

電子掲示板における用語間関係を用いたトピックと 関係のない発言の除去手法

遠藤 崇史[†] 手塚 太郎[†] 木村 文則[†] 前田 亮[†]

[†]立命館大学 情報理工学部 〒525-0058 滋賀県草津市野路東 1-1-1

E-mail: cm002068@ed.ritsumei.ac.jp, fkimura@is.ritsumei.ac.jp, {tezuka, amaeda}@media.ritsumei.ac.jp

あらまし 現在ネット上の掲示板には、荒らし行為と呼ばれる場を荒らしてスレッドを読みにくくする嫌がらせや、ただ応答するといった内容のない発言が多く存在する。そこで本研究では、大規模なコーパスを用いることで発言におけるトピックとの関連度を推定し、関連性の弱い発言に対して除去を行い、掲示板の閲覧性を高める手法を提案する。実際の電子掲示板のデータを用いた評価実験によって提案手法の有効性を示した。

キーワード 電子掲示板, Wikipedia, 不要発言の除去

Filtering Method of Irrelevant Messages to the Topic on BBS Using Term Relationship

Takafumi ENDO[†] Taro TEDUKA[†] Fuminori KIMURA[†] Akira MAEDA[†]

[†] College of Information Science and Engineering, Ritsumeikan University

1-1-1 Nojihigashi, Kusatsu-shi, Shiga, 525-8577 Japan

E-mail: cm002068@ed.ritsumei.ac.jp, fkimura@is.ritsumei.ac.jp, {tezuka, amaeda}@media.ritsumei.ac.jp

Abstract On current BBS (Bulletin Board System) or discussion forums on the Internet, there are numerous irrelevant messages such as trolls (inflammatory, extraneous, or off-topic messages) and short responses with no useful information. In this paper, we propose a method to improve the readability of BBS by eliminating irrelevant messages to the topic. We estimate the relevance of a message to the topic by utilizing Wikipedia as a large corpus. Experiments using real BBS threads in Japanese showed the effectiveness of our proposed method.

Keyword BBS, Wikipedia, Filtering of Irrelevant Messages

1. はじめに

情報交換や意見交換などの手段として電子掲示板を利用することが広く行われている。そこでは匿名性が保たれていることから、気軽に利用することができるというメリットもあり、一日数百万件のアクセスがある2ちゃんねる[1]といった大規模なサイトも存在するに至っている。そのためそれらの掲示板に蓄積されている情報量は膨大であり、知りたい情報を探す上でも重要な情報源といえる。

特に掲示板によってもたらされる効果として、コミュニティは知識の重要な源となり、人間の創造的活動はグループでより発揮されやすく[2]、さらに、インタラクティブメディア上では少数者の意見が情報発信の機会を増加させる効果を持つことが報告されており、重要な少数意見も見つけ出すことができることが指摘されている。

また掲示板内の株式や投資といったビジネスについての記事に関しては、顧客同士のインタラクション

が新しい戦略の大きなヒントになり[2]、マーケティングの分野からの利用が期待できる。

しかし上記でメリットであるかのように述べた匿名性についても、これを悪意に利用し、故意にコミュニケーションを阻害する行為[3]や、悪意は無くとも内容や意味のないような発言を書き込むなど、容易に投稿できてしまうことで逆に掲示板内において不要とされる発言が多く存在することもまた事実といえる。ネット上には学校裏サイトと呼ばれる掲示板が多数存在し、そこでは学校ごとの地域において生徒や教師への不満や誹謗・中傷が書き込まれ、中には暴力的な言葉が多く存在することも報告されている[4]。

このように閲覧者にとって迷惑とされる発言の存在は利用者に多くの不利益をもたらす。特にトピックに基づいた情報収集のみを目的とした閲覧者にとっては、「なるほど」「了解」といった単なる応答や「すごい」「まずい」といった一言で感情を表す表現、「ぎゃあああ」といった奇声など、主語や目的が明確に述べ

