Twitterの表示系を発展させスパム発見機能を強化したアプリケーションLookUpperの開発と評価

若井一樹[†1]　岡田泰輔[†2]　鎌田祐輔[†3]　佐々木良一[†4]

インターネットの普及に伴いWebサービスはその数を増やし，現在では様々な種類のサービスが人々に利用されている．世界中の人々とコミュニケーションが可能となった現代において，ソーシャルメディアは発達し注目されるようになった．例としてTwitterやFacebookなどがあり，現在でも利用者は急激に増えている．
一方でソーシャルメディアを利用した迷惑行為も増えている．その一つとしてスパム行為が挙げられる．スパム行為とは本人の許諾を得ず，一方的に営利目的の情報を送る行為である．スパム行為によってソーシャルメディア利用者は必要としない情報が大量に送られ，本来知りたい情報が埋もれてしまう．
またTwitterでは，ツイートなどは最新のものから順に表示する方法をとっている．その表示方法ではスパムを発見しても，フォロー解除やブロックなどの管理が困難である．つまり，それら管理機能が容易にできる表示系が必要である．
そこで著者らはスパム行為に悩まされることなくソーシャルメディアを活用し，世界中の人々とコミュニケーションを楽しむことを可能とするため，スパム行為を検知する手法およびその機能を持ち合わせたアプリケーションの開発を考案した．本稿では，スパム行為を検知する手法の提案と評価結果を報告するとともに，これらのスパム検知機能に表示系を発展させることによりスパム発見を容易とするように実装したアプリケーションの開発について述べる．

Development and Evaluation of the application “LookUpper” to improve the spam discovery function and display system of Twitter

KAZUKI WAKAI[†1]　TAISUKE OKADA[†2]
YUSUKE KAMATA[†3]　RYOICHI SASAKI[†4]

1. はじめに

インターネットの普及に伴い Web サービスはその数を増やし，現在では様々な種類のサービスが人々に利用されている．世界中の人々とコミュニケーションが可能となった現代において，ソーシャルメディアは発達し注目されるようになった．例として Twitter や Facebook などがあり，現在でも利用者は急激に増えている．
一方でソーシャルメディアを利用した迷惑行為も増えている．その一つとしてスパム行為（以降スパムと記す）が挙げられる．スパムとは本人の許諾を得ず，一方的に営利目的の情報を送る行為である．
スパムによってソーシャルメディア利用者は必要としない情報が大量に送られ,本来知りたい情報が埋もれてしまう．またそのような不必要な情報が大量に送信されるとネットワークに負荷がかかり，インターネットを利用する多くの人にとっても迷惑となる．

スパムはもともと電子メールやブログを使ったものが多かった.五島ら[1]のWWWシステムにおけるブログスパム検知に関する研究に代表されるように，スパム検知するための研究は多い．そのためスパム業者はソーシャルメディアの流行ともに標的を電子メールなどからソーシャルメディアへ移し，その数を増やしている[2]．特に Twitter では他のソーシャルメディアよりも容易にスパムアカウントを作りやすいと指摘されており，スパム行為が多い[3].

また Twitter の表示系は自身のフォローやフォロワー,各ユーザのツイートなどを最新のものから順に羅列して表示させるものが多く，その中でスパムを発見しても，フォロー解除やブロックなどの管理が困難である．フォローなどを最新のものから表示させるのではなく一括で表示させ，その中でスパムだけを見やすく表示するような工夫が必要である．

そこで，著者らはスパム行為に悩まされることなくソーシャルメディアを活用し，世界中の人々とコミュニケーションを楽しむことを可能とするため，スパムを検知する手法を考案した．

本稿ではソーシャルメディアの中でもスパムが多いとされる Twitter において,複数項目からスパムかどうかを総合的に判定する方式の提案と評価と，それらの機能を持ち合わせたアプリケーションの開発について述べる．

2. 先行研究

Twitter についての研究は数多く行われている.石井[4][5]は Twitter における日本人利用者の特徴についての報告や,

†1 東京電機大学大学院
†2 東京電機大学（現在，ニフティ株式会社）
†3 東京電機大学（現在，株式会社ビービーシステム）
†4 東京電機大学 教授

Facebook と Twitter の発言における特徴語の比較に対する報告を行っており，Twitter における日本人の特徴や他のソーシャルメディアとの差異に関して研究されている．榊ら[6]は Twitter のリスト機能の応用についての報告を行っており，Twitter だけに限っても多くの特徴があることがわかる．吉田ら[7]は Twitter におけるリンクを含むつぶやきの分析を行っており，ニュースなどよりも写真や動画など娯楽的サービスの URL が投稿されやすいと報告している．また Twitter をマイクロブログの 1 つと捉えての研究も行われている．Akshay ら[8]や Alexander ら[9]はマイクロブログ，主に Twitter を利用したユーザ特性や特定の言語の抽出に関する研究をしており，ユーザのコミュニティレベルで関連付けられる意図の分析や，曖昧な表現な語の分析などを行っている．

