

複数の評価項目を持つレビューの評価値の推定

Seeing Several Stars in a document  嶋田和孝, 小濱祐樹, 遠藤勉 (九工大)

はじめに

■ 評価表現の分類

- P/N 分類 or 多値分類

■ 本当に評価項目は 1 つ?

- 人間はいくつかの項目について評価しているはず

■ 各々の評価項目の評価値推定

- より詳細な評判情報の活用
- 何がどんな項目に有効な情報なのか
- どの情報がどの程度の良さ / 悪さを示すのか

■ 多項目で多値の評判情報の分類 (評価値の推定)

レビュー記事

掲載日: [2006/09/21]									投稿者:	さん	
オ	グ	音	熱	満	快	ス	(難)	クリア	プレイ時間		
2点	4点	2点	2点	3点	2点	2点	2点	未	20時間以上30時間未満		
<p>良い</p> <ul style="list-style-type: none"> ・今までのゲームの中で最高クラスの綺麗なムービーでした。 ・ガンビットシステムも初めは組み合わせる「行動」が少ないですが出来る事が増えてくと考 ・える楽しみがある。(少し面倒くさいが…) <p>悪い</p> <ul style="list-style-type: none"> ・戦闘のシステムがオンラインゲームのようになっているのはダメです。 ・良いところにも書いたガンビットの欠点 モンスターにはそれぞれの弱点や有効な攻撃(物理や魔法)があるのでガンビットでは少し厳しい。 ・召喚獣が魔法のように手軽に使えない。(召喚したキャラと召喚獣で戦う) ・ダンジョンが広くセーブポイントが少ない。町も広い。 ・宝箱もランダムだったり決まっているアイテムが出る宝箱も確率だったり意味不明。 ・このゲームの世界にはいろんな種族がいるのに仲間が人間ばかり(プレイヤー) ・敵がお金を落とさない。 <p>悪い</p> <ul style="list-style-type: none"> ・悪いところばかり書きましたが「FF」という一種のブランドを背負っている高仕方が無いですね。し ・想かしてプレイしてみると結構面白いです。このサイトの評価はあまり良くは無いですが評価だけ見 ・などやらないのは大損です。 											

項目数は 8 項目 評価値は 6 値 (0 ~ 5 まで)

手法と素性

■ Support Vector Regression (SVR) を利用

■ レビュー中の明示的な情報を利用

- 「良い所」・「悪い所」
 - ・「良い所」の単語は +1, 「悪い所」の単語は -1
- 「感想」は利用しない
 - ・今回のデータでは精度向上には繋がらないことを確認

■ 素性選択

- 全ての単語が全ての項目に貢献している?
- ある単語の各項目における信頼度
 - ・ $var(w_{c_j}) = \frac{1}{m} \sum_{i=0, w \in d_i} (real(d_i, c_j) - ave(w_{c_j}))^2$
 - ・ 出現頻度が n 回以上 (今回の実験では $n=4$)
 - ・ 評価値の平均 ≥ 3 (良い所), ≤ 3 (悪い所)

実験結果

■ 平均二乗誤差, 適合率, 再現率で評価

■ 推定結果は整数化

	Ave	SVR	var=0.75	var=0.5
オリジナリティ	1.15	0.87	0.74	0.48
グラフィックス	0.62	0.49	0.52	0.45
音楽	1.30	0.77	0.75	0.50
熱中度	1.86	1.01	0.85	0.50
満足度	2.04	0.95	0.71	0.36
快適度	1.63	1.07	0.79	0.38
ストーリー	1.43	0.78	0.62	0.32
難易度	0.66	0.65	0.71	0.57
平均	1.34	0.82	0.71	0.45
適合率 (P)	0.34	0.45	0.44	0.49
再現率 (R)	0.34	0.45	0.36	0.33
カバー率 (C)	1.00	1.00	0.83	0.66

■ ベースライン (Ave) より高い推定精度

■ 適合率はあまり向上しない

- タスクの難しさ (8 項目で 6 値分類)
- カバー率の問題
 - 根本的に推定できないと思われるレビューが 25%
 - $var=0.5$ ぐらいまでは許容範囲?
 - $var=0.25$ の場合, $MSE=0.2$, $P=0.58$, $R=0.26$, $C=0.45$

■ 素性選択手法の再考察

- n -gram の利用
- 特定の係り受け関係の利用
 - ・ 大きな精度向上はなかった

単語とある項目における平均スコアの例

単語 (項目名)	平均
採用 (オリジナル)	3.8
リアリティ (グラフィック)	4.6
圧巻 (グラフィック)	4.6
回収 (ストーリー)	1.0

今後の課題と展望

■ 十分とはいえない精度

- 問題の難しさ・推定の意義

■ 精度向上に向けて

- 学習方法の改良
 - ・ 複数の分類器の結果を利用
- より有効な素性の選別
 - ・ 各項目に重要な単語はもっと少ない?

■ 単語のスコアを計算^[1]

- P/N 分類用の手法
- データは価格 .com
 - ・ 右例は VAIO VGN UX50
- ユーザ評価とスコアの関係
 - ・ 本タスクへの適用可能性

項目	スコアを見た単語	単語のスコア	ユーザ評価
デザイン	デザイン	1.433	4.0
処理速度	速度	0.007	3.5
グラフィック性能	グラフィック	スコアなし	3.6
拡張性	拡張	-0.349	3.0
使いやすさ	機能	-2.132	2.8
携帯性	携帯	0.542	4.0
バッテリー	バッテリー	-2.284	2.2
液晶	液晶	1.535	4.0
満足度	満足	1.433	3.6

[1] 河野, 嶋田, 遠藤: スコアリングを基にした評価文の分類, 火の国シンポジウム 2007, A-3-4.