

# ネタバレを見ずにレビュー文を閲覧できるシステムの開発

## A System Enabling Users See Review Sentences without Seeing Spoilers

池田 郁    土方 嘉徳    岩井 秀成    西田 正吾\*

**Summary.** 多くの e-commerce サイトでは、購買者が商品に対するレビューを書くことができる。レビューには書き手の商品に対する評価や意見が書かれており購買の際の参考になる。しかし、小説や映画などのストーリーを持ったアイテムに対するレビューには、レビューの意見と同時にストーリーの内容（本稿では「あらすじ」と呼ぶ）が書かれている場合がある。レビューによりあらすじがわかってしまうと、実際に小説や映画を見た時の楽しみや感動が減ってしまい問題である。そこで本研究では、ストーリーを持ったアイテムに対して、あらすじを除去してレビューを見られるようにするシステムを開発する。あらすじの判定には4つの代表的な機械学習のアルゴリズムを適用する。また、あらすじ判定に特有のいくつかの特徴量を提案する。あらすじの分類精度に基づいて、提案システムの評価を行う。

### 1 はじめに

近年、多くの e-commerce サイトで、商品に対するレビューを書き込み/閲覧することができる。これらのレビューには、書き手の商品に対する評価や意見が書かれており、購買の際の参考になる。しかし、小説、映画、ゲームなどのストーリーを持ったアイテムに対するレビューには、意思決定に役立つレビューの意見が書かれている一方で、ストーリーの内容が書かれていることがある。例として、ある本に対するレビュー（ただし、本研究では英語のレビューを対象とするので、英語のレビューである）を示す。

#### <レビューの例>

Half-Blood Prince is easily one of the better books in the Harry Potter series. Several of the chapters are particularly well-written, with great suspense and imagery. The end of the book is very sad indeed; yet, I was not crying—I was merely shocked, flabbergasted at the circumstances. Before reading this book, if I had to make a list of impossible things that could never happen... Snape killing the Headmaster and fleeing the school with a bunch of Death Eaters would have been right at the top of the list. But, I'd have been wrong. I had a very strong feeling that Dumbledore would be the one to die in this book. The disturbing ending leaves you frustrated in anticipation of the next book.

(Amazon.com “Harry Potter and the Half-Blood Prince” に対するレビューより)

このレビューから、レビューがこのアイテムを高く評価しているということが分かる。このようなレ

ビューの意見は、このアイテムの購入を考えている人にとって有益である。しかし、このレビューを閲覧した人は、レビューを見るまでには予期していなかった「ダンブルドアの死」というストーリーの重要な部分がわかってしまう。このレビューにより、レビューを閲覧した人は本を読む前に結末を知ってしまい、実際に本を読む時の衝撃や楽しみが減るおそれがある。

そこで本研究では、英語のレビューを対象として、あらすじを隠してレビューを見られるようにするシステムを開発する。ここであらすじとは、ストーリーの内容についての記述を指す。提案システムは、レビュー内のあらすじを隠し、その他の部分をユーザに見せる。隠すかどうかの判断は、文またはフレーズを単位として行うこととする。この判断のため、提案システムは、各文またはフレーズをあらすじか否かに分類する（以下、あらすじ分類）。

本研究の貢献は以下の通りである。

- レビューを対象として、あらすじかどうかの判定に初めて挑戦した。
- あらすじかどうかの判定に4つの代表的な機械学習アルゴリズムを適用した。また、各アルゴリズムに対して、あらすじ分類に対する精度・再現率・F値を用いて評価を行った。
- あらすじかどうかの判定をより良く行うため、レビューの特徴を用いた改善策を提案し、その評価を行った。
- ユーザがあらすじらしさに応じて隠す範囲を調節できるインタフェースを提案した。

本稿の構成は以下の通りである。2章で関連研究、3章で本研究のアプローチを述べる。4章であらすじの判定と表示の方法について述べる。5章で4つ

Copyright is held by the author(s).

