

★ 目的

- ★ 評価分類のタスク
 - ★ P/N 分類・P/N/N 分類・さらに多い場合
 - ★ 1 文書に 1 評価?
 - ★ 人間はいくつかの項目について評価しているはず
 - ★ 多項目で多段階の評判情報の分類
 - ★ 1 つの文書が複数の評価項目を持つ
 - ★ 1 つの評価値は多段階の評価値を持つ
- Seeing Stars の拡張版 :: Seeing Several Stars**

★ レビュー記事

オ	グ	音	熱	満	快	難	クリア	プレイ時間
2点	4点	3点	4点	4点	3点	2点	済	10時間以上20時間未満
良い所								
・マリオカートらしいアイテムとテクの微妙な駆け引きは健在。 ・グラフィックは非常に綺麗で、また、1人でもバトルゲームが遊べる点も魅力。 ・新たに追加されたステージも最近のマリオのゲームをモチーフとしたものが多く、見ていて楽しめる。 ・旧作のステージが高画質で楽しめる。								
悪い所								
・ドリフトにクセがあり、しかもそれをマスターしないと絶対に早く走れない。 ・所持しているアイテムまでもが攻撃で消えることが多すぎる。 ・キラーなどの強すぎるアイテムがゲームバランスを崩している。 ・風船バトルはCPUがやたら自分ばかり攻撃してくる。 ・サンシャインバトルが簡単すぎ。 ・カートの種類はたくさんあるが、重いか軽いかぐらいでいたい内容が同じ。								
感想など								
マリオカートの正常進化版ですが、とにかく1人でもバトルゲームができる点が魅力的です。								

<http://ndsmk2.net/>

DS のゲームソフト : 1114 記事・複数のソフト
 評価項目 : 7 つ 評価値 : 6 段階 (0 ~ 5 まで)

★ 手法の比較

- ★ 手法
 - ★ Support Vector Regression (SVR)
 - ★ Support Vector Machine (SVM)
 - ★ Maximum Entropy (ME)
 - ★ 類似度に基づく手法 (SIM)
- ★ 素性
 - ★ Bag of words (BOWs)
 - ★ 名詞・形容詞・副詞・動詞
 - ★ 「良い所」「悪い所」の単語は別々に扱う
 - ★ 「感想など」は使用しない

★ 何がしたい / 何が嬉しい

- ★ 各々の評価項目の評価値を推定
 - ★ より詳細な評判情報の利用
 - ★ 項目とそれに有効な情報の関係
 - ★ どの情報がどの程度の良さ / 悪さを示すのか
- ★ 複数の項目を扱うこと
 - ★ レビュー記事中の (潜在的な) 複数の項目の評価を一気に推定できる
- ★ 多段階の評価値を扱うこと
 - ★ 詳細な要約を作成できるようになる

知識源 : P/N - good
 この PC は良い。

知識源 : 多段階 - 性能 : 5, 携帯性 : 2, 総合 : 4
 性能の面では素晴らしい。ただ、持ち運ぶにはちょっと不向き。しかし、全体的に見るといい製品だと思う。

★ 実験 1

- ★ 実験 : 5 分割交差検定
- ★ 評価尺度 : 平均二乗誤差 (MSE) ・ 標準偏差 (SD) ・ 精度

	All3	Ave	SVR	SVM	ME	SIM	
平均二乗誤差	オリジナリティ	1.26	1.54	0.88	0.91	0.98	1.03
	グラフィックス	1.03	0.85	0.74	0.78	0.82	0.84
	音楽	1.21	0.79	0.70	0.69	0.75	0.77
	熱中度	1.89	1.89	1.21	1.54	1.44	1.45
	満足度	1.97	1.77	1.22	1.54	1.57	1.42
	快適度	1.29	1.29	1.13	1.24	1.35	1.27
	難易度	1.74	1.17	1.22	1.23	1.35	1.18
	平均	1.48	1.33	1.02	1.13	1.18	1.14
標準偏差	0.17	0.24	0.12	0.19	0.19	0.20	
精度	26.60	37.69	41.37	41.76	40.23	39.47	
PNN 精度	26.60	51.98	57.41	58.43	57.05	55.71	

PNN 精度 0-2: Negative, 3: Neutral, 4-5: Positive

- ★ ベースライン
 - ★ All3 : 全て 3 だと仮定
 - ★ Ave : 訓練データの平均値で推定
- ★ SVR がベスト
 - ★ MSE が小さく, SD も小さい
- ★ Naive Bayes と C4.5
 - ★ ベースライン (Ave) よりも悪いことがある

★ 素性選択

- ★ 評価表現でないものが悪さをしている?
- ★ 高村ら [1] の PN 辞書を利用 ⇒ 精度の低下 : ドメイン特有語の欠落

[1] 高村大也, 乾孝司, 奥村学 "スピンモデルによる単語の感情極性抽出", 情報処理学会論文誌ジャーナル, Vol.47 No.02 pp. 627-637, 2006.

