

消費者の意見に基づく商品検索

杉木 健二^{†1} 松原 茂樹^{†2}

自然言語で記述された検索クエリに適応する商品検索方式を提案する。本方式では、検索クエリの適合対象として商品レビューを使用する。商品レビューは、消費者の意見など商品提供者が明示していない情報を含んでおり、これを用いることによりユーザの要求の多様性に対応することができる。システムは、検索クエリに記された条件に合致する記述が商品レビューに存在すれば、その商品は利用者の要求に適合した商品であると見なし、それを利用者に提示する。提案方式に基づく宿泊施設検索システム「宿探」を実装した。自然言語クエリを用いた検索実験により、本方式による検索可能性を確認した。

Product Retrieval Method Based on Customer Reviews

KENJI SUGIKI ^{†1} and SHIGEKI MATSUBARA ^{†2}

In this paper, we propose a method for retrieving products that match users' search queries written in natural language. We use a large amount of product reviews written by consumers, in order to match those queries with products. Since reviews include a lot of valuable information which commerce service providers don't supply, by using these reviews our system can respond to a wide variety of users' requests, especially to subjective or ambiguous requests. If a review includes a description that coincides with the condition described in a user's query, the product to which the review refers is considered the product that matches the query and is shown to the user by our system. Using our proposed method, we implemented an accommodation retrieval system named "Yado-tan". The result of search experiments using natural language queries confirmed the retrieval capability of using our method.

^{†1} 名古屋大学大学院情報科学研究科

Graduate School of Information Science, Nagoya University

^{†2} 名古屋大学情報連携基盤センター

Information Technology Center, Nagoya University

1. まえがき

近年、インターネット利用者の増大に伴い、楽天や Amazon など、電子商取引 (EC) サイトが急増している。一般に、これらのサイトで扱う商品の数は膨大であり、探したい商品に効率的にアクセスできる環境を備えていることが望ましい。しかし、その多くは、使用可能な検索項目が、品名、年式、金額など、あらかじめ定められたものに限定されており、ユーザの多様な検索要求に適応することは難しい。

これに対して、検索要求をユーザが自然言語で記述し、それをシステムへの入力とすることが考えられる。これまで、自然言語インタフェースを備えた商品検索システム¹⁾、対話的な商品推薦システム^{2),3)}などが提案されている。しかし、自然言語クエリをデータベース言語に変換する方式では、システム設計者が変換ルールを網羅的に作成することが不可欠であり、現実的ではない。

本論文では、自然言語で自由に表現された検索クエリに適応する商品検索方式を提案する。本方式では、自然言語クエリの適合対象として、商品レビューを使用する。商品レビューは、商品提供者が明示していない、消費者の視点に基づく情報を含んでおり、ユーザの多様な要求に対応することができる。商品レビューは、Amazon、楽天などの EC サイトをはじめ、@cosme、価格.com などの口コミサイト上に大量に掲載されており、また、商品に関する意見が書かれたブログも多数存在する。商品レビューは、消費者による商品購入の判断材料として大いに活用されており、このような情報を検索用データとして使用することは有意義である。

提案方式では、検索条件に合致する記述が、商品レビューテキストに存在すれば、その商品はユーザの要求に適合した商品であると見なす。この方式による検索可能性を示すために、自然言語による宿泊施設検索システム「宿探」を実現した。20,588 施設に対する 700,826 件の意見を用いて検索実験を行い、「ベッドが広くて寝心地が良い宿」など、従来の商品検索方式では対応が難しかった検索クエリに適応できることを確認した。

2. 意見情報と要求情報

提案手法では、自然言語で表現された検索クエリに対して、該当する情報がレビューに記載されている商品を検索結果として提示する。例えば、図 1 左の検索クエリのように、主観性の強い要求に対しても、プラズマテレビ「TH-42P×500」の商品のレビュー中に図 1 右の意見文が存在すれば、「TH-42P×500」はこのクエリに適合した商品であると見なすことが



