Vol.2012-IFAT-105 No.2
Vol.2012-NL-205 No.2
2012/1/20

レビュー文分類器を用いた レビュー文含有比率によるレビュー文書判定

江崎 大嗣[†1]　川場 真理子[†2]　平野 徹[†2]

レビューなどの口コミ情報が消費者の購買活動の意思決定に利用されており，商品のレビューを集めて公開したいというニーズが高まっている．しかし，消費者にレビューを書いてもらうには時間とコストを要する．このようなことから，ブログ等の CGM からレビューとして利用できる文書を自動で収集できることが望まれている．従来，このような文書分類のタスクでは，分類学習を用いた文書単位での分類が行われてきた．しかし，ブログ記事がレビュー文書か否かを判断することは，日記などのその他の情報が多く混在しているため難しい．そこで，本研究ではレビューとそれ以外の情報をより細かく見ることができるように，文単位でレビューかどうかを判定して，ブログ記事内に含まれるレビュー文の比率によって，文書がレビューかどうかを判定する．その結果，F-measure で 72.3 となり，我々の提案手法は従来の手法に比べて 26.4 上回った．

Review decision about the documents by the ratio of a review sentence by classifier using a sentence review

Hirotsugu Esaki,[†1] Mariko Kawaba[†2] and Toru Hirano[†2]

Reviews are used in purchasing decision. The need is growing to collect and publish reviews. However, it cost a lot of time and money to make them write reviews. So it is hoped that the documents which are available for reviews can be collected automatically from the CGM such as blog. Such a document classification task has been performed by using classification learning in a document level. But it is difficult to determine whether a blog is a review or not because many other information are mixed. Therefore we have done classification in a sentence level in order to classify fine-grained. And then classification is done by the ratio of a review sentence. As the result, we achieve 72.3 in F-measure. Our method is superior to conventional methods by 26.4.

1. はじめに

商品やレストランなどに関する情報は，公式サイト以外にもレビューサイトなどに口コミという形で豊富にあり，消費者の購買の意思決定に利用されている．これまでは商品やレストランなどを利用した消費者がその感想をレビューサイトに投稿することで，データが集積されていた．しかし，近年ブログなどの CGM[⋆1] の普及により，集約されない口コミ情報が増加した．そこで CGM からレビュー文書を抽出し，集約することでより多くの口コミ情報を消費者が活用できるようにしようと考えた．

本研究では，飲食店を対象として，店舗名を含むブログ記事をインターネット上から集めてきて，飲食店のレビューとそうでないものに分ける手法を提案する．飲食店のレビュー記事の例を以下に示す．この文書ではショッピングや天気などの他の話題が含まれているものの，値段や食べ物など飲食店に関して言及しているためレビューである．

> 今週はお酒とお肉が食べたい！ってことで昨日は焼肉を食べに行きました．カメラバッグを見たかったのと，彼女のテレビが寿命を迎えそうだったので，ヨドバシに行きつつブラブラしつつ．今週はずっと天気が悪くて土曜日も曇り空だったんですが．モアーズはすっかり夏モード．焼肉は横浜のトラジで．私としてはかつてない値段の焼肉屋でしたが，肉の輝きと脂の乗り方が半端じゃなかった．特に，ネギバカという大量のネギのベッドに乗せて食べたお肉は … もう忘れられない．

一方，飲食店のレビュー文書でないものは以下のようなものである．このブログ記事では飲食店についての言及はあるものの，一般的な話を述べているためレビューでない．

> 16 号線を南西に歩いていくと，なんとも妙なもの見付けた．小倉優子の焼肉屋．ケーキ屋とかレストランだったらわかるけど，なんで焼肉屋なのかな．もっとも，ケーキ屋もレストランもすでにチェーン店はあるし，新規に参入するのは難しいも

†1 奈良先端科学技術大学院大学 情報科学研究科
Nara Institute of Science and Technology Graduate School of Information Science
†2 NTT サイバースペース研究所
NTT Cyber Space Laboratories
⋆1 Consumer Generated Media

Vol.2012-IFAT-105 No.2
Vol.2012-NL-205 No.2
2012/1/20

のなのかもしれない．たぶん，いろいろやってみて焼肉屋だといちおううまく軌道に乗ったということなのだろう．

また一般的に，レビューは評判と混同されがちだが必ずしも同じではない．いつ，誰と，どういう理由でといった投稿者の状況を含むことがあり，そういった情報が含まれていないと消費者は意思決定の参考にできない．例えば，レストランを検索するときに，赤ん坊連れが可能か，そしてその利用者がどういった感想を抱いたかなどの情報が重要である．

