

新聞と Q&A サイトにおける人生相談の分析

栗山 和子^{1,a)}

概要: 本研究では、2013 年の読売新聞の「人生案内」欄と Q&A サイト「OKWave」の「恋愛・人生相談」カテゴリに投稿された質問・回答について分析を行い、その投稿者と質問・回答の特徴について比較を行った。その結果、共通点としては、配偶者か恋人かという違いはあるものの、自分のパートナーに関連する質問が多く、質問文には評価や感情を表す形容詞がよく用いられていることがわかった。相違点としては、新聞の相談者は女性の割合は高く年齢層があまり偏っていないのに対し、Q&A サイトの質問者はほぼ男女半々で 10~20 代の若年層に偏っており、回答者はそれよりも上の 20~30 代が多いことがわかった。新聞と Q&A サイトの人生相談は利用者や相談内容が完全に異なるわけではなく、10~30 代の若い世代については重複しているため、相補的に利用することで有用な回答を広く探すことが可能性であると考えられる。

1. はじめに

インターネット上では、情報を得る手段の 1 つとして、Q&A サイトと呼ばれる質問回答サイトを利用することが一般的になっている [2], [3]。

著者らは、以前の論文 [6], [7] で、Q&A サイトの質問の分析を行い、情報検索型の質問と社会調査型の質問に分けられることを示した。その際、Q&A サイトには、回答者に助言や意見などを求める社会調査型の質問が多く含まれおり、サーチエンジンを用いた情報検索や図書館のレファレンス・サービスでは回答が難しいと思われる質問が多数あることがわかった。社会調査型の質問には一つの決まった正解はないが、質問者は複数の回答から自分が納得する回答を選択して質問の解決の参考にしたり、興味や関心を充足させたりしていると考えられる。

そこで、本研究では、社会調査型の質問の中でも人生相談についての質問に着目し、印刷媒体として従来から広く普及している新聞の人生相談欄と Q&A サイトの質問・回答を比較することにより、匿名の人生相談を利用する質問者がどのような悩み・質問を抱いているのか、また、インターネット以外の媒体を参照することでより適切な回答が得られる可能性があるかどうか考察する。

2 節では、Q&A サイト上の質問・回答に関する関連研究について簡単に述べる。3 節では、利用するデータを紹介

する。4 節では、質問・回答および質問者・回答者についての基本的な統計量を示し、質問文と回答文の重要語を抽出して、それらの特徴を比較する。5 節では、本稿のまとめを述べ、今後の予定と課題を示す。

2. 関連研究

Q&A サイトに関する研究は数多く行われている。本節では、ベストアンサーの推定や質問・回答からの情報抽出を目的とする研究ではなく、質問型を考慮した質問への回答の提示や評価について焦点を当てている研究のみについて簡単に述べる。

高田ら [8] は、Q&A サイトの質問を分析し、回答者の意見を求める質問について、質問を元にしたクエリを用いて Web サーチエンジンで検索した Web 上の情報や Q&A サイト内の別の質問の回答を補完情報として提示する手法を提案している。

瀧ら [9] は、Q&A コミュニティを対象とした回答の信頼性指標を構築するために、最良回答が選ばれるときの傾向について考察している。その結果として、外部参照として他のサイトへの URL が含まれる回答は、最良回答として選ばれやすく信頼性指標になりうると述べている。

王ら [5] は、社会調査型の質問に対して、質問者が回答の信頼性を判断するための材料として、回答者の質問・回答履歴から、投稿した質問と類似した質問への回答者の回答を抽出し、質問者に提示するシステムを提案している。

渡邊ら [11] は、質問の内容と質問者の回答への期待に基づく、質問の多面的評価法について検討し、質問者の要求属性は質問タイプとは独立して判定できること、および、

¹ 東洋大学社会学部メディアコミュニケーション学科
Department of Media and Communications, Faculty of Sociology, Toyo University
a) kkuriyama@toyo.jp

質問者要求属性は質問タイプよりも詳細な判定ができることを示している。

田中ら [10] は、Q&A サイトの質問回答ログから情報検索型質問を自動抽出しクラスタリングを行い、そのクラスタを分析することで、質問者が試行錯誤する理由や求めている情報を明らかにし、Web 検索者を支援するために必要な情報について考察している。

以上のように、Q&A サイトの質問型を利用した回答の分類・評価、Web や Q&A サイト内の情報による回答の補完についての研究は見られるが、Web 以外の媒体を利用した情報の補完についての研究は少ない。そこで、本研究では、印刷媒体やオンラインデータベースとして広く利用されている新聞記事の人生相談欄との比較を行い、Q&A サイトと他の媒体上での質問・回答や質問者・回答者の違いについて考察を行い、相補的な利用の可能性について考察する。

3. 使用データ

3.1 新聞記事データ

本研究では、読売新聞社が提供し、日外アソシエーツが頒布している「読売新聞記事データ 2013 年版」に含まれる「人生案内」の記事データを使用した。2013 年版の記事の総数は 299,821 件、「人生案内」は 354 件である。

「人生案内」は、新聞休刊日を除いて毎日に掲載されており、読者が質問（相談）を投稿し、新聞社から回答を依頼されている著名人や専門家がそれに対して回答を行うという形式になっている。質問と回答は一緒に 2 段組の 1 つの枠内に収められており、全体の文字数は 800~900 字前後に統一されている。

相談募集の記事には、「生活上の悩み全般が対象。純粋な法律問題や病気の相談などは扱いません」[4] と記載されているため、質問内容は人生・生活相談に限られる。記事には、投稿先の住所とメールアドレスが示されており、相談は郵送でもメールでも投稿することができるが、「相談は 800 字程度。規定の行数に整理して掲載」[4] と文字数の制限と掲載時に編集されることが明記されており、投稿文を欄の半分程度の分量にまとめるため、質問の文章には記者の手が入れられていると考えられる。