上記のとおり Twitter には様々な特徴があり，いろいろな研究が行われている．本研究ではそれら複数の特徴を用いたスパム検知を目的としており，一つの特徴よりも複数の特徴を活用することで，よりスパム検知が可能になると考えている．

3. Twitter

3.1 Twitter について

Twitter とは「ツイート」と呼ばれる 140 字以内の短い文章で情報発信を行う特徴を持つソーシャルメディアの一種である．ユーザはツイートを見たい相手をフォローすることで，フォローした相手のツイートを自分のタイムライン上で見ることが可能となり，画像や動画の投稿機能や URL 短縮機能などの機能が実装されている．これによりユーザは自分の情報を容易に発信し，コミュニケーションを取ることが可能となる．

3.2 Twitter におけるスパムの推移

Twitter は現在なおユーザ数が増加しており，世界中でやり取りされるツイート数も増えている．Twitter 社が公表したデータによると，2010 年 1 月の時点で 1 日に約 5000 万回，1 秒あたり約 600 回ものツイートが発信されている．しかしその内の 2%はスパムツイートである[10][11] ．つまり 5000 万ツイートの内約 100 万ツイートがスパムツイートであり，毎秒約 12 回もスパムツイートがされている．

自身のタイムラインにスパムツイートが流れてくる可能性こそ低いものの，全体として見るとスパムツイートが増える分ネットワークに負荷がかかり，Twitter を使う全てのユーザに迷惑を及ぼしてしまう．

4. スパム検知時の問題点

Twitter におけるスパムは依然多く存在し，スパム排除は必要であると考えられるが，問題点が 2 点存在する．

①スパムは様々な方法で広告を発信する．ある手法で行われるスパムに対し対策しても，すぐに別の手法に切り替えてスパムが継続される可能性がある．つまりスパムの型は一定ではなく対策が一つのみであるとスパムを見逃す可能性があり，またスパムの手法が切り替わると対処できないのである．

②Twitter ではスパムを発見した場合，Twitter 社にスパムであるとレポートを送ることができる．しかし Twitter 社はそのレポートを元にアカウントを凍結するという処置を行うだけであり，まだ明らかにされていないスパムの発見を補助する機能ではない．そのため一目でスパムであるとわかりやすく表示する必要がある．

本研究ではこれらの問題点を解決するためスパム検知システムの評価を行った．次章ではその提案方式について述べる．

5. 提案方式

4 章で述べた問題点を解決するために，様々な手法を持つスパムを検知する機能と，スパムであるとわかりやすく表示する機能の実装を行った．

5.1 スパム検知手法

4 章①の問題点を解決するにあたり，スパムの持つ様々な特徴からの検知を可能とするため，スパムであると判定する項目を複数用意し，さらに項目を自由に組み合わせることにより，様々なスパムの検知を実現可能であると考えた．そこで判定項目を，Twitter 社がスパムアカウントと判断する基準を元に複数作成した[12] ．各項目の詳細を表 1 に示す．

表 1　判定項目の詳細

Table 1　Details of the judgment items.

判定項目 1（ツイート系）	同じ内容をなんども繰り返しつぶやいた場合
判定項目 2（ツイート系）	つぶやきの内容が個人的なものではなく，主にリンクばかりである場合
判定項目 3（フォロー系）	フォローしている数に対しフォロワーの数が極端に少ない場合
判定項目 4（フォロー系）	フォローに対して返してもらっている率（相互フォロー割合）
判定項目 5	特定パターンで加点（フォロー系）
判定項目 6	特定パターンで加点（ツイート系）

判定方式として項目ごとにスパムであるか採点し，合計点をスパムスコアとして算出した．スパムスコアの算出方法についての例を図 1 に示す．

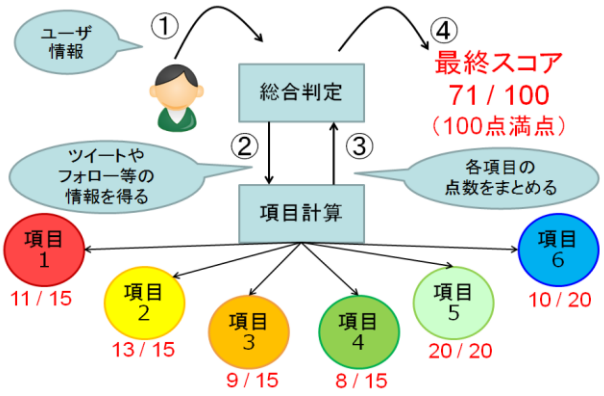

図 1　スパムスコア算出の例

Figure 1　Example of SpamScore.