図 1 検索クエリに対応するレビューテキストの例
Fig. 1 Example of a query and a review

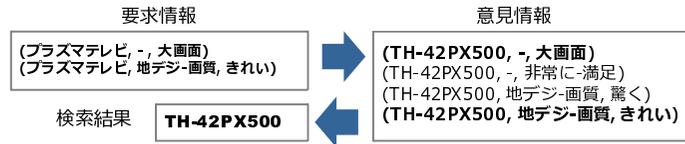


図 2 要求情報と意見情報の適合の例
Fig. 2 Example of matching query-tuples with review-tuples

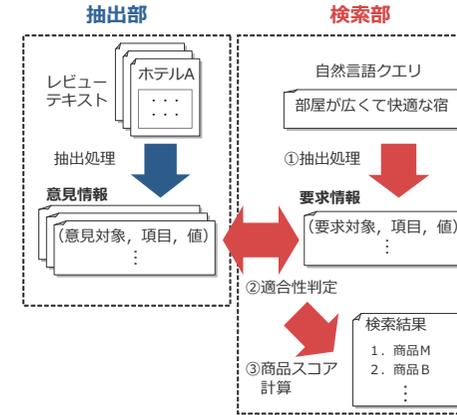


図 3 商品検索のシステム構成
Fig. 3 Overview of the product retrieval system

できる。

検索クエリに対するレビューの適合性を測るために、自然言語文を意味表現に変換する。本研究では、そのような意味表現として、**対象、項目、値**からなる3項組を用いる。ここで、「対象」は商品名あるいは商品カテゴリを、「項目」は商品の属性を、「値」は商品の属性値を意味する。これは、ユーザが商品を検索する場合、「色は赤でデザインがシンプルで音質がクリアなMP3プレイヤー」のように、属性とその値を検索条件とすることが多いためである。

商品レビューからは“(意見対象, 項目, 値)”を抽出し、これを**意見情報**と呼ぶ。例えば、商品名「TH-42P×500」の意見テキストに「画質がきれい」という記述があれば、(TH-42P×500, 画質, きれい)を意見テキストとして抽出する。一方、検索クエリからは“(要求対象, 項目, 値)”を抽出し、これを**要求情報**と呼ぶ。例えば、検索クエリ「画質がきれいなプラズマテレビ」からは、(プラズマテレビ, 画質, きれい)を要求情報として抽出する。この場合、意見情報と要求情報を比較すると、項目及び値が等しく、かつ、意見対象が要求対象に含まれるため、この意見情報は要求情報に適合すると判断できる。図2に、検索クエリ及び商品レビューから抽出された要求情報及び意見情報、ならびに、その適合の例を示す。すべての要求情報に対して、適合する意見情報が存在するため、プラズマテレビ「TH-42P×500」は、検索クエリに合致した商品として提示できる。

なお、本研究で使用する意味表現は、評判情報に関する従来の研究⁴⁾⁻⁷⁾と同一の形式で

あり、商品の特徴を表現するのに適している。これらの研究は、評判情報の抽出が目的であり、これまで、評判に関わる表現の抽出法や評判表現辞書の構築法などが検討されてきた。一方、本研究では、評判などの主観的な情報だけでなく、事実などの客観的な情報も抽出し、利用する。

3. 商品レビューを用いた商品検索

3.1 商品検索方式の概要

提案方式のシステム構成を図3に示す。以下では、対象商品として宿泊施設を例に説明する。システムは、レビューテキストから意見情報を取り出す抽出部、ならびに、入力された検索クエリに合致する商品を見つける検索部から構成される。

抽出部では、レビューテキストの各文に、係り受けパターンに基づく変換規則を適用することにより意見情報を抽出する。一方、検索部では、検索クエリから要求情報を抽出し、要求情報に対する意見情報の適合を検査し、該当する商品を検索結果として提示する。

3.2 意見情報の抽出

本研究では、商品レビューとしてECサイトに記載されている、消費者の意見文章を使用することを想定している。すなわち、意見情報の要素の一つである「意見対象」についてはレビューが対象とする商品名とすればよい。以下では、残りの要素である「項目」及び

