

ファンコミュニティ分類に基づくコミュニティ間情報補完手法

中川 竜助[†] 灘本 明代[†]

[†] 甲南大学 知能情報学部 〒658-8501 兵庫県神戸市東灘区岡本 8-9-1

E-mail: [†]ts1671087@s.konan-u.ac.jp, ^{††}nadamoto@konan-u.ac.jp

あらまし 近年プロスポーツチームにおける“にわかファン”と呼ばれるライト層のファンに注目が集まっている。にわかファンの大きな特徴として世間のブームが去った後には、そのチームに関心を持たなくなってしまうことが挙げられる。そこで本研究では、コアファンしか知らない情報を自動で抽出しにわかファンに提示することでファンの定着を促進することを目的とする。具体的には、Twitter からユーザのネットワークとツイートの内容からコアファンとにわかファンを分類する。そしてコアファンしか知らないような情報を抽出し、にわかファンに提示する手法の提案を行う。

キーワード Twitter, ファン分類, にわかファン, 情報補完

1 はじめに

野球やサッカー、ラグビーなど様々なスポーツのチームにはお気に入りのチームを応援するファンが存在する。それらファンの層は大きくコア層とライト層に分けられる。近年では、ライト層のファンに当たる人々の一部を「にわかファン」と呼び、その年話題となった新語・流行語を決定する年末恒例の『2019 ユーキャン新語・流行語大賞¹』にもノミネートされるなど、メディアで大きく注目されている。にわかファンの特徴として、世間の盛り上がりに乗じて応援することや、チームやそのスポーツに対しての知識が少ないといったことが挙げられる。そのためにわかファンは世間のブームが去った後に、チームやそのスポーツに対しての関心を失い、ファン離れを引き起こしやすい傾向にある。そこで我々は、このにわかファンに、流行り廃りに左右されずファンであり続けてもらうためには、まずチームに対する愛着を持てもらうことが必要であると考えた。チームへの愛着を深めるためには様々なアプローチが考えられるが、特にコアファンしか知らない情報をにわかファンに提供することにより、ファン離れ抑制に繋がるのではないかと考える。しかしながら、このようにコアファンしか知らない情報をにわかファンが知ることは容易ではない。そこで本研究では、コアファンしか知らない情報を自動で抽出しにわかファンに提示することでファンの定着を促進させることを目的とする。本研究では、にわかファンが知らずコアファンしか知らない情報を「コア情報」と呼び、コア情報の抽出手法の提案を行う。コア情報を抽出する際、利用ユーザが多くリアルタイムな情報発信が行われている SNS である点、テキスト形式での情報発信が主流である点、そしてコアファンとにわかファン関係なく気軽にチームに関しての内容を発信できる点の特徴より、Twitter を対象としてコア情報を抽出する。

Twitter 上のコアファン同士のコミュニティから、ユーザをチームのコアファンであるかにわかファンであるか判別を行う。

本研究の全体の流れを以下に示す。

- (1) ファンコミュニティの分類
- (2) チームに関するツイートを抽出
- (3) クラスタリングによるトピック抽出
- (4) コア情報を抽出

本研究では、近年人気が高まり新規ファン（にわかファン）の獲得が続いているプロ野球球団の広島東洋カープのファンを対象として、チームや選手のコア情報の自動抽出を行う。

以下、本論文では 2 章で関連研究について述べ、3 章ではコアファンとにわかファンの定義について述べ、4 章ではユーザのネットワークとツイート内容を用いたファン分類の手法について述べる。そして、5 章でコア情報抽出手法について述べ、6 章で提案手法の評価実験を行い、その考察について述べる。最後に 7 章で本論文のまとめと今後の課題について述べる。

2 関連研究

Twitter ユーザからコアなアカウントの抽出を行う研究として、田沼ら [1] は Twitter 機能の 1 つであるリプライに着目して、特定分野に関する語を含む割合からコアなアカウントの発見する手法を提案している。この研究はユーザ間コミュニケーションを基にした研究で、特定分野に関するツイートのリプライが多ければ多いほど発見精度が上がることを報告している。アカウント発見に関して、「分野に関するワードを含む割合の高いアカウント（コアアカウント）を発見する手法」と「コアアカウントのリプライに着目し、同様の分野に関するワードがリプライ内で含まれる割合を用いる手法」の 2 手法を統合することにより、コアアカウントであるユーザ自身のツイートの形態素解析結果では発見が困難な「特定分野に濃いユーザアカウント」を発見できることを報告している。それに対し、本研究では、リプライのみではなくユーザの発信しているツイート全てを対象とし、コアなアカウントを発見する手法の提案をしている点で異なる。

