

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

一藤 裕† 今野 将†† 曾根 秀昭†††

† 東北大学大学院情報科学研究科 〒980-8578 宮城県仙台市青葉区荒巻字青葉 6-3

†† 千葉工業大学工学部 〒275-0016 千葉県習志野市津田沼 2-17-1

††† 東北大学サイバーサイエンスセンター 〒980-8578 宮城県仙台市青葉区荒巻字青葉 6-3

E-mail: †ichifuji@mail.tains.tohoku.ac.jp, ††konno.susumu@it-chiba.ac.jp, †††sone@isc.tohoku.ac.jp

あらまし インターネット上で学校または地域ごとに集まりコミュニケーションをとる学校裏サイトのような電子掲示板では、しばしば匿名性を悪用し、陰湿ないじめにつながるが多々あり社会問題となっている。これは、電子掲示板に書き込んだ発言が他者にどう影響を与えるかを知らないことが原因の一つとして考えられる。したがって、個々の発言が他者にどう影響を与えるかの指標を確立し、その影響度を発言者へ示すことができれば、問題発言の減少につながると思われる。そこで、本稿では、電子掲示板の発言の特徴である文字と記号のみで構成されていることと、発言の長さがことになっていることに着目する。発言が短いときは、発言中の単語がその発言を象徴し、発言が長いときは、発言中の単語の組み合わせが重要であると考え、既出発言を学習させ、新たな発言が出現したとき、他者にどう影響を与えるかをベイズ理論を応用して影響度を算出する手法を提案する。単語と単語のペアの学習データをそれぞれ準備し、この手法の有効性を示す。

キーワード 電子掲示板, 社会問題, ベイズ理論

An evaluation method depending on the length of comment in BBS

Yu ICHIFUJI†, Susumu KONNO††, and Hideaki SONE†††

† Graduate School of Information Sciences, Tohoku University Aramaki Aza Aoba 6-3 Aoba-ku, Sendai, Miyagi 980-8578 Japan

†† Faculty of Engineering, Chiba Institute of Technology 2-17-1 Tsudanuma Narashino, Chiba 275-0016 Japan

††† Cyber Science Center, Tohoku University Aramaki Aza Aoba 6-3 Aoba-ku, Sendai, Miyagi 980-8578 Japan

E-mail: †ichifuji@mail.tains.tohoku.ac.jp, ††konno.susumu@it-chiba.ac.jp, †††sone@isc.tohoku.ac.jp

Abstract Electronic bulletin board systems (BBS) have become a common communication tool on the Internet. Some people who use BBS anonymously sometimes write opinions which make other people on the BBS feel bad. This is especially true as users become younger and younger and the use of BBS on the Internet and cellular phones becomes more widespread. Users don't understand how such comments can affect readers. As a result, such negative statements sometimes lead younger users to behave inappropriately on BBS. The authors have developed an assistance tool which shows users what kind of influence their statements might have on other people on the BBS. Hopefully this will make all BBS users more aware of the effects of their words on others. This should lead to a more positive atmosphere on the BBS. This assistance tool employs the Bayes theorem based on sentence length. In the case of short sentences, we treat each individual word as samples. In the case of longer sentences, we treat pairs of words as samples. Using these two types of samples, the tool we developed determine the influence of the sentence.

Key words BBS, Social issues, Bayesian theory

1. はじめに

現在、学校裏サイトと呼ばれる学校のいじめの温床となっているサイトがインターネット上に多数存在し、社会問題となっている。これらのサイトでは、各学校ごと、もしくは、その地域の学生によって、生徒や先生への不満や誹謗・中傷が書き込まれており、文部科学省の調べによると、約27%に「死ね」「殺す」といった暴力的な言葉が存在すると報告されている[1]。これは、インターネットでの発言が匿名でできること、何を発言しても問題ないという考えがあること、自分の発言が相手にどう受け取られるかを正しく理解していないことが原因の一つとして考えられる。現在、問題発言を減少させるためにNGワードフィルタの設置や掲示板へのアクセスブロック、掲示板の監視パトロールなどの対策が取られている。しかし、このような対策のみであると、だれでも気軽に利用できる掲示板の特性を失わせることになる。また、本音の討論においては、多少口汚い言葉が混じるものであり、このような言い争い（以下、「フレーミング」と呼ぶ）を全て規制してしまうと、掲示板の存在意義すら失わせるという観点から、すべてのフレーミングを規制する必要はないという意見も存在する[2][3][4]。著者らもこの意見に賛同し、規制だけでなく、ユーザへのインターネットの倫理教育が必要不可欠であると考える。