\* Kaori Ikeda, Yoshinori Hijikata, Hidenari Iwai and Shogo Nishida, 大阪大学大学院基礎工学研究科システム創成専攻

の機械学習アルゴリズムによるあらすじ分類に対する評価を行う。6章で、機械学習の際の改善策に対する評価を示す。7章で、システムのユーザ経験について述べる。最後に8章でまとめを述べる。

## 2 関連研究

我々の研究に最も近い研究分野として、意見に着目したテキストマイニングが挙げられる。この分野では大きく分けて、意見の収集、意見の分類、意見の分析の3種類の研究が存在する。

意見の収集では、文書のどの箇所に意見が書かれているかを特定し、その抽出を行う。この特定のために、文が主観的か客観的かを判定することが一般的に行われている。PangらやYuらは、レビュー中の各文に対して、機械学習により主観的な文か客観的な文かを判定している [1, 2]。Riloffらは、主観か客観かに分ける分類器を、ラベル付けされていないデータから学習するブートストラップ手法を提案している [3]。

意見の分類では、意見を肯定か否定かに分類する。Daveらは、Naive Bayes を用いてレビューの文を肯定または否定に分類している [4]。Wilsonらは、単語が本来持つ極性（肯定または否定）と文脈での極性の両方を考慮して極性判定を行っている [5]。

意見の分析では、多くの意見を集約し、わかりやすく視覚化している。Morinagaらは、ある商品と他の商品との関連性が理解できるようにレビューの要約を視覚化している [6]。Huらは、アイテムの特徴ごとにレビュー内の意見を要約している [7]。Hijikataらは、ネットオークションの出品者評価コメントから、儀礼的な記述を削除し要約する方式を開発している [8]。

我々は意見ではなくあらすじに着目している。あらすじを対象とした研究は我々の知る限りでは存在しない。意見の分類は、評価表現辞書を用いることで肯定・否定をある程度絞り込むことができる。しかし、あらすじの記述された文では、登場人物に関しての意見や行動も登場する。このため、レビューの意見や行動と登場人物の意見や行動を区別する必要があるが、登場人物の行動に関する辞書は存在しない。よって、あらすじの分類は意見の分類よりも難しい問題であると考えられる。

また、ルールベースではあるが、中村はスポーツの試合結果などを隠ぺいすることができる Web ブラウザを開発している [9]。中村は、あらかじめ特定のチーム名や選手名などでドメインを限定することで、ルールベースによるシンプルなフィルタリング方法を実現している。本研究は、機械学習により、小説や映画など、広いドメインで適用可能である。

我々は、[10]であらすじ発見の手法に機械学習を行うことを提案しているが、精度・再現率を向上させるための工夫の提案や精緻な評価は行っていない。

## 3 アプローチ

レビューを見ていると、あらすじに出やすい語とあらすじに出にくい語があるということに気づく。例として、“kill”や“island”のような語は、レビュー内であらすじの記述のため使用されることがある。だが、これらの語はあらすじ以外の部分にはめったに出現しない。逆に、“think”や“be moved”のような語は、アイテムに対するレビューの感想を表すのに多く使用され、あらすじ部分には出現しにくい。しかし、このような語を網羅的に調べ上げてルール化することは困難であると考えられる。そこで本研究では、単語を素性とした機械学習によってあらすじかどうかの判定を行う。あらすじの判定は、フレーズ単位（句読点で句切られたテキスト部分）で行う。

また、見たくないと感じるあらすじは人によって異なると思われる。あるユーザは「物語の核心さえわからなければ、多少のあらすじは見てもよい」と考えていても、別のユーザは「物語の細かな伏線も知りたくない」と考えているかもしれない。これらのユーザに同様にあらすじを隠してレビューを見せても、両方のユーザの満足は得られないと思われる。そこで、提示インタフェースでは、隠されるあらすじの量をユーザが調節できるようにする。あらすじを隠す際は、よりあらすじらしい部分から隠すことにする。そのため、あらすじ判定のモデル構築には、あらすじらしさを示すスコアや確率を得やすい機械学習アルゴリズムを用いることにする。