従来，このような文書分類のタスクにおいては，文書単位でタグを付与して，それを機械学習を用いて分類を行ってきた．しかし，上記で示したように，ブログ記事には日記などのレビュー以外の情報が多く混在しており，これに文書単位でタグを付与するとそのような情報を誤って学習してしまう恐れがある．また，そのような理由からレビューと日記などが混在している文書にレビューかどうか人手でタグ付けを行うことは難しいと考えられる．そこで，本研究ではレビューとなる文がブログ記事内にどれぐらい含まれているかで，文書をレビューかどうか判定する手法を提案する．

以下，２章では関連研究について整理をする．３章では提案手法について述べる．そして，４章で実験の詳細について述べた上で，５章で結果に対する考察を述べる．最後に６章でまとめと今後の課題を述べる．

2. 関 連 研 究

レビュー分類の関連研究では，レビューが有用かどうかを判定するものや，レビューが肯定的か否定的か分類するものがある．

前者として，文献 1)，2) などの研究がある．これらの研究では Amazon レビューを対象として，レビューとして有用性が高いかどうか SVM によって分類を行っている．これらの研究は本研究と類似しているが，対象とするデータがレビューであり，ブログ記事などの一般的な文書を対象とした我々の研究とは異なる．これらの研究ではレビューの質を判定しており，本研究ではレビューかどうかの判定を行っている．そのため，一般文が混在していることを想定していないため，単純にその手法を用いることはできない．

後者として，文献 3)，4) などの研究がある．文献 3) では，Amazon から取得した複数のドメインからなるデータを使って，SVM により文書単位で極性がポジティブかネガティブかを分類している．文献 4) では，Amazon で売られている本のレビューを使い，SVM を用いて，極性の強さを推定している．評判抽出においてはこのような極性の情報は重要とされているが，本研究で扱うレビューは評判だけでなく状況を扱うため，単純に適用することは難しい．

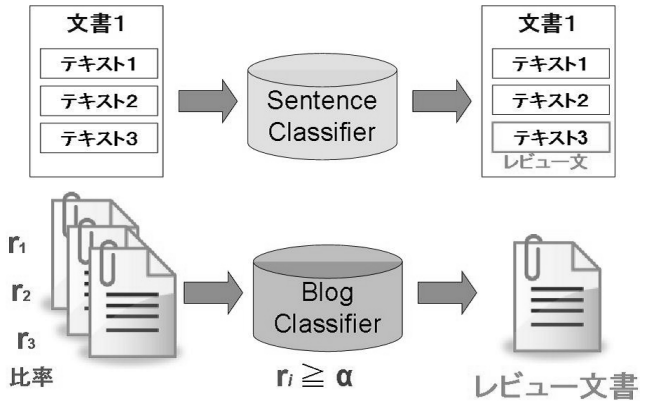

図 1 レビュー文書判定概要
Fig. 1 Outline of review detection

また，類似の研究としてブログ記事から評判を抽出するものがある．文献 5) では，ブログ記事を用いて対象，属性，評価の三項組の評判情報を検索するシステムを提案している．この研究では，三項組を抽出するために，文献 6) の評価表現辞書を用いている．しかし，ある文がレビューかどうかは，必ずしも評価表現を含むかで決められず，レビューと評判が同義ではないため本研究とは異なる．

3. レビュー文含有比率によるレビュー文書判定

本研究では，文書内に含まれるレビュー文の比率によってそのブログ記事がレビューかどうかを判定する．まず文単位でレビューかどうかの分類を行う．次にその結果を用いて，文書に含まれているレビュー文の比率を計算する．そして，その比率がある閾値以上であればレビュー文書として分類する．図１にその概要を示す．

ブログ記事には日記などの情報が多く含まれる．レビューとそのような情報が混在している文書にタグを付与すると，それらの情報も誤って学習する恐れがある．そのため，文書単位より細かい粒度で分類を行うことができるように，文単位で分類を行った．

また，文書を文単位に切り離しても十分にレビューかどうか判断することができると考えた．例えば，以下のような文書があったとする．この場合，二文目は一文目がレビューで

Vol.2012-IFAT-105 No.2
Vol.2012-NL-205 No.2
2012/1/20

あったかどうかに関わらずレビュー文と判定できる．このように，ある文がレビューかどうかは前後の文に依存することが少ないと考えた．

> お店は，横浜 月…このお店，看板は無いし，入り口のドアはどう開けたらいいかわからないし，ミステリアスなお店でした．
> お店の中は，いろんなタイプの個室に別れてて，好奇心をそそられますが，迷路みたいで迷子になりそうでした．

