしいずれも、提案手法の精度評価に主張が置かれており、不適切な投稿をほぼ 100%除去する手法は提案されていない。また、実システムに関する研究としては、馬場らの研究 [48] がある。しかし、SVM を用いた検出システムの精度 (AUC) 評価であり、実運用における有効性評価はされていない。

実システムにおいては、先行研究にはない網羅的に不適切な投稿を削除する方法が求められるが、機械学習の手法のみで不適切な投稿を 100%除去することは、実際上極めて困難である。また、人による目視ですべての投稿を検閲すれば 100%の除去が達成できるが、コミュニティQA サイトの規模に比例して、高いコストがかかる。そこで、人による目視での確認の手法を残しつつ、可能な限り機械学習を用いて、不適切な投稿を判別することにより、人的コストを抑える形で

この問題を解決できると考えた。

そこで、本章では日本最大級の女性向けコミュニティQA サイトであるママリのデータを対象として、投稿の不適切さを評価した後に、その評価値（または不適切だという評価の信頼性）を元に、3個のクラス（不適切な投稿、人による判断が必要な投稿、適切な投稿）に分類する手法を提案し、人による介入を1/10以下に低減した上で不適切な投稿をほぼ100%除去することを目指す。

本章の貢献は以下のとおりである。

1. コミュニティQA サイトにおいて、スマートデバイスの普及とコストの問題により、ニーズの増している、質問テキストのみで不適切な投稿を機械学習により分類することを提案する。特に評価実験を通じて、最も分類結果の良い学習器とシステム辞書の組み合わせを検証する
2. その評価値を元に、3個のクラス（不適切な投稿、人による判断が必要な投稿、適切な投稿）に分類する手法を提案する
3. さらに、実社会で運営されているシステムへの導入も行い、提案手法の有効性について実システムでの検証を行う

本章の構成を以下に示す。3.2節では、課題について述べる。3.3節では、提案手法の詳細について述べる。3.4節では、評価実験の内容と結果を述べ、結果を踏まえた考察を述べる。最後に3.5節で本章のまとめと今後の課題について述べる。

## 3.2 大規模コミュニティQA サイトにおける課題

未経験の人材が、投稿を監視し、不適切性を判断できる人材になるまでには一定の経験を要する。そのような人材を随時採用することもできないため、監視ができるスキルを育成する必要がある。特に大規模コミュニティQA サイトであれば、その人的リソースの教育・獲得に高いコストがかかる。

本研究で用いるデータ元の商用コミュニティQA サイトは厚生労働省発表の人口動態統計 [6] から算出すると2016年に出産予定日を迎えた人の5人に1人以上がユーザー登録をしている、大規模コミュニティQA サイトである。コミュニティQA サイトにおいては、利用ユーザー数の増加とともにコミュニティを荒らすような、ガイドラインに違反した不適切なコンテンツの投稿が多く見られるようになる。

そのようなユーザーの回答やフォロー、フォロワーといったユーザー同士の関係をはじめとした特徴量を収集し、違反ユーザーを判別するシステムを作るため

### 3.3. 質問テキストのみを用いた、不適切な投稿の検閲手法

---

にはデータ基盤の整備といったコストがかかる。また、前節でも触れたとおり、スマートデバイスの普及によるサービス提供方法の変化により、違反ユーザーとして分類するための特徴量を収集できない場合もある。従って、投稿される質問テキストのみで、その投稿が不適切な投稿であるかどうかを判別し、コミュニティを荒らす投稿を防ぐ必要がある。

投稿数の少ないサービス初期の段階では、単語の部分一致に基づくフィルターを用いて、特定の単語を含む投稿を抽出する方法で対応が可能である。これは最も簡単な対応手法の1つであり、precisionが高くはないため、事後に人間が介在して、不適切な投稿と判断された投稿が本当に不適切か否かを判断する検閲が、最終段階で必要である。また、recallも高いわけではなく、不適切だと分類されなかった投稿にも十分注意を払い、不適切な投稿と分類されなかったが実際不適切である投稿を発見し、それを不適切とするための特徴的単語を発見する必要がある。

しかし、サービスがさらに成長すると、不適切な投稿の検閲のための監視対象投稿数は、1日あたり数千件に上り、上記のフィルターと人手による方法では、1日分の監視すべき量の投稿の検閲を、1日以内に処理することができなくなる。以上のような背景から目視での投稿監視は難しく、コミュニティQAサイトにおけるこのような状況下においては、この問題を解決する必要がある。

本研究では、人間が行うべき監視の主要部分を機械学習を用いたフィルターで代替することを試みる。そのため、機械学習アルゴリズムの改善と適切な利用が必要である。特に大量の不適切な投稿に対応しうる手法を提案し、実社会で運営されているシステムに導入した上で、提案手法の有効性について実システム上での検証を行う。

## 3.3 質問テキストのみを用いた、不適切な投稿の検閲手法

本節では前節で述べた課題を解決するために、機械学習を用いて、適切な投稿か、不適切な投稿かを分類する手法を提案する。また、提案手法において、判別結果の最も良い学習器と、テキストを単語に分割する際に利用するシステム辞書の中で、最も良い結果が得られた辞書や辞書の拡張手法について述べる。さらに、機械学習を用いて学習器を構築した後に、学習器からの評価値を元に、閾値についてのチューニングを行うことについても述べる。最後に学習後の学習器で予測した評価値を用いて、投稿を3個のクラス（不適切な投稿、人による判断が必要な投稿、適切な投稿）に分類した上で、適切な recall を確保しつつ、不適切な投

稿をほぼ 100%検知する手法を提案する。

#### 3.3.1 機械学習を用いた分類

本研究では、以下の手法で学習器を構築し、投稿の検閲の自動化を試みる。

1. 形態素解析プログラム (MeCab[12] 等) を用いて質問文から単語を切り出す
2. 1 の単語から特徴量として用いる単語として、品詞が名詞、動詞、形容詞、記号とされている単語を抽出
3. 質問文毎に単語をカウントして特徴ベクトルを作成
4. 学習器を構築

未知の投稿进行分类する場合、同様に 1 から 3 の処理を行い、学習後の学習器を用いて、分類する。

なお、精度向上への取り組みとして学習器の比較・検討を行う。対象とする学習器は Naive Bayes, Decision Tree, SVM, Random Forest[20] である。

次に、システム辞書の選定を行う。形態素解析プログラム (MeCab) で質問文から単語を切り出す際にシステム辞書は利用される。比較対象となるシステム辞書は ipadic, jumandic, unidic と mecab-ipadic-NEologd[13] である。mecab-ipadic-NEologd とは、多数の Web 上の言語資源から得た新語をシステム辞書に追加することにより拡張した MeCab 用のシステム辞書である。

さらに、辞書自体の拡充を行う。コミュニティQAサイトの特徴の1つとして、コミュニティの中に独特な単語が生まれる。例えば、Yahoo!知恵袋の中では知恵袋用語集と呼ばれる用語集が存在している。具体的な例として、「削除隊」を挙げる。「削除隊」は投稿を削除する知恵袋スタッフのことである。このような用語はシステム辞書の中にもないことも多い。従って辞書の拡充が精度向上に寄与すると考えた。コミュニティQAサイト毎の特有の単語を追加・調整することにより、精度の向上に取り組む。その辞書拡充のステップを図 3.1 に示す。

辞書の拡充自体は非常に愚直な作業となる。ユーザーから違反報告があがった投稿に対して、目視で何がその違反の原因だったのかを判断し、分類に適した単語があれば辞書に追加していく。

最後に、学習後の学習器の評価値を元に閾値についてチューニングを行う。学習器を学習する際、適切な投稿と不適切な投稿の割合が 1 対 1 のデータセットを用いるが、実際には不適切な投稿の割合は、適切な投稿の割合に比べてごく少数





図 3.1: 辞書拡充のステップ

である。そこで、実験結果を確認しながら、実際のクラス割合に等しいクラス割合で構成された検証データを用意する。検証データを用いて不適切な投稿全てを検知しつつ、適切な投稿が不適切な投稿として分類される件数を大幅に削減するような閾値を探し、チューニングを行う。

#### 3.3.2 提案手法

前項の手法のみで、高い精度が出たとしても、小林ら [36] の取り組みのように 100%の精度には達せず、実運用の中ですべてを機械による判別により、自動化してもクリティカルなミスが出ることは十分に考えられる。それにも関わらず、先行研究にはない、網羅的に不適切な投稿を削除する方法が求められる。しかし、自動的に 100%不適切な投稿を除去することは困難である。

そこで適切な投稿、不適切な投稿に加えて、人による判断が必要とされるクラスを追加し、投稿を 3 個のクラスに分類した上で、人による判断が必要と分類された場合、人によって判断し、不適切な投稿をほぼ 100%検知する手法を提案する。機械学習と目視を組み合わせた提案手法を図 3.2 に示す。

まず、ユーザーからの投稿を予め学習した学習器を用いて判別する。このとき、学習器が出力する評価値 (probability) が上限閾値以上である場合は、不適切な投稿として扱い、逆に下限閾値以下であった場合には適切な投稿として扱う。下限閾値を設定する理由は、false negative な判別をしてしまった投稿を発見するためである。false negative な判別とはすなわち、本来は不適切な投稿であるが、そうでないと判別したということである。繰り返しになるが、そのような投稿はコミュニティにとってクリティカルな影響を及ぼす投稿となる。以上より、この誤判別を 0 に近づけるためにこのような二つの閾値を設定する。

学習器による評価値が二つの閾値の間に含まれた場合、その投稿を、目視での投稿内容の確認が必要な投稿として、リストに登録する。そして、そのリストに

### 3.3. 質問テキストのみを用いた、不適切な投稿の検閲手法

登録された投稿内容を人間が確認し、適切な投稿か、不適切な投稿かを判断する。

機械学習を最大限利用するが、あえてすべてでは自動化しないことで、結果としてコミュニティQAサイトにとって大事な検閲の精度を高め、運用可能な形を目指す。

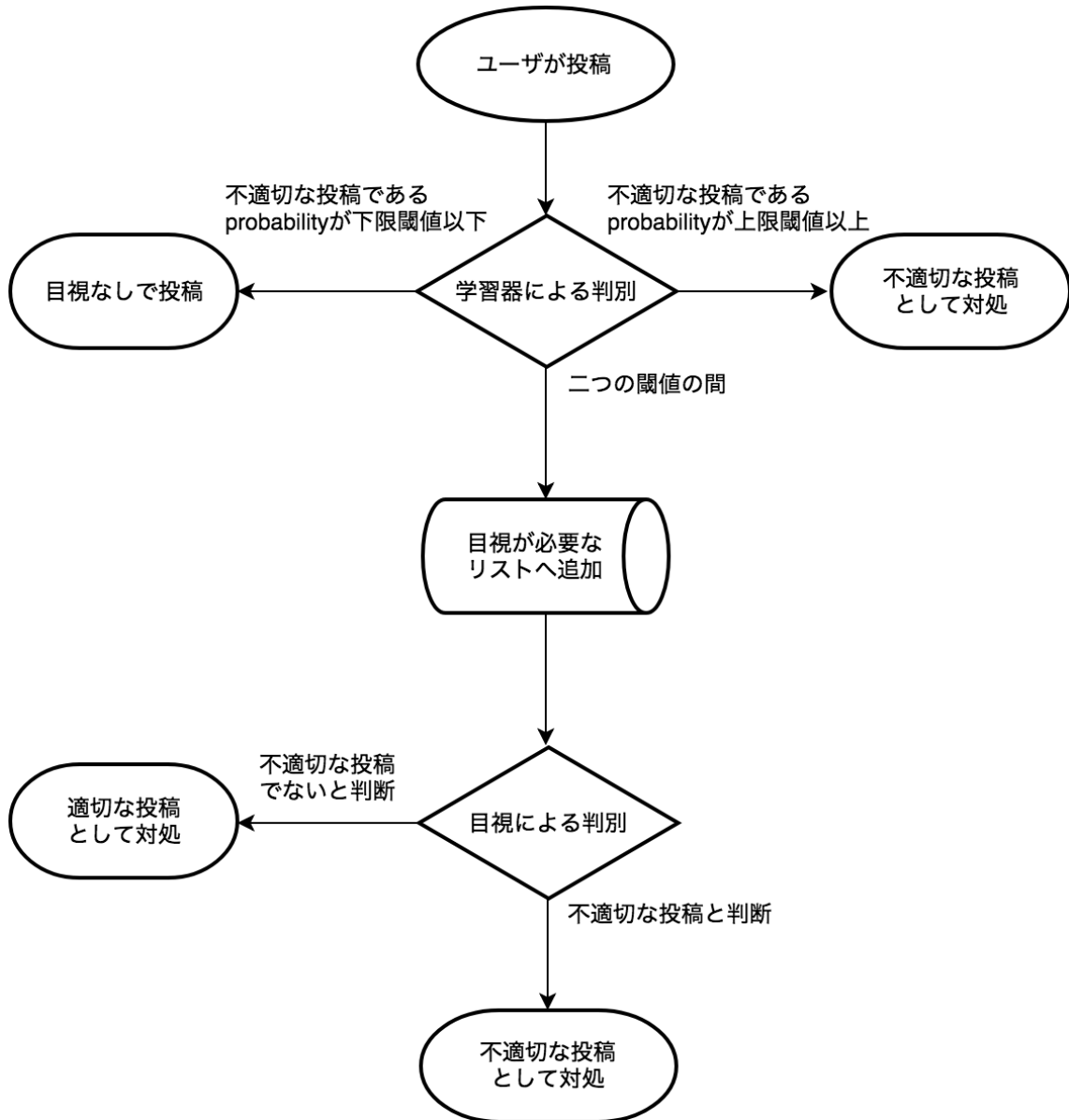


図 3.2: 機械学習と目視での投稿内容確認を組み合わせた提案手法

### 3.4 実験

#### 3.4.1 データセット

商用コミュニティQAサイトにおいて、過去実運用中に、目視により「不適切な投稿である」と判断した投稿1,000件と「適切な投稿」1,000件を用意し、合計2,000件の投稿に対して二値分類を行うことにした。

#### 3.4.2 実験手法

3.3節で述べた機械学習を用いた分類手法を用いて分類を行い、分類の精度を比較する。実装にはscikit-learn[17]を利用した。以下の項目について比較・考察を行う。なお、以後の実験での評価は、すべて5-fold cross validationで行った。

- 学習器毎の精度の比較
- MeCabを用いて質問文から単語を切り出す際に利用される、システム辞書についての精度の比較
- 機械学習を用いて学習器を構築した後に、学習器が出力する評価値を元にした閾値のチューニング
- 提案手法を実システムで運用した際の検証

#### 3.4.3 評価指標

評価指標の定義のために投稿のクラス分類の予測結果 (Prediction Outcome) と実際のクラス (Actual Value) との混同行列を図3.3に示す。Actual Valueが $p'$ の場合が真に不適切な投稿であり、Actual Valueが $n'$ の場合が真に不適切な投稿でない。Prediction Outcomeが $p$ の場合が不適切な投稿だと予測された値であり、Prediction Outcomeが $n$ の場合が不適切な投稿でないと予測された値である。また、 $P'$ が $p'$ の合計であり、 $N'$ が $n'$ の合計である。同様に、 $P$ が $p$ の合計であり、 $N$ が $n$ の合計である。

このとき、True Positive (TP), False Negative (FN), False Positive (FP), True Negative (TN) を用いて、評価指標を以下に定義する。

		Prediction Outcome		total
		p	n	
Actual Value	p'	True Positive	False Negative	P'
	n'	False Positive	True Negative	N'
total		P	N	

図 3.3: 投稿のクラス分類の予測結果 (Prediction Outcome) と実際のクラス (Actual Value) との混同行列

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.1)$$

$$Precision = \frac{TP}{TP + FP} \quad (3.2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3.3)$$

$$F1 = \frac{2Recall \cdot Precision}{Recall + Precision} \quad (3.4)$$

#### 3.4.4 目標とすべき指標数値の定義

評価指標として, precision, recall, F1, accuracy[18] を式 (3.1) から式 (3.4) に定義する. 目標とすべき指標数値として, 不適切な投稿を見逃さないことが最重要であるので, 不適切な投稿の recall を 1 に近づけることを第 1 目標とする.

またその条件下で, 適切な投稿を不必要に不適切な投稿として分類しないよう precision を実用上十分なまで上げることが第 2 目標である. これら二つの目標を達成することで, コミュニティQA サイトとして実運用が可能になり, 安心して使い続けてもらえるコミュニティを作ることに繋がる.