## 4 あらすじの判定と実装

機械学習のため、レビューの各フレーズがあらすじかどうかを複数人にラベル付けしてもらい、そして、各フレーズについて過半数がつけたラベルをそのフレーズの正解ラベルとする。このように正解ラベルを付与したレビューをもとに、単語を素性として、レビュー内の各フレーズがあらすじかどうかを分類するモデルを学習する。そしてこのモデルを用いて各フレーズに対しあらすじらしさを表すスコアを与える。機械学習アルゴリズムには、決定木 (C4.5)、Support Vector Machine (SVM)、ロジスティック回帰、Naive Bayes の4つを用いる。これらの中であらすじかどうかの判定に有効なものを使用することとする。

学習されたモデルにより、フレーズごとにあらすじらしさのスコアが算出される。レビューを表示する際には、閾値よりスコアが高いフレーズを隠すこととする。提示インタフェースには閾値設定用のスライダーを設置する。ユーザはスライダーを動かすことで、リアルタイムにあらすじの隠される量を確認することができる。実際の利用では、ユーザはすでに視聴したことがあるアイテムのレビューに対し、スライダーで上記閾値を設定しておき、その閾値を

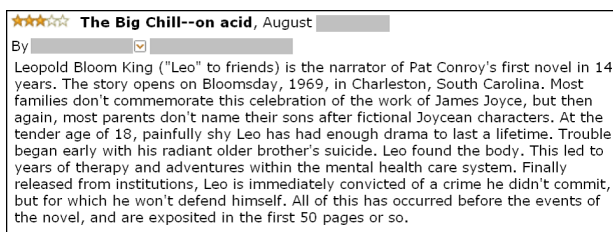


図 1. あらすじ除去前のレビュー

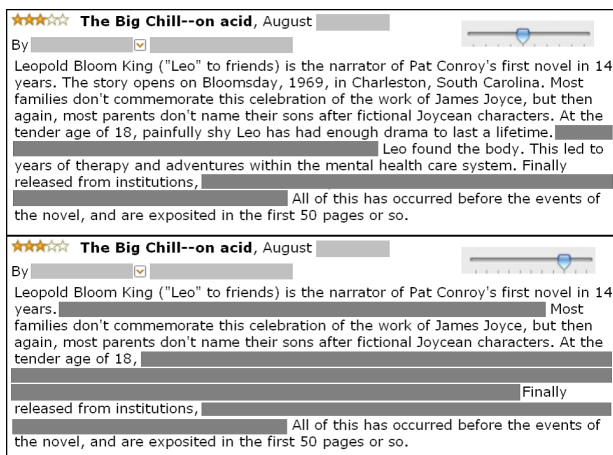


図 2. スライダーであらすじが隠されたレビュー

用いて未知のアイテムのレビューの閲覧を行う。

上述の判定方法と、あらすじ隠ぺいインタフェースを持つシステムを Web アプリケーションとして実装した。ユーザは、Web ブラウザから上記 Web アプリケーションにアクセスする。サーバは Java で実装しており、クライアントは JavaScript が埋め込まれた HTML で実装されている。サーバは、オフラインで e-commerce サイトからレビューをダウンロードし、フレーズにあらすじらしさのスコアを付与している。本実装では、e-commerce サイトとして amazon.com を対象とした。ユーザは、閾値の設定画面にて、あらすじ隠ぺいの閾値を設定する。実際にレビューからあらすじを隠す例を示す。図 1 はあらすじを隠す前のレビューである。図 2 は提案システムがユーザに提示する、あらすじが隠されたレビューである。図 2 では、スライダーでの調節により上の提示例と下の提示例で隠されるあらすじの量が変化しており、下のレビューの方がより多くのあらすじが隠されている。

図 2 の上の提示例では、“Trouble began early with his radiant older brother’s suicide.” や、“Leo is immediately convicted of a crime he didn’t commit” など、兄の自殺からトラブルが始まったことや、関わっていない犯罪の犠牲になったことなど、あらすじが書いてある部分が隠されている。

