

Twitter 上のバースト現象とコンテンツとの関係分析 —テレビドラマを例として—

佐藤 由将^{†1} 大竹 恒平^{†2} 生田目 崇^{†2}

概要：本論文ではテレビドラマを対象に、放送期間に市場においてどのようにその番組について話題が盛り上がるかを分析する。分析対象は Twitter であり、特に急な盛り上がりを示すバースト現象に注目する。対象とするテレビドラマは1クール3カ月にわたって毎週放送されるものであるが、放送されていない時間においてもツイートおよびリツイートが観測され、またバースト現象も複数回観測された。さらに、各時点でのツイート内容について自然言語処理を通じた頻出単語の抽出を行い、どのような単語がバーストに結び付くのかについて分析する。こうした分析を通じて、市場におけるバースト現象の解明とともに、番組評価の理解につながることを期待される。

キーワード：テレビドラマ、バースト検知、Twitter、自然言語処理、ランダムフォレスト

1. はじめに

インターネットおよびスマートフォンや PC などの情報端末の普及により、一般消費者による情報発信が広く行われている。ウェブサイト作成やブログサービスを介した情報発信から、現在では SNS (Social Networking Service) のようなネットワーキングサービスまで広がっている。

その中で、SNS 上で投稿された製品やサービスに対するユーザの意見やクチコミが、消費者の率直な意見を反映するものとして注目を集めている。近年では、製品のみならず政治活動や災害時の情報媒体、映画のプロモーション活動など SNS を活用する業態は多岐に渡っている。テレビ業界においては、テレビ番組に対する視聴者の反応を示す指標として広く利用されている。本論文で対象とするテレビドラマについても視聴者の SNS 上の行動が注目を集めている。SNS 上のドラマに関する投稿の分析は、視聴者が評価しているシーンやコンテンツを直接特定することができ、テレビ局、制作会社、スポンサーなど多方面のマーケティング戦略に有益な知見をもたらすことが期待される。このような状況の下、短時間で話題が急激に盛り上がるような現象はバースト現象と呼ばれ、注目を集めている [1]。バースト現象の生起は特定の話題に関して急激に興味が増加したことを表しており、対象に対する評価においても大変重要と考えられる。

本論文では、テレビドラマを対象に Twitter 上に発信されたテレビドラマに関するツイートとそのリツイートを収集し、放送期間でのバースト現象の抽出とその原因についてツイート内容や属性などから評価することを試みる。

2. 既存の研究と本論文の目的

本節では、SNS やミニブログデータを対象とした、バースト現象に関する既存の研究について論じ、本論文の目的に

ついてまとめる。

水沼 [2] は Twitter 上のバースト現象を統計的見地から各種手法を用いて比較している。具体的には、ROKU、統計的分布に基づく棄却検定 (3σ 法と増山の検定)、MAD 法を用いて比較した。結果として、統計的分布に基づく棄却検定における 3σ 法によるバースト検知が適切であると提言した。 3σ 法は、データ集合の平均値と標準偏差を用いた外れ値の閾値により各データが外れ値かどうかを検出する手法である。また、発生したバースト現象を類型化し、イベントの特徴把握を行うことで、各バーストを5種類のタイプに分類した。

前川ら [3] は、単語の共起関係を用いたマイクロクラスタリングとバースト検知を行った。さらに、各クラスターの単語数やバースト度などを特徴量とした決定木分析を行うことで、クラスターを構成する Tweet の興味深いトピックを特定する手法を提案した。また、中原ら [4] は、ソーシャルビューイングにおけるツイートを対象に、バースト時の投稿内容とリアルタイムのセリフの一致性を検証した。また、興味対象のツイートを抽出する手法を提案し、ツイートの内容を効率的に把握することを可能にした。さらに、深沢ら [5] は、スポーツ中継中の Twitter 上のバースト検知を行い、テレビ番組のイベント検出に優れた方法の開発及び検出精度の検証を行った。

これらの研究においては、放送時のリアルタイムのバースト現象に注目して分析している。これに対して本論文では、数カ月にわたって連続して放送されるテレビドラマを対象とする。大きな違いとしては、リアルタイムのツイートを分析する場合は、その時点までに放映された映像に限定されたコンテンツが主にツイートされる場合がほとんどであるのに対して、期間内に連続したコンテンツが放送されるものの、1週間のインターバルがありその間にも番組に関する直接的及び間接的な情報の両者が継続的に発信さ