られていない発言や、トピックと全くといっていいほど関係のない発言からは情報が得られないと考えられる。そこで本研究では、スレッド立案者の記述内容やアンカーで参照されている発言をトピックとして使用し、テーマであるトピックと関係性のない発言として、こうした内容のない発言や、トピックと全く関係のないことを述べている発言を不要発言とみなすこととした。本稿ではWikipediaをコーパスとして使用して関連語集合を取得し、その関連性から発言を除去することによって、結果として掲示板の閲覧性を高め、利用価値を向上させる手法を提案する。

2. 関連研究

電子掲示板に関する研究はいくつか行われている。本章ではその研究手法や違いについて説明する。

2.1. 不要発言に対する定義

電子掲示板における不要発言を除去する手法としてはいくつかのアプローチが見られるが、不要発言の定義においてはそれぞれに違いが見られる。一藤らは電子掲示板で発生する問題行為を荒らし行為とし、学校や地域ごとのコミュニケーションの場である学校裏サイトのような電子掲示板においては(1)多人数による罵り合い(2)閲覧者を不快にさせる書き込みの連続(3)荒らし行為を誘発させるような挑発的な書き込み(4)無用なコピー&ペーストの繰り返しによる閲覧の阻害といった発言を陰湿ないじめや社会問題に繋がるといった観点から不要発言とし、除去しようという試みを行っている[4]。

また掲示板内から主要書き込み人物を特定し、その人物との会話を必要発言として抽出を行っている研究もある[5]。これはその人物と絡みのない発言は不要とみなしていることになる。

このように不要発言の定義は読み手によって異なるため、ある発言が必要だという人もいれば不要だという人もいる。特に不要発言除去のアプローチの方法が多ければ除去されるパターンの種類は増える。しかし様々なアプローチにおいて、共通して不要とみなされる発言も存在する。本研究では掲示板のスレッドにおける主要なトピックに関連した情報が欲しいというユーザ層に焦点を当て、そのユーザ層にとって共通して不要とされる発言の除去を試みている。そしてその際の不要発言として、第1章で述べた内容のない発言やトピックと全くといっていいほど関係のない発言を用いる。よってたとえ学校裏サイトでは問題とされる中傷語であっても、本研究においては掲示板のトピックと関係があればそれは不要発言の対象外とみなす。

2.2. 不要発言の除去手法

一藤らは事前に人手によって不要語を特定し、発言内におけるその不要語の出現率から不要発言を推定するという手法を用いている。また推定の際に用いる方法としてベイズの定理を利用している。

本研究では各発言およびトピックに対して関連用語を推定し、その関連度を算出することで不要な発言を推定する方法を用いている。関連度の算出方法に関しては第3章で述べる。

3. 提案する不要発言の除去手法

3.1. トピックに対する定義

本研究では、発言が必要なものかを判断するために比較に用いる文章群として以下の3つのトピックを定義した。

□ スレタイトピック

掲示板のスレッドのタイトル(スレタイ)を指す。掲示板内において数あるスレタイの中からユーザが関心を持った内容。

□ 第一発言トピック

スレッド(あるいはスレ)を立ち上げた立案者の最初の発言を指す。スレッドの軸となる要素であり、ユーザはここからスレッドの全体像をイメージし、興味を持つことができれば閲覧を始めるものと考えられる。

□ アンカートピック

2ちゃんねるの場合「>>」という記号によって表されるアンカーと呼ばれる記号が存在する。これは発言のレス番号に対しリンク設定を行うことにより、そのレス番号の発言や人物に対して返信している[5]ということを表している。なお「レス」というのはレスポンスの略称で「返信」の意味の言葉を略したものである。「>>」は、この記号の後にスレッド内での発言の順番を指す発言番号を入力することで使用する。例えば、「>>20」と行えば、発言番号が20のレスに対して返信しているということになる。本研究においてもアンカーはアンカー先の発言に対しての意見だけでなく、アンカー先の発言をより詳細に説明するといった効果が期待できる点から、発言内にアンカーが含まれていれば、その発言に対してのみ「アンカートピック」として追加適用することとした。

