

旅行代理店での対話における回答方法を基準とした質問分類

柳谷百音¹ 但馬康宏²

¹岡山県立大学大学院 情報系工学研究科 ²岡山県立大学 情報システム工学科

¹sk621047@cse.oka-pu.ac.jp

概要

本研究では旅行代理店での案内役として扱われる対話ロボットに着目した。このロボットはユーザからの多数の質問に回答しなければならない。ユーザが満足する回答を示すためには質問の意図を適切に把握し、それに応じて回答手法を変えることが有効であると考え、旅行代理店の対話における回答方法を基準とした質問分類の提案およびその分類を行う手法の検討を目的としている。質問の分類として、情報提示、口コミ、意見、おすすめ、最新情報の5種類の提案を行い、分類手法として単語ベクトルを特徴量としたサポートベクターマシンを実装し、分類精度の比較実験を行った。

1 研究目的・背景

近年、様々なロボットが開発されていく中、ユーザと対話し、情報や面白さを提供することを目的とした対話ロボットの研究も進められている[1][2][3]。その中には、店内で商品の案内をするロボットや、駅の窓口で切符の買い方を案内するロボットなどユーザの目的に沿った様々な対話ロボットが存在する。

本研究では、旅行代理店などのカウンター越しに来店したお客様に対して商品の販売やサービスの提供を行う、カウンターセールスで扱う対話ロボットに着目した。これらの対話ロボットは、お客様であるユーザと対話し、ユーザからの多数の質問に答えなければならない。ユーザがする質問は商品の情報を聞きたい質問や、おすすめを尋ねる質問、相手の意見を聞きたい質問などが混在している。ユーザの対話の満足度向上には情報の正確性のみが重要ではなく、質問の意図を把握し、ユーザに寄り添った回答をすることが求められる。例えば、「東京駅には様々なお店があって楽しいですよね？」という質問の場合、質問者は「はい、東京駅には何店舗のお店があります」といった回答を望んでいるのではなく、「はい、私もよくランチをして楽しんでいます。」

のような返答が望ましい。このように回答に期待することは質問によって異なる。そこで質問を、回答方法を基準に分類し、その分類に応じた適切な回答方法で答えを提供することによって、よりユーザの満足度が高い対話が実現できると考えた。本研究では最初の段階として、旅行代理店での対話における回答方法を基準とした質問分類の提案および、その分類を行うための機械学習を用いた手法の検討を目的とする。

2 関連研究

Q&A サイトにおける質問と回答の分類に関する研究として栗山らの研究[4]がある。この研究では、Yahoo!知恵袋において質問者の目的や意図がどのように質問文に影響を与えるのかを調べるための最初の段階として、質問のタイプについての分析を行った。著者らは質問を手作業で分析し、Q&A サイトに見られる質問を大きく分けて情報検索型と社会調査型、非質問型の3種類、さらにそれぞれのタイプにおいて詳細に分類し計13種類の質問のタイプを提案している。この分類を行うために、田中ら[5]や大森ら[6]の研究では機械学習や特徴表現を用いた手法を提案している。

質問者の期待に基づく質問分類として渡邊らの研究[7][8]がある。この研究ではQA コミュニティにおいて回答者の質問の選択を容易にすることを目的とした質問タイプの分類方法について述べられている。質問タイプとして事実、根拠、経験、提案、意見へ分類し、その分類を行うためのSVMを用いた手法を提案している。SVMの入力に用いる素性として質問文から決められた品詞の組み合わせの単語ベクトルを使用した場合と、人手で各カテゴリにおいて選択語を決め素性とした場合において精度の比較を行った結果、選択語を素性とした場合のほうが単語ベクトルを素性とした場合よりもF値が2倍良好であった。

3 質問の分類

本研究において質問は質問者を満足させるための回答方法を基準に5種類への分類を提案する。質問の定義について表1に示し、それぞれの分類について説明する。

表1：提案する質問のカテゴリと定義

カテゴリ	定義
情報提示	客観的な回答が答えになる質問
口コミ	過去の経験を踏まえた相手、第三者の意見を求める質問
意見	相手の主観的な意見を求める質問
おすすめ	質問者の条件に合う回答を求める質問
最新情報	時期によって変わるような最新の情報を求める質問