## 5 あらすじ分類のためのアルゴリズム評価

この章では、前述の 4 つの機械学習アルゴリズムのうち、あらすじかどうかの判定に有効なものはどれかを検証する。

### 5.1 評価用データ

ストーリーを持ったアイテムを持つカテゴリとして小説、DVD、コミックを選択した。そして、Amazon.com から各カテゴリに属するアイテムをランダムに 100 個ずつ選択した。選択された各アイテムにつけられたレビューからランダムに 5 件ずつ選択した。このようにして各カテゴリからそれぞれ 500 件のレビューを取得し、3 人の評価者にラベル付けを行ってもらった。3 人の評価者のうち 2 人以上があらすじであるとラベル付けしたフレーズをあらすじとみなし、評価用データを作成した。取得したレビュー 500 件に含まれるフレーズ数は、小説カテゴリでは 8,513、コミックカテゴリでは 9,118、DVD カテゴリでは 8,588 であった。

### 5.2 評価尺度

評価尺度として、あらすじである部分とあらすじでない部分の判定に対する F 値（以下、分類全体に対する F 値）を用いる。F 値は精度 (Precision)、再現率 (Recall) を組み合わせた指標である。クラス  $C_i$  に対して、F 値は以下のように算出される。なお、本研究ではクラスは 2 つであり、あらすじのクラスとあらすじでないクラスである。

$$Precision(C_i) = \frac{\text{正しく } C_i \text{ に分類されたフレーズ数}}{C_i \text{ に分類されたフレーズ数}} \quad (1)$$

$$Recall(C_i) = \frac{\text{正しく } C_i \text{ に分類されたフレーズ数}}{C_i \text{ に属するフレーズ数}} \quad (2)$$

$$F\text{-Value}(C_i) = \frac{2 * Precision * Recall}{Precision + Recall} \quad (3)$$

分類全体に対する F 値 ( $F\text{-Value}(All)$ ) は、各クラスの F 値の重み付き平均として算出する。それぞれのクラスの重みは、評価用データ内でのそのクラスの存在比率としている。

ユーザが見たくないと思うようなあらすじを含むフレーズを隠すことができなければ、ユーザの満足度は大きく下がると考えられる。このため、あらすじであるフレーズをどれほど多く発見できるかが特に重要である。よって、評価データ中であらすじである部分に対する再現率（以下、あらすじ再現率。  $Recall(plot)$  と表す。）に特に注目して評価を行う。

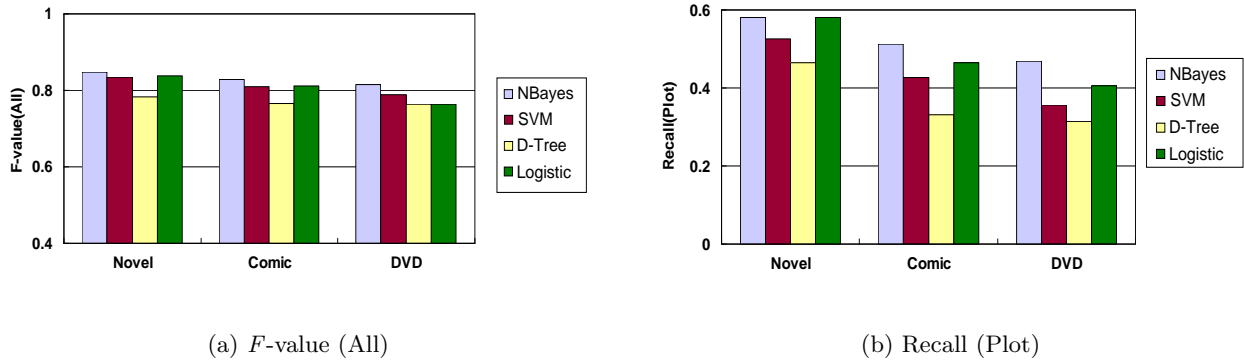


図 3. 4 種類の機械学習アルゴリズムによるあらすじ分類の結果

### 5.3 評価結果

データの平滑化のため、10 分割交差検定法 (ten-fold Cross validation) により評価を行った。なお、本評価では、より良い分類結果を得るために、素性として用いる単語の種類数を変化させて実験を行った。下記に示す結果は、最も評価が良かった場合の種類数（最適な属性数）を用いた結果である。

