DBSJ Japanese Journal
Vol. 14-J, Article No. 7, March 2016

経験的属性によるオブジェクト検索
Object Search by Experience Attributes

内田 臣了♡ 山本 岳洋♢ 加藤 誠♢
大島 裕明♢ 田中 克己♢

Shinryo UCHIDA Takehiro YAMAMOTO
Makoto P. KATO Hiroaki OHSHIMA
Katsumi TANAKA

本研究ではウェブ上のユーザレビュー文からオブジェクトの経験的属性を推定し発見する手法を提案する．既存の検索システムで「持ち運びに便利なカメラ」や「夜景に強いカメラ」といった経験的属性に基づく検索を行った場合，オブジェクトの関連文書に経験的属性が記述されていなければオブジェクトを発見することができない．また，スペック情報に含まれるような探索的属性から経験的属性を直接推定するのは，専門知識が必要となるためすべての人にとって簡単なタスクではない．本研究では収集したユーザレビュー文中の比較表現や接続詞などにより構成される論理構造を手がかりに，ある探索的・経験的属性と異なる経験的属性の間の対応規則を学習する．オブジェクトの探索的属性情報からそのオブジェクトのもつ経験的属性を推定することで，経験的属性に基づいた検索を実現する．

In this paper, we propose a method of predicting experience attributes of an object from the user reviews on the Web. Although experience attributes are quite important for users to compare objects, it is not easy for them to know their experience attributes. For example, when users search for cameras by experience attributes such as “portable cameras” or “cameras which can capture night scenes” in a e-commercial site, they cannot find such cameras unless their related documents contain such descriptions. To enable users to search for objects by their experience attributes, we propose a method to predict the value of an experience attribute of an object from its search attributes and rank the objects in accordance with their experience attribute values.

1. はじめに

Amazon.com[1]や価格.com[2]のような EC サイトの普及により，ウェブ上で異なる複数のオブジェクトを比較する機会が増えている．ユーザは数ある商品を比較し，より自身の要求に適合する商品を選択し購入する．そのために多くの EC サイトでは，様々な属性によってオブジェクトをランキングしたりフィルタリングする機能をユーザに提供している．例えば，カメラの購入を検討しているユーザは，カメラの価格，重さ，性能といった属性によりカメラをランキングすることで，自身の要求に適合したカメラを探す．こうしたオブジェクト検索は，ユーザの意思決定を支援する重要なアプローチの 1 つである．

本論文では，オブジェクトのもつ属性を探索的属性と経験的属性の 2 つに分類する．そのオブジェクトを使用せずとも評価できる属性のことを探索的属性という．例えば，オブジェクトの重さや形，機能の有無は探索的属性である．一方，人がそのオブジェクトを使うことで初めて評価できる属性のことを経験的属性という．例えば，カメラがもちうる経験的属性としては「持ち運びやすい」,「夜景が綺麗に撮れる」,「動く被写体に強い」などが考えられる．

この 2 種類の属性のうち，経験的属性はオブジェクトを探すユーザにとって重要な観点の 1 つである．一般的に，ユーザがオブジェクトを探す際はある使用目的をもっている場合が多く，その目的を達成できるオブジェクトを購入する．したがって，そのオブジェクトを使用することで得られる経験が分かると，ユーザはそのオブジェクトを購入するかどうかの意思決定を容易に行うことができる．

しかし，ユーザにとってオブジェクトの経験的属性を知ることは容易ではない．EC サイトの商品情報には探索的属性のみが記載されている場合が多く，経験的属性が記載されていることは稀であるためである．そのため，ユーザはこうした探索的属性のみから経験的属性を推定する必要がある．しかし，経験的属性の推定には，そのドメインに関する専門的な知識が必要となるため，一般的なユーザにとって難しいことも多い．例えば，撮像素子の大きさが「APS-C」のカメラと撮像素子の大きさが「中判」のカメラが存在した場合に，どちらのカメラがより綺麗な写真が撮れるか判断するのは，専門的な知識をもたないユーザにとっては難しいことである．

ユーザが経験的属性を推定する手段として，レビュー文に記述されている経験的属性に注目することが考えられる．例えば，レビュー文に「このカメラは持ち運びやすい」と記述されていた場合，そのカメラは「持ち運びやすい」という経験的属性をもつと考える．ここで問題となるのが，レビュー文にそのオブジェクトのもつ経験的属性がすべて記述されるとは限らないという点である．レビュー文に「持ち運びやすい」と記述されていなかったオブジェクトでも「持ち運びやすい」という経験的属性をもつことは多いと考えられる．そのため，このアプローチをとった場合は再現率の問題が生じる．また，このアプローチには経験的属性の大きさが分からないという問題が存在する.「持ち運びやすい」と記述されたオブジェクトが複数存在した場合，どのオブジェクトがより持ち運びやすいのか判断できない．つまり，オブジェクトを「持ち運びやすい」順に並び替えることが不可能である．

本研究では，オブジェクトのもつ探索的属性から経験的属性を推定する手法を提案し，経験的属性に基づいた検索を実現する．提案手法は，レビュー文中に記述される比較表現に注目し，オブジェクト間の順序関係を抽出する．また，レビュー文の論理構造に注目し，属性と属性の依存関係を抽出する．収集したそれらの関係を集約し，オブジェクトのもつ経験的属性を推定する．この手法により,「持ち運びやすい」といった経験的属性をクエリとして受け取り，クエリに適合する順にカメラをランキングしユーザに提示するという経験的属性に基づくオブジェクト検索が可能となる．

2. 関連研究

ウェブ上の意見文から有用な情報の抽出を行う研究について述べる．意見文からの情報抽出における主な課題は 3 つ組（対象，属性，評価）の抽出である．立石ら [10] は，文書中での共起に注

♡ 学生会員　京都大学大学院情報学研究科
uchida@dl.kuis.kyoto-u.ac.jp
♢ 正会員　京都大学大学院情報学研究科
{tyamamot, kato, ohshima, tanaka}@dl.kuis.kyoto-u.ac.jp