^{†1} 中央大学大学院

^{†2} 中央大学 (連絡先: otake@indsys.chuo-u.ac.jp)

投稿: 2017年11月30日

採択: 2018年3月1日

れている点である。したがって、リアルタイムな反応とは異なる情報発信の特徴がある可能性を指摘できる。そこで本論文においては、Twitter データを対象として、放送期間中のツイート並びにリツイートを取得し、その間に観測されるバースト現象とそのコンテンツの関係について分析することを目的とする。

3. データ概要

分析対象は2017年4月～6月に関東地方の民放局で放送された4つの連続ドラマであり、Twitter データはTwitter 社が公開している API (Application Program Interface) を用い、放送期間内を対象に各タイトルのハッシュタグによるクローリングを行っている。その結果表 1 に示すデータを得た。なお、データ取得期間は、初回ドラマ放映時間から最終回の放映日の1週間後までである。例えば、4月1日の22時に放映開始、6月10日に最終回を迎えた場合、4月1日の22時から6月16日23時59分までのデータを取得した。このデータについて、本論文では各週を1ケースとしてデータを分割している。すなわち各週の放送開始時間を区切りとして、1週間(168時間)ずつ区切って週次のデータとしている。最終回に該当するデータのみ168時間に満たないデータである。

各ドラマについて、分析に用いたデータ件数を表 1、各ドラマの内容などの概要を表 2、3にまとめる。なお、ドラマ B のみ11話、それ以外は10話の放送があった。

表 1 データ件数

ドラマ名	ツイート・リツイート件数
A	182,673
B	754,819
C	383,306
D	200,369

表 2 各ドラマの内容及び放送期間

ドラマ名	内容	放送期間
A	恋愛ドラマ	2017/04/18~2017/06/20
B	推理ドラマ	2017/04/17~2017/06/26
C	推理ドラマ	2017/04/14~2017/06/16
D	ヒューマンドラマ	2017/04/16~2017/06/18

表 3 各ドラマの放送曜日及び放送時間帯

ドラマ名	放送曜日	放送時間帯
A	毎週火曜	22:00~23:00
B	毎週月曜	21:00~22:00
C	毎週金曜	22:00~23:00
D	毎週日曜	21:00~22:00

4. 分析

本論文では、まず、ドラマ別のツイート及びリツイートデータについて1週ごとに区切る。そしてさらに1時間ごとに分ける。そのデータを用いて Kleinberg のバースト検知モデルによる各時間のバーストレベルの評価を行う。また、各時間のツイートの文章について、自然言語処理による特徴語抽出を行う。

これら2つのステップを合わせ、バーストが起こるときの特徴語が何であるかを分析することで、どのようなツイートがバーストに結び付くのかについて評価する。

これらの分析のフレームワークを図 1 に示す。

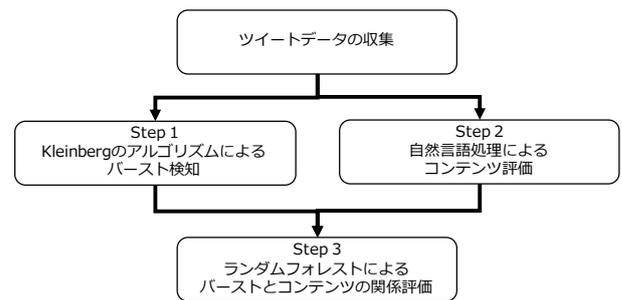


図 1 分析フレームワーク

4.1 Step1 バースト検知

トラフィック解析の分野では、短期間に大量のトラフィックが観測されるような場合をバーストと呼んでいる。バーストの状態は、通常のトラフィックをはるかに超えたトラフィックを観測した場合に抽出されるべきであり、こうしたバースト検知のためのアルゴリズムがいくつも提案されている。

本論文では、そのうち代表的なバースト検知手法である Kleinberg の手法 [6] を用いた。Kleinberg の方法では、低頻度にランダムにイベントが発生するなかで、集中的に高頻度にイベントが発生するような状態について、通常とは別の状態から発生したと考えた複状態を考える。そして、これらの状態間の遷移をオートマトンとして表現したモデルを提案している。

