

## ツイート履歴の印象極性の動向を基にしたアカウント分類

山田 紘司\*

\*東京都市大学大学院工学研究科

延澤 志保\*

\*東京都市大学知識工学部情報科学科

### 1 はじめに

ツイッターが提供している情報収集機能にリアルタイム検索がある。これはユーザの指定した検索クエリを含むツイートやアカウントを検索することができる。しかし、検索クエリを含むツイートをリアルタイムに集め表示するため、ユーザにとって有益なツイートがそれ以外のツイートに埋もれてしまうという問題がある。さらにそこからユーザにとって有益なツイートをしているアカウントを見つけることはさらに困難である。本研究ではアカウントの時系列情報とツイートの印象極性に着目し、ある特定の話題についてツイートをしているアカウントの分類について検討する。

### 2 関連研究

リアルタイム検索では、検索クエリを含むツイートをリアルタイムに集め表示するため、ユーザにとって有益なツイートがそれ以外のツイートに埋もれてしまう。そのため、ツイートやアカウントをツイート内容などによりある程度分類することができれば目的のツイート、アカウントが見つかる作業が容易になる。

既存の文書分類では教師データを基にクラスを学習させていくのが一般的である [1]。ツイートについて話題で分類しようとするとかかなり多くの教師データを用意する必要があり、現実的ではないため教師なし学習をする必要がある。

そこで、ツイートを分類するのではなくアカウントを分類することに着目した。アカウントを分類するためにはアカウントの特徴が必要である。本研究で着目したアカウントの特徴は一つの検索クエリを含むツイートの数、その印象極性と時系列である [2]。

本研究ではアカウントの時系列情報とツイートの印象極性に着目し、ある特定の話題についてツイートをしているアカウントの分類について検討する。

### 3 実験

アカウントの時系列情報とツイートの印象極性がアカウントの分類に有効な特徴なのかを調査した。調査には2016年に社会現象となった「ポケモン GO」についてツイートしている36アカウントの6月から11月までのツイート履歴を対象として行った。

あらかじめ36アカウントのツイート履歴から手作業で以下の5種類に分類した。

- 「ポケモン GO」のニュースを取り上げているアカウント
- 「ポケモン GO」のニュースの中でも事故や事件のみを取り上げているアカウント
- 「ポケモン GO」を流行りとして取り上げているアカウント
- 「ポケモン GO」をゲームの一つとして取り上げているアカウント
- 「ポケモン GO」についてだけ取り上げているアカウント

この分類されたアカウントが分類ごとに似た形のグラフを作ることがあれば、グラフを構成する情報は分類に有効な特徴であると考えられる。これらのアカウントをツイート数と時系列を軸に取ったのが図1、図2である。

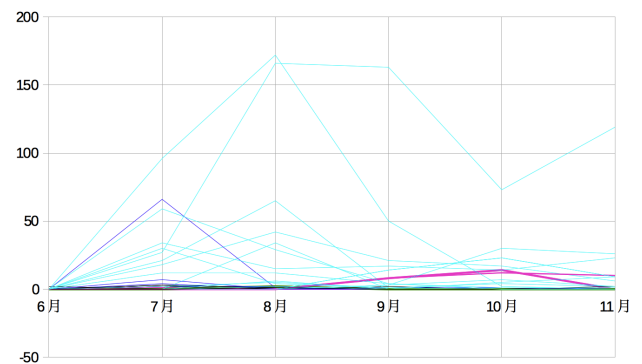


図1: アカウントごとの「ポケモン GO」に関するツイート数とその動向

Account Classification Based on Impression Polarity Shifts in Tweet Archives

Koji Yamada\* and Shiho H. Nobesawa\*.