[1]http://www.amazon.com
[2]http://kakaku.com

目することで，意見文から 3 つ組のうち対象物と評価表現を抽出し，その極性を判定した．また，立石らは [10] を発展させた研究として，意見文から属性と評価表現をブートストラップ的に抽出する手法を提案した．さらに，抽出した評価表現の極性を判定し，各属性を軸としたレーダーチャート形式での可視化を行った [11]．Liu ら [6] は supervised な相関ルールマイニングを用いてルールを生成し属性を抽出する手法を提案した．また，Liu らも立石らと同様に各属性ごとに評価値の可視化を行った．

単一のオブジェクトの評価情報にとどまらず複数のオブジェクトの比較関係を抽出する研究も行われている．Li ら [5] は weakly-supervised なブートストラップ法を用いることで，比較を意図した質問文から比較の対象となったオブジェクトの対を高い精度で抽出した．Li らでは語の一般化，具体化を行って抽出精度の向上を図っている．佐藤ら [8] はウェブ上のブログ記事から 4 つ組（対象，基準，属性，評価）または 3 つ組（対象，基準，評価）の抽出を行った．比較表現に注目したパターンマッチングに加え，構文情報とセンタリング理論を用いることで高精度での情報抽出を実現した．

本研究も意見文からの情報抽出に取り組む．本研究では，人手で作成したパターンとのマッチングにより属性間の依存関係とオブジェクト間の順序関係を抽出する．

意見文中の比較評価を集約し，オブジェクトの順位付けを行う研究について述べる．Guo ら [1] はクラウドソーシングによってある集合内での最大値をとるオブジェクトを発見する手法を提案した．また，Guo らは新たに比較評価をすると推定精度が向上するような未評価のペアを選択する手法を提案した．Guo らの手法は人による評価はすべてペアワイズであるという前提のもとで考えられている．倉島ら [9] は従来の 3 つ組（対象，属性，評価）を 4 つ組（評価対象，比較対象，属性，評価）に拡張した．意見文中の比較表現に注目しパターンマッチングを行うことで評価対象と比較対象を抽出し，相関ルール分析を用いてオブジェクト間の優劣を推定しグラフを生成する．生成したグラフに対して PageRank に基づいたアプローチを行うことでオブジェクトのランク付けを行っている．

本研究は，オブジェクト間のペアワイズな関係を集約しランキングを行うという点で Guo ら，倉島らの研究に類似している．Guo ら，倉島らの研究では比較の主題となるクエリのみを扱っている．しかし，実データを扱うにあたって，比較表現に直接出現する属性はさほど多様でないという問題がある．本研究では，この問題を解決するアプローチとして文中の順接関係に着目し，比較の主題となりにくい属性についてもランキング規則を生成する．また，本研究では比較表現のみでなく，オブジェクト自身の持つ情報も考慮し，ランキング規則の生成を行うという点で前述の研究と異なる．

3. 概念説明

本節では，本研究の基盤となる概念について説明する．まず，オブジェクトのもつ属性の分類について述べる．次に属性間の依存関係，オブジェクト間の順序関係について定義を述べる．

3.1 探索的属性と経験的属性

Nelson[7] は，消費財を探索財 (search goods) と経験財 (experience goods) の 2 つに分類している．探索財とは，購入せずとも仕様等を確認することで品質の評価が可能な消費財である．また経験財とは，その商品を購入し，実際に扱って初めて品質を評価できるような消費財である．

本研究では，Nelson による消費財の分類に基づき，オブジェクトのもつ属性を**探索的属性**と**経験的属性**の 2 つに分類する．オブジェクトのもつ属性のうち，ユーザがオブジェクトを使用せずとも評価できるような属性のことを，そのオブジェクトの探索的属性という．また，オブジェクトのもつ属性のうち，ユーザがオブジェクトを実際に使用して初めて評価できるような属性のことを，そのオブジェクトの経験的属性という．

表 1: カメラのもつ探索的属性の例

	属性値	属性表現
重さ	0.2	軽い
大きさ	0.9	大きい

表 2: カメラのもつ経験的属性の例

	属性値	属性表現
携帯性	0.7	持ち運びやすい，持ち運びに便利
夜景:強さ	0.8	夜景に強い，夜景が綺麗に撮れる

これらの属性は属性値と属性表現をもつ．属性値とは，その属性の程度の大きさであり，具体的な数値で表現できる．本研究では，属性値は $[0, 1]$ 間の実数値をとるものとする．属性表現とは，人がその属性の程度を口語的に表現したものである．探索的属性の属性値を探索的属性値，属性表現を探索的属性表現と呼ぶものとする．同様に，経験的属性の属性値を経験的属性値，属性表現を経験的属性表現と呼ぶ．カメラのもつ探索的属性の例を表 1 に，経験的属性の例を表 2 に示す．

探索的属性は探索属性表現としてレビュー文中に記述されることも多い．また，探索的属性の一部はオブジェクトのスペック情報としてウェブ上に記述される．例えば，オブジェクトの重さや大きさ，機能の有無は EC サイトなどに記述されることも多い．スペック情報を参照することで，対応するオブジェクトの探索的属性値を求めることが可能である．

経験的属性は経験的属性表現としてレビュー文中に記述される．一方，経験的属性値がウェブ上で記述されることは少ない．探索的属性と異なり，経験的属性の属性値はスペック情報などにも記述されることが少ない．そのため，ユーザがオブジェクトのもつ経験的属性値について知るのは困難である．

3.2 属性間の依存関係

ある属性の属性値が変動すると，異なる属性の属性値がそれに対応し変動することがある．例えば，オブジェクトの探索的属性「画面の大きさ」が増加すると，そのオブジェクトの経験的属性「画面の見やすさ」は増加すると考えられる．また，オブジェクトの探索的属性「重さ」が増加すると，そのオブジェクトの経験的属性「持ち運びやすさ」は減少すると考えられる．本研究ではこのような属性間の関係を依存関係と呼ぶ．