図 2 は、本論文で対象としたドラマのうち、ある週の1時間ごとのツイートおよびリツイート件数と検知したバーストレベルの比較である。左の縦軸がツイート・リツイート件数であり、右の縦軸がバーストレベルである。通常の状態はバーストレベル1であり、この例では、最大バーストレベルは7である。おおよそ、実際のツイート・リツイート件数の変化と併せてバーストが観測されているが、多少件数が増えてもバーストと判定されないこともある。また、放映中もしくは放映直後と、次週の放映直前の件数は多く、バーストレベルも高くなる。

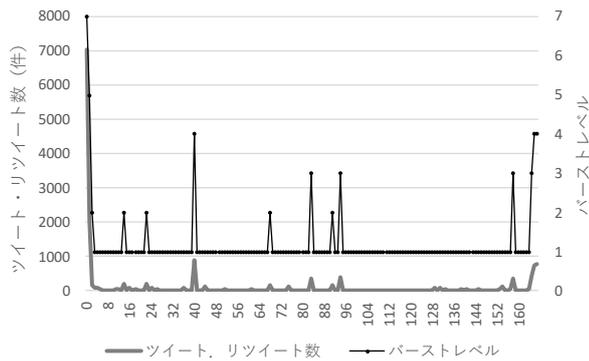


図2 バーストの推移

なお、レベル2以上のバーストと判定された時間を各ドラマ・各週に分けて集計したものが表4である。

表4 ドラマ毎に各週でのバーストした時間数

週	ドラマ			
	A	B	C	D
1	19	12	8	5
2	13	17	20	10
3	12	14	20	13
4	11	7	7	9
5	14	11	13	14
6	13	8	23	9
7	13	7	22	14
8	11	15	18	12
9	10	9	9	13
10	6	5	6	5
11		8		

全体としておよそ7%の時間にバーストが観測された。全ドラマに共通するパターンは観測されず、それぞれのドラマによって異なるバーストパターンとなっている。

なお、最終週が少ないのは、該当する時間数が少ないためである。

4.2 Step2 ツイートコンテンツの特徴分析

次に、ツイート及びリツイートの内容に着目し、自然言語処理解析によりツイートコンテンツの特徴を明らかにする。具体的には、はじめに、あるドラマについて行われたツイートを、そのドラマに関する文として捉え、ドラマを特徴的に表現することを期待した単語（特徴語）を、tf-idf法により特定する。次に、特定した特徴語をカテゴリに分類し、コンテンツの特徴について把握を試みる。

分析の手順を以下に示す。

1. 形態素解析による品詞分解
2. tf-idf法による単語への重みづけ及び特徴語の特定
3. 特徴語によるコンテンツの把握

分析には、期間中に行われた前述の4ドラマについてのツイート（表1を参照）を用いる。なお、分析にはR言語を用い、形態素解析エンジンにはオープンソースであるMeCab [7]を使用した。

4.2.1 形態素解析による品詞分解

はじめに、各ドラマについてのツイートを対象に、形態素解析を行った。形態素解析とは、自然言語（テキストデータ）から文を構成する最小要素である形態素に分割し、それぞれの形態素の品詞を判別する解析手法である。解析には、文法や辞書に定義される単語の品詞などの情報を用いる。形態素解析は自然言語処理を行う前段階として良く用いられる。

なお、品詞毎に出現頻度の集計を行った結果、名詞、動詞、形容詞の3品詞が全体の6割以上を占めている事が分かった。それぞれのドラマについて、名詞、動詞、形容詞の出現頻度の集計結果を表5に示す。本論文では、これらの3品詞に着目し、以後の分析を進める。

表5 各ドラマについての名詞、動詞、形容詞の出現頻度の集計結果

ドラマ 品詞	A	B	C	D
名詞	868,696	3,654,999	1,495,514	1,034,940
動詞	139,263	596,795	290,149	170,723
形容詞	34,957	142,176	67,330	38,303

4.2.2 tf-idf法による単語への重みづけ及び特徴語の特定

各ドラマには、そのドラマを特徴的に表現する語（特徴語）が存在すると考えられる。そこで、本論文では語に対して、ドラマを特徴的に表現する度合い（重み）を求め、特徴語の特定を行う。特徴語の特定には、tf-idf法を用いる [8]。tf-idf法は、単語に対する重み付けの指標の一種であり、単語の出現頻度 (tf) と逆文書頻度 (idf) の積により求める。

