

## 言語表現の特徴に基づくソーシャルブックマークコメントの分類

## Classification of Social Bookmark Comments Based on Linguistic Representation

高田 彰†

山田 剛一†

絹川 博之†

Akira Takada

Koichi Yamada

Hiroshi Kinukawa

## 1. はじめに

近年、Web上から発信される情報をクリッピングする場としてソーシャルブックマーク (Social Bookmark ; SBM) というサービスがある。ソーシャルブックマークではユーザのブックマークをオンライン上で保存、管理することができ、それらの情報を元にユーザ間で話題を共有することができる。従来、これらの情報は情報発見のために用いられるとされてきたが、近年では意見を表明する場、議論する場としても機能することがわかった。

このため、本論文ではコメントに着目することにし特徴を調査、分析した。この研究でははてなブックマーク [1] を対象にコメントをフィルタリングするシステムを作成することを目的としている。

## 2. SBMにおけるコメントの特徴と用途

SBMにおけるコメントの記述目的には制限がないため、記述の内容や意図はさまざまである。ここでは内容の種類に着目し、大きく3種類に分類する。

## 1) 引用、要約、単語

コメント記述者の視点による、エントリの要点が示されたもの。究極の要約として1単語で記述されることもある。コメント閲覧者はエントリの要点、注目されている部分を知ることができる。

## 2) 補足

エントリの内容に関する補足情報を提示しているもの。補足情報のありかをURLで提示していることも多い。コメント閲覧者は第三者視点による追加の情報を得ることができる。

## 3) 意見・感想

エントリ自体やエントリに書かれている事柄についての意見や感想が主であり、コメント記述者間での論争を誘発することがある。閲覧者はサイト内での意見の分布を知ることができる。

意見・感想には多くの種類があるが、特に4分類について独立して取り上げることとし、次の5つに細分類する。

3-1) 疑問: エントリの内容に対して疑問を提示しているもの。エントリの説得力や閲覧者の理解度が表れる。

3-2) 評価: エントリ自体に対する評価を示すもの。エントリに記述されている事柄に対する評価は含めない。エントリ自体の評判を表す。

3-3) 希望: エントリの内容を踏まえた、コンテンツ制作

者や閲覧者自身に対する希望を記述したもの。閲覧者の率直な思いが表れる。

3-4) 推測: エントリの内容から想像ができる状況に関して記述したもの。閲覧者の深読みや妄想が表れる。

3-5) その他の意見・感想: エントリや他のユーザのコメントに関する意見・感想で、3-1~3-4 に含まれないもの。

## 3. コメントの自動分類システム

コメントに対して前節で挙げた分類をすることにより、閲覧者のニーズに合ったコメントだけを提示することが可能となる。

## 3.1 システムの概要

コメントの分類を実現するには、その言語表現を詳細に分析する必要がある。そのため、コメントに対して形態素解析、構文解析を行う。形態素解析には茶筌、構文解析にはCaboCha[2]を用いる。また、引用か否かを判定するには、エントリの本文を取得し比較する必要がある。そのため、システムの処理の流れは図1ようになる。

## システム処理の流れ

- 1) RSS フィードからエントリのURL、コメント情報を取得。
- 2) エントリからコメント、アフェリエイトの情報を除去し、本文のみを抽出、この際、本文をCaboChaで解析し、本文に出現している表現を把握する。
- 3) 取得したコメントを CaboCha で解析し、構文情報を把握する。
- 4) 構文情報を用いてコメントを分類。

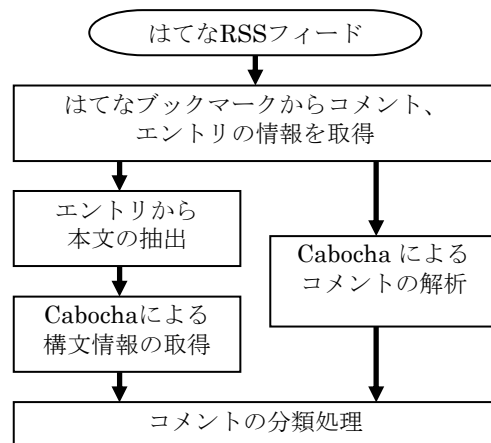


図1. システムの処理手順

†東京電機大学大学院未来科学研究科  
Graduate School of Science and Technology for Future Life,  
Tokyo Denki University

### 3.2 コメントの分類処理

前節で挙げた分類は9種類であるが、性質の異なるものも多く、それぞれ分別の手がかりになるものが異なる。そのため、一度に多クラス分類をするのではなく、分類ごとの異なる手がかりに基づき、順に分類することとした。順序はルール間の優先度に基づき、「補足」「単語」「引用」「疑問」「評価」「希望」「推測」「その他の意見・感想」「要約」の順とした。

