Vol.2012-GN-83 No.6
2012/3/21

Twitter への擬似犯罪発言抑止における リスト組み合わせ方式の提案

青柳翔（神奈川工科大学情報学部情報メディア学科）†
服部哲（同上）††　速水治夫（同上）†††

近年，SNS として流行している Twitter は，日々思ったことや出来事などを自由に文章として投稿できるサービスであり，いろいろな目的を持って利用できる．しかし，その Twitter において，悪ふざけや冗談で犯罪を告白するようなツイートをするユーザーが後を絶たない．犯罪の告白の真似をし，自らネット上における炎上を起こしたり，周囲の評判を悪くするような投稿ができるのは問題がある．そこで，Twitter を利用する上で，この問題点を解決した擬似犯罪発言抑止システムの提案をする．

The proposal of the list combination system in suppression of pseudo crime expression to Twitter

SHO AOYAGI (Kanagawa Institute of Technology)†
AKIRA HATTORI††　　HARUO HAYAMI†††

In recent years, Twitter which is in fashion as SNS is the service which can contribute having considered every day, an occurrence, etc. as a text freely, and it can use it with various purposes.
However, in the Twitter, the user who does tweet which confesses a crime with a practical joke or a joke does not sever the back.
Imitating a criminal confession, it has a problem that cause the destruction by fire on a network oneself, or contribution which worsens the surrounding reputation is made.
Then, when using Twitter, the false criminal utterance deterrence system which solved this problem is proposed.

1. はじめに

近年，流行している Twitter は，今もなおユーザーを増やしているサービスである．その Twitter において，悪ふざけや冗談で犯罪を告白するようなツイートをするユーザーが後を絶たない．上記のようなツイートは，そのユーザーのタイムラインを炎上させたり，ユーザーの周囲に悪い影響や評判を与えてしまう．

そこで，本論文では Twitter でツイートを行う際に，そのツイートが危険なものでないかを解析・判断し，必要に応じて規制する機能を持ったクライアント型の Web アプリケーションを提案する．本システムは，主にツイートをする際に機能するものであり，ユーザーが登録して利用することで自分自身を守ることができるようにする．

2. 研究対象の現状

研究対象の現状と，用語の意味を以下に示す．

2.1 Twitter

Twitter は 2009 年の春頃に流行の兆しを見せ，今もなおユーザーを増やしているサービスである [1]．これはつぶやきと呼ばれるテキストを投稿するサービスであり，ブログや SNS の日記のように各記事にタイトルを付ける必要がないことなどから，気軽に素早く投稿できるのが特徴である [2]．

2.2 ツイート

ツイートとは，Twitter 利用ユーザーが Twitter 上につぶやきと呼ばれる短い文章を投稿すること，また投稿した文章のことである．ユーザーは様々な事を自由にツイートすることができる．一度にツイートできる文章の文字数制限は 140 文字とされているため，後からツイートの本文を見ただけでも内容が分りやすい．以下，本論文では便宜上のためユーザーが Twitter 上にツイートすること，またツイートされた文章のことを「ツイート」と呼ぶことにする．

2.3 リプライ

リプライとはツイートの文章の中に「@」をつけたり，Twitter 上の返信ボタンを押して特定のユーザー宛てに投稿したツイートのことである．「@ユーザー名」のように「@」の後にユーザー名を入力することで，そのユーザー宛てにツイートすることができる．主に会話や，相手ユーザーに関係があるツイートなどに使われる．

2.4 ダイレクトメッセージ

ツイートに投稿した文章は，投稿したユーザーが非公開に設定していない限り誰でも閲覧することができる．リプライでツイートした文章も上記と同様ならば誰でも閲覧することができる．それらとは違い，ダイレクトメッセージは第三者が閲覧することのできないメッセージを特定のユーザーに直接送ることができる．またダイレクト

ⓒ 2012 Information Processing Society of Japan

メッセージを送る条件として，送る側のユーザーは，相手ユーザーからフォローされていなければならない．

2.5 フォロー/フォロワー

フォローとは，他のユーザーのツイートを自分のタイムラインに表示したい場合に利用する．また，自分がフォローされている場合の相手ユーザーをフォロワーと言う．

2.6 タイムライン

タイムラインとは Twitter のページにおいて，自分や他のユーザーのツイートが表示される部分のことである．タイムラインはページの上部に最新のツイートが表示され，下部に向かって徐々に古いツイートが表示されていく．Twitter のメインページや，ホームタイムラインに表示されるツイートは自分がフォローしているユーザーのツイートと自分のツイートである．また，特定のユーザーだけのツイートの集まりもタイムラインと言う．