Twitter 上のコミュニティ抽出の研究として、新保ら [2] はリ

¹ : <https://www.jiyu.co.jp/singo/>

ツイートを指標としてネットワークによるコミュニティ抽出を行っている。また谷口ら [3] は、Blog のリンクを自動収集するプログラムを作り自動収集させ、それらのリンクから PageRank と BetweennessClustering によりコミュニティの抽出・分析を行っている。それに対し、本研究では、フォロー関係を指標としコアコミュニティの抽出を行い、ユーザの発信する全ツイートから特定分野に関するツイートの総数とその割合を算出し、事前に決定した条件を満たすユーザをコアファンとしファン分類を行う点で異なる。

ツイートの話題の分類に関して西田ら [4] は、ツイートのデータ圧縮のしやすさに基づき、あるツイートが着目した話題に関するツイートかどうかの分類を行っている。本研究では、ツイート文中に出現する特徴語からルールベースにより着目した話題に関するツイートであるか判別する手法を提案している点で異なる。

Twitter から閲覧者にとって有益な情報の抽出を行う研究として、辻ら [5] は、閲覧者のタイムライン上の頻出語から興味を推定し、新着ツイート中に含まれる閲覧者にとって興味のありそうな話題のツイートの抽出を行っている。また Ren ら [6] は、時系列変化とユーザ間のつながりに基づくトピックモデルを提案しトピック抽出を行い、有益と判断されたツイートに対し Wikipedia の概要部分を利用し要約と説明を行っている。それに対し、本研究では、フォロー関係に着目し、ユーザをコミュニティによる分類を行った後に、それらのコミュニティ間の差分情報をコサイン類似度を用いて抽出することによって、有益な情報を抽出する点で異なる。

意外性のある情報抽出に関連する研究として、佃ら [7] は人物名、地域名、製品名、施設名、および組織名の 5 つのカテゴリの主題語を用い、Wikipedia を用いて、それぞれの同位語間の関係を考慮し、意外な情報を発見する手法を提案している。また伊藤ら [8] は、ウェブサイト All About のデータをコーパスとし、ユーザの発言から名詞、動詞、形容詞、未知語を特徴語として抽出することで、各記事との関連度、意外性を算出し、ユーザに意外性のある記事を提示する手法を提案している。本研究ではファンのツイートに着目し、ファン層によるツイートの違いからコアファンにしかない情報を抽出する点で異なる。

3 コアファン・にわかファンの定義

プロスポーツチームのファンは、そのチームに対しての思い入れにより分類すると、大きく“コアファン”と“にわかファン”に分類される。一般的にスポーツチームにおけるコアファンとは、あるチームに対し、ファンとして強い情熱を注ぎ続け、流行り廃りに左右されずにファンであり続けている人々のことを指す。その一方でにわかファンとは、巷で大いに話題となっているチームに対しテンポラリなファンであり、世間の盛り上がりに乗じて振舞う人々を指している。しかし、Twitter ユーザからコアファン・にわかファンを抽出するためにはユーザの特徴を定量的に分析する必要がある。そこで我々は、コアファンはチームに関する知識量が豊富であることや、長年ファンで

あり続けているためコアファン同士のネットワークを形成していることに着目する。本研究におけるコアファンとは、チームに対しての知識を十分に有しており、チームに対する知識が豊富な人同士のコミュニティに属しているファンとする。これに対し、にわかファンとは、チームに対しての知識をあまり有しておらず、かつコアファンのコミュニティに属していないファンとする。

本研究で用いるコアファンの条件を以下の表 1 に示す。本研究では、コアファンの中でもより熱狂的なユーザのみを抽出したいための条件である。熱狂的なユーザのみをコアファンとした理由は、コア情報を出す際に、コアファンでもなくにわかファンでもない中間層のファンがコアファンと判定されることを避けるためである。

表 1 コアファンの条件

	条件
条件 1	指定したコアファンと相互フォロー関係である
条件 2	チームツイートの割合が 50%以上である
条件 3	チームツイートの総数が 1000 件以上である