図 3(a) に F 値を図 3(b) に再現率を示す。‘NBayes’ はナイーブベイズを、‘Logistic’ はロジスティック回帰を、‘D-Tree’ は決定木を示す。図 3(a) より、4 つのアルゴリズムは全てのカテゴリにおいて、約 80 % の F 値を持っていることが分かる。一方、図 3(b) より、あらすじ再現率は最も良いナイーブベイズでも、60 % 弱とあまり高くはない。アルゴリズムの比較をすると、F 値とあらすじ再現率の双方においてナイーブベイズが最も良くなっている。このことより、あらすじ判定にはナイーブベイズが最も良いことが分かる。カテゴリ別には、小説が最も良い結果となっている。

## 6 あらすじ判定のための改善

機械学習の際にあらすじ判定に特有の特徴量を用いることで、あらすじ分類の結果を改善する。

### 6.1 単語の一般化

レビューを調査した結果、登場人物名が現れる文は比較的あらすじであることが多いということがわかっていて、逆に著者名が現れる文は、あらすじになりにくいということがわかっていて、しかし、登場人物名と著者名は多様に存在し、同じ語が使われることは少ない。そのため、学習の際にこれらの語を含むフレーズに高いスコアが与えられず、分類に寄与しなかったことが考えられる。改善策として登場人物名と著者名の一般化を行う。すなわち、“Peter” や “Sally” と言った登場人物名を指す固有名詞は、登場人物名を表す抽象化したタグで表現することにす

る。また、レビューがアイテムに対する感想を述べる場合に、レビュー中でそのアイテム名を使用することが多い。よってアイテム名の一般化も行う。

登場人物名の一般化のため、米国国勢調査局の人名データベース<sup>1</sup>と株式会社アルクの提供するオンライン辞書<sup>2</sup>、また Amazon.com で提供されているアイテムの紹介文とを利用した。アイテムの紹介文には、アイテムの簡単な内容が記述されている。まず、人名データベースに含まれる語から、オンライン辞書で人名以外の語として掲載されている語を除き、残った語を人名辞書として保存した。そして各アイテムの紹介文と人名辞書とを照合し、人名辞書とアイテムの紹介文の両方に出現している語を登場人物名として使用することとした。例として、“Peter finally got married with his rival Sally.” というフレーズに出現する登場人物名を一般化すると、“<character> finally got married with his rival <character>.” というフレーズが得られる。ここで、<character> は単語が登場人物名であることを表すタグである。

著者名とアイテム名も同様に一般化を行う。著者名とアイテム名については、Amazon.com のアイテムの紹介ページ中に決まった形式で記されているため、そのページから抽出して使用した。なお、レビューは、アイテム紹介文に出現しない登場人物名をレビューに書くこともある。そこで、アイテム紹介文に登場する登場人物名以外の人名を、氏名として一般化することも試みた。

### 6.2 意見らしさの利用

筆者が調査を行ったところ、レビュー中では、意見とあらすじは離れた段落に書かれることが多く、1 つのフレーズに意見とあらすじが共存することはほとんどないことがわかった。意見らしさを推定し、そ