属性間の依存関係を以下のように定義する．ある属性 a の属性値が増加すると異なる属性 a' の属性値もまた増加するとき，またはある属性 a の属性値が減少すると異なる属性 a' の属性値が減少するとき，a と a' の間に正の依存関係が存在するといい，

$$a \underset{+}{\rightarrow} a' \tag{1}$$

のように表記する．また，ある属性 a の属性値が増加すると異なる属性 a' の属性値が減少するとき，またはある属性 a の属性値が減少すると異なる属性 a' の属性値が増加するとき，a と a' の間に負の依存関係が存在するといい，

$$a \underset{-}{\rightarrow} a' \tag{2}$$

のように表記する．ただし，全属性集合 A について $a \in A$，$a' \in A$ である．

3.3 オブジェクト間の順序関係

複数のオブジェクトに対し，ある属性の属性値に関して大小関係をつけることが可能である．これをある属性のもとでのオブジェクト間の順序関係と呼ぶ．

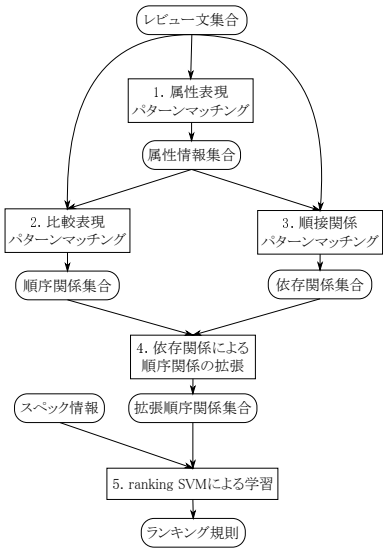

図 1: 手法の流れ

表 3: レビュー文の例

(1)	持ちやすく撮影に集中できます
(2)	ファインダーが見やすいのでピントが合わせやすいです
(3)	グリップが抜群に持ちやすいので、移動の際も手が疲れにくいですね
(4)	ファインダーは X3 に比べて見やすいと思う

オブジェクト間の順序関係を以下のように定義する．属性 a について，オブジェクト o がオブジェクト o' より高い属性値をもつとき，a の基で o と o' の間に順序関係が存在しているといい，

$$o \underset{a}{\succ} o' \tag{3}$$

のように表記する．ただし，全オブジェクト集合 O について，$o \in O$, $o' \in O$ である．また，全属性集合 A について，$a \in A$ である．

4. 経験的属性の推定

本節では我々が提案する経験的属性値の推定手法について詳細を述べる．本研究の目的は，オブジェクトの探索的属性から経験的属性値を推定し，経験的属性によるオブジェクト検索を実現することである．本手法は以下の 5 つのステップから構成される．

1. レビュー文中の属性表現からの属性情報抽出
2. 文の順接関係からの依存関係抽出
3. レビュー文中の比較表現からの順序関係抽出
4. 新たな順序関係の生成
5. 経験的属性値に基づくランキング規則の推定

手法の流れを図 1 に示す．図 1 の中の角丸はデータを，矩形は処理を表す．矢印は処理の入出力を表す．

本手法ではウェブ上のレビュー文から属性間の依存関係やオブジェクト間の順序関係を抽出する．レビュー文の例を表 3 に示す．

より正確な情報抽出を行うためにレビュー文に対し前処理を施す．まず，レビュー文を読点などの文章の区切りとなる箇所で分割し，それぞれの文に対して形態素解析と係り受け解析を行う．これにより，各文は閉路のない有向グラフで表現される．グラフの各ノードは文の各文節に対応する．ノードは複数の単語のシーケンスで構成される．また，各エッジは文節の係り受け関係に対応する．さらに有向グラフを複数の有向道に分解することで，1 文を有向道の集合で表現する．4.1，4.2，4.3 節では，この処理で得られた有向道と抽出パターンとのマッチングを行うことで，属性情報，属性間の依存関係，オブジェクト間の順序関係の抽出を行う．

本手法では，集めた順序関係を訓練データとして機械学習を行うことで，オブジェクトのもつ探索的属性から経験的属性の値を推定する．

表 4: 属性表現の抽出パターン

パターン	属性名	評価値
⟨[*x*/名詞][は | が | に | も/助詞]⟩?⟨[*y*/形容詞]⟩ ⟨[*y*/動詞 | 形容詞][*/名詞・接尾語][は | が | に | も/助詞]⟩⟨[*z*/形容詞]⟩	*x* : *y* *y*	高 *z*
⟨[*x*/名詞][は | が | に | も/助詞]⟩?⟨[*y*/名詞・形容動詞語幹][*/助動詞]⟩ ⟨[*y*/動詞 | 形容詞][*/名詞・接尾語][は | が | に | も/助詞]⟩⟨[*z*/名詞・形容動詞語幹][*/助動詞]⟩	*x* : *y* *y*	高 *z*
⟨[*x*/名詞][は | が | に | も/助詞]⟩?⟨[*y*/動詞][やすい/形容詞]⟩	*x* : *y*	高
⟨[*x*/名詞][は | が | に | も/助詞]⟩?⟨[*y*/名詞][できる/動詞]⟩	*x* : *y*	高
⟨[*x*/名詞][は | が | に | も/助詞]⟩?⟨[*y*/動詞][にくい/形容詞]⟩	*x* : *y*	低
⟨[*x*/名詞][は | が | に | も/助詞]⟩?⟨[*y*/動詞][づらい/形容詞]⟩	*x* : *y*	低

4.1 レビュー文中の属性表現からの属性情報抽出

レビュー文にはオブジェクトのもつ属性について言及した文，すなわち属性表現が多く記述される．また，属性表現からはその属性の属性値を推定することが可能である．

本手法では，レビュー文に対し，人手で用意した属性表現の抽出パターンとのマッチングを行うことで，レビュー文中の属性表現からオブジェクトのもつ属性の情報を抽出する．ここで抽出する属性の情報は属性名と評価値の 2 つ組である．評価値とは，その属性の属性値をおおまかに高低の 2 値に分類したもののことを言う．

