

マイクロブログからの映画評判情報の抽出

倉橋 宏幸[†] 青野 雅樹[‡]

[†] 豊橋技術科学大学情報工学課程 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

[‡] 豊橋技術科学大学情報・知能工学系 〒441-8580 愛知県豊橋市天伯町雲雀ヶ丘 1-1

E-mail: [†] kurahashi@kde.cs.tut.ac.jp, [‡] aono@tut.jp

あらまし 近年、インターネットの普及に伴って、Web 上の Blog や電子掲示板等において、個人の様々な意見が取得できるようになった。本研究では、Web 上で盛んに利用されているサービスのひとつである Twitter に注目し「公開中の映画」を対象に評判情報を抽出する。具体的には、映画のタイトルを含む Tweet を収集し、意見を含む「レビュー」と、それ以外の「非レビュー」に分類する。文書分類に用いる特徴量について、文書中の BOW を基本的な特徴量としたうえで、本研究ではそれ以外にいくつかの素性を提案する。分類を行う手法としては、SVM と Random Forest による機械学習、素性のスコアリングによる手法を用いて行った。いずれの手法においても F 値が 0.8 に近い値となり、BOW のみで分類する場合よりも提案素性を組み合わせることによって分類精度を向上することができ、提案素性の有効性を確認できた。

キーワード Twitter, 評判情報, SVM, Random Forest

Extraction of movie reputations from microblog

Hiroyuki KURAHASHI[†] Masaki AONO[‡]

[†] Department of Information and Computer Sciences, Toyohashi University of Technology

1-1 Hibarigaoka Tenpaku-cho, Toyohashi, Aichi, Japan

[‡] Department of Computer Science and Engineering, Toyohashi University of Technology

1-1 Hibarigaoka Tenpaku-cho, Toyohashi, Aichi, Japan

E-mail: [†] kurahashi@kde.cs.tut.ac.jp, [‡] aono@tut.jp

Abstract Recently, microblog (e.g. Twitter) has been very effectively used for communicating information in real time. In this research, we will focus on Twitter as a typical example of microblog. We have collected movie reputations from Twitter. The purpose of this research is to classify Twitter that includes movie comments into either ‘review type’ or ‘non-review type’, and to quickly grasp latest movie review. To classify Twitter, we proposed several unique boolean features, such as “not including adjective” and “frequently used verbs plus several varying terminating words” (e.g. “Mita”, “Mitai”), in addition to BOW (Bag Of Words). We employed three different classification methods including SVM, Random Forest, and Fujimura’s method. From our experiments, we have achieved approximately 0.8 (in F-measure).

Keyword Twitter, Extraction of reputation, SVM, Random Forest

1. 研究背景

近年、インターネットの普及に伴って、様々な情報を Web 上から入手することが可能になった。特に Blog や電子掲示板等では、個人の様々な意見が取得でき、企業にとってはマーケティングやクレーム処理、個人にとっては目的の評判を探す際の情報源として利用することができる。また、評判情報を探す際の主な情報源としてレビューサイトの利用があげられる。しかし、発売されたばかりの製品や、公開されたばかりの映画など即時性の求められる対象について評判情報が不足しがちであるという問題点があり、即時的により多く

のレビューを集めることが求められている。

そこで本研究では Web 上で盛んに利用されているサービスのひとつであるマイクロブログに注目する。マイクロブログとは、チャットとブログの中間的なサービスであり、近年そのコンテンツは急増している。本論文ではマイクロブログの中でも最も有名なものの 1 つである Twitter を研究対象とする。

Twitter は 140 文字という書き込みの字数制限による気軽さ、携帯端末からの書き込みの容易さから、即時性の高い情報発信ツールとして活用されている。ユーザが投稿する 1 つの文章を Tweet と呼び、投稿される

Tweet は多岐にわたる．例えば，ユーザ自身に関連すること以外にも，ユーザが興味を持った映画や書籍，ニュースサイト等，他のメディアに関する多くの感想や意見が投稿されている．

