

# レビュー文を対象としたあらすじ分類手法の提案と あらすじ非表示システムの開発

岩井 秀成<sup>1,a)</sup> 池田 郁<sup>1</sup> 土方 嘉徳<sup>1,b)</sup> 西田 正吾<sup>1</sup>

**概要:** Amazon.com のようなショッピングサイトでは、アイテムに対してレビューを簡単に作成・閲覧できる機能が提供されている。レビューに書かれた他者の意見は有用であるが、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、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分かると、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、レビュー文を対象としたあらすじ分類手法を提案する。あらすじの判定には5つの代表的な機械学習アルゴリズムを適用した。そして、人名と特有語の一般化による分類性能の向上を図り、F 値を用いて評価を行った。また、あらすじを非表示にするシステムを開発した。最後に、被験者実験によりシステムの評価を行った。

## Proposal of a Plot Classification Method for Reviews and Development of a Plot Hiding System

HIDENARI IWAI<sup>1,a)</sup> KAORI IKEDA<sup>1</sup> YOSHINORI HIJIKATA<sup>1,b)</sup> SHOGO NISHIDA<sup>1</sup>

**Abstract:** Users can write and read reviews of items easily on many Internet shopping sites. Reviews on items including stories like novels and movies contain spoilers (undesired plot descriptions) along with the opinions of the review author. In this paper, we propose a system that helps users see reviews without seeing plot descriptions. This system classifies each sentence in user reviews to plot or non-plot and hides plot descriptions from user reviews. We tested five common machine-learning algorithms and our improving methods for plot classification. Finally, we conducted a user study on how much spoilers are deleted and how much useful comments are remained when using our proposed system.

### 1. はじめに

近年、インターネットでのショッピングが盛んになりつつある。ある商品やコンテンツ（以降、アイテム）を購入する際、多くの人は情報収集を行う。そして得られた情報をもとに、購入するかどうかの意思決定を行う。Amazon.com や eBay のようなショッピングサイトは、アイテムに対してレビューを簡単に作成・閲覧できる機能を提供している。これらのサイトでは、実際にアイテムを購入した多くの人の意見を知ることができるため、ユーザにとって有益

である。

しかし、コミック、小説、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、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。例として次のレビューを挙げる。（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）

〈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. After completing this book, I was in a state of total shock. 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*

<sup>1</sup> 大阪大学大学院基礎工学研究科，豊中市  
Graduate School of Engineering Science, Osaka University,  
1-3 Toyonaka-shi, 560-8531, Japan

a) iwai@nihilab.sys.es.osaka-u.ac.jp

b) hijikata@sys.es.osaka-u.ac.jp

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. However, I never saw it coming the way it happened. 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)

このレビューからは、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見てとれる。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を知ってしまう。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。株式会社社会情報サービスが運営している、様々なアンケートをインターネット上で行うサイト\*1では、アイテムのネタバレに関するアンケートを行っている。そのアンケートには、「レビューでストーリーの結末が書かれていてショックだった」、「刑事映画の犯人を知ってしまい落胆した」などの意見が寄せられている。また、半数以上の人がレビューによってストーリーの内容を知りたくないと答えている。

そこで本研究では、英語のレビューを対象として、ストーリーの内容を表す部分であるあらすじを検出する手法を提案する。具体的には、レビューを文単位で、あらすじか否かに分類（以下、あらすじ分類）する手法を提案する。あらすじ部分を特定することで、レビューに書かれたストーリーの重要な部分を非表示にするシステムを開発することを目的とする。本研究の貢献は以下の通りである。

- レビュー文を対象として、あらすじ分類に初めて挑戦した。
- あらすじ分類に5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F値を用いて評価を行った。
- あらすじ分類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。
- あらすじを非表示にするシステムを開発し、被験者実験により評価を行った。

本稿の構成は以下の通りである。2章で関連研究を述べる。3章で5つ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。4章でレビューの特徴を用いた改善策について述べ、その評価を行う。5章であらすじを非表示にするシステムの実装について述べる。6章で被験者実験によりシステムの評価を行う。最後に7章でまとめを述べる。

\*1 <http://www.enquete.ne.jp/hundred/>

## 2. 関連研究

インターネット上のレビュー文に対する研究はテキストマイニングの分野で広く行われている。中でも、我々の研究に最も関連する分野は文書内の意見に着目した研究分野である。Pangら [1] は、レビュー中の各文に対してSVM, Naive Bayesを用いて主観的な文か客観的な文かを判定することで、どの箇所意見が書かれているか特定している。Yuら [2] は、上記の判定を文単位と文書単位の両方でできるようにしている。Riloffら [3] は、上記判定のためにラベル付けされていないデータから学習するブーストラップ手法を提案している。

