

少量の実況ツイートからの状況推定モデルとスポーツ観戦支援

藤本 和† 牛尼 剛聡††

†九州大学大学院芸術工学府 815-8540 福岡県福岡市南区塩原 4-9-1

††九州大学大学院芸術工学研究院 815-8540 福岡県福岡市南区塩原 4-9-1

E-mail: †fujimoto.nodoka.017@s.kyushu-u.ac.jp, ††ushiana@design.kyushu-u.ac.jp

あらまし スポーツの試合は、録画よりもリアルタイムに視聴することに価値を感じる人が多いが、視聴者が忙しい時や試合が停滞してつまらない場合、視聴を続けることは難しい。スポーツ中継において、視聴者にとって価値が高いと思われる場面を検出し、リアルタイムに通知ができれば、視聴者の効果的なリアルタイム視聴を支援できる。本論文では、プロ野球中継を例題として、最新の少量の実況ツイートから、状況をリアルタイムにモデル化する手法を提案する。本手法の状況のモデル化では、実況ツイートの特徴を反映した行列を生成するが、テキストの表記揺れやツイート固有の独特な表現などに対応可能な、文字レベルのオートエンコーダを利用する。提案手法の有用性を評価するために、試合の盛り上がりの大きさを反映すると考えられるツイート投稿数などの特徴量を推定するタスクを利用し、実験の結果に基づいて提案手法の有効性を評価する。

キーワード SNS, 実況ツイート, オートエンコーダ, 深層学習

1 はじめに

現在、世の中には多くの動画コンテンツが存在し、情報収集、学習、娯楽等様々な目的のために重要な役割を果たしている。動画コンテンツは、オンデマンド型とリアルタイム型の2種類に大別できる。オンデマンド型のコンテンツは、ユーザは観たい時に視聴できるという利点がある。一方、リアルタイム型のコンテンツは、実世界のイベントを撮影した動画をそのままユーザに配信するものである。リアルタイム型のコンテンツは、実世界のイベントをリアルタイムに配信しているため、遠隔地でも同時にイベントを体験できる。

リアルタイム型の動画コンテンツとして配信されるイベントには様々なものが存在するが、代表的なものにスポーツの試合がある。スポーツの試合をリアルタイム配信するスポーツ中継では、視聴者は試合中に発生する様々な出来事をリアルタイムに体験したいという要求が多い。試合の結果を確認するだけであれば、ニュースやインターネットなどで結果やダイジェストを観ればよいにもかかわらず、TVやインターネットではスポーツ中継が放送されており、多くの人がリアルタイムに楽しんでいる。これは、スポーツの試合には、録画よりもリアルタイムに視聴することに視聴者は価値を感じる事が理由だと考えられる。

しかし、スポーツ中継をリアルタイムに楽しむには、いくつかの問題がある。その一つに、視聴者が、観たい場面を見逃さないようにするためには、中継をリアルタイムに視聴し続ける必要があることである。視聴者が忙しい時などは常に中継を視聴し続けることが困難であり、また、忙しくなくても、試合が停滞してつまらない状況で視聴し続けることは、視聴者にとって負担になる。

このような背景の下、本件研究では、視聴者のリアルタイム

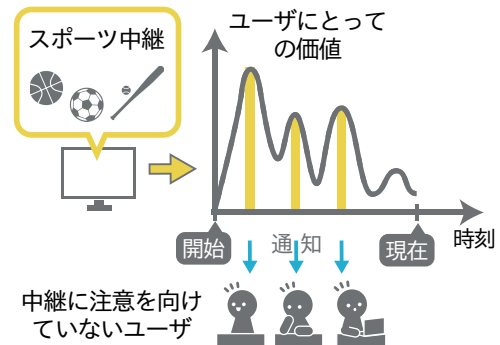


図1 提案システムの概要

動画の効果的な視聴を支援するシステムの開発を行う。リアルタイム動画の視聴を支援するためには、様々なアプローチが考えられる。我々は、スポーツ中継を対象に、視聴者にとって価値が高い状況を自動的に検出し、リアルタイムに視聴者に通知するシステムの開発を目標とする。このシステムの概要を図1に示す。本システムを利用することによって、ユーザはスポーツ中継を視聴し続けることなく、システムからの通知によって、価値の高い場面のみをリアルタイムに体験可能となる。

ユーザにとって価値ある場面をリアルタイムに通知するためには、システムが対象となるスポーツの試合の状況をリアルタイムに理解し、ユーザにとって価値を判定する必要がある。本論文では、代表的なスポーツ中継の一つであるプロ野球を対象に、試合の状況を適切にモデル化するための手法を提案する。スポーツの試合の状況を理解するためには、様々な方法が考えられる。本研究では、SNSから取得可能なデータを利用して、試合の状況を推定する。

近年、SNSの普及に伴って、SNSにコメントを投稿しながらリアルタイムに動画を視聴する、ソーシャルビューイングという視聴形態が注目されている。ソーシャルビューイングで

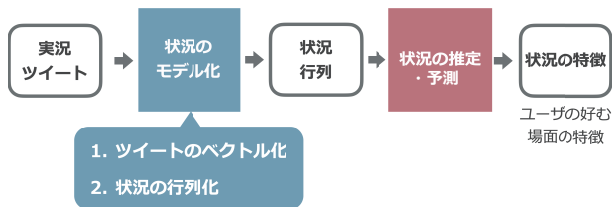


図 2 システムの概要

は、動画コンテンツに対する感想や応援などの投稿し、共有することで動画内のイベントを楽しむことができる [1, 2]。ソーシャルビューイングで利用される SNS の中で代表的なものは Twitter¹である。Twitter におけるソーシャルビューイングのための投稿（ツイート）を「実況ツイート」と呼ぶ。実況ツイートには、対象とする動画コンテンツを区別するためにハッシュタグが付与されている場合が多く、動画コンテンツに関するハッシュタグで検索することによって、対象とするリアルタイム型コンテンツの実況ツイートを簡単に閲覧できる。

