

# 電子掲示板の発言判別システムの話題依存性の一考察

## A study for dependency of topics in an evaluation system for BBS's comments

小熊健太\* , 一藤裕\*\* , 曾根秀昭\*\*\*

Kenta Oguma\* , Yu Ichifuji\*\* , Hideaki Sone\*\*\*

\*東北大学工学部, \*\*東北大学大学院情報科学研究科, \*\*\*東北大学サイバーサイエンスセンター

\*Department of Engineering of Tohoku University, \*\*Information department of scientific research of Graduate school of Tohoku University, \*\*\*Cyber science center of Tohoku University

キーワード： 電子掲示板 (Bulletin Board System) , キーワード分析(Keyword analysis) , 教育支援システム (Educational support system)

連絡先： 〒980-8578 仙台市青葉区荒巻字青葉6-3 東北大学 工学部 情報知能システム総合学科 曾根・水木研究室

小熊健太 , Tel.: 022-795-6094 E-mail: a5tb2060@cs.he.tohoku.ac.jp

### 1. はじめに

近年、インターネット上での誹謗中傷が社会的問題となっている。ニュースでも大きく取り上げられた「学校裏サイト」によるいじめや、匿名掲示板での誹謗中傷、犯罪予告による逮捕など、インターネット上でのモラルに欠ける行動が取り上げられ、問題視されるようになった。これらの問題の対策として、学校等の各種機関でネット上のモラルに関する教育を行い、ユーザーのモラル意識を高める必要がある。しかし、個々にこうした教育を行うことは非常に膨大な手間がかかってしまうため、学校等の機関のみで十分なモラル教育を行うことは、現実的には容易ではない。この問題を受け、インターネットを利用する際のモラル教育を自動でサポートすることを目的とし、学校裏サイトを対象とした電子掲示板の発言判別シ

ステム(以下、判別システム)<sup>1)</sup>が作成された。

ネット上のモラル欠如が原因で引き起こされる問題は、学校裏サイトのみならず、多種の掲示板で問題となっている。よって、既存の判別システムを学校裏サイト以外のトピックへ応用することが可能となれば、モラル教育のサポートの幅が大きく広がることが考えられる。しかしながら、上記の判別システムは「学校裏サイト」のみを対象としたものであり、他トピックへの応用は現在考えられていない。

本研究では判別システムを学校裏サイト以外の掲示板に応用するための足がかりを得ることを目的とする。

判別システムの他掲示板応用の際、学校裏サイトには見られなかった専門用語の出現によって判別結果が左右されてしまうことが考えられる。そ

のため本研究では目的達成の手段として、判別システムのデータベースから専門用語を除外することを提案する。

## 2. 既存システムと問題

### 2.1 専門用語の定義

本研究では、専門用語を次のように定義する。

- そのトピック以外ではほぼ出現することのない単語
- 一般的に発言の判別を左右しない単語

### 2.2 問題発言の定義

本研究における問題発言の定義は、「掲示板の他ユーザーを批難する発言」及び「一般に他者が読んで不愉快な発言」である。掲示板での議論対象に対する批判（例：政治トピックでの法案に対する批判）やフレーミングは通常発言として許容する。また、通常発言の定義は、全発言から問題発言・コピーアンドペースト・アスキーアートを除いた発言である。

### 2.3 既存の判別システムの概要

判別システムの動作は、学習フェーズと判別フェーズに分かれる。

学習フェーズでは、判別システムのデータベースに掲示板の発言データを格納する。判別フェーズでは、判別システムのデータベースを用いて掲示板の発言の自動判別を行う。

#### 2.3.1 学習フェーズ

学習フェーズでは、掲示板の発言を判別する前段階として、人間の手で掲示板の発言を通常発言・問題発言に分類し、結果を判別システムのデータベースに格納する。

学習フェーズの概要をFig. 1に示す。学習フェーズにおいて、判別システムに発言データが格納される際、データは単語及び二つの単語を組み合わせたペア単位で格納される。そのため、はじめに掲示板の発言は、形態素解析に基づいて単語に分解される。続いて、単語を二つずつ全パターンで組み合わせた単語のペアが生成される。これらの単語・ペアはデータベースに格納される。発言が通常発言（問題のない発言）ならば、これらの単語・ペアについて、データベース内の「通常発言での出現回数」が出現した回数ずつ増える。問題発言ならば、「問題発言での出現回数」が出現した回数ずつ増える。

