

# マイクロブログ上のサイドビジネス商法による 要注意ユーザの発見手法に関する研究

寺口敏生<sup>†</sup> 田中成典<sup>‡</sup> 高嶋小百合<sup>‡</sup>

関西大学先端科学技術推進機構<sup>†</sup> 関西大学総合情報学部<sup>‡</sup>

## 1. はじめに

Twitter 等のマイクロブログの普及に伴い、詐欺を目的とした投稿を行うユーザが増加している。特に、短期間で多額の利益が発生すると謳い高額な教材を販売するサイドビジネス商法が流行しており、国民生活センターが警告[1]を発している。本研究では、国民生活センターが情報提供しているサイドビジネス商法の特徴や手口に基づき、マイクロブログ上からサイドビジネス商法に関する投稿を行っている要注意ユーザを発見する手法を考案する。

既存研究では、サイドビジネス商法を対象としたものは見当たらなかった。関連分野として、Twitter におけるスパムアカウントを判別する手法[2][3][4]が提案されている。既存研究[2]と発展研究[3]では、Twitter 社が公表する基準[5]に基づき、スパムアカウントの特徴を計 29 種類定義し構築した分類器を用いて判別する手法が提案されている。これらの既存研究では主に、有害なリンクの投稿や同じ投稿内容の重複をスパムアカウントの特徴として重要視している。しかし、これらの既存研究では、スパムアカウントの発見は対象としていない。スパムアカウントの効率的な発見を目的とした既存研究[4]では、スパムアカウント間の言及関係を分析する手法の有用性が検証されている。しかし、この手法では、他のスパムアカウントと連携していないアカウントを発見することが難しい。

本研究では、Twitter 上にてサイドビジネス商法を行う要注意ユーザを自動的に判別・発見することを目的としているため、既存手法をそのまま適用することは難しい。そこで、サイドビジネス商法を行う要注意ユーザを対象とした判別・発見手法を提案する。

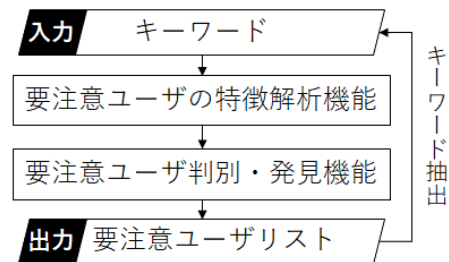


図1 本提案手法の概要

## 2. 研究の概要

本研究では、国民生活センターが公開しているサイドビジネス商法に関する記述から抽出したキーワードを入力とし、要注意ユーザリストを出力する判別・発見手法を提案する。本提案手法(図1)は、要注意ユーザの特徴解析機能と要注意ユーザ判別・発見機能から構成される。

### 2.1 要注意ユーザの特徴解析機能

本機能では、キーワードを基に Twitter 上から収集した要注意ユーザの特徴を解析する。解析の結果、本研究では、①他のサイトへのリンクが含まれる投稿の割合、②全ツイートの投稿内容の多様性、③内容が全く同じツイートの平均重複投稿数、④ツイートの最頻重複投稿数の4項目を要注意ユーザの特徴として扱う。

### 2.2 要注意ユーザ判別・発見機能

本機能では、キーワードを用いて Twitter 上から要注意ユーザを収集し、要注意ユーザリストを作成する。この時、既知の要注意ユーザから得られたキーワードを基に検索するだけでは、誤判定が数多く含まれると想定される。そこで、本研究では、キーワード以外の第2.1節で解析した特徴に基づき、要注意ユーザを判別する手法を提案する。要注意ユーザの判別には、既存手法[2]と同じくデータマイニングツール Weka に内蔵された決定木アルゴリズム J48 を用いる。また、解析結果より作成した要注意ユーザリストに基づき収集した要注意ユーザの投稿からキーワードを抽出する。そのキーワードを用いて新たに収集したツイートを解析することで、日々更新されるサイドビジネス商法の商材の変化に柔軟に対応し、要注意ユーザ候補を的確に発見する技術が実現できる。

Research for Discovering Users to Notice with Side Business Methods on Microblog

<sup>†</sup> Toshio Teraguchi

Organization for Research and Development of Innovative Science and Technology, Kansai University

<sup>‡</sup> Shigenori Tanaka, Sayuri Takashima

Faculty of Informatics, Kansai University

### 3. 実証実験

#### 3.1 実験内容

本提案手法により要注意ユーザが発見できることを検証するため実証実験を行った。本実験では、要注意ユーザの判別結果と要注意ユーザの発見効率との2項目について実験を行った。