以下にコアファンの条件の詳細を述べる。

条件 1 指定したコアファンと相互フォロー関係である
コアファンのフォロー関係を調べ、相互フォロー関係であるユーザをコアファン候補とする。

条件 2 チームツイートの割合が 50%以上である
コアファンの相互フォロー先は必ずしも同じチームのファンであると限らないため、そのユーザがチームのファンであるか否かを判別するために、条件 1 を満たしたユーザのうち、そのスポーツチームに関連する内容のツイート（以下、チームツイート）が全体の半数を超えていたら、そのチームのコアファンである可能性が高いと考える。

条件 3 チームツイートの総数が 1000 件以上である
条件 1 のユーザが条件 2 を満たしていたとしても、全体のツイート数が少ない場合はコアファンであるとはいいがたい。そこで、事前に手動でコアファンと思われる人を無作為に調査したところ、チームツイート数が 1000 を超えているユーザが多いことより、条件 3 を設定した。

条件 1 を満たすコアファン候補の内、条件 2 と条件 3 を満たすユーザをコアファンとして抽出する。

次に、本研究で用いるにわかファンの条件を以下の表 2 に示す。

表 2 にわかファンの条件

	条件
条件 1	相互フォロー先にコアファンが存在しない
条件 2	チームツイートの割合が 10%以上 50%未満である
条件 3	チームツイートの総数が 50 件以上 1000 件未満である

以下ににわかファンの条件の詳細を述べる。

条件 1 相互フォロー先にコアファンが存在しない
指定ユーザのフォロー関係を調べ、相互フォロー関係である

ユーザにコアファンの条件2と条件3を満たすユーザがいないユーザをにわかファン候補とする。

条件2 チームツイートの割合が10%以上50%未満であるにわかファン候補がチームのファンであるか否かを判別するために、チームツイートの割合が全体の10%を超えていたら、そのチームのファンである可能性が高いと考えられる。しかしながら、その割合が50%を超えていたら、コアファンである可能性が高くなるため、チームツイートの割合が10%以上50%未満とした。

条件3 チームツイートの総数が50件以上1000件未満である条件1のユーザが条件2を満たしていたとしても、全体のツイート数が少ない場合はチームのファンであるとはいえない。そこで、事前に手動でチームのファンであるアカウントと思われる人を無作為に調査したところ、チームツイート数が50件を超えているユーザが多いことより、条件2を設定した。しかしながら、その数が1000件を超えていたら、コアファンである可能性が高くなるため、チームツイートの総数が50件以上1000件未満とした。

条件1を満たすにわかファン候補の内、条件2と条件3を満たすユーザをにわかファンとして抽出する。

4 ファンコミュニティ分類

4.1 ファンコミュニティ分類のシステムフロー

コアファンとにわかファンの分類を各々の条件に基づき行う。コアファンとにわかファン両方の抽出において、各々の条件1ではユーザのフォロー関係に着目し、相互フォロー関係のユーザにコアファンの条件を満たすユーザが存在するかどうかによりコアファンかにわかファンかを決定する。さらに各々の条件2と条件3においては、チームツイートの総数と全ツイートに対する割合からコアファンとにわかファンを分類する。図1の(a)にコアファン抽出のシステムフローを、図1の(b)ににわかファン抽出のシステムフローを示す。

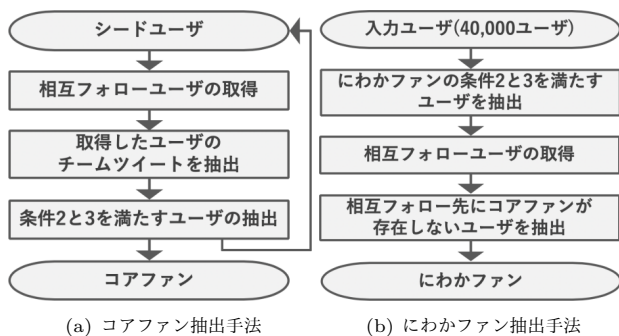


図1 ファンコミュニティ分類のシステムフロー

4.2 フォロー関係に基づくコアファンとにわかファンの決定 コアファンの決定

フォロー関係に基づくコアファンの決定は、はじめに任意のユーザを指定し、これをシードユーザとする。このシードユーザは、チームのコアファンであると主観で判断したユーザを用

