

SNS 上での BOT 検出

A Study on Detecting BOTs in SNS

○岡部茂樹*, 趙強福*

○Shigeki Okabe, Qiangfu Zhao

*会津大学システム知能学講座

*System Intelligence Lab in University of Aizu

キーワード：機械学習, クラスタリング, K-means 法, Twitter, BOT

連絡先：福島県会津若松市一箕町鶴賀

公立大学法人会津大学 システム知能学講座

岡部 茂樹 電話番号：080-6030-1995, Email: m5171105@u-ziau.ac.jp

1. はじめに

近年, ソーシャルネットワーキングサービス(SNS)が世界中で普及している. 個人個人で同じ趣味や地域同士で繋がり, その繋がりをスマートフォンやタブレットなどの端末が普及した事により更に加速した. 個人から情報発信は容易になり, 情報が膨大化する事になった. 個人が自分の探したいものや他のユーザを大量のデータから探せる利点もあるが, 大量のデータから探し出す事は欠点でもある. 多くのユーザの中には, 悪意のあるユーザも存在しており, それらの被害も報告されている. SNS の一つである Twitter サービスでは, 利用者である一般的なユーザの他に Bot サービスというものが存在している. Robot の短縮形で Bot と呼ばれ, 個人法人間

わず製作されユーザに為になる情報を発信している. 個人で製作可能なプログラムである為, 悪意のあるユーザに製作される可能性がある. 意味のない冗長的なメッセージを連続で送る, 個人情報抜き取るサイトに誘導する, 迷惑なメッセージをそのユーザ自身に発信させるなどと実際の被害が確認されている. この問題を解決する一つの方法として機械学習を用いて Bot を検出し, 一般的なユーザと分けよう考えた.

本研究では, 教師なし機械学習の非階層型クラスタリング手法である K-means 法を用いて一般的なユーザと Bot を判別している. ユーザと Bot それぞれのメッセージを基にデータベースを自作し, Bot の特徴を利用しようとしている. Bot の特徴として, プログラムを利用している為発信されるメッセー

ジの内容の重複度合いや定期的な発信に着目し、それらをクラスタリング生成に用いた。

2. 提案手法

ユーザと Bot を判別する為、教師なし機械学習の非階層型クラスタリング手法である K-means 法を用いている。このアルゴリズムは以下のステップで K 個のクラスタリングを生成する。

1. 各データに対してランダムにクラスタを割り振る
2. 割り振ったデータをもとに重心を計算する。
3. 各データと各重心との距離を求め、各データに最も近い重心に割り当て直す
4. 重心の更新がなくなった場合終了する

K-means でクラスタリング生成の為、Bot の特徴を用いた。Bot の特徴として、プログラムから発信されている為メッセージが同じで内容であるという点。そして、定期的な発信である点に着目した。同じ内容である場合、完全に同じメッセージが有るならその数をカウントし、その割合を算出した。定期的な発信の場合、メッセージ間の時間差を用いてその標準偏差を算出することによって求めた。

$$s = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - m)^2}$$

x_i : 発信の時間差

3. 実験

先程上げた特徴を用いて、クラスを分類す

る。ユーザと Bot それぞれのデータベースを作り、1つのデータベースを1人のユーザ、もしくは1つの Bot のメッセージで構築されている。データベース数とそれらに含まれている要素は以下の通りである。

Table 1 Database Name and Number of Data

データベース名	データ数
ユーザ	271
Bot	186

それぞれのデータベースには共通の要素を含み、それらの要素を使い特徴を算出する。以下に要素を示す。

Tweet Text: 140 文字以下の文字列

Tweet At: 発信した時刻

これらのデータを 500 から 3000 まで 500 刻みで特徴として使用し、それぞれの特徴数のクラスタリング生成を行う。

4. 結果と考察

それぞれのデータベースのデータを 500 から 3000 まで 500 刻みで使用した K-means クラスタリングを次の頁に示す。加えて、各使用データ数の正答率を下記の表で示す。

Table 2 Recall of Bots and Users

データ数	Bot	ユーザ
500	0.538	0.967
1000	0.538	0.963
1500	0.511	0.963
2000	0.435	0.956
2500	0.425	0.967
3000	0.425	0.970

