

Q&A サイトで複数のアカウントを不正に用いるユーザの検出: 複数の回答者であるかのように装うユーザについて

Detection of Illegal Multiple Account Users in a Q&A site: Multiple Account Users Who Submitted Multiple Answers to a Question Repeatedly

石川 尚季 (Naoki ISHIKAWA)¹・梅本 顕嗣 (Kenji UMEMOTO)²・西村 涼 (Ryo NISHIMURA)³・
渡辺 靖彦 (Yasuhiko WATANABE)⁴・岡田 至弘 (Yoshihiro OKADA)⁵
^{1 2 3} 龍谷大学大学院 理工学研究科 情報メディア学専攻 修士課程・
⁴ 龍谷大学 理工学部 情報メディア学科 講師・⁵ 龍谷大学 理工学部 情報メディア学科 教授

[Abstract] Some users in a Q&A site abuse the anonymity and attempt to manipulate communications in the Q&A site. These users and their submissions discourage other users, keep them from retrieving good communication records, and decrease the credibility of the Q&A site. To solve this problem, we conducted an experimental study to detect users suspected of using multiple user accounts and submitting multiple answers to a question repeatedly.

[キーワード] 多重アカウント、Q&A サイト、なりすまし、信頼性

1. はじめに

Q&A サイトにおける質問や相談などのコミュニケーションでは匿名であることが重要である。匿名のコミュニケーションであることで、照れ、恥かしさ、評判などを気にしないで質問、相談、回答、議論などを行うことができるからである。しかし、匿名性を悪用して、Q&A サイトにおけるコミュニケーションを妨害しようとするユーザもいる。したがって、Q&A サイトにおけるコミュニケーションを促進するためには匿名性を悪用するユーザを検出することが重要である。

匿名性の悪用の1つに、1人のユーザが複数のアカウントを利用してコミュニケーションを操作しようとするものがある。オンラインメッセージのリンク可能性[1]に着目し、機械学習および統計的手法によって著者を識別する方法はこれまで数多く研究されてきた[2][3][4][5][6][7]。しかし、複数のアカウントを利用してQ&A サイトでのコミュニケーションを操作しようとするユーザを検出する方法はまだあまり研究されていない。そこで石川らは、複数のアカウントを用いて質問とその回答を繰り返し投稿し、回答者としての自分の評価を操作しようとするユーザを、その投稿履歴と文体の類似判定を手がかりに検出する方法について検討した[8]。本研究では、複数のアカウントを用いて1つの質問に複数の回答を投稿し、自分の意見や考えを正当化・擁護しようとしたり、議論を妨害しようとすることを繰り返すユーザを検出する方法について検討する。

複数のアカウントを利用して不適切な投稿を行うユーザについての情報(人数、目的、方法など)は公開されていない。このため、提案手法の精度を議論することはむずかしい。そこで本研究では実験結果を詳しく示し、提案手法の有効性について議論する。本研究で検出された疑わしいユーザについての情報は、複数のアカウントを利用してQ&A サイトでのコミュニケーションを操作しようとするユーザの目的やふるまいを検討する手がかりになると考えられる。

¹ t10m096@mail.ryukoku.ac.jp

² t11m074@mail.ryukoku.ac.jp

³ r_nishimura@afc.ryukoku.ac.jp

⁴ watanabe@rins.ryukoku.ac.jp

⁵ okada@rins.ryukoku.ac.jp

表-1 Yahoo!知恵袋の「パソコン、周辺機器」、「病気、症状、ヘルスケア」、「政治、社会問題」カテゴリにおける質問および回答の投稿者の異なりと投稿件数。 U_{qst} と U_{ans} はそれぞれ、質問者と回答者の異なり。 N_{qst} と N_{ans} はそれぞれ、質問と回答の投稿件数。それぞれのユーザについて一緒に回答した質問の数が最も多いユーザを調べ、その件数を合計したものが N_{mfe} である。

カテゴリ	U_{qst}	N_{qst}	U_{ans}	N_{ans}	N_{mfe}
パソコン、周辺機器	43493	171848	27420	474687	67846
病気、症状、ヘルスケア	29954	84364	38223	289578	54949
政治、社会問題	13259	78777	25766	403306	74781

2. 複数のアカウントを利用したコミュニケーションの操作

本研究では、Q&A サイトの例としてYahoo!知恵袋^(注1)を取り上げる。Yahoo!知恵袋は、登録されたユーザが質問と回答のメッセージをやりとりすることができるQ&A サイトである。このYahoo!知恵袋に2004年4月から2005年10月までに投稿された質問約311万件、回答約1347万件のメッセージが国立情報学研究所から公開されている^(注2)。この公開データに収録されている質問には1件以上の回答が必ずあり^(注3)、そのうち1件が質問者によってベストアンサーに選ばれている。このYahoo!知恵袋のデータの「パソコン、周辺機器」「病気、症状、ヘルスケア」「政治、社会問題」のカテゴリに質問あるいは回答を投稿したユーザの異なりと投稿件数を表1に示す。さらにそれぞれの回答者について、一緒に回答した質問の数が最も多い回答者とその件数も調べた。例えば、ユーザ691911が「政治、社会問題」のカテゴリで一緒に回答した質問の数が最も多いのはユーザ802184で、その件数は47件であった。各カテゴリのすべての回答者について一緒に回答した質問の数が最も多いユーザを調べ、その件数を各カテゴリで合計したものが表1の N_{mfe} である。したがって、「政治、社会問題」のカテゴリであるユーザが回答を100件投稿しているとする、そのユーザにとって一緒に回答した質問の数が最も多いユーザとは

$$\frac{N_{mfs}}{N_{ans}} \times 100 = \frac{74781}{403306} \times 100 = 18.5$$

件の質問と一緒に回答していると期待できる。ユーザ691911は「政治、社会問題」のカテゴリの質問54件に回答していて、そのうち47件はユーザ802184と一緒に回答しているので、これは期待値よりもかなり多い。