いる。このユーザのフォロー関係を、TwitterAPIのFollowers/idsとFriends/idsを用いて取得する。このフォロー関係において相互フォローユーザをコアファン候補とする。次にコアファン候補のツイートを取得し、ツイートからチームツイートを抽出した後に、条件2と条件3を満たすユーザをコアファンとする。チームツイートの抽出については4.3章で詳しく述べる。このシステムを繰り返す行うことで、多くのコアファンを抽出することを可能としている。2回目以降の入力ユーザは、前回判定されたコアファンとする。実際に、提案手法を3回繰り返し、コアファン476ユーザを獲得した。

にわかファンの決定

事前に用意した任意のチーム専門用語集を検索クエリとしてそのチームに関連する内容のツイートを取得し、そのツイートを発信しているユーザから無作為に抽出した40,000ユーザを入力ユーザとする。その入力ユーザのツイートを取得し、ツイートからチームツイートを抽出した後に、条件2と条件3を満たすユーザをにわかファン候補とする。次に、にわかファン候補のフォロー関係を取得し、相互フォローユーザを抽出する。最後に、抽出されたにわかファン候補の相互フォロー先にコアファンの条件2と条件3を満たすユーザが存在しなければ、そのユーザをにわかファンとして決定する。実際に提案手法により、入力ユーザ40,000アカウントの内、1,846ユーザをにわかファンとして獲得した。

4.3 チームツイート抽出

コアファン、にわかファン各々のツイートがすべてチームに関連する内容のツイートであるとは限らない。そこで本研究では、チームツイートの抽出手法を提案する。具体的には、ルールベースによる手法と機械学習による手法の、2種類のチームツイートの抽出手法を提案する。

ルールベースによるチームツイート抽出

ルールに用いる辞書として以下の3つの辞書を用いる。

- (1) チーム確定語辞書 [選手フルネームなど195語]
- (2) 選手名辞書 [姓・名別288語]
- (3) 野球用語辞書 [356語]

チーム確定語とは、その単語が含まれていれば、そのチームに関連する内容のツイートであると断言できる語のことである。例えば、所属選手のフルネーム等のチームを特徴づける単語であり、広島東洋カープの場合、チームが勝利した際に頻繁にツイートされる「こいほー」という言葉が該当する。チーム確定語辞書は、事前に手動でチームのコアファンであると思われるユーザのツイートを調査し、そのチームのツイートだと思われるツイートに共通する単語を抽出して作成した。選手名辞書は、チームに所属している選手や監督、過去に所属していた有名な選手など、姓名別で作成された辞書である。野球用語辞書とは、プロ野球で使用される野球用語を集めた辞書であり、“プロ野球データ管理室 プロ野球用語辞典²⁾”というWebサイトを参考として作成した。ルールベースによるチームツイート抽出手

2: <http://www.din.or.jp/nakatomi/rule/>

法の流れについては、以下の図 2 に示す。

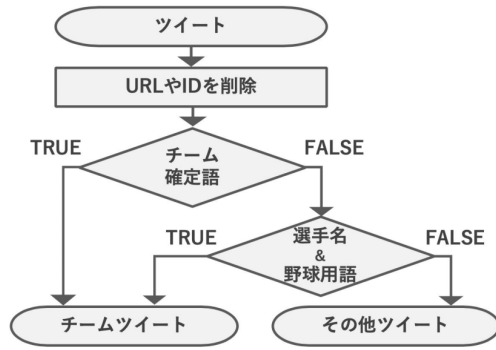


図 2 ルールベースによるチームツイート抽出手法の流れ

ルールベースによるチームツイート抽出手法では、まず、ツイート内にアカウント ID や URL が含まれている場合、誤判定の原因となるため正規表現を用いて削除する処理を行う。次にチーム確定語を含むツイートをチームツイートとして抽出する。さらに、チーム確定語を含まないツイートから、選手名と野球用語どちらも含むツイートもチームツイートとして抽出する。実際にファンコミュニティ分類によって獲得したコアファン 476 ユーザ、にわかファン 1846 ユーザのツイートを取得して、チームツイートを抽出したところ、コアファンからは 557,179 件、にわかファンからは 501,439 件のチームツイートを獲得した。

機械学習によるツイートの分類

Support Vector Machine(SVM) を用いてチームツイートを抽出する。形態素解析器は Mecab³ を用い、全品詞を対象として Word2Vec⁴ によりベクトル化を行う。形態素解析には新語や固有表現に強い mecab-ipadic-Neologd⁵ の辞書を用いる。SVM の実装には LIBSVM⁶ を使用し、カーネルは RBF カーネルを用いる。ハイパーパラメータはグリッドサーチにより、コストパラメータ C は 2048.0、カーネルパラメータ g は 0.03125 と決定する。