本研究では，上記のような特性を持つ **Twitter** を対象とし，「公開中の映画」を対象に評判情報を抽出することを目的とする．具体的には，映画のタイトルを含む **Tweet** を収集し，意見を含む「レビュー」と，それ以外の「非レビュー」に分類する．これらを行うことで，より即時的に評判情報を集めることができるものと考ええる．また，**Tweet** から取得できる内容語に注目し，いくつかの素性を提案することで，分類精度の向上を目指す．

2. 関連研究

マイクロブログの発達に伴い **Twitter** の注目度が高くなったことで，**Twitter** を対象とした研究が増加している．また，評判抽出に関する先行研究としては **Blog** や **Web** ページから意見を抽出する研究が多くなされている．ここでは，**Twitter** の即時性に着目した研究，**Blog** や **Web** ページからの評判抽出に関する研究を紹介する．

小林ら[1]は，評判分析に用いるために，ユーザがイベントを閲覧しているかを **Twitter** の書き込み内容から自動で識別する実験を行っている．ここでは多数のユーザが同時に閲覧するイベントの例として，**TV** ドラマの実況書き込みを扱っている．識別には **SVM** を用いており，単語の出現頻度上位を要素とする特徴ベクトルで表している．

坂本ら[2]は，**Twitter** の即時性に着目し，イベントの要約生成システムを試作している．具体的には，**Twitter** の記事から単一のトピックに関して抽出された記事を対象とし，リアルタイムにバーストを監視することでトピック内に起こったイベントを検出する．これを要約の対象とすることで，随時イベントの要約を生成して並べる手法となる．

これらのように，**Twitter** の即時性に着目した研究が多くなされている．本研究では，**Twitter** を利用し評判情報を抽出することで，より即時的な評判抽出が行えるのではないかと考えている．

立石ら[3]は **Web** ページから意見文を抽出することを目的として研究を行っている．意見情報を「対象物，属性，評価」の3つ組で定義し，この定義を用いてあらかじめ特徴語辞書を作成する．この表現と抽出ルールを手掛かりに **Web** ページから意見文を抽出している．

藤村ら[4]は肯定的／否定的文書中の出現頻度を用いることで，単語のスコアリングを行い，電子掲示板

の評判情報を文単位で肯定と否定に分類している．スコアリングは式(1)により行っている．

$$score(w_i) = \frac{P_P(w_i) - P_N(w_i)}{P_P(w_i) + P_N(w_i) + k} \quad (-1 \leq score(w_i) \leq 1) \quad (1)$$

$$Score(d) = \sum_{ALL w_i} score(w_i) \quad (2)$$

$$\begin{cases} \text{if } Score(d) > 0 \rightarrow \text{positive} \\ Score(d) < 0 \rightarrow \text{negative} \end{cases} \quad (3)$$

$P_P(w_i)$ は肯定的な評判に属性 w_i が出現する確率， $P_N(w_i)$ は否定的な評判に属性 w_i が出現する確率を表しており，これらの確率を利用し属性 w_i のスコア $score(w_i)$ を求めている．最終的な分類は各文書に含まれる属性のスコアの総和が 0 より大きければ，肯定的な評判であるとし，0 より小さければ，否定的な評判であるとしている．

3. 提案手法

本章では，**Twitter** における映画評判情報を含む **Tweet** を収集する為の流れを示す．以下より時系列の処理手順を述べる．

3.1 データセットの作成

分類を行うための **Tweet** を **Twitter API** により収集する．ここでは日本語の **Tweet** を対象とする．また，URL を含むものは宣伝目的のものが多く，**RT(Retweet)** を含むものは他人の引用を含むため本人の意見としては相応しくないと考え，これらは対象外とする．また，訓練・辞書作成用データと実験評価用データとで，別々の映画タイトルを含む **Tweet** を収集するものとする．データセットの作成のため，収集した **Tweet** を読み「レビュー」か「非レビュー」かを判断し，人手でラベル付を行う．実際に収集した **Tweet** の映画タイトルと，その **Tweet** 数を表 1，表 2 に示す．

表 1 訓練，辞書作成用データ

訓練・辞書作成用		
タイトル	review	nonreview
モテキ	479	778
一命	350	666
電人ザボーガー	346	665
夜明けの街で	342	673
キャプテン・アメリカ	128	124
ザ・ウォード	277	354
ラビット・ホラー	124	218
ワイルド・スピード	151	364
恋の罪	374	587
ツレがうつになりまして	429	571
合計	3000	5000