インターネットの倫理教育は、ユーザ全体へ行うものと、個々の発言を評価し教育するものの2通りに分けられる。全体教育では、ネチケットや匿名性の正しい認識を教育することが可能であると考えられる。しかし、個々の発言に対しては、明確な基準がないため、その影響力を示し、発言の良し悪しを教育することは難しい。したがって、個々の発言が他者へどう影響を与えるかを示す指標を確立し、その影響力をユーザへ示すことが今後必要である。ユーザへ発言の影響力を示すことにより、自分の発言が相手にどう受け取られるかを正しく認識させることができ、結果として、誹謗中傷発言を減少させることが期待できる。

本稿では、発言者へ自身の発言の影響力を示すために、掲示板の既出発言を誹謗中傷発言のような他者に不快感のみを与える発言（以後、「問題発言」と呼ぶ）とそれ以外の発言の2つのクラスに分類し、それを基準として利用し新たな発言を評価する手法を提案する。具体的には、既出発言を分類後、単語に分解し、それぞれの単語が問題発言にどの程度出現するかを算出して学習させる。その後、新たな発言に対し、学習データを用いて、その発言全体の影響度をベイズ理論を応用して算出する。ここで、電子掲示板の発言の特徴の1つである、発言者や発言内容によって、発言の長さが異なることに着目し、発言の長さを考慮して影響度を算出する手法を提案する。

2. 掲示板の特徴と着眼点

2.1 ベイズ理論

ベイズ理論は、過去に起きた事象の確率を利用して、未来に起こる事象の確率を予測する理論である。ある事象Aが生じた時の事象Bが生起する条件付確率（事後確率： $P(B|A)$ ）と事

象Aと事象Bの生起頻度の2つの確率（ $P(A)$ 、 $P(B)$ ）から、事象Bが生じた時の事象Aが生起する条件付確率（事後確率： $P(B|A)$ ）を導き出す定理である。 $P(A) > 0$ とすると、以下の式が成り立つ。

$$P(B|A) = \frac{P(A|B) * P(B)}{P(A)} \quad (1)$$

この理論は、SPAMメールの判定などにも用いられており、学習データを増やすことにより優秀な判定率を誇っている[5][6]。本稿では、ベイズ理論を応用し発言の判定を行う。

2.2 電子掲示板の発言の特徴と対象

一般の電子掲示板の発言は、文字や記号を使って構成されているため、発言が与える影響を考慮する場合、出現単語が非常に重要である。また、発言の長さは、一言二言で終わる短い発言からある程度長い発言まで存在しているため、発言が短い場合、出現単語がその発言の特徴を象徴するものと考えられる。逆に、長い発言の場合、短い発言の場合とは異なる場合が考えられる。例として、ある発言の特定の部分に反対意見を述べる場合を考える。このとき、どの部分に反対意見を述べるかを明示するため、引用する可能性がある。そのため、単語ごとに発言を見た場合、逆の結果が得られることがある。したがって、発言の長さを考慮して、掲示板の発言の評価を行う必要がある。