実況ツイートには、スポーツの試合に対する感想や応援がテキストデータとして含まれているため、試合の状況やコメントを投稿する視聴者にとっての価値が反映されている。これらの投稿を分析することによって、スポーツの試合の状況の視聴者にとっての特徴を抽出することが期待できる [3, 4]。そこで、本論文では、スポーツの試合の状況を推定するための状況のモデル化に、実況ツイートを用いる。本研究では、対象とするスポーツとしてプロ野球のデータを用いるが、野球固有のドメイン知識を用いずに、一般的な形のモデル化を行い、他のスポーツや他のジャンルの動画コンテンツにも適用可能な汎用的な手法の開発を目指す。提案する状況のモデル化の概要を図 2 に示す。

スポーツの試合によっては、試合中に大量の実況ツイートが投稿されるため、全てのツイートを取得して特徴の抽出を行うのは、処理コストが大きい。そのため、一定時間に投稿されたツイートの一部をサンプリングしたものから、その瞬間の状況を推定できることが期待される。一般に、人は、実況ツイートのタイムラインの全てのツイートを見なくても一部分のツイートのみからその場面を想像できることが多い。そこで、本手法では、対象とする時間区間に投稿された一部の実況ツイートを用いて、状況のモデル化を行う。そして、モデル化された状況を用いて、リアルタイムに現在の状況の特徴を推定する、または未来の状況の特徴を予測することで、モデルの有用性を評価する。

本論文の貢献は、実況ツイートを利用したリアルタイムの動画コンテンツの状況を表現するモデルを開発したことにある。このモデルの特徴は以下の通りである。

- 少量のツイートを用いた状況の推定と、単純な処理によって、ツイートの収集と処理のコストを抑えている。
- 人名や専門用語などのドメイン固有の知識を使わないことで、様々なドメインに対応可能な汎用性のあるモデルである。

2 関連研究

2.1 テキストの分散表現生成

本論文で提案する手法では、ツイートに含まれるテキストから生成された埋め込みベクトルを利用する。テキストのベクトル化に関しては、これまで様々な手法が提案されており、大きく分けて、単語レベルの手法と文字レベルの手法に大別できる。

単語レベルの処理でテキストのベクトルを生成する手法は、単語のベクトルを用いて文章全体のベクトルを生成する手法と、直接文章のベクトルを生成する手法がある。単語の意味的な情報をベクトル化する単語埋め込みの手法として、代表的な手法に Word2Vec [5] や fastText [6] などがある。文章の分散表現を生成する手法においては、単語の分散表現を用いる手法 [7-9] や、深層学習を用いて直接文章をベクトル化する手法 [10, 11] が存在する。これらのような単語レベルの処理を行う手法には、日本語のテキストであれば、形態素解析を行って単語分割を行う必要がある。

文字レベルで埋め込みベクトルを生成する手法では、単語分割を行う必要がなく、表記揺れや独特な表現のある文章にも対応できる。文字レベルの one-hot ベクトルの系列を生成し、再帰的ニューラルネットワークや畳み込みニューラルネットワーク等の深層学習によって、テキストの分類などを行う手法が提案されている [12]。また、CHARAGRAM [13] は、文字レベルの n-gram を次元削減して、テキストの埋め込みベクトルを生成している。これらの研究は、文字の種類数が少ない英語のテキストや、日本語のテキストの読みをアルファベットに変換したものを対象にしている。本論文では、それらよりも文字種類の多い日本語のテキストをそのまま扱っており、日本語のテキストにも適用可能な手法を提案する。

2.2 ソーシャルビューイングデータの活用

大量の実況ツイートから特徴を抽出し、イベントの検出や要約をすることで、ユーザがコンテンツを効率的に体験することを支援する手法は、これまでもいくつか提案されている。代表的な手法として、スポーツのイベント検出を目的として、ツイートの割合の急激な増加を見つけ、テキストの内容を教師あり学習で分類することでイベント内容を特定する手法が提案されている [3, 14]。教師なし学習を用いた手法の例としては、トピックモデルに基づいてイベントを特定する手法が提案されている [15, 16]。これらは、イベントが終わった後に検出することを前提としており、本研究で対象とするようなリアルタイムのイベント検出を対象としていない。

ツイートのデータをリアルタイムに処理をする試みとしては、リアルタイムにイベントの要約を生成する研究 [17]、投稿数の変化をリアルタイムに検知してイベント検出する研究 [18] 等が行われている。これらの研究では、リアルタイムに全てのツイートが収集できることを前提にしている。

本研究では、サンプリングされた少量のツイートを用いて、対象とする状況の特徴を抽出する。

1: <https://twitter.com>

3 実況ツイートによる状況のモデル化

3.1 ツイートのベクトル化

実況ツイートの内容的特徴の抽出方法については、これまでに提案された手法の多くが、ツイートに含まれる単語の意味に基づいて特徴を解析するアプローチを採用している [19,20]. しかし、スポーツ中継の実況ツイートは、表記揺れやノイズ、Twitter 固有の特殊な表現が多く、一般的な形態素解析を使った単語の意味による解析手法は適切に適用できない場合がある. そこで、本研究では単語レベルの処理ではなく、文字レベルの処理でツイートの埋め込みベクトルを生成する.

Vosoughi ら [12] は、英語で書かれたツイートを対象に文字レベルのベクトル化を行う Tweet2Vec 手法を提案している. 本研究では、Tweet2Vec を日本語に適用した手法によりツイートの埋め込みベクトルを生成するアプローチを採用する.