Yahoo!知恵袋では、質問と回答を匿名で投稿することができる。匿名での投稿は、照れ、恥かしさ、評判などを気にしないで質問、相談、回答、議論などを行うことができるので、Q&A サイトでのコミュニケーションを促進するには重要である。しかし、匿名で投稿した質問あるいは回答であってもアカウントによってリンク可能であるので、投稿を重ねれば個人を識別できる情報が蓄積されるおそれがある。このため、コミュニケーションを操作することが目的ではなく、個人を識別されないために複数のアカウントを利用しているユーザも少なくないと考えられる。一方、コミュニケーションを操作することが目的で、以下に示すTYPE QAやTYPE AAなどの投稿を行うために複数のアカウントを利用しているユーザもいると考えられる。

TYPE QA 1人のユーザが複数のアカウントを利用して質問とその回答を投稿する。このタイプの投稿は、回答している自分の評価を高めることが目的と考えられる。例えば、Yahoo!知恵袋では1つの質問に対して必ず1つの回答が質問者によってベストアンサーに選ばれる。自分が投稿した回答をベストアンサーにするため、このタイプの投稿が行われている可能性がある。

TYPE AA 1人のユーザが複数のアカウントを利用して1つの質問に対して複数の回答を投稿する。このタイプの投稿は、コミュニケーションが行われている場の支配や妨害を目的としていると考えられる。具体的には、

(注1) <http://chiebukuro.yahoo.co.jp>

(注2) <http://research.nii.ac.jp/tdc/chiebukuro.html>

(注3) 回答が投稿されなかった質問は削除されているので、この公開データに収録されている質問には1つ以上の回答が必ずある。

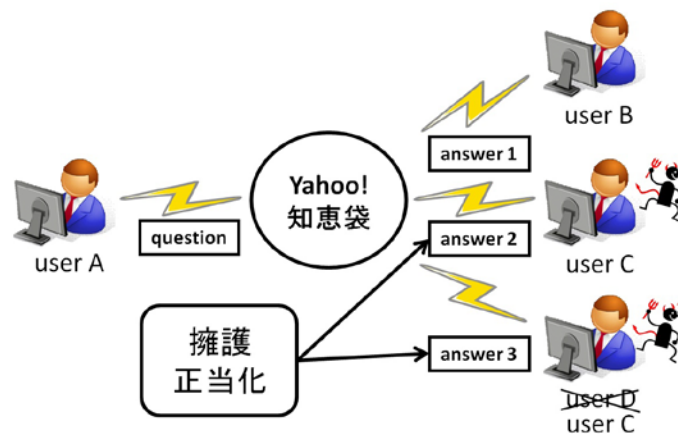


図-1 TYPE AA の投稿の例: 1 人のユーザ(user C)が1つの質問に対し複数の回答を複数のアカウントを利用して投稿し、自分の意見や考えを正当化・擁護しようとしたり、議論を妨害しようとしている。

- 自分の意見を正当化・擁護して、議論の場を支配する
 - ふさわしくない発言をくりかえして、議論の秩序を乱す
- ことなどが目的と考えられる(図1)。

こうした投稿を繰り返すユーザがいると、まじめに議論しようとするユーザは新たにメッセージを投稿しようとする意欲を失い、結果としてそのQ&Aサイトは利用されなくなるおそれがある。また、コミュニケーションが操作されたQ&Aサイトの質問と回答の内容は信頼できない。これは、Q&Aサイトに投稿された質問と回答を知識として2次利用する場合に問題になる。したがって、コミュニケーションを操作しようとするユーザを検出することは重要である。

これまでに石川らは、Yahoo!知恵袋でTYPE QAの投稿を繰り返すユーザ、すなわち、複数のアカウントを利用して質問とその回答を繰り返し投稿し、回答者としての自分の評価を操作しようとするユーザを、その投稿履歴と文体の類似判定を手がかりに検出する方法について検討した[8]。本研究では、TYPE AAの投稿を繰り返すユーザ、すなわち、同一人物が投稿したとはわからないように複数のアカウントを利用して複数の回答を1つの質問に対して投稿し、自分の意見や考えを正当化・擁護しようとしたり、議論を妨害しようとすることを繰り返すユーザを検出する方法について検討する。