3. 本提案手法

本判別手法は、図1のように発言を判別するための学習を行う学習フェーズと学習データを用いて判別を行う判別フェーズの2つによって構成されている。以下で説明を行う。

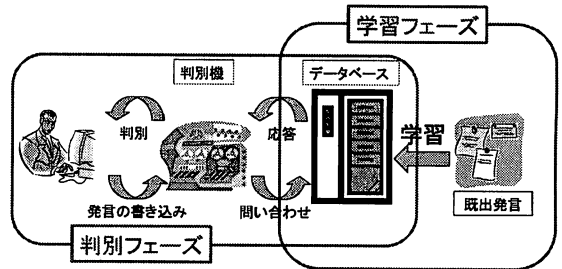


図1 提案手法の概要

3.1 学習フェーズ

発言を判別するための学習方法について述べる。前処理を行った発言を、形態素解析し単語に分解する。このとき、助詞・助動詞はデータとして用いないこととする。これは、事前実験により、どちらの発言にも同じ程度出現し、差異が見られなかったためである。その後、トークンを単語とする場合は単語を、トークンを単語のペアとする場合は単語のペアを作成し、データベースに登録し学習する。具体例を図2に示す。

まず、既出発言を問題発言と通常発言の2クラスに人手を用いて分類し、それぞれのクラスごとに処理を行う。1発言ごとに形態素解析を行い、必要とする品詞の単語を抜き出す。単語の学習データベースには、抜き出した単語とそれぞれのクラス

での出現数を登録する。単語のペアの学習データベースには、抜き出した単語ですべてのペアを作成し、その後、単語のペアのデータベースに単語のペアとそれぞれのクラスでの出現数を登録する。最終的に、表1のように、単語の学習データベースには、単語、問題発言に出現した総数、通常発言に出現した総数が、単語のペアの学習データベースには、単語のペア、問題発言に出現した総数、通常発言に出現した総数が格納されることとなる。同時に、学習させた発言総数もクラスごとに登録する。

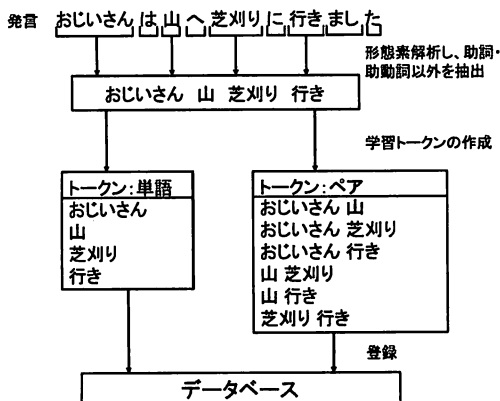


図2 データベースへ登録するための分解図

表1 単語のデータベースの登録例

登録単語	問題発言クラス	通常発言クラス
おめでとう	12	298
ありがとう	15	224
.	.	.

発言の長さに応じて学習データを準備することにより、反対意見を述べることやフレーミングなどを容認しつつ、誹謗中傷の発言を判別することを目指す。

3.2 判別フェーズ

投稿される発言が誹謗中傷発言クラスか問題の無い発言クラスのどちらに分類されるか数値で算出し判別を行う。判別の流れは、図3のように、長さに応じた判別を行う。

新たな発言があるとき、まず、その発言を形態素解析し、必要な品詞の単語を抜き出す。抜き出した単語数から、発言の長さを決定する。その後、長さにあった学習データを用い、発言の影響度を算出し、その結果を発言者へ示す。

単語の場合、発言の影響度を算出するために、まず、出現単語の影響度をそれぞれ算出する。表1の問題発言クラスにおける x^i の総数を $S_b(x^i)$ 、通常発言クラスにおける x^i の総数を $S_g(x^i)$ 、学習させた問題発言の総数を BN 、通常発言の総数を GN とすると、ある単語 x の影響度 $\pi(x^i)$ は式2で算出される。

$$\pi(x) = \frac{\frac{S_b(x^i)}{BN}}{\left(\frac{S_b(x^i)}{BN}\right) + 2 \times \frac{S_g(x^i)}{GN}} \quad (2)$$