3.1.1 文字レベルの処理による特徴行列の生成

まず、実況ツイートを構成する個々のツイートの文字レベルの特徴を表現することを考える. Vosoughi らが提案した Tweet2Vec 手法では、ASCII 文字から構成されるツイートの文字レベルの特徴を表すために、ツイートを構成する文字を one-hot ベクトルとして表現し、文字の one-hot ベクトルを出現順に並べて行列としたものをツイートの特徴行列としている.

本研究では、日本語のツイートを対象とするが、日本語のテキストは、ひらがな、カタカナ、漢字等で表されるために、英語の文章よりも、文字の種類が非常に多い. そのため、全ての文字を対象とした one-hot ベクトルを作るのは使用するメモリ及び処理時間の観点から現実的でない. そこで、本手法では、コーパスを対象に文字の出現頻度を集計し、頻出する文字のみを対象とする. 具体的には、頻出上位 d 件の文字に加え、 d 件の中に含まれない未知文字、文字数を揃えるためのパディング用の記号を使い、1文字を $(d+2)$ 次元の one-hot ベクトルで表す. one-hot ベクトルの生成に利用される文字種類数を d としたとき、ツイートテキストの i 番目の文字を表す one-hot ベクトルは、 $\mathbf{x}_i \in \{0, 1\}^{(d+2)}$ となる.

実況ツイートから、試合の盛り上がりのような特徴を予測する場合、論理的な考察をしている長いテキストの投稿よりも、投稿者の感情的な反応が見られる短いテキストの投稿の方が試合の状況を予測する際に有効である可能性が高い. そこで、リアルタイムに処理可能となるように処理速度を削減するため、投稿テキストの先頭の c 文字のみを使用する. なお、Twitter で1度に投稿できる最大文字数は140であるため、 $c < 140$ である. このとき、 c 文字からなるツイートを行列 $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_c)$ として表現する.

3.1.2 オートエンコーダによるツイートのベクトル化

3.1.1 で示したツイートの行列表現 \mathbf{X} から、オートエンコーダ [21] を用いてツイート埋め込みベクトルを生成する. オートエンコーダは、ニューラルネットワークを用いて特徴ベクトルを生成するための代表的な手法の一つである. オートエンコーダでは、入力データ自体を正解ラベルとしてエンコーダとデ

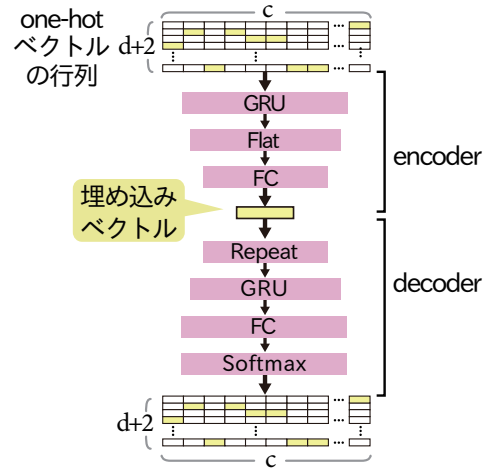


図3 AE-GRUの構造

コーダを学習させる. 特徴ベクトルを生成する場合には、学習済みのオートエンコーダのエンコーダ部に対して、目的とするデータを入力として前向きな推論を行い、出力結果として得られたベクトル (エンコードされたベクトル) を埋め込みベクトルとする.

ニューラルネットワークを利用してエンコーダとデコーダを実現する場合、様々なモデルを考えることができる. 本研究では、GRU [22] を利用し、このモデルを AE-GRU と呼ぶ. AE-GRU の構成図を図3に示す. GRU は、再帰的ニューラルネットワーク (RNN) の一種であり、同じく RNN の一種である LSTM よりも計算量が少なく高速である [23]. 図中において、**Flat** は1次元のテンソルへの変換、**FC** は全結合層、**Repeat** は同じベクトルを重ねた2次元のテンソルへの変換である. 学習済みのオートエンコーダのエンコーダ部を利用して、1ツイートに対する埋め込みベクトルを生成する. 1ツイートは行列 \mathbf{X} として表され、エンコーダによって n 次元の埋め込みベクトル \mathbf{v} に変換される.

3.2 状況のモデル化

3.2.1 対象とするデータの形式化

実況ツイートのタイムライン tl は、ツイート tw_i の系列 $tl = (tw_1, tw_2, \dots, tw_n)$ として表現する. いま、ツイート tw の投稿時刻を $\text{time}(tw)$ と表記すると、タイムライン tl に含まれる任意の2つのツイート tw_i, tw_j は、 $i < j$ のとき、 $\text{time}(tw_i) < \text{time}(tw_j)$ を満足する.

実況ツイートが対象としているリアルタイム型動画コンテンツの状況を表すために、状況セグメントの概念を導入する. 状況セグメントとは、対象とするリアルタイムコンテンツ内の時間区間であり、 $ss(tl, st, et)$ として定義する. ここで、 tl は対象とするタイムライン、 st はセグメントの開始時刻、 et は終了時刻を表す.

本研究では、対象とするリアルタイム型動画コンテンツを、時間窓を利用して同じ粒度の単位状況セグメントに分割し、それを単位として特徴をモデル化する. 単位状況セグメントを生成するためには、リアルタイム型動画コンテンツに対して、窓

幅 m の幅の時間窓を適用し、 $m/2$ の幅で移動させ、窓が半分重なるようにすることで、セグメントの境界部分も適切にモデル化できるようにする。対象とするリアルタイム型動画コンテンツに対して、単位状況セグメントを定義すると、コンテンツの状況は、単位状況セグメントの系列として表現できる。以降、特に明示しない場合には、状況セグメントは単位状況セグメントを表す。それぞれの状況セグメント ss に対して、それに対応する実況ツイート系列 $tw(ss)$ 考えることができる。

3.2.2 モデル化のアプローチ

本論文では、対象とするリアルタイムコンテンツに含まれる任意の状況セグメント ss に対して、そのセグメントが対象とする実況ツイートタイムラインの部分系列 $tw(ss)$ を利用して、特徴ベクトルを生成する。