ここで、tfは文章中に出現する回数が大きいほど重要度が高いことを、idfは複数の文書で横断的に使われている単語は特定の文書に特有のものではないとするペナルティを表す。本論文では、tf-idf値を各ドラマについて特徴的な語（特徴語）を特定する際の指標として用いる。なお、各ドラマについての1ツイートを1文、ツイート群を1文章として捉え、全ての名詞、動詞、形容詞に対してtf-idf値の算出を行い、tf-idf値が上位50位までを特徴語と定めた。

4.2.3 特徴語によるコンテンツの把握

次に、tf-idf 値により定めた特徴語を用い、コンテンツの内容の把握を試みる。はじめに、それぞれのドラマの上位 50 単語を、その単語の内容に基づき、「登場人物_主要」、「登場人物_その他」、「役者_主要」、「役者_その他」、「地名・場所」、「タイトル」、「ポジティブ感情」、「ネガティブ感情」、「主要トピック」、「放送トピック」、「放送属性」、「放送局」、「他番組」の 13 カテゴリに分類し、これらの特徴語とした。なお、単語の内容を判断する際には、それぞれのドラマについての公式 Web サイト、フリー百科事典 (Wikipedia など) を参考にした。表 6 に、各ドラマについての、13 カテゴリに分類した特徴語数を示す。

表 6 特徴語として取り上げた単語数

カテゴリ	ドラマ			
	A	B	C	D
登場人物_主要	4	6	6	5
登場人物_その他	2	0	1	0
役者_主要	7	5	10	11
役者_その他	4	1	3	3
地名・場所	0	0	0	1
タイトル	4	2	2	1
ポジティブ感情	10	14	9	10
ネガティブ感情	4	5	4	2
主要トピック	2	3	1	4
放送トピック	9	12	3	11
放送属性	6	8	16	7
放送局	1	1	0	0
他番組	0	0	0	2

次に、表 6 に分類された特徴語を用い、各ドラマについて、カテゴリ毎の特徴語の出現比率をまとめる。なお、出現数の算出には、ある特徴語について、1 ツイート中に 1 回以上出現していた場合をカウントし、特徴語が属するカテゴリについて集計した値である。

表 7 より、全体の傾向として「タイトル」の構成比率が高いものの、他のカテゴリについてはドラマ毎に構成比率が高いカテゴリが存在することが分かる。例えばドラマ C では、ドラマ中の主要な登場人物についての特徴語である「登場人物_主要」や、そのキャストを演じている芸能人についての特徴語である「役者_主要」、「放送属性」についてのカテゴリの出現比率が、他のドラマに比べて高い。これに対し、ドラマ D では、ポジティブな感情 (楽しい、面白い、幸せなど) に関する特徴語である「ポジティブ感情」についての特徴語カテゴリの出現比率が他のドラマに比べて高く、また、他の番組名に関する特徴語である「他番組」が唯一出現している。これらの特徴語の集計結果から、ド

ラマによって話題となる内容が異なることが推察される。そこで、次節では、ツイートの内容によりバースト現象に与える影響が異なるという考えの下、特徴語を用いたバーストレベルへの影響について分析を行う。

表 7 頻出特徴語の集計結果

	A	B	C	D
登場人物_主要	0.068	0.043	0.120	0.110
登場人物_その他	0.021	0.000	0.005	0.000
役者_主要	0.099	0.068	0.192	0.144
役者_その他	0.029	0.007	0.020	0.010
地名・場所	0.000	0.000	0.000	0.004
タイトル	0.572	0.730	0.429	0.449
ポジティブ感情	0.072	0.060	0.053	0.080
ネガティブ感情	0.029	0.007	0.012	0.004
主要トピック	0.018	0.016	0.014	0.055
放送トピック	0.054	0.037	0.021	0.092
放送属性	0.033	0.031	0.134	0.038
放送局	0.006	0.000	0.000	0.000
他番組	0.000	0.000	0.000	0.013