表 2 実験評価用データ

実験評価用		
タイトル	review	nonreview
アジョシ	375	625
猿の惑星	413	587
探偵は BAR にいる	478	522
アントキノイノチ	311	689
マネーボール	375	625
合計	1952	3048

レビューの基準としては「映画を観たうえで、感想を語っていることが想像できるもの」とし、非レビューの基準としては「感想を述べていない、映画を未見であることが想像できるもの」としている。

3.2. 特徴語辞書の作成

訓練・辞書作成用に収集した Tweet を形態素解析器 MeCab にかけて必要な品詞を取り出し、分類指標となる特徴語辞書を作成する。ここで MeCab により取り出す品詞としては内容語である「名詞、動詞、形容詞、形容動詞」とする。取り出した内容語を元に特徴語辞書を作成するが、ここではその手順を述べる。

①語の出現頻度を閾値とすることで特徴語を選定

ここでは出現頻度の低い形態素は特徴語として相応しくないと考え除去する。また、藤村らのスコアリング手法による形態素のスコアが零付近の形態素を除去する(素性のスコアリングについては次節で述べる)。スコアが零付近の形態素は特徴語として不適切であり、スコアの絶対値が 1 付近の形態素は特徴語として適切である可能性が高いためである。

②映画の内容に依存する形態素、非自立語等を除去

例えば映画のタイトルや俳優名、役名などの映画に依存する形態素を除去する。また、名詞である「こと」や、動詞である「する」などの形態素は、それだけでは意味をもたないために除去する。

以上のような手順をふむことで、主に評価表現を中心に構成された特徴語辞書を作成することができる。

3.3. BOW(Bag Of Word)の作成

特徴語辞書を元に BOW(Bag Of Word)の作成を行う。具体的には、特徴語を素性とするベクトルモデルの作成、藤村らのスコアリング手法を参考とし、素性のスコアリングを行った特徴語スコア表を作成する。ベクトルモデルの作成には素性出現の Boolean を与えるものとする。また特徴語のスコアリングは式(4)を用いる。

$$score(w_i) = \frac{P_R(w_i) - P_N(w_i)}{P_R(w_i) + P_N(w_i) + k} \quad (4)$$

$$(-1 \leq score(w_i) \leq 1)$$

$P_R(w_i)$ はレビューにて属性 w_i が出現する確率である。同様に $P_N(w_i)$ は非レビューにて属性 w_i が出現する

確率である。また、 k は例えば $P_R(w_i)$ が 0 であった際、 $P_N(w_i)$ によらずスコアが 1 になってしまう問題を解決するために分母に設けられた実数である。この式を用いると、レビューについて特徴的な語句は正の値をもって現れ、非レビューについて特徴的な語句は負の値をもって現れることになる。実際にスコアリングされたものの例を表 3 に示す。

表 3 スコアリングの例

type	語	score
review	素晴らしい	0.834016
	良い	0.774638
	よい	0.749395
nonreview	早い	-0.41892
	はやい	-0.40367
	忙しい	-0.4

3.4. 分類手法

作成した BOW と、次節で述べる提案素性を元に、実験評価用データにおいて「レビュー」と「非レビュー」の分類を行う。分類手法としては、機械学習手法として SVM, Random Forest を用いたものと、藤村らのスコアリング手法(以降 Score 法とする)を用いて行う。SVM, Random Forest には R 言語とそのパッケージである”e1071”及び”randomForest”を利用する。SVM のカーネルには Sigmoid を利用する。また、スコアリングによる分類には以下の式(5)を用いる。

$$Score(d) = \sum_{ALL w_i} score(w_i) \quad (5)$$

$$\begin{cases} \text{if } Score(d) > 0 \rightarrow \text{review} \\ Score(d) \leq 0 \rightarrow \text{nonreview} \end{cases} \quad (6)$$

各 Tweet に含まれる特徴語のスコアの総和が 0 より大きければレビュー、0 以下ならば非レビューと分類する。特徴語が 1 語も現れない Tweet がある場合は 0 となり非レビューと判定する。例えば、感嘆表現や名詞句だけで Tweet を構成している場合がそれにあたる。