本システムではリアルタイム性が必要とされるため、すべての実況ツイートを処理することは処理コストの面から現実的でない。そこで、セグメントに含まれるツイート $tw(ss)$ のうち投稿時刻が遅い k 個のツイートをサンプリングし、サンプリングされたツイートから、そのセグメントの状況を表す特徴行列を生成するアプローチを取る。

3.2.3 状況行列の生成

$tw(ss)$ に含まれる k 個のツイートを利用し、まず、3.1 で示した手法で k 個の埋め込みベクトルを生成する。次に、それらのベクトルを結合し、セグメントの特徴を表す状況行列 M を生成する ($M = (v_1, v_2 \dots v_k)$)。ここで、 M は $n \times k$ の行列となる。

4 状況の特徴の推定

セグメントの特徴行列 M を利用して、スポーツ中継における試合の状況の特徴量を推定する事を考える。ここにおける状況の特徴量は、ユーザが視聴しているリアルタイムコンテンツの状況が、ユーザにとってどの程度の価値があるかを推定するための基準となるものである。代表的な特徴として、「試合の盛り上がり」や「得点シーン」が考えられる。本論文で提案する状況行列は、状況の特徴を表す一般的なモデルであり、様々な特徴量の推定に利用可能であることが期待できる。

いま、状況セグメント ss の状況行列を $M(ss)$ と表記し、推定する特徴量を y とし、 $M(ss)$ から y を推定する関数を $F()$ とするとき、この推論は $y = F(M(ss))$ と表記できる。本手法では、関数 $F()$ をニューラルネットワークによって実現する。このニューラルネットワークのモデルは、全結合層や Dropout 層 [24] を用いた深層学習のモデルを扱う。このモデルによって回帰問題および分類問題を解く。本研究では、状況セグメントに対応付けられる実況ツイートの投稿数の推定と、得点シーンの判定、シーンの極性の分類をすることを考える。

4.1 ツイートの投稿数

ツイートの投稿数は、スポーツ中継の盛り上がりを表す特徴量である。実況ツイートの投稿数は教師データの収集が容易であり、提案する状況行列の性能を評価することに適切であると

考えられる。正解ラベルとする実況ツイート投稿数は、対象とする試合によって異なるため、絶対値を予測するのではなく、正規化することが必要である。そこで、学習データでは、対象とする試合における状況セグメントの最大ツイート投稿数で割ることによって、正規化を行う。

3.2.3 で示したことから、ひとつの状況セグメント内のツイート投稿数を推定するには、セグメント内で投稿時刻が遅い k 個のツイートを扱う。この入力と出力のペアだけでは、ツイート投稿数の推定モデルを学習する際には、データ数がセグメントの数だけに制限されてしまうため少ない。そこで、ある試合の実況ツイートの系列を k 個ずつに分割し、学習の際の入力とする。出力となる正解は、 k 個のツイートの中で一番投稿時刻が遅いツイートが含まれるセグメントのツイート投稿数とする。具体的には、 k 個のツイートの k 番目のツイートが、 $ss(tl, st + \frac{m}{2}, et)$ に含まれていれば、セグメント $ss(tl, st, et)$ のツイート投稿数を正解と定める。

4.2 得点シーンの判定

野球やサッカーなどの得点を競うスポーツにおいては、得点シーンは重要なシーンとなる。得点シーンのような特定のシーンは、起こった時刻さえわかれば正解ラベルを作ることができ、得点シーン以外の推定にも応用できると考えられる。サンプリングしたツイートから、言及されている場面が得点シーンまたはそれ以外のシーンかを判定する。

得点シーンの時刻がセグメント $ss(tl, st + \frac{m}{2}, et)$ に含まれているとき、同じセグメント内の投稿時刻が遅い k 個のツイートから生成された状況行列を入力として、得点シーンとしての正解ラベルを与える。学習データでは、得点シーンを正解ラベルとするデータセットを増やすために、 $ss(tl, st, et)$ に含まれる得点シーンの時刻以降に投稿されたツイートから、 k 個のツイートのセットを複数取り出し、それぞれを入力の特徴行列とする。

4.3 シーンの極性の分類

様々なイベントにおいて、視聴者の感情はあるシーンを表す重要な要素である。また、4.1 で述べた盛り上がりは、ポジティブな意味での盛り上がりだけではなく、ネガティブな意味での盛り上がりなど様々あると考えられる。そこで、サンプリングされたツイートから、視聴者の感情を判定することを考える。ここでは単純に、極性（ポジティブとネガティブ）の2値分類を行う。

この分類モデルを生成するための学習データとして、ポジティブとネガティブの正解ラベルを与える必要がある。しかし、多くのシーンそれぞれに対して、主観的な評価をつけることはコストがかかる。そこで、明らかにポジティブだと考えられる得点シーンと、明らかにネガティブだと考えられる被得点シーンの場面のみを、学習データに含める。つまり、得点/被得点シーンの時刻がセグメント $ss(tl, st + \frac{m}{2}, et)$ に含まれているとき、同じセグメント内の投稿時刻が遅い k 個のツイートから生成された状況行列を入力として、得点/被得点シーンとしての正解ラベルを与える。そして、4.2 と同様に、正解ラベルと