レビューが否定的な文か肯定的な文かを判定する研究もある。Daveら [4] は、Naive Bayesを用いてレビュー文を肯定的か否定的かに分類している。Wilsonら [5] は、単語と文脈の両方の極性を考慮し、フレーズごとに分類している。

我々の研究は、本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を対象とし、レビューからあらすじを見つけ出すことを目的としている。本研究は、レビューから意見を見つけ出して極性を判定する研究と類似している。しかし、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存在しない。意見は、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。レビューからあらすじに関する記述を発見するには、レビューの行動と登場人物の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、レビューからあらすじを見つけ出すことは、意見を見つけ出すことよりも難しい問題であると考えられる。

また、レビューの効率的な閲覧を支援する研究もある。Titovら [6] は、レビューからアイテムの属性を発見するトピックモデルを提案している。Huら [7] は、重さやバッテリーの持ちの良さなどのアイテムの特徴ごとにレビュー内の意見を要約する方法を提案している。Huangら [8] は、レストランの属性を学習し、スマートフォン上で推薦と簡潔な意見表示を行うインタフェースを開発している。Hijikataら [9] 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する方法を提案している。

中村ら [10] は、スポーツの試合結果に関する記述を非表示にする方法として、該当箇所を消去する、黒塗りにするなどの方法を提案し、検証している。

我々の研究では、文中からあらすじに関する記述を発見し、非表示にして提示することで、ストーリーに関する記述を読みたくないユーザでもレビューを閲覧できるシステムを開発する。

なお、本論文は [11] で発表された成果を発展させたものである。[11] では単純に単語ベクトルから機械学習したもの

のでシステムを実装しているが、本論文ではより詳細な機械学習アルゴリズムの調査を行い、人名と特有語の一般化、詳細な分類性能の評価、被験者実験を追加している。

### 3. 機械学習によるあらすじ分類

この章では、あらすじ分類を行うための基本となる手法について説明する。また、その基本手法に対して、あらすじ分類の正確性を評価する。

#### 3.1 アプローチ

レビューからどのようにあらすじを特定するのかを考える必要がある。レビューには、あらすじに出やすい語とあらすじに出にくい語がある。例として、“kill”や“friend”のような語は、あらすじの記述のために多く使用され、あらすじ以外には使用されにくい。逆に、“think”や“read”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、あらすじには使用されにくい。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、単語を素性として機械学習を行う。そして学習されたモデルを用いて各文にスコア付けを行い、あらすじであるかどうかを判定する。機械学習を行うために、我々は文  $p$  を bag-of-words にて表現する。すなわち、文  $p$  は、単語  $w_m$  により、 $p = \langle w_1, w_2, \dots, w_M \rangle$  のように表現される。また、文  $p_n$  中で各単語  $w_m$  が出現した回数  $x_{n,m}$  も記録しておく。

機械学習アルゴリズムは、学習されたモデルによってあらすじらしさを示すスコアを得やすいものを選択する。具体的には、Naive Bayes (NBayes) [12], Support Vector Machine (SVM) [13], ロジスティック回帰 (Logistic) [14], 決定木 (D-tree, 具体的には C4.5 を使用) [15], k-nearest neighbor 法 (k-NN) [16] の 5 つを用いる。

#### 3.2 正解データ

評価者にレビュー内の各文があらすじであるかどうかの判断を行ってもらい、正解データを作成した。我々はまず、ストーリーを持ったアイテムを持つカテゴリとしてコミック、小説、DVD 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た。3 人の評価者にラベル付けを行ってもらった。各評価者は、各文があらすじであるか否かを判定する。3 人の評価者のうち 2 人以上があらすじであると判定した文をあらすじとみなし、正解データを作成した。表 1 に、属性数 (出現単語数)、出現回数が 1 回のものを除いた属性数、あらすじである文の数と、あらすじでない文の数を示す。なお、3 人の評価者の

表 1 各カテゴリの正解データの詳細

Table 1 Statistics of ground truth

	コミック	小説	DVD
属性数	6414	6334	6966
属性数 (出現回数 2 回以上)	3603	3539	3761
あらすじである文の数	1523	1602	1357
あらすじでない文の数	3484	3225	3445
意見である文の数	2770	2567	3065
意見でない文の数	2237	2260	1737

判定の一致の程度を調べるために Fleiss の  $\kappa$  係数 [17] を計算した。すると、コミックのカテゴリで 0.612, 小説のカテゴリで 0.544, DVD のカテゴリで 0.466 となった。

#### 3.3 評価尺度

評価尺度として、あらすじである文の判定に対する F 値を用いる。F 値 (F-Value) は精度 (Precision) と再現率 (Recall) の調和平均により算出される指標であり、情報検索の分野で広く用いられている [18]。