属性表現の抽出パターンを表 4 に列挙する．レビュー文を前処理して得られた各有向道に対し，それぞれのパターンとのマッチングを行う．パターン中の変数 x, y, z に当てはまる単語は抽出され，パターンごとの設定に従って，属性名または評価値に格納される．ただし，抽出した語 z が評価値に対応している場合は，語の評価値を求めて高低の 2 値に変換し格納する．例えば,「重さがとんでもない」という文からは，属性名「重さ」，評価値「高」という情報を抽出する．これは，評価表現「とんでもない」の評価値が高く設定されているためである．また，パターンの末尾に否定の助動詞が接続されていた場合は，評価値を反転させる．

表 3 の (1) に対して，表 4 で示した属性表現のパターンを適用した場合，(持ちやすさ，高）と (撮影:集中しやすさ，高) が抽出される．また，表 3 の (2) に属性表現のパターンを適用した場合，(ファインダー:見やすさ，高) と (ピント:合わせやすさ，高) が抽出される．

5.1 節で述べるユーザレビューに対し表 4 で列挙した属性表現抽出のパターンを適用し，属性情報の抽出を行った．表 4 で示したパターンを適用することで，203,624 件の属性情報が抽出された．出現回数の多かった属性情報の一部を表 5 に示す．

では，得られた属性情報を利用して，文の順接関係から属性間の依存関係を抽出する．では，得られた属性情報を観点にもつ比較表現からオブジェクト間の順序関係を抽出する．

4.2 文の順接関係からの依存関係抽出

本手法では，接続助詞により構成される順接関係に注目することで属性間の依存関係の生成を行う．接続詞や接続助詞により文と文の間には論理関係が構築される．特に，文と文の間に順接関係が存在する場合は,「前件が後件の順当な原因・理由になっている」

表 5: 出現数の多い属性の例

属性名	評価値	出現数	属性名	評価値	出現数
大きさ	高	6,490	高さ	高	2,291
小ささ	高	4,101	暗さ	高	2,236
重さ	高	3,634	綺麗さ	高	2,179
軽さ	高	3,474	使いやすさ	高	2,167
明るさ	高	2,697	十分さ	高	2,108

表 6: 順接関係の抽出パターン

⟨(e_a/属性表現)[ので/助詞]⟩⟨($e_{a'}$/属性表現)⟩
⟨(e_a/属性表現)[から/助詞]⟩⟨($e_{a'}$/属性表現)⟩

[3]と考えられる．そのため，属性表現を含む文節と属性表現を含む文節の間に順接関係が存在する場合は，それらが表す属性の間に依存関係が構築されていると考えられる．

属性表現を抽出した有向道に対して順接関係の抽出を行い，属性の依存関係を取得する．表 6 に順接関係の抽出パターンを示す．

パターン中の属性表現 e_a と $e_{a'}$ にそれぞれ対応する属性 a，a' の間の依存関係を取得する．a と a' の評価値が一致していた場合は正の依存関係 $a \xrightarrow[+]{} a'$ が成立していると考える．また，a と a' の評価値が異なっていた場合は負の依存関係 $a \xrightarrow[-]{} a'$ が成立していると考える．例えば，表 3 の (2) にパターンを適用した場合，正の依存関係「ファインダー:見やすさ」$\xrightarrow[+]{}$「ピント:合わせやすさ」が抽出される．同様に，表 3 の (3) にパターンを適用した場合，負の依存関係「グリップ:持ちやすさ」$\xrightarrow[-]{}$「手:疲れやすさ」が抽出される．

抽出の結果，同一の属性の間で正の依存関係と負の依存関係が同時に抽出されることがある．正負の依存関係のうち，一方は必ず誤った順序関係を生成する．誤った順序関係はランキング学習の精度を低下させるため，正しく矛盾を解消する必要がある．本手法では，正負の関係のうち多く出現する方を信頼できるものとみなし，出現数の少ない方をすべて消去することで矛盾の解消を図る．ただし，出現数の少ない方の数だけ多い方の関係も消去する．

5.1 節で述べるユーザレビューのそれぞれの文のうち，属性表現が複数記述されていた文に対し，表 6 で列挙した順接関係のパターンを適用し，属性間の依存関係の抽出を行った．出現回数の多かった依存関係の一部を表 7 に記す．

抽出された依存関係は 4.4 節で順序関係の拡張に用いる．

4.3 レビュー文中の比較表現からの順序関係抽出

本手法では，レビュー文中の比較表現に注目することでオブジェクト間の順序関係の抽出を行う．

ユーザは使用したオブジェクトと他のオブジェクトとの比較を行い，その関係をレビュー文に記述する．ユーザの記述した比較表現からは，ある属性に関するオブジェクトの優劣，すなわち，ある属性に関するオブジェクトの順序関係を抽出できる．また，一般にレビュアーは自らの経験に基づきレビューを行うと考えられるため，比較表現に注目することで経験的属性に関する順序関係を抽出できる．経験的属性に関する順序関係を集約して得られる知見は，経験的属性値の推定に有用であると考えられる．

Jindal ら [2] は比較表現を以下の 4 つに分類している．

- **Non-Equal Gradable**: ある属性に関して，複数のオブジェクト間の優劣関係を表現したもの．

[3]順接 (ジュンセツ) とは - コトバンク, https://kotobank.jp/word/順接-530004

表 7: 出現数の多い依存関係

依存関係			出現数	依存関係			出現数
確認しやすさ	$\xrightarrow[+]{}$	便利さ	5	軽さ	$\xrightarrow[+]{}$	携帯性:良さ	4
操作しやすさ	$\xrightarrow[+]{}$	便利さ	5	撮影しやすさ	$\xrightarrow[+]{}$	便利さ	4
大きさ	$\xrightarrow[+]{}$	持ちやすさ	5	小ささ	$\xrightarrow[+]{}$	持ち運び:楽さ	4
小ささ	$\xrightarrow[+]{}$	仕方:なさ	4				