3.5 提案素性

分類精度向上のため以下の素性を提案する。

- (1)作成した特徴語辞書にない形容詞を含むかどうか
- (2)形容詞を含むかどうか
- (3)「みる」「いく」に付属語を考慮したもの
- (4)形容詞に頻出するモダリティを考慮したもの

以上の素性について、ベクトルモデルの作成には Boolean を与え、スコアリングには式(4)を用いる。

3.5.1 評価表現には形容詞が多く存在することに着目した素性

評価表現には形容詞が多く存在する。例えば「面白い」や「素晴らしい」などである。レビューに関して

は形容詞が用いられる傾向が多くみられ、逆に非レビューでは形容詞が用いられる傾向が少ない。

そこで提案した素性が(1),(2)である。素性(1)は訓練データにより網羅されていない評価表現を考慮するために提案する。素性(2)はレビューには形容詞が出現しやすい特性があるために提案する。

3.5.2 頻出する動詞に着目した素性

頻出する動詞に着目した素性が(3)となる。TwitterではBlogによる特性から評価対象に対しての行動が動詞として現れる。特に「みる」「いく」についてはTweet中に頻出する傾向がみられ、これらについて付属語を考慮したものを素性とする。以下の表4に「見る」について付属語を考慮した例を示す。

表4 「見る」に付属語を考慮した例

	+	た	=	見た
見る	+	たい	=	見たい
	+	ます	=	見ます

付属語を考慮しない場合、「見る」という形態素については、これから映画を鑑賞する場合と、既に映画を鑑賞済みの場合と二通りの使用が想定される。このような理由から付属語を考慮した素性が有効になると考えた。

この素性では合計22素性を提案素性として用いる。

3.5.3 形容詞の語尾に着目した素性

形容詞の語尾に着目した素性が(4)となる。Twitterではチャットの性質から表5に示すモダリティが出現しやすい。

表5 形容詞に頻出するモダリティ

	+	そう	(推量)
形容詞	+	かな	(疑問)
	+	みたい	(伝聞)
	+	らしい	(伝聞)

これらは形容詞の語尾に現れやすい。例えば「面白そう」という推量や「面白いかな」という疑問のように、周りの反応を伺うようなTweetが多くみられる。形容詞はレビューに多く用いられる語であるが、このようなモダリティ表現はレビューでみられることが少なく、逆に非レビューに関して多くみられる表現である。そのため、これらの素性が有効になるものと考えた。

この素性では合計4素性を提案素性として分類に用いる。

4. 評価実験

本研究の提案手法により、Twitterにおける映画タイトルを含む日本語のTweetを収集し、「レビュー」と「非レビュー」の分類を行った。また、分類を行ったTweetの正解・不正解を確認し、提案手法の精度をみた。以下に詳細を示す。

4.1 評価方法・実験方法

提案素性の有効性を検証する為に、BOWによる素性(681素性)の分類を比較手法とした。BOWに加え、提案素性(28素性)を用いたものを提案手法とする。分類には機械学習手法であるSVMとRandom Forest、スコアリング手法であるScore法により行った。特にRandom ForestはGini係数による提案素性の有効性を検証できるために用いた。評価尺度としては、適合率、再現率、F値による評価を行うものとした。

4.2 実験結果

4.2.1 分類精度

BOWのみで分類を行った場合と、BOWに提案素性を加えたもので分類を行った場合の実験結果を表6から表8に示す。

表6 適合率による評価

	SVM	RF	Score法
BOWのみ	0.815	0.762	0.690
BOW+提案素性	0.839	0.813	0.754

表7 再現率による評価

	SVM	RF	Score法
BOWのみ	0.720	0.763	0.855
BOW+提案素性	0.774	0.788	0.865

表8 F値による評価

	SVM	RF	Score法
BOWのみ	0.765	0.762	0.763
BOW+提案素性	0.805	0.800	0.806

これらより適合率、再現率、F値ともに提案素性を組み合わせることで精度を向上することができ、提案素性の有効性が確認できたといえる。また、SVM、Random Forest、Score法のいずれもF値が0.8に近い値となった。それぞれの手法を比べてみると、適合率は機械学習手法の精度が良く、再現率はスコアリング手法の精度が良い結果となった。このような結果となった要因として、訓練データの割合について非レビューの割合が大きかったことにより、機械学習において過学習が起きていたことが想定される。スコアリング手法で再現率に偏る結果となった要因としては、レビューについては話題の統一性があるが、非レビューに