#### 3.4 学習条件

データの平滑化のため、10 分割交差検定法 (ten-fold cross validation) [19] で評価を行った。なお、本データセットは 2 つのクラスの出現頻度が異なる不均衡データであるため、学習時にはダウンサンプリングの手法により均衡データ化している。また一般に、機械学習の性能は用いる属性数に依存する。そこで、属性数を 10 個から 2000 個、そして属性数の最大値まで変化させて評価を行った。属性数は、100 個までは 20 個おきに、100-1000 個までは 200 個おきに、1000 個以上は 2000 個おきに取らる属性数の最大値まで変化させた。なお、ロジスティック回帰は計算に時間がかかり、属性数の最大値まで取ると計算が終了しなかったため、属性数 2000 個までを実行した。この際、学習に用いる属性は、Glover らの手法 [20] に従い、あらすじのラベルとの相互情報量の高い順に選択した。また、データセット中、出現回数が 1 回しかないものは、その相互情報量によらず、属性としては採用しないことにした。そのため、属性数 4000 個の前に属性数の最大値に達する (表 1 参照)。

#### 3.5 ベースライン

本評価の目的は、あらすじ分類に適した学習アルゴリズムが何かを明らかにすることである。しかし、ランダムな判定方法と、従来の意見分類をあらすじ分類に応用した方法と比べて、各学習アルゴリズムの結果がどれだけ良いのか知っておくことは重要である。意見分類とは、レビュー文がレビューの意見が書かれたものであるか否かを分類するものである。

ランダムであらすじかどうかを決めたときの F 値を示す。ランダムな判定方法はあらすじか否かを 50% ずつの確

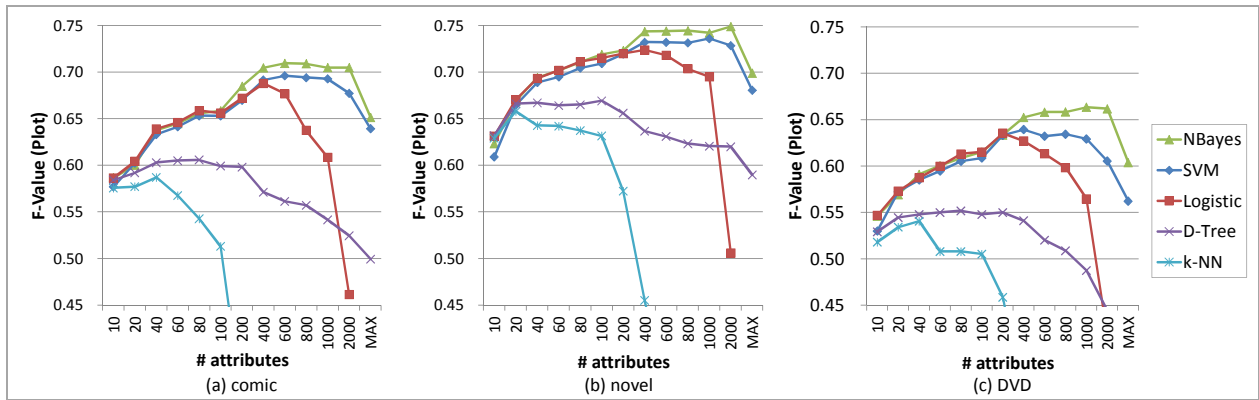


図 1 各アルゴリズムの F 値  
Fig. 1 F-value of each algorithm

率で決定するものである。この F 値の結果は、コミックカテゴリで 0.378、小説カテゴリで 0.399、DVD カテゴリで 0.361 となった。

また、従来の意見分類をあらすじ分類に応用した方法は、レビューの各文を意見分類し、意見と判定された文をあらすじでないとみなし、意見と判定されなかった文をあらすじとみなすものである。本研究で行うあらすじ分類の方法と同じように(すなわち機械学習にて)意見分類を行った。機械学習で意見分類を行うには、各文に意見か否かを判定した正解データを与えておく必要がある。あらすじ分類用の評価データを作成した場合と同様に、同じレビューデータに対して、同じ評価者を用いてラベル付けを行った。3 人の評価者は、各文が意見であるか否かを判定する。表 1 に意見である文の数、意見でない文の数を示す。Fleiss の  $\kappa$  係数 [17] を計算したところ、コミックカテゴリで 0.531、小説カテゴリで 0.551、DVD カテゴリで 0.463 となった。

### 3.6 評価結果

図 1—(a), (b), (c) にそれぞれ、コミック、小説、DVD の各カテゴリの結果を示す。k-NN では、k を 1, 3, ..., 39, 41 と試したが、最も結果の良かった 19 (コミック)、5 (小説)、23 (DVD) の時の結果を示す。いずれのカテゴリにおいても属性数が 200 程度までは、NBayes, SVM, Logistic の 3 つの手法間には差はほとんど見られない。ただし、400 を超えてくると、NBayes がやや優勢になり、差が開いてくることが分かる。最も高くなった時の値で比較すると、いずれのカテゴリにおいても、NBayes, SVM, Logistic の順で良いことが分かる。一方、k-NN と D-tree は、上記 3 手法よりもかなり低い値になっている。