複数のアカウントを用いて1つの質問に複数の回答を投稿することを繰り返すユーザがいるのなら、

(予想1) あるユーザが回答した質問に、特定のユーザも回答していることが異常に多い

ことが考えられる。しかし、特定のトピックに関する質問に熱心に回答を投稿するユーザもいるので、そうしたユーザとTYPE AAの投稿を繰り返すユーザを区別するため、

(予想2) 1人のユーザが複数のアカウントを利用して投稿した回答の文体は類似する

(予想3) 評価が公平であるなら、1人のユーザが複数のアカウントを利用して投稿した回答への評価には差がない

ことも考慮する。そこで、

- 一緒に回答した質問が非常に多いという異常な投稿を行うユーザのペアの検出
- 文体の類似判定
- 回答への評価の比較

によって複数のアカウントを利用して1つの質問に複数の回答を投稿し、自分の意見や考えを正当化・擁護しようとしたり、議論を妨害しようとした疑いのあるユーザを検出する方法について検討する。

3. 複数のアカウントを用いて1つの質問に複数の回答を投稿しようとするユーザの検出

表-2 入力として与えられた回答の文体が特定のユーザの文体に類似しているか判定するために用いる素性

素性	説明
S1	すべての文に含まれる形態素
S2	それぞれの文に含まれる形態素とその文番号
S3	すべての文から取り出した文字3-gram
S4	それぞれの文から取り出した文字3-gramとその文番号
S5	それぞれの文の文頭の1~10文字
S6	それぞれの文の文末の1~10文字
S7	PrefixSpanを利用して取り出した、出現回数5回以上、アイテム数3以上、最大ギャップ数1、最大ギャップ長1の系列パターン

不適切な内容のメッセージが投稿されるのを抑制するため、利用者にユーザ登録することを求める Q&A サイトが多い。しかし、ユーザ登録の情報は、不適切な内容のメッセージを投稿するユーザを検出するには有効であるが、複数のアカウントを用いて他人になりすまし、コミュニケーションを巧妙に操作しようとするユーザを検出する手がかりとしては不十分である。そこで、投稿履歴と文体の特徴を手がかりに、

- 異常な投稿を行っているユーザのペアの検出
- 検出されたユーザペアに対する文体の類似判定
- 検出されたユーザペアに対する回答への評価の比較

を行い、複数のアカウントを用いて1つの質問に複数の回答を投稿した疑いのあるユーザを検出する。

3.1 異常な投稿を行っているユーザのペアの検出

ユーザ i が、ユーザ j の回答した質問に対して異常に多くの回答を投稿していないかを、以下の仮説 AA を用いて判定する。

仮説 AA ユーザ j の回答した質問に対して、ユーザ i が異常に多くの回答を投稿していないならば、ユーザ i はユーザ j の回答した質問に $N_{AA}(i)$ 回の回答を投稿していると期待できる。

$$N_{AA}(i) = \frac{N_{mfe}}{N_{ans}} \times ans(i)$$

ここで $ans(i)$ とはユーザ i によって対象カテゴリに投稿された回答の数で、 N_{ans} とは対象カテゴリに投稿されたすべての回答の数である。また、対象カテゴリのすべてのユーザについてそれぞれ一緒に回答した質問の回数が最も多いユーザを調べ、その回数を合計したものが N_{mfe} である (表 1)。

もしこの仮説が片側二項検定で棄却されれば、ユーザ i はユーザ j の回答した質問に異常に多くの回答を投稿していると判定する。

仮説 AA を同じ質問に回答を投稿したすべてのユーザのペアに適用して、仮説が棄却されたユーザペアを異常な投稿を行ったユーザペアとして取り出す。

3.2 文体の類似判定

異常な投稿を行っているとして検出されたユーザペアに対して文体の類似判定を行う。文体の類似判定には以下の仮説 WS を用いる。

仮説 WS ユーザ i はユーザ j よりも多くの回答を対象カテゴリに投稿しているとする。また、分類器 $classifier(i)$ は入力された回答の文体がユーザ i の文体に類似しているかどうか判定する分類器とする。このときもし、ユーザ i とユーザ j の文体が類似していないならば、ユーザ j の回答の文体がユーザ i の回答の文体と類似していると分類器 $classifier(i)$ が判定するのは N_{WS} 回であると期待できる。

$$N_{WS} = (1 - P_{WS}) \times m$$

ここで P_{WS} は分類器 $classifier(i)$ の分類精度で、 m は文体の類似判定のために $classifier(i)$ に与えられたユーザ j の回答の件数である。

もしこの仮説が片側二項検定で棄却されれば、ユーザ i とユーザ j の文体は類似していると判定する。

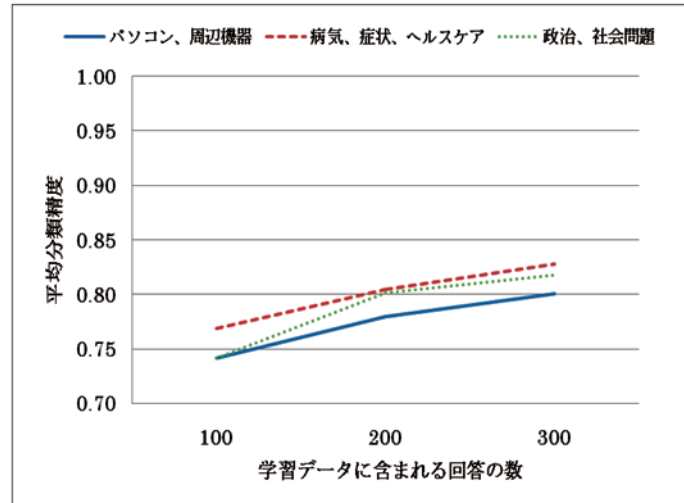


図-2 対象カテゴリで質問を 10 件以上、回答を 160 件以上投稿したユーザ 30 人の回答を対象に作成したユーザ分類器の平均分類精度。学習には、100 件(対象ユーザ 50 件とその他 50 件)、200 件(対象ユーザ 100 件とその他 100 件)、300 件(対象ユーザ 150 件とその他 150 件)の回答からなる学習データを用いた。