については話題の統一性がなく、特徴的な表現がとりづらいことがあげられる。このことにより、分類に用いる素性のプラススコア（レビューの特徴となる）、マイナススコア（非レビューの特徴となる）の割合に偏りが発生した。分類に用いる合計 709 素性において、プラススコアが 467 素性、マイナススコアが 242 素性となっており、レビューと誤判定をする要因になっていると考えられる。

機械学習手法とスコアリング手法において、F 値に大きな差がでない結果となった。しかし、実際の Tweet では、レビューに対し非レビューの投稿が多くみられる傾向がある。これを考慮すると、機械学習手法の方が誤判定を防ぐことができ、スコアリング手法より有効であると考えられる。

レビューであると判定された Tweet の一部を表 9 に、非レビューと判定された Tweet の一部を表 10 に示す。

表 9 レビューと判定された Tweet

「アジョシ」はアクションは最高！でも話が雑持のような。
『アジョシ』は全く感情移入出来ないまま終わってしまって「なんだかなあ…」ってカンジだったのだが、隣のおっさん（多分）がラストシーンで泣きだしたから驚いた。

表 10 非レビューと判定された Tweet

アジョシとっても面白そうだなあ…
たまの休みに 1 人映画 2 本だては『モテキ』と『アジョシ』を観た。

4.2.2 素性の重要度

Random Forest により素性の重要度の順位(709 素性中)を求めた結果を述べる。Gini 係数による上位 20 素性の重要度を図 1 に示し、上位 20 位内に現れた提案素性の順位を表 11 に示す。

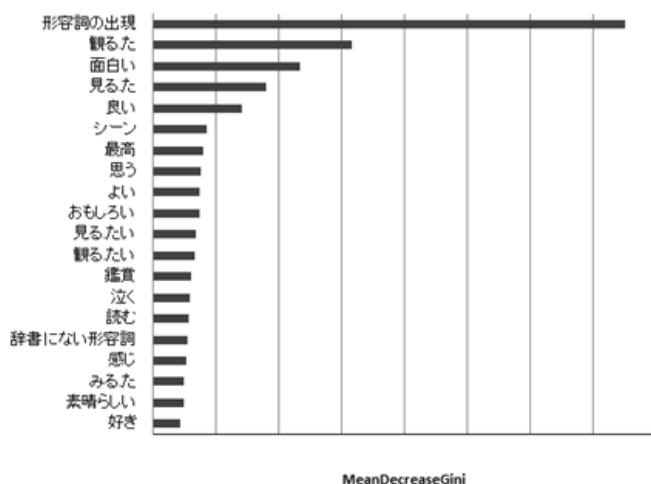


図 1 Gini 係数による素性の重要度

表 11 上位 20 素性中の提案素性

素性名	順位
形容詞を含むかどうか	1
観る＋た	2
見る＋た	4
見る＋たい	11
観る＋たい	12
特徴語辞書にない形容詞を含むかどうか	16
みる＋た	18

上位 20 素性に提案 28 素性中 7 素性を包含する結果となった。また、最も素性重要度の高かった素性は「形容詞を含むかどうか」であった。

提案素性(4)について上位 20 素性に現れていないため、その順位を表 12 に示す。

表 12 提案素性(4)の順位

素性名	順位
そう	31
かな	54
みたい	89
らしい	197

全体の素性数が 709 素性であることを考えると、これらの素性も上位に現れているものがあり、分類に有効であると考えられる。

これらの結果により、提案素性の有効性が確認できた。

5. 考察

5.1. ラベル比を揃えた実験

機械学習手法について、精度が再現率よりも適合率に偏った結果を受け、訓練データのレビュー／非レビューの割合を揃えた実験を行った。SVM における実験結果を表 13 に、Random Forest における実験結果を表 14 に示す。