機械学習においては、一般的にデータが高次元になると、次元の呪いにより汎化性能が低下する傾向がある。しかし、NBayes と SVM は、次元の呪いを緩和する特徴を持つため、高次元になった時に F 値が低下するタイミングを遅らせ、またその程度が低くなっているのだと思われる。

表 2 意見分類による方法との比較 (最適属性数)

Table 2 Method using opinion classification vs. method using plot classification

カテゴリ	アルゴリズム	あらすじ分類による方法	意見分類による方法
コミック	NBayes	<b>0.709</b>	0.646
	SVM	<b>0.696</b>	0.586
	Logistic	<b>0.688</b>	0.583
小説	NBayes	<b>0.749</b>	0.695
	SVM	<b>0.736</b>	0.627
	Logistic	<b>0.724</b>	0.621
DVD	NBayes	<b>0.663</b>	0.599
	SVM	<b>0.639</b>	0.536
	Logistic	<b>0.636</b>	0.538

各カテゴリ及び各アルゴリズムで F 値が最も高くなった属性数(以降、最適属性数)の時に t 検定を行い、アルゴリズム間で有意差があるか否かを検証した。その結果、コミック、小説、DVD の各カテゴリにおいて、NBayes は、SVM, Logistic, k-NN, D-tree のそれぞれの手法との間に有意差が見られた ( $p < 0.05$ )。

また、それぞれの最適属性数の時に、意見分類による方法と比較した表を表 2 に示す。意見分類による方法も、用いる属性数を 100 個までは 20 個おきに、100–1000 個までは 200 個おきに、1000 個以上は 2000 個おきに取る属性数の最大値まで変化させて調べた。属性は意見のラベルとの相互情報量の高い順に選択した。この結果から、意見か否かの正解データから意見分類を行い、意見と判定された時にあらすじではないとみなす方法は、あらすじか否かの正解データから直接にあらすじ分類を行う方法に比べて、F 値が低いことが分かる。各カテゴリ、各アルゴリズムにおいて t 検定を行ったところ、あらすじ分類の方法と意見分類の方法の間に、統計的有意差が見られた ( $p < 0.01$ )。

これらの結果をまとめると、NBayes は、最適属性数の F 値が最も高く、なおかつ他の全てのアルゴリズムに対して、統計的有意差が見られた。また、NBayes は、学習に

用いる属性数が増えたと他手法よりも安定して良い値を示す傾向にあった。このことから、あらかじめ分類に最も適したアルゴリズムはNBayesであることが分かる。また、あらかじめ正解データから機械学習する方法は、意見分類による方法よりも良い結果となることが分かった。

## 4. 人名と特有語の一般化

この章では、機械学習の際にレビューの特徴を用いた工夫を行い、あらかじめ分類の結果を改善することを考える。我々は改善策として、人名と特有語の一般化を提案する。

### 4.1 方法

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かじめ分類されることが多く、逆に著者名が現れる文は、あらかじめ分類されにくいということが分かった。しかし、登場人物名と著者名は多様に存在し、それぞれの語の出現回数は少ないため、学習に向かない。

改善策として登場人物名、著者名の一般化を行う。すなわち、登場人物名（または著者名）を指す固有名詞は、登場人物名（または著者名）を表す抽象化したタグで表現することにしよう。例として、“Peter finally got married with his rival Sally.” という文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).” という文が得られる。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。著者名については (author) というタグで表現する。なお、DVD カテゴリでは俳優名も獲得できるので、俳優名についても一般化する。また、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、上記以外の人名を指す固有名詞についても、その他人名として一般化する。

人名の一般化のため、米国国勢調査局の人名データベース<sup>\*2</sup>と株式会社アルクの提供するオンライン辞書<sup>\*3</sup>、また、Amazon.com で提供されているアイテムの紹介文を利用した。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。著者名と俳優名については、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。

上記の方法で、一般的な名前（主としてヨーロッパ系の名前）は抽出して一般化できると考えられる。しかし、アジアやアフリカなど、その他の地域の名前はカバー

されていない。また、コミックにおいては、現実には使われていない名前も登場する（例えば、「ドラえもん」や「ポニョ」など）。このような名前は、辞書が整備されていない上に、次々に新しい名前が誕生するため、対応することが難しい。そこで、あるアイテムへのレビュー文にしか出てこない語を、そのアイテム特有の単語として一般化することにした。このようなアイテム特有の単語には、上記のような特有の登場人物名が含まれるかもしれない。3.2 節で用いたデータ（全 100 アイテムに対するレビュー）において、複数のアイテムに出てきた単語を非特有語としてリスト（非特有語リストと呼ぶ）に登録した。次に、レビュー中の各文において、それぞれの単語を調べ、非特有語リストに登録されていなければ、特有語として一般化した。

