

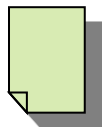
レビュー文中における ネタバレ非表示システムの検討

研究概要

ECサイト等で用いられるレビュー機能では商品の情報を詳しく知ることができ、売り上げ増加や、ユーザ満足度の向上といった効果をもたらすことが確認されている。しかしながら、小説のような物語性を持つ商品のレビューでは、詳細過ぎる情報がかえってユーザの購買意欲を削いでしまう恐れがある。

本研究では、このような「ネタバレ」を機械学習アルゴリズム(Naive Bayes)を用いて判別し、除去する方法について検討する。

レビュー文



文ごとにクラスを推定

文1: 犯人が〇〇で驚いた

文2: 評判どおりの面白さだった

ネタバレ

非ネタバレ

非表示化

Naive Bayes

ベイズの定理に基づいた確率モデル

今回はこのアルゴリズムを基にレビュー文の判別を行った

文が与えられたとき、その文が属するクラスを推定する

$$\hat{c} = \arg \max_c P(c) \prod_{i=1}^n P(w_i | c)$$

ネタバレ文には頻出しやすい語とそうでない語が存在するため単語を素性として学習を行う

固有名詞の汎化

作品名、登場人物などの固有名詞は作品ごとに異なり学習に用いるには不適切である

そのため固有名詞を一般的な語に変換する

レビュー文: そして誰もいなくなったを読んだが、アガサクリスティ最高傑作と言われるだけあって面白かった

汎化後: <作品名>を読んだが、<作者名>最高傑作と言われるだけあって面白かった

ネタバレ文の除去

- ・現在構想段階中
- ・Naive Bayesを基にスコアを算出
- ・基準値を超える文をネタバレとする
- ・基準値はユーザが調節可能とする
- ・ネタバレは黒塗りで非表示化

★★★★☆ Posted by ムロロガ 2015年09月23日

蒼井上鷹は、たまたま本屋で見かけた「九杯目には早すぎる」をきっかけに読み始めた作家。読みやすい文章とヒネリの利いたプロットの作品が多く、かなり好きな作家である。

この「最初に探偵が死ぬ」という部分までヒネリが効いているという秀逸な作品だ。

設定: この設定をベースに、蒼井上鷹らしいヒネリの効いたストーリーが進んでいく作家・星野万丈の莫大な遺産を受け継いだ内野崇也には、血の繋がらない養子達。その中で殺人事件が起きる。

除去方法の例

実際の判別例(成功例)

ネタバレ文として分類

- ・死者の正体は榊原の副担任であり叔母の三神怜子だった。
- ・一貫して被害者の立場であった檜山が一転加害者の立場に立たされる事になります。

非ネタバレ文として分類

- ・読み終わって、いまいちスッキリしなかった。

今後

- ・システムの実装を進める
- ・ネタバレ文以外の情報を活用
 - 意見・感想などのネタバレになりにくい文の活用
 - あらすじのような隠す必要のない内容部分の活用