「値」の抽出方法について述べる。

3.2.1 意見情報への変換規則

意見情報を構成する項目と値の組は、レビューテキストでは主に、

- 主語・述語の関係

「料理がおいしい」⇒(A ホテル, 料理, おいしい)

- 修飾・被修飾関係

「きれいな部屋」⇒(B ホテル, 部屋, きれい)

として出現する。これらは一般に係り受けの関係にあることが多く、本研究では、係り受けの出現パターンに基づいて、図4に示す4つの変換規則を作成した。ここで、 $X \rightarrow Y$ でXがYに係ることを示し、 $A \Rightarrow B$ は「Aという係り受け関係があれば、意見情報Bを抽出する」ことを意味する。

ただし、本研究では、より詳細な情報を抽出するため、以下の場合には、項目や値をそれぞれ複数の文節から抽出する。すなわち、変換規則(1),(2)では、Xの名詞として「名詞 + の + 名詞」のような名詞句、またYとして「副詞 + 形容詞」「名詞 + (に/で/を) + 動詞」のような被修飾語句も対象とする。例えば、「フロントの対応 → 親切」からは(フロント-対応, 親切)を、「コンビニが → 近くにある」からは(コンビニ, 近く-ある)を抽出する。

3.2.2 意見情報抽出の手順

意見情報を抽出する手順は以下の通りである。

前処理 商品レビューから抽出した文に対して係り受け解析を実行し、意見情報の抽出対象とする。ただし、「～して欲しい」「～が望ましい」「～ば嬉しい」など、期待、願望、依頼を表す文末表現を含む文は、現行の商品に備わっていない特徴が記述されており、対象として適さないため除去する。

抽出処理 図4に示した4つの変換ルールを適用し、意見情報を抽出する。項目と値が主語・述語関係となる場合には(1),(2)が、項目と値が被修飾・修飾関係となる場合には(3)が適用され、項目候補が存在しない場合は(4)が適用されることになる。

後処理 特定の品詞や不要語を除去する。項目としては、名詞及び未知語以外の品詞を、また、値としては、名詞、形容詞、動詞、副詞、接頭辞、接尾辞以外の品詞を除去する。また、「方」、「こと」、「もの」、「思う」、「考える」など、意見情報として重要な意味をもたない語を除去する。

抽出例として、「部屋も広く、浴室の浴槽も広く、ゆったりとくつろげるので大変気に入

<p>(1) $X \rightarrow Y \Rightarrow (O, X, Y)$ X: 名詞 + (は/が/も) Y: 動詞, 形容詞, サ変名詞+する 例) 部屋が → きれい ⇒ (A ホテル, 部屋, きれい)</p>
<p>(2) $X \rightarrow Y_1 \rightarrow Y_2 \Rightarrow (O, X, Y_1), (O, X, Y_2)$ X: 名詞 + (は/が/も) Y₁, Y₂: 動詞, 形容詞, サ変名詞+する 例) 部屋が → きれいで → 快適でした ⇒ (A ホテル, 部屋, きれい), (A ホテル, 部屋, 快適)</p>
<p>(3) $Y \rightarrow X \Rightarrow (O, X, Y)$ X: 名詞 + (は/が/も/を/に/だ/です) Y: 形容詞 例) きれいな → 部屋でした ⇒ (A ホテル, 部屋, きれい)</p>
<p>(4) $Y \Rightarrow (O, , Y)$ Y: 動詞, 形容詞, サ変名詞+する 例) 親切で ⇒ (A ホテル, , 親切)</p>

O: 意見対象

図4 意見情報への変換規則

Fig.4 Transformation rules into review-tuples

ています。」というレビューテキストの文から意見情報を抽出する過程を図5に示す。係り受け解析により生成された係り受け構造に対して、変換規則の適用可能な部分構造を取り出す。規則を適用した結果、5つの意見情報を抽出することができる。

3.3 商品の検索

入力された検索クエリをもとに、商品がユーザの検索要求を充足する程度(以下、商品スコア)を計算し、スコア順に従って商品を提示する。以下では、商品スコアを算出する方法について述べる。