### 3.4.5 実験結果

#### (1) 学習器の選定

それぞれの学習器で3.4.1のデータを分類した結果を表3.1に示す。システム辞書はMeCab標準のシステム辞書であるipadicを用いた。最も良い結果の得られた学習器はRandom Forestであった。以後の実験についてはRandom Forestを学習器として使用する。しかし、それでもprecision, recallともに0.7程度であり、網羅的に不適切な投稿を検閲できる精度に達するには、学習器の選定だけでは難しいことが分かった。

#### (2) システム辞書の選定

次にシステム辞書自体の精査を行った。mecab-ipadic-NEologdを含めた各辞書で最も良い辞書を選定するために、(1)の学習器の比較で最も結果の良かったRandom Forestを用いて辞書の比較を行った。ipadic, jumandic, unidic, ipadicにmecab-ipadic-NEologd加えてそれぞれで実際に判別を行った結果を表3.2に示す。

表3.2よりimecab-ipadic-NEologdが最も良い結果となったことがわかる。スマートデバイスベースのサービスであるママリには絵文字や顔文字、Web上で

表 3.1: 学習器毎の各指標の比較

分類手法	Precision	Recall	F1	Accuracy
Naive Bayes	0.65	0.63	0.62	0.66
Decision Tree	0.66	0.66	0.66	0.66
SVM	0.69	0.57	0.50	0.58
Random Forest	<b>0.71</b>	<b>0.70</b>	<b>0.69</b>	<b>0.70</b>

表 3.2: システム辞書毎の各指標の比較

システム辞書	Precision	Recall	F1	Accuracy
ipadic	0.71	0.70	0.69	0.70
jumandic	0.72	0.71	0.71	0.71
unidic	0.71	0.70	0.70	0.70
mecab-ipadic-NEologd	<b>0.73</b>	<b>0.72</b>	<b>0.72</b>	<b>0.72</b>

### 3.4. 実験

よく使用されるスラングやより一般的な会話の中で使われる単語等が利用されるため、結果に改善が見られたと考えられる。

具体例として、

「携帯一つで簡単で安心安全です！良かったメールレディご紹介しますよ！（笑顔絵文字顔文字）」

という投稿について、mecab-ipadic-NEologd を利用して MeCab を実行した結果と ipadic で実行した結果をそれぞれ図 3.4, 図 3.5 に示す。

メールレディとは男性とメールでやりとりをする在宅ワークのことである。この二つの図を見ると「メールレディ」という単語が ipadic では「メール」という単語と「レディ」という単語に分割され、mecab-ipadic-NEologd では「メールレディ」という 1 単語に分割されていることがわかる。

ママリにおいては「メールレディ」に関する投稿は不適切な投稿である割合が高いため、mecab-ipadic-NEologd の方が正しい分割をしていると考えられる。従って、mecab-ipadic-NEologd の方が文章の内容をより理解するための情報を得ることができる。以上より、mecab-ipadic-NEologd をシステム辞書として採用

```
携帯 名詞,サ変接続,*,*,*,*,携帯,ケイタイ,ケイタイ
一つ 名詞,一般,*,*,*,*,一つ,ヒトツ,ヒトツ
で 助詞,格助詞,一般,*,*,*,*,で,デ,デ
簡単 名詞,形容動詞語幹,*,*,*,*,簡単,カンタン,カンタン
で 助動詞,*,*,*,特殊・ダ,連用形,だ,デ,デ
安心 名詞,サ変接続,*,*,*,*,安心,アンシン,アンシン
安全 名詞,形容動詞語幹,*,*,*,*,安全,アンゼン,アンゼン
です 助動詞,*,*,*,特殊・デス,基本形,です,デス,デス
！ 記号,一般,*,*,*,*,!,!,!
良かつ 形容詞,自立,*,*,形容詞・アウオ段,連用タ接続,良い,ヨカツ,ヨカツ
たら 助動詞,*,*,*,特殊・タ,仮定形,た,タラ,タラ
メール 名詞,一般,*,*,*,*,メール,メール,メール
レディ 名詞,一般,*,*,*,*,レディ,レディ,レディ
ご 接頭詞,名詞接続,*,*,*,*,ご,ゴ,ゴ
紹介 名詞,サ変接続,*,*,*,*,紹介,シヨウカイ,シヨウカイ
し 動詞,自立,*,*,サ変・スル,連用形,する,シ,シ
ます 助動詞,*,*,*,特殊・マス,基本形,ます,マス,マス
よ 助詞,終助詞,*,*,*,*,よ,ヨ,ヨ
！ 記号,一般,*,*,*,*,!,!,!
😊 記号,一般,*,*,*,*,*
```

図 3.4: ipadic を利用して MeCab による形態素解析を実行した結果

### 3.4. 実験

```
携帯 名詞,サ変接続,*,*,*,*,携帯,ケイタイ,ケイタイ
一つ 名詞,一般,*,*,*,*,一つ,ヒトツ,ヒトツ
で 助詞,格助詞,一般,*,*,*,*,で,デ,デ
簡単 名詞,形容動詞語幹,*,*,*,*,簡単,カンタン,カンタン
で 助動詞,*,*,*,特殊・ダ,連用形,だ,デ,デ
安心 名詞,サ変接続,*,*,*,*,安心,アンシン,アンシン
安全 名詞,形容動詞語幹,*,*,*,*,安全,アンゼン,アンゼン
です 助動詞,*,*,*,特殊・デス,基本形,です,デス,デス
! 記号,一般,*,*,*,*,!,!,!
良かつ 形容詞,自立,*,*,形容詞・アウオ段,連用タ接続,良い,ヨカッ,ヨカッ
たら 助動詞,*,*,*,特殊・タ,仮定形,た,タラ,タラ
メールレディ 名詞,固有名詞,一般,*,*,*,メールレディ,メールレディ,メールレディ
ご 接頭詞,名詞接続,*,*,*,*,ご,ゴ,ゴ
紹介 名詞,サ変接続,*,*,*,*,紹介,ショウカイ,ショーカイ
し 動詞,自立,*,*,サ変・スル,連用形,する,シ,シ
ます 助動詞,*,*,*,特殊・マス,基本形,ます,マス,マス
よ 助詞,終助詞,*,*,*,*,よ,ヨ,ヨ
! 記号,一般,*,*,*,*,!,!,!
😊 名詞,一般,*,*,*,*,笑顔,エガオ,エガオ
```

図 3.5: mecab-ipadic-NEologd を利用して MeCab による形態素解析を実行した結果

することにした。以後の実験については mecab-ipadic-NEologd をシステム辞書として使用する。

#### (3) 辞書の拡充

コミュニティQA サイト毎の特有の単語を追加・調整することにより、精度の向上に取り組んだ。ユーザーからの違反報告に向き合い、合計約 2,000 単語について検討、精査を行い、辞書の拡張を続けた。3.4.1 のデータセットに対して拡張した辞書を利用し、学習器には Random Forest を用いて判別を行ったところ、各指標は表 3.3 に示す値となった。

辞書の拡張の具体例として、「ビジネスパートナー」という単語を辞書登録する

表 3.3: 辞書拡充前後の各指標の比較

辞書拡充の前後	Precision	Recall	F1	Accuracy
辞書拡充前	0.73	0.72	0.72	0.72
辞書拡充後	<b>0.81</b>	<b>0.80</b>	<b>0.80</b>	<b>0.80</b>

かの判断について説明する。ママリにおける「ビジネスパートナー」という単語を用いた投稿はユーザーをネットワークビジネスへ勧誘する投稿であることが多い。例えば以下のような投稿である。

「はじめまして、私は今妊娠中ですが家からの仕事で旦那より稼いでます。一緒に仕事をするビジネスパートナーを募集中です。もし興味があれば連絡先を交換しませんか？」

この文章を形態素解析すると、「ビジネス」と「パートナー」という二つの名詞に分かれる。このような単語の場合、「ビジネスパートナー」を1語として認識させるために、単語として辞書登録を行う。以上のような辞書登録を経た結果、precision, recallともに0.8以上となった。約2万件近くの違反報告を精査することは愚直な取り組みではあるが、コミュニティQAサイトの不適切な投稿の分類においても、精度向上に寄与することがわかった。

#### (4) 機械学習で用いた評価値を元にした閾値チューニング

(3)までは、学習器の学習のために、適切な投稿と適切でない投稿の割合が1対1のデータを用いるが、実際には不適切な投稿の割合は適切な投稿に比べてごく少数である。直近で行われた投稿約100万件の実データから1,000件をランダムサンプリングし、内容を確認したところ不適切な投稿は12件であった。この1,000件に対して、学習済みの学習器により判別を行ったところ表3.4に示す結果となった。

不適切な投稿の recall が1.0であることから、学習器による判別によって不適切な投稿が全て検知できていることがわかる。一方で、本来は不適切でない投稿として判別されるべき一部のデータが不適切な投稿として判別されているため、不適切な投稿における precision の値が低い。

そこで、検証データを用いて不適切な投稿全てを検知しつつ、適切な投稿が不適切な投稿として判別される件数を大幅に削減するような閾値を調査し、チューニングを行う。

scikit-learn の RandomForestClassifier を用いて probability を出力させ、その判別の閾値を調整することにした。probability を確認すると、不適切でない投稿であるが、不適切な投稿として判別された投稿の多くは probability が0.5から0.74という評価値であることがわかった。そこで、デフォルトの設定における閾値は0.5であるが、0.74という閾値に変えた結果、表3.5に示す結果となった。これにより、学習器による評価値を元に閾値についてチューニングを行うことで、



### 3.4. 実験

表 3.4: サンプルングしたデータに対する判別結果

	Precision	Recall	F1	全体件数
適切な投稿	1.00	0.82	0.90	988
不適切な投稿	0.06	1.00	0.12	12

表 3.5: サンプルングしたデータに対する、閾値調整後の判別結果

	Precision	Recall	F1	全体件数
適切な投稿	1.00	<b>0.99</b>	0.99	988
不適切な投稿	<b>0.55</b>	1.00	0.71	12

すべての不適切な投稿を検知しつつ、適切な投稿が不適切な投稿として判別される件数を大幅に削減できることがわかった。

#### 3.4.6 提案手法の有効性についての実システムでの検証

3.3.1 の手法と 3.3.2 の手法を実システムとして運営されている商用コミュニティQA サイトに導入し、検証を行った。

##### (1) クリティカルな分類ミスの存在

3.3.1 の手法で実際にコミュニティQA サイトに提案手法を導入した場合、分類ミスは少ないが、運用上極めて不適切な投稿でもあるにも関わらず、適切と分類してしまう場合が生じた。例えば、以下の投稿は一見普通の投稿に見えるが、実は投稿文中のリンク先は、ユーザーを不適切なコンテンツへと誘導させている。

「みなさん子育てで参考になっている本や人、ブログなどありますか??おすすめなどあれば教えてほしいです(^-^)/私はこの方のブログを参考にさせてもらってます!優しい気持ちになれるので、めちゃめちゃお勧めです!(^^) <http://xxx.xxx.xxx> (不適切なコンテンツの閲覧へ誘導させるリンク)」

このような、リンクによる不適切なサイトへの誘導を検出するには、記述を見るだけでは不十分であり、リンク先のコンテンツをダウンロードし、その内容を解析し、その適否を判断する必要がある。一般にこうしたサイトは巧みに書かれて

いるため、その違反性の判別は簡単ではない。1つの簡便な方法は、ブラックリストによる方法である。こうしたブラックリストは極めて有用ではあるが、最新のブラックリストを作成、更新し続けることはコスト的に困難である。別の方法は、人間に頼るものであり、目視での内容確認を組み合わせた3.3.2の手法が挙げられる。

#### (2) 投稿を不適切、人による判断、適切な3つへ分類

3.3.1の手法のみでは、実際にコミュニティQAサイトでは運用上極めて不適切な投稿でもあるにも関わらず、適切と分類してしまう場合が生じた。そこで、投稿を3個のクラス（不適切な投稿、人による判断が必要な投稿、適切な投稿）に分類し、その結果、不適切な投稿をほぼ100%検知する手法（図3.2）を導入した。

機械学習によるフィルタリングによって、怪しくはあるが不適切とは断定できない投稿を、不適切、適切とは別のクラスとして分類し、その後、そのクラスに分類された投稿を人間が最終判別するシステムを作成した。実運用では3.4.5(4)と同様に scikit-learn の RandomForestClassifier を用いて、投稿毎に不適切な投稿である probability を出力させた。それが一定の閾値を越える場合に人に判断を促すようなアラートを出すという手法を行うために、scikit-learn における RandomForestClassifier が出力する、不適切な投稿である確率の推定値に対する上限閾値を3.4.5(4)に沿い0.74に設定し、人に判断を促すアラートを出す下限閾値は0.6として設定した。

このシステム導入後、人手による判別が必要と判断される投稿は、1日あたり200件程度であった。この件数は全投稿の1%以下である。これにより、1日数回、リストに登録された投稿をチェックする程度で検閲が完了している。3.3.2の手法により、機械学習を最大限利用するが、あえてすべては自動化しないことで、結果としてコミュニティQAサイトにとってクリティカルな問題となる、不適切な投稿に対する検閲の精度を高め、実サービスでも運用可能なことがわかった。

## 3.5 まとめと今後の課題

コミュニティQAサイトのような、活発に質問と回答が行われる環境下においては、適切でない投稿の検知が課題となる。しかし、先行研究で行われてきたユーザーの行動履歴を使用した方法は大きなコストがかかる。また、スマートデバイスの普及によりそもそも不適切な投稿を行うユーザーの行動履歴が取りづらくなったという課題も存在している。

### 3.5. まとめと今後の課題

---

それらの課題に対し、限られたリソースの中で機械学習を用いて人間が行うべき監視の主要部分を代替するためには、機械学習を用いた上での工夫が必要である。そこで本章では、課題解決のための手法を提案し、大規模なコミュニティQAサイトにおける、不適切な投稿に対応した手法を、実社会で運営されているシステムへの導入を行うことで、提案手法の有効性を確認した。

具体的には、コミュニティQAサイトによる不適切な投稿の分類に用いる学習器として Naive Bayes, Decision Tree, SVM, Random Forest を比較、さらに形態素解析を使用する際に使われるシステム辞書について ipadic, jumandic, unidic と多数の Web 上の言語資源から得た新語を追加することにより拡張した MeCab 用のシステム辞書である mecab-ipadic-NEologd を利用した。学習器は Random Forest, システム辞書としては mecab-ipadic-NEologd の精度が高いことがわかった。

多くの研究では数%の精度差について最新の学習アルゴリズムを用いて競い合うことが多い。しかしながら現実に適用する際には、非常に愚直で時間はかかるものの、「ビジネスパートナー」といったコミュニティQAにおけるシステム辞書の中になが不適切な投稿の分類のために重要と思われる単語を追加し、辞書を整備することが確実に効果的であることがわかった。

また、学習器が出力した評価値を元に閾値についてチューニングを行うことで、不適切な投稿の recall を 1 に近づけつつ、不適切でない投稿を不必要に不適切な投稿として分類しないよう precision を実用上十分な値まで上げることが可能なことを確認することができた。

最後に、機械学習での一定の成果に基づき、機械学習によるフィルタリングによって、不適切な投稿とは判断できない投稿を、人による判断が必要な投稿と分類し、その後、そのクラスに分類された投稿を人間が最終判別するシステムを作成し、実システムとして利用者の多い大規模コミュニティQAサイトでその手法の効果を示した。結果として、先行研究にはない、「網羅的に不適切な投稿を削除する方法」を提案することができた。

現在も継続的かつ定期的に辞書のアップデートを行いながら今の閾値が適切かどうかの確認を行っている。今後は画像付きの投稿に対しても機械学習を用いたシステムによる画像判別により、不適切な投稿を検知する仕組みを導入していきたいと考えている。

## 第4章 コミュニティQAサイトにおける共感を求める質問の認識

本章では、回答率の低いと推測される質問の抽出を行うために、共感を求める質問とそれ以外の質問の定義を述べ、その上で機械学習を用いて、共感を求める質問と共感以外を求める質問かを分類する手法について述べる。

### 4.1 はじめに

大規模コミュニティQAサイトでは、様々な社会的・心理的・経済的な問題を抱えているユーザーや、健康上の問題を抱えているユーザーが存在している。コミュニティQAサイトに投稿を行うユーザーの中には、相談するあてのない、そうした質問を投稿することがある。そのような質問に対しては、答えが得られれば十分というのではなく、その応答を通して、やりとりをしている同士が繋がっているという感覚が重要である。その感覚から、孤独ではないという心理的支援を得られることが、投稿を行うユーザーにとって必要である。そのためには、コミュニティQAサイトに質問が投稿されてから短時間で回答が得られ、その回答は心のこもった内容であることが望ましい。