表 8: 比較表現のパターン

主語	⟨[s/名詞][の/助詞]⟩⟨[方/名詞][が/助詞]⟩⟨(e_a/属性表現)⟩
基準	⟨[c/名詞][より/助詞]⟩⟨(e_a/属性表現)⟩
	⟨[c/名詞][と	に/助詞]⟩⟨[比べる/動詞]⟩⟨(e_a/属性表現)⟩
	⟨[c/名詞][と/助詞]⟩⟨[比較/名詞][する/動詞]⟩⟨(e_a/属性表現)⟩

- **Equative**: ある属性に関して，2 つのオブジェクトの等位関係を表現したもの．
- **Superlative**: ある属性に関して，あるオブジェクトの最上位性を表現したもの．
- **Non-Gradable**: ある属性に関して，複数のオブジェクトを比較したものだが，優劣関係については言及していないもの．

本手法では，Non-Equal Gradable な比較表現のみを扱う．ただし，Non-Equal Gradable な比較表現の中でも，具体的な 2 つのオブジェクトを比較しているものに限る．

4.2 節と同様に，属性表現を抽出した有向道に対してのパターンマッチングにより比較表現を発見する．比較表現の抽出パターンを表 8 に示す．

比較表現のパターンは主語抽出パターンと基準抽出パターンからなる．主語抽出パターン中の s に対応した語は，その比較文の主語とみなされる．また，基準抽出パターン中の c に対応した語は，その比較文の比較対象とみなされる．属性 a の評価値が高かった場合は順序関係 $s \underset{a}{\succ} c$ を，a の評価値が低かった場合は順序関係 $c \underset{a}{\succ} s$ を抽出する．

上記の手法により，レビュー文から順序関係を抽出するが，レビュー文では比較文の主語や基準が省略されることも多い．そうしたレビュー文に対しては，省略された比較の対象を推測し補完する必要がある．比較文の主語が省略されている場合，そのレビューの対象であるオブジェクトで主語を補完する．これの補完は，比較表現の主語はレビュー対象のオブジェクトであることが多いという推測に基づく．なぜなら，レビューの主題はレビュー対象のオブジェクトに関する評価であるためである．また，比較文の基準が省略されていて主語がレビュー対象のオブジェクトと異なる場合，レビュー対象のオブジェクトで基準を補完する．これは，主語の補完と同様の考えに基づく．

比較文の基準が省略されている場合でも，レビュアーは暗黙的にオブジェクトを他の何かと比較し評価しているものと考えられる．暗黙的な比較基準としては，同価格帯のオブジェクト，ユーザが過去に使っていた同ドメインのオブジェクト，ユーザがオブジェクトに対し抱いていたイメージなどが考えられる．しかし，暗黙的な比較基準を正確に推定するのは困難である．本研究では，上述の補完を行うにとどまり，暗黙的な比較対象の推定については主として取り扱わないものとする．

また，本研究では，具体的な 2 つのオブジェクトを対象とする比較表現のみを取り扱う．パターン中の s または c には，具体的なオブジェクト以外を表す語が対応することもある．例えば，「A 社のカメラは B 社のカメラより使いやすい」という文にパターンを適用した場合，s = A 社のカメラ，c = B 社のカメラ が抽出される．このとき，s,c はともにオブジェクトのメーカに関する部分集合を指している．部分集合を指す語については，メーカの他に，「デジタル一眼カメラより」のようなカテゴリの細分類に関す

表 9: 抽出された順序関係の例

順序関係			レビュー文
OLYMPUS SZ-31MR	$\succ_{\text{軽さ}}$	OLYMPUS STYLUS SH-50	SH-50 よりも軽いし、決して携帯性は悪くないのだが
LUMIX DMC-GX7 ボディ	$\succ_{\text{モノ:撮りやすさ}}$	LUMIX DMC-GX1 ボディ	飛びモノは GX1 より確実に撮りやすくなりました

るもの,「もっと大きい方が」のようなオブジェクトの属性値に関するもの,「昔使っていたカメラより」のようなユーザの過去に所有していたオブジェクトに関するものなどが挙げられる．このような集合を対象とする比較表現は本研究では取り扱わず今後の課題とする．

5 節で述べるレビュー文に対し，表 8 で列挙した比較表現のパターンを適用し，オブジェクト間の順序関係の抽出を行った．比較表現のパターンを用いると比較対象を表現した文字列が得られる．これを便宜上，比較対象文字列と呼ぶ．順序関係を構築するには，比較対象文字列からオブジェクトの実体を特定する必要がある．実体の特定を誤った場合，得られる順序関係が不正なものとなり，ランキング規則学習の精度に影響をあたえる．本実験では，精度を高水準に保つため，得られた比較対象文字列が特定の具体的なオブジェクトを表していると判断される場合に限り，実体の特定を行った．比較対象文字列に含まれる数値とオブジェクト名に含まれる数値が一致するオブジェクトのうち，最も比較対象文字列と類似しているオブジェクト名をもつオブジェクトを比較対象とした．

パターンマッチングによって 15,457 件の比較表現が得られた．その比較表現のうち比較対象の実体の特定を行って得られた 3,542 件の順序関係の一部を表 9 に示す．

抽出した順序関係は，4.4 節で述べる順序関係の拡張のためのデータとして用いられるほか，4.5 節で述べるランキング規則学習の訓練データとして用いられる．

4.4 新たな順序関係の生成

4.3 節でレビュー文からオブジェクト間の順序関係を抽出する手法について説明した．得られた順序関係を訓練データとして用いて，4.5 節で述べる Ranking SVM による学習を行うことでランキング規則を生成することができる．しかし，レビュー文に記述される各属性あたりの順序関係の数は多いとは言いがたい．特に，レビュー文における経験的属性表現の出現数は探索的属性表現に比べて少ないため，ランキング規則学習の訓練データとして用いるには数が不十分であることが多いと考えられる．そのため，ランキング規則の学習を行うにあたって訓練データの数を増やす必要がある．

本手法では，属性間の依存関係を用いて，対応する属性に関するオブジェクト間の順序関係の拡張を行い，新たな訓練データの獲得を試みる．以下の仮説に基づき，4.2 節で抽出した属性間の依存関係と 4.3 節で抽出したオブジェクト間の順序関係から新たな順序関係を生成する．