入力された回答の文体が対象ユーザであるユーザ i の文体と類似しているかどうかを判定する分類器 $classifier(i)$ は、ユーザ i の文体に対する最大エントロピー法による学習にもとづいて作成する。学習に用いるデータは、

- 対象ユーザが対象カテゴリに投稿した m 件の回答から無作為に取り出した n 件の回答
- 対象ユーザ以外のユーザが対象カテゴリに投稿した回答から無作為に取り出した n 件の回答

の合計 $2n$ 件の回答から構成されている。学習に用いる素性を表 2 に示す。S1 と S2 は回答を形態素解析した結果得られた形態素を素性に利用している。ただし、S1 ではどの文に含まれる形態素であるか区別しないが、S2 では回答のどの文に含まれる形態素であるか区別する。S3 と S4 で文字 n -gram として 3-gram を使用しているのは、小高らが文字 n -gram のうち日本語には 3-gram が効果的であると報告しているからである [9]。ただし、S3 ではどの文に含まれる文字 3-gram であるか区別しないが、S4 では回答のどの文に含まれる文字 3-gram であるか区別する。S5 と S6 でそれぞれの文の文頭および文末の 1~10 文字を素性としているのは、回答の著者それぞれの特徴的な表現が文頭および文末に現れることが多いと考えたからである。S7 は PrefixSpan^(注4) を利用して取り出した素性である。本研究では、出現回数 5 回以上、アイテム数 3 以上、最大ギャップ数 1、最大ギャップ長 1 の形態素の系列パターンを PrefixSpan を用いて取り出し、S7 の素性とした。

分類器の分類精度 P_{WS} を測定するため、Yahoo! 知恵袋の「パソコン、周辺機器」「病気、症状、ヘルスケア」「政治、社会問題」のカテゴリを対象に実験を行った。各カテゴリにおける対象ユーザは、質問投稿数が 10 件以上、回答投稿数が 160 件以上のユーザ 30 人とした。実験では最初に、各ユーザの回答から 10 件を無作為に取り出し、評価用データとした。次に、

- 対象ユーザが対象カテゴリに投稿した回答(評価用データとして取り出した回答 10 件は除く)から無作為に取り出した n 件の回答
- 対象ユーザ以外のユーザが対象カテゴリに投稿した回答から無作為に取り出した n 件の回答

の合計 $2n$ 件の回答 ($n = 50, 100, 150$) から構成される 3 種類の学習用データを作成した。この学習データに対して最大エントロピー法による学習を行った。最大エントロピー法による学習には maxent^(注5) を用いた。また、形態素解析には Mecab^(注6) を用いた。作成した分類器の分類精度を図 2 に示す。図 2 に示すように、300 件の回答(対象ユーザの回答 150 件を含む)から構成される学習データを用いれば、どのカテゴリでも 80% 以上の精

(注4) <http://prefixspan-rel.sourceforge.jp/>

(注5) <http://mastarpj.nict.go.jp/~mutiyama/software/maxent/>

(注6) <http://mecab.sourceforge.net/>

表-3 ユーザ分類器の分類精度 P_{WS}

学習に用いた回答の件数		分類精度 P_{WS}
～	99	—
100	～ 199	0.753
200	～ 299	0.796
300	～	0.816

度で対象ユーザの回答であるかどうかを判定できる。表2の素性の中では、それぞれの文の文頭および文末の1～10文字(S5とS6)が有効であった。

3.3 回答への評価の比較

質問者が回答を公平に評価している場合、1人のユーザが複数のアカウントを用いて投稿した回答への評価には差がないと考えられる。すなわち、回答への評価に差があるのなら、それらの回答は1人のユーザによって投稿されたのではないと考えられる。そこで、異常な投稿を行っているとして検出されたユーザペアについて、そのユーザらが投稿した回答への評価に差がないか比較する。回答への評価の比較には仮説DEを用いる。

仮説 DE ユーザ i とユーザ j は $N_{DE}(i, j)$ 件の質問と一緒に回答を投稿しているとする。この $N_{DE}(i, j)$ 件の質問に対してユーザ i が投稿した回答のうち、ベストアンサーに選ばれたものが $N_{DEbestans}(i, j)$ 件あるとする。このときもし、ユーザ i とユーザ j の回答への評価に差がないならば、この $N_{DE}(i, j)$ 件の質問に対してユーザ j が投稿した回答のうちベストアンサーに選ばれるのは $N_{DEbestans}(i, j)$ 件であると期待できる。