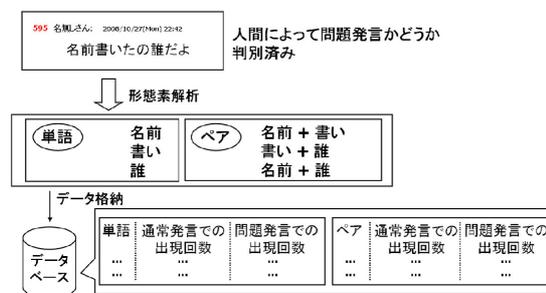


Fig. 1 学習フェーズの概要

#### 2.3.2 判別フェーズ

判別フェーズでは、学習フェーズで生成されたデータベースを基に掲示板の発言が通常発言であるか問題発言であるかの判別を行う。

判別フェーズの概要をFig. 2に示す。判別フェーズでは、発言内の単語・ペアの判別結果によって発言の判別がなされる。はじめに、学習フェーズと同様に発言は単語・ペアに分解される。続いて、分解された単語・ペアはデータベースに基づき、ベイズの理論<sup>2)</sup>によって0から1の範囲の値を付けられる。この際、通常発言に多く出現する単語・ペアほど0に近い値がつけられ、問題発言に多く出現する単語・ペアほど1に近い値がつけられる。そして、各単語・ペアにつけられた値から、単語

による判別結果、ペアによる判別結果の二つの値が算出される。

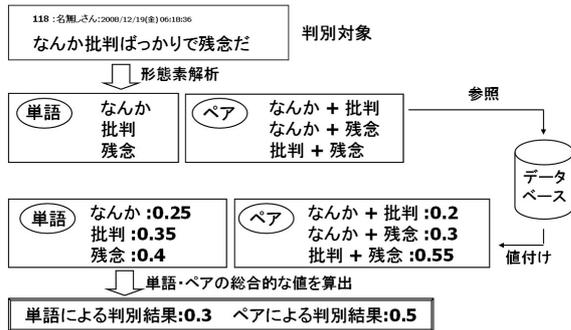


Fig. 2 判別フェーズの概要

## 2.4 他トピック応用上の問題点

前節までに記したとおり、判別システムは単語・ペアによる判別結果を出力するため、他掲示板への応用の際に出現する専門用語が問題となってしまう可能性がある。一部の専門用語が極端に低い値、または高い値を持ってしまい、判別フェーズの際に発言判別に影響を及ぼしてしまう恐れがある。この問題について次章で詳しく述べる。

## 3. 専門用語除外の提案

判別システムの他掲示板への応用の際、トピックによっては専門用語が出現する。

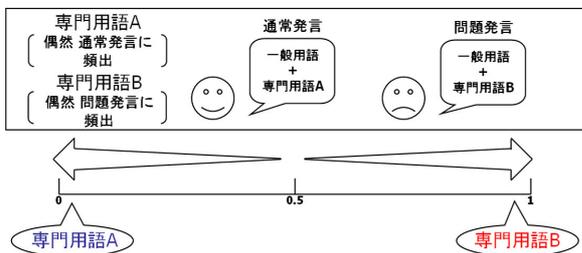


Fig. 3 特定の専門用語を持つ値

Fig.3は、特定の専門用語が極端な値を持ってしまうケースを図式化したものである。トピックによっては、同じテーマを話題とする発言が大量に

続くことがあり、その際に偶然、一部の専門用語が通常発言あるいは問題発言に頻出となってしまう可能性がある。それらの専門用語は学習フェーズの際、極端に低い値あるいは極端に高い値を持ってしまう。こうした専門用語を含む発言は、その単語の持つ極端な値によって判別結果が左右されてしまう。

### 通常発言

景気回復してから消費税増税っていつたが、増税の時期を明記するってことはすなわち景気回復の時期も明記ってことかい  
 単語 : 0.997  
 ペア : 0.999

単語		ペア
回復 0.001	し 0.367	こと 景気 0.381
明記 0.291	て 0.369	いつ 景気 0.738
いっ 0.330	消費 0.381	景気 明記 0.787
税 0.361	こと 0.403	景気 税 0.852
する 0.361	増税 0.451	
	景気 0.519	