3.2 Q&A サイトの質問データ

本研究では、Q&A サイトのデータとして、OKWave[2] に 2013 年 1 月 1 日から 2013 年 12 月 31 日までの 1 年間に投稿され、2014 年 8 月 21 日の時点で同サイトに掲載されていた質問と回答のページ、および、その質問・回答の投稿者のプロフィールのページを収集して使用した。

OKWave では、質問に連続した番号が付けられているため、2013 年 1 月 1 日 0 時台に最初に投稿された質問から 2013 年 12 月 23 時台に最後に投稿された質問まで連続

表 1 カテゴリごとの質問件数

| カテゴリ | 質問件数 | 割合 (%) |
|--------------------|---------|--------|
| ライフ | 127,231 | 29.5 |
| デジタルライフ | 74,474 | 17.0 |
| 趣味 | 27,565 | 6.4 |
| エンターテイメント | 18,485 | 4.3 |
| 美容・健康 | 37,783 | 8.8 |
| ビジネス・キャリア | 19,302 | 4.5 |
| 社会 | 26,980 | 6.3 |
| マネー | 12,933 | 3.0 |
| 学問・教育 | 47,873 | 11.1 |
| 地域情報 | 27 | 0.0 |
| [地域情報] 旅行・レジャー | 12,514 | 2.9 |
| [技術者向] コンピューター | 14,392 | 3.3 |
| 大規模災害 | 61 | 0.0 |
| 商品ウォッチ | 0 | 0.0 |
| アンケート | 10,874 | 2.5 |
| この Q&A コミュニティーについて | 1,194 | 0.3 |
| おしゃべり広場 | 59 | 0.0 |
| 企業公式 | 113 | 0.0 |
| 特別企画 | 103 | 0.0 |
| 合計 | 431,028 | 100.0 |

した番号の質問 538,304 件を収集したが、その時点で、存在しない番号の質問が 107,212 件あったため、実際に収集できた質問は 431,093 件である。そのうち、65 件は、「現在、この質問はサポートで内容を確認中です。」という理由で質問文が公開されていなかったためデータからは除き、431,028 件を使用した。

表 1 に、最上位の 19 カテゴリについてカテゴリごとの質問件数と質問件数合計に対する割合 (%) を示す。

表 1 から、1 年間の質問全体の約 3 分の 1 が「ライフ」カテゴリに質問されたものであり、投稿数の多い上位カテゴリ「ライフ」「デジタルライフ」「学問・教育」で全体の半数以上を占めることがわかる。

「ライフ」のサブカテゴリは、「恋愛・人生相談」、「結婚」、「出産・育児」、「料理・グルメ」、「家電製品」、「住まい」、「暮らし・生活お役立ち」、「ペット」、「交通」、「その他(ライフ)」の 10 カテゴリであるが、全て含めると、人生相談ではない質問が多数含まれてしまうため、新聞記事データの「人生案内」と比較するために、「恋愛・人生相談」カテゴリの質問 65,424 件を使用した。これは、「ライフ」カテゴリの質問の 51.4%、質問全体の 15.2%にあたる。

「恋愛・人生相談」には、11 のサブカテゴリが含まれる。表 2 にサブカテゴリごとの質問件数と「恋愛・人生相談」の質問件数合計に対する割合 (%) を示す。

4. 新聞と Q&A サイトの質問・回答の比較

4.1 質問者

読売新聞の「人生案内」の相談者（以下では、呼称を統一するため「質問者」と呼ぶ）は匿名になっている。相談

表 2 「恋愛・人生相談」のサブカテゴリの質問件数

| カテゴリ | 質問件数 | 割合 (%) |
|---------------|---------------|-------------|
| 恋愛相談 | 33,988 | 52.0 |
| 友達・仲間関係 | 5,165 | 7.9 |
| 夫婦・家族 | 7,979 | 12.2 |
| 社会・職場 | 5,615 | 8.6 |
| シニアライフ | 262 | 0.4 |
| 人生相談 | 1,478 | 2.3 |
| LGBT | 39 | 0.1 |
| DV (家庭内暴力) | 38 | 0.1 |
| いじめ相談 | 471 | 10.7 |
| 不倫関係 | 184 | 0.3 |
| その他 (恋愛・人生相談) | 10,205 | 15.6 |
| 合計 | 65,424 | 100.0 |

者名は原則として住所の地域と姓名の名の部分のアルファベットの頭文字と名の一部で構成されており、例えば、「千葉・W子」、「兵庫・I夫」のような仮名で表されている。質問者の地域を表 3,4 に示す。

表 3 「人生案内」質問者の地域

| 地域 | 件数 | 地域 | 件数 | 地域 | 件数 | 地域 | 件数 |
|-----|-----------|-----|-----------|-----|-----------|-----|----|
| 北海道 | 8 | 東京 | 44 | 滋賀 | 4 | 愛媛 | 1 |
| 青森 | 2 | 神奈川 | 30 | 京都 | 6 | 徳島 | 1 |
| 岩手 | 0 | 山梨 | 4 | 大阪 | 29 | 高知 | 0 |
| 秋田 | 2 | 長野 | 1 | 兵庫 | 23 | 福岡 | 8 |
| 宮城 | 5 | 新潟 | 7 | 奈良 | 3 | 佐賀 | 0 |
| 山形 | 1 | 富山 | 7 | 和歌山 | 1 | 長崎 | 2 |
| 福島 | 1 | 石川 | 1 | 鳥取 | 0 | 熊本 | 3 |
| 茨城 | 16 | 福井 | 0 | 島根 | 2 | 大分 | 2 |
| 栃木 | 4 | 静岡 | 1 | 岡山 | 7 | 宮崎 | 1 |
| 群馬 | 9 | 愛知 | 5 | 広島 | 5 | 鹿児島 | 0 |
| 埼玉 | 49 | 岐阜 | 1 | 山口 | 6 | 沖縄 | 0 |
| 千葉 | 27 | 三重 | 4 | 香川 | 2 | 未記載 | 19 |