発言 y の影響度を $p(y)$ とすると、 $\pi(x^i)$ 、 GN 、 BN を用い

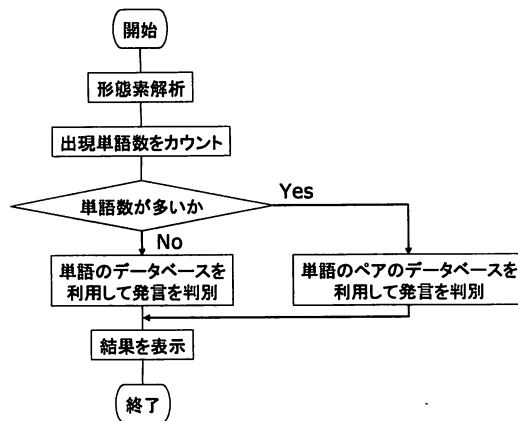


図3 発言の判別の流れ

て、式3.2で算出される。

$$PB = \frac{BN}{BN + GN} \quad (3)$$

$$p(y) = \frac{(PB)^{1-m} \prod_{i=1}^m \pi(x^i)}{(PB)^{1-m} \prod_{i=1}^m \pi(x^i) + (1 - PB)^{1-m} \prod_{i=1}^m (1 - \pi(x^i))} \quad (4)$$

このようにして、発言者へ、自身が発した発言が他者にどう受け取られるかを式で示す。単語のペアの場合も同様に算出する。算出された値が1に近いほど、誹謗中傷発言と判別され、0に近いほど問題のない発言クラスに分類される。

4. 検証実験

本稿では、長さに応じて学習データを変化させ判別を行う手法を提案している。そのため、長さのことなる発言に対し、それぞれの学習データを用いて判別し、どの程度判別率に差があるかを検証する。

4.1 実験準備

今回、いわゆる学校裏サイトと呼称される掲示板の発言を検証実験に利用している。2004年11月から2008年4月までに書き込まれた発言を対象とし、人手を用いて誹謗中傷発言を520発言、問題のない通常発言を1276発言抽出し、学習させている。形態素解析は、“mecab”という形態素解析エンジンを使用し、ネットで多様される造語などをあらかじめ登録している[7]。

4.2 実験内容

発言の長さに応じた判別手法の有効性を示すため、長い通常発言・短い通常発言・長い誹謗中傷発言・短い誹謗中傷発言を用意した。それぞれに対し、単語の学習データを用いた場合と単語のペアの学習データを用いた場合で新規発言が問題発言か通常発言かの判別を行い、どちらが有効かを検証する。

4.3 実験サンプル

検証実験に使用した掲示板の発言は2008年3月から2008年4月までの1300個である。実験に使用したデータのサンプル

の1部を以下に示す。データは、書いた人物を特定されないため、一部改変している。実験には、オリジナルデータを利用している。

表2 実験サンプルの一部

発言例	発言内容
通常発言	けれど別に、誰かさん以外の人に反対しているわけじゃないの。もちろん、みんなが言う意見に賛成できる時もある！いじめなんて、弱い人がやることだと思う。そりゃ、人間なんだから、好きな人・嫌いな人はそれぞれいると思うけれど、だからって陰でこそそとと靴を隠したり、ここで悪口を言ったりするのは、弱いと思う、
問題発言1	醜いあらそいだな。どうせきもいとか言うお前らのほうが超きもかったりするんだよ
問題発言2	うげえ。

(1)は、いじめはよくないということを伝える発言である。しかし、単語に分解し単語のみを見た場合、「いじめ」「悪口」「よわい」「嫌い」など問題発言ととられると判断される可能性がでてくる。しかし、今回の提案手法による単語のペアをみることにより、「反対-じゃない」「いじめ-なんて」といったものに変わり、発言の本質が単語よりも見えるようになる。また(3)のようにペアが作れないような発言も多くあるため、単語のペアだけではなく、単語単体を学習データとして使う必要がある。

4.4 実験結果

長い問題発言と短い問題発言、長い通常発言と短い通常発言を7つ抜き出し、判別結果を表3に示す。

表3 実験結果

発言番号	単語数	単語	ペア	誹謗 or 通常
1	44/596	1.0	0.00013	通常
2	14/29	1.0	1.383E-10	通常
3	5/9	0.1427	0.9019	通常
4	51/599	0.000138	1.0	誹謗
5	12/46	0.0599	1.0	誹謗
6	18/83	0.9989	0.000138	通常
7	1/0	0.999	0	誹謗