3.3.1 要求情報の抽出

自然言語で記された検索クエリから要求情報の集合を抽出する。本研究では、「部屋が広い宿」のように、検索条件及び要求対象から構成される名詞句によってクエリを表現することを想定しており、名詞句の主要語(head)を要求対象として抽出する。また、項目、値に

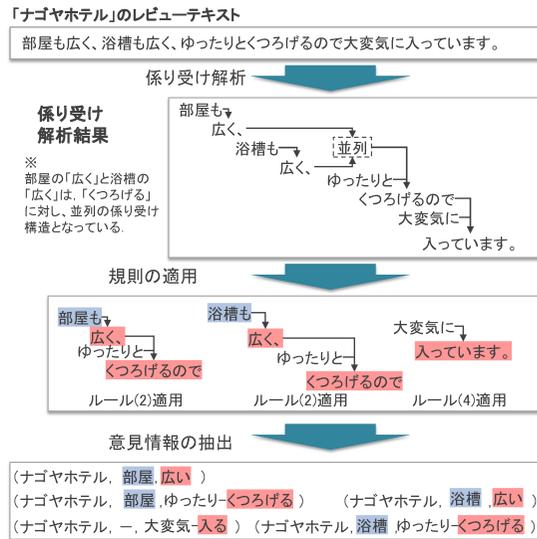


図5 抽出処理の例
Fig. 5 Example of review-tuples extraction

については、意見情報の抽出と同様に、変換規則の適用により獲得する。検索クエリ「部屋が広くくつろげる宿」からは、要求情報集合 $\{(宿, 部屋, 広い), (宿, 部屋, くつろげる)\}$ を抽出する。

3.3.2 要求情報に対する意見情報の適合性判定

獲得した各要求情報に対して、意見情報の適合性を計算する。図6に要求情報 q と意見情報 o との適合度 $c_rate(q, o)$ の計算法を示す。すなわち、項目及び値の形態素レベルでの適合率の積により算出する。例えば、要求情報 $q = (宿, 部屋, とても-広い)$ に対する意見情報 $o = (A \text{ ホテル}, 部屋, 広い)$ の適合度は $c_rate(q, o) = 1 \times \frac{1}{2}$ となる。ただし、要求情報に項目が存在しなければ、項目に関する適合度を $\alpha (0 \leq \alpha \leq 1)$ として計算する。また、適合度計算は、意見対象が要求対象に合致した意見情報に対してのみ実行する。

3.3.3 商品スコアの計算

各商品に対する意見情報が、検索クエリから抽出された要求情報をどの程度充足するかに従って、商品スコアを以下の通り算出する。検索クエリ q から抽出した要求情報の集合を $R(q)$ 、商品 p のレビューから抽出した意見情報の集合を $O(p)$ とするとき、 q に対する p の

$$c_rate(r, o) = \begin{cases} \frac{I_{ro}}{I_r} \times \frac{V_{ro}}{V_r} & (\text{項目がある}) \\ \alpha \times \frac{V_{ro}}{V_r} & (\text{項目がない}) \end{cases}$$

I_{ro} : 要求情報の項目中の一致形態素数
 I_r : 要求情報の項目中の全形態素数
 V_{ro} : 要求情報の値中の一致形態素数
 V_r : 要求情報の値中の全形態素数

※ただし、値は末尾の形態素が一致する必要あり

図6 要求情報と意見情報の適合度計算

Fig. 6 Calculation of concordance rate between a query-tuple and a review-tuple

$$Score(q, p) = \sum_{r_i \in R(q)} PF_i \cdot IOF_i$$

$$PF_i = \sum_{o_j \in O(p)} pf_j \times c_rate(r_i, o_j)$$

$$IOF_i = \log \frac{ON}{of_i}$$

pf_j : 商品 p のレビューでの意見情報 o_j の出現頻度
 of_i : クエリ中の要求情報 r_i との適合率 > 0 である意見情報が出現する意見対象数
 $|P|$: 意見対象の総数