表 3 からわかるように、質問者の地域では、群馬、栃木を除く関東と関西の一部（大阪、兵庫）が比較的多いが、これは、2013 年に朝刊販売部数で読売新聞が 1 位になっている地域とほぼ重なっており [1]、新聞自体の読者数の偏りが影響していると考えられる。

質問者の属性は、質問文の第 1 文に「14 歳の男子中学生」、「60 代の主婦」、のように示されているため、年代、性別、職業をスクリプトで抽出した。第 1 文に性別のわかる語が含まれていない 4 件の記事については、仮名から判別した。職業は第 1 文にも質問文全体にも含まれていない場合は抽出できないため、不明とした。

質問者の性別は、354 人中、女性 299 人、男性 55 人であった。年代と職業をそれぞれ表 4,5 に示す。「パートで働く主婦」のように書かれている場合は「主婦」とした。「その他」には、派遣社員、契約社員、カルチャー講師、建設業、養護教諭、学校事務員、臨時職員、非常勤、会社顧問、保育士が含まれる。

表 4 「人生案内」質問者の年代

| 年代 | 人数 | 割合 (%) |
|------|----|--------|
| 10 代 | 21 | 5.9 |
| 20 代 | 33 | 9.3 |
| 30 代 | 50 | 14.1 |
| 40 代 | 63 | 17.8 |
| 50 代 | 53 | 15.0 |
| 60 代 | 66 | 18.6 |
| 70 代 | 46 | 13.0 |
| 80 代 | 19 | 5.4 |
| 90 代 | 3 | 0.8 |

表 5 「人生案内」質問者の職業

| 職業 | 人数 |
|-----------------|------------|
| 主婦・主夫 | 128 |
| 会社員 | 39 |
| 公務員 | 5 |
| 自営業 | 12 |
| 生徒・学生 | 23 |
| パート・アルバイト・フリーター | 12 |
| 無職・年金暮らし | 10 |
| その他 | 11 |
| 不明 | 114 |

表 4 から、質問者の年齢の分布は、特定の年代に偏っておらず、中高年が多いものの、30 代以下の若者も約 3 割いることがわかった。また、表 5 からわかるように、職業が抽出できた 240 件の中では、主婦・主夫が半数以上と非常に多かった。

OKWave の「恋愛・人生相談」カテゴリの質問 65,424 件の質問者数は 29,947 人である。「恋愛・人生相談」カテゴリで質問している利用者のこのカテゴリでの質問回数は、1 回が 20,417 人 (68.2%)、2~9 回が 8,715 人 (29.1%)、10 回以上 815 人 (2.7%) である。また、同じ利用者 29,947 人の全カテゴリでの質問回数は、1 回が 14,116 人 (47.1%)、2~9 回が 12,459 人 (41.6%)、10 回以上 3,372 人 (11.3%) であり、約半数は「恋愛・人生相談」カテゴリのみで 1 回だけ質問している。

OKWave では、利用者は利用登録時に自分のプロフィールを登録することになっているが、「恋愛・人生相談」カテゴリの質問者 29,947 人中では、質問ページを収集した時点で OKWave を退会していた者が 3,653 人、プロフィールが存在する者が 25,294 人であり、そのうち性別、都道府県、生年、職業がわかるプロフィールを登録している者は、3,085 人 (質問者の 10.3%) であった。3,085 人のうち、女性は 1,639 人 (53.1%)、男性は 1,446 人 (46.9%) であり、「人生案内」の質問者とは異なり、男女がほぼ半々であることがわかった。3,085 人のプロフィールについて、表 6 に地域を、表 7 に生年から計算した年代を、表 8 に職業を示す。

表 6 からわかるように、「恋愛・人生相談」の質問者の地

表 6 「恋愛・人生相談」質問者の地域

| 地域 | 件数 | 地域 | 件数 | 地域 | 件数 | 地域 | 件数 |
|-----|-----|-----|-----|-----|-----|-----|-----|
| 北海道 | 194 | 東京 | 547 | 滋賀 | 26 | 愛媛 | 25 |
| 青森 | 28 | 神奈川 | 253 | 京都 | 80 | 徳島 | 15 |
| 岩手 | 26 | 山梨 | 24 | 大阪 | 239 | 高知 | 13 |
| 秋田 | 22 | 長野 | 39 | 兵庫 | 104 | 福岡 | 105 |
| 宮城 | 59 | 新潟 | 38 | 奈良 | 31 | 佐賀 | 4 |
| 山形 | 22 | 富山 | 8 | 和歌山 | 14 | 長崎 | 17 |
| 福島 | 31 | 石川 | 26 | 鳥取 | 8 | 熊本 | 34 |
| 茨城 | 54 | 福井 | 16 | 島根 | 9 | 大分 | 15 |
| 栃木 | 43 | 静岡 | 83 | 岡山 | 33 | 宮崎 | 16 |
| 群馬 | 31 | 愛知 | 150 | 広島 | 45 | 鹿児島 | 24 |
| 埼玉 | 201 | 岐阜 | 33 | 山口 | 18 | 沖縄 | 32 |
| 千葉 | 155 | 三重 | 34 | 香川 | 20 | 海外 | 41 |

表 7 「恋愛・人生相談」質問者の年代

| 年代 | 人数 | 割合 (%) |
|---------|-------|--------|
| 10 歳未満 | 3 | 0.1 |
| 10 代 | 1,142 | 37.0 |
| 20 代 | 1,280 | 41.5 |
| 30 代 | 384 | 12.4 |
| 40 代 | 157 | 5.1 |
| 50 代 | 40 | 1.3 |
| 60 代 | 18 | 0.6 |
| 70 代 | 7 | 0.2 |
| 80 代 | 1 | 0.0 |
| 90 代 | 2 | 0.1 |
| 100 歳以上 | 51 | 1.7 |