一方、回答が得られないと、逆に負の心理的影響を受け、単にコミュニティQAサイトから離れるという、コミュニティQAサイトの運営者にとって望ましくない結果となる。このような結果は、コミュニティQAサイトを作成し、運営する側にとっては容認できない。このような状況になってしまう1つの原因として、前述の相談するあてのない質問のような、直接的な回答を求めない質問投稿が挙げられる。回答が必要にも関わらず、回答が難しいため、このような質問に対しては回答が与えられにくいという特徴がある。例えば以下のような質問である。

「子育てに疲れてイライラを人にぶつけてしまいます..手をあげたりはしないけど、ついイライラして怒鳴ってしまう。旦那の転勤で引っ越してきてから、イライラがひどくなってしまって…こんな自分に嫌気がさします。正直辛い。」

#### 4.1. はじめに

---

表 4.1: 共感を求める質問と共感以外を求める質問それぞれ 300 件に対する未回答の数

質問の種別	回答あり (1 件以上)	未回答
共感を求める質問	162	138
共感以外を求める質問	299	1

質問の具体的な答えそのものよりも、自分の思いへの共感を求めるこのような質問は、商用データ内においては、全質問の約 2 割を占めている。しかし、このような質問は回答者からすると回答しにくい質問である。

共感を求める質問と共感以外を求める質問とを商用コミュニティQA サイトの中から手作業でそれぞれを 300 件取り出し、未回答質問数と回答者数を調べた。その結果を表 4.1, 表 4.2 に示す。共感を求める質問と共感以外を求める質問とを手作業で振り分ける方法については、4.3 で言及する。なおコミュニティQA サイトの性質を踏まえて、本論文での「質問」とは、本論文で注目している、共感を求める質問と共感以外を求める質問の両者を意味するものとする。

表 4.1 から、共感を求める質問では 300 件中 138 件が未回答のままであるが、共感以外を求める質問では未回答のままのものは 1 件しかない。共感を求める質問は、共感以外を求める質問に比べて、回答を得られないままの状態であることが多いことがわかる。

それぞれ 300 件の質問に対して、共感を求める質問と共感以外を求める質問の回答者数は表 4.2 より、共感を求める質問には 309 人、共感以外を求める質問に対しては 847 人である。質問数よりも回答者数が多いのは、1 つの質問に対して複数人の回答者がいる場合があるからである。2 つの値を比較すると、約 2.7 倍の違いがあり、共感を求める質問には共感以外を求める質問に比べて、回答者数が少ない。

ママリが提供されているアプリマーケットの 1 つである App Store[54] におけるママリのレビューにおいて、5 段階評価でもっとも高い評価である 5 のついたレビューからランダムに 100 件のレビューを選択したところ、そのうち 58 件が「回答がもらえることが嬉しい」という主旨であった。

以上から、回答が得られることは、質問を投稿したユーザーの満足度が向上するための重要な要素の 1 つであることがわかる。そこで本章では、共感を求める質問に対しての回答率を上げることを目的とする。

表 4.2 の回答者属性の割合より、共感を求める質問と共感以外を求める質問とで、回答者属性の割合が違うことがわかる。共感以外を求める質問では 67% の回

## 4.2. 共感の定義

表 4.2: 共感を求める質問と共感以外を求める質問それぞれ 300 件に対する回答者数 (人) と回答者属性の割合 (%)

ユーザー属性	共感を求める質問	共感以外を求める質問
出産経験者	255 人 (83%)	565 人 (67%)
出産未経験者	54 人 (17%)	282 人 (33%)
合計	309 人 (100%)	847 人 (100%)

答者が出産を経験したユーザーだったが、共感を求める質問では 83% の回答者が出産を経験したユーザーであった。このことから、出産を経験したユーザーに対して、共感を求める質問に対する回答を促すことで、共感を求める質問に対しての回答を得る可能性を上げることができるのではと考えた。

図 4.1 のように、スマートフォンにおける、ママリを含むコミュニティサービスのユーザーインターフェースの 1 つとして、ユーザーが投稿した質問が一覧で表示される形式が存在している。指を上から下へフリックすることにより画面が遷移し、新たな質問の閲覧が可能である。ユーザーは質問の中で興味のある質問をクリックし、回答を行う。この質問一覧画面において、ユーザー属性に応じて表示する質問の種類をコントロールすることによって回答率の改善が可能であると考えた。

そこで、質問文から抽出した特徴と機械学習を用いて、共感を求める質問と共感以外を求める質問を分類した後、高い回答率を持つユーザー（この場合、出産経験者）に共感を求める質問を提示することによって、質問の回答率が高まるものと考えた。そのために、共感を求める質問と共感以外を求める質問を適切に分類することを目指す。

本章の構成を以下に示す。4.2 節では、共感の定義について述べる。4.3 節では、共感を求める質問と共感以外を求める質問の定義について述べる。4.4 節では、提案手法について述べる。4.5 節では、評価実験の内容と結果を述べ、結果を踏まえた考察を述べる。最後に 4.6 節で本章のまとめと今後の課題について述べる。

## 4.2 共感の定義

4.3 にて共感を求める質問と共感以外を求める質問それぞれについて定義を行うために、本節では「共感」の定義を確認する。

「共感」という言葉の定義として広辞苑 [55] では“他人の体験する感情や心的状態、あるいは人の主張などを、自分も全く同じように感じたり理解したりする

## 4.2. 共感の定義



図 4.1: ママリの画面図

こと、同感”という定義をしている。これまで、「共感」については様々な研究が行われ、研究者たちはそれぞれの解釈を元に定義を行ってきた。

Köhler[56]は“相手と感情を共有することよりも、それらを理解することが重要”としている。Dymond[57]は“他者の考え、感情、行為の中に自分自身を想像的に置き換え、その人の世界を構成すること”としている。Stotland[58]は“他者が感情を体験しているか、体験しようとしていることを、理解したために起こる、観察者の感情的な反応”としている。Wispé[59]は“個人的な見解、または、基準に基づいて判断するのを控えた上で、相手を理解しようとする試み”としている。それぞれの研究者の定義から共通項を鑑みると、「相手の感情を理解すること」が「共感」であると考えられる。この「共感」の定義を前提として、4.3にて共感を求める質問と共感以外を求める質問それぞれについて定義を行っていく。

## 4.3 質問分類タスクの定義

### 4.3.1 分類の定義

本研究では、下記の2つに質問を分類した。

#### 共感を求める質問

本研究では、「共感を求める質問」を以下のいずれかの要素を持つ質問と定義した。

- 質問文に対して、「わかります」という理解を示す文章のみで回答を与えることができる
- 疑問を解くような具体的な回答や提案を与える必要がない

質問者は自分の質問を肯定もしくは理解されることを求めて質問すると考えられる。

#### 共感以外を求める質問

本研究で述べる「共感以外を求める質問」とは以下の要素を持つ。

- 質問文に対して、「わかります」という理解を示す文章のみでは回答として不十分である



### 4.3. 質問分類タスクの定義

---

- 疑問を解くような具体的な回答や提案を与える必要がある

質問者は回答者の多様な意見もしくは、正しい事実を求めて質問すると考えられる。

#### 4.3.2 各分類の具体的な例

「子育てに疲れてイライラを人にぶつけてしまいます..手をあげたりはしないけど、ついイライラして怒鳴ってしまう。旦那の転勤で引っ越してきてから、イライラがひどくなってしまって…こんな自分に嫌気がさします。正直辛い。」

という質問は共感を求める質問の1例である。「辛いですね、わかります」という回答をもらったこのようなユーザーにインタビューを行なうと、「世の中に同じ立場の人が居り、同じ悩みを共有することで不安が和らぎ、安心感を覚えた」というような回答が得られた。一方、

「つわりがひどいのですが、楽になる方法はありませんか？」

という質問は共感以外を求める質問の1例であり、つわりを楽にする具体的な方法を求めている。

#### 4.3.3 質問分類タスクにおける議論

投稿の中には判断が非常に難しい投稿も存在する。例えば以下のような投稿である

「ほんと旦那が家事しなくてムカつく!!!料理毎日してるのでゴミ捨て一つしないってどーゆうことなん!?皆さんそう思いませんか？」

この文章は「皆さんそう思いませんか？」という部分を回答者に対しての問いかけととり、多様な意見を求める質問として分類することも考えうるが、一般的にこのような質問を問いかけるユーザーは不満を投げかけ、それに対する理解を求めている。よって、本研究においては共感を求める質問として分類した。

## 4.4 分類手法

### 4.4.1 特徴語句による分類

共感を求める質問かどうかを特定のキーフレーズ、特徴語句を用いて分類を行なう。この特徴語句による分類方法をベースラインと呼ぶ。表 4.3 に本手法で用いる特徴語句のリストを示す。この特徴語句のリストは栗山ら [8] の研究で「情報検索型」の質問と「社会調査型」の質問に特徴的な表現（名詞・形容詞・形容動詞・文末表現）のパターンを抜き出したものである。このリスト中の文字列（103 個）が質問文に含まれない場合、共感を求める質問と分類する。

表 4.3: 特徴語句リスト

---

何時	どこ
誰	だれ
名前	場所
金額	料金
いくら	可能ですか
できない	出来ない
出来ますか	できますか
出来ませんか	できませんか
本当ですか	ホントですか
何から	違い
理由	原因
なぜ	何故
方法	手立て
やり方	対策
方策	し方
法	どうやって
どれくらい	どのような
どうすればいい	どうやったら
どちらがいい	べきですか
べきじゃないですか	た方が良い
た方が悪い	ほかない
しかない	大丈夫
した方がいい	ていいですか
たほうがいい	どう思いますか
一番の	良い
いい	悪い
だめ	ダメ
良くない	よい
おいしい	美味しい
美味しく	うまい
まずい	おいしく
うれしい	おもしろい
楽しい	おすすめ
お勧め	お薦め

表 4.3: 特徴語句リスト (続き)

---

オススメ	有名な
話題の	評判の
人気	魅力
好き	嫌い
気に入り	評価
どうですか	どうでした
ことありますか	ってありますか
していましたか	していますか
てましたか	しますか
しませんか	した方がいたら
した方は	ている方
た人	てる人
経験	一般論
みなさん	皆さん
あなたは	思う
思える	思われる
だから	なんで
よね.	

---

### 4.4.2 機械学習を用いた分類

共感を求める質問を機械学習を用いて判別する。自然言語処理における分類ではいくつかの手法が存在するが、3章と同様に、下記に示す方法を基本とし、共感を求める質問を判別する。

1. 形態素解析プログラム (MeCab 等) を用いて質問文から単語を切り出す
2. 1の単語から特徴量として用いる品詞として、名詞、動詞、形容詞、記号を抽出
3. 質問文毎に単語をカウントして特徴ベクトルを作成
4. 学習器を構築

未知の質問を分類する場合、同様に1から3の処理を行い、学習後の学習器を用いて、分類する。また、MeCabを用いて質問文から単語を切り出す際に利用される、システム辞書についても検証する。

システム辞書には MeCab 標準のシステム辞書である ipadic と mecab-ipadic-NEologd を利用する。3章で述べたように mecab-ipadic-NEologd は、多数の Web 上の言語資源から得た新語を追加することにより拡張した MeCab 用のシステム辞書である。

学習器は Logistic Regression, SVM, Gaussian Naive Bayes, k-Nearest Neighbors, Random Forest, AdaBoost, Decision Tree の7つを用いて、結果を比較する。SVM については rbf, linear の2種類のカーネルを用いる。以後、この分類手法を  $M$  とし、学習器とシステム辞書を含めて示す際には  $M$ -**学習器 (システム辞書)** として表すこととする。例えば学習器には SVM, カーネルは rbf, システム辞書は ipadic を用いた場合、 $M$ -SVM(rbf)(ipadic) と表記する。

### 4.4.3 word2vec 用いた質問文拡張

ユーザーはスマートフォンで質問の投稿を行うため、質問文が短文であることが多い。そのため、投稿文中の単語だけでは、文の特徴を表現するには不十分であると考えた。

そこで、前節の方法において特徴語として選択した単語に対して word2vec[14] を用いて、word2vec における類似度が最も高い10単語を特徴語として追加することにより、質問文拡張を行うことで文の特徴を表現する情報量を増やすことにした。word2vec のモデル作成のために、ママりに投稿された約100万件の質

## 4.5. 実験

---

問（以下，mamari コーパスと呼ぶ）と wikipedia で提供されているダンプデータ [15] をコーパス（以下，wikipedia コーパスと呼ぶ）としてそれぞれ利用し，比較することにした．具体的に前節の方法に 2 つ工程を加えた．以下に流れを示す．

1. コーパスから word2vec のモデルを作成
2. 形態素解析プログラム（MeCab 等）を用いて質問文から単語を切り出す
3. 2 の単語から特徴量として用いる品詞として，名詞，動詞，形容詞，記号を抽出
4. 選択された単語を元に word2vec で単語を拡充
5. 質問文毎に単語をカウントして特徴ベクトルを作成
6. 学習器を構築

以後，word2vec 用いた質問文拡張を  $\mathbf{W}$  (利用コーパス名) として表すこととする．

## 4.5 実験

### 4.5.1 データセット

4.3 で述べた基準に則り，2 人の判別者により手作業で共感を求める質問を分類し，ラベル付けを行った．

#### (1) ラベル付けの方法

ランダムに抽出した 1,600 件の質問に対して以下の 4 段階で表 4.4 左のラベル付けを行った．

Definitely Yes (DY) は判別者が自信を持って共感を求める質問であると判断した際に用いられる．Probably Yes (PY) は DY よりも自信がないが共感を求める質問であろうと判断した際に用いられる．逆に自信を持って共感を求める質問でないと判断した場合には Definitely No (DN) とラベル付けし，DN よりも自信がない場合には Probably No (PN) を用いる．判別者 2 人の組み合わせに応じたラベル付けが表 4.4 右である．表中の Y は共感を求める質問として分類し，N は共感以外をを求める質問と分類する．D となっている部分に関してはラベル付けの段階で分類不可能と判断し，データセットから除いた．

表 4.4: 4段階の評価によるラベル付け

DY: Definitely Yes	DY	PY	PN	DN
PY: Probably Yes	Y	Y	Y	D
PN: Probably No	Y	Y	D	N
DN: Definitely No	Y	D	N	N
	D	N	N	N

## (2) ラベル付けの結果

Yと判別された質問数は311, Nと判別された質問数は1,206であった。ラベル付けの段階で分類不可能と判断し、データセットから除いた質問は83であった。そのうちPY/PNの組み合わせは79件, DY/DNの組み合わせは4件だった。判別者2人の判別が合わなかった理由として、文意によっては共感を求めているとも、意見を求めているともとれる質問があったことが挙げられる。

ラベル付けに際して、判別者2名によるラベル付けの結果の一致度をみるためにkappa係数[16]を用いた。その結果0.88と高い一致が見られることがわかった。

以上より、1,517件の質問を実験に用いる。1,017件を教師データ, 500件をテストデータとする。1,517件のうち、311件(約2割)が共感を求める質問である。311件のうち211件は教師データ, 100件はテストデータに利用する。

### 4.5.2 実験手法

4.4で述べた3つの分類手法(ベースライン, 学習器, word2vecによる単語拡充)を用いて分類を行い, 分類の精度を比較する。実装にはscikit-learnを利用した。以下の項目について比較を行う。

- 教師データのクラス毎の割合を調節することで, 精度に差が出るか検証
- MeCabを用いて質問文から単語を切り出す際に利用される, システム辞書についての精度の比較
- 各学習器毎の精度の比較
- word2vec コーパスの精度の比較

### 4.5.3 目標とすべき指標数値の定義

共感を求める質問を正しく判別していることを定量的に計るため、precision, recall, F1 を利用 (3.4.3 を参照) する。共感を求める質問は全質問の約 2 割を占めているが、表 4.1 に示す通り、未回答で終わってしまう質問の数が多い。しかし、質問者を満足させるコミュニティを実現するためには、これらの質問にも回答が与えられるべきである。そこで、可能な限り、共感を求める質問を回答する傾向が、相対的により高い出産経験者 (表 4.2) に重点的に表示させることが共感を求める質問の回答率を上げることに繋がる。

しかし、出産を経験したユーザーの数は限られている。仮に出産を経験したユーザーにすべての質問を割り振った場合、回答候補者 1 人当たりの質問数が多くなるため、回答する頻度や質が低下する恐れがある。しかしそれを実測する方法はないため、本研究では 1 人あたりが回答することのできる質問の許容量を現在の回答数とする。ママリにおいては、ユーザーの半数が出産を経験したユーザーである。共感を求める質問が全質問の 2 割であることから、共感を求める質問の recall を限りなく 1 に近づけた上で、precision を 0.4 に近づけることで、回答候補者 1 人当たりの質問数を今と同じ水準に保ったまま、出産を経験したユーザーへ、共感を求める質問を割り振ることができる。これらの値を本実験における目標とすべき値とする。

