

レビューの構造と評点を利用したアスペクトと語の対応付け

山海 匡史[†] 嶋田 和孝[‡] 遠藤 勉[‡]

[†]九州工業大学大学院情報工学府情報科学専攻 [‡]九州工業大学情報工学部

{m.yamaumi, shimada, endo}@pluto.ai.kyutech.ac.jp

1 はじめに

現在, Web 上には多種多様な製品やサービスに関するレビューが日々蓄積されている. これらのレビューにはその対象に関する個人の評価が多く記述されており, この個人の評価に関する情報 (評価情報) は対象の利用者や提供者の双方にとって有益な情報となる. そのため, レビュー集合から製品やサービスに関する評価情報を自動的に抽出・整理し, 提示するシステムの研究が活発化している [1].

レビューには, 対象そのものだけでなく, その対象が持つ側面や性質に関する様々な記述が存在する. レビュー要約の代表的な手法には, アスペクトに基づくレビュー要約が挙げられる. 評価の視点 (アスペクト) に基づきレビュー要約を行うことで, 視覚的に分かりやすく, 情報の偏りが少ない要約の提示が可能になる. アスペクトに基づくレビュー要約の実現には, アスペクトに関連する記述または語句を特定する必要がある.

現在, 様々なレビューサイトでマルチアスペクトレビューが利用されている. マルチアスペクトレビューとはあらかじめ提示されている複数の評価視点に評点を付与したレビューのことを指す. 本研究では, マルチアスペクトレビューを対象とし, アスペクトに関連する項目 (以下, 関連項目) の抽出を行う.

2 提案手法の概要

関連項目抽出手法として, 本稿では 2 つの手法を提案する. 単語単位で抽出を行うスコアリングに基づく手法 (スコアリング手法) と, 単語クラスタ単位で抽出を行う単語クラスタリングに基づく手法 (クラスタリング手法) である. これら 2 つの提案手法では, レビューの構造と評点に基づいて特徴語を抽出し, 関連項目抽出に利用する. 提案手法の概要を図 1 に示す.

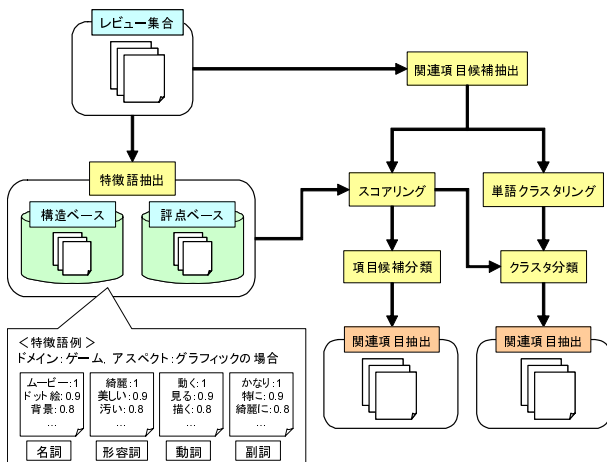


図 1: 提案手法の概要

スコアリング手法では, まず, レビュー集合からレビューの構造と評点に着目して各アスペクトに関する特徴語の抽

出を行う. 抽出した特徴語を用いて項目候補の各アスペクトに対するスコアを算出する. そのスコアが最大となるアスペクトに項目候補を分類し, そのアスペクトの関連項目とする.

クラスタリング手法では, 項目候補集合から単語クラスタを形成し, 前述の項目候補のスコアを用いて, クラスタのアスペクトに対するスコアを算出する. そのスコアが最大となるアスペクトをクラスタに割り当て, 単語クラスタ単位で関連項目抽出を行う. 類似した傾向を持つ単語の集合を作成し, 抽出する.

3 特徴語抽出

レビューの構造と評点の 2 点に着目し, 各アスペクトに関する特徴語の抽出を行う. 特徴語の品詞は名詞 (未知語を含む), 形容詞 (形容動詞を含む), 動詞, 副詞とし, 後述のアスペクトに関する特定の記述またはレビューからそれぞれ抽出する. 形態素解析器には茶筌¹を使用しており, 連続する名詞は複合名詞として扱っている.

