

レビューの構造と評点に着目した関連項目抽出

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

1 はじめに

近年、レビュー集合の要約システムに関する研究が活発になりつつある。Blair-Goldensohnらは評価の視点(以下、アスペクト)に基づく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を可能としている[1]。アスペクトに基づくレビュー要約では、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。

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。関連項目抽出手法として、レビューの構造と評点に着目した手法を提案している。

2 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抽出を行う。特徴語は名詞、未知語、形容詞、形容動詞、動詞とし、特徴語抽出にはtf-idf法を利用した。連続する名詞は複合名詞として扱っている。

2.1 レビュー構造に着目した特徴語抽出

マルチアスペクトレビューでは、レビューの書き手は箇条書き等でアスペクトごとに評価を整理して記述する傾向がある。このような構造を抽出し、各アスペクトに関する特徴語を抽出する。

2.2 レビュー評点に着目した特徴語抽出

あるアスペクトの評点が際立って高い/低いレビューには、そのアスペクトに関連した記述が存在する傾向があることを利用する。例えば、あるアスペクトAの評点のみが高い/低いレビューをAが特徴的なレビューとし、このレビューから抽出した特徴語をAに関するものとする。

3 関連項目抽出

関連項目の抽出では、抽出対象となるレビュー集合に含まれる項目候補を各アスペクトに分類することで行う。処理の流れは以下の通りである。

- レビュー集合から項目候補(名詞・未知語)を抽出
- 抽出した各項目に対し単語特徴ベクトルを作成
- 各特徴語を用いて単語特徴ベクトルをスコアリング
- 各項目をスコアが最大になるアスペクトに分類

単語特徴ベクトルの素性には対象語の周辺語を用いる。スコアリングでは、単語特徴ベクトルの有効な素性に対し、各アスペクトの特徴語でそれぞれスコアリングを行う。素性が特徴語に含まれているとき、その特徴語のtf-idf値をスコアとし、各素性のスコアを足し合わせる。各アスペクトで合計スコアを算出し、それが最大となるアスペクトに項目を分類する。

4 評価実験

4.1 実験データ

本研究では、現在、ゲームレビューサイト¹から様々なゲームのレビューを収集している。実験に使用するマルチ

アスペクトレビューの例を図1に示す。収集したレビュー集合から、あるゲームのレビュー集合130件を対象とし実験を行った。

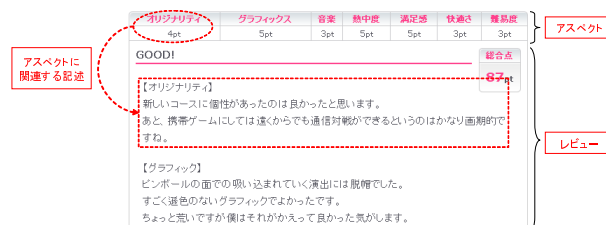


図1: マルチアスペクトレビュー

4.2 実験内容

分類対象となる項目候補の異なり単語数は1822語である。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。特徴語抽出や単語ベクトル作成には収集した全レビュー集合のデータを利用する。

各単語を「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)」のいずれかのアスペクトに分類し、抽出精度を調査した。

4.3 結果・考察

それぞれの分類結果から単語をランダムで30個選び、その単語が正しく分類されているか調査した。主観的な評価であり、その単語によってアスペクトを想起できるのであれば正しいものとした。実験の結果を表1に示す。

表1: 実験結果

	o	g	m	a	s	c	d
正解数	6/30	13/30	8/30	10/30	6/30	8/30	15/30
単語数	1026	69	33	102	175	347	43

関連項目を抽出できることは確認できたが、全体的に精度は低かった。この原因としては、すべての名詞や未知語を分類対象としたことが挙げられる。そのため、スコアに閾値を設定し、分類する単語の制限を行うことで精度向上が期待できる。また、大多数が「オリジナリティ」に分類された原因としては、そのアスペクト自体の曖昧性が挙げられる。これは「オリジナリティ」が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。精度向上のためには特徴語の改善や重み付けも行う必要があることがわかった。

5 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。評価実験では、関連項目を抽出できることを確認した。

参考文献

- [1] Sasha Blair-Goldensohn, et al., "Building a sentiment summarizer for local service reviews", WWW2008: NLPix Workshop, 2008.

¹Nintendo DS mk2 (<http://ndsmk2.net/>)