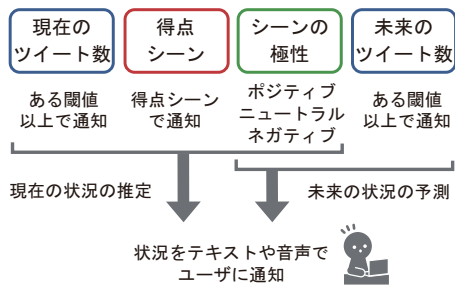


図4 通知システムの構成図

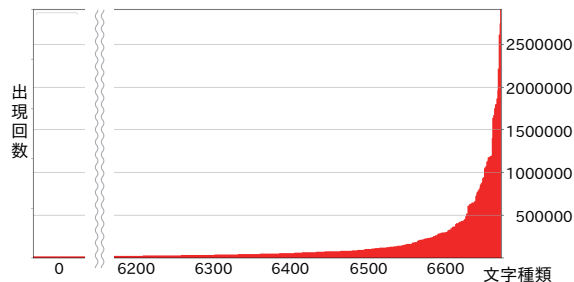


図5 実況ツイート中の文字種の出現頻度

するデータセットを増やすために、 $ss(tl, st, et)$ に含まれる得点シーンの時刻以降に投稿されたツイートから、 k 個のツイートのセットを複数取り出し、それぞれを入力状況行列とする。

4.4 未来のツイート投稿数の予測

4.1では、現在のツイート投稿数の推定を行う手法を提案した。ここでは、未来のツイート投稿数の予測を行う。

学習データの生成やデータの正規化方法は4.1と同じで、正解データが未来のデータとなる。 em 分後のツイート投稿数を予測する場合、状況セグメント $ss(tl, st, et)$ 内のツイートをを用いて、 $ss(tl, st + em, et + em)$ 内のツイート数を予測する。

5 盛り上がり予測システム

図4は、状況の特徴の推定値によって、ユーザーにスポーツ中継の状況を通知するまでの流れを表した、システムの構成図である。それぞれの推定値および予測値に対して閾値を設定することで、ユーザーに通知するかどうか判定する。

6 評価実験

プロ野球中継を対象に提案手法の有効性を評価する4つの実験を実施した。全ての実験に置いて、セグメントに含まれるツイートのサンプリング数 k を $k = 10$ 、セグメントの幅 m を $m = 2$ (分間)に設定した。

6.1 データセットと前処理

TwitterAPI²を用いて、プロ野球の試合時間中の実況ツイートのタイムラインを取得した。対象とした試合は、2019/4/14から2019/10/10に開催された日本野球機構(NPB)に所属する全12球団の試合である。

取得した実況ツイートを、学習データの試合(全718試合/全4,000,000ツイート)とテストデータの試合(全18試合/108,000ツイート)に分割し、これを利用して実験を行った。そして、全てのツイートのテキストに対して、URLやハッシュタグ等の除去、全半角の統一などの前処理を行った。

次に、オートエンコーダの入力および出力となる、行列の生成について説明する。学習データにて実況ツイート中の文字の出現頻度を集計したところ、6,669種類の文字種が存在した。図5に文字の出現頻度のヒストグラムを示す。これらの中

表1 オートエンコーダの精度(交差エントロピー誤差: CEL)

d	100	250	500	1000	1500
CEL($\times 10^{-2}$)	0.5445	0.7936	2.728	5.509	6.143

正解	高山もホームラン(((o(*^▽^*)o)))
復元例	高山もホームラン(((o(*^▽^*)o)))
正解	岩崎さんも完璧👏👏一刻も早く同点から脱したい
復元例	岩崎さんも完璧👏👏休刻も早く同点から脱したい
正解	7回裏虎4-4 燕勝ち越すぞ!
復元例	7回裏虎4-4 燕勝ち越すぞ!
正解	明日の三壘スタメンは■だな・・・。
復元例	明日の三壘スタメンは■だな・・・。

図6 オートエンコーダによるテキストの復元例

から、上位 d 文字種を使用し、パディング(PAD)と未知文字(UNK)を加え、先頭から50文字($c = 50$)を、次元数が $(50, d + 2)$ の行列 \mathbf{X} として表現した。

6.2 オートエンコーダの性能評価

6.2.1 実験手法

3で提案したオートエンコーダを学習させ、エンコーダ部を用いて未知のテキストのベクトル化し、それに対してデコーダ部デコーダ部を利用して復元した際に、元のテキストに対する復元の正確さについて性能を評価した。入力となるone-hotベクトルの行列 \mathbf{X} には、それぞれの出現頻度上位 $d = 100, 250, 500, 1000, 1,500$ の文字種を使用して比較実験を行った。

学習の際には、学習データのツイートを20,000個ずつ学習させ、バッチサイズは100、最適化手法はAdam[25]、損失関数は交差エントロピー誤差(Cross-entropy Loss)、20%を検証データとして、学習を繰り返した。各学習は、エポック数は最大50として、検証データの誤差が小さくなったらモデルを保存した。検証データの誤差が減少しない学習が25エポック続いた場合は、その学習を終了して次のデータの学習を行った。最終的なモデルは、テストデータを用いて交差エントロピー誤差で評価した。

6.2.2 実験結果と考察

オートエンコーダの性能を評価において、結果を表1に示す。結果から、設定したどの d の値が小さいほど、テキストの復元精度が高い結果となった。 $d = 1,500$ での、テキストの復元例を図6に示す。“■”は未知文字、パディング部分は省略してある。

オートエンコーダの性能評価においては、どのモデルでも文

2: <https://developer.twitter.com>

表 2 ツイート投稿数の推定精度 (MSE)

AE-GRU					Word2Vec
$d = 100$	$d = 250$	$d = 500$	$d = 1000$	$d = 1500$	
0.02899	0.02429	0.02765	0.02782	0.02922	0.06155

字種類数 d の値が小さいほど復元精度が高い結果となった。これは、文字が復元される際に、候補となる文字が多い方が正確な推定が困難になるためだと考えられる。しかし、 $d = 1,500$ の場合であっても、テキスト復元例から適切なベクトル化ができてきていることが確認できる。