### 4.2 評価

3 章の学習アルゴリズム間の比較実験の時と同じデータセットを用いて評価した。また、3.4 節で示した方法で、属性数を変化させて、単語の一般化なしの場合（“ORG” と表記）、人名（登場人物名、著者名、俳優名とその他人名）を一般化した場合（“CAN” と表記）、特有語を一般化した場合（“PEC” と表記）、人名と特有語を一般化した場合（“CAN+PEC” と表記）の 4 つを比較した。分類性能の良かった NBayes, SVM, Logistic に対して、単語の一般化の有効性を検証した。それぞれの最適属性数の時の結果を図 2 に示す。

まず、人名の一般化 (CAN) に注目する。コミックと小説のカテゴリにおいては、ORG よりも CAN の方が F 値が高いことが分かる。また、コミックと小説を比較すると、小説の方が一般化の効果が大きいことが分かる。これは、小説の方が一般的な名前の登場人物が多いためと思われる。コミックは、辞書に載っていない名前の影響で、人名の一般化の効果があまり高くなかったと思われる。コミックと小説に対して、DVD は一般化の効果があまり見られない。その理由として、DVD の商品紹介文には非常に短いものが多く、登場人物名がうまく抽出できなかったことが挙げられる。そのかわりに、DVD においては俳優名が登場人物名としてあらかじめ現れることが多く、俳優名タグが登場人物名タグと同じような働きをしている。しかし、俳優名を含む文には俳優自体に対する批評もあるため、俳優名を一般化することによる効果は、コミックや小説において登場人物名を一般化することによる効果に比べて少なかったといえる。

次に、特有語の一般化 (PEC) に注目する (図 2 参照)。小説カテゴリでは、PEC は ORG よりも F 値が高いが、CAN とほとんど差が見られない。一方、コミックカテゴリにおいては、PEC は CAN よりも F 値が高い。このことから、小説カテゴリでは人名の一般化が十分に機能していたが、コミックでは人名と認識されていなかった人名が

\*2 <http://www.census.gov>, 1995 年度版

\*3 <http://www.alc.co.jp>

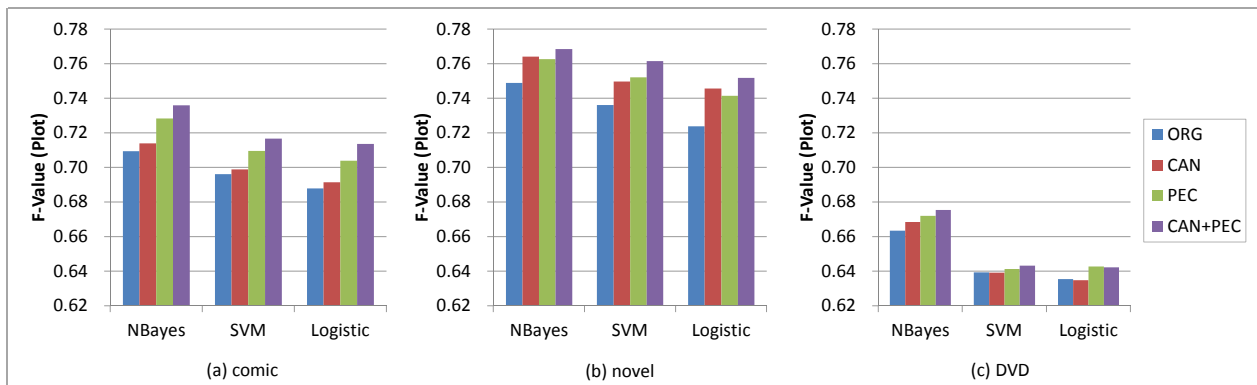


図 2 一般化における F 値 (最適属性数)

Fig. 2 F-value for each algorithm with generalization (under the optimal number of attributes)

多くあったと思われる。実際に特有語として一般化された単語を詳しく調べると、Negi, Husky, Cooro, Anima, Grumpy など、人名辞書には存在しないキャラクター名が多く一般化されていた。また、人名の一般化と特有語の一般化の両方を行った CAN+PEC では、いずれのカテゴリにおいても最も良くなった。

NBayesにおいて、人名と特有語の一般化による F 値の向上に統計的有意差があるかどうかを、t 検定を用いて検証した。その結果、コミックカテゴリでは ORG-PEC 間, ORG-CAN+PEC 間, CAN-CAN+PEC 間, PEC-CAN+PEC 間, 小説カテゴリでは ORG-CAN 間, ORG-CAN+PEC 間に統計的有意差が得られた ( $p < 0.01$ )。