「情報提示」は客観的な回答が求められ、回答者による差はない質問である。このカテゴリの質問の場合には、質問者は知識や情報量が少ない人と考えられ、回答方法としてはインターネットの情報などを分かりやすく提示するのがよいと考えられる。

「口コミ」は回答者によって答えが異なる質問で、質問者は経験について知りたいことが多い。そのため、回答手法としては口コミサイトなどからの引用が効果的であると考えられ、数が多い、信頼性が高い情報が求められる。

「意見」も回答者によって答えが異なる質問であるが、「口コミ」とは異なり、経験を含まない質問も存在する。また、一人の回答者であっても、時期によって答えが異なる特徴がある。

「おすすめ」は質問者がおすすめを尋ねる質問で、レコメンドシステムのように質問者に応じた回答が求められる。そのため、「おすすめ」の質問の場合には、すぐに回答するのではなく、質問者の好みや条件を絞るような質問をすることが効果的であると考えられる。

「最新情報」は旅行代理店特有の時期に応じた質問が多いことに着目しカテゴリとして提案する。旅行や観光における質問には、その時によって回答が異なることや、インターネットよりも最新の情報を求める場合が多い。そこでこのカテゴリの質問に対しては、SNSなどの最新情報を得られるツールでの回答や、現地の人が回答することによってより質問者の求める新鮮な情報が得られると考えた。

各カテゴリにおける旅行代理店で考えられる質問

例を表2に示す。

表2：各カテゴリにおける質問文の例

カテゴリ	質問例
情報提示	東京駅までの行き方を教えてください。
口コミ	スカイツリーって楽しめますか。
意見	北海道と沖縄県どちらが好きですか。
おすすめ	地酒を楽しめるおすすめのお店はありますか。
最新情報	最近の東京ディズニーランドって混雑していますか。

関連研究と異なる点は「事実」、「根拠」の情報検索型に分類される質問を一つのカテゴリとして設定したことである。これは、「事実」、「根拠」のどちらのカテゴリにおいても回答方法はウェブサイトの記事からの引用など同じであると判断したため、今回の回答方法を基準とした分類では一つのカテゴリとした。また、今回は関連研究にはなかった「最新情報」のカテゴリを追加した。このカテゴリの追加により、質問を時間軸で分類することが可能となり、リアルタイムな情報が求められる旅行代理店では有効でありユーザの満足度が向上すると考えられる。

4 分類手法

4.1 使用したデータ

次に、本研究の分類を行うための機械学習を用いた手法について説明する。

はじめに、機械学習に用いたデータについて説明する。旅行代理店のカウンターで考えられる質問としては観光地に関する質問、旅行に関する質問が考えられる。そこで質問文のデータとしてはYahoo!知恵袋の「地域、旅行、おでかけ>観光地、行楽地」カテゴリ内の2020年12月1日から2020年12月13日までに投稿された回答済みの質問501個を用意した。旅行代理店のカウンターでは対話が行われているため、ユーザは質問を短い文で尋ねると考えられる。そこで、用意した各質問の中心となる質問部分を1文、または2文を抜粋し、今回の質問文のデータとした。そして各質問文に対して形態素解析器McCabを用いて形態素解析を行い、質問文を形態素ごとに分解し、品詞を付与した。各カテゴリの質問数および合計を表3に示す。学習データとして各カテゴリの質問の7割を使用した。

表 3: 各カテゴリにおける質問数

カテゴリ	質問数
情報提示	203
口コミ	109
意見	50
おすすめ	71
最新情報	68
合計	501

4.2 単語のベクトル化

形態素ごとに分割された質問文に対して単語ベクトルの付与を行った。今回、ベクトル化には tf-idf ベクトルと Word2Vec[9]を用いた。それぞれのベクトル化について説明する。

tf-idf ベクトルとはいくつかの文書があった場合、各文書におけるキーワードの重要性を示すための手法である。今回は各カテゴリの質問文全文をひとつの文書としてとらえ、各カテゴリにおける単語の tf-idf ベクトルの計算を行った。tf(term frequency)は単語の出現頻度で、各文書においてその単語がどれほど出現したかを表し、ある単語*i*の文書*j*内での出現回数を n_{ij} 、文書*j*のすべての単語の出現回数の和を $\sum_k n_{kj}$ とすると、式(1)で求められる。