要注意ユーザの判別結果に関する実験では、サイドビジネス商法に関する投稿から取得したキーワード(表1)を用いてTwitterから収集したアカウントを解析対象とした。そして、それらのアカウントを対象に目視で判別した要注意ユーザと一般ユーザについて、適切に判別できるかを適合率、再現率とF値の3つの指標を用いて評価した。要注意ユーザの発見効率に関する実験では、発見した要注意ユーザのアカウントから抽出したキーワードを基に提案手法を再適用し、試行回数と対象ユーザ数、および要注意ユーザの割合から、提案手法の発見効率が再適用回数によって、どのように変化するかを評価した。本実験では、提案手法の再適用回数を2回目までとした。

#### 3.2 実験結果

要注意ユーザの判別結果に関する実験結果(表2)より、本研究で扱う4つの特徴を基に判定を行うことで、要注意ユーザを適切に判別できていることが明らかとなった。その一方で、一般ユーザの再現率が低い結果となった。これは、要注意ユーザに見られる重複投稿を行っている一般ユーザが一定数存在するためと考えられる。そこで、今後の発展として、各特徴の判別結果を詳細に分析し、複数の特徴を組み合わせる手法や特徴に重み付けを行い、要注意ユーザのみを効果的に判別する手法の検討に取り組む必要があると考えられる。

要注意ユーザの発見効率に関する実験結果(図2)から、本提案手法は試行回数を増加させることで、より広範のキーワードを対象とした検索が実行でき、多くのユーザを対象として取得することが可能となる。また、実験で求めた要注意ユーザの割合から、取得するユーザ数の増加に伴い、要注意ユーザの発見効率が改善されることが明らかとなった。しかし、試行回数の増加に伴って、一般に使用される語句の含有率が増加すると、誤判定数も増加してしまうことが危惧される。そのため、発見された要注意ユーザの判別基準を更新し続ける必要があると考えられる。

#### 4. おわりに

本研究では、サイドビジネス商法を行う要注

表1 要注意ユーザ発見に用いたキーワード

試行数	特徴語句
1回目	楽, 月収
	簡単, 副業, ネットビジネス
2回目	在宅, 収入
	ビジネス, ネット, お金

表2 要注意ユーザの判別結果

分類	適合率	再現率	F値
要注意	0.850	0.931	0.889
一般	0.820	0.655	0.728

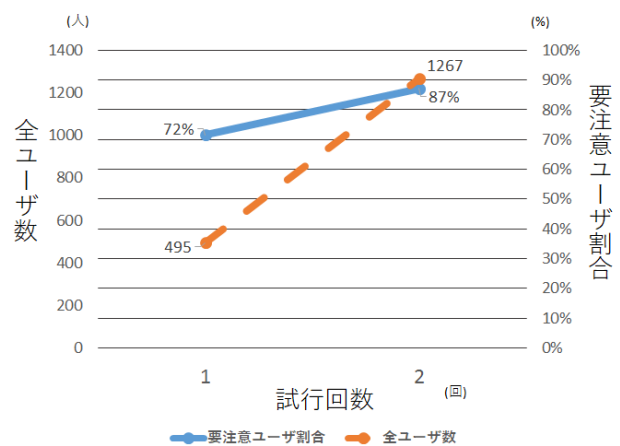


図2 要注意ユーザの発見効率

意ユーザの発見を目的とした手法を提案し、実験を通じてその有用性を検証した。今後の展開として、本研究を通じて明らかになった課題を解決するとともに、サイドビジネス商法以外の悪徳商法にも適用範囲を拡大することを考えている。

#### 参考文献

- [1] 国民生活センター：2016年度版暮らしの豆知識，2015.
- [2] 中村悠一，山田剛一，絹川博之：Twitterにおけるスパムユーザの分別，情報科学技術フォーラム，FIT運営委員会，Vol.10，No.2，pp.157-158，2010.
- [3] 中村悠一，山田剛一，絹川博之：Twitterにおけるスパムユーザフィルタの開発とその評価，情報科学技術フォーラム，FIT運営委員会，Vol.11，No.2，pp.99-100，2011.
- [4] 菊池望，吉村博幸：Twitterにおける言及関係によるクラスタリングを利用したスパムアカウント判定手法の検討，研究報告情報システムと社会環境研究会研究報告，情報処理学会，Vol.2014-IS-130，No.1，pp.1-4，2014.
- [5] Twitter社：Twitterルール，Twitterのポリシー：<http://support.twitter.com/articles/253501>，(入手2017.1.12)