これらの結果をまとめると、人名の一般化は、小説カテゴリで有効であることが分かった。特有語の一般化は、コミックカテゴリで有効であることが分かった。人名と特有語の両方を一般化した場合は、全てのカテゴリにおいて、最も F 値が良くなることが分かった。特に、コミックカテゴリと小説カテゴリで有効であることが分かった。

## 5. あらすじ非表示システムの実装

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実装した。提案システムは、Web アプリケーションとして実装した。サーバサイドプログラムは Java サーブレットで実装しており、クライアントサイドプログラムは JavaScript が埋め込まれた HTML で実装されている。サーバサイドプログラムは、保存されている e-commerce サイトのレビューに対して、各文を (span) タグで括り、属性値にあらすじらしさのスコアを埋め込んでいる。本実装では、e-commerce サイトとして Amazon.com を対象とした。あらすじらしさのスコアの算出には、人名と特有語の一般化を行った NBayes のアルゴリズムを用いている。

学習されたモデルにより、文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は、閾値よりスコアが高い文をクライアントサイドプログラムで隠すこと

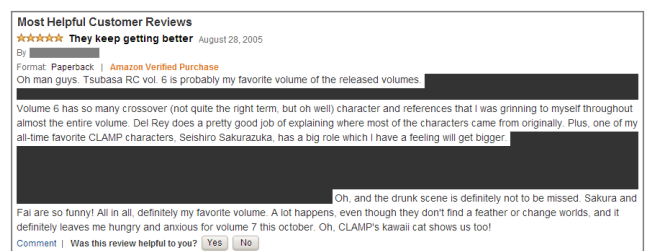


図 3 提案手法を組み込んだ Amazon.com のレビュー

Fig. 3 Reviews displayed by our proposed system in Amazon.com

とする。閾値を変化させることで遮蔽量を変化させることができる。隠し方についてはあらすじと推定された文を消去したり、黒く塗りつぶしたりすることにより非表示にすることが考えられる。黒く塗りつぶした場合、提案システムでレビューは図 3 のように表示される。

## 6. 被験者実験

### 6.1 実験の目的

本論文では、ストーリーに関する重要な記述でユーザができれば見たくないものをネタバレと定義する。4 章にて、提案システムがあらすじを除去できることは示したが、ネタバレを除去できなければ有用とはいえない。そこで本章では、どれだけネタバレとなる文を除去できるかを検証する。おそらく多くのユーザは、高いネタバレ遮蔽率で、レビュー文を読みたいと思われる。しかし 100%近いネタバレ遮蔽率でレビュー文を非表示にすると、ほとんどレビュー文が残らなくなることも考えられる。そこで、なるべく多くのレビュー文を表示しつつ (すなわち高いレビュー表示率で)、多くのネタバレとなる文を除去できることが重要である。

また、レビューを見た際に、ネタバレとなる文が遮蔽されているだけでなく、商品購入の意思決定において参考になる文 (以降、有用な文) が表示されている方がよい。さらに、一部非表示になっていたとしても、レビューの文脈

が崩れておらず、スムーズに読める方が良い。すなわち、システムは高いネタバレ遮蔽率を、より高いレビュー表示率で、またより高い有用な文の表示率（以降、有用文表示率）で実現し、なおかつレビューの文脈が維持されているほうが良い。

本章では、あるネタバレ遮蔽率を達成した時のレビュー表示率、またその時の有用文表示率に関して、従来の意見分類を応用した手法、ランダムで判断する手法と比較する。特に、高いネタバレ遮蔽率において、どの程度これらに差があるかを調べる。また、レビュー文を非表示にしたときに、どの程度文脈が保たれているかについて調査する。

なお、どの程度ストーリーの詳細さについて書かれているとネタバレと感じるかは、人によって異なると思われる。そこで、これまでのような多数決で判定したラベルによるデータセットで評価を行うのではなく、ユーザごとにどれだけネタバレが遮蔽できるかを調べる被験者実験で評価を行う。

## 6.2 ネットバレである文と有用な文の判定

ネタバレ遮蔽率に対するレビュー表示率、有用文表示率の関係を調査し、あらすじ分類、意見分類、ランダム分類の性能比較を行う。

被験者実験に利用するデータセットは、最もあらすじ分類が有用に働いた小説カテゴリを用いる。3章に用いた小説カテゴリの100個のアイテムからランダムに10個のアイテムを選択した。被験者は7人とし、TOEICスコアの平均は753.6点である。被験者は、レビューの全文を見て、各文がネタバレかどうかと有用な文かどうかを判定する。各被験者の判定それぞれを評価データとみなし、全被験者の平均値を評価値とする。