なお、コメントが複数の文で構成されていた場合には、まずその各文に対して分類を行う。その後それぞれの分類のうち優先順位の高い分類をそのコメントの分類として扱う。

以下に各分類処理における代表的なルールを示す。

#### 3.2.1 「補足」のコメント分類処理

1) 外部リンクを表す「http」「tp」といった記述を含む場合「補足」とする。

#### 3.2.2 「単語」のコメント分類処理

1) 文に動詞、助動詞が存在しない場合「単語」とする。

#### 3.2.3 「引用」のコメント分類処理

1) コメント文とコンテンツの文章の名詞、動詞の出現順序を比較し、一致する場合「引用」とする。

#### 3.2.4 「疑問」のコメント分類処理

1) 文末に「？」が存在する、もしくは「何故」「どうすれば」などの疑問を表す語が存在する場合「疑問」とする。

#### 3.2.5 「希望」のコメント分類処理

1) 文末に「～して欲しい」「～して下さい」といった表現が存在する場合「希望」とする。

#### 3.2.6 「推測」のコメント分類処理

以下すべての条件を満たした文は「推測」とする。

- 1) 「推測」を表すことの多い助詞が存在する(例「～かも」)
- 2) 1)の助詞が文末に存在する。ただし、文末の「～と思う」「～と感じる」などの表現はないものとして扱う。
- 3) 1)の前にある語が「感情」を示す語ではない。  
(感情を示す語の例「嬉しい」「仕方ない」)
- 4) 一人称やコンテンツ配信者を指す語がない。  
(例「私」「筆者」)

#### 3.2.7 「評価」のコメント分類処理

以下すべての条件を満たした文は「評価」とする。

- 1) 「評価」を示す形容詞、形容動詞がある。  
(例「素晴らしい」「酷い」)
- 2) 1)の後に「評価」の対象が記述されており、それがエントリを示している。  
(例「記事」「文章」「まとめ」)

#### 3.2.8 「その他の意見・感想」のコメント分類処理

以下いずれかの条件を満たした文は「その他の意見・感想」とする。

- 1) 「その他の意見・感想」の手がかりとなる語、絵文字がある。  
(例「www」「(´ω´)」)
- 2) 主部が一人称、コンテンツ配信者を示している。  
(例「俺」「作者」)
- 3) 述部がユーザの行動、考えだとわかる語である。  
(例「懐かしい」「笑える」)

#### 3.2.9 「要約」のコメント分類処理

「要約」は要約元エントリの内容により表現が左右され

表層的な特徴では判断し難いため、他の分類に認定されなかったコメントを要約としている。

## 4. 評価実験

### 4.1 実験方法

今回、はてなブックマークに存在する8つのカテゴリそれぞれから1つのエントリ、計8 エントリを抽出しそれぞれのコメントについて分類の内訳を調査した。なお、条件としてブックマーク数が100を超えているエントリの中から無作為に選択し、そのエントリに存在するコメントに対して、本システムを用いて解析し精度、再現率を求める。

$$\text{精度} = \frac{\text{システムによって検出された正解コメント数}}{\text{システムによって検出された数}}$$

$$\text{再現率} = \frac{\text{システムによって検出された正解コメント数}}{\text{そのエントリに存在する正解コメント数}}$$

### 4.2 実験結果

本システムを用いて解析した結果、表2の結果になった。

	システム出力数	正解数	出力中の正解数	精度	再現率
補足	7	8	6	0.857	0.750
単語	59	56	51	0.864	0.911
引用	157	174	150	0.955	0.862
疑問	90	60	48	0.533	0.800
評価	13	7	5	0.385	0.714
希望	14	18	11	0.786	0.611
推測	28	25	15	0.536	0.600
意見感想	423	543	349	0.825	0.643
要約	161	61	31	0.193	0.508

表2. コメント分類の精度、再現率

### 4.3 考察

今回、精度が低い分類として「要約」「疑問」「評価」「推測」が挙げられる。「要約」では誤りの多くが「その他の意見・感想」に分類されており、今後「その他の意見・感想」の適合率を上げる必要がある。「疑問」では文末に「？」があっても疑問を表していないコメントがあり精度を下げている。「評価」では評価対象であるエントリを表現する語彙を捉えきれていなかった。「推測」ではユーザ自身についての推測を排除しきれていなかった。

## 5. おわりに

今回、SBMを分類するにあたり各コメントの分類手法を調査、実験を行い、その結果から今後「要約」「疑問」「評価」「推測」の適合率を上げる必要があるとわかった。

### 参考文献

- [1] はてなブックマーク : <http://b.hatena.ne.jp/>
- [2] CaboCha : <http://sourceforge.net/projects/cabocha/>