3.1 レビュー構造に着目した特徴語抽出

レビューはレビューを書く際に箇条書き等で評価情報を整理して記述することがある. 特にマルチアスペクトレビューでは, 各アスペクトを見出しとして評価情報を記述する傾向が見られる. このような構造に着目し, アスペクトごとの記述から特徴語を抽出する. レビュー構造に着目して抽出した特徴語をここでは StWord と呼ぶ.

手順としては, まず, 記述の構造や記号の出現パターンに基づいてレビュー中の見出し付き箇条書きの抽出を行う. 以下に示すような構造パターンを見出し付き箇条書きとして抽出している. 見出し付き箇条書きの抽出後, その見出しとアスペクトとのマッチングを行い, アスペクトに関する記述の抽出を行う.

(行頭記号) + 見出し + [区切り記号 | 改行] + 内容

3.2 レビュー評点に着目した特徴語抽出

評点が高い/低いレビューには, 特定のアスペクトに関する記述が存在すると考えられる. そこで, あるアスペクトの評点の特徴的なレビューにはそのアスペクトに関する記述が多く存在すると仮定し, そのレビュー中から特徴語を抽出する. あるアスペクトの評点の特徴的なレビューとは, 他のアスペクトは平均的な評点であるが, そのアスペクトの評点だけ際立って高得点または低得点であるレビューのことを指す. レビュー評点に着目して抽出した特徴語をここでは EvalWord と呼ぶ.

EvalWord はノイズも多く含むため, 抽出した特徴語の絞込みを行う. 絞込みには各アスペクトにおける評点の相関関係を利用する. アスペクトの相関関係を用いた候補の絞込みでは, 例えば, アスペクト A とアスペクト B で相

¹<http://chasen-legacy.sourceforge.jp/>

関係がみられなかった場合、お互いが持つ共通の特徴語はそれぞれの評点に無関係であるものとして除外する。

3.3 特徴語のスコア付け

抽出した各特徴語はそれぞれ tf-idf 法を用いてスコア付けを行う。付与したスコアは各特徴語の各品詞で [0,1] の範囲に正規化する。提案手法ではレビュー集合を対象に特徴語抽出を行い、この特徴語のスコアに基づき関連項目抽出を行う。

4 スコアリングに基づく関連項目抽出

関連項目の抽出は、項目候補を各アスペクトに分類することで行う。まず、レビュー集合から項目候補を抽出する。次に、抽出した各候補に対し、前節で得た特徴語を基に素性を作成し、項目アスペクトスコアを算出する。項目アスペクトスコアとは、項目が持つ各アスペクトらしさのスコアである。各スコア算出後、各候補を項目アスペクトスコアの最大値を持つアスペクトに分類し、完了とする。関連項目抽出の概要を図 2 に示す。

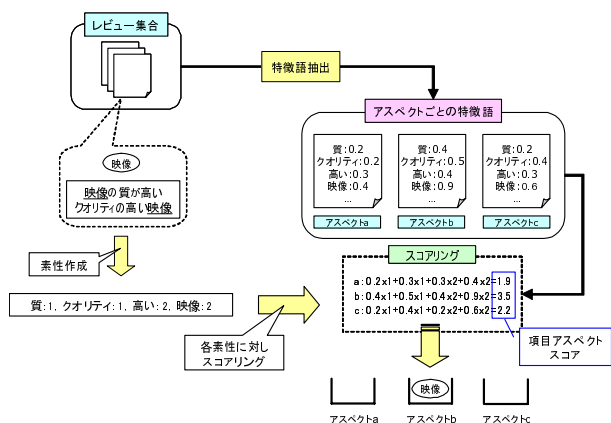


図 2: 関連項目抽出例

4.1 項目アスペクトスコアの算出

素性は分類対象語の周辺語（名詞，形容詞，動詞，副詞）とし、素性の値には出現頻度を用いる。また、対象語自身も素性に追加し、その値は素性中の最大値と同じにする。動詞は他の品詞と比べて高頻出語が多いため、動詞で高頻出かつ評価表現になりにくい語は stopword として除外している²。