もしこの仮説が片側二項検定で棄却されれば、ユーザ i とユーザ j の回答への評価には差があると判定する。そして、回答への判定に差があるユーザペアは同一人物ではないと判定する。回答への評価が公平でない場合については、4章で議論する。

仮説 AA と仮説 WS が棄却され、仮説 DE が棄却されなかったユーザペアは、

- 異常な投稿を繰り返している
- 文体が類似している
- 回答への評価に差があるとはいえない

ことから、TYPE AA の投稿を繰り返した疑い、すなわち、1人のユーザが複数のアカウントを利用して1つの質問に複数の回答を投稿することを繰り返した疑いがあると判定する。

4. 実験と検討

提案手法の有効性を評価するために、Yahoo!知恵袋(286 カテゴリ)に回答を投稿したすべてのユーザ(183242 ユーザ)を対象に TYPE AA の投稿を繰り返した疑いのあるユーザ、すなわち、複数のアカウントを利用して1つの質問に複数の回答を投稿することを繰り返した疑いのあるユーザを検出する実験を行った。実験で用いた有意水準は、仮説 AA では0.000005、仮説 WS では0.005、仮説 DE では0.01 である。仮説 AA の検定に用いた有意水準0.000005は一般的に用いられる有意水準に比べて極めて小さい。これは、極めて異常な投稿を検出するためである。表3に仮説 WS の検定で用いるユーザ分類器の分類精度 P_{WS} を示す。分類精度 P_{WS} には、3.2節で測定した「パソコン、周辺機器」「病気、症状、ヘルスケア」「政治、社会問題」カテゴリにおける分類精度(図2)の平均値を用いる。 P_{WS} の値は、表3に示すように、学習に用いる回答の数によって変化する。

実験ではまず(予想1)にもとづいて、異常な投稿を行っているユーザペア、すなわち、一緒に回答することが非常に多いユーザのペアを仮説 AA を用いて検出した。その結果、Yahoo!知恵袋の286カテゴリから767組のユーザペアによる790件の異常な投稿が検出された。異常な投稿を行っているとして検出された767組のユーザペアは以下の2つのタイプに分類できる。

type a 異常な投稿が複数のカテゴリで検出されたユーザペア(13組)。例えば、ユーザ 691911 とユーザ 802184 のユーザペアは「野球」「政治、社会問題」「テレビ、ラジオ」など11個のカテゴリで異常な投稿が検

表-4 異常な投稿を行っているユーザペアの文体の類似判定と回答への評価の比較の結果

(a) すべてのカテゴリ (286 カテゴリ)

ユーザペア の異なり		回答への評価の比較結果						合計
		type 0			type X			
		文体の類似判定結果			文体の類似判定結果			
		type 1	type 2	type 3	type 1	type 2	type 3	
type a	13	6	5	1	16	7	1	36
type b	754	168	339	10	55	182	0	754
合計	767	174	344	11	71	189	1	790

(b) 政治、社会問題カテゴリ

ユーザペア の異なり		回答への評価の比較結果						合計
		type 0			type X			
		文体の類似判定結果			文体の類似判定結果			
		type 1	type 2	type 3	type 1	type 2	type 3	
type a	1	0	0	0	1	0	0	1
type b	19	5	4	3	5	2	0	19
合計	20	5	4	3	6	2	0	20

表-5 「政治、社会問題」カテゴリで異常な投稿を行っているユーザペアの文体の類似判定と回答への評価の比較の結果

(a) 人手による調査で同一人物と判定されたユーザペア (8 組)

ユーザペア の異なり		回答への評価の比較結果						合計
		type 0			type X			
		文体の類似判定結果			文体の類似判定結果			
		type 1	type 2	type 3	type 1	type 2	type 3	
type a	1	0	0	0	1	0	0	1
type b	7	4	0	3	0	0	0	7
合計	8	4	0	3	1	0	0	8

(b) 人手による調査で同一人物と判定されなかったユーザペア (12 組)

ユーザペア の異なり		回答への評価の比較結果						合計
		type 0			type X			
		文体の類似判定結果			文体の類似判定結果			
		type 1	type 2	type 3	type 1	type 2	type 3	
type a	0	0	0	0	0	0	0	0
type b	12	1	4	0	5	2	0	12
合計	12	1	4	0	5	2	0	12

表-6 石川ら[8]によって同一人物の疑いがあると判定されたユーザペアの文体の類似判定と回答への評価の比較の結果

ユーザペア の異なり		回答への評価の比較結果						合計
		type 0			type X			
		文体の類似判定結果			文体の類似判定結果			
		type 1	type 2	type 3	type 1	type 2	type 3	
type a	2	1	0	0	11	0	1	13
type b	8	4	1	0	1	2	0	8
合計	10	5	1	0	12	2	1	21