3.2. 不要発言の判別手法

不要な発言かどうかを判別する方法として、本研究ではトピックと各発言との間の関連度を求める。関連度は発言ごとに付与され、トピックと各発言がもつ関

連用語から推定される。本節ではその関連用語の選出から関連度の算出に至るまでの具体的な手法について説明する。

3.2.1. 大規模なコーパスの利用

大規模な文書集合であるコーパスを使用し、あらかじめ各単語に対する関連語を求めておき、どれだけ多くの関連語を含むかによって発言間の関連度の推定を行う。本システムでは広く使用されているオンライン百科事典であるWikipedia[6]をコーパスとして用いた。

Wikipediaは記事の集合によって構成されている。各記事には項目名となる用語があり、記事内ではその用語についての説明が述べられている。Wikipediaは全世界の多数の言語における版が存在しているが、日本語版においても2009年7月時点で50万を超える膨大な記事数を所有している。そのため多種多様に渡る用語を網羅しており、このボキャブラリーの多さは掲示板のような様々な分野に関する話題が展開されている場においても、その出現用語に対して同様に多くの網羅が期待できると考えられる。

3.2.2. 用語間関係の抽出方法

本項では、用語間関係の抽出にあたってWikipediaをコーパスとしてどのように用いるかについて説明する。まず本研究のタイトルでもある用語間関係についてであるが、ここでいう「用語」の定義としてWikipediaの項目名を使用する。3.2.1でも述べたが、Wikipediaには各記事に対して項目名が存在し、記事内ではその項目名の意味を説明するという構成になっている。そして記事内に着目すると、そこにはその記事の項目名以外にも、他の記事で項目名となっている用語が存在する。これが項目名に対する関連用語となる。

以下の表1を例にあげると、例えば項目名が「夕焼け」の記事内には「日没」、「現象」といった用語が存在する。これはWikipediaの他の記事において「夕焼け」と同様、項目名となっている用語である。そしてこれらが関連用語であり、この用語間の繋がりが、ここで述べている用語間関係となる。

表1: 項目名に基づく用語間関係

項目名	記事に含まれる項目名
夕焼け	日没, 現象, 波長, 地平線, ...
競馬	騎手, 馬, スポーツ, ...
時計	時間, 機械, 電気, ぜんまい...
携帯電話	電話機, バッテリー, ...
ペットボトル	容器, 飲料, コーラ, ...

3.2.3. トピックとの関連度の算出

本項では前項で説明した用語間関係を基に導き出された関連用語からどのようにして関連度を算出するかについてさらに詳しく説明する。

まず発言およびトピックからそこに含まれる項目名を抽出し、そこから関連用語を推定する。そして発言とトピックの間でお互いがどの程度同じ関連用語を持っているかの割合によって関連度を算出する。その式は以下のようになる。

$$\text{関連度} = \frac{\text{トピックと発言が共に持つ関連用語数}}{\text{発言の持つ関連用語数}} \times 100$$

発言が持っている関連用語の延べ数を分母とし、その関連用語の中でトピックにも含まれる用語の合計数を分子とした。

なお、発言とトピックとの間に関連性がないと判断する基準値として、関連度に対して閾値を設けるものとした。つまり閾値を下回った発言、すなわち関連度の低い発言がトピックと関係のない不要発言となる。

4. システム概要

本章では実験で使用したシステムの概要図を図1に示し、その仕組みについて説明する。

まずスレッド内からトピックとなる要素を抽出する。一般的にはスレタイと最初の発言のみとなるが、3.1で述べたように判別対象の発言がアンカーを含んでいた場合、そのアンカー先の発言もトピックに加えるものとしている。続いてトピックと各発言をそれぞれ形態素解析にかけて項目名を抽出し、データベースに問い合わせることでその関連用語を抽出している。そして3.2.3に基づき、トピックと各発言における関連度を推定する。