> 正の依存関係 (1) と順序関係 (3) がともに成立しているとき，順序関係 (4) が成立する．
>
> $$o \underset{a'}{\succ} o' \tag{4}$$
>
> 負の依存関係 (2) と順序関係 (3) がともに成立しているとき，順序関係 (5) が成立する．
>
> $$o \underset{a'}{\prec} o' \tag{5}$$

例えば，表 3 の (2), (4) からはそれぞれ，依存関係 ファインダー:見やすさ $\rightarrow$ ピント:合わせやすさ と順序関係 $o \underset{\text{ファインダー:見やすさ}}{\succ}$ X3 が抽出される．ここで，c は (4) のレビュー対象であるオブジェクトである．このとき，上の仮説を適用すると，新たに経験的属性「ピント:合わせやすさ」に関する順序関係 $o \underset{\text{ピント:合わせやすさ}}{\succ}$ X3 が得られる．このように，レビュー文中に「ピント:合わせやすさ」に関する o と X3 の順序関係が出現していなくとも，依存関係に基づく順序関係の拡張を行うことで,「ピント:合わせやすさ」に関する順序関係が増え，訓練データを獲得することができる．

4.2 節で述べた属性間の依存関係を用いて，オブジェクト間の順序関係の拡張を行った．拡張により，1,533 属性に関する 3,542 件の順序関係が 1,982 属性に関する 40,930 件に増加した．本実験では，依存関係 $a \rightarrow a'$1 つにつき，a に関する順序関係集合に含まれるすべての順序関係 $o \underset{a}{\succ} o'$ に対し，$o \underset{a'}{\succ} o'$ を生成した．つまり，複数回出現する因果関係は信頼度が高いとみなされ，1 回のみ出現した因果関係より多くの順序関係を生成する．

4.5 経験的属性値に基づくランキング規則の推定

4.3 節，4.4 節でオブジェクト間の順序関係の収集を行った．

順序関係を訓練データとした機械学習により，オブジェクトの探索的属性から経験的属性値に基づくランキング規則を導出する．

各オブジェクトを探索的属性のベクトルで表現する．全オブジェクト集合を $O = \{o_1, o_2, \ldots, o_m\}$, オブジェクトのもつ探索的属性の集合を $A_S = \{a_s^{(1)}, a_s^{(2)}, \ldots, a_s^{(n_s)}\}$ とすると，オブジェクト o_i に関する探索的属性のベクトル $\boldsymbol{o}_i$ は以下のように表現される．

$$\boldsymbol{o}_i = (a_{s,i}^{(1)}, a_{s,i}^{(2)}, \ldots, a_{s,i}^{(n_s)})^{\mathrm{T}} \in \mathbb{R}^{n_s}$$

ここで，$a_{s,i}^{(j)}$ はオブジェクト o_i の探索的属性 $a_s^{(j)}$ の属性値であり，$0 \leq a_{s,i}^{(j)} \leq 1$ を満たす実数である．ただし，$i = 1, 2, \ldots, m$, $j = 1, 2, \ldots, n_s$ である．

探索的属性の一部はスペック情報に記載される．記載されたスペック値を $[0, 1]$ の実数値に正規化したものを探索的属性値とし，ベクトルの各要素とする．

オブジェクトのもつ経験的属性の集合を $A_E = \{a_e^{(1)}, a_e^{(2)}, \ldots, a_e^{(n_e)}\}$ と表記する．経験的属性 $a_e^{(k)}$ の属性値を，$a_e^{(k)}$ に関する順序関係集合 $T^{(k)} = \{t_1^{(k)}, t_1^{(k)}, \ldots, t_l^{(k)}\}$ を訓練データとした機械学習により $a_e^{(k)}$ の属性値関数を学習する．ここで，$k = 1, 2, \ldots, n_e$ である．各訓練データは，$t_k^{(z)} = \boldsymbol{o}_{i_z} \underset{a_e^{(k)}}{\succ} \boldsymbol{o}_{i'_z}$ のように表される．ここで，$z = 1, 2, \ldots, l$ である．また，$i_z, i'_z = 1, 2, \ldots, m$, $i_z \neq i'_z$ であり，$\boldsymbol{o}_{i_z}$, $\boldsymbol{o}_{i'_z}$ がそれぞれ表すオブジェクト o_{i_z}, $o_{i'_z}$ に対し，$o_{i_z}, o_{i'_z} \in O$ である．

学習には Ranking SVM[3] を用いる．Ranking SVM は，2 クラス分類の学習器である SVM をランキング問題に応用した機械学習法である．Ranking SVM を用いて学習を行うことで，最終的に以下の関数を得ることができる．

$$\mathrm{f}_{a_e^{(k)}} : \mathbb{R}^{n_s} \rightarrow \mathbb{R} \tag{6}$$

式 (6) はオブジェクト o_i の探索的属性ベクトル $\boldsymbol{o}_i$ を入力として受け取り，そのオブジェクトの属性 $a_e^{(k)}$ の属性値を推定する関数である．この関数を用いることで，オブジェクトの探索的属性値から経験的属性 $a_e^{(k)}$ に基づいたランキングを行うことが可能になる．

5. 実験

5.1 概要

オブジェクトのスペック情報とユーザレビューは価格.com から収集を行った．文の形態素解析には，形態素解析エンジン MeCab[4] を

[4] https://code.google.com/p/mecab/

利用した．また，係り受け解析には，日本語係り受け解析器 CaboCha[5]を利用した．

価格.com から抽出したデータのうち，カメラカテゴリに属するオブジェクト 619 件をオブジェクトのデータとして用いた[6][7]．また，上記 619 件のオブジェクトを対象としたレビュー 17,061 件をユーザレビューとして用いた．

ランキング規則の学習および規則の適用には SVM^{rank}[8]を利用した [4]．SVM のカーネルとしては RBF カーネルを用いた．また，その他のパラメタはデフォルト値を採用した．Ranking SVM での学習を行うにあたり，オブジェクトを探索的属性のベクトルで表現する必要がある．本実験では，スペック情報から抽出した以下の値を素性として用いて，オブジェクトを 49 次元のベクトルで表現した．