判定するアカウントからツイートなどの情報を取得し，その情報をもとに各項目で採点する．例えば判定項目 2 では，取得したツイートからどれだけリンクつきツイートをしているかで採点する．そして各項目で採点された結果を加算してスパムスコアとして算出する．スパムスコアは 100 点満点で算出され，スコアが高いほどスパムの可能性は高くなり，スパムスコアが 60 点以上だとスパムであると判定する．

判定項目はツイートに関するもの，フォロー数などに関するものの二種類に分けて，4 個項目を作成した．またツイート系の判定項目，フォロー系の判定項目それぞれにおいて，各項目における採点が高い場合にさらに加点する判定項目を 2 個作成した．

各項目の内容を容易に変更できるようにすることで，新たなスパムの手法にも対応が可能であると考えた．また項目ごとに判定比重を変更可能にすることで，様々なスパムの検知が可能であると考えた．

5.2 インタフェース

4 章②の問題点を解決するにあたり，スパムの検知だけでなく，検知結果をわかりやすく表示することが必要であると考えた．そこで，アカウントと同時にスパムスコアを同時に表示するようにした(図 2)．

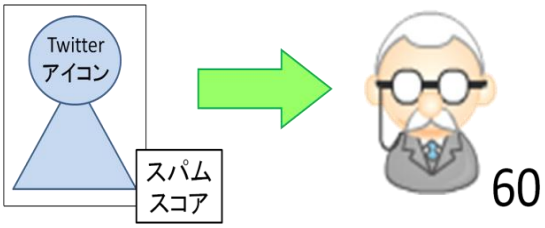

図 2　スパムスコアの表示例

Figure 2　Example displays the spam score.

また他のアカウントと見比べられるようにすることで，どれほどスパムの可能性があるか比較できるようにした(図 3)．図の例では右側の老人のアカウントの方が，左の女性のアカウントよりもスパムの可能性が高いことを表している．

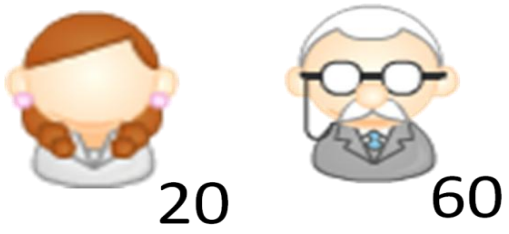

図 3　スパムスコアの比較表示例

Figure 3　Example of a comparison display the spam score.

6. 新たな判定手法

スパムスコアを算出する判定項目は Twitter 社がスパムアカウントと判断する基準を元に作成した．そのため Twitter 社が見逃してしまうスパムは判定し損ねる可能性がある．例えば Twitter 社がスパム報告を受けたにも関わらず，アカウント凍結し損ねているアカウントである．この問題に対し Twitter 社とは別の視点からスパムを検知する手法が必要である．

6.1 新たな判定手法の考案

Twitter 社のスパムの基準は主にフォロー数などからなるもの，リンク付きツイートなどツイート内容に着目したものなどである．しかしこれだけではスパム検知には不十分でありスパムを見逃してしまうと考え，別の視点からの判定手法として，投稿元のクライアント名を使った判定手法，自己紹介文を使った判定手法を考案した．

6.2 投稿元のクライアント名を使った判定

投稿元のクライアント名を使った判定では，ユーザが使用するクライアント名を元に判定する．クライアントとは Twitter のサービスを利用し独自機能を搭載したクライアントソフトウェアの総称である．

投稿元のクライアント名は主に「via ○○」の形で記される．クライアントの例として「web」や「TweetDeck」，「Janetter」などがある．

クライアント名には様々な種類があり，スパムアカウント特有の傾向があると考えた．この判定では投稿元クライアントの数や，他のアカウントも同じクライアントを使っているか，などを判定の基準として実験を行った．

6.3 自己紹介文を使った判定

自己紹介文を使った判定では，アカウントのプロフィール欄のうち自己紹介文を元に判定する．自己紹介文ではユーザが自分の趣味や興味あるものなど自由に記載する．

趣味や興味あるものを記しているアカウントの例として，「○○に興味アリ」など文として記すもの，「/○/△/」など単語を斜線で区切って記すものなどがある．

自己紹介文には様々な記入方法があり，スパムアカウント特有の傾向があると考えた．この判定では広告を宣伝するような語が使用されているか，宣伝広告するサイトへ誘導する URL が記入されているか，などを判定の基準とした．