2.7 リツイート

リツイートとは，Twitter 利用ユーザーが他の Twitter 利用ユーザーのツイートを引用形式で自分のアカウントから発信することである．RT と表記されることが多い．これにより自分が興味を持ったツイートを，手軽にフォロワーへと流すことができる．

2.8 Twitter にまつわる問題

Twitter で起きている問題には，デマの拡散，広告などの迷惑なツイート，社内文書の漏洩，犯罪の告白などがある．これらはツイートに 140 文字までという字数以外の制限がないことから，内容に関わらずツイートできる仕様により起きている．

近年では，特に悪ふざけやいたずらで犯罪の告白を含むツイートをするユーザーが後を絶たない [3),4),5),6)]．犯罪の告白には主に飲酒運転，未成年飲酒，カンニング，窃盗，放火予告，殺人予告などがあげられる．このようなツイートは即座に拡散し，それによって発言したユーザーのタイムライン，ダイレクトメッセージなどは炎上する．中には本当に犯罪をして告白しているユーザーも存在するが，多数のユーザーは悪ふざけでツイートしている．悪ふざけで嘘の犯罪の告白をするのは本人の自由ではあるが，自ら炎上させる目的でツイートするユーザーはそうはいないと考えられる．そこで本研究におけるシステムでは，一般人が多く用いる，冗談やいたずら目的の犯罪に関連する発言を主な対象とする．

3. 問題点と解決策

3.1 研究対象の問題点と解決策

Twitter に関連する問題点は，前述した通りである．その中でも犯罪の告白に関連することを本研究では主な対象とする．いままでは文字数以外の制限がないことから，多くのユーザーが悪ふざけや冗談で犯罪に関連するツイートを行っていた．このツイートは，そのユーザーのタイムラインを炎上させたり，そのユーザーの周囲に悪い影響や評判を与えてしまう．

この点に関して，本研究では Twitter でツイートをする際における文章に制限をかけることで，犯罪に関連する危険なツイートを規制する．また，それによってユーザーがシステムに登録するだけで自分自身を守れる形式を取る．

本研究では，まず，システムにツイートを解析する機能を持たせ，ツイートに含まれる単語が個別で検出されるようにする．次いで，用意された単語が含まれていた場合は規制するようにする．また，ツイートした内容の目視が可能でないと何をツイートしたのかが確認できないため，タイムラインの表示も行う．

3.2 関連サービス

関連サービスとして，プラスアルファ・コンサルティングのカスタマーリングスを挙げる．カスタマーリングスは Twitter の場合，ツイートを収集・蓄積し，テキストマイニングしてツイートを分類・分析するシステムである [7)]．事前設定された複数キーワードで分類し，ネガティブ・ポジティブのつぶやきや，企業に宛てた質問・クレーム・要望・部門別などのフォルダに分ける機能を持つ．この機能によって緊急性の高いツイートを早く発見し，企業などが素早く対応することができる．

しかし，このシステムはキーワードの発見を行い分類・分析するだけであり，実際にシステムそのものには規制などを行う機能は持たないため，犯罪に関するツイートを止めることができない．

ⓒ 2012 Information Processing Society of Japan

4. 使用技術

4.1 使用技術概要

本研究では，Web ブラウザ上で動作するクライアントとして開発を行う．クライアントとしての形態を取るのは，一度認証を行うだけでツイートやダイレクトメッセージなどの多様な機能を使用する際に，すぐに使用できるようにするためである．

4.2 Apache

Apache(アパッチ)は Web サーバソフトウェアの一つである．代表的なオープンソース・ソフトウェアでもあり，非常に多くのサイトに利用されている．

4.3 MySQL

MySQL はリレーショナルデータベース管理システムであり，Windows や Unix 系など，多くのプラットフォームで動作する．オープンソース・ソフトウェアである．

4.4 PHP

PHP は，動的に Web ページを生成するサーバの拡張機能の一つである．HTML に埋め込むサーバサイド・スクリプト言語として分類され，Web サーバ上で動作し，クライアントなどからの要求されるたびに，動的な結果を生成する．Javascript やデータベースとの連携に優れている．