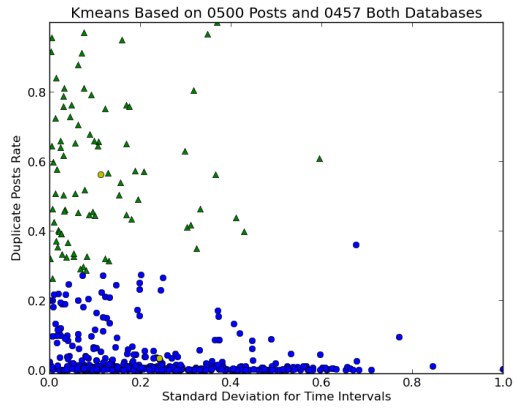


Fig 1 K-means by using 500 data

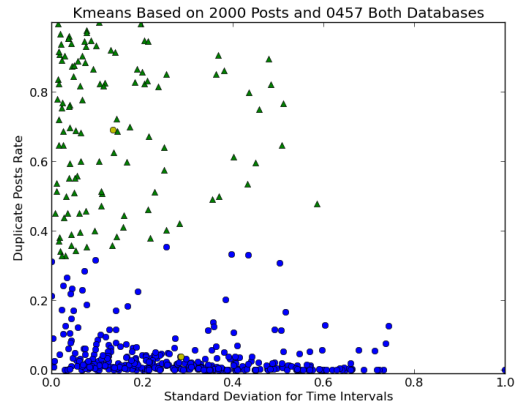


Fig 4 K-means by using 2000 data

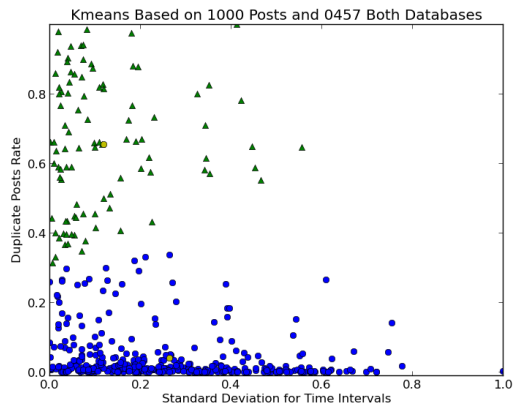


Fig 2 K-means by using 1000 data

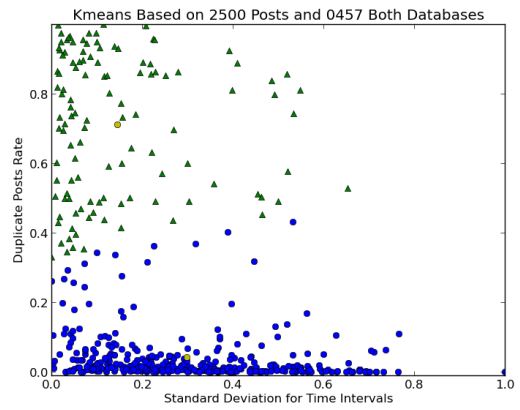


Fig 5 K-means by using 2500 data

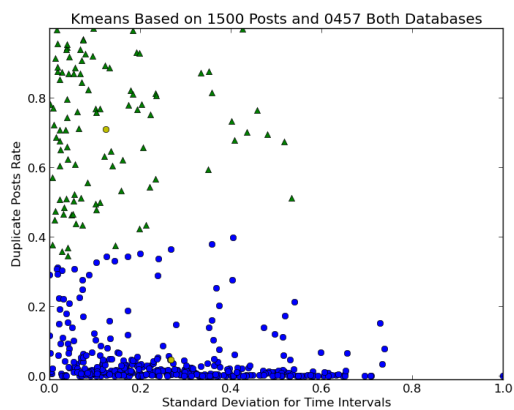


Fig 3 K-means by using 1500 data

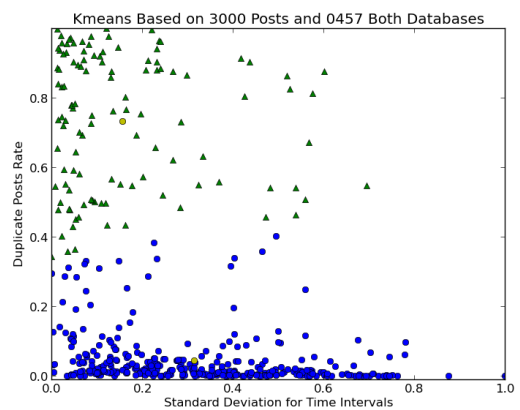


Fig 6 K-means by using 3000 data

一般的なユーザは高い正答率が得られたが、逆に Bot に対しては高い正答率は得られなかった。その理由として、今回の実験の場合同じ内容と定期的な発信を元にクラスタリング生成を行った。ユーザと Bot 共に定期的な発信ではそれぞれ顕著な結果は得られなかった。しかし、同じ内容についてはユーザが多くの場合低い値を示し、広範囲に広がる Bot とは全く異なる結果が得られた。ユーザが同じ内容を何度も発信する可能性は低い事が示された。Bot に対しては同じ内容を発信する Bot が多くを占めていない事が示された。考えられる可能性として、同じ内容を発信するのではなくそれぞれの Bot が持つ規則性によって発信が行われている事である。今回の実験では、完全に一致する場合カウントすることにした為、ある規則性を持ちながら内容を少し変更した場合カウントはされない。よって、規則性を持つ場合別な機械学習を用いてその規則性を学習させれば正答率もよくなると考える。

5. まとめ

今回の研究では、2つの特徴を用いて使用したデータ数毎の K-means クラスタリングを生成した。ユーザクラスの正答率は高い値を示したが、Bot クラスの正答率に問題が残る結果となった。

今後の課題として、Bot の特徴となる仮定した2つの項目を使用してクラスタリングを生成したが、定期的な発信については高い値を示さなかった。よって、この特徴以外の Bot の特徴として使える項目を用いて新たにクラスタリングを行いたいと考えている。

また、今回のクラスタリングは教師なしの機械学習の1つとして用いたが、別な教師ありの機械学習を用いて、未知の問題に対応出来る様にする必要性を考える。考察で述べた通り、Bot にある規則性が存在していた場合、その都度学習させ適応させる事ができ

ば正答率が向上すると考える。

参考文献

- 1) 人工知能に関する断創録
<http://aidiary.hatenablog.com/>
- 2) 大倉一馬, 塩田茂雄 “Twitter における時系列変化解析に関する研究” (2013)
- 3) GeSource
<http://www.gesource.jp/>