7. 検証実験

7.1 実験概要

新たに考案した手法を実装し，スパムを検知できるのか検証実験を行った．

実験ではすでにスパムであると公表されているスパムアカウント 29 件を対象に，新たに作成した判定項目でスパムを検知できるか検証を行った．

7.2 実験結果

実験結果を図 4,5 に示す．

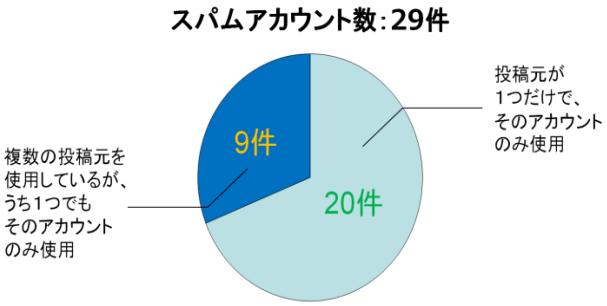

図 4　投稿元のクライアント名を使った判定

Figure 4　Judgment by using the client name of the post.

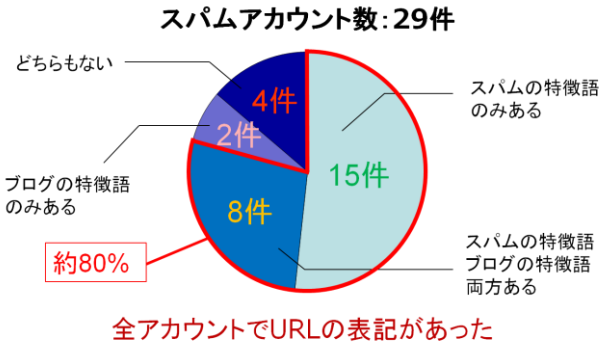

図 5　自己紹介文を使った判定

Figure 5　Judgment by using the self-introduction statement.

投稿元のクライアント名を使った判定について，投稿元が 1 つだけでそのアカウントのみ使用しているものが全体の約 70%であった．またすべてのアカウントにおいて，独自の投稿元クライアントを用いていることがわかった．

以上の結果からスパムアカウントの傾向として，スパムアカウントのみ使用する独自のクライアント名が多い，クライアントを複数使用するが独自のクライアント名を持ち使用しているアカウントが多いことがわかった．

よってこの判定では，投稿元クライアントが 1 つだけでそのアカウントのみ使用しているか，複数の投稿元を使用していても 1 つでもそのアカウントのみ使用している投稿元があるかを判定の基準とする．

自己紹介文を使った判定について，「無料」や「稼ぐ」など宣伝広告で多用される語（以降これらの語をまとめてスパムの特徴語と記す）を使用するアカウントが全体の約 80%であった．また「ブログ」や「Facebook」など自身のサイトを人に紹介する語（以降これらの語をまとめてブログの特徴語と記す）を使用するアカウントも存在した．また全てのアカウントで自己紹介文に URL が含まれていた．

以上の結果からスパムアカウントの傾向として，宣伝広告で多用されるようなスパムの特徴語を使用するアカウントが多い，自身のサイトを人に紹介するようなブログの特徴語を使用するアカウントが存在する，宣伝広告するサイトへ誘導する URL を記入するアカウントが多いことがわかった．

よってこの判定では，スパムの特徴語を多用しているか，URL が含まれているかを判定の基準とする．ただし URL の含まれているとき，ブログの特徴語のみを使用しているアカウントは単に自分のサイトを紹介する一般ユーザの場合もあるため，ブログの特徴語の時は判定の比重を中程度とし，スパムの特徴語のみ使用している時は判定の比重を大きくする．スパムの特徴語とブログの特徴語の両方を使用している場合は，広告宣伝の語で人を煽り宣伝先のサイトへ勧誘していると考えられるので，判定の比重を大きくする．また URL が含まれていない場合において，スパムの特徴語とブログの特徴語の両方を使用している場合は判定の比重を大きく，スパムの特徴語のみ使用している場合は判定の比重を中程度とし，ブログの特徴語のみ使用している場合は判定の比重を小程度にする．

8. 評価実験

7 章の検証実験から，新たに考案した判定でスパムを検知できることがわかった．続いて新たに考案した判定を新規作成項目として実装し，Twitter 社がスパムアカウントと判断する基準を元に作成した判定項目で見逃していたスパムが発見可能か，評価実験を行った．