評価する分類手法は、あらすじ分類、意見分類、ランダム分類の3つである。あらすじ分類は、NBayesのアルゴリズムを用い、人名と特有語の一般化を行って付与したスコアをもとに分類し、ネタバレ遮蔽率を変化させて評価する。意見分類は、3.5節の方法で求めたスコアをもとに分類し、ネタバレ遮蔽率を変化させて評価する。ランダム分類は遮蔽するか否かを一定確率で決定し、その確率を変化させて評価する。

ネタバレを遮蔽した際のレビュー表示率を図4に示す。70%以上の高いネタバレ遮蔽率の時、あらすじ分類の方が意見分類より高いレビュー表示率となっている。これは、ユーザが高いネタバレ遮蔽率を望んだ場合、より多くのレビュー文が表示できているあらすじ分類の方が優れていることを示している。95%から急激にレビュー表示率が減少することから、ユーザはネタバレ遮蔽率を70%~90%の範囲で設定するとシステムがより効果的に機能するといえる。

ネタバレを遮蔽した際の有用文表示率を図5に示す。ネタバレ遮蔽率にかかわらず、あらすじ分類と意見分類の有用

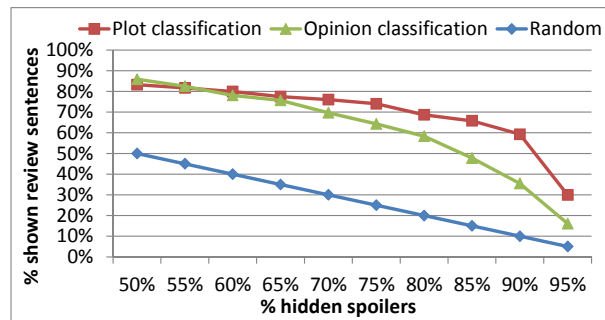


図4 ネットバレを遮蔽した際のレビュー表示率  
Fig. 4 Ratio of shown review sentences

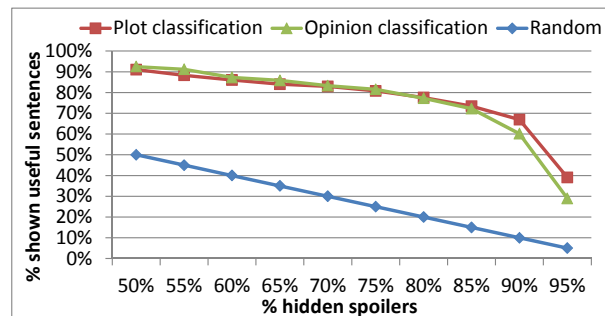


図5 ネットバレを遮蔽した際の有用文表示率  
Fig. 5 Ratio of shown useful sentences

用文表示率はほぼ同じ曲線を描いている。これはすなわち、同じだけネタバレを消したときに、あらすじ分類と意見分類はほぼ同数の有用な文を残せることを示している。95%から急激に有用文表示率が減少することから、ユーザはネタバレ遮蔽率を90%までの範囲で設定するとシステムがより効果的に機能するといえる。

以上より、あらすじ分類は意見分類よりネタバレの遮蔽に対し有効に働くことが分かった。また、ネタバレ遮蔽率を70%~90%の範囲で設定するとシステムがより効果的に機能すると分かり、その時あらすじ非表示システムは、高いネタバレ遮蔽率と高いレビュー表示率と高い有用文表示率を同時に実現できていることが分かった。

## 6.3 あらすじを非表示にした際のレビューの文脈

あらすじ非表示システムを用いてレビューを見た際に、どれだけ文脈が保たれているかを調査する。

被験者実験に用いるデータは、小説カテゴリのうち文の合計が50から60であるアイテムの中から、6アイテムをランダムに選択した。この6アイテムを2つのデータセットに分割し、3アイテムずつにしたものをそれぞれ Dataset A, Dataset B と呼ぶことにする。文の合計を50から60に限定した理由は、データセット間の文の数による違いを抑えるためである。また、データセットの全てのレビューはもともと文脈の崩れた部分がないことを3人により確認している。

被験者は6人とし、TOEICスコアの平均は753.3点で

ある。被験者のうち3人はあらすじ分類で Dataset A, ランダム分類で Dataset B を評価し, 残りの3人があらすじ分類で Dataset B, ランダム分類で Dataset A を評価する。被験者は, レビューをどの手法で見ているかは知らされない。被験者は, あらすじ非表示システムを通して読んだレビューで, 文脈の崩れている箇所を判定する。図3ではあらすじと判定された文を黒く塗りつぶすことで非表示にしたが, これではどの文が隠されたか分かってしまうため, 被験者実験においてはあらすじと判定された文を消去することで非表示にする。そのため, 被験者はどこが消去されたのか分からない。