4.5 MeCab

MeCab は京都大学情報学研究科によって開発されたオープンソース形態素解析エンジンである [8]. 本研究で辞書として登録されているのは Wikipedia とはてなキーワードであり，これらに含まれている単語で主な形態素解析を行う．また，形態素解析とは自然言語処理の一環であり，辞書を元に文章を，言語として意味を持つ最低限の状態に分割することである．

4.6 OAUTH 認証

OAUTH 認証(オーオースにんしょう)は， Web アプリケーションなどに API の認証手段を提供するオープンプロトコルである．Web 上のサービスが WebAPI を用いて連動する際に，ユーザーID とパスワードを一元管理することができる．

4.7 TwitterAPI

TwitterAPI は API を利用することでさまざまなアプリケーションを作り出すことができる．TwitterAPI では，ツイートのことを status（ステータス）と言う．

5. 使用技術

5.1 システム概要

本システムは，Twitter でツイートをする際において，炎上するようなツイートを，悪ふざけや冗談でしてしまうのを防ぐシステムである．

利用する上で必要な OAUTH 認証は，システムを利用する前に行ってもらう．これによりシステムの利用は Twitter の ID とパスワードのみで可能となる．

認証後には，ホームタイムラインを取得し，メインページが表示される．メインページには上からユーザー情報タグ，ツイート用のテキストボックス，その下に実際にツイートする文章を解析し，規制する単語のリストのデータベースと見比べて検出された結果が表示されるスペース，ツイートできるかできないかの結果、タイムラインの順で表示される．また，上部タブではダイレクトメッセージの送受信と，関連ツイート，フォロー，フォロワーなどのユーザー情報の確認ができる．実際のメインページを図 5.1，システム構成を図 5.2 に示す．