8.1 概要

現行の判定項目のみ使用した判定に，新たに作成した 2 項目を追加した判定方法で，スパム検知率と誤検知率についての実験を行った．

実験対象として明示されているスパム，明示されていないスパム，一般アカウントの 3 種類のアカウントを用いた．明示されているスパムとは既に公表されているスパムアカウント，明示されていないスパムとはスパムであると公表されてはいないが，明らかにスパム行為をしているスパムアカウント，一般アカウントはスパム行為をしない普通のアカウントとし，サンプル数をそれぞれ 20 件，30 件，30 件，合計を 80 件とした．なお明示されていないスパムは東京電機大学情報セキュリティ研究室の 4 名でスパムであるかどうか判断した．

また本実験において誤検知とは，一般アカウントをスパムであると判定した場合，スパムアカウントを一般アカウントであると判定した場合を指す．

8.2 結果

実験結果を表 2 に示す．

表 2　現行の判定項目に新規項目を追加した実験の結果
Table 2　Results of experiments combined new judgment items and current item.

	現行の判定項目		現行の判定項目 +新規項目	
	正検知	誤検知	正検知	誤検知
明示されている スパム(20)	20 (100%)	0 (0%)	20 (100%)	0 (0%)
明示されていない スパム(30)	21 (70%)	9 (30%)	28 (93%)	2 (7%)
一般アカウント (30)	30 (100%)	0 (0%)	30 (100%)	0 (0%)
合計(80)	71 (89%)	9 (11%)	78 (97%)	2 (3%)

現行の判定項目に新規項目を追加しても明示されているスパムアカウントをすべて検知することが可能であり，一般アカウントを誤検知することもなかった．また現行の判定項目のみのものより 7 件多くスパムアカウントを検知し，検知率は 93%となった．しかしすべての明示されていないスパムアカウントを検知することはできなかった．

8.3 実験結果の考察

検知できなかった明示されていないアカウント 2 件の特徴として，どちらもフォロー数・フォロワー数が多く，ツイート数が比較的少ないことがわかった．また各新規項目から判明したこととして，投稿元のクライアント名を使った判定からは，ボットと併用してスパムを行うアカウントが多いことがわかった．ボットとは一定時間ごとに自動でツイートするアカウントのことである．ツイートを半自動化することで，容易にスパムを行えるようにしていると考えられる．

自己紹介文を使った判定からは，ブログに誘導しようと見られる記載や，相互フォローについての記載されているアカウントが多いことがわかった．特にフォローよりもフォロワーが多いアカウントは，相互フォローについて記載しているアカウントをフォローしているものが多いことがわかった．フォローよりもフォロワーが多いアカウントは有名人など知名度の高いアカウントが多いため，スパムアカウントは相互フォローと記載しているアカウントをフォローし，フォローを返された後，フォローを外すことで，知名度の高いアカウントに似せようとしていると考えられる(図 6)．

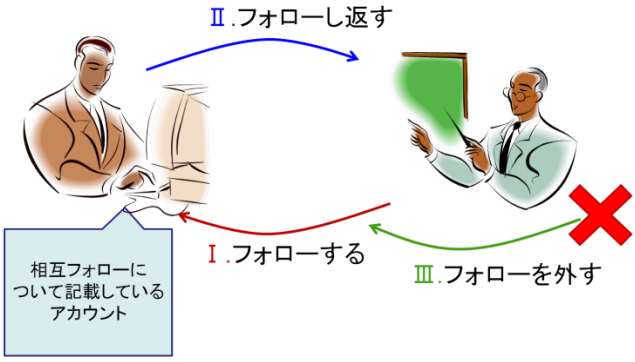

図 6　スパムアカウントのフォロワー獲得イメージ
Figure 6　Image of follower acquisition by spam account.

9. 数量化理論の適用

6 章から 8 章にわたり，スパム判定の新しい項目の考案と各実験について述べた．実験結果から新たに考案した判定項目はスパム発見に有効であることがわかった．しかし単に判定項目を増やし続けても検知しやすくなるのは当然であり，むしろ一般ユーザまでスパムであると検知してしまうなどの誤検知が増えてしまう危険性がある．

そこで筆者らは，どの判定項目の組み合わせが誤検知を最も低くすることができるのか，数量化理論Ⅱ類を適用して検証した．

9.1 検証方法

数量化理論Ⅱ類を適用するにあたり，目的変数は「スパムである」または「スパムではない」の 2 群とした．また説明変数のアイテム数は「判定項目 1」から「判定項目 8」までの計 8 個とし，カテゴリー数は全てその判定項目でスパムを「検知できた」または「検知できなかった」の 2 個とした．適用対象はスパムアカウント 50 件，一般アカウント 50 件，計 100 件のアカウントを用いた．