4.3 Step3 バーストの要因分析

本節では、バーストに対して影響を与える内容の発見を目的に、ランダムフォレストによる特徴語カテゴリの重要度の把握を試みる。ランダムフォレストとは、アンサンブル学習手法の一つであり、複数の決定木を弱識別機として用い判別を行う [9]。ここでは、Step1 にて求めた時間毎のバーストレベルが 2 以上を 1、レベル 1 を 0 にカテゴリ化した値を目的変数に、Step2 にて求めた各ドラマについての 13 の特徴語カテゴリの出現頻度を時間毎に集計した値を説明変数として、ドラマ毎に判別モデルを作成した。表 8 に判別モデルの Accuracy (正答率)、Recall (再現率)、Precision (適合率)、F-value (F 値) を、表 9 に判別モデル中において重要度が高い説明変数を示す。なお、重要度としては該当変数を除いた際の正解率の減少についての平均値をもとに算出した。

表 8 判別精度の比較

分類精度	ドラマ			
	A	B	C	D
Accuracy	0.937	0.952	0.908	0.951
Recall	0.738	0.736	0.508	0.791
Precision	0.267	0.371	0.221	0.343
F-value	0.392	0.494	0.308	0.479

表9 ランダムフォレストから得られた重要度

ドラマ カテゴリ	A	B	C	D
登場人物_主要	1.493	8.408	-4.315	3.361
登場人物_その他	2.347	6.099	-11.393	0.000
役者_主要	7.818	8.848	-8.653	18.507
役者_その他	4.976	6.097	-14.172	1.874
地名・場所	0.000	6.526	0.000	-1.151
タイトル	0.178	4.338	1.690	6.539
ポジティブ感情	3.780	5.329	-11.073	1.770
ネガティブ感情	4.387	5.138	-3.017	3.149
主要トピック	0.528	7.480	-0.061	4.956
放送トピック	-0.539	5.907	-7.480	0.694
放送属性	10.697	5.898	-8.626	0.144
放送局	2.338	0.000	0.000	0.000
他番組	0.000	0.000	0.000	2.369

表8より、すべての判別モデルにおいて、Accuracyの値が90%を超えていることが分かる。ただし、バーストレベルをカテゴリ化した際、目的変数の比率は1:1ではなく、0に偏ったデータセットになっている。F値の値に注目すると、ドラマA、ドラマB、ドラマDについては概ね40%~50%であった。一方、ドラマCについては30%であり、最も判別精度が低かった。

表9より、特徴語カテゴリの重要度は、ドラマによって異なることが分かった。それぞれのドラマの特徴語カテゴリの傾向について、以下に述べる。

ドラマAでは、「放送属性」、「放送局」といった、ドラマの制作側に関する特徴語カテゴリが、バーストに対して影響を与えていることが分かった。また、「タイトル」に関する特徴語カテゴリの重要度が他のドラマと比べて低い。これにより、単純にドラマのタイトルに関するキーワードを含んだ話題の集計値だけではなく、制作側や登場人物を演じる役者（「役者_主要」）や、ポジティブ・ネガティブな感想（「ポジティブ感情」、「ネガティブ感情」）を含んだ話題を用いることにより、より精緻なバースト検知モデルが作成されていると推察される。

ドラマBでは、「登場人物_主要」や「役者_主要」の特徴語カテゴリの重要度が、バーストに対して影響を与えていることが分かった。ドラマBは主演として、有名アイドルグループのメンバーや、人気女優をキャストしている。そのため、ドラマ内のキャスト並びにそのキャラクターを演じている役者に関する話題が、バーストが起こる際のきっかけとなっていると推察できる。また、主に撮影場所に関する「地名・場所」が特徴語として抽出され、バーストへの影響を与えている。

ドラマCにおいては、「タイトル」以外の重要度が0、若

しくはマイナスとなっている。これは、「タイトル」以外の特徴語カテゴリでは、判別が困難であることを示している。表8の結果と合わせると、ドラマCについては判別モデルの作成において、学習が十分に行われていない可能性もある。

ドラマDでは、「役者_主要」がバーストに対して高い影響を与えている。一方で、「登場人物_主要」の重要度は特筆して高いわけではない。ドラマDは、実力派俳優を主演にキャストしている。また、他のドラマと比べて最も登場人物が多く、1話辺りに出演する役者数も多い。過去の出演作も多いことから、作中の登場人物よりも実際の役者についての話題が、バーストが起こる際のきっかけとなっていると推察できる。これは、ドラマBとは対照的といえる。また、他のドラマにはない「他番組」との比較が行われているという特徴もある。これは、前述の通り、俳優が出演した過去の作品との比較が主な話題となっていると推察される。