6.3 ツイート投稿数の推定

6.3.1 実験手法

対象とするチームの例として、「阪神タイガース」のタグを付けられて投稿された実況ツイートを対象に、ツイート投稿数を推定した。

比較手法として、単語レベルの埋め込みベクトルを用いた手法を利用した。比較手法では、まず、前処理をしたテキストに対して形態素解析を行い、自立語（動詞、形容詞、名詞など）のみを抽出した。これらの単語をコーパスとし、学習データのツイートを gensim³ の Word2Vec のモデルを学習させた。単語ごとに出力されるベクトルは、オートエンコーダによる埋め込みと同じ 300 次元とし、テキストに含まれる単語のベクトルの平均をテキストのベクトルとした。そして、10 個 ($k = 10$) のツイートから状況行列を生成した。

テストデータとした 18 試合において、実際のツイート投稿数と推定値との平均二乗誤差 (MSE) で、テキストをベクトル化した AE-GRU と Word2Vec のモデルを評価した。

6.3.2 実験結果と考察

ツイート投稿数の推定精度は、テキストをベクトル化した AE-GRU と Word2Vec のモデルにおいて、表 2 の通りとなった。AE-GRU を用いた場合、ツイート数の推定モデルにおいて、 $d = 250$ のときの状況行列が一番精度が良い結果となった。

テストデータの試合で、AE-GRU ($d = 250$) を用いて、ツイート投稿数を推定したグラフの 2 試合の例を次の図 7 に示す。推定値には途中で線画途切れているところがあるが、これはテキストのあるツイートが 10 件に満たなかったときである。

ツイート投稿数を推定する実験において、AE-GRU で $d = 250$ のときが最も高い結果となった。これは、文字種類数が多い方がテキストを表す情報量が多いため、状況を推定する上で効果が大きいと考えられる。また、Word2Vec を使ったモデルよりも、提案した AE-GRU のモデルの方が高い精度を示した。Word2Vec を使ったモデルの場合は、学習の際の誤差が収束しなかった。このことから、スポーツ中継の実況ツイートの状況を表すモデルにおいて、ツイートテキスト中の単語の意味よりも、テキストの文字列のパターンが重要だと言える。テストデータの試合において、推定値のグラフでは、ツイート投稿数が少ない場面では誤差があるものの、盛り上がりを反映した形状になっていることが観測できる。

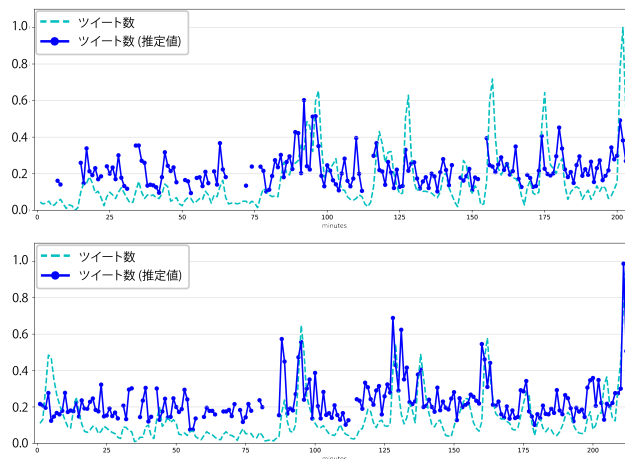


図 7 ツイート投稿数とその推定値

表 3 得点シーンの評価

AE-GRU					random
$d = 100$	$d = 250$	$d = 500$	$d = 1000$	$d = 1500$	
0.46	0.41	0.40	0.40	0.55	0.065

6.4 得点シーン判定

6.4.1 実験手法

ある特定のシーンを判定することは、ユーザにとって有益である。そこで、「阪神タイガース」のタグを付けられて投稿された実況ツイートを対象に、得点シーンとそれ以外のシーンの分類を行った。

4.2 に示した手法によって、入力となるツイート群と正解ラベルのペアを定めた。実験に用いるテストデータには、テストデータの試合における得点シーン 39 個、それ以外のシーンが 992 個含まれている。このテストデータにおいて、適合率と再現率から計算される F 値で評価を行った。評価するモデルは、テキストをベクトル化した AE-GRU と、1/2 の確率でランダムに推定を行うモデルを仮定しベースラインとして比較を行った。

6.4.2 実験結果と考察

d の値を変化させたモデル AE-GRU とランダムに推定するモデル random で比較した結果は、表 3 の通りである。結果から、AE-GRU の $d = 1,500$ のモデルが一番良い結果となった。

AE-GRU ($d = 1,500$) を用いた状況行列を使った分類モデルは、誤って得点シーンだと判定した場面であっても、得点をとった後の反応がまだ続いている場面が多かった。また、誤って得点シーン以外だと判定した場面の 1/3 は、相手のミスによる得点シーンであった。これは、応援しているチームの選手が直接成果をあげた訳ではないため、盛り上がりの種類が違ったり、困惑の反応がツイートに見られるからだと考えられる。学習データのバリエーションを増やすことによって、この分類モデルの精度が改善できる見込みがある。

6.5 シーンの極性の判定

6.5.1 実験手法

シーンの極性は、どのような盛り上がりかを判定する上で重

3 : <https://radimrehurek.com/gensim/>

表 4 極性判定の評価 (正解率)

AE-GRU					random
$d = 100$	$d = 250$	$d = 500$	$d = 1000$	$d = 1500$	
0.76	0.78	0.79	0.76	0.74	0.50

要な特徴である。そこで、「阪神タイガース」のタグを付けられて投稿された実況ツイートを対象に、シーンの極性（ポジティブ/ネガティブ）の分類を行った。

本実験では、6.3において、ツイート数を推定する一番良い精度のモデルを使った際に、推定値が 0.45 以上の 112 シーンを用いた。正解データは、主観評価によって定めた。まず、112 シーンに対して、出来事の内容を表すラベル（「攻撃時、バント成功、チャンスをつくる」、「守備時、リリーフピッチャーに交代」等）を付与した。そのラベルの種類は 55 個あり、これらに対して、被験者 3 人がポジティブ/ニュートラル/ネガティブの 3 段階の主観評価をするアンケートを行い、3 人の評価から多数決で出来事の極性を決定した。評価指標は、ネガティブシーンをネガティブ、ポジティブシーンをポジティブと正しく推定できた正解率とした。