<sup>1</sup> <http://www.census.gov>, 1995 年度版

<sup>2</sup> <http://www.alc.co.jp>

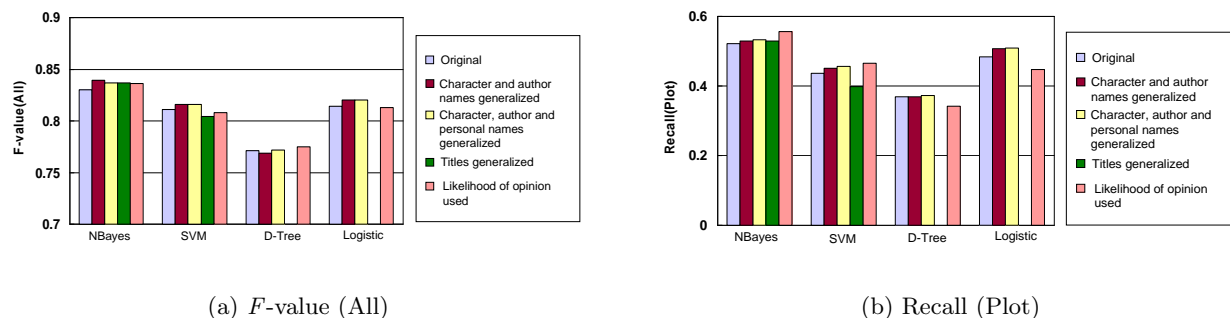


図 4. 改善策を用いたあらすじ分類の評価

の結果を利用することで、あらすじであるフレーズがより良く発見できる可能性がある。意見らしさの利用のため、レビュー中の各フレーズに対して意見であるかどうかを評価者に判断してもらい、評価用データを作成した。この評価用データを用いて Naive Bayes, SVM, 決定木, ロジスティック回帰の4つの機械学習アルゴリズムで学習を行い、レビューの各フレーズに対して意見らしさのスコアを算出した。あらすじらしさについても同様に、上記4つの機械学習アルゴリズムによりあらすじらしさのスコアを算出した。そして、意見らしさのスコアとあらすじらしさのスコアを属性として、ロジスティック回帰で学習を行い、レビューの各フレーズがあらすじかどうかを判定することとした。ただし、この学習の際、属性として使用する意見らしさのスコアとあらすじらしさのスコアは、同一の機械学習アルゴリズムから算出されたものを使用した。

### 6.3 評価結果

各改善策について、最適な属性数を用いた場合の結果を図4(a)と図4(b)に示す。ここで、‘original’は前章の方法の結果である。タイトル名を一般化したものは、決定木とロジスティック回帰においては、高い情報量を持つものがなかった。そのため、タイトル名の一般化は、決定木とロジスティック回帰では用いられなかった。人名を一般化したもの(登場人物名と著者名を一般化したものと、登場人物名と著者名と氏名を一般化したもの)のF値とあらすじ再現率は、決定木を除いて、オリジナルの方法よりも良かった。人名を一般化することは、あらすじ判定に有効であることが分かる。タイトル名の一般化は、ナイーブベイズではF値もあらすじ再現率も向上しているが、SVMでは両方とも悪化している。そのため、タイトル名の一般化は、ナイーブベイズでのみあらすじ判定に有効であることが分かった。

意見らしさを用いた場合は、ナイーブベイズでは、F値もあらすじ再現率も向上している。SVMでは、あらすじ再現率は向上しているもののF値は下がっ

ている。決定木では、F値は向上しているものの、あらすじ再現率は悪化している。ロジスティック回帰では、F値とあらすじ再現率の両方で悪化している。これらのことから、意見らしさの利用はナイーブベイズでのみ有効であることが分かる。

## 7 ユーザ経験

提案システムを3人のユーザに使ってもらった。ここでは、閾値設定画面でサンプルのレビューを用いてスライダーで閾値を設定し、その後数個のレビューを閲覧してもらった。利用後にインタビューしたところ、スライダーで閾値を調節できることについてはポジティブな評価を得た。特に、見たくないフレーズにだけ着目しておき、それが隠されるところまでスライダーを動かして設定するのが、便利であったとのことであった。一方、設定された閾値で他のレビューを閲覧すると、あらすじが隠されずに残されているのが目につくとの意見もあった。それが本人にとって致命的なネタバレになっているのかどうかを知るには、より多くのレビューを用いて実験する必要がある。

## 8 おわりに

本研究では、ストーリーを持ったアイテムに対するレビューについて、あらすじを見ずにレビューが見られるシステムを開発した。あらすじの判定には4つの代表的な機械学習アルゴリズム(Naive Bayes, サポートベクターマシン, 決定木, ロジスティック回帰)を適用し、フレーズに対してあらすじらしさのスコア付けを行うことを考えた。また、付けられたスコアをもとに、ユーザがどの程度あらすじを隠すかをスライダーで調節できるようなインターフェースを開発した。あらすじとあらすじでない部分の判定に対するF値(およびあらすじに対する再現率)により評価したところ、Naive Bayesが最も良い結果を示した。また、あらすじ判定のための改善策として、登場人物名、著者名、氏名の一般化を行う