- サブカテゴリ（デジタル一眼レフカメラかコンパクトデジタルカメラか）
- メーカー
- 重さ
- 画素数
- 幅，高さ，厚さ
- 体積
- シャッタースピード (最遅，最速)
- フレームレート
- 連写性能
- 撮影感度
- モニター (インチ数，解像度)
- 手ブレ補正機能 (有無，何式か)

数値で表現されている値は最大値で割ることで [0, 1] の範囲に正規化を行った．欠損値は学習に大きな影響を与えないよう，全オブジェクトの属性値の平均で補った．サブカテゴリやメーカーなど，複数の値から 1 つを選ぶカテゴリカルなデータについては，対応する値ごとに次元を設け，その項目がその値を持つときその次元に 1，それ以外の次元に 0 を格納するものとした．今回のデータセットでは欠損値は存在しなかったが，存在した場合は最頻値の対応次元に 1 を格納する．

5.2 ベースライン

提案手法と比較するベースラインについて述べる．ベースラインには，属性の言及数によるランキングを採用した．例えば,「使いやすさ」に関するランキングであれば，レビュー中で「使いやすい」とより多く言及されているオブジェクトを上位にランキングする．これは，一般的なユーザが商品を比較する際に用いるランキング手法であると考えられる．

5.3 評価方法

依存関係による拡張を施す以前に順序関係が多く得られた属性のうち，経験的属性であると人手で判断したものを評価クエリとして採用した．評価に用いるクエリとその属性に関する順序関係数，および順序関係拡張後の順序関係数を表 10 に示す．表 10 の「順序関係数」は，レビュー文から抽出されたそのクエリに関する順序関係の数を表す.「拡張順序関係数」は，依存関係により拡張されたそのクエリに関する順序関係の数を表す.

本評価手法ではレビュー中に記述されたオブジェクトの順序関係を正解とみなすこととする．正解とされたオブジェクトの順序関係の 1 つを訓練データセットから除外し学習を行い，生成されたランキング規則が除外した順序関係を正しく導出できるかを確かめる．すべての正解データに対して上記の処理を実行し，正しく導出できた割合を精度とする．

表 10: 評価に用いるクエリとその数

クエリ	順序関係数	拡張順序関係数
使いやすさ	60	372
持ちやすさ	54	1,551
見やすさ	43	393
画質:良さ	28	170
撮影しやすさ	26	580
ホールドしやすさ	18	266
操作しやすさ	12	19
扱いやすさ	8	75
携帯性:良さ	8	1,004

表 11: 評価結果

クエリ	ベースライン	提案手法 (拡張なし)	提案手法 (拡張あり)
使いやすさ	0.53	0.70	0.43
持ちやすさ	0.50	0.56	0.59
見やすさ	0.56	0.67	0.70
画質:良さ	0.46	0.68	0.68
撮影しやすさ	0.50	0.54	0.69
ホールドしやすさ	0.56	0.67	0.67
操作しやすさ	0.50	0.67	0.33
扱いやすさ	0.50	0.63	0.75
携帯性:良さ	0.25	0.63	0.38
マクロ平均	0.48	0.64	0.58
マイクロ平均	0.51	0.64	0.58

評価実験は，それぞれのクエリに対し，依存関係による順序関係の拡張を行うか行わないかの 2 通りの場合について実施する．Ranking SVM のカーネルには RBF カーネルを用いた．パターン抽出で得られた拡張以前の順序関係集合から 1 つを取り出し正解データとし，残りの順序関係を用いてランキング規則の学習を行う．取り出した関係を構成するオブジェクト対に対し規則を適用することで属性値を推定し，正解データと順序が一致しているかを確認し正答率を求める．この処理を順序関係集合の各要素についてそれぞれ行う．

5.4 結果

5.1, 5.2 節で述べた実験の結果とそれに対する考察を述べる．

実験結果を表 11 に示す．表 11 の「ベースライン」はベースラインの精度を表す.「提案手法 (拡張なし)」は順序関係の拡張をせずに学習を行った場合の精度を表す.「提案手法 (拡張あり)」は拡張した順序関係によって学習を行った場合の精度を表す.

実験の結果，マクロ平均，マイクロ平均ともに提案手法がベースラインを上回る精度を確保した．特に，順序関係の拡張を行わない手法を適用した場合は，すべてのクエリにおいてベースラインを上回る精度が得られたほか，マクロ平均とマイクロ平均においても最も高い精度が得られた．順序関係の拡張を行った場合は，4 つのクエリにおいて，順序関係の拡張を行わない場合と比べて精度の向上が確認されたが，3 つのクエリにおいては，精度の低下が確認された．

レビュー中に記述される経験的属性の数は探索的属性に比べて少ない．表 5 に出現数の多かった属性表現の上位 10 件を列挙したが，経験的属性とみなせるものは「綺麗さ」,「使いやすさ」の 2 件であり，最も出現数の多かった「綺麗さ」も全体では 8 位である．経験的属性表現の中でも比較表現に結びついているものは更に少ない．表 10 に記載したとおり，経験的属性に基づく順序

[5]https://code.google.com/p/cabocha/
[6]http://kakaku.com/camera/digital-slr-camera/itemlist.aspx
[7]http://kakaku.com/camera/digital-camera/itemlist.aspx
[8]http://www.cs.cornell.edu/people/tj/svm_light/svm_rank.html

関係のうち最も多い「使いやすさ」に関する順序関係でも，わずか 60 件しか存在しなかった．順序関係のみに基づいてランキング規則の学習を行った場合は，学習データの数が不足し，得られる規則が汎用性を失う可能性が高い．そのため，順序関係の拡張は必要であると考えられる．

実験の結果，依存関係による順序関係の拡張を行った場合，精度の低下がたびたび確認された．順序関係の拡張の行った場合の精度の低下は，依存関係に混入した不正確なデータが誤った順序関係を生成したため発生したと考えられる．本手法では，比較表現から抽出された順序関係と依存関係によって拡張された順序関係をすべて同じ重要度で扱っている．依存関係に信頼度を導入し，より信頼できる依存関係から生成された順序関係の重要度をより高くすることで，精度の向上が実現できると考えられる．