### 4.5.4 実験結果

すべての組み合わせについて実験を行ったが、組み合わせの数が、学習器 (7 種類) × MeCab のシステム辞書 (2 種類) × 分類手法 (3 種類) × 教師となるデータのクラス毎の割合 (4 種類) より 168 種類あるため、実験結果の要約として目標値へ最も近づいた組み合わせをベースラインとともに表 4.5 に示す。ベースラインの結果は **B** と示す。**B** 以外の結果について、上段の評価値との差分を括弧内に示す。

表 4.5 より、ベースラインでは recall は 0.4, precision は 0.13 となった。一方で教師データのクラス毎の割合が 1:1, 辞書が mecab-ipadic-NEologd, 学習器として SVM(rbf) を利用し、word2vec(mamari コーパス) で質問文を拡張した場合に、recall は 0.96, precision は 0.39 となった。十分にベースラインを上回り、かつ目標の指標数値に非常に近い値となった。このとき、SVM で学習およびテストを行う際の特徴語の特徴ベクトルの次元数は 9,799 次元であった。次節以降で教師となるデータのクラス毎の割合, 辞書, 学習器, word2vec に利用したコーパスについて考察していく。



## 4.5. 実験

表 4.5: 教師データ中のクラス毎のサンプル数比を 1:1 にした際の、テストデータにおける共感を求める質問の分類結果 (**B**: ベースライン, **M**-学習器 (システム辞書): 4.4.2 の手法, **W**(コーパス): word2vec 用いた質問文拡張)

分類手法	Precision	Recall	F1
<b>B</b>	0.13	0.40	0.20
<b>M</b> -SVM(rbf)(ipadic)	0.30 (+0.17)	0.93 (+0.53)	0.46 (+0.26)
<b>M</b> -SVM(rbf)(ipadic) + <b>W</b> (mamari)	0.38 (+0.08)	0.94 (+0.01)	0.54 (+0.08)
<b>M</b> -SVM(rbf) (mecab-ipadic-NEologd) + <b>W</b> (mamari)	<b>0.39</b> (+0.01)	<b>0.96</b> (+0.02)	<b>0.55</b> (+0.01)

### (1) 教師データ中の各クラスのサンプル数比

教師データ 1,017 件のうち、約 2 割が共感を求める質問である。これはママリ内の実際の割合と同じであるが、教師データ中の各クラスのサンプル数比を適切に設定することにより、精度の向上が可能であると考えた。そこで教師データ中の共感以外を求める質問サンプル数を 200 件ずつ変化させた。すべての学習器について、4.4.2 で述べた手法を用いて実験を行った結果を図 4.2 に示す。

なお、MeCab のシステム辞書については、2 種類のシステム辞書による結果の平均値とした。縦軸は、共感を求める質問の recall である。横軸は、教師データの共感以外を求める質問の件数である。図 4.2 より、どの学習器においても、第一目標 (recall を 1.0 に近づける) を実現できているのは、その割合が 1:1 の場合であることがわかる。

教師データ中の共感以外を求める質問のサンプル数が 400 件以上のときに、SVM(rbf) のみ recall が 0 となっているが、これはすべての投稿を、共感以外を求める質問として分類してしまったためである。しかし、教師データ中の共感以外を求める質問のサンプル数を 200 件にすると、最も高い recall が得られた。4.4.3 で述べた word2vec を用いた手法でも同様の結果が得られた。また、どの辞書を用いても同様の結果が見られた。

以上の結果より、表 4.5 より得られた結果に加えて、以降の実験では教師データ中の各クラスのサンプル数比を 1:1 として実験を行う。

### (2) MeCab のシステム辞書についての考察

図 4.3 に 2 種類のシステム辞書 (ipadic, mecab-ipadic-NEologd) の結果を示す。

## 4.5. 実験

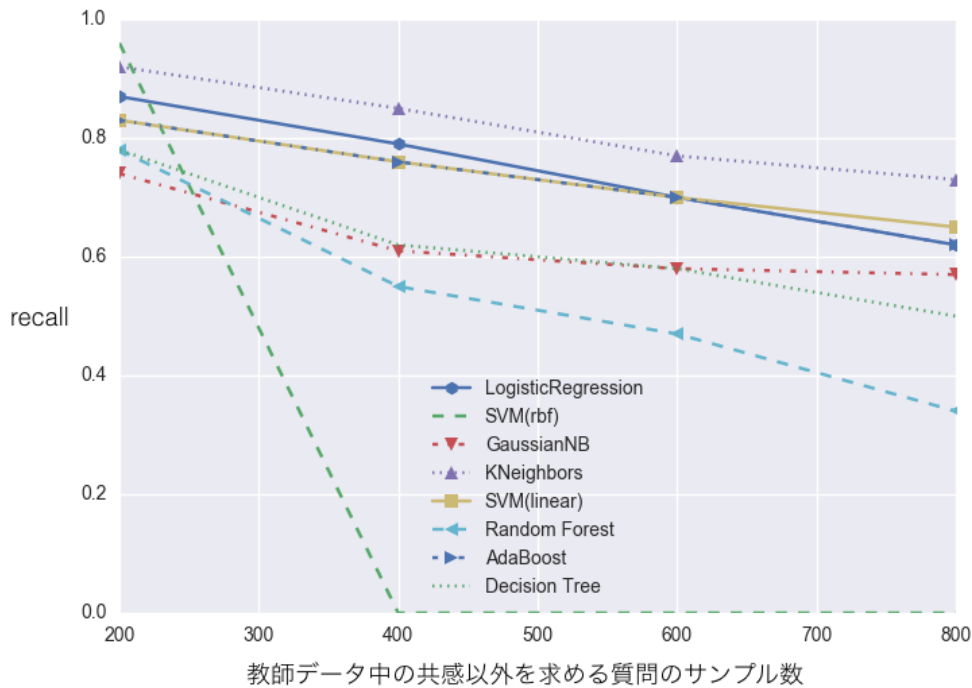


図 4.2: 教師データ中の共感以外を求める質問サンプル数を変化させたときの、テストデータにおける recall

ここでは、最も結果の良い  $\mathbf{M}$ -SVM(rbf) +  $\mathbf{W}$ (mamari) を用いた場合の結果を示す。word2vec の利用の有無に関わらず、7つの学習器と同様の傾向が見られた。precision, recall, F1 共に、mecab-ipadic-NEologdの方が良い結果が得られた。これは、3.4.5でも述べたように、スマートデバイスでQ&Aが行われることが関係している。

スマートデバイスにおける商用コミュニティサイトでは、絵文字や顔文字、Web上でよく使用されるスラングや、より一般的な会話の中で使われる単語等が利用される。従って、共感を求める質問の分類という課題についても、ipadicと比べてそれらの単語が登録されているmecab-ipadic-NEologdの方が良い結果になったと考えられる。

### (3) 各学習器についての考察

共感を求める質問の recall を限りなく1に近づけた上で、precisionを0.4に近づけるといふ今回の目標に沿い、行った分類実験結果の中で、recallが0.95以上となった組み合わせは、表4.5の  $\mathbf{M}$ -SVM(rbf)(mecab-ipadic-NEologd) +  $\mathbf{W}$ (mamari) という組み合わせのみであった。

図4.4に分類手法毎の recall を、図4.5に分類手法毎の precision を、図4.6に分

## 4.5. 実験

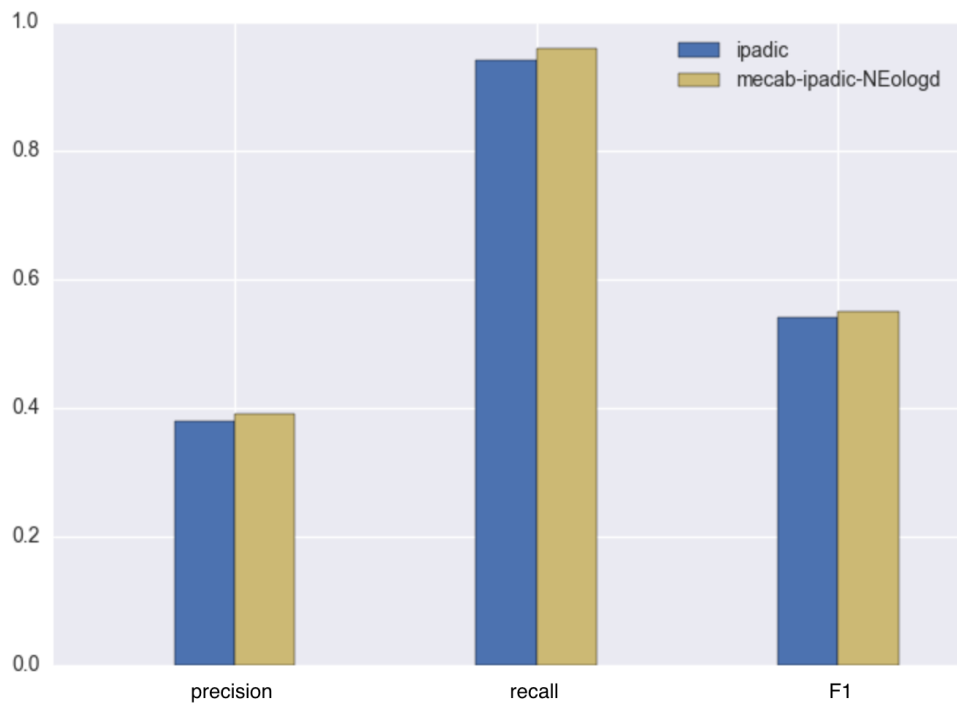


図 4.3: MeCab のシステム辞書における評価値の比較

類手法毎の F1 を示す。図 4.4 から、学習器としては、SVM(rbf) が目標値に近い recall で分類を行うことができ、今回の目標に適した学習器であることがわかった。また、k-Nearest Neighbors も SVM(rbf) に次いで 0.9 を超える高い recall 値を持つことがわかる。

ここで今後、本研究手法を実際のプロダクトへ適用した場合を想定する。共感を求める質問に対して出産を経験したユーザーに回答してもらえよう、質問を振り分けた結果、出産を経験したユーザー 1 人あたりの回答することのできる質問の許容量が、現在よりも小さくなる可能性がある。その場合には、図 4.5 より、他の学習器よりも precision が大きい、Logistic Regression, SVM(linear), Random Forest が学習器として有用であることがわかる。

## 4.5. 実験

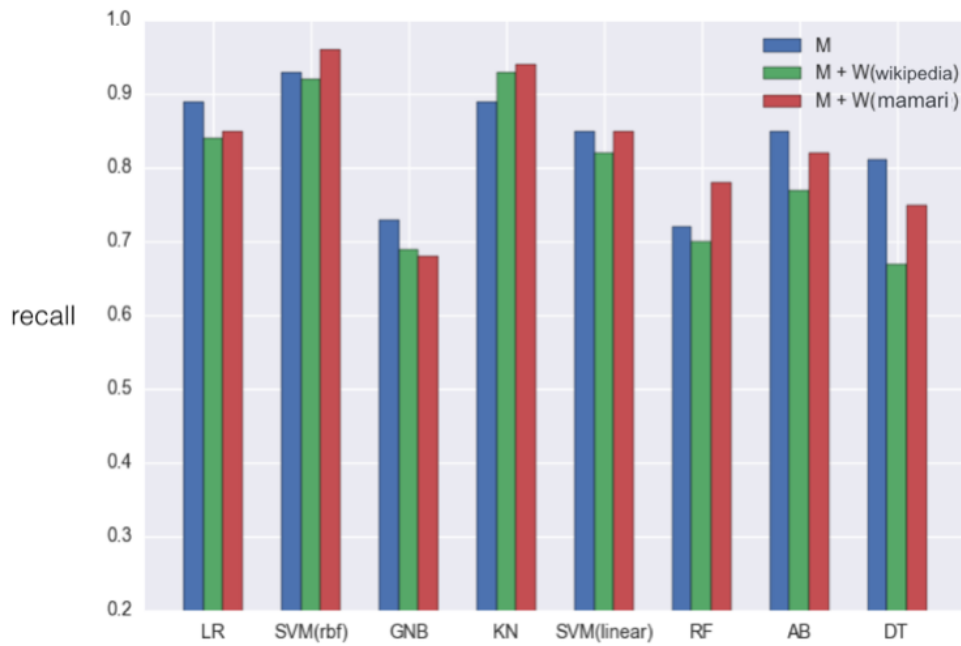


図 4.4: 各分類手法における recall (LR: Logistic Regression, GNB: Gaussian Naive Bayes, KN: k-Nearest Neighbors, RF: Random Forest, AB: AdaBoost, DT: Decision Tree)

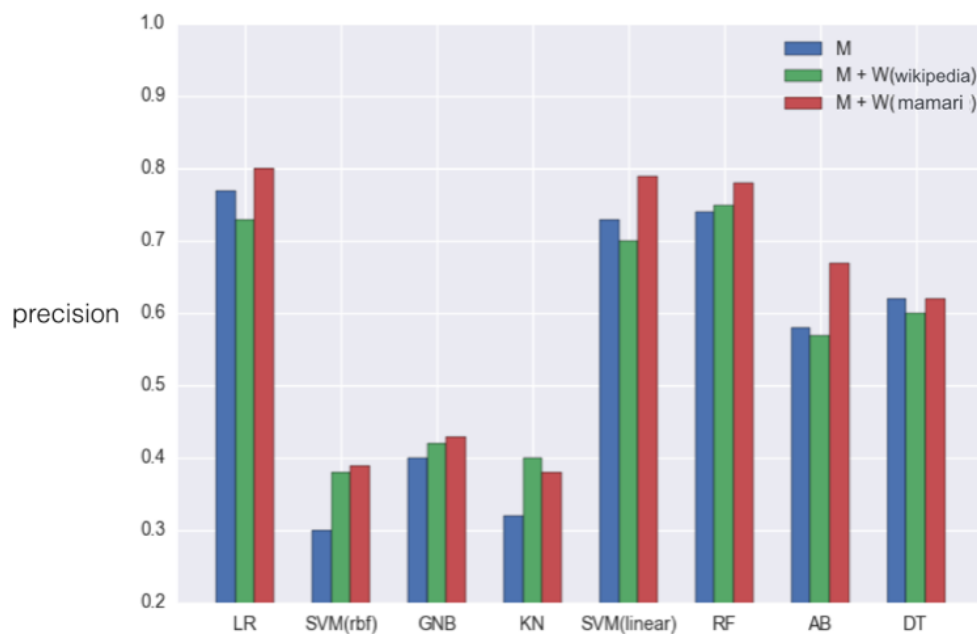


図 4.5: 各分類手法における precision (LR: Logistic Regression, GNB: Gaussian Naive Bayes, KN: k-Nearest Neighbors, RF: Random Forest, AB: AdaBoost, DT: Decision Tree)

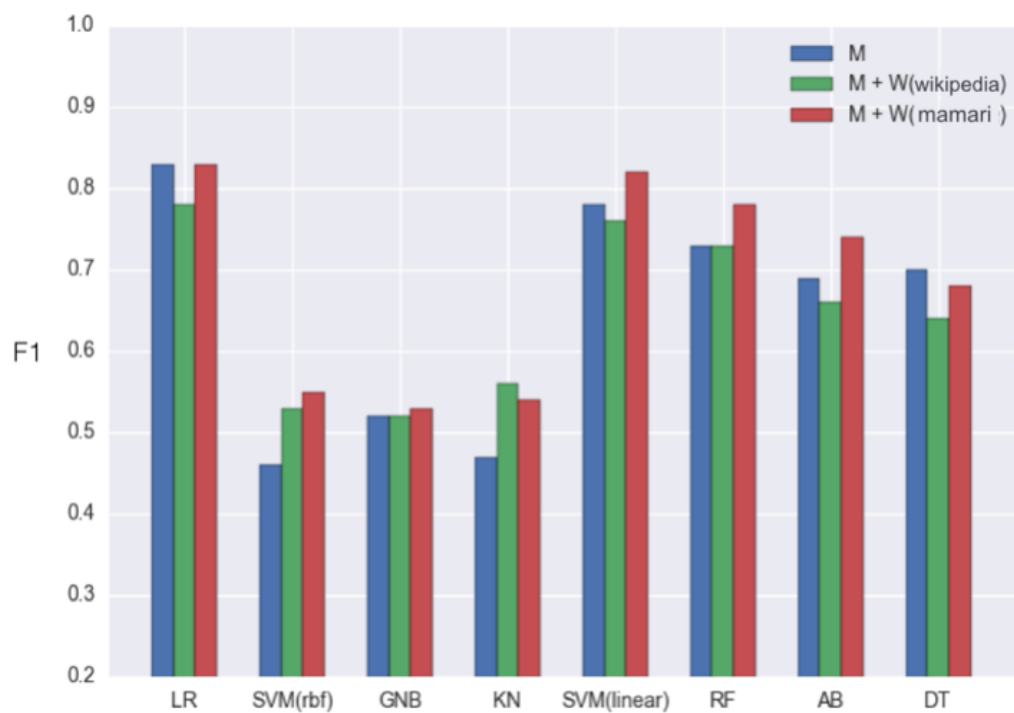


図 4.6: 各分類手法における F1 (LR: Logistic Regression, GNB: Gaussian Naive Bayes, KN: k-Nearest Neighbors, RF: Random Forest, AB:AdaBoost, DT:Decision Tree)