表3の単語数欄は、単語の出現数単語のペアの出現数を示しており、単語欄は、単語の学習データを用いて判別したときの影響度を示している。ペア欄、単語のペアの学習データを用いて判別したときの影響度を示し、誹謗 or 通常欄は、人間がその発言を読んだときの判別結果を示している。

表3の発言番号3, 7より、短い発言(単語数が少ない)の場合、単語の学習データを用いて判別を行ったほうが良い傾向が得られた。発言番号1, 2, 4, 5, 6より、長い発言(単語数が多い)の場合、単語のペアの学習データを用いた判別を行ったほうが良い傾向が得られた。また、発言7のように、単語のペアができないような発言も多数存在するため、提案した手法のように、発言の長さを考慮する判別手法が有効であると思われる。

実験サンプル掲示板で検証した結果を図4に示す。横軸が出

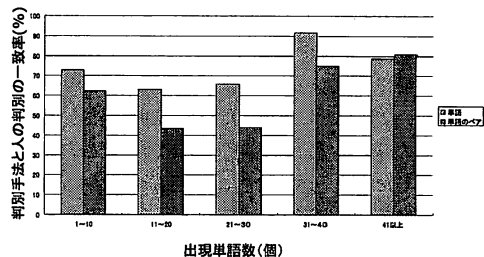


図4 単語数による学習データの違いによる判別一致率

現した単語数を示し、縦軸は人間の判断と発言の判別との一致率を示している。結果より、発言中の単語数が増えるにつれ、単語のペアの学習データを利用した場合の正答率が上がっていくという傾向が見える。単語の学習データを利用した場合は、単語数が増加してもある程度の正答率を保っている。誤判定が出た発言も複数存在しているため原因を調査したところ、形態素解析できなかった単語や学習データに未登録の単語が存在したためであった。今後、学習データを増やし、形態素解析できなかった単語の辞書登録などを行うことにより、回避できると思われる。

以上の実験より、実験サンプルで示した発言のタイプに対しては、長さを考慮した判別を行うことで間違った判別を減らすことが示された。また、掲示板全体の結果より、単語数が増加するにつれ、単語のペアの正答率が上昇していくことから、発言が長い場合、単語単体だけでなく、単語のペアの学習データも利用し、組み合わせることで判断することが必要である。

5. おわりに

本稿では、学校裏サイトにおける誹謗中傷発言を、発言の長さ(発言中の単語数)に応じて使う学習データを変化させるベイズ理論を応用した発言の判定手法を提案した。その結果、短い発言には、単語を学習データとして判別するほうが良い傾向が、また、長い発言には、単語のペアを学習データとして判別するほうが良い傾向が得られた。

文献

- [1] 文部科学省, <http://www.mext.go.jp/>
- [2] 大澤幸生, 松村真宏, 中村洋, “フレーミングは議論を阻害するか-2ちゃんねるは何故面白い?”, IEICE technical report. IA, Vol.102, No.143, pp.55-60, 2002.
- [3] 柴内康文, “言い争う-「フレーミング論争の検証」”, 現代のエスプリ, 川浦康至(編), vol.370, 至文堂, 1998.
- [4] 松村真宏, 三浦麻子, 柴内康文, 大澤幸生, 石塚満, “2ちゃんねるが盛り上がるダイナミズム”, 情報処理学会誌, vol.45, no.3, pp.1053-1061, 2004.
- [5] 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”, 情報の科学と技術, 56(10), pp.464-468, 2006.
- [6] 岩永学, 田端利宏, 櫻井幸一, “ベイジアンフィルタリングを用いた迷惑メール対策における多言語環境でのコーパス分離手法の提案と評価”, 情報処理学会論文誌, 46(8), pp.1959-1966, 2005.
- [7] 形態素解析エンジン “mecab”, <http://mecab.sourceforge.net/>