出されていて、type a に分類されている。このタイプのユーザペアによる異常な投稿は 39 件検出された。

type b 異常な投稿が 1 つのカテゴリでだけ検出されたユーザペア (754 組)。このタイプのユーザペアによる異常な投稿は 754 件検出された。

この検出された 767 組のユーザペアによる 790 件の異常な投稿すべてのケースについて、(予想 2) にもとづく文体の類似判定を仮説 WS を用いて行った。例えば、ユーザ 691911 とユーザ 802184 のユーザペアは、「野球」「政治、社会問題」「テレビ、ラジオ」など 11 個のカテゴリで異常な投稿が検出されているので、その文体の類似判定は合計 11 回行っている。文体の類似判定の結果は以下の 3 つのタイプに分類できる。

type 1 仮説 WS が棄却されるケース。対象のユーザペアの文体は類似していると判定される。

type 2 仮説 WS が棄却されないケース。対象のユーザペアの文体は類似しているとはいえないと判定される。

type 3 文体の類似判定を行わなかったケース。このケースはどのユーザも回答投稿数が 50 件未満の場合で、学習データが小さすぎるので分類器を作成しなかった。

さらに、(予想 3) にもとづく回答への評価の比較も、この 790 件の異常な投稿すべてのケースについて仮説 DE を用いて行った。回答への評価の比較の結果は以下の 2 つのタイプに分類できる。

type 0 仮説 DE が棄却されないケース。対象のユーザペアのそれぞれのユーザが投稿した回答への評価には差があるとはいえないと判定される。

type X 仮説 DE が棄却されるケース。対象のユーザペアのそれぞれのユーザが投稿した回答への評価には差があると判定される。

表 4(a) に実験結果を示す。この 767 組のユーザペアによる 790 件の異常な投稿すべてについて同一人物による投稿であるかを検討するのはむずかしい。そこで、「政治、社会問題」のカテゴリで異常な投稿を行った 20 組のユーザペアについて同一人物であるか詳しく検討する。「政治、社会問題」カテゴリで異常な投稿を行った 20 組のユーザペアに対する実験結果を表 4(b) に示す。

「政治、社会問題」カテゴリで異常な投稿が検出された type a のユーザペアは (691911×type 1→802184) の 1 組である。ここで数字 (例: 691911) はユーザ ID を、双方向の矢印とラベル (例: ←type 1→) は異常な投稿が検出されたユーザペアであることとその文体の類似判定の結果 (この場合は type 1) を示す。この例のように矢印の上に × があれば、それは回答への評価の比較の結果が type X であることを示す。したがって、このユーザペア (691911×type 1→802184) は文体は類似している (type 1) が回答への評価に差がある (type X) ので、(予想 1) ~ (予想 3) にもとづく方法では同一人物ではないと判定された。しかし、type a に分類されているこのユーザペアは「政治、社会問題」カテゴリだけでなく、「野球」「テレビ、ラジオ」など合計 11 個のカテゴリで異常な投稿が検出されている。異なるユーザが、複数の異なるカテゴリの質問と一緒に回答する回数が非常に多いというのは不自然である。この点と投稿内容を人手で調査した結果を考慮して、このユーザペアは同一人物の疑いがあると判定した。人手による調査で同一人物の疑いがあると判定されたこのユーザペアで、(予想 3) に反して回答への評価に差があることについては後で詳しく議論する。

複数のカテゴリの質問と一緒に回答する回数が非常に多いユーザペアは不自然である。なぜなら、異なるカテゴリに投稿される質問の内容はさまざまであるのに、異なるユーザの興味や関心が不自然なまでに一致して一緒に回答しているからである。同様に、1 つのカテゴリに投稿された質問であっても、さまざまな内容の質問と一緒に回答する回数が非常に多いユーザペアもまた不自然である。「政治、社会問題」カテゴリで異常な投稿を行っているとして検出された以下の 2 組のユーザペアでは、一緒に回答している質問の内容がそれぞれさまざまであった。

(126206←type 1→115030) (126206←type 1→380592)

異なるユーザの興味や関心がここまで一致するのは不自然であり、これらのユーザペアはそれぞれ同一人物の疑いがあると判定した。この 2 組のユーザペアは、文体は類似している (type 1)、回答への評価に差はあるとはいえない (type 0) と判定されていて、(予想 1) ~ (予想 3) にもとづく方法でも同一人物の疑いがあると判定されている。この 2 組のユーザペアのようにさまざまなトピックに関する質問に回答を投稿するユーザペアがいる一方で、特定のトピックに関する質問と一緒に回答することが非常に多くて、それが異常な投稿として検

