

商品カテゴリの階層に着目した商品の評価基準抽出手法

山崎義隆[†] 鈴木優[†] 川越恭二[†]

[†] 立命館大学 情報理工学部

1 はじめに

近年、Web 上からの評判情報の抽出に関する研究が多く行われている [1]. その中でも、ある対象のレビューから対象についてどのような評価のポイントがあるのかを抽出する研究が行われている. 赤木ら [2] の研究では、ある対象についてのレビューを記載したページ集合を一定数集め、それらから対象について評価基準を抽出している.

本稿では、評価基準を利用者が商品を購入する際に評価する基準と定義する. 例えば、iPod の場合、デザインや音質、本体の大きさ、液晶などの単語から iPod の商品进行评估することが多い. 従来の手法では、商品固有のレビューを対象に評価基準を抽出する. そのため、レビューが少数の商品について、評価基準を抽出することが困難である.

そこで、本研究では商品が属する商品カテゴリに着目し、商品の評価基準を抽出する手法を提案する. 多くのレビューサイトでは、商品カテゴリが用いられている. また、同一カテゴリに含まれる商品が持つ評価基準は類似していると考えたためである. これにより、商品についてレビューが少数の場合においても、評価基準を抽出できると考える.

2 提案手法

本研究では、商品についてレビューが少数の場合においても商品の評価基準を抽出するために、商品カテゴリに着目する.

提案手法を図 1 に示す. まず、Web 上のレビューサイトから商品カテゴリと商品のレビューを取得する. 取得した商品カテゴリを利用した木構造の葉ノード以外の各ノードをカテゴリとし、葉ノードは商品とする. 同一カテゴリの商品が持つ評価基準は類似していると考え、各カテゴリのレビューを対象に評価基準を抽出する. 2.1 節において評価基準の抽出手法を示す. 各カテゴリにおいて抽出された評価基準を木構造の葉ノード以外の各ノードが共有する. 2.2 節で評価基準の共有手

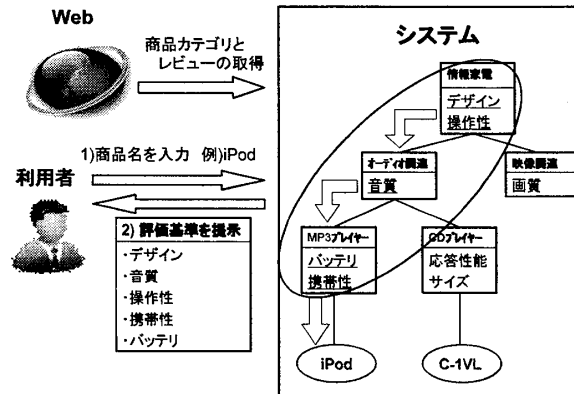


図 1: システムの流れ

法を示す. そして、利用者がシステムに商品名を入力する. 最後に、システムが木構造の根ノードから利用者の入力した商品が属するノードまで辿ることにより、取得した評価基準を利用者に提示する.

2.1 商品の評価基準抽出手法

本研究では、商品の評価基準は名詞であるとする. 形容詞は名詞の性質や状態を表すため、商品の評価基準抽出において形容詞と共起している名詞に着目した. まずは、形態素解析を用いてレビューから登場回数が上位の形容詞を求める. レビューにおいて登場回数が上位にある形容詞は、商品の評価基準を形容している可能性が高いと考える. 登場回数が上位の形容詞において、一文において共起している形容詞の前後 3 個の名詞を抽出する. 登場回数が上位の形容詞と多く共起している名詞は商品の評価基準である可能性が高いと考える.

形容詞の登場頻度が上位 30 件以降は評価基準を形容してないと思われるノイズの出現頻度が高いため、登場回数の上位 30 件の形容詞を $A_x (1 \leq x \leq 30)$ とする. 各文において形容詞 A_x と名詞 n が共起した回数を C_{xn} とし、名詞 n の共起度を C_n とする.

$$C_n = \sum_{i=1}^{30} C_{in} \quad (1)$$

Evaluation Item Extraction Method with Product Category Tree

Yoshitaka YAMASAKI[†], Yu SUZUKI[†] and Kyoji KAWAGOE[†]

[†]College of Information Science and Engineering, Ritsumeikan University.