表 8 「恋愛・人生相談」質問者の職業

| 職業 | 人数 | 割合 (%) |
|-------|-------|--------|
| 主婦 | 122 | 4.0 |
| 会社員 | 879 | 28.5 |
| 公務員 | 89 | 2.9 |
| 自営業 | 119 | 3.9 |
| 学生 | 1,513 | 49.0 |
| フリーター | 251 | 8.1 |
| 無職 | 112 | 3.6 |

域は大都市圏が多いが、国内ではインターネットの利用環境に居住地の影響は少ないので、この偏りは人口の多さを反映しているためと考えられる。また、表 7 では、質問者の年代は 10 代と 20 代の若者が 8 割近くを占めている。これは、「人生案内」の質問者の年齢があまり偏っていないのとは対照的であり、若年層が多いのは、インターネットの利用がより身近であるためと考えられる。表 8 から、質問者は学生が約半数であり、年齢層と同様に、若者の利用が多いことがわかる。しかし、3 割近くは会社員・公務員であり、社会人も多く利用している。

以上のように、読売新聞の「人生相談」の質問者は、世代にはあまり偏りがなく、職業としては主婦が多いのに対し、OKWave の質問者は 10~20 代の学生が多く、質問者層は一部重複するものの、異なっていると考えられる。

4.2 回答者

読売新聞の「人生相談」の回答者は、年によって異なる。2013 年の記事 354 件の回答者 11 人の氏名、職業（専門分野）、生年、2013 年の回答回数を表 9 に示す。

表 9 「人生案内」回答者

| 氏名 | 職業（専門分野） | 生年 | 回答回数 |
|-------|----------------|------|------|
| 大日向雅美 | 大学教授（発達心理学） | 1950 | 29 |
| 海原純子 | 医師（心療内科） | 1952 | 35 |
| 最相葉月 | ライター（ノンフィクション） | 1964 | 32 |
| 出久根達郎 | 作家（小説・随筆） | 1944 | 34 |
| 土肥幸代 | 弁護士 | 不明 | 34 |
| 野村総一郎 | 医師（精神科） | 1949 | 38 |
| 樋口恵子 | 評論家（女性・福祉・教育） | 1932 | 26 |
| 久田恵 | 作家（ノンフィクション） | 1947 | 33 |
| 増田明美 | スポーツ解説者（長距離走） | 1964 | 24 |
| 眉村卓 | 作家（SF 小説） | 1934 | 18 |
| 山田昌弘 | 大学教授（家族社会学） | 1957 | 30 |
| 鷲田清一 | 哲学者（臨床哲学・倫理学） | 1949 | 21 |

表 9 からわかるように、回答者の年代は 40 代後半から 80 代前半であり、60 代以上が 9 人と 4 分の 3 を占めている。（土肥幸代の生年は不明だが、著書のプロフィールに「1963 年東京弁護士会に弁護士登録」と書かれているため、生年は 1940 年代以降と推測される。）

OKWave の「恋愛・人生相談」カテゴリの質問 65,424 件の回答者は 7,203 人である。その中でプロフィールページが存在する者が 6,433 人、性別、都道府県、生年、職業がわかるプロフィールを登録している者は 539 人（回答者の 7.7%）であった。539 人のうち、女性は 219 人（40.6%）、男性は 320 人（59.3%）であり、男性の方が多いことがわかった。539 人のプロフィールについて、表 10 に生年から計算した年代を、表 11 に職業を示す。

表 10 からわかるように、「恋愛・人生相談」の回答者の年代は、10~40 代が 9 割を占めており、特に 20~30 代が半数以上と多い。質問者の 8 割が 10~20 代であるのに対し、回答者はそれよりも上の年齢層にシフトしていると考えられる。表 11 から、回答者の年齢層が質問者の年齢層よりも少し上であるため、学生よりも社会人が増えていることがわかる。この傾向は、読売新聞の「人生案内」の回答者が高齢者ばかりであるのとは大きく異なる。

4.3 質問

読売新聞の「人生案内」の質問 354 件と OKWave の「恋愛・人生相談」の質問 65,424 件について、形態素解析ソフトウェア MeCab を用いて形態素解析を行い、質問中に出現する語を抽出した。解析用の辞書は、IPADIC をそのまま使用した。形態素解析と語の抽出は、タイトルと本文に分けて行った。

「人生案内」と「恋愛・人生相談」それぞれのタイトル

表 10 「恋愛・人生相談」回答者の年代

| 年代 | 人数 | 割合 (%) |
|--------|-----|--------|
| 10代 | 112 | 20.8 |
| 20代 | 176 | 32.7 |
| 30代 | 122 | 22.6 |
| 40代 | 75 | 13.9 |
| 50代 | 25 | 4.6 |
| 60代 | 20 | 3.7 |
| 70代 | 4 | 0.7 |
| 80代 | 1 | 0.2 |
| 90代 | 1 | 0.2 |
| 100歳以上 | 3 | 0.6 |

表 11 「恋愛・人生相談」回答者の職業

| 職業 | 人数 | 割合 (%) |
|-------|-----|--------|
| 主婦 | 44 | 8.2 |
| 公務員 | 17 | 3.2 |
| 会社員 | 203 | 37.7 |
| 自営業 | 59 | 10.9 |
| 学生 | 140 | 26.0 |
| フリーター | 42 | 7.8 |
| 無職 | 34 | 6.3 |

と本文の文字数の平均, 最小, 最大, 標準偏差を表 12 に示す. ただし, 「人生案内」は記事全体で 800~900 文字程度なので, 質問の文字数はその半分程度が上限であり, OKWave では質問本文は 4000 文字までに投稿が制限されている. 表 13 に, 質問のタイトルと本文の延べ語数と異なり語数とその平均を示す.