5 コア情報抽出

5.1 クラスタリングによるトピック抽出

トピック毎にコアファンとにわかファンの差分情報であるコア情報を抽出するために、ツイートをトピック毎にクラスタリングする。本研究では、シンプルな構成で大規模なデータでも実用的なスピードで実行できるクラスタリングツールの bayon を使用する。予備実験より、対象とする品詞を固有名詞のみとする。また入力 of 固有名詞は Word2Vec によりベクトル化する。提案手法により獲得した各々のファンのチームツイートの内、無作為に抽出した 500,000 件ずつのツイートを使用し、Repeated Bisection [9] を用いてクラスタリングによるトピ

クの抽出を行う。ハイパーパラメータはグリッドサーチにより、クラス数 100 と決定する。

実際に抽出されたトピックを以下の表 3 に示す。

表 3 抽出されたトピック

カテゴリ	トピック数	例
カープ選手	55	エルドレッド・菊池涼介
野球関連	13	交流戦・NPB・CS
チーム	10	カープ・マツダスタジアム
球団名	9	ヤクルト・巨人・阪神
他球団選手	6	山田哲人・菅野・青木
その他	7	笑・18 時・RCC

5.2 コサイン類似度を用いたコア情報抽出

最後にクラスタリングによって抽出されたトピック別に、コア情報の抽出を行う。具体的には、コアファンのツイートとにわかファンのツイートから類似しているツイート（以下、類似ツイート）を取り除くことによってコア情報の抽出を行う。本研究における類似ツイートとはコサイン類似度が閾値 α 以上のツイートとする。まず各トピックのツイートをコアファンのツイートとにわかファンのツイートに分ける。次にコアファンのツイート同士のコサイン類似度を求める。そして類似ツイートがクラスタ内ツイートの n % 以上あるツイートはありふれた情報であると考え削除する。残ったツイートをコア情報抽出の対象ツイートとする。次に対象ツイートとにわかファンのツイートのコサイン類似度を求める。その結果、類似ツイートがクラスタ内ツイートの m % 以下であるツイートはコアファン特有の情報とみなしコア情報として抽出する。予備実験により本研究では α は 0.1、 n と m は 1 と決定した。

実際に抽出されたコア情報の一部を以下の表 4 に示す。

表 4 抽出されたコア情報

トピック	コア情報
エルドレッド	エルドレッドはとんこつラーメンが好物で、お気に入りの店には同僚の外国人選手を連れていっとるらしい
誠也	鈴木誠也選手は昔「出沒！アド街ック天国」に出演し、実家が荒川にあることから「巨人の星」ならぬ「平成の星」として紹介された
カープ	永川選手はかつて、弟（彼も元カープ所属）に「兄ちゃんパワプロだとフォークしか変化球がないから使えない」と言われたことがきっかけで、スライダーを取得した

6 実験

提案手法の有用性を示すために、(1) チームツイート抽出、(2) ファンコミュニティ分類、(3) コア情報抽出の 3 つの実験を行った。すべての実験において、対象のチームはプロ野球チームの広島東洋カープ（以下、カープと呼ぶ）とした。

3 : <https://taku910.github.io/mecab/>

4 : <https://skymind.ai/japan/wiki/word2vec>

5 : <https://github.com/neologd/mecab-ipadic-neologd>

6 : <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>

6.1 チームツイート抽出の評価実験

ルールベースによる抽出と機械学習による抽出の、2種類のチームツイート抽出手法の有用性を評価することを目的とした実験を行った。

実験条件

被験者はカーブのコアファン3名である。コアファンを被験者にした理由は、ツイートにはマイナー選手や選手の愛称が含まれているものも多く、チームに関する知識量が多いコアファンの方が、回答の正確率が高いと考えたためである。実験に用いたツイート文は、提案するファンコミュニティ分類手法によって獲得したコアファン・にわかファンのツイートから無作為に抽出した4000ツイートである。