## (4) word2vec での質問文拡張と拡張に利用されたコーパスについての考察

図 4.4, 図 4.5, 図 4.6 から分類手法について, 各学習器毎の結果を比較した場合, 24 個中 17 つで mamari コーパスを用いた質問文拡張を行った手法が, 他の手法よりも良い結果となっている。

逆に, wikipedia コーパスを用いて質問文拡張を行った手法では, 拡張を行っていない手法と比較して, 24 個中 17 個が同じかそれ以下の値となっており, 性能が低下している。この理由として, コーパスの影響が考えられる。

word2vec を用いて, 「つわり」に類似度の高い 10 個の単語 (拡張単語) を, 例として表 4.6 に示す。wikipedia コーパスには, ユーザーが普段コミュニティ QA サイトの中で用いることが少ない「糖尿病性ケトアシドーシ」といった病名が含まれている。これは wikipedia に存在する医学的な記述などから得られたと考えられる。

他方, mamari コーパスを用いた場合では, つわりの表記揺れ単語や, それに伴う症状が列挙されており, 質問文拡張として適切な単語が並んでいると言える。これらより, 次のことが言える。(1) wikipedia コーパスを用いた場合には, 不適切な単語を用いた質問文拡張となり, 質問文拡張を行っていない 4.4.2 の手法よりも性能が低くなった。(2) mamari コーパスを用いた場合には, 適切な単語が用いられ, 他の手法よりも性能が高くなった。

表 4.6: 「つわり」の拡張単語

rank	wikipedia	mamari
1	便秘	悪阻
2	高血圧	ツワリ
3	脂肪肝	づわりが
4	更年期障害	胸焼け
5	糖尿病性ケトアシドーシ	胃痛
6	膀胱炎	吐き気
7	けいれん	胸やけ
8	アトピー性皮膚炎	づわりも
9	心因性	づわり
10	骨粗鬆症	乗り物酔い

## 4.6 おわりに

本章ではコミュニティQAサイトにおいて、回答が有意に少ないとされる共感を求める質問の回答率を上げることを目的として、質問文から抽出した特徴から機械学習を用いて、共感を求める質問と共感以外を求める質問の分類を行った。

7種類の学習器、形態素解析プログラム (MeCab) を用いて質問文から単語を切り出す際に利用されるシステム辞書、word2vecによる質問文拡張、さらにこれらの手法の組み合わせについても考察を行い、最も精度の良い分類手法を調べた。

その結果、特徴語句による分類と比べて大幅に良い分類結果を得ることができた。今後は実際にプロダクトに組み込み、ユーザーへの影響を見ながら、回答率を上昇させることでユーザーの満足度を向上させることに取り組んでいきたいと考えている。

# 第5章 コミュニティQAサイトにおける質問テキストのみを用いたユーザー属性の抽出

本章では、ユーザーが投稿した質問から、機械学習を用いてユーザー属性を予測する方法について述べる。

## 5.1 はじめに

多くのオンラインサイトにはユーザー属性の登録機能があり、コミュニティQAサイトでも自分のユーザー属性を入力できるサイトが多い。質問者としては、ユーザー属性は質問に対する良い補助情報となり、回答者に適切な回答を促すことにつながる。また回答者としては、ユーザー属性は回答への付加情報として、自身の属性を明示することができる。このような効用を意識して、能動的にユーザー属性の入力を行うユーザーも一定数存在している。

サービス提供側としては、ユーザーが入力をしたユーザー属性を用いることにより、上記のように、質問者となる場合、回答者となる場合のいずれにおいても、ユーザー自身が求める情報のマッチングが可能になるため、より良い質問ルーティングを実現することが可能である。

しかし、ユーザーがこのようなユーザー属性を入力するには、入力の面倒さが障壁となっている。さらに、スマートデバイスの普及により、入退会が容易になった。そのため、登録したユーザー属性を今までよりも容易に削除することが可能になった。

そこで、より良い質問ルーティングを実現するために、ユーザー自身がユーザー属性を入力せずとも、ユーザーが投稿した質問から機械学習を用いてユーザーの属性を予測する手法の考案に取り組む必要がある。本研究においては、コミュニティQAサイトにおける質問テキストのみを用いた、ユーザー属性の抽出の達成を目指す。

本章の構成を以下に示す。5.2節では、ユーザー属性の活用と取得に関する課



題について述べる。5.3節では、ユーザー属性の取得のための、質問文の特徴について分析した結果と考察について述べる。5.4節では、提案手法について述べる。5.5節では、評価実験の内容と結果を述べ、結果を踏まえた考察を述べる。最後に5.6節で本章のまとめと今後の課題について述べる。

## 5.2 ユーザー属性の活用と取得に関する課題

### 5.2.1 ユーザー属性の活用

第3章や第4章では、より良い質問ルーティングを実現するために、質問を分類する手法を提案した。一方、本章では質問文からのユーザーの属性の抽出に取り組む。なぜなら、質問ルーティングの実現においては、質問の分類とユーザー属性の抽出という2つの課題に取り組む必要があるからである。

加えて、ユーザー属性を用いることで、ユーザーの満足度を向上させる施策や、パーソナライズされた広告の適切な配信も可能になる。特に結婚や出産と言った、人の生涯で発生する重要な出来事（ライフイベント）を起点として、様々な商品・サービスを必要とすることが知られている。

広告配信の例として、例えば出産したことがわかれば生命保険や学資保険に関する広告を配信するということが考えられる。また、出産からある程度時間が経過したユーザーには幼児教育に関するコンテンツを提供することで、ユーザーの潜在ニーズに応え、より長くサービスを利用してもらうことに繋げることができる。

妊娠して経過した月数や週数によって行われるコンテンツのレコメンデーションの具体的な画面を図5.1に示す。このようにユーザー属性には多くの活用先がある。

### 5.2.2 ユーザー属性の取得に関する、コミュニティQAサイトの課題

ユーザーが入力をした属性を用いることにより、ユーザー自身が求める情報コンテンツを受け取ることが可能になるが、ユーザーが自身の属性を入力するには、プライバシーの乱用に関する懸念といった精神的なハードル、入力の面倒さといった障壁がある。スマートデバイスを用いることにより、容易に入退会ができ、登録したユーザー属性を今までよりも容易に消すことが可能になったため、ユーザー属性を活用しにくいという課題がある。



図 5.1: 妊娠週数に合わせたコンテンツやレコメンデーションの例

### 5.3. 質問文の特徴分析

---

一方、ユーザー属性を登録しないユーザーでも、質問文中には前提条件を説明するために、ユーザー属性を記載する機会が多いことがわかった。そこで、より良い質問ルーティングを実現するために、ユーザー自身がユーザー属性を入力せずとも、ユーザーが投稿した質問から機械学習を用いてユーザーの属性を予測する手法を考案し、コミュニティQAサイトにおける質問テキストのみを用いたユーザー属性の抽出の実現を目指す。

なお、5.5.3では、具体的な質問の分類結果について考察しているが、プライバシー配慮のために個人を特定できないよう、本章で紹介する質問本文については一部文章の修正を行っている。

## 5.3 質問文の特徴分析

### 5.3.1 目的

本節では、コミュニティQAサイトにおける質問文において、どのような質問文であれば投稿者の正しいユーザー属性を把握することができるか、分析を行う。その上で、課題があればそれらを明らかにし、どのようなアルゴリズム・前処理等が有効かを検討する。

### 5.3.2 対象とするデータ

本研究においては、出産予定日をユーザー属性として、抽出することを目的とする。対象とするユーザーは、自分の子どもの生年月日をユーザー属性として、入力したユーザーとする。そして、それらのユーザーが出産以前に投稿した質問を対象データとして扱う。ユーザーの子どもの生年月日から、妊娠経過週を割り出すことにより、各質問の投稿時の妊娠経過週がわかる。これを教師ラベルとする。対象ユーザー数は1,081人、対象ユーザーが妊娠したときから出産までにした、対象の質問数は21,886件である。

### 5.3.3 問題設定

#### (1) 目標とする指標の定義

本研究ではユーザーの妊娠経過週とともに、時系列で、どの程度正確なユーザーの妊娠経過週の予測ができるのかを指標とする。ここで言う妊娠経過週とは、ユーザーの子どもの出産日から280日を引いた日を起算日とした、経過週数

### 5.3. 質問文の特徴分析

---

である。この妊娠経過週を「出産日を起算日とした妊娠経過週」とする。従って、「出産日を起算日とした妊娠経過週」が正しく予測できれば、実際の出産日も算出することが可能である。「出産日を起算日とした妊娠経過週」毎に予測を行うのは、早い段階で「出産日を起算日とした妊娠経過週」を予測できた方が、その分長い期間、ユーザーニーズにマッチしたコンテンツや広告等を提供できるためである。

また、実際の出産日を起算日とした値ではなく、最終月経の初日から数えて280日目を出産予定日と言う。この出産予定日から、280日を引いた日、すなわち、最終月経の初日を起算日とした、妊娠経過週を「予定日を起算日とした妊娠経過週」とする。

#### (2) 正期産を考慮した、ユーザーの出産日を起算日とした妊娠経過週予測

予定日を起算日とした妊娠経過週が37週0日から41週6日の間にお産があることを「正期産 [52]」と言う。多くの人は正期産を迎える。この期間に生まれた新生児は、体の各機能も十分に成熟しているので、母体の外での生活にもスムーズに適応することができる。

正期産を考慮し、予測結果の妊娠経過週の3週間前から、予測結果週の妊娠経過週の1週間と6日後の期間を、出産日を起算日とした妊娠経過週の予測結果として、正しい予測結果であるとする。

具体的な例として、例えば質問投稿者の出産日を起算日とした妊娠経過週が39周目でユーザーが質問の投稿を行ったとする。この場合、正期産を考慮し、36周目から40周と6日の間の週数であると、出産日を起算日とした妊娠経過週が、この質問から予測されれば、出産日を起算日とした妊娠経過週の予測結果としては正しい予測結果であるとする。

#### (3) 対象ユーザー率と正解率の定義

5.3.2で述べた、対象とするデータ内の全ユーザー数に対して、出産日を起算日とした妊娠経過週の予測を行ったユーザー数を「対象ユーザー数」とし、データ内の全ユーザーに対する割合を「対象ユーザー率」と定義する。対象ユーザー率が100%とならない具体的な事例については5.3.5で述べる。また、対象ユーザー数に対して、5.3.3(2)で述べた、正期産を考慮したユーザーの、出産日を起算日とした妊娠経過週が、正しく行われたユーザー数の割合を「正解率」と定義する。

#### 5.3.4 特徴の分析方法

まずは質問文を目視で確認し、特徴の分析を行う。その中で、ユーザーの出産日予測に有効と考えられる方法について仮説を立て、実際にその方法で予測が可能かどうかを確認する。

#### 5.3.5 分析結果

質問を分析すると、具体的に自身の妊娠週を記し、それに応じた内容を回答者に尋ねている質問が存在することがわかった。例えば以下のような質問である。

「私は現在4wの初妊婦です。最近妊娠がわかったのですが、つわりがとてもひどいです。皆さんはどうやってつわりを乗り越えましたか？」

質問文内の「4w」とは、妊娠を話題としたテキストコミュニケーションの中で、予定日を起算日とした妊娠経過週が4週であることを示すために、ユーザーが使う表現である。よってこの例の場合は、このユーザーは妊娠4週であると判断することが可能である。このように自身の妊娠週が質問文に含まれている場合は、正しくユーザーの出産日を起算日とした妊娠経過週を予測できる可能性が高いと考えた。

以上より < 数値 > [w|週] という組み合わせが質問文内にある場合に正規表現を用いてその数値を週数として抜き出し、抽出した週数を出産日を起算日とした妊娠経過週とした。

妊娠してから出産に至るまで、ユーザーは週数に応じた質問をコミュニティの中で尋ねる。そして、多くの場合、出産までの間にユーザーは複数回質問を行う。ユーザー1人に対しての出産日を起算日とした妊娠経過週の予測は、予測する時点までに投稿されていた質問を用いることが可能である。

本節では、予測時点で対象となっている質問の中で、最も新しい質問を用いて予測することにした。結果を図5.2に示す。なお、各プロット上部の数値は出産日を起算日とした妊娠経過週を表す。

図5.2より、40週目時点で対象ユーザー率は95%以上となった。100%となっていないのは、ユーザーの中には、< 数値 > [w|週] という文字列が質問文内に含まれた投稿を一度もせずに、出産を迎えるユーザーも存在しているからである。図5.2を見ると正解率は80%前後であった。これは自身の予定日を起算日とした妊娠経過週以外の週を記述するために、< 数値 > [w|週] という文字列を使用しているためだと考えられる。具体的な質問文を下記に示す。

### 5.3. 質問文の特徴分析

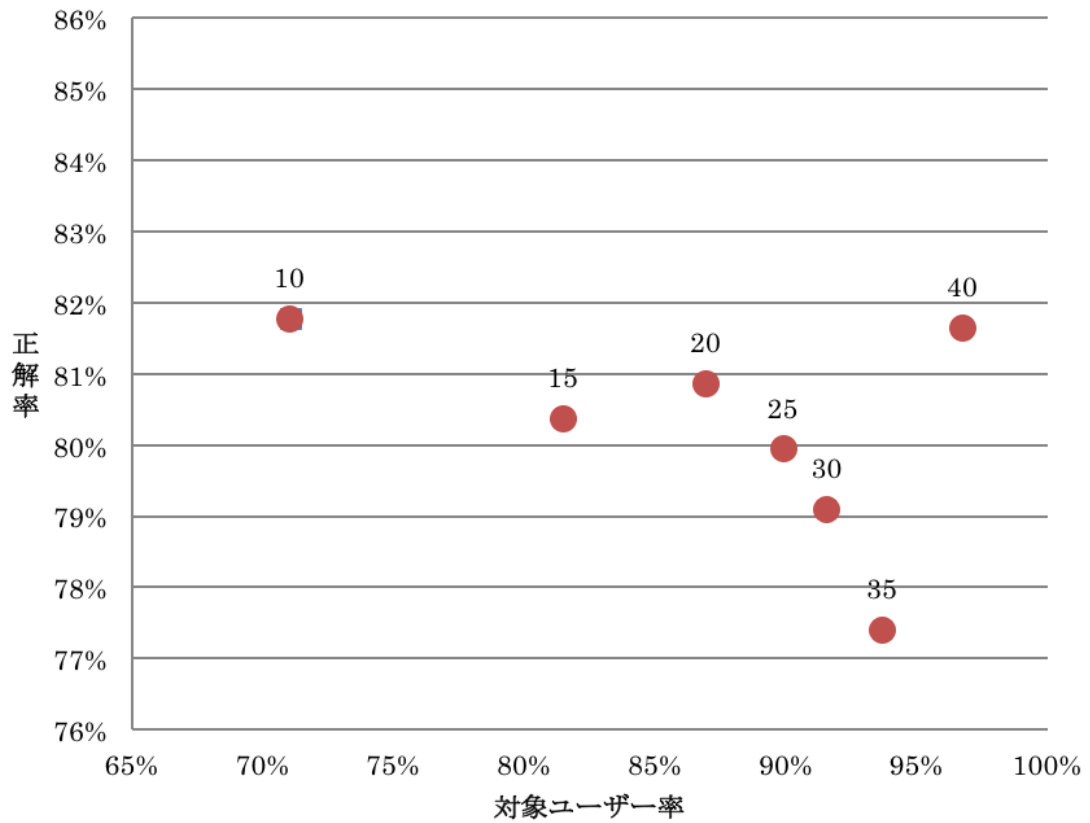


図 5.2: < 数値 > [w 週] という特徴語句を持つ質問から出産日を起算日とした妊娠経過週を抜き出し予測した正解率と対象ユーザー率 (プロット上部の数値は出産日を起算日とした妊娠経過週)