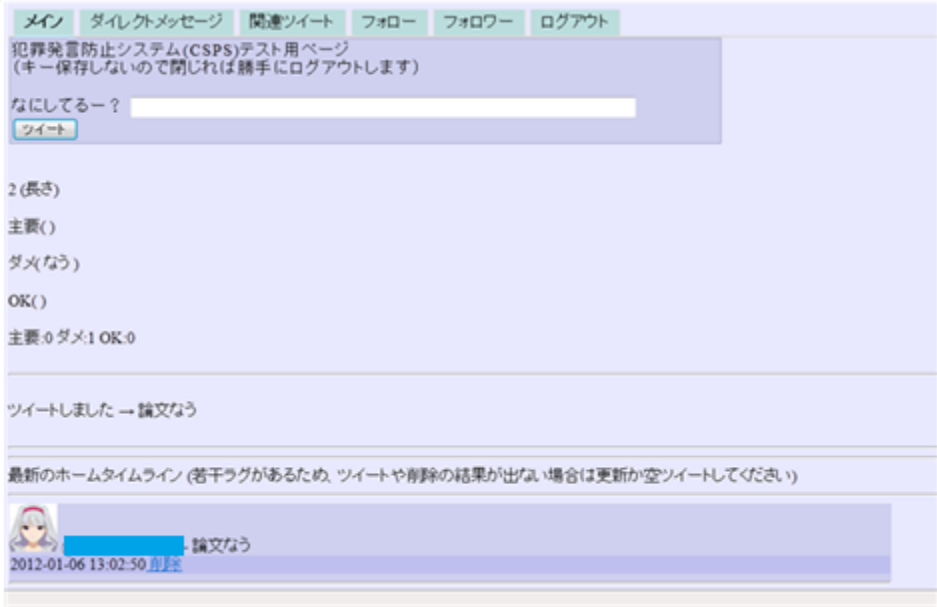

図 5.1 実際のメインページ

ⓒ 2012 Information Processing Society of Japan

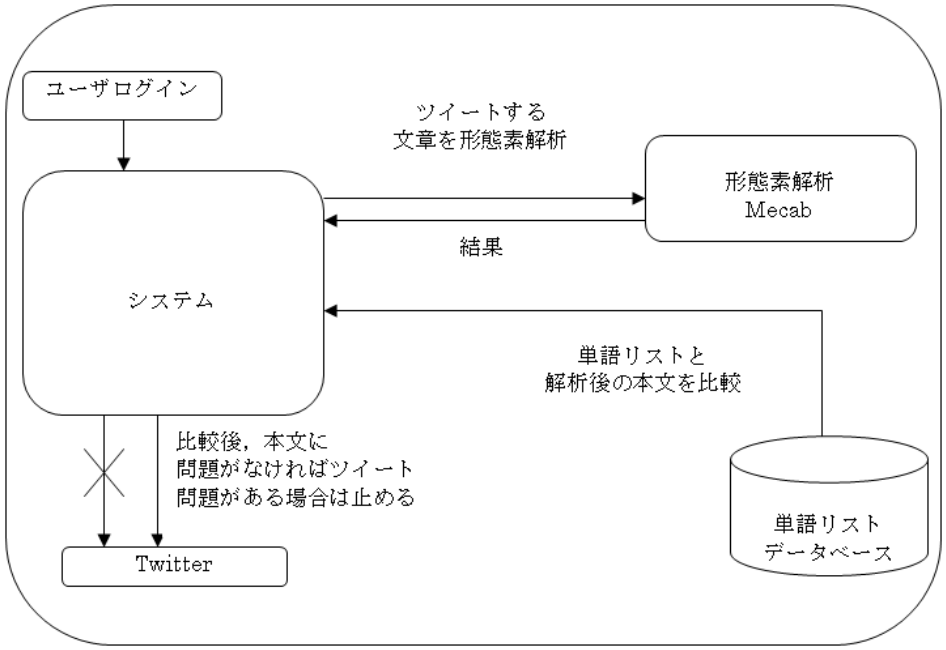

図 5.2　システム構成

5.2　規制手段

規制する手段には，ブラックリスト方式，オートマトン方式がある．しかしこれらは本システムにおいて用いる場合，欠点がある．以下で各手段の主な利点と欠点を示し，それらを踏まえた上で新しく提案した方式を示す．

5.2.1　ブラックリスト方式

ブラックリスト方式による規制は，主に 1 つだけの単語で規制する手段である．これは規制手段としては非常に簡潔であり，新語が増えた場合にもリストへの追加が非常に容易である．しかし，単語 1 つで規制を行うと，全く意図していないツイートで規制することがあり得るため，実際にシステムに用いる上には弊害がある．つまり，新語の追加は容易であっても規制手段としてはふさわしくない．

5.2.2　オートマトン方式

オートマトン方式による規制は，複数の単語によって規制する手段である．これは一定の単語同士の組み合わせによって規制を行うため，精度としては非常に高い．しかし，新語が増えた場合にリストへの追加を行うと，組み合わせの種類も増えるため，ソースプログラムにも変更が必要になる．つまり，精度は高いがメンテナンス面において効率が悪い．

5.2.3　リスト組み合わせ方式

リスト組み合わせ方式による規制は，ブラックリスト方式に補助のリストを追加することで文章を分類し，必要に応じて規制する手段である．これは一定の単語同士の組み合わせでなく，分類された複数のリストの単語の組み合わせで規制するため，オートマトン方式と比較すると精度はやや下がるものの，ソースプログラムの変更を行わずに幅広い文章を比較できる．新語の追加については，リストに種類別に追加していくだけでリスト間の組み合わせのパターンが増えるため，容易に行える．本システムでは，精度が高く，メンテナンスも容易であるリスト組み合わせ方式を用いる．なお，具体的なリストの構成と組み合わせについては，5.3 に従って記述する．

5.3　規制するツイートの検出

ツイートする文章を形態素解析した結果と，単語のリストを比較し，単語が検出された場合，リストの組み合わせでツイートを規制する．データベースに登録されている単語のリストを表 5.1 に示す．

表 5.1　単語リスト

リスト A	主要的な単語に該当する単語をまとめたリスト
リスト B	リスト A の単語と共に含まれていた場合 問題なくツイートできる単語をまとめたリスト
リスト C	リスト A の単語と共に含まれていた場合 規制されツイートできなくなる単語をまとめたリスト

リスト A の主要的な単語に該当する単語をまとめたリストには，「飲酒運転」や「カンニング」などの主語にあたるものが該当する．リスト B の，リスト A と共に含まれている場合，問題なくツイート出来る単語のリストには「してはいけません」や「違法です」などの，リスト A の単語に対する否定や，注意を促す単語や文節が含まれる．リスト C の，リスト A と共に含まれている場合，規制されツイートできなくなる単語をまとめたリストには「余裕でした」や「なう」などの，リスト A の単語にあたることを行う予定，あるいは実際に行っているということを表す単語や文節が含まれる．