単語 t_i のアスペクト asp に対する項目アスペクトスコア $t_i\text{-Score}_{asp}^{(s)}$ は以下の式によって求める。

$$t_i\text{-Score}_{asp}^{(s)} = \sum_{f_j \in F_i} fval_j(S_{j,asp} + E_{j,asp}). \quad (1)$$

各アスペクトで項目アスペクトスコアを算出し、このスコアが最大となるアスペクトに単語を分類する。 F_i は単語 t_i の素性集合を表し、 $fval_j$ は素性 f_j の値を表す。 $S_{j,asp}$ は StWord における f_j の値、 $E_{j,asp}$ は EvalWord における f_j の値を表し、各特徴語に f_j が含まれない場合の値は 0 である。

²4174 件のレビューでの高頻出語の上位 5 件はすべて動詞であり「する」「ある」「なる」「思う」「できる」となっている。この中で「する」「なる」「思う」を除外している。

4.2 特徴語の重み設定

分類に利用する特徴語について、どの品詞も等しく関連項目の推定に貢献するとは限らない。さらに、素性内の品詞の割合が異なるため、割合の大きい品詞に分類結果が左右されることもある。そのため、スコアリングの際には特徴語の品詞ごとに重みの設定を行う。また、EvalWord は StWord に比べてノイズが多い場合があるため、EvalWord 全体の重みを調整する。StWord の信頼度をさらに高めるために、StWord にも重みの調整を行う。StWord 内の語の各アスペクトにおける出現分布を考慮して、その語の重みを調整する。

人手で訓練データを与えて重みを設定する方法や経験的に重みを設定する方法はコストがかかる。この節では自動的に重みを設定する方法について述べる。

4.2.1 品詞別の重み設定

人手で訓練データを与える代わりに、StWord から抽出できる高信頼度のデータを訓練データとすることで自動的に重みの設定を行う。StWord の tf-idf スコア上位のものは関連項目としての信頼度が高い。その中からさらに評価項目に絞ることでより信頼度の高い関連項目の抽出を行い、訓練データとして使用する。

まず、StWord から評価項目の抽出を行う。評価項目の抽出には以下の共起パターンにより評価項目を抽出する。評価表現に関しては、小林らの評価表現辞書を利用した [3]。

[評価項目] が / は / も / に / を [評価表現]

例 [グラフィック] が [綺麗]

[評価表現] [評価項目]

例 [綺麗な] [グラフィック]

各アスペクトの StWord から抽出した評価項目の tf-idf スコア上位 10 件をそのアスペクトの訓練データとして使用する。他のアスペクトと重複した項目は削除して下位の項目を繰り上げる。各品詞の重みを訓練データに対して累計分類誤差が最小となるように機械的に調整する。分類誤差とは誤って分類した語のアスペクトスコアから分類されるべきだった語のアスペクトスコアの差のことを指す。tf-idf スコア上位 5 件に対しても同様のことを行い、それぞれの累計分類誤差の和が最も小さくなる重みの組み合わせを採用する。

4.2.2 EvalWord・StWord の重み設定

EvalWord に関しては、前節と同様に累計分類誤差が最小になるように重みの調整を行う。StWord はアスペクトに関する記述から抽出した特徴語であるため高信頼度ではあるが、各アスペクトで重複するものも存在する。特徴語の信頼度をさらに高めるため、スコアリングの際には重複するものに対して特徴語スコアの補正を行う。アスペクト数の 3/4 以上のアスペクトに出現する特徴語には 0.25 の重み付けを、1/2 以上のアスペクトに出現する特徴語には 0.50 の重み付けを、1/4 以上のアスペクトに出現する特徴語には 0.75 の重み付けをそれぞれ行う。

5 単語クラスタリングに基づく関連項目抽出

スコアリングに基づく関連項目抽出では、単語単体に対してスコアリングを行っているが、ここではある基準に基づき似ている単語をまとめ、単語の集合に対してスコアリングを行う。