また，文より文書にタグを付与するほうが難しいと考えた．これはブログの一つの記事に，レビューとそれ以外の話題が混在している場合があるからである．そのため，文書がレビューかどうか判定する際に，レビュー情報とそれ以外の話題のどちらが主な内容か判断する必要がある．上記の理由から，本研究では文単位で文書のレビュー分類を行った．

3.1 レビュー文分類器の構築

まず，文単位でレビューかどうか判定する分類器を構築する．分類器の構築には人手で飲食店のレビュー文かどうかタグを付与したものを用いる．

3.2 レビュー文書判定

上記で述べたレビュー文分類器を用いて，文書がレビューかどうか判定する．まず，ブログ記事の文全てに対してレビューかどうかを分類する．次に，文書内の全文に対するレビュー文の割合 r を計算する．そして，閾値 α を設定して，それ以上であればレビュー文書とする．

4. レビュー文書判定実験

分類器がレビューとそうでないものを正しく分類できているか調べるために実験を行う．そして，我々の提案手法が従来手法より優れていることを示す．

4.1 ブログ記事の収集

横浜にある 120 店舗の名前をキーワードとして，飲食店に関係するブログ記事を Web 上から収集した．そして，収集したブログ記事からランダムに 100 記事を抽出した．

4.2 文書，文に対するタグの付与

ブログ記事 100 文書に人手で，飲食店のレビュー文書かどうかの二つのタグを付与した．評価者には ”店舗の話が話題の中心となっていればレビューである ”とするように，タグの付与を行ってもらった．

そして，ランダムに 100 記事抽出した文書に対して，評価者二名にタグを付与してもらい一致度を調べた結果 $\kappa = 0.654$ となった．以下にタグを付与した結果，飲食店のレビューとされた文書の例を示す．

> 横浜うかい亭（鉄板料理）先日，大切な方の結婚式に出席した時に忘れられない様な料理をいただいたので 紹介したいと思います．場所は神奈川県大和市 横浜うかい亭 オーシャントラウトとイチジクのカルパッチョ フォアグラのソテー

同様に，飲食店のレビューでないとした文書を次に示す．

> 今日は大学時代からの友人の結婚式．久々に色んな友人にも会えた．場所は南青山ル・アンジェ教会ラ・ロシェル南青山．どういう結婚式が「良い結婚式」という定義なのかは知らないが，僕は今日は「良い結婚式」だったと思う．

これらのデータから，レビューとされた文書には，店舗の場所やメニューなどが記述されていることが分かる．一方，レビューでないとされた文書は，話題の内容がその店舗に関してではなく，それ以外のものになっていることが分かる．

次に，ブログ記事 100 文書に相当する全文に人手で，飲食店のレビュー文かどうかタグを付与した．また，文書を文に区切るのに句点を利用した．タグは ”A 飲食店のレビューである ”，”B 飲食店のレビューでない ”，”C 飲食店のレビューか不明である ”の三つ用意した．そして，評価者には ”食べ物，建物・場所，サービス，店の雰囲気に関するレビューである ”場合には A のタグ，”飲食店のレビューでない ”ものには B のタグ，”飲食店のレビューか不明，あるいは文の区切りミスなどのノイズと思われる文 ”には C のタグを付与するように指示を行った．”A 飲食店のレビューである ”とタグを付与された例は以下のようなものである．

- Hard Rock Cafe 横浜店 に行って来ました．
- お店は，横浜 月…このお店，看板は無いし，入り口のドアはどう開けたらいいかわからないし，ミステリアスなお店でした．

同様に，”B 飲食店のレビューでない ”とタグを付与された例は以下のようなものである．

- お祝い♪本日 雨が降ってて とても寒い日でした．
- 今日は大学時代からの友人の結婚式．

Vol.2012-IFAT-105 No.2
Vol.2012-NL-205 No.2
2012/1/20

また，”C 飲食店のレビューか不明 ”であるとされたものは次のようなものである．

- …
- 魚素材に愛情を込めて，毎日が真剣です．

また，評価者二名に全文からランダムに抽出した 100 文にタグを付与してもらって，一致度を調べた結果 $\kappa = 0.730$ となった．

4.3 SVM による分類実験

ランダムに 100 記事抽出したものにタグの付与を行った結果，25 %の文書がレビューとなり，16 %の文がレビューとなった．したがって，文書 100 件中 25 件を正例として実験を行い，文のレビュー分類には，1278 件中 201 件を正例として実験を行った．指標としては Accuracy, Precision, Recall, F-measure を用いて，評価では 5 分割交差検定を行った．