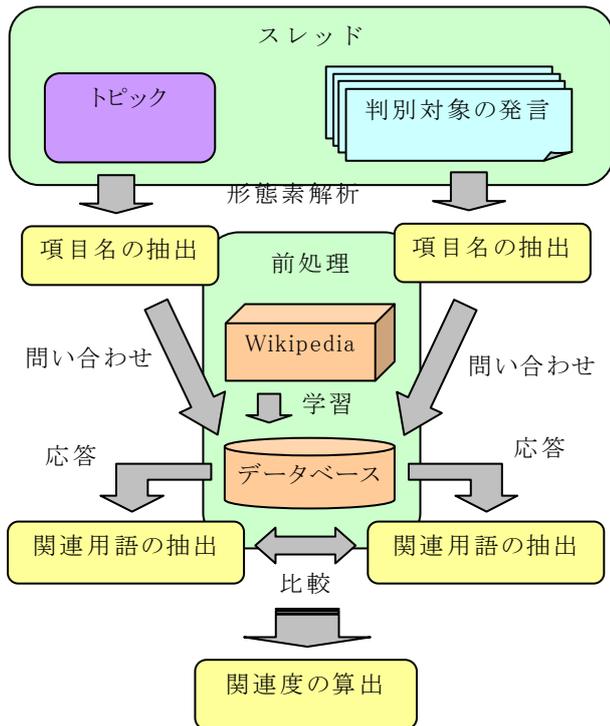


図1：システムの概要図

4.1. データベースの構築

アクセス時間の短縮や連続アクセス防止のため、前処理として事前にWikipediaの記事群から用語間関係を抽出し、3.2.2で述べた要領で項目名ごとに関連用語を推定し、データベースに格納した。これにより各項目名は、データベースに問い合わせることで自身の関連用語を速やかに抽出できるようになる。また登録する項目名の条件として以下の2つのパターンを用いた。第5章ではこのパターンを用いたシステム検証を行っている。

□ パターン①

全ての項目名をそのまま使用する。

□ パターン②

記号・数字のみで構成されている、または人称代名詞・指示代名詞となる語を除外する。

4.2. 形態素解析

項目名を抽出するために形態素解析を行うが、その解析にはMecabを使用している。またデフォルトの辞書をそのまま使うのではなく、実験で使用する項目名のみに登録し直すことにより、適切に項目名だけの抽出を行えるようにした。

5. 実験と考察

5.1. スレッドの選定

実験の対象とする掲示板としては2ちゃんねるを用いた。第1章で述べたように一日数百万件のアクセスがあるということ、また全体的に荒れた掲示板が多く、不要発言の出現割合が多いことから本手法の対象として適している。

そしてスレッドの条件としては、タイトルトピックおよび第一発言トピックから読み取れる内容が明確化されており、また除去された不要発言が実際に不要であるかどうかの判定を行う評価者が知識を持つ分野から6スレッド(A~Fとする)を選択した。そもそも本研究はトピックとの関連性のない発言を除去するというものであるため、そのトピックを構成するはずのスレタイやスレッド立案者の最初の発言からトピックがイメージできないもの、不要かどうかの判定を行う人間があまり詳しくないテーマの話題に対しては対象外とみなした。

5.2. システムの検証

4.1で述べたパターン①、②を用いたシステムをそれぞれ提案手法1、提案手法2とし、システムの検証を行った。

5.2.1. フィルタリング性能の評価

システムのフィルタリング性能について評価を行った。3.2.3で説明したように、本システムは関連度が閾値を下回った発言に対しフィルタリングを行うというものである。そこでまず各発言に対し、システムが出力した関連度を基に昇順に並べた。言い換えるとフィルタリングされる順番に並べたことになる。そして提案手法1、2のシステムが実際にスレッド内の発言を一つずつフィルタリングしていくにつれ、そのスレッド内の不要発言がどのタイミングでフィルタリングされていくかについての経過を表している。その際、同じ関連度に不要発言と不要でない発言が含まれる場合はその出現数の割合に応じて並べるものとしている。ここでは代表例としてスレッドBにおける経過を以下の図2に示し、これを基に解説する。(青が提案手法1、赤が提案手法2となっている。)