Fig. 4 専門用語による誤判定例

Fig.4は、極端な値を持つ専門用語が一因となり判別システムが高い値を出力した例である。専門用語を含むペアが高い値を持つことで、ペア全体の値が引き上げられ、結果として通常発言に対し、ペアとして高い値を出力している。データベースに専門用語を含むことで、こうした誤判定が下されてしまう場合がある。

判別フェーズにおいて、通常発言が問題発言と認識されてしまうことがある。この問題をFalse Positive(以下 FP)と呼ぶ。また、逆に問題発言が通常発言と認識されてしまうことがある。この問題をFalse Negative(以下 FN)と呼ぶ。前述の通りデータベースに専門用語が存在することで、FPやFNを誘発してしまう。専門用語を除外することにより、こうした一部の極端な値を持つ専門用語による誤判定の誘発が解消されると考えられる。

上記の理由から、専門用語の除外によって、より安定した結果の出力が期待できる。よって本研究では、専門用語による誤判定の誘発を防ぐため、

データベースからの専門用語除外の提案を行う。

## 4. 検証実験

### 4.1 実験内容

本研究では、以下の二点について実験を行った。

- 既存の判別システムで他トピックの判別が可能かどうかを確認する
- 専門用語の不要性を確認する

特に本研究では、後者である専門用語の不要性の確認に重点を置いた。

既存の判別システムを他トピックにそのまま適用することが可能であるかを確かめるため、学校裏サイト用のデータベースを持つ判別システムを用いて、学校裏サイト、政治トピック、議論トピックの3種を評価した。

また、専門用語の不要性を確かめるため、政治トピック用のデータベースを持つ判別システムを用いて、政治トピックの発言を評価した。その後、政治トピック用のデータベースから専門用語を除外し、改めて政治トピックの発言を評価した。

### 4.2 実験環境

本研究では、2ちゃんねるの政治トピック及び議論トピックを実験の対象とした。

また実験に用いた判別システムの発言学習数はTable.1の通りである。

Table 1 各判別システムの発言学習数

	通常発言	問題発言
学校裏サイト用データベース	1276	520
政治トピック用データベース	1325	268

実験において、問題発言の基準値を単語・ペア共に0.75以上と定めた。単語・ペアの両方で0.75以

上の値を出力したときのみ、判別システムによる判定は問題発言となる。

### 4.3 判別システムによる各トピックの評価

本研究において、判別システムの評価指標には、問題発言を通常発言と誤判定した割合であるFalse Negative発生率(以下FN発生率)と、通常発言を問題発言と誤判定した割合であるFalse Positive発生率(以下FP発生率)を用いた。この二つの指標を用いることにより、問題発言を判別できない「判別漏れ」及び通常発言を問題発言扱いしてしまう「濡れ衣」がどの程度発生しているのかを明らかにした。

### 4.4 実験結果

既存の判別システムを他トピックにそのまま適用することが可能であるかを確かめるための実験結果として、学校裏サイト用のデータベースを持つ判別システムで「学校裏サイト」「政治トピック」「議論トピック」を評価した際のFN発生率及び、専門用語の不要性を確認するための実験結果として、政治トピック用のデータベースを持つ判別システムで「政治トピック」を評価した際のFN発生率をTable.2に示す。また、同様にFP発生率をTable.3に示す。

Table 2 各トピックのFN発生率

	学校裏サイト用データベース	政治トピック用データベース
学校裏サイト	47.4%	-
政治トピック	63.5%	23.6%
議論トピック	68.5%	-

学校裏サイト用のデータベースを用いたとき、FN発生率は対象が学校裏サイトの場合が最も低く、対象が他トピックの場合は若干高くなってし

Table 3 各トピックのFP発生率

	学校裏サイト用データベース	政治トピック用データベース
学校裏サイト	13.2%	-
政治トピック	13.4%	49.8%
議論トピック	6.7%	-

まうという結果が得られた。つまり、既存の学校裏サイト用データベースでは、他トピックの問題発言の検知漏れが発生してしまう。よって、既存の学校裏サイトのデータベースをそのまま他分野に応用するのは難しいと言える。