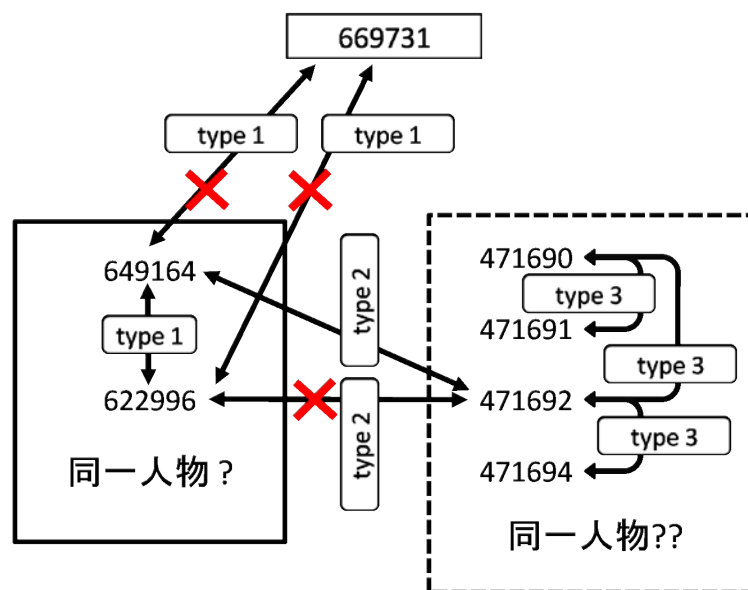


図-3 「政治、社会問題」カテゴリで特定のトピックに関する質問に繰り返し回答を投稿していた 8 組のユーザペアに対して、(1) 同一人物かどうかを手で判定した結果と (2) 文体の類似判定および回答への評価の比較の結果を示す。同一人物の疑いがあると判定されたユーザのグループは、図中で太線あるいは破線で囲まれている。双方向の矢印とラベル(例: ←type 1→)は異常な投稿が検出されたユーザペアであることとその文体の類似判定の結果を示す。矢印の上にある×は回答への評価に差があること (type X)、すなわち、同一人物ではないと考えられることを示す。

出されたユーザペアが「政治、社会問題」カテゴリでは 9 組いた。このうち (76731←type 1→267614) は、在日外国人問題に関する質問と一緒に回答することが非常に多くて、それが異常な投稿として検出されたユーザペアである。このユーザペアは、意見がよく一致していること、文体も類似していること、特に他のユーザが「政治、社会問題」カテゴリではほとんど使わないトピック(「個人の HP」「漫画家」など)がそれぞれのユーザの回答の中で取り上げられていることなどから、人手による調査では同一人物の疑いがあると判定した。また、このユーザペアは、文体が類似していて (type 1)、回答への評価に差があるとはいえない (type 0) と判定されていて、(予想 1)～(予想 3) にもとづく方法でも同一人物の疑いがあると判定されている。一方、残りの 8 組のユーザペアは、特定の宗教団体に関する質問と一緒に回答することが非常に多くて、それが異常な投稿として検出されたユーザペアである。この 8 組のユーザペアについての調査結果を図 3 に示す。この中でユーザペア (622996←type 1→649164) は意見が特に一致していた。このユーザペアで意見が対立したのは、他のユーザから同一人物ではないかと指摘を受けた直後の 1 回だけであった。そこで、このユーザペアは同一人物の疑いがあると判定した。このユーザペアは (予想 1)～(予想 3) にもとづく方法でも同一人物の疑いがあると判定されている。このユーザペアと同様に、図 3 の以下の 3 つのユーザペア

(471690←type 3→471691) (471690←type 3→471692) (471692←type 3→471694)

を構成する 4 人のユーザ(ユーザ 471690、ユーザ 471691、ユーザ 471692、ユーザ 471694)の意見もよく一致していた。そこで、これらのユーザも同一人物の疑いがあると判定した。この 4 人のどのユーザも回答投稿件数が 50 件未満であるため、機械学習による文体の類似判定は行えなかったが、人手による調査ではこの 4 人のユーザの文体は類似していた。なお、同一人物であるかの判定には考慮しなかったが、この 4 人のユーザ ID は不自然なほど連続している。この原因として、ユーザ登録の時期が非常に近かったとか、ハンドルネームが非常に似ているなど、同一人物であることに由来するものが考えられる。実際、ユーザ 471692 は「字数制限あるから ID 変えて続きを書く」と述べて、ユーザ ID が隣接しているユーザ 471693 と同一人物であることを明言している。ユーザ 471693 は「政治、社会問題」カテゴリに 8 件しか回答を投稿していないため、仮説 AA

では検出できなかった。しかし、その 8 件の回答のうち 6 件はユーザ 471692 と一緒に同じ質問に投稿したものである。一方、図 3 の以下の 4 つのユーザペアでは意見は一致せず、対立することが多かった。このため、これらのユーザは同一人物ではないと判定した。

(622996×(type 2)→471692) (622996×(type 1)→669731)
(649164←(type 2)→471692) (649164×(type 1)→669731)

特に、ユーザ 471692 はユーザ 622996 および 649164 とおたがいに相手の投稿内容を否定したり非難することが多かった。この 4 組のユーザペアは、文体が類似するとはいえない(type 2)、あるいは回答への評価に差がある(type X)と判定されたため、(予想1)～(予想3)にもとづく方法でも同一人物ではないと判定されている。

このように、「政治、社会問題」カテゴリで異常な投稿を行っているとして判定された 20 組のユーザペアのうち同一人物の疑いがあると人手による調査で判定したのは 8 組であった。この 8 組のユーザペアとそれ以外の 12 組のユーザペアそれぞれに対する文体の類似判定および回答への評価の比較の結果を表 5 に示す。表 4(b)に示すように、文体が類似していて(type 1)そして回答への評価にも差があるとはいえず(type 0)、(予想1)～(予想3)にもとづく方法では同一人物の疑いがあると判定されたユーザペアは 5 組あった。表 5(a)に示すように、そのうち 4 組が人手の調査でも同一人物であると判定された。投稿件数が少なく type 3 に分類されている 3 組のユーザペアも、調査期間を延長して十分な数の回答を集めることができれば、同一人物の疑いがあるユーザペアとして検出できる可能性がある。