5.1 単語クラスタリング

単語特徴ベクトルの素性には、周辺語の出現頻度を用いる。動詞には 4.1 節と同様に stopword を設定している。

クラスタリングには汎用的データクラスタリングツールである bayon³ を使用している。bayon はクラスタリング手法に Repeated Bisection 法を採用し、データ集合を繰り返し 2 分割することでクラスタリングを実行する。適切なクラスタ数を決めるのは難しいため、bayon の機能を使い、ある程度適切な分割が完了するまで分割する設定にしている⁴。

5.2 クラスタへのアスペクト割り当て

クラスタ C_k のアスペクト asp に対するクラスタアスペクトスコア $C_k-Score_{asp}$ は以下の式によって求める。

$$C_k-Score_{asp} = \sum_{t_i \in C_k} t_i-Score_{asp}^{(s)} \delta_{ik} \quad (2)$$

このスコアが最大となるアスペクトをクラスタに割り当てる。 $t_i-Score_{asp}$ は単語 t_i の項目アスペクトスコアであり、 δ_{ik} は単語 t_i のクラスタ C_k に対する帰属度である。アスペクトの割り当ての概要を図 3 に示す。

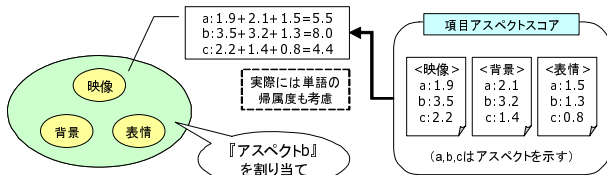


図 3: アスペクトの割り当て例

5.3 項目アスペクトスコアの作成

アスペクトを割り当てたクラスタを基に新たな項目アスペクトスコア $t_i-Score_{asp}^{(c)}$ を生成する。以下の式のように単語の各アスペクトが割り当てられたクラスタへの帰属度の最大値をスコアとする。

$$t_i-Score_{asp}^{(c)} = \max_{t_i \in T} \delta_{ik_{asp}} \quad (3)$$

$\delta_{ik_{asp}}$ は単語 t_i のアスペクト asp が割り当てられたクラスタ C_k に対する帰属度を表す。

6 評価実験

本研究では、現在、ゲームレビューサイト⁵のレビューを研究対象として扱っている。対象レビューには「オリジナリティ(o)」、「グラフィックス(g)」、「音楽(m)」、「熱中度(a)」、「満足感(s)」、「快適さ(c)」、「難易度(d)」がアスペクトとして与えられている。研究対象であるマルチアスペクトレビューの例を図 4 に示す。

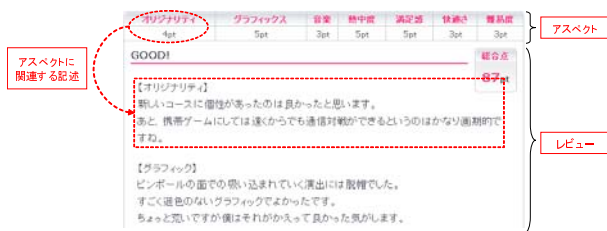


図 4: マルチアスペクトレビュー

³ <http://alpha.mixi.co.jp/blog/?p=1049>

⁴ 分割ポイントは 1.0 に設定、-idf オプションを使用

⁵ Nintendo DS mk2 (<http://ndsmk2.net/>)

評価実験に使用するレビューとしては 4174 件収集しており、複数のタイトルが含まれている。特徴語抽出や素性作成には収集した全レビュー集合のデータを利用する。

評価実験では、スコアリング手法とクラスタリング手法の精度を比較する。まず、あるタイトルのレビュー集合を対象とし、人手で正解データを作成し、得られた関連項目の精度を比較・検証する。しかし、一般に、抽出されたすべての関連項目を客観的に評価するのは難しく、正解データの作成も高コストである。そこで、提案手法によって得られた関連項目を評価文のアスペクト分類のタスクに適用し、文のアスペクト分類精度を比較することでそれぞれの手法の有効性を検証する。