政治トピック用のデータベースを用いたときのFN発生率は、対象が政治トピックの場合23.6%と非常に低い値であった。しかし、逆に各トピックのFP発生率が49.8%と非常に高い値を示し、結果的には発言内容に関わらず、多くの発言に対して問題発言であるという判断を行ってしまうことが明らかとなった。

政治トピック用のデータベースから専門用語を除外した判別システムで政治トピックのFN発生率、FP発生率をTable.4に示す。

Table 4 専門用語除外後のデータベースによる判定結果

政治トピック用データベース(除外後)	
FN発生率	44.4%
FP発生率	35.0%

Table.4から読み取れる通り、専門用語を除いた政治トピック専用データベースを用いて、政治トピックの発言を評価した結果、FP発生率が49.8%から35.0%と減少した。しかし、FN発生率が23.6%から44.4%と増加してしまうという結果が得られた。結果としては、通常発言を問題発言と誤判定してしまう件数は大幅に減少したが、逆に問題発言を通常発言と誤判定してしまう件数が増加してしまった。

Fig.5はデータベースから専門用語を削除したこ

通常発言

景気回復してから消費税増税っていったが、増税の時期を明記するってことはすなわち景気回復の時期も明記ってことかい

除外前

単語:0.997  
ペア:0.999

除外後

単語:0.657  
ペア:0.753

単語		ペア
回復 0.001	し 0.367	こと 景気 0.381
明記 0.291	て 0.369	いつ 景気 0.738
いっ 0.330	消費 0.381	景気 明記 0.787
税 0.361	こと 0.403	景気 税 0.852
する 0.361	増税 0.451	
	景気 0.519	

Fig. 5 専門用語除外による値の変化例

とによって判別結果が変化した例である。この例では極端に高い値を持つ専門用語を含むペアが除外されたことにより、ペアの値が低下している。この例のように、専門用語除外により値が引き下げられることが多く、FPは抑えられる傾向にある。ただし一方で、FNは増加している。

## 5. おわりに

### 5.1 まとめ

本研究では、ネット上のモラル欠如が原因で生じる種々の問題の対策として、ネット上におけるモラル教育をサポートするため、学校裏サイトのみを対象とした既存の判別システムを他掲示板へ応用する足がかりを得ることを目的とした。

その手段として、他分野に特化した判別システムのデータベースから専門用語を除外することを提案した。

検証実験として、学校裏サイト用のデータベースを持つ判別システムを用いて学校裏サイト及び各種トピックの発言を評価し、既存の判別システムを他トピックにそのまま応用できるかどうかを確認した。また、政治トピック用のデータベースを持つ判別システムを用いて政治トピックの発言を評価し、その後に同データベースから専門用語を除外し、再度政治トピックを評価することで専門用語除外の効果を確認した。

学校裏サイト用のデータベースを持つ判別システムを用いて他トピックの発言を評価すると、FN発生率が高くなってしまい、問題発言の検知漏れが多くなることが明らかとなった。また、政治トピック用のデータベースを持つ判別システムを用いて政治トピックの発言を評価すると、FP発生率が非常に高くなり、通常発言を約半数を問題発言だと判別してしまうことが明らかとなった。さらに、同データベースから専門用語を除外し、再度政治トピックを評価した結果、FP発生率は抑えられたものの、FN発生率が増加してしまうという結果が得られた。こうした新たな結果を得ることにより、判別システムの他掲示板への応用に向け前進した。

## 5.2 今後の課題

今回の実験において、専門用語除外後、FN発生率が上昇してしまうという問題が発生した。この問題に関し、今後詳細に理由を分析する必要がある。

また、学校裏サイトとその他の掲示板では、語彙の違い、中傷の形態の違い等のさまざまな差異が存在する。判別システムを他分野へ応用するためには、それらの要因の検証も必要である。

## 参考文献

- 1) 一藤裕,今野将,曾根秀昭:”発言の長さに応じた電子掲示板における発言の評価方法に関する研究”,信学技報, vol. 108, no. 75 31/34(2008)
- 2) 田端利宏:”SPAMメールフィルタリング:ベイジアンフィルタの解説”,社団法人情報科学技術協会 Vol.56 464/468,(2006)