表 12 質問のタイトルと本文の文字数 (平均/最少/最大/標準偏差)

| 文字数 | 人生案内 | | 恋愛・人生相談 | |
|------|------|-------|---------|--------|
| | タイトル | 本文 | タイトル | 本文 |
| 平均 | 17.7 | 398.5 | 13.9 | 545.4 |
| 最小 | 13 | 337 | 1 | 2 |
| 最大 | 22 | 458 | 35 | 4000 |
| 標準偏差 | 1.49 | 25.08 | 5.95 | 453.11 |

表 12 からわかるように, 「人生相談」の質問は, 記事全体の長さの制限によってタイトルも本文も短くなっているが, 「恋愛・人生相談」の質問は, 投稿可能な文字数の上限が大きいため, 長さのばらつきが大きい. 表 13 から, 質問文の長さが異なるので, 延べ語数と異なり語数には差があるが, 質問件数で平均をとると, タイトルでは語数の差が小さいことがわかる. このことから, 質問者がタイトルをある程度自由に付けられる場合でも, 長さはタイトルとして適当な長さに収まるものと考えられる.

質問の特徴的な語を抽出するために, 語の重みとして $TF-IDF$ を計算した. $TF-IDF$ の計算法にはいくつかのヴァリエーションがあるため, 本研究では, 文書 d_i における語 t_j の $TF_{ij} \times IDF_j$ の計算法として, 以下の式を用いる.

表 13 質問のタイトルと本文の使用語数 (延べ/異なり)

| 語数 | 人生案内 | | 恋愛・人生相談 | |
|----------|-------|--------|---------|------------|
| | タイトル | 本文 | タイトル | 本文 |
| 延べ | 2,653 | 90,901 | 545,737 | 21,581,878 |
| 異なり | 794 | 5,119 | 13,476 | 51,270 |
| 延べ (平均) | 7.5 | 256.8 | 8.3 | 329.9 |
| 異なり (平均) | 7.4 | 114.5 | 7.9 | 126.2 |

表 14 質問中の名詞の統合 $TF-IDF$ の上位 25 件

| 順位 | 人生案内 | | 恋愛・人生相談 | |
|----|------|-------------|---------|-------------|
| | 名詞 | 統合 $TF-IDF$ | 名詞 | 統合 $TF-IDF$ |
| 1 | 夫 | 4.01 | の | 900.08 |
| 2 | 私 | 3.94 | 私 | 718.88 |
| 3 | の | 3.85 | こと | 612.68 |
| 4 | 人 | 2.70 | 人 | 574.70 |
| 5 | こと | 2.65 | 彼 | 521.13 |
| 6 | 0 | 2.55 | ん | 416.55 |
| 7 | 母 | 2.31 | 彼女 | 364.80 |
| 8 | 息子 | 2.22 | 自分 | 362.16 |
| 9 | 娘 | 1.99 | よう | 348.79 |
| 10 | よう | 1.96 | 女性 | 333.31 |
| 11 | 2 | 1.93 | 好き | 331.34 |
| 12 | 1 | 1.89 | 事 | 301.83 |
| 13 | 代 | 1.86 | 男性 | 291.36 |
| 14 | 子ども | 1.83 | 方 | 281.11 |
| 15 | 彼 | 1.81 | 何 | 279.00 |
| 16 | 年 | 1.73 | メール | 272.07 |
| 17 | 自分 | 1.70 | 友達 | 255.91 |
| 18 | 女性 | 1.61 | 時 | 240.10 |
| 19 | 彼女 | 1.61 | 今 | 235.16 |
| 20 | 家 | 1.52 | 彼氏 | 225.02 |
| 21 | 3 | 1.45 | (| 224.11 |
| 22 | 今 | 1.45 | 気 | 224.10 |
| 23 | 結婚 | 1.42 | 仕事 | 223.21 |
| 24 | 何 | 1.41 | お願い | 219.70 |
| 25 | 仕事 | 1.41 | 相手 | 218.97 |

$$TF_{ij} \times IDF_j = \frac{tf_{ij}}{\sum_{i=1}^n tf_{ij}} \times (\log \frac{N}{df_j} + 1) \quad (1)$$

ここで, tf_{ij} は文書 d_i における語 t_j の出現頻度, n_i は文書 d_i の異なり語数, df_j は単語 t_j が出現する文書数, N は全文書数とする. 質問の長さのばらつきが大きいため, TF_{ij} は相対出現頻度として, 出現頻度 tf_{ij} を文書 d_i におけるすべての語の出現頻度の総和 $\sum_{i=1}^n tf_{ij}$ で割ったものを使用する.

表 14,15 に, 「人生案内」と「恋愛・人生相談」の質問本文に出現する語のうち, 名詞と形容詞について, 全文書に対する $TF-IDF$ を足し合わせた統合 $TF-IDF$ の上位 25 件を示す.

表 14 から, 「人生案内」の質問には, 「夫」「母」「娘」「息子」「子ども」「彼」「彼女」のような配偶者や恋人や家族などの身近な人を表す語が多く出現することがわかる. 「恋愛・人生相談」でも, 「彼」「彼女」「彼氏」「友達」「相手」

表 15 質問中の形容詞の統合 TF-IDF の上位 25 件

| 順位 | 人生案内 | | 恋愛・人生相談 | |
|----|------|-----------|---------|-----------|
| | 名詞 | 統合 TF-IDF | 形容詞 | 統合 TF-IDF |
| 1 | ない | 1.52 | いい | 338.28 |
| 2 | 良い | 1.20 | ない | 336.21 |
| 3 | いい | 0.93 | 良い | 187.95 |
| 4 | ほしい | 0.80 | 悪い | 122.94 |
| 5 | 悪い | 0.73 | 多い | 103.29 |
| 6 | よい | 0.63 | よい | 99.48 |
| 7 | 優しい | 0.50 | 無い | 85.98 |
| 8 | 寂しい | 0.50 | 嬉しい | 85.24 |
| 9 | つらい | 0.50 | すごい | 81.43 |
| 10 | 楽しい | 0.49 | 欲しい | 81.20 |
| 11 | 早い | 0.40 | ほしい | 80.73 |
| 12 | 怖い | 0.38 | 楽しい | 80.61 |
| 13 | 若い | 0.38 | 辛い | 71.92 |
| 14 | 多い | 0.36 | 怖い | 63.78 |
| 15 | うまい | 0.32 | 忙しい | 63.16 |
| 16 | 悲しい | 0.31 | 優しい | 61.20 |
| 17 | 苦しい | 0.30 | 早い | 54.27 |
| 18 | 強い | 0.29 | 可愛い | 54.17 |
| 19 | 情けない | 0.25 | うまい | 50.31 |
| 20 | おかしい | 0.24 | 長い | 49.91 |
| 21 | 悔しい | 0.24 | 強い | 49.88 |
| 22 | うれしい | 0.24 | 高い | 48.82 |
| 23 | 幼い | 0.24 | おかしい | 46.80 |
| 24 | ひどい | 0.24 | 少ない | 45.57 |
| 25 | 遅い | 0.24 | 寂しい | 45.09 |