「政治、社会問題」カテゴリのユーザペア (691911×(type 1)→802184) についても一度検討する。このユーザペアは、文体は類似する(type 1)が回答への評価に差がある(type X)と判定されたため、(予想1)～(予想3)にもとづく方法では同一人物ではないと判定された。しかし、人手による調査ではこのユーザペアは同一人物の疑いがあると判定されている。さらに、ユーザ 691911 がユーザ 802184 の回答を異常に繰り返しベストアンサーに選んでいて、石川らはこのユーザペアは同一人物の疑いがあると指摘している[8]。このようにユーザ 691911 と 802184 が同一人物であるのなら、それらの回答への評価にはなぜ差があるのかが問題になる。ユーザ 691911 と 802184 が「政治、社会問題」カテゴリで一緒に回答した 47 件の質問を調べると、そのうち 44 件がユーザ 425288 の質問であった。この 44 件の質問のうち、ユーザ 425288 は 43 件でユーザ 802184 の回答をベストアンサーに選んでいた。このようにユーザ 425288 もユーザ 802184 の回答を異常に繰り返しベストアンサーに選んでいて、石川らはこのユーザペアも同一人物の疑いがあると指摘している[8]。したがって、ユーザ 691911 と 802184 の回答への評価に差があるのは、ユーザ 425288 が評価を公平に行っていないことが原因と考えられる。

(予想1)にもとづく仮説 AA によって検出した 767 組のユーザペアを調べると、この (691911×(type 1)→802184) のように、一方のユーザがもう一方のユーザの回答を異常に繰り返しベストアンサーに選んでいて、石川ら[8]によって同一人物の疑いがあると判定されたユーザペアが 10 組見つかった。この 10 組のユーザペアによる 21 件の異常な投稿が仮説 AA によって検出されていて、それらに対する文体の類似判定および回答への評価の比較の結果を表 6 に示す。表 6 に示すように、これら 21 件のケースのうち、回答への評価に差があるもの(type X)が 15 件あった。「政治、社会問題」カテゴリの (691911×(type 1)→802184) のケースのように回答への評価が不公平に行われていないか調べると、回答への評価に差があるとはいえない(type 0)と判定されたケースで 2 件、回答への評価に差がある(type X)と判定されたケースで 14 件、合計 16 件のケースで回答への評価が不公平に行われている疑いがあった。そこで不公平な評価が行われている疑いのある回答をとりぞいて回答への評価の比較を行った。その結果、回答への評価に差があるケースは 2 件だけだった。このことから、評価が公平であるなら、同一人物による回答への評価には差がないという(予想3)が妥当であるといえる。

5. おわりに

本研究では Yahoo!知恵袋を対象に複数のアカウントを用いて 1 つの質問に複数の回答を投稿することを繰り返すユーザを

(予想 1) あるユーザが回答した質問に、特定のユーザも回答していることが異常に多い

(予想 2) 1 人のユーザが複数のアカウントを利用して投稿した回答の文体は類似する

(予想 3) 評価が公平であるなら、1 人のユーザが複数のアカウントを利用して投稿した回答への評価には差がない

の3つの予想にもとづいて検出する方法を提案した。本研究では「政治、社会問題」カテゴリについて実験結果を詳しく検討し、提案手法の有効性について議論した。第4回知識共有コミュニティワークショップ(主催:情報社会学会)でも指摘されたように、今後は複数のカテゴリでの検出結果を比較して、疑わしいユーザをより精度よく検出できる方法について検討する。

[参考文献]

- [1] 折田, 江木: リンク不能性および一覧性の観点による匿名性の分類, 情報処理学会研究報告, 2007-EIP-37, (2007).
- [2] de Vel, Anderson, Corney, and Mohay: Mining e-mail content for author identification forensics, ACM SIGMOD Record, 30(4), (2001).
- [3] Koppel, Argamon, and Shimoni: Automatically Categorizing Written Text by Author Gender, Literary Linguistic and Computing, 17(4), (2002).
- [4] Corney, de Vel, Anderson, and Mohay: Gender-Preferential Text Mining of E-mail Discourse, ACSAC 2002, (2002).
- [5] Argamon, Saric, and Stein: Style mining of electronic messages for multiple authorship discrimination: first results, 9th ACM SIGKDD, (2003).
- [6] Zheng, Li, Chen, and Huang: A Framework for Authorship Identification of Online Messages: Writing Style Features and Classification Techniques, Journal of the American Society for Information Science and Technology, 57(3), (2006).
- [7] 坪井, 松本: Authorship Identification for Heterogeneous Documents, 情報処理学会研究報告, 2002-NL-148, (2002).
- [8] 石川, 梅本, 西村, 渡辺, 岡田: Q&A サイトにおける自分の投稿に対する評価を操作するために複数のアカウントを利用するユーザの検出, 情報社会学会誌, Vol. 5 No. 3, (2011).
- [9] 小高, 村田, 高, 諏訪, 白井, 高橋, 黒岩, 小倉: n-gram を用いた学生レポート評価手法の提案, 電子情報通信学会論文誌, Vol. J86-D-I No. 9, (2003).

(2012年1月26日受理)