実験の流れ

被験者は、提示されたツイート文を読んだ後、アンケートに答える。アンケート内容はその文書は「カーブに関連する内容である」「カーブに関連する内容ではない」の2つの選択肢から一つ選択する。これを4000ツイート行った。3名中2名以上がカーブに関連するツイートであると判別したものを正例、それ以外のツイートを負例とした。結果として、4000ツイートの内1345ツイートが正例、2655ツイートが負例となった。正解データは、正例と負例のデータ数を揃えるためにアンダーサンプリングを行い、負例の中から無作為に抽出した1345ツイートと正例1345ツイートの合計2690ツイートを使用する。この正解データを用いて、5交差検定によりルールベースと機械学習の各々の分類器の適合率、再現率、F値を算出する。

結果と考察

表5にルールベースによるチームツイート抽出の実験結果を、表6に機械学習によるチームツイート抽出の実験結果を示す。ここでの適合率、再現率、F値は5交差検定の平均である。

表5 チームツイート抽出の実験結果 (ルールベース)

予測 実際	Positive	Negative
Positive	1255	90
Negative	149	1196
適合率	0.894	
再現率	0.933	
F 値	0.913	

表6 チームツイート抽出の実験結果 (機械学習)

予測 実際	Positive	Negative
Positive	603	742
Negative	398	947
適合率	0.602	
再現率	0.448	
F 値	0.513	

ルールベースによるチームツイート抽出手法では、適合率・再現率・F値共に90%を超えており、良い結果となった。その要因としては、カーブに関連するツイートには特定の単語が含

まれているものがほとんどであることが考えられる。ルールベースとは対照的に、機械学習によるチームツイート抽出手法の結果は、全て悪い結果となっている。その要因としては、カーブに関連するツイートには特徴が見られないことなど挙げられるが、一番大きな要因としては、学習させる訓練データが少なかったことだと考えられる。本研究では、ルールベースによるチームツイート抽出手法が良い結果となったため、こちらの手法を採用する。

6.2 ファンコミュニティ分類の評価実験

提案手法でのファンコミュニティ分類手法の有用性を評価することを目的とした実験を行った。

実験条件

被験者は実験1と同様にカーブのコアファン3名である。実験に用いたアカウントはファンコミュニティ分類により獲得したユーザから無作為に抽出したコアファン・にわかファンそれぞれ100ユーザである。

実験の流れ

被験者は、提示されたTwitterアカウントのタイムラインを1分間閲覧した後、アンケートに回答した。アンケートは、閲覧したアカウントは「コアファンである」「コアファンではない」「カーブファンではない」「unknown」の4つの選択肢から一つ選択する。これを200アカウント行った。にわかファンであるという選択肢を「コアファンではない」といった選択肢に置き換えた理由は、被験者にとって閲覧したTwitterアカウントをにわかファンであるかどうか判断することが難しいためである。提示されたアカウントが何かしらの影響で閲覧できなかった場合の選択肢として「unknown」を追加した。被験者3名の内、2名以上が選択した選択肢を正解データとする。この正解データを用いて、コアファン・にわかファン各々の適合率を算出した。

結果と考察

表7にコアファンの実験結果を、表8ににわかファンの実験結果を示す。

表7 ファンコミュニティ分類の実験結果 (コアファン)

コアファン			
判定	選択肢	選択数	適合率
正解	コアファン	86	0.863
不正解	コアファンではない	13	
	カーブファンではない	0	
-	unknown	1	

表8 ファンコミュニティ分類の実験結果 (にわかファン)

にわかファン			
判定	選択肢	選択数	適合率
正解	コアファンではない	43	0.453
不正解	コアファン	16	
	カーブファンではない	36	
-	unknown	5	

コアファンの適合率が高い要因としては、我々の提案するコアファンの条件 2, 3 が厳しく設定されていたことが考えられる。コアファンの分類において誤判定されたユーザの多くは、タイムラインの大半がチームに関するリツイートによるもので、チームのファンであることは確認できるが、コアファンであると判断する材料がなかったことが考えられる。にわかファンの適合率が低い要因としては、世界野球などの影響で、チームのファンではないユーザがチームに関するツイートを多くしたこと、誤判定が生じたためと考えられる。

6.3 コア情報抽出の評価実験

提案手法でのコア情報抽出手法の有用性を評価することを目的とした実験を行った。

実験条件

実験対象のチームはプロ野球チームの広島東洋カープとした。被験者はカープのコアファンとにわかファン各々1名ずつである。実験に用いたデータは、クラスタリングされた100トピックの内、カープの選手名である55トピックを対象とする。そして、コア情報抽出手法により獲得したコア情報をトピック毎に25件ずつ無作為に抽出したツイートである。