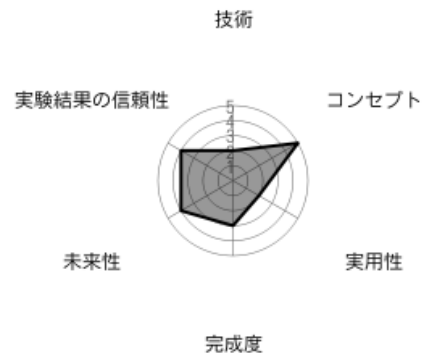
ことが有効である可能性が示された。今後は、実際に提案システムを用いた被験者実験を行い、あらすじを隠してレビューを見せる本システムが購入の意思決定の際に有効に利用できるかを検証する予定である。

## 参考文献

- [1] Pang, B., Lee, L.: A Sentimental Education: Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts. *Proc. of ACL'04*, pp. 271-278 (2004).
- [2] Yu, H., Hatzivassiloglou, V.: Towards Answering Opinion Questions: Separating Facts from opinions and Identifying the Polarity of Opinion Sentences. *Proc. of EMNLP'03*, pp.129-136 (2003).
- [3] Riloff, E., Wiebe, J.: Learning Extraction Patterns for Subjective Expressions. *Proc. of EMNLP'03*, pp. 25-32 (2003).
- [4] Dave, K., Lawrence, S., Pennock, D.M.: Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. *Proc. of WWW'03*, pp. 519-528 (2003).
- [5] Wilson, T., Wiebe, J., Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis. *Proc. of HLT/EMNLP'05*, pp. 347-354 (2005).
- [6] Morinaga, S., Yamanishi, K., Tateishi, K., Fukushima, T.: Mining Product Reputations on the Web. *Proc. of KDD'02*, pp. 341-349 (2002).
- [7] Hu, M., Liu, B.: Mining and Summarizing Customer Reviews. *Proc. of KDD'04*, pp. 168-177 (2004).

- [8] Hijikata, Y., Ohno, H., Kusumura, Y., Nishida, S.: Social Summarization of Text Feedback for Online Auctions and Interactive Presentation of the Summary. *Knowledge-Based Systems*, Vol. 20, No. 6, pp. 527-541, 2007.
- [9] 中村聡史：ネタバレ防止ブラウザの実現，WISS 2010，2010.
- [10] Ikeda, K. and Hijikata, Y. and Nishida, S.: Proposal of Deleting Plots from the Reviews to the Items with Stories. *Proc. of SNSMW'10*, 2010.

## アピールチャート



## 未来ビジョン

何かモノを買う時に、ユーザレビューを見ることは当たり前になっている。人によっては、カタログも見ずに、ユーザレビューだけで、購入するか否かの意思決定をしている。しかし、映画や小説、漫画と言ったストーリーを持つアイテムのレビューを見ていると、時にネタバレを含んでいることがある。特に、ストーリーの結末について書かれていることもあり、それを見てしまうと、購入する意欲が一気になくなってしまふ人もいるだろう。

本研究は、そのようなネタバレを含む可能性のある、ストーリーの内容に関する記述（あらすじ）をフレーズ単位で検出するものである。提案システムでは、発見されたあらすじを隠して、レビューを読むことができる。あらすじを見て許容できる範囲は人によって異なる可能性もあるため、あらすじらしさを表すスコ

アによって、調節することができる。

自然言語処理の研究分野において、あらすじの判定に挑戦したのは、本研究が初めてである。主観的文書の判定や、その中での肯定的意見や否定的意見の判定は、よく行われてきた。これらの判定では、単語の極性（肯定的／否定的）を表す辞書が使える。しかし、あらすじ判定用の辞書は存在せず、自然言語処理の研究分野においてあらすじ判定は非常に難しい研究課題と言える。

本研究は機械学習を用いているが、機械学習に完璧なものではなく、あらすじを隠そうとすると、意見まで隠してしまうこともある。逆に、致命的なあらすじが残されたままになることもある。このような根本的な問題を内在しつつも、取り組むべき課題なのかどうか議論できればと思う。そして、どういう形で、このような機能をユーザに提供すべきなのか議論できればと思う。