\* Graduate School of Engineering, Tokyo City University

\* Faculty of Knowledge Engineering, Tokyo City University

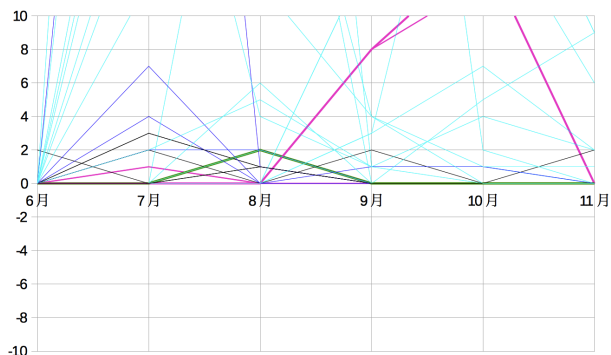


図 2: 図 1 を拡大したグラフ

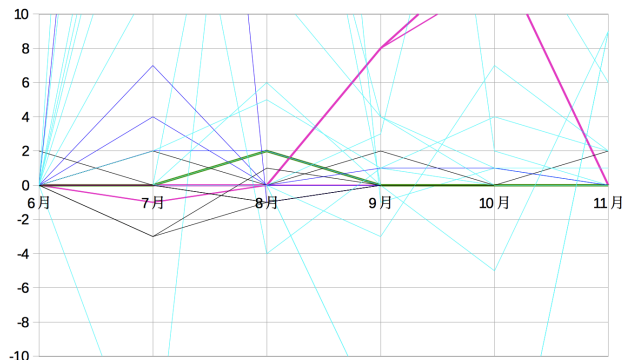


図 4: 図 3 を拡大したグラフ

このグラフからは7月から8月にかけて「ポケモン GO」が大きな話題になっていることがわかる。しかし、これだけではアカウントが分類できているようには見えない。

次にこれらのアカウントの「ポケモン GO」についてのツイートの印象極性を日本語評価極性辞書 [3][4] を用いて取得する。このグラフに印象極性を加えたグラフが図 3, 図 4 である。これらのグラフは、各アカウ

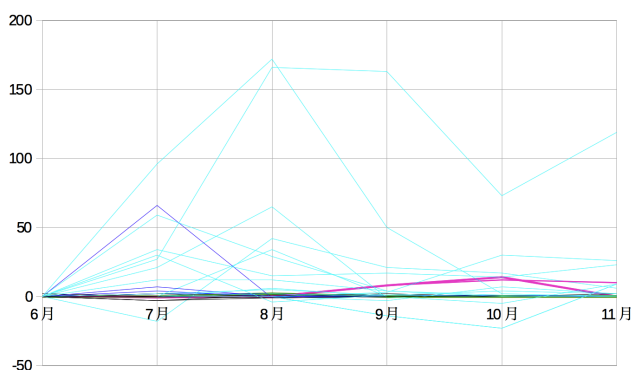


図 3: アカウントごとの「ポケモン GO」に関するツイート数, 印象極性とその動向

ントごとにポジティブなツイートが多い月のツイート数を正, ネガティブなツイートが多い月のツイート数を負で表している。図 1, 図 2 は図 3, 図 4 に比べあらかじめ分類したアカウント同士が似たグラフの形を取っており, 印象極性を用いることでツイート数と時系列のみのものよりもアカウントの特徴がグラフに現れている。

#### 4 まとめ

アカウントの時系列情報とツイートの印象極性に着目し, ある特定の話題についてツイートをしているアカウントの分類について検討した。ツイート数と時系列を用いた結果, 特定の話題がどの時期に盛んになっているのかなど話題の動向を読み取ることはできるが, これだけではアカウントの分類に有効ではなかった。また, ツイート数と時系列に印象極性を用いることでアカウントの分類に有効であることがわかった。

#### 参考文献

- [1] 古宮嘉那子, 高村 大也, 奥村 学. "文書分類のための Negation Naive Bayes." 自然言語処理 Vol.20, No.2 pp.161-182, 2013.
- [2] Koji Yamada, Shiho Nobesawa "Visualization of Keyword Transition for Twitter Account Recommendation", 2016 International Symposium on Multimedia and Communication Technology, pp125-128, 2016.
- [3] 小林的ぞみ, 乾健太郎, 松本裕治, 立石健二, 福島俊一, "意見抽出のための評価表現の収集", 言語処理学会 Vol.12, No.3 pp.203-222, 2005.
- [4] 岩崎祐貴, 折原 良平, 清 雄一, 中川 博之, 田原 康之, 大須賀昭彦. "CGM における炎上の分析とその応用." 人工知能学会論文誌 Vol.30, No.1 pp.152-160, 2015.