4.3.1 レビュー文分類実験

文がレビューであるかどうかの分類実験を行った．SVM には TinySVM[⋆1]を用いて，カーネルには線形カーネルを用いた．また，素性には単語を見出し語にした，bag-of-words を用いた．

4.3.2 レビュー文書分類実験

文書がレビューかどうか分類実験を行った．4.3.1 節のレビュー文分類の結果を用いて，各文書内のレビュー文の比率を計算した．そして，閾値 α を設定して，それ以上レビューを含む文書をレビューとした．また，閾値 α の値を 0 から 1 まで 0.01 刻みに変化させて，各指標の変化を調べた．

ベースラインには文書全体の bag-of-words を SVM によって学習，分類を行ったものを用いた．先と同様に，SVM には TinySVM を用いて，カーネルに線形カーネルを用いた．

5. 結果・考察

ランダムに 100 記事抽出したものに対して分類実験を行った結果を示す．

5.1 レビュー文分類結果

レビュー文分類結果の分割表を表 1 に示す．また，各指標の値を表 2 に示す．これらより，Accuracy が 84.9% に対して，F-measure が 42.9 であることから負例に傾いた分類器ができたことが分かる．これは全文 1077 中 1278 件が負例であったためだと考えられ，正例が少ないことで学習が難しかったことが分かる．

⋆1 http://chasen.org/ taku/software/TinySVM/

5.2 レビュー文書分類結果

閾値 α を変化させていったときの Accuracy 曲線と F-measure 曲線を図 2，3 に示す．横軸が閾値 α，縦軸がそれぞれ Accuracy，F-measure となっている．また，文分類の再現率・適合率が 100%のときに文書分類したときの値を上限値とした．実線が実験値であり，点線が上限値を示している．閾値が $\alpha = 0.20$ のとき F-measure で分類性能が最も高かった．閾値が $\alpha = 0.20$ のとき，上限値で F-measure が最大だった閾値 $\alpha = 0.30$ のときのレビュー文書分類結果を表 3 に示す．

このグラフより文書中にレビュー文が 20%以上含まれていることがレビューの判断基準になっていることがわかる．baseline と比較すると Accuracy で 11，F-measure で 26.4 高くなっており，同じ学習データ数だと文単位でレビューを判定して，その比率によって分類する方が優れていることが分かった．上限値に関しては，$\alpha = 0.29$ であったとき最大となり，そのとき F-measure が 96.2 であった．これより，上限ではレビュー文が約 30%以上含まれていることが判断基準になっていることがわかる．実験値最大のときの閾値 α が，上限値最大のときの閾値 α より，0.09 低いことは，レビュー文分類器が負例に偏っているため，レビューでないと分類することが多いためだと考えられる．

また，タグの付与に関して，文書に対して $\kappa = 0.654$ であり，文に対して $\kappa = 0.730$ であることから，文にタグを付与するほうが揺れが少ないことが分かった．また，閾値 α の設定の仕方で性能が変化するため，閾値 α を頑健に決定する手法が必要なことがわかった．

5.3 エラー分析

Accuracy，F-measure 共に値が変化しなくなった，閾値を $\alpha = 0.65$ としたときのレビュー文書判定が間違っていた 4/100 件に関して分析を行った．4 件とも文書が文に上手く区切れておらず，1 文から成り立っており，その文が誤って分類された結果，Accuracy が下がっ