のような身近な人についての語が多く出現しているが、家族を表すような語は上位にはないため、家族よりも恋人や友達に関する質問が多いと考えられる。

表 15 からわかるように、形容詞では「人生案内」と「恋愛・人生相談」の両方に共通して、「いい」「よい」「良い」「悪い」「優しい」「強い」「おかしい」のような評価を表す語、「つらい」「辛い」「楽しい」「怖い」のような感情を表す語が含まれる。しかし、「人生案内」では、「寂しい」「つらい」「辛い」「怖い」「悲しい」「苦しい」「情けない」「悔しい」「ひどい」のようなネガティブな感情を表す語が多く含まれるのに対し、「恋愛・人生相談」では、「嬉しい」「すごい」「楽しい」「優しい」「可愛い」「うまい」のようなポジティブな感情や評価を表す語が多く含まれる。これは、「人生相談」の質問が夫婦や家族のような固定された人間関係についてのものが多いのに対し、「恋愛・人生相談」の質問が恋人や友人のような解消可能な人間関係についてのものが多く、質問者の年齢が若いため、前向きな表現が含まれているからだと思われる。

「人生案内」にはカテゴリはないが、「恋愛・人生相談」と比較するために、手作業で同じカテゴリを付与した。表 16 に、「人生案内」のカテゴリごとの質問件数と質問件数合計に対する割合 (%) を示す。

表 16 から、「人生案内」の半数以上の質問が「夫婦・家

表 16 「人生案内」のサブカテゴリの質問件数

| カテゴリ | 質問件数 | 割合 (%) |
|---------------|------|--------|
| 恋愛相談 | 15 | 4.2 |
| 友達・仲間関係 | 18 | 5.1 |
| 夫婦・家族 | 194 | 54.8 |
| 社会・職場 | 36 | 10.2 |
| シニアライフ | 17 | 4.8 |
| 人生相談 | 30 | 8.5 |
| LGBT | 0 | 0.0 |
| DV (家庭内暴力) | 0 | 0.0 |
| いじめ相談 | 3 | 0.8 |
| 不倫関係 | 4 | 1.1 |
| その他 (恋愛・人生相談) | 37 | 10.5 |
| 合計 | 354 | 100.0 |

表 17 「恋愛・人生相談」の質問中の年齢・年代表現

| 年代 | 件数 | 年代 | 件数 |
|-----------------|--------|--------------|-------|
| 10 歳未満 | 2230 | 中学 1 年/中 1 | 290 |
| 10 代 (10~19 歳) | 2,786 | 中学 2 年/中 2 | 406 |
| 20 代 (20~29 歳) | 10,571 | 中学 3 年/中 3 | 441 |
| 30 代 (30~39 歳) | 5,811 | 高校 1 年/高 1 | 430 |
| 40 代 (40~49 歳) | 2,217 | 高校 2 年/高 2 | 463 |
| 50 代 (50~59 歳) | 753 | 高校 3 年/高 3 | 493 |
| 60 代 (60~69 歳) | 345 | 大学 1 年生/1 回生 | 120 |
| 70 代 (70~79 歳) | 154 | 大学 2 年生/2 回生 | 108 |
| 80 代 (80~89 歳) | 111 | 大学 3 年生/3 回生 | 87 |
| 90 代 (90~100 歳) | 39 | 大学 4 年生/4 回生 | 145 |
| 小学 1 年/小 1 | 31 | 大学 5 年生/5 回生 | 2 |
| 小学 2 年/小 2 | 26 | 大学 6 年生/6 回生 | 1 |
| 小学 3 年/小 3 | 40 | 大学生 | 2,229 |
| 小学 4 年/小 4 | 24 | アラサー | 220 |
| 小学 5 年/小 5 | 30 | アラフィフ | 35 |
| 小学 6 年/小 6 | 97 | アラフォー | 357 |

族)に関するものであることがわかる。

前述のように、「恋愛・人生相談」の質問者は全員のプロフィールがあるわけではないため、年代は一部の質問者についてしかわからず、また、自己申告のため、実際とは異なるプロフィールを登録している可能性もある。質問には、質問者本人の年齢・年代が書かれているとは限らないが、質問中に出てくる年齢・年代に関する表現を抽出し、年代ごとにまとめたものを表 17 に示す。例えば、「25 歳」のように年齢がそのままの表現は「20 代 (20~29 歳)」としてカウントした。また、「アラサー」「大学生」のように年齢・年代が曖昧な表現や「中学 1 年生」「高 2」のような学年の表現はそのままカウントした。

表 17 から、10 代から 20 代の若い年代について言及している質問が多く、次に 30 代~40 代についての質問が多いことがわかる。この結果は、表 7 に示した、プロフィールからわかる質問者の年代と矛盾しない。プロフィールには実際と異なる情報を登録している利用者もいると考えられるが、質問内容から、実際に、若い年代の利用者が質問を

表 18 回答の文字数 (平均/最少/最大)

| 文字数 | 人生案内 | 恋愛・人生相談 |
|------|-------|---------|
| 平均 | 464.2 | 642.9 |
| 最小 | 396 | 1 |
| 最大 | 659 | 17,542 |
| 標準偏差 | 32.36 | 622.43 |