図7 商品スコアの計算

Fig. 7 Calculation of a product score

商品スコア $Score(q, p)$ を、図7に従って計算する。ここで、 PF_i 、 IOF_i は、それぞれ

- ある商品に対して、同一の意見を記す消費者が多ければ、それはその商品の特徴付ける意見である。
- ある要求に適合する意見が少なければ、それはその商品の特徴付ける意見である。

という見方を反映している。すなわち、要求情報に適合する意見情報が多く、かつ、適合しにくい要求を満たす商品ほど、商品スコアが大きくなる。

Q#	検索クエリ	正解件数
(1)	チェックアウト後も荷物を預かってもらえる宿	10/10
(2)	食事が和食と洋食で選べる宿	6/10
(3)	チェックアウト時間が遅い宿	8/10
(4)	風呂が広くてアメニティが充実している宿	9/10
(5)	部屋の照明が明るい宿	10/10
(6)	周辺が閑静な雰囲気、落ち着いて過ごせる宿	8/10
(7)	コンビニやレストランが近くて、食事に困らない宿	10/10
(8)	出張に快適で、アクセスが便利な宿	10/10
(9)	料理がおいしくて、お風呂が貸し切りできる宿	7/10
(10)	ベッドが広くて、寝心地がよい宿	10/10
検索精度 (上位 10 件の精度)		0.86

図 8 実験に用いた 10 個の検索クエリ検索結果
Fig. 8 Retrieval results for 10 queries

4. 宿泊施設検索システム

4.1 システムの概要

提案方式の実現可能性を示すために、宿泊施設を対象とした検索システム「宿探」を実現した。利用者は、宿泊施設の検索クエリを自然言語で自由に記すことができる。図 9 に検索クエリ「チェックアウト後も荷物を預かってもらえる宿」に対する検索結果の提示画面を記す。

意見レビューとして、宿泊予約サイト「楽天トラベル」の意見投稿サイト「お客さまの声」^{*1}を使用した。実装では、20,588 施設に対する意見テキスト 700,826 件を登録している。なお、係り受け解析には KNP⁸⁾を使用し、図 6 の係数 α の値は 0.02 とした。

4.2 検索実験

4.2.1 実験方法

宿泊施設が提供する情報だけでは、検索することが難しいと思われる条件を含む検索クエ

リを、レビューとは独立に 10 個設定し、検索結果の正しさを評価した。使用した検索クエリを図 8 左に示す。評価は、商品スコアの高い上位 10 件の宿泊施設を対象にそのレビューテキストをすべて閲覧し、検索クエリに対する施設の適合性を判定することにより行った。

4.2.2 実験結果

システムが提示した上位 10 件の宿泊施設のうち、クエリに適合した件数を図 8 右に示す。上位 10 件の精度にて 86.0% を示しており、自然言語で表現されたクエリに対する本手法による検索可能性を確認できた。

検索クエリに適合した宿泊施設では、以下のように要求に合致する意見文がみられた。

- チェックアウト後も、午後 10 時過ぎまで預かっていただいた上に、大変親切に対応していただきました。(クエリ 1)
- 部屋が広く、アメニティも充実しており、きもちよく宿泊できました。(クエリ 4)
- 室内の照明が明るいのが助かりました。(クエリ 5)
- 駅にも近く、都心施設へのアクセスが便利なので満足しています。周囲は落ち着いていますし、コンビニが数店あり、レストランも多く、出張するのが楽しみなホテルのひとつです。(クエリ 8)

一方、クエリに適合しない商品が提示された理由として、

- 要求に完全に合致する意見が少なかったため、適合性の低い商品が検索結果の上位に提示された。
- 要求情報に反する意味をもつ意見情報を適合性が高いと判定し、その商品が提示された。などが挙げられる。いずれも、要求情報と意見情報との間の適合性計算に起因しており、今後、自然言語の意味的等価性とその判定について改善を図る必要がある。