6.1 関連項目抽出実験

提案手法を用いて関連項目の抽出を行う。ここでは、明確な関連項目に関する分類精度に関して評価・考察を行う。

6.1.1 実験内容

収集したレビュー集合から、あるタイトルのレビュー集合 130 件を対象とし、実験を行った。分類対象となる項目候補の異なり単語数は 1817 語である。ただし、素性に特徴語を含まないものは分類対象から除外している。素性の範囲は文を区切りとして対象語の前後 2 語とした。4.2 節の方法で自動的に設定した各特徴語の重みは、それぞれ、EvalWord:1.0, 名詞:0.5, 形容詞:1.0, 動詞:0.1, 副詞:0.3 となった。

各項目候補を「o」、「g」、「m」、「a」、「s」、「c」、「d」のいずれかのアスペクトに分類した。事前に各アスペクトの関連項目を 20 個ずつ主観的に選出してもらい、その項目が正しく分類されているかを調査した。なお、特徴語の重み設定で使用した訓練データは評価データには含まない。

6.1.2 実験結果・考察

スコアリング手法の実験結果を表 1 に示し、クラスタリング手法の実験結果を表 2 に示す。各 20 語を正解データとした場合の再現率を上段に、適合率を中段に示す。下段にはそのアスペクトに分類された単語数を示す。

表 1: 実験結果 (スコアリング)

	o	g	m	a	s	c	d
再現率	75% (15/20)	90% (18/20)	100% (20/20)	75% (15/20)	45% (9/20)	75% (15/20)	45% (9/20)
適合率	56% (15/27)	90% (18/20)	91% (20/22)	58% (15/26)	82% (9/11)	75% (15/20)	75% (9/12)
単語数	421	112	131	386	157	356	225

スコアリング手法では、全体で 72.1% (101/140) の再現率を得ることができた。重みを設定しなかった場合の全体の再現率は 62.9% であり、最適な重みが設定された場合の全体の再現率は 75% であった。

表 2: 実験結果 (クラスタリング)

	o	g	m	a	s	c	d
再現率	45% (9/20)	55% (11/20)	85% (17/20)	60% (12/20)	30% (6/20)	70% (14/20)	40% (4/20)
適合率	31% (9/29)	65% (11/17)	63% (17/27)	43% (12/28)	67% (6/9)	67% (14/21)	57% (4/7)
単語数	506	79	106	439	128	392	298

クラスタリング手法では、全体で 52.1% (73/140) の再現率となり、スコアリング手法よりも 20% の低下が見られた。

「オリジナリティ」の適合率が低いことからわかるように、誤った事例の多くは「オリジナリティ」に分類されていた。これは、「オリジナリティ」が曖昧なアスペクトであり、他のアスペクトにも分類できるような項目も含んでいるためである。評価文に人手でアスペクトを割り当てるタスクにおいても「オリジナリティ」と「満足感」は混同されやすい傾向にあることがわかっており [4]、関連項目の分類においても「満足感」の多く項目が「オリジナリティ」の項目として分類されていた。同様な傾向は「難易度」と「快適さ」にも見られる。「快適さ」が「操作視点の快適さ」の側面だけでなく、「ゲーム進行の快適さ」の側面も含むことができることが挙げられる。

スコアリング手法と比較してクラスタリング手法での全体の再現率は大きく低下してしまった。しかし、分類結果の内訳を調べると、スコアリング手法で正しく抽出できなかった項目の抽出ができていたことが確認できた。

クラスタリング手法による分類結果は形成されたクラスタにも依存するため、更なる精度向上のためには素性の選択やクラスタの形成方法を改善する必要がある。

6.2 評価文アスペクト分類実験

ここでは提案手法で抽出した関連項目を利用して評価文のアスペクト分類を行い、それぞれの手法の精度を比較・考察し、得られた関連項目の有効性を検証する。

6.2.1 実験内容

アスペクトがタグ付けされた評価文 700 文を対象とし、評価文のアスペクト分類を行った。この 700 文には各アスペクトに関連する評価文が 100 文ずつ含まれる。