評価する手法は, あらすじ分類とランダム分類の2つである。あらすじ分類は, NBayes のアルゴリズムを用い, 人名と特有語の一般化を行って付与したスコアをもとに消去する。ネタバレ遮蔽率は80.6%とする(3章の機械学習の標準の閾値を用いたため)。ランダム分類は, あらすじ分類で消去された文の数と同様の割合でランダムに消去する。

被験者実験の結果から, あらすじ分類とランダム分類に対して文脈欠損率を求めた。各手法で文脈が崩れた回数を各手法で表示する文の数で割った値を文脈欠損率とする。結果として, ランダム分類の文脈欠損率は15.3%, あらすじ分類の文脈欠損率は9.1%となった。

あらすじ分類もランダム分類も同じ割合で文を消去しているにも関わらず, あらすじ分類の方がランダム分類よりも文脈が崩れる割合が少ない。これは, あらすじでない文の方が, あらすじである文と比べてレビューの文脈の維持に必要であるからだと考えられる。

## 7. 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文を対象としたあらすじ分類手法の提案を行った。あらすじの判定には5つの代表的な機械学習アルゴリズム(Naive Bayes, Support Vector Machine, ロジスティック回帰, 決定木, k-nearest neighbor 法)を適用した。結果として, Naive Bayes が最も良かった。

分類性能向上のための改善策として, 人名(登場人物名, 著者名, 俳優名とその他人名)と特有語(アイテム特有の単語)の一般化を行った。結果として, 人名と特有語の一般化を行うことがあらすじ分類に有効であることが分かった。

また, あらすじを非表示にするシステムを実装し, 被験者実験により評価した。結果として, あらすじを非表示にするシステムは, ネタバレの遮蔽率が高く, なおかつ, レビュー及び有用な文の表示率が高くなるようなレビューの提示を実現できていることが分かった。

今後は, ネットバレーとネットバレーでないあらすじの違いを検討し, ネットバレーの本質を調査することで効率的なネタバレの遮蔽を目指す。また, ブログやTwitter などショッピングサイトのレビュー以外への適用にも取り組む予定である。

## 参考文献

- [1] Pang, B. and 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. and 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. and Wiebe, J.: Learning Extraction Patterns for Subjective Expressions, Proc. of EMNLP'03, pp.105–112, (2003).
- [4] Dave, K., Lawrence, S. and 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. and Hoffmann, P.: Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis, Proc. of HLT'05, pp.347–354, (2005).
- [6] Titov, I. and McDonald, R.: Modeling Online Reviews with Multi-grain Topic Models, Proc. of WWW'08, pp.111–120, (2008).
- [7] Hu, M. and Liu, B.: Mining and Summarizing Customer Reviews, Proc. of KDD'04, pp.168–177, (2004).
- [8] Huang, J., Etzioni, O., Zettlemoyer, L., Clark, K. and Lee, C.: RevMiner: An Extractive Interface for Navigating Reviews on a Smartphone, Proc. of UIST'12, pp.3–12, (2012).
- [9] Hijikata, Y., Ohno, H., Kusumura, Y. and Nishida, S.: Social Summarization of Text Feedback for Online Auctions and Interactive Presentation of the Summary, Proc. of IUI'06, pp.242–249, (2006).
- [10] 中村聡史, 小松孝徳: スポーツの勝敗にまつわるネタバレ防止手法の検討, インタラクシオン 2012 予稿集, pp.9–16, (2012).
- [11] 岩井秀成, 池田郁, 土方嘉徳, 西田正吾: ネットバレーを見ずにレビュー文を閲覧できるシステムの開発, WISS2011 論文集, pp.36–41, (2011).
- [12] John, G.H. and Langley, P.: Estimating Continuous Distributions in Bayesian Classifiers, Proc. of UAI'95, pp.338–345, (1995).
- [13] Platt, J.C.: Fast Training of Support Vector Machines Using Sequential Minimal Optimization, Advances in Kernel Methods, MIT Press, pp.185–208, (1999).
- [14] Le Cessie, S. and Van Houwelingen, J.C.: Ridge Estimators in Logistic Regression, Applied Statistics, Vol.41, No.1, pp.191–201, (1992).
- [15] Quinlan, J.R.: C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, (1993).
- [16] Dasarathy, B.V., Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, IEEE Computer Society Press, (1990).
- [17] Siegel, S. and Castellan, N.J.Jr.: Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, (1988).
- [18] 北研二, 津田和彦, 獅々堀正幹: 情報検索アルゴリズム, 共立出版, (2002).
- [19] Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, Proc. of IJCAI'95, Vol.2, pp.1137–1145, (1995).
- [20] Glover, E.J., Tsioutsoulouklis, K., Lawrence, S., Pennock, D.M. and Flake, G.W.: Using Web Structure for Classifying and Describing Web Pages, Proc. of WWW'02, pp.562–569, (2002).