6.5.2 実験結果と考察

極性を推定した結果を表 4 に示す。正解率としては、 $d = 500$ の時が一番高かった。しかし、得点シーンや被得点シーン以外のシーンも含まれているため、正しい推定が困難なシーンがあった。得点シーンと被得点シーン以外のシーンのデータも、学習データに入れていくことで改善されることが期待される。また、テストデータにおいて、試合中の一部のシーンに絞って評価したため、今後はより多くのデータで評価をする必要がある。

6.6 未来のツイート投稿数数の予測

6.6.1 実験手法

リアルタイムに盛り上がるシーンを視聴するためには、未来の盛り上がりを予測する必要がある。そこで、「阪神タイガース」のタグを付けられて投稿された実況ツイートを対象に、未来のツイート投稿数を予測した。何分後の未来を予測するかどうかの変数 em (分) には、 $em = 1, 2, 3$ で比較を行った。つまり、現在の状況から 1, 2, 3 分後のツイート数を予測するということである。

ツイート投稿数を予測するモデルに、6.3 で用いた AE-GRU ($d = 250$) と同様の状況推定モデルを用い、評価は平均二乗誤差 (MSE) で行った。

6.6.2 実験結果と考察

未来のツイート投稿数を予測したときの予測値と正解データとの MSE を表 5 に示す。 $em = 1, 2, 3$ において、どの予測も、6.3 で示した現在の状況を推定するよりも精度は下がっており、未来に行くほど予測が難しいという人間の感覚に合っている。この実験では、一つの状況セグメントからしか予測を行っていないが、複数の過去のセグメントからの予測や、別の要素を入力として加えることで、精度が改善される見込みがある。

表 5 ツイート投稿数の予測精度 (MSE)

em	1	2	3
MSE	0.02521	0.02653	0.02365

6.7 盛り上がり予測システムの評価

6.7.1 実験手法

5 で述べた提案システムについて、推定された現在の状況および、予測された未来の状況が正しいかどうかを評価した。

評価の対象は 2019/09/29 の「阪神タイガース」対「中日ドラゴンズ」の試合である。本実験では、「阪神タイガース」に対するのツイートをを用いて行い、「阪神タイガース」のファンにとって価値ある場面を通知するためのシステムとする。

6.7.2 実験結果と考察

図 8 は、横軸は試合の開始時間から終了時間までの、それぞれ推定または予測された特徴を表しており、吹き出しは実際に発生した主なイベントを示している。ツイート投稿数の予測に関しては、 $em = 1, 2, 3$ (分後) 予測するモデルのいずれかが盛り上がっていると予測した時間に色をつけた。

全体的に、ツイートの投稿数が多い場面においては、推定と予測共に精度が高い予測が可能であることが示された。この実験では、1 つの試合での評価しか行っていないため、今後は他の試合でも同様の結果が出るかを検証する必要がある。

7 まとめ

本研究では、スポーツ中継のイベントに対する実況ツイートから、文字レベルのオートエンコーダを用いて状況を表す状況行列を生成する手法を提案した。また、その行列からツイート数を推定することによって、状況行列の有用性を示した。

提案した手法は、ツイートのテキストデータのみを扱っており、選手名やイベント名などの事前知識が必要ない。また、ツイートをベクトル化する際の文字列の長さ c 、状況行列を生成するツイートの個数 k や時間窓 w の値は可変であり、ドメインの特徴によって変更することができる。そのため、野球中継以外の他のドメインであっても、適切な値に設定することで効果的な状況のモデル化が可能となる可能性がある。野球以外の対象についての提案手法の有効性の評価は今後の課題である。また、今後は、リアルタイムに通知するシステムを被験者実験によって評価することや、ユーザ個人に合わせた通知システムの検討をする必要がある。

謝 辞

本研究は JSPS 科研費 19H04219 の助成を受けたものです。

文 献

- [1] X. Wang. Applying the integrative model of behavioral prediction and attitude functions in the context of social media use while viewing mediated sports. *Computers in Human Behavior*, Vol. 29, pp. 1538–1545, 2013.
- [2] Q. Ji and A. A. Raney. Morally judging entertainment: A case study of live tweeting during downton abbey. *Media Psychology*, Vol. 18, pp. 1–22, 2015.