## 5.4. 実システムへの導入のための判別手法の提案

---

「今月4週目の土曜日に家族でディズニーランドに行く予定です！皆さんはもうハロウィンイベントに行かれましたか？」

といった文である。そこで、正解率を上げるために、4w2dと言った形で< 数値 > [w|週]< 数値 > [d|日]という表現を持つ質問に絞って週数を抜き出すことを試みた。「4w2d」とは、妊娠を話題としたテキストコミュニケーションの中で妊娠経過期間が4週2日であることを示すためにユーザーが使う表現である。

図5.3に抜き出した結果をまとめた。ここでは、各週数目における< 数値 > [w|週]を●とし、< 数値 > [w|週]< 数値 > [d|日]を■としてプロットしている。それぞれの組み合わせから週数を抜き出し、予測した正解率と対象ユーザー率を示した。各週数目における●から■へ矢印を引き、それぞれの週数における改善を示した。

正解率が上昇したことから、日単位までの情報を入れることで自身の状態について詳細に述べている質問文から週数を抽出する方が、より正確に抽出できることがわかる。一方で週数の抽出条件を厳しくしたために、40週時点での対象ユーザー率は7割程度となった。ここで対象ユーザー率よりも正解率を優先して考えることに関しては5.4.1で解説を行う。

## 5.4 実システムへの導入のための判別手法の提案

### 5.4.1 目標とすべき指標数値の定義

5.3.5の手法により、正解率としては90%を超える結果となった。しかし、それでも誤判別は約1割存在している。ママリの会員登録数は約100万人以上である。従って、この正解率でシステムを構築し、導入すると約10万人以上のユーザーには望まないコンテンツが表示され、ユーザーの不利益になる可能性がある。そこで、この正解率を可能な限り100%に近づけることを、目標とすべき指標数値とする。

### 5.4.2 提案手法の概要

5.4.1で定めた目標とすべき指標に近づくために、機械学習を用いる。5.3.3で述べた手法を行った後、出産日を起算日とした妊娠経過週を、質問文から特徴語句を用いて抽出する前に、機械学習を用いてフィルタリングを行うことにした。

#### 5.4. 実システムへの導入のための判別手法の提案

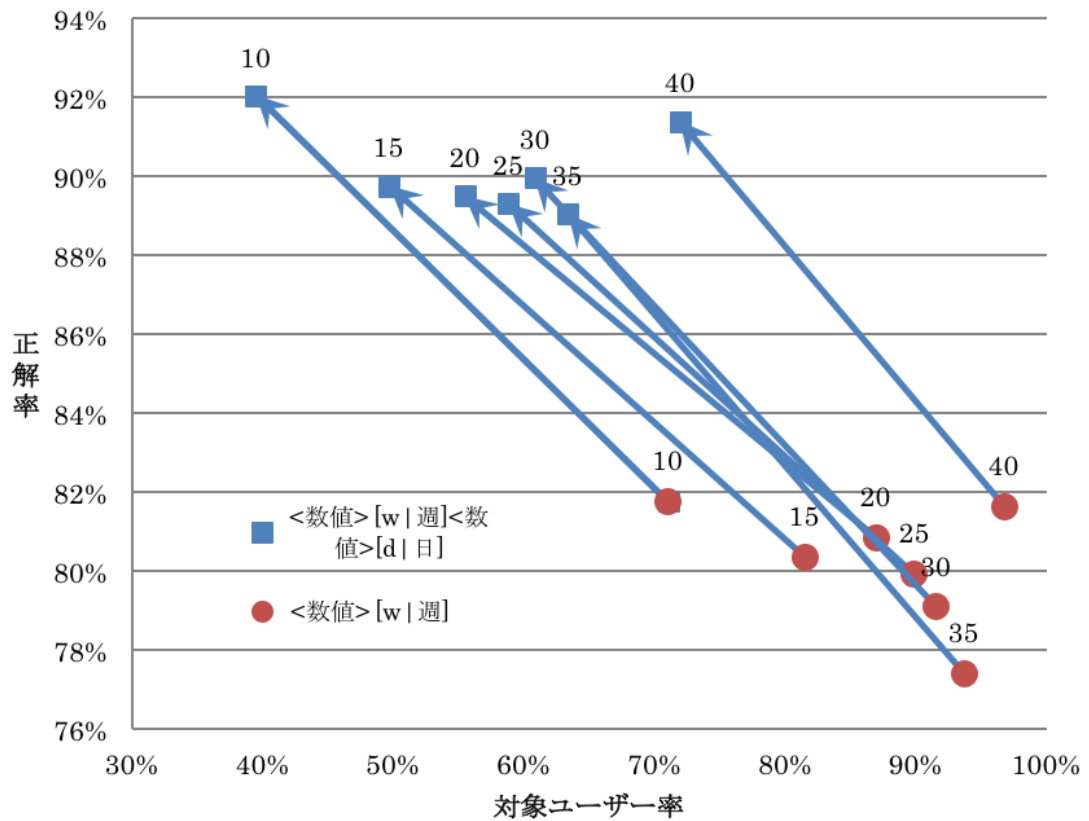


図 5.3: <数値> [w | 週] と <数値> [w | 週]<数値> [d | 日] それぞれの特徴語句を持つ質問から出産日を起算日とした妊娠経過週を抜き出し予測した正解率と対象ユーザー率（プロット上の数値は出産日を起算日とした妊娠経過週）



## 5.4. 実システムへの導入のための判別手法の提案

特徴語句を含んだ質問に対して、週数の推定として対象の質問が利用できるかどうかを機械学習によって分類する。提案手法のフローチャートを図5.4に示す。

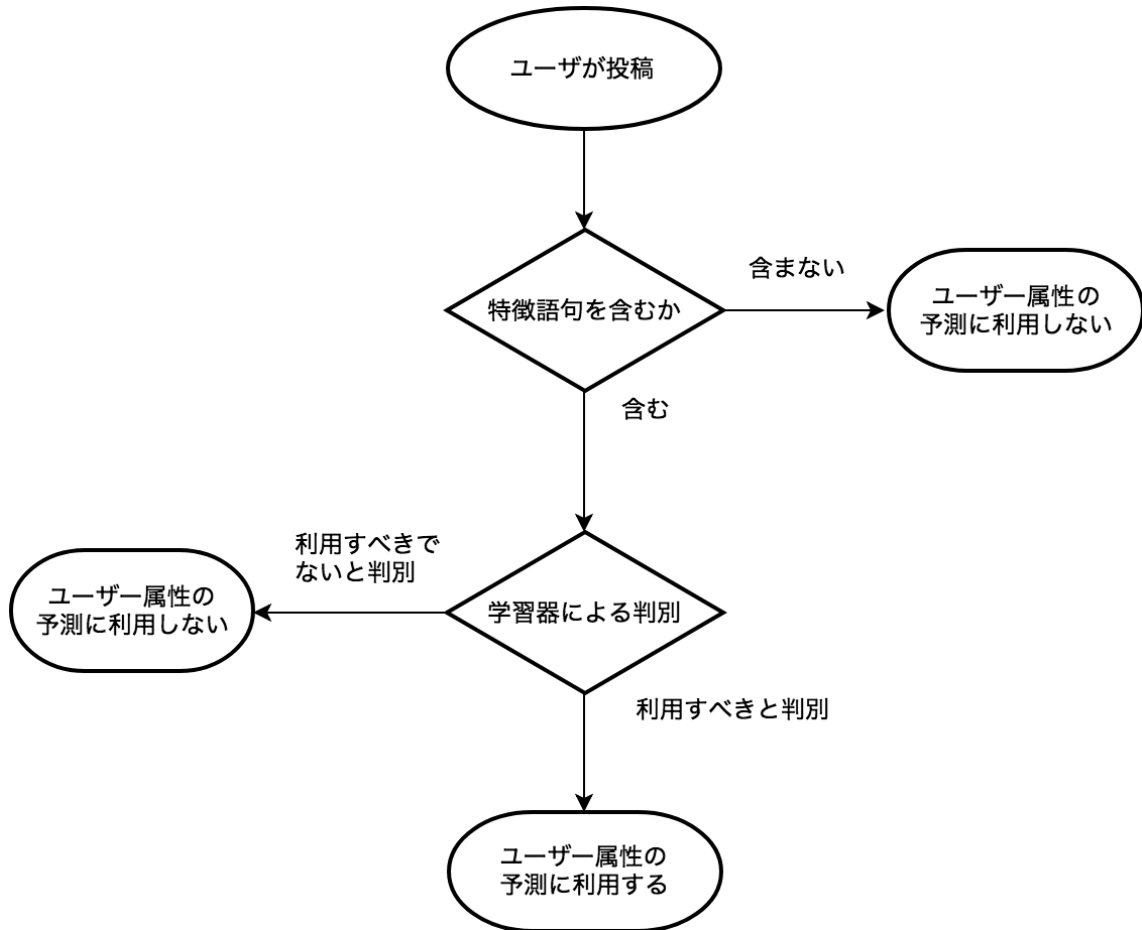


図 5.4: 提案手法のフローチャート

### 5.4.3 機械学習を用いた分類

5.3章で述べた特徴語句による質問文からの誕生日予測を行うべきかどうかを質問毎に機械学習を用いて分類する。自然言語処理における分類手法にはいくつかの手法が存在するが、下記に示す方法を用いて、質問からの誕生日予測を行うべきか判別する。

1. 形態素解析プログラム (MeCab 等) を用いて質問文から単語を切り出す

## 5.5. 実験

---

2. 1の単語から特徴量として用いる品詞として、名詞、動詞、形容詞、記号と定め、この品詞の単語を抽出
3. 各単語を数値ベクトルへ変換する（これを単語ベクトルと呼ぶ）
4. 質問文毎に単語ベクトルの平均値を算出する
5. 4を特徴量として学習器を構築

未知の質問を分類する場合、同様に1から4の処理を行い、学習後の学習器を用いて、分類する。用いる単語の品詞としては、内容語である名詞、動詞、形容詞および、商用データ内では感情を表すものとしてよく使用されている記号を用いた。後者の例としては、「至急です！」中の「！」や「夜泣きで本当に悩んでいます。。」の「。」が挙げられる。

また、MeCabを用いて質問文から単語を切り出す際に利用されるシステム辞書にはmecab-ipadic-NEologdを利用する。

各単語を単語ベクトルへ変換する方法として、word2vec, Latent Semantic Indexing (LSI) [50], Latent Dirichlet Allocation (LDA) [51]の中で、最も適したものを採用することとした。なお、LSI, LDAを使用する際には、単語文書行列には、単語のtf-idf[53]値を用いた。

学習器には、linear, rbfカーネルを用いたSVMとRandom Forestの中で最も精度の高い学習器を採用する。

手順3で利用するword2vec, LSI, LDAそれぞれのモデル作成には、ママリに投稿された約100万件の質問をコーパスとし、学習データとして利用した（以下、ママリコーパスと呼ぶ）。word2vecのモデル作成において、特徴ベクトルの次元数は100、文脈学習時の前後の対象単語の幅は5とし、ママリコーパスに出現する頻度が5回未満の単語は使用しなかった。

また、LSIのモデル作成におけるトピック数は200、LDAのモデル作成におけるトピック数は100である。実装にはgensim[49]を利用した。学習器と単語ベクトルの生成方法の組み合わせについて実験を行い、性能を比較し検証する。

## 5.5 実験

### 5.5.1 データセット

5.3.2で述べたデータを用いる。商用データ内において、自分の子どもの生年月日を入力したユーザーの投稿データであるため、それぞれの質問には実際の出産

日を起算日とした妊娠経過週のラベルが付されている。データセットから 5.3.5 の分析の結果発見した、 $\langle \text{数値} \rangle [w | \text{週}] \langle \text{数値} \rangle [d | \text{日}]$  という表現を持つ質問が、本実験で利用するデータとなる。

$\langle \text{数値} \rangle [w | \text{週}] \langle \text{数値} \rangle [d | \text{日}]$  という表現を持つ質問の数は 5,198 件である。その内 5.3.3 (2) で述べた正期産を考慮した出産日を起算日とした妊娠経過週予測で、正しく週数を抜き出した質問の数は 4,686 件、誤って抜き出してしまった質問の数は 512 件である。質問文から抽出した名詞、動詞、形容詞、記号の数はそれぞれ名詞 590,065 語、動詞 297,273 語、形容詞 43,979 語、記号 244,440 語である。これらの質問に対して機械学習を用いて、出産日を起算日とした妊娠経過週の予測に利用するかどうかの分類を行う。

### 5.5.2 実験手法

5.4 章で述べた提案手法に従い、5.5.1 のデータセットに対して、機械学習を用いる。出産日を起算日とした妊娠経過週の予測に利用するかどうかの分類を行う。実装には scikit-learn を利用し、学習器それぞれのパラメータについては、学習器と単語ベクトルの生成方法の組み合わせ毎に学習データを用いてグリッドサーチを行い、最適な値を求めた。なお、以後の実験での評価は、すべて 10-fold cross validation で行った。

また、scikit-learn の SVC, RandomForestClassifier を用いて probability を出力し、probability がある値（以下、閾値）以上であれば出産日を起算日とした妊娠経過週の予測に利用し、そうでなければ利用しない。この閾値を調節することにより、出産日を起算日とした妊娠経過週予測の正解率を向上させることにした。probability を利用した理由は、同一人物が特定の出産日を起算日とした妊娠経過週において、予測の対象となる質問が、複数件見つかる場合があるからである。その場合は、特定の出産日を起算日とした妊娠経過週で、最も probability が高い質問を採用することで、正解率を上げることができると考えた。

なお、10-fold cross validation で、probability の閾値を学習データから動的に出力するために、学習データの中で 10-fold cross validation を再度行い、正解率が 100% になった閾値の平均値を採用することにした。probability の閾値の導出方法を図 5.5 に示す。

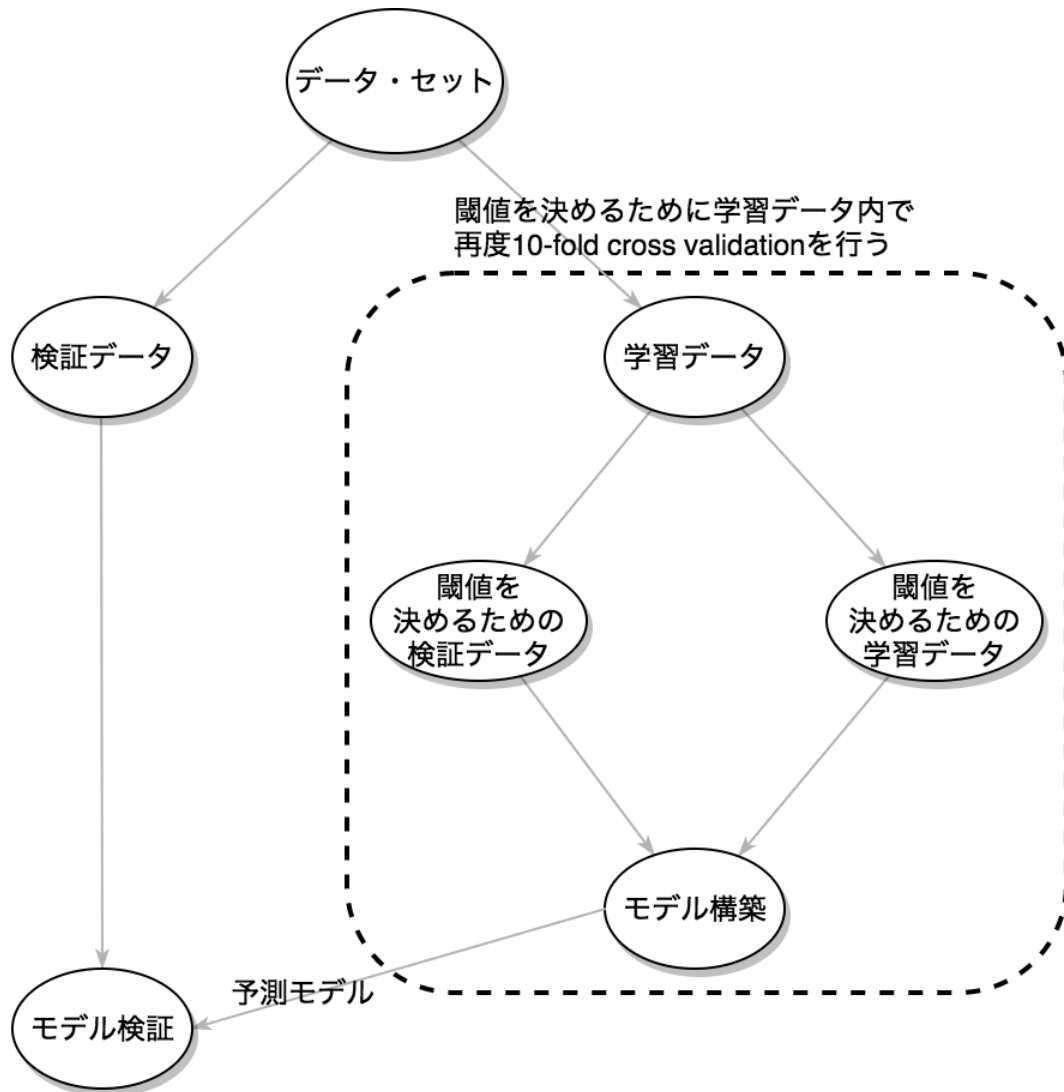


図 5.5: probability の閾値の導出方法

### 5.5.3 実験結果

図 5.6 に出産日を起算日とした妊娠経過週が 40 週時点の、各手法による正解率と対象ユーザー率を示す。図 5.6 中で、RandomForest は RF, word2vec は W という略称を用いた。