なお、このスレッドA~Fに関しては5.1で説明した基準を基に選んだものである。

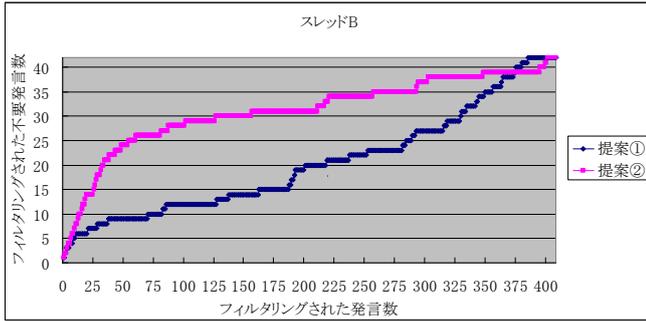


図2：スレッドBに関するフィルタリング性能評価

□ 提案手法 1 に関して

「不要な発言」と「不要でない発言」がほぼ均等な割合でフィルタリングされてしまっている。これは他のスレッドにおいてもほぼ同様の結果となった。これではランダムに発言を除去するのと特に変わらず、システムのフィルタリング性能としては、不十分な結果である。

□ 提案手法 2 に関して

「不要でない発言」に対し「不要な発言」の方が多くフィルタリングされるという結果が出た。こちらも他のスレッドにおいてほぼ同様の結果となった。不要でない発言もフィルタリングされてしまっているという点では不十分ではあるが、「不要な発言」の関連度が「不要でない発言」の関連度よりも全体的に低いという結果を表している点に関しては良いといえる。

この実験では、スレッド内における発言数と不要発言数から提案手法の評価・比較を行った。しかしこれらの数はスレッドによって異なるため、全体的な比較を行う上で不便である。そこで再現率、誤フィルタリング率という評価指標を定義し用いることでスレッド間での統一を行った。これにより上記の違いに囚われることなくスレッド間の比較ができ、より視覚的にわかりやすい評価が可能になる。ここで定義した再現率、誤フィルタリング率という指標について説明する。

□ 再現率

人手によって不要であるとみなされた発言が、システムによってどの程度フィルタリングできたかの割合。

□ 誤フィルタリング率

人手によって不要でないとみなした発言に対し、システムが誤ってフィルタリングしてしまった発言の割合。

スレッドBの提案手法 2（図2を参照）を例にとって

説明する。

例えばフィルタリングされた発言数が200個の時点において、システムは人手によって不要であるとみなされた発言42個に対し、31個をフィルタリングできたのでその再現率は73.81%ということになる。一方、人手によって不要でないとみなされた発言367個に対し、システムは誤って169個の発言をフィルタリングしてしまったため誤フィルタリング率は46.05%となる。

つまり再現率の増加につれ、誤フィルタリング率も増加していくということになる。この誤フィルタリング率の増加に伴い、どのように再現率が増加していくかという経過を追ったのが図3～図5である。これにより下限～上限が0～100%で表され、図2で示したフィルタリングされた発言数と不要発言数の間における関係性を崩さずに各スレッドが統一された。そのためスレッドの発言数の量に関係なく比較できるようになった。