6.2.2 実験結果・考察

評価文の分類は、特徴語、スコアリング手法およびクラスタリング手法で得られたスコアに基づき行われる。まず、3.3 節の特徴語スコアを用いて、評価文 S_i のアスペクト asp に対する特徴語アスペクトスコア $S_i-Score_{asp}$ を算出する。特徴語の重みは最適になるように調整した。

特徴語アスペクトスコアと各提案手法での項目アスペクトスコアを統合したアスペクトスコア $TS_i-Score_{asp}$ を以下の式によって求める。

$$TS_i-Score_{asp} = \alpha S_i-Score_{asp} + \beta \sum_{t_k \in S_i} t_k-Score_{asp}^{(s|c)}. \quad (4)$$

この統合アスペクトスコアにより、評価文のアスペクト分類を行う。 α 、 β は各アスペクトスコアに対する調整パラメータである。 α 、 β を最適に設定した場合の分類精度を表 3 に示す。

表 3: 評価文分類結果

	統合アスペクトスコア
スコアリング手法	65.71%
クラスタリング手法	66.14%

ここで、 $\beta = 0$ のとき、つまり特徴語アスペクトスコアのみでの分類精度は 65.47% であった (ベースライン)。このベースラインと比較すると、2 つの提案手法とも、項目

アスペクトスコアを考慮することで若干ではあるが分類精度の向上が見られた。また、6.1 節の実験では、クラスタリング手法の精度は十分ではなかったにもかかわらず、評価文分類ではスコアリング手法を上回っていた。これは、クラスタリング手法が潜在的に有効な関連項目をうまく抽出できていたためだと考えられる。分類結果の内訳を調べると、特徴語アスペクトスコアで誤った分類が項目アスペクトスコアを統合することによって正しく分類されるようになる事例は、クラスタリング手法を使用した場合の方が多かった。

7 関連研究

近年、アスペクトに関する記述や語句を特定する研究が行われており、Blair-Goldensohn らは主に機械学習を用いてアスペクトに関連する記述を特定している [2]。しかし、訓練データの作成は一般に高コストである。一方、Ivan らは文書集合の生成モデルである LDA を拡張した Multi-Grain LDA を用いて、訓練データを必要とせず、ある話題で共通する単語の集合を作成し、アスペクトに話題ごとの単語を割り当てている [5]。これらの手法を参考にすることで、今後更なる精度向上を目指す。

8 おわりに

本研究では、レビューの構造情報と評点に着目し、アスペクトに関連する項目の抽出手法を提案した。評価実験により、提案手法を用いて関連項目の抽出ができることを確認した。また、抽出した関連項目を評価文のアスペクト分類に適用することで、その有効性の確認もできた。

今後は、単語特徴ベクトルの素性の検討を行い、クラスタリング手法の再現率の向上を目指す。また、スコアリング手法とクラスタリング手法を統合した新たな手法についても検討し、関連項目抽出及び評価文アスペクト分類の更なる精度向上を目指す。

謝辞

この研究の一部は栢森情報科学振興財団の助成を受けて遂行された。

参考文献

- [1] 乾孝司, 奥村学. テキストを対象とした評価情報の分析に関する研究動向. 自然言語処理, Vol.13, No.3, pp.201-242, 2006.
- [2] Sasha Blair-Goldensohn, et al., Building a sentiment summarizer for local service reviews. WWW2008: NLPix Workshop, 2008.
- [3] 小林のぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一. 意見抽出のための評価表現の収集. 自然言語処理, Vol.12, No.2, pp.203-222, 2005.07
- [4] Tadano et al., Effective construction and expansion of a sentiment corpus using an existing corpus and evaluative criteria estimation. PA-CLING2009.
- [5] Ivan Titov, Ryan McDonald., A Joint Model of Text and Aspect Ratings for Sentiment Summarization. Association for Computational Linguistics (ACL), 2008.