

目的指向商品推薦システムのための商品属性値の自動決定手法

安藤 哲志† 伊藤 孝行†§ 高橋 侑也† 見並 史彬† 小林 幹門‡
 †名古屋工業大学 ‡北陸先端科学技術大学院大学 §MIT スローン経営大学院

1 はじめに

近年、Web 上ではショッピングサイトが増加しており、ユーザーの目的や状況によって商品を推薦する目的指向商品推薦システムが試作されている [1]。目的指向商品推薦システムでは商品データベースに含まれる商品の属性値によって推薦を行っている。属性値は商品の特徴を数値で表したものであり、現在は属性値の決定を人の手で行っている。人の手による決定では商品の属性値を客観的に決定することは難しく、商品の数が増えるにつれ商品データベースの作成が困難になる。そこで本稿では商品の属性値を自動的に決定し、商品データベースを構築する手法を提案し、実装する。商品データベースの作成には単語の印象を数値で表した印象辞書を用いる。また実装したデータベースを用いた商品の推薦方法を提案し、被験者実験の結果を示す。

2 関連研究

2.1 目的指向商品推薦システム

目的指向商品推薦システムとは商品を使用する目的や状況によって推薦するシステムである。本稿では取り扱う商品を衣服とする。目的指向商品推薦システムではユーザーにキーワードではなく「中学の同窓会に着ていく」など服を着ていく目的や状況を入力させ商品を推薦する。ユーザーは商品の適切なキーワードやカテゴリーを知らなくても欲しい商品を探し出す事が可能となるため、検索時のユーザーの負担が軽くなる。目的指向商品推薦システムでは、「同窓会」などイベントとなる単語を含むイベント単語データベースと商品データベースにより構成されている。2つのデータベースには属性値も含まれている。商品の推薦はユーザーからの入力に含まれるイベント単語の属性値と商品の属性値のマッチングにより行われる。

2.2 印象辞書

印象辞書 [2] とは単語の対となる印象を [0-1] の実数で表したものであり、多くの文書を用いる事で自動的に

印象	印象語群
暑い	暑い, 蒸暑い, 蒸し暑い, お暑い, 猛暑, 夏, 真夏, あつい, 厚手, 厚み
寒い	寒い, 肌寒い, お寒い, 寒冷, 冬, 真冬, 厳冬, さむい, 薄手, 薄い
フォーマル	フォーマル, 上品, 正式, ブラックフォーマル, 上品さ
カジュアル	カジュアル, 非公式, 普段, 普段着, 日常, 日常的, カジュアルウェア, カジュアル感
装飾性	装飾, 装飾品, 飾り, 飾られる, ファッション, 飾る, 飾
運動性	運動, 動く, 動きやすい, 動ける

表 1: 使用した印象語群

作成可能である。印象辞書は単語の心的印象は対比的な 2 つの印象語群との共起の仕方によって決まるという考えのもと、提案されている。本稿では印象に [暑い ⇔ 寒い], [フォーマル ⇔ カジュアル] および [装飾性 ⇔ 運動性] を用いた。印象語群は印象の代表となる単語群である。本稿で使用した印象語群を表 1 に示す。[2] では元となる文書に日経新聞全文記事データベースを用いた年ごとの計算を行って作成している。本稿では印象辞書の元となる文書に amazon.co.jp* から取得した商品のレビューと Wikipedia† を使用し、年ごとの計算を行えなかったため、年ごとの計算を省いた式を用いた。形態素解析は Mecab‡ を使用している。作成した印象辞書の作成方法を示す。印象語群 e_1 に含まれる印象語のいずれかを含む文書の数を $df(e_1)$ とする。印象語群 e_1 に含まれる印象語と単語 w の両方を含む文書の数を $df(e_1 \& w)$ とする。印象語群 e_1 のいずれかが現れたときに単語 w も現れる確率 $P(w \& e_1)$ は式 (1) で求める。

$$P(w \& e_1) = \frac{df(e_1 \& w)}{df(e_1)} \quad (1)$$

次に、対となる印象 [$e_1 \Leftrightarrow e_2$] を構成する印象語群 e_1 , e_2 に対し、単語 w の印象語群 e_1 に対する出現確率 $P(e_1 \& w)$ と印象語群 e_2 に対する出現確率 $P(e_2 \& w)$ の内分比 $S(e_1, e_2, w)$ を式 (2) で求める。

$$S(e_1, e_2, w) = \frac{P(w \& e_1)}{P(w \& e_1) + P(w \& e_2)} \quad (2)$$

$S(e_1, e_2, w)$ は単語 w の印象 [$e_1 \Leftrightarrow e_2$] に対する印象値として扱う。単語には印象と関係の深い単語とそうでない単語がある。そこで単語 w の対となる印象 [$e_1 \Leftrightarrow e_2$]

†Satoshi ANDO †Takayuki ITO †§Yuya TAKAHASHI
 †Humiaki MINAMI †Mikito KOBAYASHI
 †Nagoya Institute of Technology
 †Japan Advanced Institute of Science and Technology
 §MIT Sloan School of Management

*<http://www.amazon.co.jp>

†<http://ja.wikipedia.org>

‡<http://mecab.sourceforge.net/>

に対する重み $M(e_1, e_2, w)$ を式 (3) のように定義する.

$$M(e_1, e_2, w) = \log(df(e_1 \& w) + df(e_2 \& w))$$

$S(e_1, e_2, w)$ と $M(e_1, e_2, w)$ を対となる印象すべてについて求めたものが印象辞書の要素である.