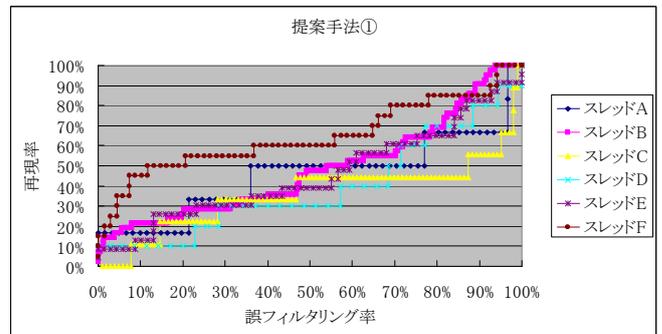


図3：提案手法 1 における再現率と誤フィルタリング率の関係

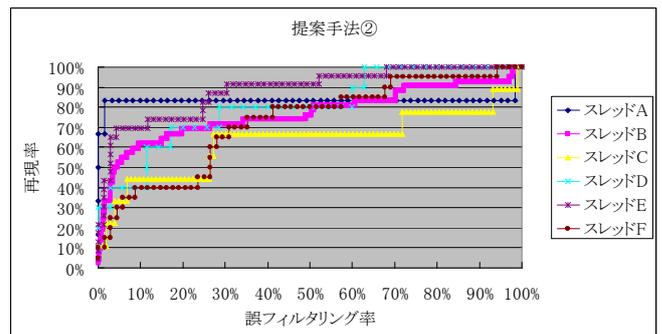


図4：提案手法 2 における再現率と誤フィルタリング率の関係

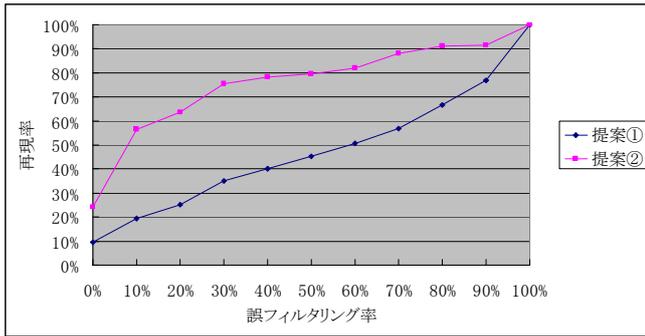


図 5：提案手法 1 と 2 の平均値による比較

まず図3～図4から読み取れることとして、提案手法 1 は誤フィルタリング率の増加に対する再現率の増加が全体的にほぼ一定であるのに対し、提案手法 2 は誤フィルタリング率が低い段階から全体的に再現率が大きく上昇していることがわかる。これはやはり先ほど述べた「不要でない発言」に対し「不要な発言」の方が割合多く優先的にフィルタリングされているということを裏付けている。特にスレッド数の多いBのスレッドからはその様子がよくわかる。

また精度に関してはAのスレッドで最も良い結果がでている。ただシステムにとっての最良の結果があくまで「誤フィルタリング率0%」での「再現率100%」であることから判断するとまだ到達できていない。

次に図5についてであるが、これは提案手法 1, 2 において各スレッドの誤フィルタリング率10%ごとに対する再現率の平均値を取り、出力したものである。ここからわかることとして、同じ誤フィルタリング率の地点において、提案手法 2 が提案手法 1 を明らかに上回っていることがわかる。これは提案手法 1 に対して提案手法 2 の方が優れているということを裏付けている。

そこで提案手法 1, 2 においてどの程度それが実現できているのかを検証した。図3～図4を基に、誤フィルタリング率10%ごとの再現率を抽出した結果を表2～表3に示す。

表 2：提案手法 1 における誤フィルタリング率と再現率の関係性

提案手法1						
	A	B	C	D	E	F
誤フィルタリング率	再現率					
0.00	0.17	0.02	0.00	0.10	0.04	0.05
0.10	0.17	0.21	0.11	0.10	0.13	0.45
0.20	0.17	0.26	0.22	0.10	0.26	0.50
0.30	0.33	0.30	0.33	0.30	0.30	0.55
0.40	0.50	0.40	0.33	0.30	0.35	0.60
0.50	0.50	0.50	0.44	0.30	0.39	0.60
0.60	0.61	0.52	0.44	0.40	0.52	0.65
0.70	0.50	0.55	0.44	0.50	0.61	0.80
0.80	0.67	0.69	0.44	0.70	0.65	0.85
0.90	0.67	0.90	0.56	0.80	0.83	0.85