$$tf(i, j) = \frac{n_{ij}}{\sum_k n_{kj}} \quad (1)$$

次に idf(inverse document frequency)は逆文書頻度で、ある単語が出現する文書頻度の逆数を表す。この値は複数の文書にわたって多く出現する単語は値が小さくなるのに対して一つの文書のみ何回も出現する単語は値が大きくなる。全文書数を*D*、単語*i*が出現する文書数を $df(i)$ としたときの idf は式(2)によって求められる。

$$idf(i) = \log \frac{D}{df(i) + 1} \quad (2)$$

tf 値と idf 値を利用して文書*j*における単語*i*の tf-idf 値は式(3)で求められる。

$$tf - idf_{i,j} = tf(i, j) \times idf(i) \quad (3)$$

この値を各カテゴリにおけるそれぞれの単語において求め単語ベクトルとして使用した。今回はベクトル化する単語の品詞を名詞、名詞(数、固有を除く)、自立語、全品詞の4パターン作成し、精度の比較を行った。次に、Word2Vec について説明する。Word2Vec とは文章中の単語を数値ベクトルに変換

してその意味を把握するものである。tf-idf ベクトルでは学習データに使用されている単語のみのベクトル化になるため、未知語がテストデータに含まれた場合に対処できない。そのため精度が下がることがある。そこで膨大なテキストデータを学習させた Word2Vec を使用することで未知語にも対応するようになった。Word2Vec の訓練済みベクトルとして日本語 Wikipedia エンティティベクトルを使用した。またベクトル化する単語の品詞は名詞+動詞を使用した場合と自立語を使用した場合の2パターンを作成して精度の比較を行った。

4.3 使用した分類器

次に質問文の単語ベクトルを入力とした分類器を作成し、質問文の分類を行った。本研究の分類にはサポートベクターマシンを使用した。サポートベクターマシンは通常2値分類を行う分類器である。そこで今回は各カテゴリに属するか否かの分類を行う実験1、およびサポートベクターマシンを多クラス分類に応用した実験2を行った。サポートベクターマシンを多クラス分類に適応した仕組みについて説明する。本研究ではサポートベクターマシンのツールとして scikit-learn ライブラリの SVC(Support Vector Classification)を使用した。決定関数は ovr(one versus the rest)とし、ある特定のクラスに入るか、他のどれかのクラスに入るかの2値分類をとく分類器を作成した。そして多クラス分類に適応させるために、分類の確信度のスコアとして decision_function の値を使用した。decision_function の値は入力はそのカテゴリにおける信頼度をレンジなしで示すものであり、値が大きいほどそのクラスに対する確信度が高い。そこで入力に対して最も大きい値を出力したクラスをその質問文のカテゴリとして多クラス分類を行った。

5 実験結果・考察

提案手法を用いて素性別、カテゴリ別の分類精度について実験した。使用したデータは Yahoo!知恵袋の「地域、旅行、おでかけ>観光地、行楽地」カテゴリ内の2020年12月1日から2020年12月13日までに投稿された回答済みの質問501個で、7割を学習データ、3割をテストデータとし、3回の交差検証を行った。

ⁱ <https://github.com/singletongue/WikiEntVec>

表 4：実験 1 における各カテゴリ素性別の正解率

	TF-IDF ベクトル				Word2Vec ベクトル			
	BL	名詞	名詞(数,固有除く)	自立語	全品詞	名詞+動詞	自立語	
情報提示	0.407	0.604	0.624	0.578	0.593	0.667	0.664	
ロコミ	0.780	0.678	0.613	0.700	0.718	0.709	0.711	
意見	0.900	0.813	0.809	0.829	0.833	0.802	0.798	
おすすめ	0.860	0.909	0.922	0.902	0.887	0.856	0.862	
最新	0.867	0.827	0.858	0.813	0.813	0.851	0.849	