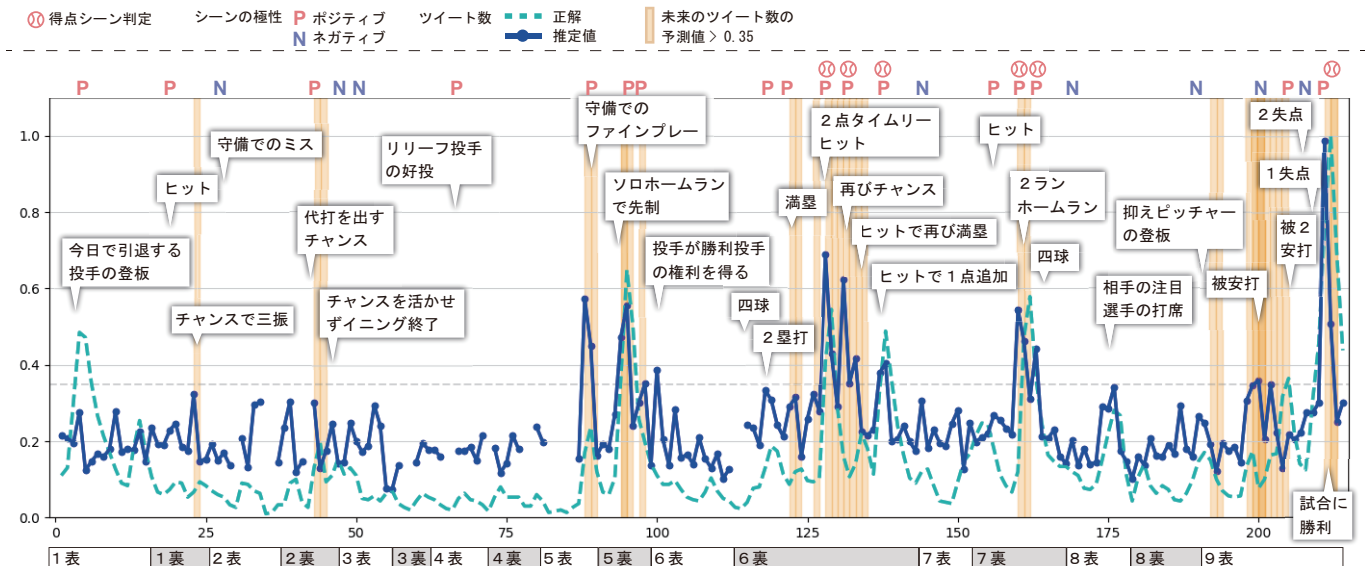


図 8 推定・予測された状況の特徴

- [3] S. Zhao, L. Zhong, J. Wickramasuriya, and V. Vasudevan. Human as real-time sensors of social and physical events: A case study of twitter and sports games. *CoRR*, Vol. abs/1106.4300, , 2011.
- [4] E. Kalampokis, E. Tambouris, and K. Tarabanis. Understanding the predictive power of social media. *Internet Research*, Vol. 23, , 2013.
- [5] T. Mikolov, I. Sutskever, K. Chen, G. Corrado, S., and J. Dean. Distributed representations of words and phrases and their compositionality. In *Proc. NIPS*, pp. 3111–3119, 2013.
- [6] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov. Enriching word vectors with subword information. *Transactions of the Association for Computational Linguistics*, Vol. 5, pp. 135–146, 2017.
- [7] R. Tian, N. Okazaki, and K. Inui. The mechanism of additive composition. *Machine Learning*, Vol. 106, pp. 1083–1130, 2017.
- [8] M. Dheeraj, G. Vivek, P. Bhargavi, and K. Harish. Scdv : Sparse composite document vectors using soft clustering over distributional representations. In *Proc. of the 2017 Conf. on Empirical Methods in Natural Language Processing*, pp. 659–669, September 2017.
- [9] S. Dabiri and K. Heaslip. Developing a twitter-based traffic event detection model using deep learning architectures. *Expert Systems with Applications*, Vol. 118, pp. 425–439, 2019.
- [10] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *Proc. of the 31st International Conf. on Machine Learning*, pp. 1188–1196, 2014.
- [11] R. Kiros, Y. Zhu, R.R. Salakhutdinov, R. Zemel, R. Urtasun, A. Torralba, and S. Fidler. Skip-thought vectors. In *NIPS 2015*, 2015.
- [12] S. Vosoughi, P. Vijayaraghavan, and D. Roy. Tweet2vec: Learning tweet embeddings using character-level cnn-lstm encoder-decoder. *SIGIR '16*, 2016.
- [13] J. Wieting, M. Bansal, K. Gimpel, and K. Livescu. Charagram: Embedding words and sentences via character n-grams. In *Proceedings of the 2016 Conf. on Empirical Methods in Natural Language Processing*, p. 1504–1515, 2016.
- [14] G. Bekoulis, J. Deleu, T. Demeester, and Devellder C. Sub-event detection from twitter streams as a sequence labeling problem. In *Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Vol. 1, pp. 745–750, 2019.
- [15] P.K. Srijith, M. Hepple, K. Bontcheva, and D. Preotiu-Pietro. Sub-story detection in twitter with hierarchical dirichlet processes. *Information Processing & Management*, Vol. 53, No. 4, pp. 989–1003, 2017.
- [16] C. Xing, Y. Wang, J. Liu, Y. Huang, and W. Ma. Hashtag-based sub-event discovery using mutually generative lda in twitter. In *Proc. of the Thirtieth AAAI Conf. on Artificial Intelligence*, AAAI Press, pp. 2666–2672, 2016.
- [17] A. Chellal, M. Boughanem, and B. Dousset. Multi-criterion real time tweet summarization based upon adaptive threshold. In *2016 IEEE/WIC/ACM International Conf. on Web Intelligence*, pp. 264–271, 2016.
- [18] J. Kannan, AR. M. Shanavas, and S. Swaminathan. Twitter sports: Real time detection of key events from sports tweets. *Transactions on Machine Learning and Artificial Intelligence*, Vol. 5(6), No. 36, 2017.
- [19] E. Aramaki, S. Maskawa, and M. Morita. Twitter catches the flu: Detecting influenza epidemics using twitter. In *Proc. 2011 Conf. on Empirical Methods in Natural Language Processing*, pp. 1568–1576, 2011.
- [20] K. Lee, D. Palsetia, R. Narayanan, Md. Mostofa Patwary, A., A. Agrawal, and A. Choudhary. Twitter trending topic classification. In *11th IEEE International Conf. on Data Mining Workshops*, 2011.
- [21] G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, Vol. 313, No. 5786, pp. 504–507, 2006.
- [22] K. Cho, B. Merriënboer, V., D. Bahdanau, and Y. Bengio. On the properties of neural machine translation: Encoder-decoder approaches. In *SSST@EMNLP*, 2014.
- [23] J. Chung, C. Gulcehre, K. Cho, and Y. Bengi. Empirical evaluation of gated recurrent neural networks on sequence modeling. In *NIPS 2014 Workshop on Deep Learning*, 2014.
- [24] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958, 2014.
- [25] P. D. Kingma and J. Ba. Adam: A method for stochastic optimization. *International Conf. on Learning Representations*, 12 2014.