表3:提案手法 2 における誤フィルタリング率と再現率の関係性

提案手法2						
	A	B	C	D	E	F
誤フィルタリング率	再現率					
0.00	0.67	0.05	0.11	0.30	0.22	0.10
0.10	0.83	0.62	0.44	0.40	0.70	0.40
0.20	0.83	0.69	0.44	0.70	0.74	0.40
0.30	0.83	0.71	0.67	0.80	0.87	0.65
0.40	0.83	0.74	0.67	0.80	0.91	0.75
0.50	0.83	0.76	0.67	0.80	0.91	0.80
0.60	0.83	0.81	0.67	0.80	0.96	0.85
0.70	0.83	0.83	0.67	1.00	1.00	0.95
0.80	0.83	0.90	0.78	1.00	1.00	0.95
0.90	0.83	0.93	0.78	1.00	1.00	0.95

□ 誤フィルタリング率10%時点

提案手法 1 に関しては、スレッドFを除いて再現率を10～20%程度しか得ることができなかった。それに対し提案手法 2 に関しては再現率を最低4割、スレッドAでは8割強を得ることができた。

□ 誤フィルタリング率90%時点

提案手法 1 に関しては高いもので9割、スレッドCに至っては5割強しか再現率を得ることができなかった。一方の提案手法 2 に関しては、スレッドD, Eにおいて誤フィルタリング率が90%以前の70%の時点で再現率100%を得ることができている。提案手法 1 で5割強しか再現率が得られなかったスレッドCに関しても8割弱と増加した。

5.2.2. 発言と関連度の関係性の評価

発言と関連度の関係性を調べるため、各スレッドにおける関連度に対する発言数の分布図を作成した。以下に代表例としてスレッドBの分布図6を示し、考察する。

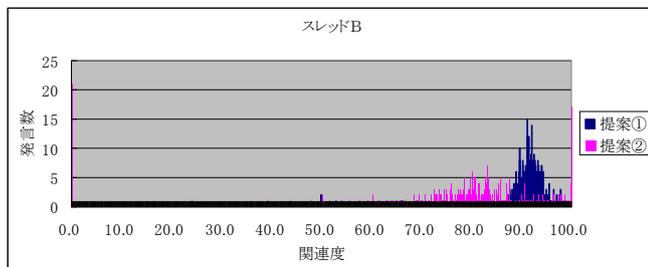


図6：スレッドBに関する関連度と発言の分布

□ 関連度0の発言の存在

まず関連度0の発言が存在しているのがわかる。これは特に提案手法2に至っては他のどのスレッドに関しても同様に存在していた。これは発言内に項目名が存在しなかったということを示している。基本的に内容のない不要発言はここに含まれるため、誤フィルタリング率が低い段階での急激な再現率の向上はこれが大きな要因といえる。

しかし逆に、ここに含まれるのが全て不要発言ではないという問題もある。通常、掲示板内では用語を正しく正式名称で書き込まず、略称を用いることが多い。Wikipediaにおいてもその記事数の多さからいくつかの項目名に関する略語は存在しているが、全てではない。そのため略語の検出漏れという事態が発生する。その他にも発言者が用語を打ち間違えるということもある。こうした曖昧語の問題解決は「不要でない発言」に対する関連度の増加が期待できるため、今後改善する必要がある。

□ 関連度100の不要発言の存在

逆に関連度100に不要な発言が含まれていることも確認できる。これはトピックを構成する上でほとんど必要性のない項目名のみが発言とマッチしたものと考えられる。こうした問題は発言に含まれる項目名が少ないほど起こりやすいといえる。そこで今後、関連用語の少ない発言についてはその項目名の重要性を考慮するなどの判断が必要になる。

□ 関連度に対する発言の分布状況

提案手法1に関しては関連度がどの発言も90付近の高い地点に密集するという結果が出た。一方の提案手法2では関連度の平均値付近を中心に比較的穏やか

な山なりが形成された。この違いは4.1で述べたデータベースを構築する際に、パターン②を適用するにあたって除外した記号などの関連用語がWikipedia内かなりの割合で含まれていたということを示している。提案手法1において全体的に発言の関連度がかなり高くなっているのはそのためといえる。