次いで，組み合わせのパターンを表 5.2 に示す．

ⓒ 2012 Information Processing Society of Japan

表 5.2　単語組み合わせパターン簡易表

リスト A の単語	リスト B の単語	リスト C の単語	ツイート	システム動作
ヒット 例：飲酒運転	―	―	△ 可能	注意
ヒット 例：飲酒運転	ヒット 例：違法です	―	○ 可能	なし
ヒット 例：飲酒運転	―	ヒット 例：余裕でした	× 不可能	規制
ヒット 例：飲酒運転	ヒット 例：違法です	ヒット 例：余裕でした	× 不可能	注意・規制
―	ヒット 例：違法です	―	○ 可能	なし
―	ヒット 例：違法です	ヒット 例：余裕でした	○ 可能	なし
―	―	ヒット 例：余裕でした	○ 可能	なし

これらのうち，特に規制をされずに標準通りにツイートされるパターンは，主要的な単語を含んでいない場合や，主要的な単語があり，それを否定・注意する単語や文節が含まれている場合である．また，解析後の判断が難しい主要的な単語のみが含まれるツイートに関しては注意を促し，必要に応じて削除をするようにすることで標準通りにツイートが可能な形式にした．更に，本システムの主となる，規制をされ，ツイートが止められるパターンは，主要的な単語を含んでおり，かつそれを行う予定，あるいは行ったという単語や文節が含まれている場合である．

5.4 データベース構成

データベース構成を以下に示す．すべてのテーブルには単語のリストが含まれている．

- リスト A を保存する kisei_t テーブル（表 5.3）
- リスト B を保存する okkisei_t テーブル（表 5.4）
- リスト C を保存する damekisei_t テーブル（表 5.5）

表 5.3　kisei_t テーブル

カラム名	タイプ	用途	備考
tid	int(3)	単語 1 つ 1 つを区別するための ID	主キー
tango	varchar(50)	解析後に比較する主要的な単語	

表 5.4　okkisei_t テーブル

カラム名	タイプ	用途	備考
okid	int(3)	単語 1 つ 1 つを区別するための ID	主キー
tid	int(3)	kisei_t との関連付け	
oktango	varchar(50)	解析後に組み合わせで比較	

表 5.5　damekisei_t テーブル

カラム名	タイプ	用途	備考
dameid	int(3)	単語 1 つ 1 つを区別するための ID	主キー
tid	int(3)	kisei_t との関連付け	
dametango	varchar(50)	解析後に組み合わせで比較	

5.5 ログイン方法

ユーザー認証には，TwitterAPI を利用し，Twitter を経由して OAUTH 認証を行い本システムにログインする．本システムを利用するには必ず Twitter のユーザーID が必須となるため，本システムを利用する前に Twitter のユーザーID を用意しておかないといけない．

ユーザーは初めに，本システムのログインページにアクセスし，「OAUTH 認可する」から Twitter アカウントへのログイン画面へ遷移する．遷移先でユーザーは TwitterID とパスワードを入力し．Twitter アカウントへログインする．Twitter にログインが成功した後，本システムによるアプリケーションの許可が一括で行われるため，そこで許可した場合本システムを利用できるようになる．図 5.3 に「OAUTH 認可する」ページを示す．

また，ユーザーが既に Twitter にログイン中である場合は Twitter アカウントへのログインページは表示されなく，自動的にアプリケーションの許可画面が表示される．更に，本システムにおいてはアプリケーションの許可情報の保存を行わないため，一度ログアウトすると再び利用の際には再度許可が必要となる．

ⓒ 2012 Information Processing Society of Japan

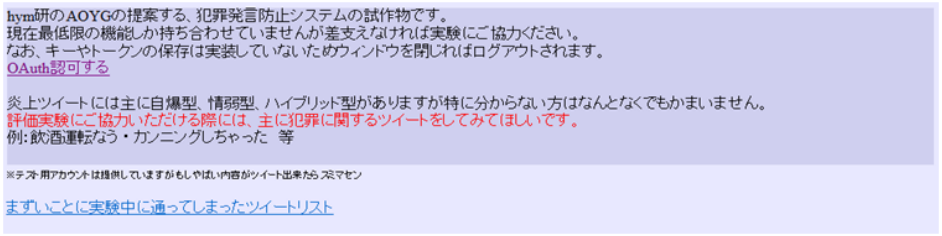

図 5.3 OAUTH 認可画面

5.6 ユーザー情報表示部

メインページ上部のタブより，ダイレクトメッセージの送受信，関連ツイート（@自分のユーザー名のツイート）の確認，フォローとフォロワーの確認，システムからのログアウト（アプリケーションの許可を切ること）の各機能が使用できる．