本評価手法では，ユーザによってレビューに記述されたオブジェクト間の順序関係を正解としている．しかし前述のとおり，経験的属性に基づく順序関係は少ないため正解数が十分ではない問題がある．より正確な評価のため，レビューの記述から正解を定めるのではなく，専門家による評価を実施し正解を用意する必要があると考えている．また，他のドメインに対する提案手法の有効性を検証するため，カメラに関するドメインだけではなく他のドメインに関する実験も今後行う必要がある．

6. おわりに

本研究では，オブジェクトの探索的属性から経験的属性に基づくランキング規則を推定する手法を提案し，経験的属性による検索を可能にした．提案手法では，あらかじめ用意した抽出パターンとのマッチングによって，レビュー文から属性情報，属性間の依存関係，オブジェクト間の順序関係の抽出を行った．次に，収集した依存関係と順序関係から順序関係集合を拡張することで，ランキング規則学習の新たな訓練データを獲得した．さらに，得られた順序関係を訓練データとして用いて Ranking SVM による学習を行い，経験的属性に基づくランキング規則を推定した．

また本稿では，EC サイト価格.com から収集したカメラに関するスペック情報，レビューデータを用いて評価実験を行った．評価実験の結果，順序関係を拡張する必要性を確認した．依存関係による順序関係拡張が必ずしも有効に作用するとは限らないことを確認した上で，依存関係による拡張手法の改善案について考察した．

本手法の課題としては，属性の粒度の問題が挙げられる．本実験では,「使いやすさ」や「持ちやすさ」など粗い粒度の経験的属性のみを取り扱った．しかし，一般により細かい粒度の経験的属性による検索を行いたいという要求は存在する．例えば,「海での使いやすさ」や「片手での持ちやすさ」は上記の属性をより細かく表現したものである．扱う属性を粒度を荒くすると，属性表現の情報量を失ってしまう恐れがある．一方で，属性の粒度が細かくなると，属性表現が細分化されるため，それぞれの属性表現の出現数が減少し，属性ごとに得られる順序関係がさらに減少するという問題がある．今後は，対象とするユーザレビューの拡充や専門家によるオブジェクト評価によって，より多くのテストデータ，正解データの収集に取り組み，より多くのデータを対象に本手法の有用性を検証する予定である．

[謝辞]

本研究の一部は，文科省科研費基盤 (A)「ウエブ検索の意図検出と多元的検索意図指標にもとづく検索方式の研究」(24240013, 研究代表者：田中克己）によるものです．ここに記して謝意を表します．

[文献]

[1] Stephen Guo, Aditya Parameswaran, and Hector Garcia-Molina. So who won?: dynamic max discovery with the crowd. In *Proceedings of the 2012 ACM SIGMOD International Conference on Management of Data*, pp. 385–396. ACM, 2012.

[2] Nitin Jindal and Bing Liu. Identifying comparative sentences in text documents. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 244–251. ACM, 2006.

[3] Thorsten Joachims. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 133–142. ACM, 2002.

[4] Thorsten Joachims. Training linear svms in linear time. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 217–226. ACM, 2006.

[5] Shasha Li, Chin-Yew Lin, Young-In Song, and Zhoujun Li. Comparable entity mining from comparative questions. *Knowledge and Data Engineering, IEEE Transactions on*, Vol. 25, No. 7, pp. 1498–1509, 2013.

[6] Bing Liu, Minqing Hu, and Junsheng Cheng. Opinion observer: analyzing and comparing opinions on the web. In *Proceedings of the 14th international conference on World Wide Web*, pp. 342–351. ACM, 2005.

[7] Phillip Nelson. Information and consumer behavior. *The Journal of Political Economy*, pp. 311–329, 1970.

[8] 佐藤敏紀, 奥村学. blog からの比較関係抽出. 情報処理学会自然言語処理研究会, pp. 7–14, 2007.

[9] 倉島健, 別所克人, 内山俊郎, 片岡良治. 比較評価情報の抽出とそれに基づくランキング手法の提案. 第 18 回データ工学ワークショップ (DEWS 2007), 2007.

[10] 立石健二, 石黒義英, 福島俊一. インターネットからの評判情報検索. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション, Vol. 101, No. 189, pp. 75–82, 2001.

[11] 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. 情報学基礎研究会報告, Vol. 2004, No. 93, pp. 1–8, 2004.

内田 臣了　Shinryo UCHIDA
京都大学大学院情報学研究科社会情報学専攻修士課程在学中．おもにテキストマイニングやオブジェクト検索の研究に従事．日本データベース学会学生会員．

山本 岳洋　Takehiro YAMAMOTO
京都大学情報学研究科社会情報学専攻助教．2011 年京都大学大学院情報学研究科博士課程修了．博士（情報学）．おもに情報検索におけるユーザインタラクションやユーザ理解に関する研究に従事．情報処理学会，日本データベース学会，ACM 各会員．

加藤 誠　Makoto P. KATO
京都大学大学院情報学研究科社会情報学専攻特定助教．2012 年京都大学大学院情報学研究科博士後期課程修了．博士（情報学）．情報検索，おもに対話的情報検索の研究に従事．情報処理学会，日本データベース学会，人工知能学会，ACM 各会員．

大島 裕明　Hiroaki OHSHIMA
京都大学大学院情報学研究科社会情報学専攻特定准教授．2007 年京都大学大学院情報学研究科博士後期課程修了．博士（情報学）．主に情報検索，ウェブマイニング，デザイン学の研究に従事．電子情報通信学会，日本データベース学会，ACM 各会員．

田中 克己　Katsumi TANAKA
京都大学大学院情報学研究科社会情報学専攻教授．1976 年京都大学大学院博士前期課程修了．博士（工学）．主にデータベース，マルチメディアコンテンツ処理，ウェブ検索の研究に従事．IEEE Computer Society，ACM，人工知能学会，日本ソフトウェア科学会，情報処理学会，日本データベース学会各会員．