適用方法は，初めにスパムアカウント，一般アカウントそれぞれ 40 件ずつ計 80 件を用いて数量化理論を適用し，その後残りの 20 件を用いて実際にスパムアカウントと一般アカウントを判別できるか調査する．

初めに実装したすべての判定項目を使用する状態で数量化理論を適用した．使用した判定項目を表 3 に示す．

表 3　使用した判定項目の詳細
Table 3　Details of the used judgment items.

判定項目 1（ツイート系）	同じ内容をなんども繰り返しつぶやいた場合
判定項目 2（ツイート系）	つぶやきの内容が個人的なものではなく，主にリンクばかりである場合
判定項目 3（フォロー系）	フォローしている数に対しフォロワーの数が極端に少ない場合
判定項目 4（フォロー系）	フォローに対して返してもらっている率 （相互フォロー割合）
判定項目 5	投稿元のクライアント名を使った判定項目
判定項目 6	プロフィールの特徴を使った判定項目
判定項目 7	特定パターンで加点（フォロー系）
判定項目 8	特定パターンで加点（ツイート系）

実装したすべての判定項目で数量化理論を適用した場合の判別的中率は 98.8%であった．よってすべての判定項目を用いてもスパム判定は可能であると考えられる．

しかし今後さらにスパムであるか判定する際，項目数が多い状態では誤検知が増えてしまう可能性がある．つまりなるべく少ない項目数で，的中率の高い組み合わせを選択する必要がある．この問題に対し，赤池情報量基準（以降 AIC と記す）を適用することで，判定項目の最適な組み合わせがあるか調査した．

9.2 結果

調査の結果を表 4 に示す．

表 4　AIC による判定項目の組み合わせ
Table 4　Optimal Combination of judgment items based on AIC.

判定項目	AIC	判別的中率
1268	1.673	98.8%

AIC はモデルの当てはまりの程度を表す指標として用いられる[13]．AIC が低いほど当てはまりは良いとされる．

調査の結果，判定項目の全ての組み合わせ 255 パターンの中で最も低い AIC の値は 1.673 となり，その時の組み合わせは判定項目 1,2,6,8 で，判別的中率は 98.8%であった．

あった．

項目数が少なく判別的中率も高いので，判定項目 1,2,6,8 の組み合わせをスパム判定項目の最適な組み合わせとする．表 5 に判定項目の最適な組み合わせを示す．

表 5　判定項目の最適な組み合わせ
Table 5　Optimal combination of judgment items.

判定項目 1（ツイート系）	同じ内容をなんども繰り返しつぶやいた場合
判定項目 2（ツイート系）	つぶやきの内容が個人的なものではなく，主にリンクばかりである場合
判定項目 6	プロフィールの特徴を使った判定項目
判定項目 8	特定パターンで加点（ツイート系）

9.3 調査結果の考察

最適な組み合わせから，フォロー数などを用いた判定がなくても的中率が高いことがわかった．今後はフォロー数などを使った判定でも確度の高い判定を行えるよう改良が必要だと考えられる．

また新たに作成した判定項目が最適な組み合わせとして同時に選ばれることはなかった．そこで新規項目においてスパムアカウントをスパムでないと誤検知したアカウントについて調査したところ，以下 3 点の見解が得られた．

①投稿元のクライアント名を使った判定では，投稿元をボットと併用しているスパムアカウントが多かった．今後ボットと併用しているアカウントの判定比重を大きくする必要があると考えられる．

②自己紹介文を使った判定では，相互フォローについて記載されているスパムアカウントが多かったため，相互フォローと記載しているアカウントの判定比重を大きくする必要があると考えられる．

③全体的にブログに誘導しようとするスパムアカウントが多かったため，ブログに過剰に誘おうとするアカウントの判定比重を大きくする必要があると考えられる．

10. 表示系の発展

前章までスパム発見方法や，判定項目の作成，実験，数量化理論への適用について述べた．様々な特徴を持つスパムを検知するため判定項目は 8 項目実装し，97%の確度でスパムを検知できることを実証した．また数量化理論への適用により全項目を使用しても判別的中率が 98.8%あることを確認し，さらに AIC を適用しても判別的中率を 98.8%に維持したまま最適な判定項目の組み合わせを選出することができた．これはスパム検知に対し有効であると考えられる．しかしスパム検知機能を実装する際，検知結果の表示方法について問題がある．

