

客観的事実の抽出と利用

- 客観的情報 1 : 製品のスペック
 - 製品のスペック情報抽出
 - スペック情報に基づく製品特徴の抽出
 - 抽出データに基づく要約文の生成

	PC1	PC2	PC3	PC4
重量	1.58kg	1.41kg	1.32kg	1.65kg
スコア	2.2点	7.6点	10点	0点

特徴データ
製品名: PC3 属性名: 重量 属性値: 1.32kg
製品名: PC2 属性名: 重量 属性値: 1.41kg

1st: この中で [属性名] が最も軽いのは [製品名] で [属性値].
2nd: [製品名] も比較的軽い.
補足: 条件 (属性名 : 重量 & 順位 : 1st)
携帯性には最も優れている

この中で重量が最も軽いのは PC3 で 1.32kg.
携帯性には最も優れている.
PC2 も比較的軽い.

- 客観的情報 2 : 売り上げなど
 - シンプルなテンプレートベースの情報抽出 (製品) の [売り上げ | 出荷台数] は (事実 : 台数) 台

重要文抽出

- 基本的な考え方
 - シンプルな従来の重要度算出のモデルで実装
 - 評価文は抽出済み & アスペクト推定済みという前提
 - tf・idf ベースのスコアリング

$$tf_j^i = \frac{\log(\text{対象レビュー}j\text{内での単語}i\text{出現回数}+1)}{\log(\text{対象レビュー}j\text{内での単語の異なり数})}$$

$$idf_i = \log \frac{\text{全レビュー数}}{\text{単語}i\text{を含むレビュー数}} + 1$$

-文sの重要度

$$\frac{\text{文}s\text{中の単語の}tfidf\text{値の和}}{\text{文}s\text{の総単語数}}$$

熱中度 (a)
コイン集めなどにはまった
スターコインを集めるといやりこみ要素も充実
快適度 (c)
甲羅マリオは使いにくいし、デカマリオは場所を選ぶ
「マリオVSルイージ」の土管を壊して土管の位置にいと埋まってしまう
難易度 (d)
これから難しくなるだろう・・・と思っていたらワールド8になってクッパを倒してしまった
ラスボスも、巨大マリオで一撃ですぐに倒してしまう
グラフィックス (g)
水中面も綺麗
グラフィックは もちろん、アクションも豊富で、マリオが進化したって実感しました
音楽 (m)
特に、水中面やマリオVSルイージのバトルゲームの曲は最高
ステージ音楽が心に残らない
オリジナリティ (o)
巨大マリオ、豆マリオが使えない
チビマリオ、コウラマリオ、デカマリオが全く使えない
満足度 (s)
やっぱり決めてはビッグキノコ
スターコイン探しが面白い

アスペクトの関連情報の抽出

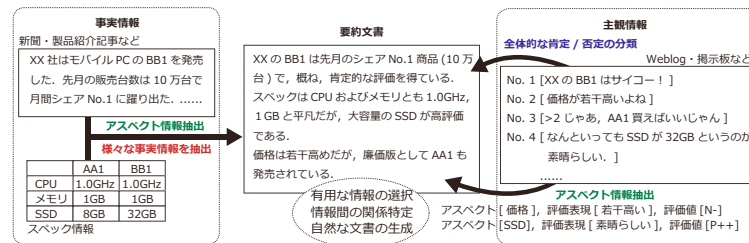
- アスペクトに関連する語の推定
 - ゲームのレビュー (約 4000 件)
 - レビューの構造情報に着目
 - アスペクトごとの記述 = 高信頼度な情報
- 箇条書き抽出
 - シンプルな抽出パターン (記号)+見出し+区切り記号+内容
- tf・idf でソート

熱 (a)	快 (c)	難 (d)	グ (g)	音 (m)	オ (o)	満 <s>
クリア	戦闘	難易度	ムービー	曲	マッピング	満足
要素	ロード	敵	絵	音楽	システム	満足感
ジョブ	移動	ボス	ドット絵	画面	ボリューム	満足感
プレイ	操作	レベル	背景	音	マップ	ゲーム
難易度	タッチペン	ラスボス	グラフィック	アレンジ	戦闘	要素
ゲーム	表示	全滅	ドット	雰囲気	パズルゲーム	難易度
中毒性	テンボ	クリア	キャラクター	音質	パートナー	内容
ステージ	ストレス	手ごたえ	ポリゴン	名曲	ボクモン	ボス
ストーリー	操作性	ボス戦	表現	サントラ	発想	マップ
対人戦	敵	ミニゲーム	モンスター	マッチ	操作	達成感

事実と主観のアスペクトに基づく情報要約

嶋田和孝・唯野良介・山海匡史・波多野 匡 (九州工業大学)

主観情報のみではなく、客観的なデータも統合して情報を要約する枠組みの提案



まとめ・今後の課題

- 関連情報抽出
 - 抽出データをシードにし、より規模の大きい関連項目辞書の作成
- アスペクト推定
 - 構造情報の利用と教師なし学習の適用
- 主観情報の重要文抽出
 - アスペクト推定との統合
 - 各アスペクトの割合、内容の重複度
- 全体としての要約生成
 - 主観情報と客観情報の統合とそのバランス

評価文のアスペクト推定

- 評価文のアスペクトを推定
- 第1ステップ
 - 機械学習の手法を比較
 - SVM, 最大エントロピー (ME), NaiveBayes, C4.5
 - BOW モデル
- アスペクト: 7 種類
 - 熱中度 (a), 快適度 (c), 難易度 (d), グラフィックス (g), 音楽 (m), オリジナリティ (o), 満足度 (s)
 - マルチアスペクトあり (例) 難しすぎて面白くない
 - 難 (d) -> Negative, 満 (s) -> Negative
 - クラス数: 20 クラス
 - 事例数が 10 未満のものを除く
 - データ数: 4652 文
- 評価尺度: 適合率
 - 完全一致: 全てのアスペクトが正解
 - 部分一致: 1つのアスペクトは正しく推定

学習器	完全一致	部分一致
SVM	62.73	79.79
ME	61.29	79.21
NaiveBayes	58.43	76.00
C4.5	55.27	74.13

コーパス作成ツール

- 学習データなどの必要性
 - アスペクト・評価の効率的なタグ付け