実験の流れ

被験者は、提示されたツイート文を読んだ後、アンケートに答える。アンケート内容はその文書は「知っている情報である」「知らない情報である」の2つの選択肢から一つ選択する。これを1トピックあたり25ツイートを55トピックの、合計1375ツイート行った。コアファンが「知っている情報である」と回答し、かつにわかファンが「知らない情報である」と回答したツイートを正解とする。この正解データを用いて、コア情報抽出の適合率を算出した。

結果と考察

表9にコア情報抽出の実験結果を示す。

表9 コア情報抽出の実験結果		
	にわかファン	コアファン
知っている情報	89	457
知らない情報	1	828
適合率	0.332	

適合率は0.332とあまり良くない結果となった。適合率が低い要因は、コアファンとにわかファンどちらも知らない情報であると回答したツイートが、データの過半数を超えていることにあると考えられる。各々のファンが知らないと回答した情報の多くは、試合の実況ツイートが多く含まれており、それらのツイートは知っているか否かで回答することが出来ないものであった。しかしながら、コアファンとにわかファンどちらも知っていると回答した情報に比べ、コアファンのみが知っていると回答した情報が多く抽出されていることから、コアファン特有の情報抽出の有用性は示すことが出来たと考えられる。

7 まとめと今後の課題

本研究では、ユーザ同士のフォロー関係に着目しファンコミュニティの分類から、ファンの定着に有効的なコア情報の抽出を行う手法を提案した。ファンコミュニティ間の差分情報の抽出を行うことは出来たが、抽出されたコア情報の多くはにわかファンが興味を示す内容であるとは言い難い情報であった。

今後の課題として、コア情報からにわかファンのファン定着に有効的である情報を抽出する手法を提案する必要がある。また提案手法により抽出されたコア情報をにわかファンへ提示するためのインタフェースの開発についても今後の課題である。

謝 辞

本論文の一部はJSPS 科研費17K00430, 19H04218, 19H04221 及び、私学助成金（大学間連携研究補助金）の助成によるものである。ここに記して謹んで感謝の意を表する。

文 献

- [1] 田沼勇輝, 鈴木政巳, 小林亜樹, “Twitterにおける特定分野に「濃い」アカウントの発見手法”, 第3回データ工学と情報マネジメントに関するフォーラム (DIEM Forum 2011), 5 pages, A10-4.
- [2] 新保直樹, 織田瑞夫, 城沙友梨, 米山照彦, 水野誠, “Twitterネットワーク上のユーザコミュニティ抽出と話題分析”, 2013年度人工知能学会全国大会 (JSAI 2013), 3 pages, 3I4-OS-14b-4, 2013.
- [3] 谷口智哉, 松尾豊, 石塚満, “Blogコミュニティ抽出と分析”, 第6回セマンティックウェブとオントロジー研究会 (SWO 2004), 6 pages, SIG-SWO-A401-08, 2004.
- [4] 西田京介, 坂野遼平, 藤村考, 星出 高秀, “データ圧縮によるTwitterのツイート話題分類”, 第3回データ工学と情報マネジメントに関するフォーラム (DIEM Forum 2011), 7 pages, A1-6.
- [5] 辻一明, 宝珍輝尚, 野宮 浩揮, “新着ツイート群からの興味をひくツイートの抽出に関する考察”, 情報処理学会関西支部平成23年度支部大会, C-7, 2011.
- [6] Zhaochun Ren, Shangsong Liang, Edgar Meij, Maarten de Rijke, “Personalized Time-Aware Tweets Summarization”, Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval, pp.513-522, 2013.
- [7] 佃洗撰, 大島裕明, 山本光穂, 岩崎弘利, 田中克己, “語の認知度と同意語間の関係に基づく意外な情報の発見”, 日本データベース学会論文誌 Vol.11, No3pp.21-26, 2013.
- [8] 伊藤直之, 西川侑吾, 大野和久, 松本征二, 中川修, “語の意外度に基づき話題展開する非タスク指向型対話システム”, 2015年度人工知能学会全国大会 (JSAI 2015), 4 pages, 2L5-OS-07b-2, 2015.
- [9] Ying Zhao and George Karypis, “Comparison of agglomerative and partitional document clustering algorithms”, Technical report, Department of Computer Science, University of Minnesota, Minneapolis, MN 55455, 2002.