投稿していると考えられる。

以上のことから、「人生案内」と「恋愛・人生相談」の質問は、配偶者か恋人かという違いはあるものの、自分のパートナーに関連する質問が多く、質問文には評価や感情を表す形容詞がよく用いられていることがわかった。質問者の年齢構成は、「人生案内」の方がより幅広いが、重複する層もあるため、質問内容によっては、相互に質問・回答を参照することによって、質問者にとって有用な回答を探せる可能性もある。

4.4 回答

読売新聞の「人生案内」の回答 354 件と OKWave の「恋愛・人生相談」のベストアンサー 38,253 件について、質問と同様に形態素解析を行い、回答中に出現する語を抽出した。OKWave では、質問者は回答者が投稿した回答からベストアンサーを選ぶことができるが、ベストアンサーを選ばなくても回答の投稿を締切ことができるため、ベストアンサーが決まっていなかった質問も存在する。「恋愛・人生相談」の質問 65,424 件中、ベストアンサーが選ばれていたのは 38,253 件 (58.5%) であった。

「人生案内」の回答と「恋愛・人生相談」のベストアンサーの回答の文字数の平均、最小、最大、標準偏差を表 18 に示す。また、表 19 に、回答のタイトルと本文の延べ語数と異なり語数とその平均を示す。

表 19 回答の使用語数 (延べ/異なり)

| 語数 | 人生案内 | 恋愛・人生相談 |
|----------|--------|------------|
| 延べ | 99,077 | 15,231,497 |
| 異なり | 6,241 | 47,845 |
| 延べ (平均) | 279.9 | 398.2 |
| 異なり (平均) | 125.1 | 114.5 |

表 18 からわかるように、質問と同様、「人生相談」の回答は長さのばらつきは小さいが、「恋愛・人生相談」のベストアンサーは、質問と比べて平均長も長く、長さのばらつきも大きい。しかし、異なり語数の平均はベストアンサーの方が少なく、長い回答があるからと言って、語の種類が増えるわけではなことがわかった。

質問と同様に、回答の特徴的な語を抽出するために、語の重みとして $TF-IDF$ を計算した。表 20 に「人生案内」の回答と「恋愛・人生相談」のベストアンサーに出現する語のうち、名詞について全文書に対する $TF-IDF$ を足し合わせた統合 $TF-IDF$ の上位 25 件を示す。

表 20 回答中の名詞の統合 $TF-IDF$ の上位 25 件

| 順位 | 人生案内 | | 恋愛・人生相談 | |
|----|------|-------------|---------|-------------|
| | 名詞 | 統合 $TF-IDF$ | 名詞 | 統合 $TF-IDF$ |
| 1 | の | 4.70 | の | 505.81 |
| 2 | こと | 4.08 | こと | 412.63 |
| 3 | あなた | 3.97 | 人 | 348.11 |
| 4 | 人 | 2.61 | あなた | 286.64 |
| 5 | 夫 | 2.57 | ん | 280.04 |
| 6 | さん | 2.29 | 彼 | 264.59 |
| 7 | 自分 | 2.28 | 自分 | 254.56 |
| 8 | よう | 2.25 | 事 | 245.80 |
| 9 | 息子 | 1.58 | 私 | 226.05 |
| 10 | 娘 | 1.49 | よう | 221.00 |
| 11 | もの | 1.46 | 相手 | 192.73 |
| 12 | それ | 1.41 | 貴方 | 188.97 |
| 13 | これ | 1.40 | さん | 177.97 |
| 14 | 私 | 1.37 | それ | 176.97 |
| 15 | 子ども | 1.32 | 彼女 | 171.29 |
| 16 | 方 | 1.30 | 方 | 169.85 |
| 17 | 今 | 1.29 | 好き | 163.75 |
| 18 | 的 | 1.26 | 女性 | 161.84 |
| 19 | 心 | 1.19 | 者 | 153.33 |
| 20 | 相手 | 1.17 | 気 | 152.30 |
| 21 | 気持ち | 1.16 | 何 | 146.30 |
| 22 | 何 | 1.10 | 質問 | 145.54 |
| 23 | 生活 | 1.08 | 的 | 145.43 |
| 24 | 関係 | 1.06 | 気持ち | 140.21 |
| 25 | 親 | 1.05 | 今 | 138.36 |

表 20 から、質問と回答に出現する語は類似するが多いことがわかる。回答に共通する特徴的な語としては、「あなた」「相手」「気持ち」が挙げられる。「あなた」という語が多く含まれる理由は、回答者が質問者と呼びかける形で回答を行っているため、質問者の呼称として「あなた」、質問者の配偶者や恋人の呼称として「相手」という語が多く含まれ、感情について言及するために「気持ち」という語が含まれていると考えられる。

4.5 Q&A サイトの質問者の質問行動

新聞にはない Q&A サイトに特有の質問者に可能な行動として、質問のレベルの設定と回答の評価がある。OKWave では、質問者が投稿時に「困り度」を設定できる。困り度は、level01:「暇なときにでも」、level02:「困ってます」、level03:「すぐに回答を！」の 3 つのレベルから選択できる。表 21 に、「恋愛・人生相談」カテゴリの質問者のこのカテゴリ内での質問回数を 1 回、2~9 回、10 回という 3 つのグループに分けたときの、人数と困り度ごとの質問件数と困り度内での割合を示す。質問者数は、1 回 20,417 人、2~9 回 8,715 人、10 回以上 815 人である。

また、OKWave では、質問者は、質問に対する回答を受け付ける必要がなくなったときには、回答を締め切ることができ、また、回答に満足した場合は、ベストアンサーを

選ぶことができる。質問者が回答にどれくらい満足しているか考えるため、表 22 に、質問回数ごとの質問者グループで分けたときの、質問件数合計、ベストアンサー（以下、BA）の件数合計と質問件数合計に対するその割合（BA 付与率）（%）を示す。