5. おわりに

本論文では、ユーザの検索要求の多様性に対応できる商品検索方式を提案した。本方式では、消費者が投稿した意見情報を活用することにより、商品提供者が明示していない情報に対しても、条件に該当する商品を提示できる。本方式に基づいて、宿泊施設検索システムを実装した。自然言語クエリを用いた検索実験により、本方式による検索可能性を確認した。

今後の課題として、まず、検索の再現性の向上が挙げられる。意見情報と要求情報の適合は、現在、項目及び値の一致により行っているが、表現の同義性や類義性を考慮することにより、要求に合致する意見を広く取得できる可能性がある。また、意見テキストを対象とした検索のスコアリング手法の検討が必要である。本研究では、各商品に対する意見全体を一

*1 http://travel.rakuten.co.jp/auto/tabimado_bbs_top.html



図9 宿泊施設検索システム「宿探」の検索結果画面
Fig.9 Screenshot of accommodation retrieval system “Yado-tan”

つのテキストとみなしてスコア計算を定めたが、実際には、意見テキストは、複数の消費者の意見の集まりであり、互いに矛盾することもある。このようなテキストの特徴をふまえたスコアリング方式を考案する必要がある。さらに、大量の検索クエリを用いた定量的評価、及び、被験者による評価実験も今後の課題である。

謝辞 日頃から御指導頂いている名古屋大学大学院情報科学研究科教授の渡邊豊英先生に感謝致します。

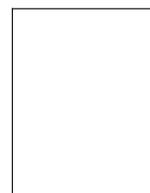
参考文献

- 1) Dittenbach, M., Merkl, D., and Berger, H.: A Natural Language Query Interface for Tourism Information, *Proceedings of the 10th International Conference on Information Technologies in Tourism (ENTER-2003)*, pp.152-162 (2003).
- 2) Chai, J., Horvath, V., Nicolov, N., Stys, M., and Kambhatla, N.: Natural Language Assistant: A Dialog System for Online Product Recommendation, *AI Magazine*, Vol. 23, No. 2, pp.63-75 (2002).
- 3) Mcsherry, D.: Explanation in Recommender Systems, *Artificial Intelligence Re-*

view, Vol. 24, No. 2, pp.179-197 (2005).

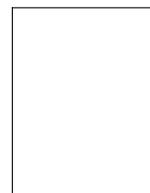
- 4) Hu, M. and Liu, B.: Mining Opinion Features in Customer Reviews, *Proceedings of the 19th National Conference on Artificial Intelligence (AAAI-2004)*, pp.755-760 (2004).
- 5) Popescu, A. and Etzioni, O.: Extracting Product Features and Opinions from Reviews, *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP-2005)*, pp.339-346 (2005).
- 6) 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一: テキストマイニングによる評価表現の収集, *情報処理学会研究報告*, NL-154, pp.77-84 (2003).
- 7) 立石健二, 福島俊一, 小林のぞみ, 高橋哲朗, 藤田篤, 乾健太郎, 松本裕治: Web 文書集合からの意見情報抽出と着眼点に基づく要約生成, *情報処理学会研究報告*, NL-163, pp.1-8 (2004).
- 8) 黒橋禎夫: 日本語構文解析システム KNP version 2.0 (1998).

(平成 19 年 10 月 15 日受付)
(平成 20 年 1 月 8 日採録)



杉木 健二 (学生会員)

2005 年名古屋大学工学部電気電子・情報工学科卒業。2007 年同大学院情報科学研究科社会システム情報学専攻博士前期課程修了。現在、同後期課程に在学中。自然言語処理, 情報検索, デジタル図書館の研究に従事。日本図書館情報学会会員。



松原 茂樹 (正会員)

1998 年名古屋大学大学院工学研究科博士後期課程修了。博士 (工学)。同年, 同大学言語文化部助手。2002 年名古屋大学情報連携基盤センター助教授。現在に至る。この間, 日本学術振興会特別研究員, ATR 音声言語コミュニケーション研究所客員研究員, 情報通信研究機構専攻研究員を兼任。自然言語処理, 音声言語処理, 情報検索, デジタル図書館の研究に従事。IEEE, ACM, 電子情報通信学会, 言語処理学会等会員。