グリッドサーチによるパラメータ探索の結果、用いる学習器のパラメータを示す。

- SVM(rbf)-LDA:  $C = 1000$ ,  $gamma = 0.001$
- SVM(rbf)-LSI:  $C = 10$ ,  $gamma = 0.001$
- SVM(rbf)-W:  $C = 1000$ ,  $gamma = 0.0001$
- SVM(linear)-LDA:  $C = 1$
- SVM(linear)-LSI:  $C = 1$
- SVM(linear)-W:  $C = 10$
- RF-LDA:  $min\_samples\_leaf = 200$ ,  $n\_estimators = 2000$
- RF-LSI:  $min\_samples\_leaf = 200$ ,  $n\_estimators = 1000$
- RF-W:  $min\_samples\_leaf = 1000$ ,  $n\_estimators = 3000$

図 5.6 で最も正解率の高かった、RF-W (Random Forest と word2vec の組合せ) を提案手法として採用し、 $\langle$  数値  $\rangle [w]$  週  $\langle$  数値  $\rangle [d]$  日 の特徴語句を持つ質問から、出産日を起算日とした妊娠経過週を抜き出し予測した正解率と、対象ユーザー率の比較を図 5.7 に示す。

矢印の始点となっている■の点は、5.3.5 の手法による結果であり、 $\langle$  数値  $\rangle [w]$  週  $\langle$  数値  $\rangle [d]$  日 という表現を持つ質問に絞って、週数を抜き出した結果を示す。▲が提案手法の結果となっており、提案手法による精度の向上を矢印で示す。

提案手法は 40 週時点では正解率 95% を超える結果となった。また、妊娠 40 週時点での対象ユーザー率は約 20% となった。これは 5.4.1 で述べた通り、望まないコンテンツのレコメンデーションはユーザーにとって不利益となるため、対象ユーザー率よりも正解率を優先し、モデル作成の中で正解率が 100% になった閾値を採用した結果である。

提案手法により正しく妊娠週数を抽出できている質問を見ると、正期産でなかった兆候を質問文から判別できていたことがわかった。具体的には以下のような質問である。

「33w2d です。前回の検診では大きな問題はなかったのですが、先生からコメントいただいた点で気になったのは子宮頸管の長さが正常の中でも短い方と言われま

## 5.5. 実験

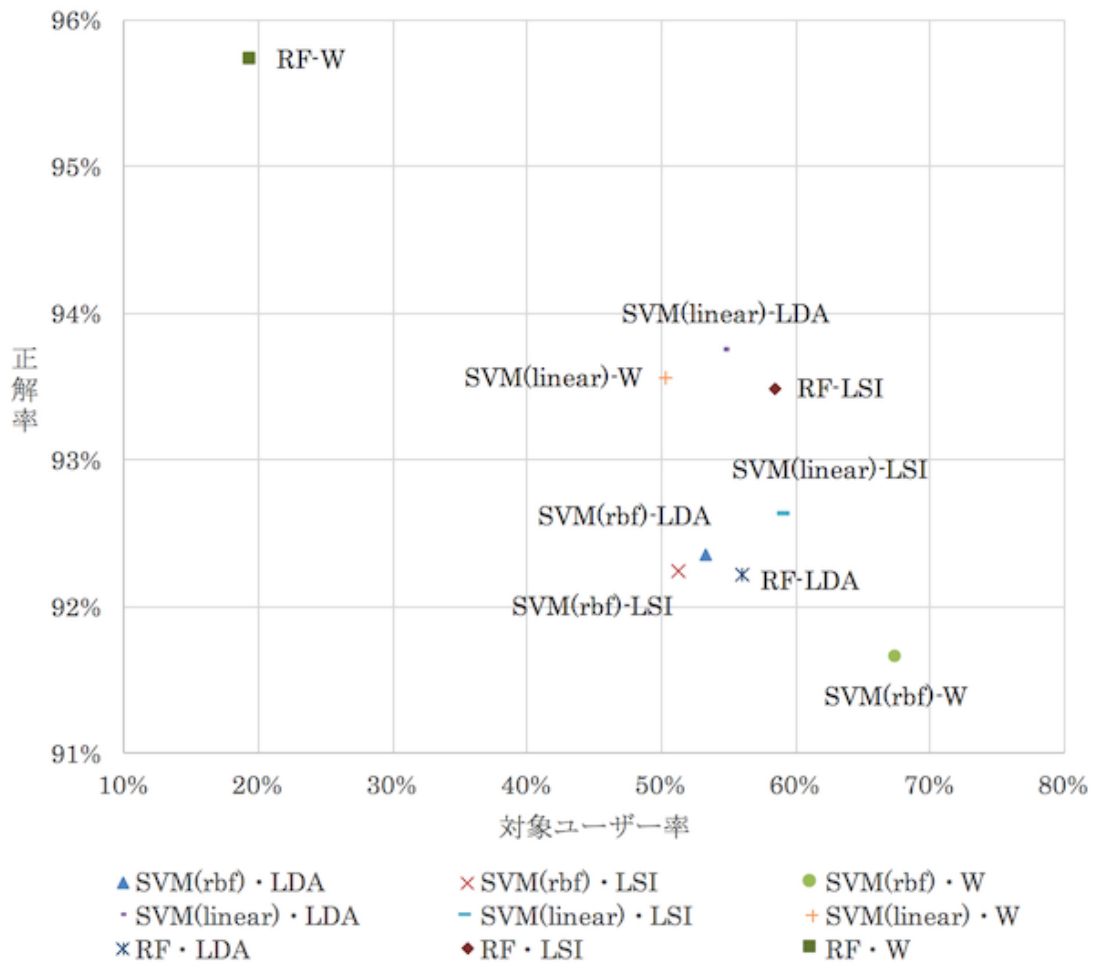


図 5.6: 学習器と単語ベクトルの生成方法の組み合わせによる，出産日を起算日とした妊娠経過週が40週時点の正解率と対象ユーザー率の比較（RF: Random-Forest, W: word2vec）

した。同じように言われた方はいらっしゃいますか？少し歩くだけでお腹が張ってしまうので，今は安静にしています。次の検診は明日です。今少し張りが強くて，病院に電話しようか迷っています。皆さんはどんな症状で病院に電話しましたか？もしできれば，同じような経験をされた方からコメントもらえると嬉しいです。」

このような質問を投稿したユーザーはこの投稿後の2週間後（35週目）で出産を迎えたので，早産の兆候を含む質問を的確に抽出できた例だと考えられる。

一方で提案手法でも誤判別をしてしまった質問を見ると，正期産でなく早産であったり，妊娠満42週以降の出産であったりするユーザーの質問であり，その

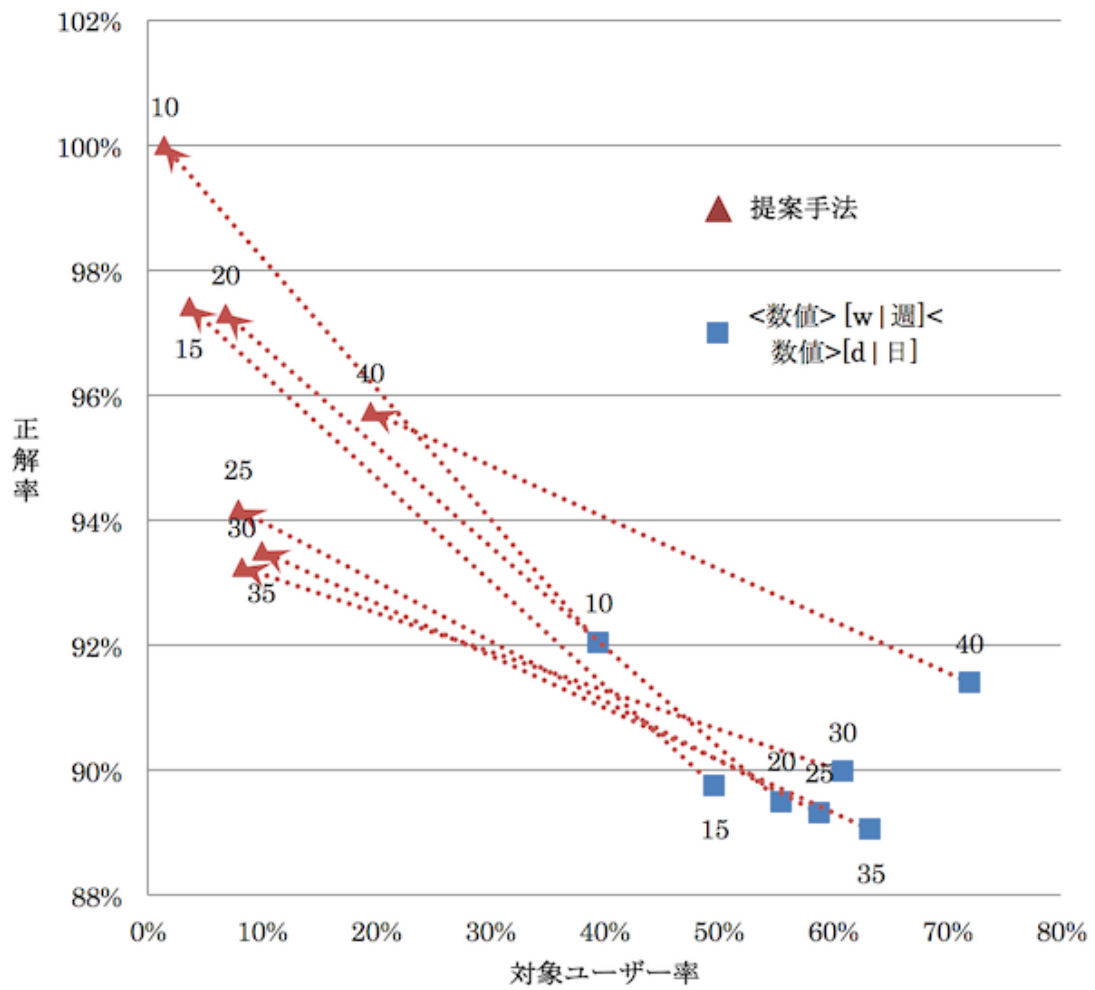


図 5.7: 提案手法と <数値> [w|週]<数値> [d|日] の特徴語句を持つ質問から出産日を起算日とした妊娠経過週を抜き出し予測した正解率と対象ユーザー率の比較

## 5.6. 終わりに

---

兆候は人が見ても質問文から判別しづらい質問であることがわかった。具体的には以下のような質問である。

「24w5dです！最近食欲がやばいです！とくに今週…!!! 元々スレンダーではない体系なのですが、頑張って、かなりいろいろ大変だったのですが、なんとか体重の増加は1.4程度におさまっています＼(^o^)/でも出産までまだ数ヶ月あるのが恐怖でなりません(T\_T;) 皆さんどうやって体重コントロールされましたか？」

このような投稿を行ったユーザーの中には44週目で出産を迎えているユーザーが存在している。正期産でないが、文章からその兆候が見られない。このような質問文の存在が、妊娠経過週によって100%を達成できなかった理由であると考えられる。

## 5.6 終わりに

本章ではコミュニティQAサイトにおいて、より良い質問ルーティングを実現するために、ユーザー自身がユーザー属性を入力せずとも、ユーザーが投稿した質問から、機械学習を用いてユーザー属性を予測する手法を考案した。その結果、コミュニティQAサイトにおける、質問テキストのみを用いた、ユーザー属性の抽出を実現した。

具体的には、質問文の特徴分析では、ユーザー属性の予測に役立つ要素表現を見つけ、検証を行った。さらに正解率を高めるために、機械学習を用いた事前の分類フィルタをユーザー属性の抽出処理フローに組み込んだ。その結果、95%を超える正解率でユーザー属性の予測を行うことができた。今後は実際にプロダクトに組み込み、ユーザーへの影響を見ながら、ユーザーの満足度を向上させることや、収益化に取り組んでいきたいと考えている。





## 第6章 結論

本章では、本論文のまとめと今後の課題について述べる。

### 6.1 まとめ

近年のインターネットやスマートデバイスの普及により、コミュニティQA サイトは人々の生活の中で身近に利用されている。一方で、スマートデバイスの普及によるサービス提供方法の変化により、スマートデバイスに保持したIDのみで多くのサービスの利用が可能になったため、コミュニティQA サイトを利用するユーザーは、入退会を繰り返すことで今までよりも簡易にユーザー履歴を消すことが可能になった。

コミュニティQA サイトの主な運営目標は、投稿された質問に最短時間で最も適切な回答を提供することである。この目的を達成するために、多くの既存研究では、ユーザー履歴を利用した質問ルーティングの手法の提案がされている。

しかし前述の通り、簡易にユーザー履歴を消すことが可能な近年のコミュニティQA サイトにおいても、実用可能な質問ルーティングの手法の提案が求められている。本論文では、質問テキストのみを用いた問ルーティングに関する取り組みを行うことで、より良い質問ルーティングを行う手法の提案を行った。

具体的には、重要性和課題解決の可能性を踏まえて以下の3つの課題に対して、機械学習を用いた手法を提案し、課題の解決に結びつけた。

1. 「不適切な質問」の抽出
2. 「回答率が低いと推察される質問」の抽出
3. 「ユーザー属性」の抽出

第一の課題に対しては、ユーザーが投稿した質問文を機械学習によって不適切な投稿かどうかを判別するために、最も最適な学習器、システム辞書の組み合わせを発見した。また、実際のシステムに導入し利用できる形にするため、目視を

## 6.2. 今後の課題

---

組み合わせた不適切な投稿の検閲システムを構築・提案し、業務フローに組み込み、実社会で成果を出すことができた。

第二の課題に対しては、コミュニティQAサイトに投稿される質問に関し、共感を求める質問とそれ以外を求める質問を定義し、有意に回答数が少ない、共感を求める質問を機械学習を用いて分類する手法について提案を行い、目標として目指す値に近い値を出すことができた。

第三の課題に対しては、質問文を分析した結果に基づき、特徴語による情報抽出を行うかどうかを機械学習によって判別する手法を提案し、十分な値の正解率を出すための手法を提案した。

本研究では、上記3つの課題に対して解決手法を提案することで、質問テキストのみで、より良い質問ルーティングを実現することができた。機械学習を用いて、コミュニティQAサイトにおける品質向上に対して一定の有効性を示したと言える。

## 6.2 今後の課題

本研究で取り組んだ課題について、トピック毎に今後の課題を述べる。「不適切な質問」をどう抽出するかという課題に対しての今後の課題は以下の通りである。

1. さらに精度の向上を行う。具体的には、現状は目視のみから1/10の監視対象数になったが、1/20を目指す
2. 画像つき投稿に対しては、テキスト+画像の情報を用いて学習器による判別を行う
3. 質問の回答に対しても同様のフィルターを導入して効果を確認する

「回答率が低いと推察される質問」をどう抽出するかという課題に対しての今後の課題は以下の通りである。

1. 今回の研究結果を実社会で提供中のサービスへ組み込み、ユーザーの反応を得る
2. 共感を求める質問に対して、満足度の高い回答を与える方法を開発する

ユーザー属性の抽出をどう行うかという課題に対しての今後の課題は以下の通りである。

## 6.2. 今後の課題

---

1. 今回の研究結果を実社会で提供中のサービスへ組み込み，ユーザーの反応を得る
2. ユーザー属性を回答から抽出する

上記に加えて，コミュニティQA サイトにおける課題はこれらに限らず多岐にわたるため，新たな課題への取り組みも含めて提案手法の応用・改善を行ってきたい。



## 参考文献

- [1] ママリ: <https://qa.mamari.jp>, 2018.7.23 参照.
- [2] Yahoo!知恵袋: <https://chiebukuro.yahoo.co.jp/>, 2018.7.23 参照.
- [3] Yahoo Answers: <https://answers.yahoo.com/>, 2018.7.23 参照.
- [4] Quora: <https://www.quora.com/>, 2018.7.23 参照.
- [5] Stack Overflow: <https://stackoverflow.com/>, 2018.7.23 参照.
- [6] 厚生労働省 平成 28 年人口動態統計月報年計 (概数) の概況: <http://www.mhlw.go.jp/toukei/saikin/hw/jinkou/geppo/nengai16/index.html>, 2018.7.23 参照.
- [7] Soojung Kim, Jung Sun Oh, and Sanghee Oh: Best-Answer Selection Criteria in a Social Q&A site from the User-Oriented Relevance Perspective, American Society for Information Science and Technology (ASIS&T) 2007 Annual Meeting, Milwaukee, Wisconsin, ASIS&T, 2007.
- [8] 栗山和子, 神門典子: Q&A サイトにおける質問と回答の分析. 情報処理学会研究報告, Vol.2009-FI-95, No.19, 2009.
- [9] Bo Qu, Gao Cong, Cuiping Li, Aixin Sun, and Hong Chen: An Evaluation of Classification Models for Question Topic Categorization, Journal of the American Society for Information Science and Technology, Vol. 63, No. 5, pp. 889-903, 2012.
- [10] Naoyoshi Aikawa, Tetsuya Sakai, and Hayato Yamana: Community QA Question Classification: Is the Asker Looking for Subjective Answers or Not?, IPSJ Online Transactions, Vol. 4, pp.160-168, 2011.
- [11] Tom Chao Zhou, Xiance Si, Edward Y. Chang, Irwin King, and Michael R. Lyu: A Data-Driven Approach to Question Subjectivity Identification in Community Question Answering, In Proceedings of the Twenty-Sixth

Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence, 2012.