5.1 実験 1

各カテゴリに属するか否かの分類を行った実験結果について、各カテゴリの正解率を表 4 に示す。ベースラインはすべての質問を情報提示に分類した場合の正解率である。

正解率における素性別の結果については、tf-idf ベクトルを使用した場合において素性による差がほぼ見られなかった。情報提示、おすすめ、最新情報では名詞（数，固有を除く），ロコミと意見においては全品詞の正解率が最も高かった。Word2Vec では名詞+動詞を素性とした場合が良好であるが tf-idf ベクトルとほぼ同じ正解率であった。カテゴリ別における結果についてはどのカテゴリにおいても 0.6 を上回る正解率の素性が見られた。おすすめ，最新情報の値が高く，情報提示が最も低い正解率であった。交差検証において分散が最も小さいのは Word2Vec の名詞+動詞を素性とした場合で平均値も最も高かった。カテゴリ別の比較ではおすすめが他と比べて圧倒的に高いが素性によるばらつきは見られた。ロコミ，意見はどの素性においても低い結果であった。

5.2 実験 2

次にサポートベクターマシンを用いて多クラス分類を行った。その結果を表 5 に示す。

表 5 は各カテゴリの再現率，および全体の正解率を示している。実験 2 のにおいては名詞，名詞 2（数，固有を除いた名詞），自立語，全品詞を素性とした tf-idf ベクトルと自立語を素性とした Word2Vec を単語ベクトルとして使用した。

表 5：実験 2 における再現率，全体の正解率

	名詞	名詞 2	全品詞	自立語	W2V
情報提示	0.361	0.333	0.366	0.322	0.404
ロコミ	0.242	0.323	0.313	0.293	0.333
意見	0.244	0.289	0.378	0.378	0.311
おすすめ	0.841	0.841	0.825	0.873	0.714
最新情報	0.550	0.450	0.383	0.417	0.450
正解率	0.416	0.413	0.422	0.411	0.429

再現率の結果について素性別にみると，情報提示，ロコミでは Word2Vec を使用した場合，意見，おすすめでは自立語，最新情報では名詞を用いた場合の再現率が最も高くなった。しかし，平均を取ると素性による差はほぼ見られない。また，カテゴリ別に比較すると，おすすめの再現率が 0.8 と圧倒的に高く，ロコミ，意見は 0.3 前後と低い。全体の正解率においては素性別の差はほぼ見られず，0.4 前後であった。

6 まとめ

本研究では旅行代理店での対話における回答方法を基準とした分類の提案，および機械学習を用いてその分類を行う手法の検討を目的とした。回答方法を基準とした分類として，「情報提示」「ロコミ」「意見」「おすすめ」「最新情報」の 5 種類への分類を提案した。分類手法としては tf-idf ベクトルや Word2Vec の単語ベクトルを特徴量とした SVM の実装を行い，素性別やカテゴリ別における分類精度を比較した。実験の結果素性による差は大きくはないものの，名詞が分類を左右する特徴になっていることが分かった。今後の課題として分類器にサポートベクターマシン以外を使用すること，精度を高めるための素性の探究が挙げられる。

参考文献

- [1] 吉田裕介, 萩原将文, “複数の言語資源を用いたユーモアを含む対話システム”, 日本知能情報フエジ学会誌, vol.26, No.2, pp.627-636, 2014.
- [2] 藤倉将平, 小川義人, 菊池英明, “非タスク指向対話システムにおけるユーモア応答生成手法”, 人工知能学会全国大会論文集, pp.1-4, 2015.
- [3] 赤崎智, 鍛冶伸裕, “知的対話アシスタントにおける雑談を目的としたユーザ発話の検出”, 情報処理学会研究報告, vol.2017-NL-231, No.18 pp.1-9, 2017.
- [4] 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.2019-F1-95, No.19 pp.1-8, 2009.
- [5] 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”, 第 74 回全国大会講演論文集, pp.529-530, 2012.
- [6] 大森勇輔, 森田和宏, 泓田正雄, 青江順一, “疑似訓練データを用いた Q&A サイトの質問分類”, 言語処理学会第 21 回年次大会発表論文集, pp.489-492, 2015.
- [7] 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”, DEIM Forum 2011, B5-1, 2011.
- [8] 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.
- [9] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean, “Efficient Estimation of Word Representations in Vector Space” ICLR Workshop, 2013.