本研究においてフィルタリングは関連度に対して閾値を引くことで行うため、その関連度の低い順に発言が不要とみなされていくわけであるが、そもそも発言数に対する不要発言の割合はスレッドごとに異なる。そのため発言間の関連度の開き具合は発言の重要性をより具体化し、フィルタリングの際の閾値を決定する判断材料になる。さらに関連度が近似したもの同士を集合体として考えるなど、閾値の判断に対して応用した方法もとりやすくなるといえる。つまり提案手法2は提案手法1に比べて発言に対する関連度の分布が大きく見られるため、その観点において優れているとみなすことができる。

5.3. 考察

フィルタリング性能については提案手法1に関してはあまり良くない結果が出たため、この状態での実用性はほぼないと考える。一方そこから「! “ # \$」といった記号や、「オレ」「彼」といった人称代名詞、「これ」「あっち」といった指示代名詞などを除外して検証を行った提案手法2については、「不要でない発言」に対し、「不要な発言」を比較的多くフィルタリングできたという点で悪くない結果が出たとみている。しかし5.2.1でも述べたが、本研究の最終目的はあくまで「誤フィルタリング率0%」で「再現率100%」を実現すること、すなわち「不要な発言」のみを全て除去しようというものであるため、比較的良いとは言え、「不要でない発言」が除去されてしまっている状況は決して良いとは言えない。

関連度と発言における分布関係からは、提案手法1に対し提案手法2の方が全体に広く散らばっており、閾値の判断に適していると考えられる点から優れているとみなせる。しかし、不要な発言の中には関連度100という数値を記録しているものが存在するという問題が残った。

今後システムを改善する手段として、5.2.2で述べた曖昧語の問題の解決や項目名に対して重要性を考慮するといった方法が必要になると思われる。

6. まとめ

本論文では掲示板においてスレッドのテーマをトピックとし、大規模なコーパスから用語間関係を推定し用いることでそれと関係性のない発言を掲示板から

除去する手法を提案した。用語間関係の推定には Wikipedia の項目名を用語として用い、その記事内に出現した用語がその関連用語であるとした。

実験結果からは、提案手法 1 に関しては不十分な結果となったが、もう一方の提案手法 2 に関しては、不要発言のフィルタリングという点において比較的よい結果が出たといえる。

今後の課題としては、5.3 で述べた曖昧語が正式な名称と認識されないことが精度の減少を招くと考えられる問題や、またトピックに存在する用語がどの程度トピックを構成する上で重要であるのか考慮されていない問題などについて、その解決法を探る必要がある。

謝辞

本研究の一部は文部科学省私立大学戦略的研究基盤形成支援事業「芸術・文化分野の資料デジタル化と活用を軸とした研究資源共有化研究」の支援を受けている。

参考文献

- [1] 2ちゃんねる, <http://www.2ch.net/>
- [2] 松尾豊, 大澤幸生, 石塚満, “電子掲示板における会話からのトピックの発見と要約”, 第 16 回人工知能学会全国大会論文集, pp206-209, 2002
- [3] 一藤裕, 今野将, 曾根秀昭, “BBS における発言の相関に基づいたコミュニケーション阻害行為の発見”, 電子情報通信学会技術研究報告. SITE, 技術と社会・倫理, pp.17-21, 2007
- [4] 一藤裕, 今野将, 曾根秀昭, “掲示板の発言に対する自動判別を用いたユーザ教育支援手法の改良”, 電子情報通信学会技術研究報告. SITE, 技術と社会・倫理, pp.219-223, 2009
- [5] 田村幸寛, 砂山渡, “電子掲示板の主要人物に関するストーリー要約”, 第 23 回人工知能学会全国大会, 1C3-1, 2009
- [6] Wikipedia(ja), <http://ja.wikipedia.org/wiki/>
- [7] 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