Twitter を利用し続けていけばユーザ自身のフォローや

フォロワーは多くなっていく．フォローやフォロワーが多い時，スパムが紛れていると探し出すのが困難である．多くの対象の中からでもスパムの発見が可能となる表示系が必要である．また Twitter クライアントの多くはツイートを表示するタイムラインの機能に特化しており，フォローやフォロワーなどの管理に優れているものは少ない．Twitter のインタフェースと差別化を図り，単にタイムラインを表示するのではなく自分のフォローやフォロワーを可視化することで，より Twitter のコミュニケーション機能やフォローやフォロワーの管理機能を有効活用できるアプリケーションが必要である．

この問題を解決するにあたり，これらの機能を実装し Twitter の表示系を発展させたアプリケーションとして LookUpper を開発した．次章で詳細について述べる．

11. LookUpper

11.1 LookUpper について

10 章で述べた問題を解決するアプリケーションとして LookUpper を開発した．

LookUpper では自分のフォローやフォロワーを対象に複数項目からスパムであるか判定し，結果をスパムスコアとして 100 点満点で算出する (図 7)．このスパムスコアと同時にフォローやフォロワーを，自分を中心に同心円状に配置する．このときスパムスコアが小さいほど自分に近い位置に，スパムスコアが大きいほど自分より遠い位置に配置する．

図 7　LookUpper のメイン画面

Figure 7　Main screen of LookUpper.

ユーザが LookUpper を使いスパムアカウントを発見した場合，ユーザはそのスパムアカウントを画面左側にあるメニュー欄にドラッグ&ドロップすることで，フォローの解除やブロックなどの各種管理を行うことが可能となる．また管理だけでなく普通にツイートすることも，リプライやダイレクトメールなどを送ることも可能である．

このように配置することにより，フォローやフォロワーの中にどのようなアカウントがいるのか分かりやすくなる．また中心からの距離が遠くなるにつれスパムの可能性が高くなるため，ユーザが一目でスパムかどうか判断することができる．

既存の Twitter のアプリケーションに多い，単にツイートを表示させるだけのものではなく，わかりやすく表示することに重点を置いた．これにより LookUpper ではスパム検知と同時に Twitter を見て楽しむことを可能にした．

11.2 LookUpper の判定対象

Twitter の API は使用時に制限が設けられている[14]．LookUpper ではこの制限に対処するため，以下の 3 種類のアカウントを判定対象外とした．

①フォロワー20000 人以上のアカウント

②ボットアカウント

③鍵付きアカウント

なお鍵付きアカウントとはユーザ自身のツイートを閲覧許可したフォロワーにのみ公開しているアカウントのことである．

①を判定対象外としたのは，フォロワーが非常に多いアカウントは有名人や公式機関などの知名度の高いアカウントが多いためである． Twitter では 20000 人以上フォローしたい場合，フォロワーの 1.1 倍までしかフォローすることができないという制限[15]があり，フォロワーが 20000 人以上にならなければ，フォローできる人数の上限は 20000 人のままである．

基本的にスパムを発信するアカウントは，自分の宣伝広告を見せるためより多くの人をフォローすると考えられる．ただし 8 章 1 節 2 項で述べたとおり，スパムアカウントの中には意図的にフォロワー数を増やし，フォロー数が極端に多いように見せているアカウントもあるため，LookUpper ではフォローできる人数の上限を増やすことができるフォロワー20000 人を知名度の高いアカウントの目安とし，フォロワー20000 人以上のアカウントは判定対象外とした．

②を判定対象外としたのは，ボットには一定の興味や関心がある集団向けに作られたアカウントが多いためである．例えば大学受験者向けに作られたボットは，受験問題や用語の語呂合わせなどをツイートするアカウントである．これらボットは様々な種類があり，また多数の人から需要があることから，LookUpper ではボットはスパムではない可能性が高いと考え判定対象外とした．ただし 8 章 1 節 2 項で述べたとおり，スパムアカウントにはボットと併用してスパムを行うアカウントもあるので，一概にボットを判定対象外にするのではなく，ボット機能の他に別のクライアントを用いてツイートしていた場合などは判定対象にするなど例外処理を備えた．

③を判定対象外としたのは，ツイートをフォロワーにのみ公開する鍵付きアカウントが，宣伝広告するスパムと反するものだからである．LookUpper では鍵付きアカウントにスパムが存在する可能性は低いと考え判定対象外とした．

12. おわりに

本研究では Twitter におけるスパム発見手法の開発と評価，数量化理論への適用と，これらスパム発見手法用いて Twitter の表示系を発展させたアプリケーション LookUpper の開発を行った．