3 商品属性値の自動決定手法

本章では印象辞書を用いた商品データベースの作成方法および作成したデータベースを使用した商品の推薦方法を提案する. 本稿で作成する商品データベースは amazon.co.jp の API から取得したデータに属性値を追加したものを使用している.

商品データベース中の商品の属性値は商品レビューの印象値を用いる. 文書の対となる印象 $[e_1 \leftrightarrow e_2]$ に対する印象値 $O(e_1, e_2, text)$ は単語の印象値 $S(e_1, e_2, w)$ と重み $M(e_1, e_2, w)$ を使い式 (3) で求める. ただし, 式中では $O(e_1, e_2, text)$ を O , $S(e_1, e_2, w)$ を S および $M(e_1, e_2, w)$ を M と表記する. また $text$ は文書を形態素解析で分解した単語 w の集合である.

$$O = \frac{\sum^{text} (S * |2S - 1| * M)}{\sum^{text} (|2S - 1| * M)} \quad (3)$$

$|2S - 1|$ は印象語群との関係が乏しい一般的な単語が O の平均に及ぼす悪影響を削減するために導入している.

商品の推薦は以下の方法を用いる. 1. ユーザーからの入力文を形態素解析し, 単語毎に分割する. 2. 分割した単語群から入力文の各印象値を求める. 3. 求めた各印象値に属性値が近似する商品を推薦する.

2. の入力文の印象値は式 (3) を使用し求める. 3. の商品の推薦では式 (4) の値が最も少ない商品から順に推薦を行う. 式中の $O_{item}(e_1, e_2)$ は商品の印象 $[e_1 \leftrightarrow e_2]$ の属性値, $O_{input}(e_1, e_2)$ は入力文の印象 $[e_1 \leftrightarrow e_2]$ の印象値, $M_{sum}(e_1, e_2)$ は入力文の印象 $[e_1 \leftrightarrow e_2]$ の重みの合計である. E は対となる印象の集合である.

$$\sum_{e_1, e_2 \in E} \{|O_{item}(e_1, e_2) - O_{input}(e_1, e_2)| * M_{sum}(e_1, e_2)\} \quad (4)$$

式 (4) は入力文の印象値と商品の属性値の差に重みの合計をかけた値である. 入力文の中で重要となる印象は重みが大きいと考え, 重み $M_{sum}(e_1, e_2)$ を導入している.

4 実験結果と考察

本章では商品レビューを元に作成した印象辞書と Wikipedia から作成した印象辞書の比較を行い, 行った被験者実験の結果を示す.

商品レビューおよび Wikipedia の文書数および含まれる単語数を表 2 に示す. 商品レビューと比較し, Wikipedia

元となる文書	文書数	単語の総数
Wikipedia	379335	2223120
商品レビュー	22654	23438

表 2: 元文書の数の比較

の方が文書数では約 16 倍, 単語の総数では約 95 倍多い結果となった.

被験者実験は男性 44 名, 女性 8 名の合計 52 名に対して行った. 被験者には入力文に対し推薦された商品のうち「正しいと思われる商品」がいくつあるかを回答してもらった. 男性にはメンズ商品を推薦し, 女性にはレディース商品を推薦している. 本稿で作成した商品データベースはメンズ商品 5909 点, レディース商品 12193 点の合計 18102 点を取り扱っている. 結果を表 3 に示す. 表中の入力文は「～に着ていく」を省略して表記している.

データベース	入力文	男性平均	女性平均
商品レビュー	友人の結婚式	9.15	7.88
Wikipedia	友人の結婚式	1.04	2.88
商品レビュー	夏祭り	10.84	11.22
Wikipedia	夏祭り	9.75	14.22
商品レビュー	海外旅行	13.80	14.66
Wikipedia	海外旅行	11.84	14.77

表 3: 正しいと思われる商品の平均数

表 3 から商品レビューを用いた推薦の方が Wikipedia を元文書とした推薦より良い推薦が行えていると思われる. また, 商品レビューを元文書とした推薦ではほぼ半数以上の商品が「正しいと思われる商品」であることがわかった.

5 まとめ

本稿では目的指向商品推薦システムのための商品の属性値の自動決定方法の提案および作成したデータベースを使用した商品の推薦方法の提案を行った. 提案手法を用いた被験者実験では多くの場合, 商品レビューを元文書に用いた推薦の方が Wikipedia を元文書に用いた推薦より良い結果が得られた.

参考文献

- [1] 小林幹門, 見並史彬, 伊藤孝行, "概念辞書を用いたユーザの状況を想定した目的指向衣服推薦システムの実装", 第 70 回情報処理学会全国大会, 2008
- [2] Tadahiko Kumamoto, Katsumi Tanaka, "Proposal of Impression Mining from News Articles", KES '05