6. 評価実験

6.1 評価方法

本システムの評価は、6 人の実験協力者に実際にシステムを利用してもらい，5 段階評価と，アンケートを行った．質問内容は以下の 4 つである．

1. 犯罪に関するツイートを見たことがあるか
2. 犯罪に関するツイートをしたことがあるか
3. 犯罪に関する，炎上するようなツイートをシステムは止めていたか
4. 犯罪に全く関連性のない，通常のツイートをする際に問題はないか

6.2 犯罪に関するツイートについて

本システムの評価の前に，実験協力者に犯罪に関連するツイートに対する認識を確認し，システムの有用性を調査した．

	2 者択一	
質問内容	いいえ	はい
犯罪に関連するツイートを見たことがあるか	2 人	4 人
犯罪に関連するツイートをしたことがあるか	4 人	2 人

調査の結果，犯罪に関連するツイートを冗談やいたずらでしたことがあると答えたのは 6 人中 2 人であったが，犯罪に関連するツイートを見たことがあると答えたのは 6 人中 4 人と多かった．これにより，小規模ではあるが本研究の対象となるツイートは多くされていることが分かる．

6.3 システム評価

本システムを利用し，犯罪に関連するツイートに対する対策ができているかの評価を 5 段階で行った．

評価内容	1	2	3	4	5
犯罪に関連するようなツイートをシステムが止めてくれたか	0 人	1 人	3 人	1 人	1 人
犯罪に関連しない通常のツイートをする際に弊害はないか	0 人	0 人	0 人	2 人	4 人

評価の結果，犯罪に関するツイートをシステムが規制してくれたかという項目では平均 3.4 点，犯罪に関連しない通常のツイートをする場合に邪魔をされないかという項目では平均 4.7 点とどちらも高い評価が得られた．今回はシステムでは犯罪に関するツイートの解析・比較と規制を主に行ったが，比較に用いるデータベースの単語の種類（単語の登録数）がまだ甘いという意見が多かった．更に，規制をかける場合には完全にツイートできないという形式ではなく，ツイートするかしないかの選択性にすべき，という意見も頂いた．また，単語同士の組み合わせによる規制はしっかりとできているため，単語の登録数を増やしていけば良いシステムになる，という意見も頂いた．

6.4 実際のツイートによる実験

5.3 で示した，検出する機能の効果を調査するために，実際に存在した犯罪に関する（冗談やいたずらによる）ツイートと，それらを否定，もしくは反対するツイートをシステムでツイートする場合の結果を以下で図示する．中央部の最新のホームタイムラインという項目の上部にある文章が結果である．