6 章から 8 章では新たに考案した判定手法の検証実験と評価実験によって，より高い精度でスパムを発見することが可能となり，スパム発見手法として有効であった．

9 章では各判定項目を数量化理論Ⅱ類に適用し，スパムの判別的中率を高い確度で算出する判定項目の組み合わせを選出した．

11 章ではスパム発見手法を Twitter の表示系を発展させたアプリケーション LookUpper に実装し，スパム検知と同時に Twitter を見て楽しむことを可能にした．

今後の課題として，スパム発見手法についてはさらにスパムに対する新たな判定方法を考案していきたい． 2 章でも述べたように Twitter には様々な特徴があり，それらの特徴を使ったスパムの判定方法もあると考えられる．別の視点からスパムであるか判定する手法の考案と評価を重ね，Twitter におけるスパム検知率を向上させたい．また新たな判定項目による実験についてサンプル数を増やして実験し，より精度の高いものにしたい．各項目合計で算出したスパムスコアに関しても同様にサンプル数を増やして実験し，スパムであるか，またはスパムではないかの閾値の設定をしていきたい．

表示系についてはさらに LookUpper を改良し，Twitter におけるスパムの見つけやすさや，コミュニケーションとして楽しむ機能を追求していきたい．

参考文献

1) 五島 優美， 井下 善博， 岡崎 直宣，“WWW システムにおけるブログスパム検知手法に関する研究”，Memoirs of the Faculty of Engineering， University of Miyazaki，Vol.39, pp.279-286(2009)，〈http://hdl．handle．net/10458/3225〉

2) 米シマンテック：「スパムの数が 3 年ぶりに減少した」と調査結果を発表，電子メールからソーシャル・メディア・サイトへ標的を変え，成功を収めたスパム業者，〈http://www.computerworld.jp/topics/632/201263〉（参照 2013-05-14）

3) Washington Post：Twitter reaches 500 million user mark，〈http://www.washingtonpost.com/business/technology/twitter-reaches-500-million-active-users-140-million-in-the-us/2012/07/30/gJQAVdIMLX_story.html〉（参照 2013-05-14）

4) 石井健一，“マイクロブログ Twitter における日本人利用者の特徴”，Department of Social System and Management Discussion Paper Series， No.1277，pp.1-7(2011)，〈http://hdl.handle.net/2241/115334〉

5) 石井健一，“Facebook と Twitter の発言における特徴語の比較”，Department of Social System and Management Discussion Paper Series, No.1279，pp.1-10(2011)，〈http://hdl.handle.net/2241/115339〉

6) 榊剛史，松尾豊，“ソーシャルブックマークとしての Twitter リスト機能の応用”， The 24th Annual Conference of the Japanese Society for Artificial Intelligence，Vol.2010，No.3B3-2，pp.1-3(2010)，

7) 吉田光男，乾孝司，山本幹雄，“リンクを含むつぶやきに着目した Twitter の分析”，DEIM Forum 2010， 第 2 回データ工学と情報マネジメントに関するフォーラム，Vol.2010，No.5A-1，pp.1-3(2010)，〈http://hdl.handle.net/2241/111107〉

8) Java, A., Song, X., Finin, T., and Tseng, B. : Why We Twitter: Understanding Microblogging Usage and Communities, In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis. *ACM*, pp.56-65(2007).

9) Pak, A., and Paroubek, P. : Twitter based system: Using Twitter for disambiguating sentiment ambiguous adjectives, In Proceedings of the 5th International Workshop on Semantic Evaluation, *Association for Computational Linguistics*, pp.436-439(2010).

10) Twitter：Twitter Blog: Measuring Tweets，〈http://blog.twitter.com/2010/02/measuring-tweets.html〉（参照 2013-05-14）

11) Twitter：Twitter Blog: State of Twitter Spam，〈http://blog.twitter.com/2010/03/state-of-twitter-spam.html l〉（参照 2013-05-14）

12) Twitter：Twitter ルール，〈https://support.twitter.com/articles/253501-twitter〉（参照 2013-05-14）

13) 統計 WEB 統計：WEB | Excel やエクセル統計を使った統計解析，〈http://software.ssri.co.jp/statweb2/tips/tips_10.html〉（参照 2013-05-14）

14) Overview： Version 1.1 of the Twitter API，〈https://dev.twitter.com/docs/api/1.1/overview〉（参照 2013-05-14）

15) Twitter：Why can't I follow people? ，〈https://support.twitter.com//forums/10713/entries/66885〉（参照 2013-05-14）