表 21 「恋愛・人生相談」カテゴリの質問回数グループごとの困り度

| 質問回数 | level01 | level02 | level03 |
|--------|--------------|---------------|---------------|
| | 暇なときにでも | 困ってます | すぐに回答を！ |
| 1 回 | 2,867 (14.0) | 9,752 (47.8) | 7,798 (38.2) |
| 2~9 回 | 4,216 (15.3) | 10,945 (39.6) | 12,462 (45.1) |
| 10 回以上 | 2,840 (16.3) | 5,758 (33.1) | 8,786 (50.5) |
| 合計 | 9,923 (15.2) | 26,455 (40.4) | 29,046 (44.4) |

表 22 BA の付与率

| 質問回数 | 質問件数 | BA 付与件数 (%) |
|--------|--------|---------------|
| 1 回 | 20,417 | 9,120 (44.7) |
| 2~9 回 | 27,623 | 16,669 (60.3) |
| 10 回以上 | 17,384 | 12,464 (71.7) |

表 21 からわかるように、どの質問回数グループでも、質問群の 4 割程度は困り度の高い「すぐに回答を！」である。しかし、「すぐに回答を！」の割合は、質問回数が多い質問者グループの質問群ほど高くなっている。表 22 から、質問回数が少ない質問者の方が、BA 付与率が低いことがわかる。質問回数が 10 回以上の 2 つの質問回数グループでは、BA 付与率が 7 割以上と高く、質問回数が多くても、質問をしたままにせず、回答を締め切って BA を付与するという Q&A サイトの利用ルールを守っていることがわかる。これらの結果から、複数回質問するような回答者は、回答をすぐにほしいと思っており、回答を得られた場合には、BA を付与するというルールを守ることで、Q&A サイトのコミュニティをうまく利用していると言える。

5. おわりに

本研究では、読売新聞の「人生案内」欄と OKWave の「恋愛・人生相談」カテゴリの質問と回答、質問文と回答文を分析することにより、匿名の人生相談を利用する質問者がどのような悩み・質問を抱いているのか、また、新聞と Q&A サイトでは質問と質問者がどう異なるのか、そして、インターネット以外の媒体を参照することでより適切な回答が得られる可能性があるかどうかを考察した。

その結果、共通点としては、人間関係に関する質問、特に、自分のパートナー（配偶者や恋人）に関連する質問が多く、質問文にはパートナーや身近な人間を表す語や評価や感情を表す語がよく用いられていることがわかった。相違点としては、新聞の相談者は女性の割合は高く年齢層があまり偏っていないのに対し、Q&A サイトの質問者はほぼ男女半々で 10~20 代の若年層に偏っており、回答者はそれよりも少し上の 20~30 代が多いことがわかった。

新聞の人生相談は一方通行であり、質問者と回答者が直接やりとりをしたり、回答を評価したりすることはできないため、同じ事柄に関する質問を連続して、あるいは、継続的に投稿したい質問者や、回答者と直接やりとりしたい質問者は、Q&A サイトの利用の方が利便性が高いと考えられるが、新聞と Q&A サイトの人生相談は利用者や相談内容が完全に異なるわけではなく、10~30 代の若い世代については重複しているため、相補的に利用することで有用な回答を広く探すことが可能性であると言える。

今後の課題として、Q&A サイトにおいて、質問投稿回数の多い質問者は、連続した質問や繰り返し同じ事柄についての質問を投稿している可能性があるため、そのようなサイトの常連化している質問者がどのような回答を求めて質問をしているのか、質問行動の詳細な分析を行いたい。また、今回は、質問投稿回数が少ない利用者に人気のある「恋愛・人生相談」カテゴリについてのみ質問の特徴を調べたが、同じ社会調査型の質問が多く、質問投稿件数も多い「社会問題」や「アンケート」の質問について、利用者の質問・回答行動や質問内容についても調べる必要があると考えられる。

参考文献

- [1] 日本 ABC 協会：都道府県別上位 3 紙（朝刊販売部数・世帯普及率）、詩文発行社レポート 普及率 2013 年 1 月～6 月平均 (2013).
- [2] OKWave: OKWave , OKWave (オンライン), 入手先 (http://okwave.jp/) (参照 2015-03-01).
- [3] Yahoo!JAPAN: Yahoo!知恵袋, Yahoo!JAPAN (オンライン), 入手先 (http://chiebukuro.yahoo.co.jp/) (参照 2015-03-01).
- [4] 読売新聞社：人生案内相談募集, 読売新聞 2013 年 2 月 9 日朝刊, p. 21 (2013).
- [5] 王曦虹, 小林智也, 小倉加奈代, 西本一志：Q&A サイトにおける社会調査型質問への回答者に対する信頼性判断支援システム, 情報処理学会研究報告. HCI, ヒューマンコンピュータインタラクション研究会報告, Vol. 2012, No. 18, pp. 1-8 (2012).
- [6] 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. 情報学基礎研究会報告, Vol. 95, pp. 1-8 (2009).
- [7] 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 (4)-質問タイプ分類の一致度について-, 情報処理学会研究報告. 情報学基礎研究会報告, Vol. 2010, No. 5, pp. 1-8 (2010).
- [8] 高田夏希, 山本祐輔, 小山聡, 田中克己：質問応答コンテンツに対する Web による回答補完, *DEIM 2009*, Vol. C4-6 (2009).
- [9] 瀧寛文, 森崎修司, 大平雅雄, 松本健一：Q&A コミュニティを対象とした回答の信頼性指標構築に向けた分析, 情報社会学会誌, Vol. 4, No. 2, pp. 49-57 (2009).
- [10] 田中友二, 望月崇由, 八木貴史, 徳永幸生, 杉山精：Q&A サイトにおける情報検索型質問の自動抽出, 全国大会講演論文集, Vol. 2012, No. 1, pp. 529-531 (2012).
- [11] 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).