- [12] Kudo, Taku, Kaoru Yamamoto, and Yuji Matsumoto: Applying Conditional Random Fields to Japanese Morphological Analysis, Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP-2004), pp. 230-237, 2004.
- [13] Sato Toshinori: Neologism Dictionary Based on the Language Resources on the Web for Mecab, 2015: <https://github.com/neologd/mecab-ipadic-neologd>, 2018.7.23 参照.
- [14] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space, International Conference on Learning Representations Workshop, 2013.
- [15] jawiki: <https://dumps.wikimedia.org/jawiki/>, 2018.7.23 参照.
- [16] J. Richard Landis, and Gary G. Koch: The Measurement of Observer Agreement for Categorical Data, *Biometrics*, Vol.33, No.1, pp.159-174, 1977.
- [17] scikit-learn: <http://scikit-learn.org/>, 2018.7.23 参照.
- [18] Claude Sammut, and Geoffrey I. Webb (Eds.): *Encyclopedia of Machine Learning*, pp.291-292, 2010.
- [19] Corinna Cortes, and Vladimir Vapnik: Support-vector network. *Machine Learning*, Vol.20, pp.273-297, 1995.
- [20] Leo Breiman: Random Forests, *Machine Learning*, Vol.45, No.1, pp.5-32, 2001.
- [21] Ivan Srba, and Maria Bielikova: A Comprehensive Survey and Classification of Approaches for Community Question Answering, *ACM Transactions on the Web*, Vol.10, No.3, 2016.
- [22] Bojan Furlan, Bosko Nikolic, and Veljko Milutinovic: A Survey and Evaluation of State-of-the-Art Intelligent Question Routing Systems, *International Journal of Intelligent Systems*, Vol.28, pp.686-708, 2013.

- [23] Tom Chao Zhou, Michael R. Lyu, and Irwin King: A Classification-Based Approach to Question Routing in Community Question Answering, *WWW Companion*, pp.783-790, 2012.
- [24] Baichuan Li, and Irwin King: Routing Questions to Appropriate Answerers in Community Question Answering Services, In *Proceedings of the ACM 19th Conference on Information and Knowledge Management*, pp.1585-1588, 2010.
- [25] Lichun Yang, Shenghua Bao, Qingliang Lin, Xian Wu, Dingyi Han, Zhong Su, and Yong Yu: Analyzing and Predicting Not-Answered Questions in Community-Based Question Answering Services, In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.
- [26] 渡邊直人, 島田諭, 関洋平, 神門典子, 佐藤哲司: コミュニティQAにおける質問の多面的評価法の検討, *情報知識学会誌*, Vol.21, No.2, pp.163-168, 2011.
- [27] 渡邊直人, 島田諭, 関洋平, 神門典子, 佐藤哲司: QA コミュニティにおける質問者の期待に基づく質問分類に関する一検討, 第3回データ工学と情報マネジメントに関するフォーラム, B5-1, 2011.
- [28] 三浦麻子, 川浦康至: 人はなぜ知識共有コミュニティに参加するのか: 質問行動と回答行動の分析, *社会心理学研究*, Vol.23, No.3, pp.233-245, 2008.
- [29] Baoli Li, Yandong Liu, and Eugene Agichtein: CoCQA: Co-Training Over Questions and Answers with an Application to Predicting Question Subjectivity Orientation, In *Proceedings of the Conference on Empirical Methods in Natural Language Processing-EMNLP'08*, Association for Computational Linguistics, pp.937-946, 2008.
- [30] F. Maxwell Harper, Daniel Moy, and Joseph A. Konstan: Facts or Friends? Distinguishing Informational and Conversational Questions in Social Q&A Sites, In *Proceedings of the 27th International Conference on Human Factors in Computing Systems-CHI 09*, 2009.
- [31] Eduarda Mendes Rodrigues, and Natasa Milic-Frayling: Socializing or Knowledge Sharing?: Characterizing Social Intent in Community Question Answering, In *Proceedings of the 18th ACM Conference on Information and Knowledge Management-CIKM'09*, 2009.



- [32] Long Chen, Dell Zhang, and Mark Levene: Understanding User Intent in Community Question Answering, In Proceedings of the 21st International Conference Companion on World Wide Web-WWW'12, 2012.
- [33] Fan Bu, Xingwei Zhu, Yu Hao, and Xiaoyan Zhu: Function-Based Question Classification for General QA, In Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing, 2010.
- [34] F. Maxwell Harper, Joseph Weinberg, John Logie, and Joseph A. Konstan: Question Types in Social Q&A, Vo.15, No.7, 2010.
- [35] Hapnes Toba, Zhao-Yan Ming, Mirna Adriani, and Tat-Seng Chua: Discovering High Quality Answers in Community Question Answering Archives Using a Hierarchy of Classifiers, Information Sciences, Vol.261, pp.101-115, 2014.
- [36] 小林大祐, 松村真宏, 木戸冬子, 石塚満: 知識検索サイトにおける不適切な投稿の分類, 第 21 回人工知能学会全国大会, 2007.
- [37] 朱成敏, 武田英明: コミュニティQA における文章の表層的特徴に基く質問文と回答文の分析, 第 29 回人工知能学会全国大会, 2015.
- [38] Kyosuke Nishida, and Ko Fujimura: Hierarchical Auto-Tagging: Organizing Q&A Knowledge for Everyone, In Proceedings of the 19th ACM International Conference on Information and Knowledge Management-CIKM'10, 2010.
- [39] Li Cai, Guangyou Zhou, Kang Liu, and Jun Zhao: Large-Scale Question Classification in CQA by Leveraging Wikipedia Semantic Knowledge, In Proceedings of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11, 2011.
- [40] Amit Singh, and Karthik Visweswariah: CQC: Classifying Questions in CQA Websites, In Proceedings of the 20th ACM International Conference on Information and Knowledge Management-CIKM'11, 2011.
- [41] Kyoungman Bae, and Youngjoong Ko: An Effective Category Classification Method Based on a Language Model for Question Category Recommendation on a CQA service, In Proceedings of the 21st ACM International Conference on Information and Knowledge Management-CIKM'12, 2012.

- [42] Wen Chan, Weidong Yang, Jinhui Tang, Jintao Du, Xiangdong Zhou, and Wei Wang: Community Question Topic Categorization via Hierarchical Kernelized Classification, In Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management-CIKM'13, 2013.
- [43] Liqiang Nie, YiLiang Zhao, Xiangyu Wang, Jialie Shen, and Tat-Seng Chua: Learning to Recommend Descriptive Tags for Questions in Social Forums, ACM Transactions on Information Systems, Vol.32, No.1, 2014.
- [44] 石川大介, 酒井哲也, 関洋平, 栗山和子, 神門典子: コミュニティQAにおける良質回答の自動予測, 情報知識学会誌, Vol.21, No.3, pp.362-382, 2011.
- [45] Imrul Kayes, Nicolas Kourtellis, Daniele Quercia, Adriana Lamnitchi, and Francesco Bonchi: The Social World of Content Abusers in Community Question Answering, In Proceedings of the 24th International Conference on World Wide Web-WWW'15, pp.570-580, 2015.
- [46] Otsuka Takanobu, Deng Deyue, and Ito Takayuki: Text Filtering for Harmful Document Classification Using Three-Word Co-Occurrence and Large-Scale Data Processing, Electronics and Communications in Japan, Vol.98, No.10, pp.31-40, 2015.
- [47] Pooja Revar, Arpita Shah, Jitali Patel, and Pimal Khanpara: A Review on Different Types of Spam Filtering Techniques, International Journal of Advanced Research in Computer Science, Vol.8, No.5, 2017.
- [48] 馬場雪乃, 鹿島久嗣, 木下慶, 山口豪志, 秋好陽介: 機械学習による不適切なクラウドソーシングタスクの検出, 第5回データ工学と情報マネジメントに関するフォーラム, 2013.
- [49] gensim: <https://radimrehurek.com/gensim/index.html>, 2018.7.23 参照.
- [50] Christos H. Papadimitriou, Prabhakar Raghavan, Tamaki Hisao, and Santosh Vempala: Latent Semantic Indexing, A Probabilistic Analysis, Journal of Computer and System Sciences, Vol.61, No.2, pp.217-235, 2000.
- [51] David M. Blei, Andrew Y. Ng, and Michael I. Jordan: Latent Dirichlet Allocation, Journal of Machine Learning Research, Vol.3, pp.993-1022, 2003.
- [52] 岡井崇, 綾部琢哉 (編), 標準産科婦人科学, 医学書院, 2014.

- [53] Christopher D. Manning, Prabhakar Raghavan, and Hinrich Schütze: Introduction to Information Retrieval, Cambridge University Press, 2008.
- [54] App Store: <https://www.apple.com/jp/ios/app-store/>, 2018.7.23 参照.
- [55] 新村出 (編), 広辞苑 第七版, 岩波書店, 2018.
- [56] Wolfgang Köhler, Gestalt psychology, New York : H. Liveright, 1929.
- [57] Dymond, Rosalind F.: A Preliminary Investigation of the Relation of Insight and Empathy, Journal of Consulting Psychology, Vol.12, pp.228-233, 1948.
- [58] Ezra Stotland: Exploratory Investigations of Empathy, Advances in Experimental Social Psychology, Vol.4, pp.271-314, 1969.
- [59] Wispé, Lauren: The Distinction Between Sympathy and Empathy: To Call Forth a Concept, A Word is Needed, Journal of Personality and Social Psychology, Vol.50, No.2, pp.314-321, 1986.
- [60] Aditya Pal, Rosta Farzan, Joseph A. Konstan, and Robert Kraut: Early Detection of Potential Experts in Question Answering Communities, In Proceedings of the 19th International Conference on User Modeling, Adaptation, and Personalization-UMAP'11, pp.231-242, 2011.
- [61] Aditya Pal, F. Maxwell Harper, and Joseph A. Konstan: Exploring Question Selection Bias to Identify Experts and Potential Experts in Community Question Answering, ACM Transactions on Information Systems (TOIS), Vol.30, No.2, 2012.
- [62] Dana Movshovitz-Attias, Yair Movshovitz-Attias, Peter Steenkiste, and Christos Faloutsos: Analysis of the Reputation System and User Contributions on a Question Answering Website: StackOverflow, In Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining-ASONAM'13, 2013.
- [63] Siqi Song, Ye Tian, Wenwen Han, Xirong Que, and Wendong Wang: Leading Users Detecting Model in Professional Community Question Answering Services, In Proceedings of IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing, pp.1302-1307, 2013.

- [64] Adabriand Furtado, Nazareno Andrade, Nigini Oliveira, and Francisco Brasileiro: Contributor Profiles, their Dynamics, and their Importance in Five Q&A Sites, In Proceedings of the 2013 Conference on Computer Supported Cooperative Work-CSCW'13, pp.1237-1252, 2013.
- [65] Gideon Dror, Dan Pelleg, Oleg Rokhlenko, and Idan Szpektor: Churn prediction in new users of Yahoo! answers, In Proceedings of the 21st International Conference Companion on World Wide Web-WWW'12 Companion, pp.829-83, 2012.
- [66] Jagat Pudipeddi, Leman Akoglu, and Hanghang Tong: User Churn in Focused Question Answering Sites: Characterizations and Prediction, In Proceedings of the 23st International Conference Companion on World Wide Web-WWW'14 Companion, pp.469-474, 2014.
- [67] TurboTax Live Community: <http://ttlc.intuit.com/>, 2018.7.23 参照.



# 研究業績

## 査読付き国内雑誌

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. 情報処理学会 デジタルプラクティス, Vol. 8, No. 1, 92-96 (2017)

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 (知能情報ファジィ学会論文誌) , Vol. 29, No. 4, 611-618 (2017)

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 (採録 2018年2月14日付)

## 査読付き国際学会

\*Tatsuro Shimada, Akito Sakurai. Recognition of Empathy Seeking Questions in One of the Largest Woman CQA in Japan. Proc. ACIIDS2017, LNCS 10191, 641-650, Springer (2017)



# 謝辞

本研究を行うにあたり、多くの方々のご指導、ご鞭撻を賜りましたことを、この場をお借りして心より感謝いたします。

指導教官の櫻井彰人先生には、学部時代の卒業研究に始まり、修士研究、そして今回の博士研究まで本当にお世話になりました。修士課程の途中では、突然会社を立ち上げ、一度研究を諦めるという選択肢までご相談させていただいたのは、今ではとても良い思い出です。ご迷惑ばかりかけてしまい、大変申し訳ありませんでした。しかし、櫻井先生のおかげで仕事にも研究にも熱意を持って取り組むことができました。深く、感謝を申し上げます。櫻井先生から学んだことを、しっかりとこの先、社会に対して還元していきたいと思えます。

また、櫻井先生のご退官後に、主査を引き受けてくださった、鈴木先生にも深く感謝申し上げます。特にプレゼンテーションに関するアドバイスを多分にいただき、研究発表のクオリティを大きく向上させることができました。重ね重ね、まことに有難うございました。

お忙しい中、副査を快諾していただき、本論文の執筆にあたり、有益なご助言、およびご指導を頂いた、篠沢先生、萩原先生、中西先生にも厚く御礼申し上げます。松林先生には授業に招致いただき、お話をさせていただく貴重な機会をいただき大変感謝しております。

櫻井研究室先輩方には大変お世話になりました。研究分野にとどまらず、様々な分野で活躍される先輩方の背中には私にとって、とても刺激的でした。特に数原氏、deng氏には研究において示唆に富んだアドバイスを多数いただき、本当に感謝しております。共に切磋琢磨した研究室同期、そして後輩のみなさまにも大変お世話になりました。研究室生活を振り返ると楽しい思い出ばかりなのは、みなさまのおかげです。

短い間ではありましたが、鈴木研究室の方々にも大変お世話になりました。櫻井研究室から突然やってきた私を受け入れてくださり、研究発表に対してもとても有益な助言をいただきました。本当に有難うございました。

インターン時代にお世話になった同僚、先輩方にも感謝致します。学生時代のインターン活動で得た経験は、研究活動に大きな影響を与えてくれました。インターンを卒業しても、声をかけてくださる皆様にいつも支えられています。



Connehito 株式会社社会の皆様には大変お世話になりました。共に起業し、会社を運営するパートナーとして大変お世話になった大湯氏に心から感謝致します。研究を進める上で、様々な後押しをして頂き、いつでも柔和に逆算試行を持ってものごとにあたるその姿を見て、日々尊敬の念を抱いております。

最後に、私の人生を支えてくれた家族に感謝します。尊敬する両親、兄、姉や従兄弟をはじめ、小さいときから私を見守ってくださる祖母という家族があつて、今の自分があります。そして、この春無事に産まれてくれた息子の顕太郎と、妻にも深く感謝致します。妻と彼の笑顔のおかげで、研究と仕事の忙しい日々を乗り越えることができましたと思います。たくさんの負担をかけてしまい申し訳ありません。いつも、本当に有難うございます。これからも、家族で楽しい日々を過ごそう。