★素性選択の改良

★素性拡張

★BOWs + α : 単語の系列

★長さ l の 2 単語の共起 (1 文内に限定)

入力 :: Wi-Fi での対戦は白熱!

素性 :: Wi:Fi, Wi: 対戦, Wi: 白熱, Fi: 対戦, Fi: 白熱, 対戦: 白熱

★素性選択

★全ての単語 (系列) が全ての項目に関連している??

★ある単語 (系列) の各項目における信頼度

★評価値のブレを基準にする

$$var(w_{c_j}) = \frac{1}{m} \sum_{i=0, w \in d_i}^n (real(d_i, c_j) - ave(w_{c_j}))^2$$

[いい:曲]に関する var の値

オ:0.9, グ:1.8, 音:0.2, 熱:0.8, 満:1.2, 快:1.4, 難:1.6

閾値 (0.25) :: 音楽の学習データのみ利用

★出現頻度と出現場所

★出現頻度が f 回以上

★評価値 ≥ 3 (良い所) 評価値 ≤ 3 (悪い所)

★実験 2 ・手法は全て SVR

★ var の値と精度変化

★ $l = 4, f \geq 1$

var	0.25	0.5	0.75
MSE (平均)	0.99	0.99	0.99
MSE (標準偏差)	0.12	0.12	0.12
精度	41.40	41.59	41.56
PNN 精度	57.49	57.49	57.46

★長さ (l) の値と精度変化

★ $var = 0.5, f \geq 1$

l	1	2	6
MSE (平均)	1.02	1.00	0.99
MSE (標準偏差)	0.12	0.12	0.13
精度	41.39	41.39	41.46
PNN 精度	57.08	57.34	57.26

★頻度と精度変化

★ $l = 4, f \geq 2$

var	0.25	0.5
MSE (平均)	1.02	1.01
MSE (標準偏差)	0.12	0.12
精度	41.11	41.32
PNN 精度	56.98	57.05

★BOWs は頻度と位置のみの制約・ var は系列だけに利用

★パラメータを変更しても平均二乗誤差, 精度ともに変化なし

★素性選択の対象を変更

★ var をコーパス全体から算出して使用

	Bow1	Bow2	Comb1	Comb2	Comb3
MSE (平均)	0.72	0.89	0.74	0.84	0.99
MSE (標準偏差)	0.12	0.13	0.11	0.17	0.13
精度	50.72	44.65	52.72	46.96	42.39
PNN 精度	69.41	61.70	69.18	61.74	58.30
処理対象	0.70	0.87	0.86	0.93	1

Bow1 :: BOWs のみ $var=0.25, f \geq 2$

Bow2 :: BOWs のみ $var=0.5, f \geq 2$

Comb1 :: Bow1+ 系列 $var=0.25, f \geq 2, l=4$

Comb2 :: Bow2+ 系列 $var=0.5, f \geq 2, l=4$

Comb3 :: BOWs は選択なし + 系列 $var=0.5, f \geq 2, l=4$

★全データから var を算出すると有効

★訓練データのみに対して $comb1$ と同条件で実験すると

MSE = 1.19 (SD=0.23), 精度 = 38.91, PNN 精度 = 54.63

★ var を全ての素性の制約にすると処理できない文書が生まれる

★まとめ・今後の課題と展望

★ベースラインよりは高精度

★SVR が最も安定している

★十分とはいえない精度

★精度は 4 割程度. MSE も決して小さくない

★PNN 精度でさえ 6 割未満

★素性選択・拡張

★精度に大きな変化はなし

★全データを使用

★概ね精度は向上

★訓練データ数の問題?

★精度と処理対象数のトレードオフ

★評価項目に対して記述のないレビューもある

★30%前後の項目は例え人間でも根本的に推定できない

★より有効な分類手法と素性選択の確立