[†]yamasaki@coms.ics.ritsumei.ac.jp,

[†]{yusuzuki,kawagoe}@is.ritsumei.ac.jp

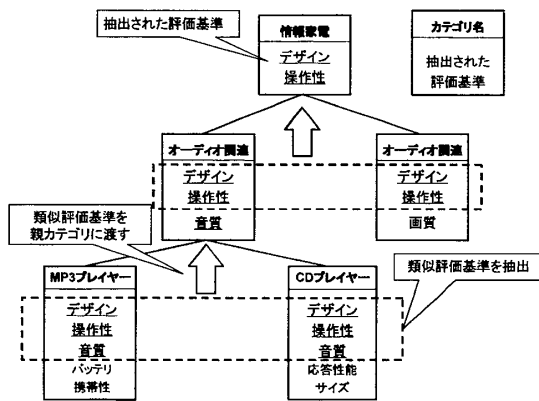


図 2: 商品カテゴリの階層に基づいた評価基準共有手法

2.2 商品カテゴリの階層に基づいた評価基準共有手法

本稿では、価格.com¹などのレビューサイトから取得した商品カテゴリの木構造を利用する。また、木構造の葉ノード以外の各ノードをカテゴリとし、葉ノードは商品とする。

親ノードが同一である葉ノードの商品は類似する商品の評価基準を持つ可能性が高いと考え、親ノードが同一である葉ノードの商品のレビューから商品の評価基準を抽出する。また、多くの子ノードが持つ評価基準が類似しているならば、それらの子を持つ親ノードの評価基準になると考える。図2に示すように、MP3プレイヤーが持つ評価基準とCDプレイヤーが持つ評価基準が類似しているならば、その類似している評価基準は親ノードであるオーディオ関連のノードが持つ評価基準とする。この手法を用いることにより、商品固有のレビュー数に依存することはなく、商品について、レビューが少数の場合においても商品の評価基準を抽出できる。また、新しいカテゴリが登録される際、根ノードからそのカテゴリであるノードまでの各ノードの評価基準を利用することにより、レビューが少数のカテゴリにおいても評価基準を抽出できると考える。

3 実験

3.1 実験環境

提案手法の有効性を示すために評価実験を行う。商品のカテゴリやレビューといった評価データは価格.comから取得する。本実験では、従来手法と提案手法の比較により、レビューが少数の商品を対象に利用者に提示される商品の評価基準の違いを示す。従来手法として、2.1節において述べた手法を用いる。「商品名の～」と書いて意味が通じるものを評価基準である考える。

¹http://kakaku.com/

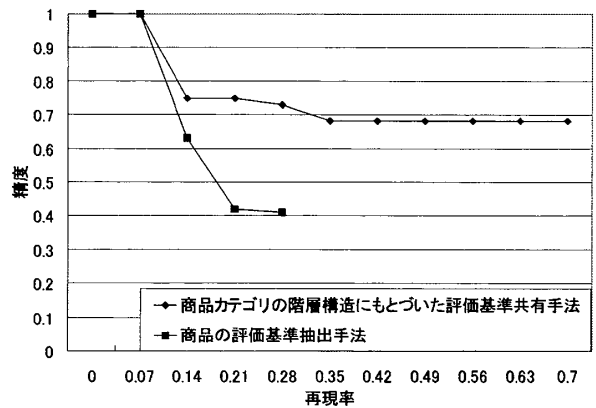


図 3: 再現率・精度曲線

3.2 実験結果と考察

実験結果を図3に示す。2.2節において、類似している評価基準の上位50件以降は評価基準ではないノイズの出現頻度が高い。よって本実験では、上位50件の類似している評価基準を利用者に提示する評価基準と考えた。レビューが少数の場合、2.1節では評価基準が上位に出現しない。つまり、上位30件の結果では、再現率、精度共に低い結果になったと考えられる。また、2.2節では、レビュー数が十分なカテゴリも考慮するため、2.1節よりも再現率、精度共に高い結果を得ることができたと考えられる。

4 おわりに

本稿では、利用者が入力した商品が属するカテゴリのレビューから商品の評価基準を抽出し、商品カテゴリの木構造を用いて評価基準を共有する手法を提案した。これにより、その商品に対するレビューが少数の場合においても、商品の評価基準を抽出できる手法を提案した。

また今後の課題として、2.1節について、レビューページのHTMLソースにおいてタグBRやH等の特定タグに着目することにより、評価基準の抽出精度が向上すると考える。

参考文献

[1] 赤木, 大島, 小山, 田島, 田中: “レビューページ例からの属性抽出に基づくレビューページ検索”, 第17回データ工学ワークショップ (DEWS'06) 論文集 (2006).

[2] 立石, 石黒, 福島: “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, **101**, 189, pp. 75-82 (20010709).