表 13 SVM における実験結果

review : nonreview	3 : 5	1 : 1
適合率	0.839	0.802
再現率	0.774	0.825
F 値	0.805	0.813

表 14 Random Forest における実験結果

review : nonreview	3 : 5	1 : 1
適合率	0.813	0.747
再現率	0.788	0.865
F 値	0.800	0.802

これらより、SVM では適合率に精度が偏ることが改善され、全体的な精度も向上する結果を得ることができたことが分かる。一方 Random Forest では、適合率に精度が偏っていないものの、逆に再現率に精度が偏ってしまった。分類精度は訓練データの割合に大きく依存することが分かり、分類対象を統一することで適切な結果が得られることが分かった。

5.2. 映画別の実験結果

本研究では分類において、訓練・辞書作成用データと実験評価用データとで別の映画タイトルを含むものを用意している。これは、提案手法が様々な映画に対応できることが望ましいために行ったものである。

実験評価用データとして用意した 5 種類の映画それぞれの分類精度を F 値により評価したものを表 15 に示す。

表 15 映画別の実験結果

	SVM	RF	Score 法
アジョシ	0.759	0.763	0.751
アントキノイノチ	0.78	0.772	0.785
マネーボール	0.818	0.806	0.837
猿の惑星	0.821	0.799	0.808
探偵は BAR にいる	0.839	0.842	0.856

F 値について、どの映画も 0.75 を超える精度となった。しかし、映画により精度に大きな差がでる結果となった。要因としては、訓練・辞書作成用のデータによって評価表現が網羅できていない場合に、映画の内容に依存してしまう可能性があげられる。そのために、今回の結果のような差が生じてしまうのではないかと考えられる。

6. まとめと今後の課題

本研究では、マイクロブログサービス Twitter に注目し「公開中の映画」を対象とした評判情報を抽出する研究を行った。具体的には、映画のタイトルを含む Tweet を収集し、意見を含む「レビュー」と、それ以外の「非レビュー」に分類するものである。

分類手法としては SVM と Random Forest による機械学習手法、藤村らのスコアリング手法を用いた。分類に用いる素性としては、BOW(681 素性)と提案素性(28 素性)により行い、全ての分類手法において F 値が 0.8 に近い値となった。

各素性の重要度を Random Forest により求めると、提案素性である「形容詞を含むかどうか」が最も重要な素性として判定され、上位 20 素性に提案 28 素性中、7 素性を包含する結果を得ることができた。

今後の課題としては、意見として抽出された Tweet の分析があげられる。今回は意見として抽出したのみ

であったが、これを肯定／中立／否定というような分類を行うことが求められる。また、本研究では映画というドメインに依存し研究を行った。しかし、分類のためのベクトルモデル作成や辞書作成には膨大な手間がかかる。そのことを考慮し、ドメインに依存しない分類の研究を行うことが求められる。

参 考 文 献

- [1] 小林尊志, 野田雅文, 出口大輔, 井出一郎, 村瀬洋. Twitter における実況書き込み検出手法の検討. 映像情報メディア学会技術報告 34(25), 129-130, 2010-06-29.
- [2] 坂本翼, 横山昌平, 福田直樹, 石川博. マイクロブログを対象としたリアルタイムな要約生成システムの試作. DEIM Forum, F5-5, 2011
- [3] 立石健二, 石黒義英, 福島俊一. インターネットからの評判情報検索. 情報処理学会研究報告, NL-144-11, pp.75-82, 2001.
- [4] 藤村滋, 豊田正史, 喜連川優. Web からの評判および評価表現抽出に関する一考察. 電子情報通信学会研究報告, Vol. 104, No. 176, pp. 141-146, 2004.
- [5] 青木優, 山本和英. 構文片を用いた分野の同定を必要としない意見・評判情報抽出. 電子情報通信学会技術研究報告. NLC, 言語理解とコミュニケーション 107(480), 7-12, 2008-01-31.
- [6] 川口敏広, 松井藤五郎, 大和田勇人. SVM と新聞記事を用いた Weblog からの意見文抽出. 人工知能学会全国大会論文集, 巻:20th 頁:1A3-3, 2006.
- [7] 佐藤智博, 青野雅樹. リンク情報に基づいたウェブスパム検出. DEIM Form, B3-5, 2011.
- [8] 山本裕介. Twitter API ポケットリファレンス. 技術評論社, 2011-7-15.