表 1 分割表
Table 1 Contingency table

	正例	負例
正例であると予測	72	63
負例であると予測	129	1014

表 2 レビュー文分類結果
Table 2 Result of review classification of sentence

Accuracy[%]	Precision[%]	Recall[%]	F-measure
84.9	53.3	35.8	42.9

Vol.2012-IFAT-105 No.2
Vol.2012-NL-205 No.2
2012/1/20

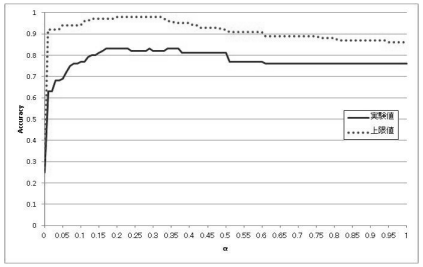

図 2 Accuracy グラフ
Fig. 2 Graph of Accuracy

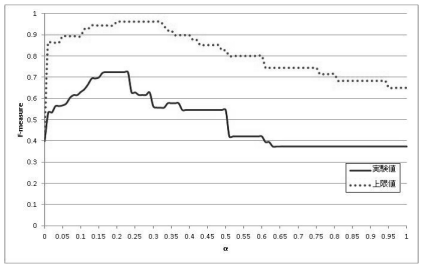

図 3 F-measure グラフ
Fig. 3 Graph of F-measure

てることが分かった．判定が間違ってた文書の例を以下に示す．この文書では，ブログの著者が文の区切りに句点を用いてないため，文の分割に失敗している．また，内容に飲食物の話があがっているが，飲食店の記述ではなく，お祭りの出店についての記述が主になっているため，この記事は飲食店のレビューとはされなかった．このように，店舗以外の話題にも関わらず，飲食物の話題が含まれているときに分類が失敗していることが分かる．

> お祭りこんばんわ 今日は 六本木ヒルズ で行われているお祭りへ 会社帰りに husband と待ち合わせして けやき坂に面したアリーナで開催 ヒルズの中にあるレストランが多数，出店してます 梅蘭 で焼きそば食べることにしたよ〜 中にあんかけの具がたっぷり入っておいしかったよ〜 でも結果的には Rigoletto に寄っちゃった 相変わらず Bar section は 外国人 と日本人 の出会いの場になってますね〜 見てると 楽しい です ピザをオーダー ここのは 窯焼き Pizza だからおいしいっ 2人で食べすぎた ので反省しながら歩いて帰ったのでありました Good night

本研究ではブログ記事を実験対象としているため，文が句点で区切られていない記事が多くあるため，性能を改善するには文書を文に上手く区切る必要があることが分かった．

表 3 レビュー文書分類結果 ($\alpha = 0.20$)
Table 3 Result of review classification of document($\alpha = 0.20$)

	Accuracy[%]	Precision[%]	Recall[%]	F-measure
baseline	72.0	44.4	48.0	45.9
Proposed method(α =0.20)	83.0	60.7	89.5	72.3
Proposed method(α =0.29)	83.0	59.0	66.6	62.8

6. おわりに

本研究ではブログ記事がレビュー文書となっているかを SVM を用いて実験を行った．実験では，まず文単位にレビューかどうかを学習させ分類を行い，それを用いて文書に含まれるレビューの比率を計算して，それが閾値 α 以上を超えていればレビューとして出力することで分類を行った．この手法は，同じデータ量の場合，単純に文書にタグを付与して学習させたものよりも優れていることが分かった．

今後の課題としては，閾値 α を最適に決定する手法を見つけることや，他のドメインで閾値 α の最適値がどうなっているかを調べることが考えられる．また，学習量を増やしていったときに指標がどう変化していくか調べることが必要だと考えられる．そして，単文から成り立っている文書を分類するには，文の分類器の性能を上げる必要があるため，文に有効な学習の手法を考える必要があることが分かった．また，今回は文書がレビューであるかどうかの分類を行ったが，最終的にはその文書が知りたい店舗のレビュー記事かどうかの分類を行う必要があると考えられる．

参 考 文 献

1) Jingjing, L. et al.: Low-Quality Product Review Detection in Opinion Summarization, *Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* , Prague, ACL, pp.334–342 (2007).
2) 山澤 美由起, 吉村 宏樹, 増市 博.: Amazon レビュー文の有用性判別実験, 情報処理学会研究報告, NL173, pp.15–20 (2006).
3) Maria T. et al.: Automatic Sentiment Classification of Product Reviews Using Maximal Phrases Based Analysis, *Proceedings of the 2nd Workshop on Computational Approaches to Subjectivity and Sentiment Analysis* , Portland, Oregon, USA ACL-HLT 2011, pp.111–117 (2011).
4) Daisuke Okanohara and Jun'ichi Tsujii.: Assigning Polarity Scores to Reviews Using Machine Learning Techniques, *IJCNLP 2005, LNAI 3651* , pp.314–325 (2005).
5) M. Tsuchida,H. Mizuguchi and D. Kusui.: Ranking Method of Object-Attribute-Evaluation Three-Tuples for Opinion Retrieval, New Frontiers in Artificial Intelligence: LNAI 5447,pp. 87-98, 2009.
6) 小林のぞみ，乾健太郎，松本裕治，立石健二，福島俊一.：意見抽出のための評価表現の収集， 自然言語処理，Vol.12，No.2，pp.203-222，2005.