完全に規制され，ツイートできないパターンを図 6.1，安全だと判断し，規制や注意を行わない通常のパターンを図 6.2 に示す．

ⓒ 2012 Information Processing Society of Japan

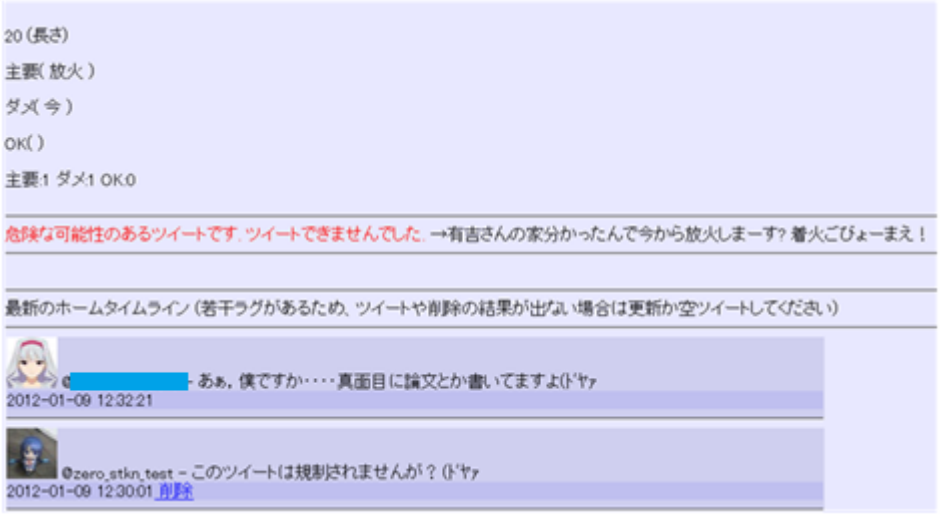

図 6.1 規制パターン

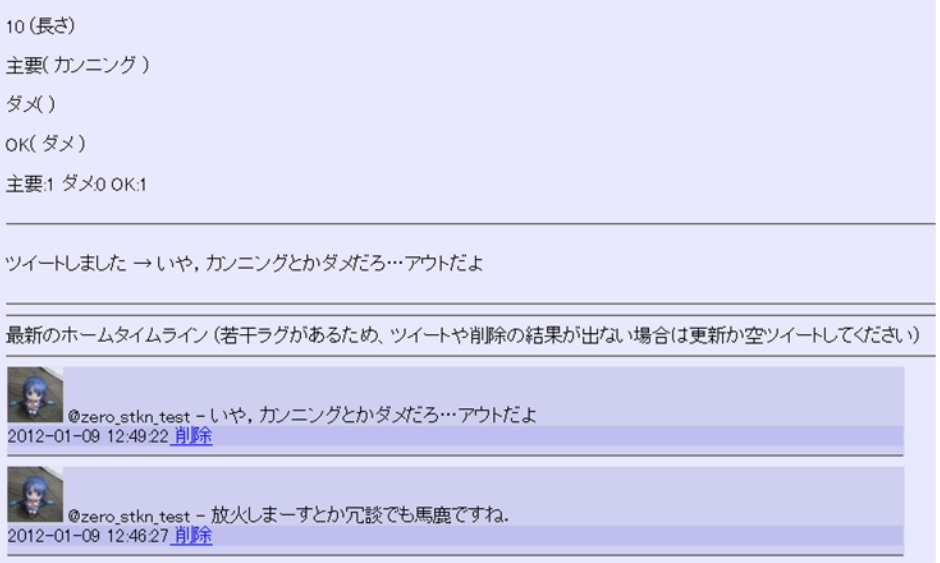

図 6.2 通常パターン

実験の結果，犯罪に関連する単語や文章を含んでおり危険だと判断され，規制されたツイートは 71 件中 52 件となり，およそ 7 割は止めることができた．この実験には，実際に冗談でツイートされ炎上しまとめられていたものから抜粋したものを利用した．

更に，逆に特に危険性がない通常のツイートをする際に規制されてしまったという件数に関しては 51 件中 1 件であり，通常のツイートをする際にもほぼ弊害はないという結果となった．

7. おわりに

現在 Twitter のような多くの SNS が流行している．それらを利用する目的はユーザーにより異なるが，Twitter において冗談やいたずらの目的で犯罪に関連するツイートをするのは，ただ自らのタイムラインを炎上させたり，所属する団体の周囲からの評価を下げるだけである．本研究では，犯罪に関連するツイートを規制するシステムを提案した．実験結果から，本研究で提案したシステムを利用すると本来のツイートをする際の邪魔をせずに，対象ツイートを規制することができるという結果を得られた．

今後は，システムの機能の改善，具体的にはクライアントとしての機能の搭載，ユーザーインターフェースの改良を行っていきたい．

参考文献

1) Twitter
http://twitter.com/
2) 今更聞けない Twitter の常識：Twitter とは 国内で”再流行”，一般化の兆しも
http://www.itmedia.co.jp/news/articles/0907/28/news011.html
3) なんちゃって鶯卿 半世紀 recollection
http://uguisu.skr.jp/recollection/index.html
4) 有吉さんに Twitter で「今から放火しまーす」と放火予告！ネットユーザー「放火って犯罪と同じぐらいの重罪」
http://rocketnews24.com/2011/12/27/167392/
5) 日本はツイッター入門を必修にすべき？！今度は学生が「犯罪のデパート」公表
http://www.j-cast.com/2011/07/27102642.html
6) Twitter 炎上まとめ～口は災いの元～
http://matome.naver.jp/odai/2130600556121282301
7) Twitter サポート品質を握る新たな体制と運営支援ツール ～その２～
http://japan.internet.com/wmnews/20111221/1.html
8) MeCab：Yet Another Part-Of-Speech and Morphological Analyzer
http://mecab.sourceforge.net/

ⓒ 2012 Information Processing Society of Japan