5. 考察

時間別のツイート・リツイート数については、概して夜中の時間帯の投稿数が少ない。また、放映前後の時間について投稿数が大きく増える。また、すべての週でバーストを検知しており、放映されていない時間も何らかの話題の盛り上がりが見られた。このようにバーストが検知される理由については、番組オフィシャルのアカウントによる情報提供や、俳優が別番組で番宣を行うといったことも考えられるが、実際に投稿を行ったアカウントに注目すると一般消費者のアカウントも多くみられた。

また、頻出特徴語を見ると、ドラマごとに大きく様子が異なっていることもわかり、ドラマそのものの内容の差異もさることながら、プロモート方法の違いも反映しているものと思われる。

バーストに影響を与える特徴語カテゴリについては、各ドラマによって大きく異なることが分かった。全体の傾向としては、主演俳優（役者_主要）並びに主演俳優が演じる登場人物（登場人物_主要）に関する特徴語カテゴリが、バーストに対して影響を与えていることが分かった。また、ポジティブ感情についての特徴語カテゴリは、ネガティブ感情についての特徴語カテゴリに対して、必ずしも重要であるわけではないことが分かった。むしろドラマA、ドラマDについては、ネガティブ感情を有する特徴語カテゴリの方が、重要度が高い。これは、ドラマBに比べ、ドラマA、Dはシリアスなテーマを主題としている点に起因するものであると考えられ、特徴語カテゴリによるコンテンツの把握は、各ドラマの特徴の一端を正確に捉えられていると考える。また、先行研究では、ある放送や事象に対するリアルタイムな状況把握としてバースト検知を試みたものが多いが、本論文においては数カ月わたって連続して放

送された、4つのテレビドラマを対象とした。期間中に行われた全てのツイートデータを用いたことで、多様なカテゴリからコンテンツの評価を行うことができたと考える。例えば、限定的に出演した役者の、次回放送回への出演報告の投稿や、情報番組・バラエティー番組での番組宣伝に関する投稿、出演者同士の舞台裏について語った記事・番組に関する投稿など、放送前後の事象についてのツイートが確認された。加えて、バーストの要因分析における判別モデルの作成においては、10ないし11週連続で放送されたドラマのデータを用いることで、各週におけるストーリーの盛り上がりを加味した、より精緻なモデルの作成を行うことができたと考える。

6. まとめと今後の課題

本論文ではテレビドラマを対象として、Twitterにおけるバースト検知とバーストが検知された時点におけるツイートのコンテンツに関する分析を行った。バーストは放送前後に観察されたが、それ以外の時間についてもかなりの時間検知された。また、本論文の分析結果から、ドラマごとのバーストについて、特定の特徴語との関係があることが示唆された。バーストに関係する特徴語をもとに、市場において興味を持たれる情報を発信することで、ドラマ放映時でなくても市場の興味を喚起する可能性がある。

本論文ではバースト検知について Kleinberg の指標を用いた。ただし、この指標は、各ケースについて個別に行うため、同じレベルでもツイートの絶対数が異なるといった欠点もある。また、オートマトンを個別に構築するため、サンプルサイズが増えると計算効率が大きく下がる。バースト検知に関する簡便的な方法もいくつか提案されているため、どのような検知方法が適しているのかを今後議論する必要がある。

さらに、本論文のデータについてバースト時間と視聴率の変化について分析した(図3)。図3の縦軸は、各週の視聴率とその前週の視聴率の差分を、視聴率の変化として表している。また、横軸は、表4に示した各週のバーストした時間数を表している。しかしながら、図3において、全データを通しての明らかな関係は見いだせなかった。ただし、図3を見てわかるように図の左上及び右下の少数のデータを除くとバースト時間と視聴率変化について正の関係があるようにも見えることから、バースト状態が頻出することにより、ドラマへの注目が集まりそのことで次回の放映時の視聴率が上昇するといったことも考えられる。ただしバースト状態とドラマへの注目のどちらが先かといった議論も必要なことからドラマのコンテンツとツイートの関係を分析することで、バーストと視聴率変化について有意な関係見いだせる可能性がある。これらは今後の課題としたい。

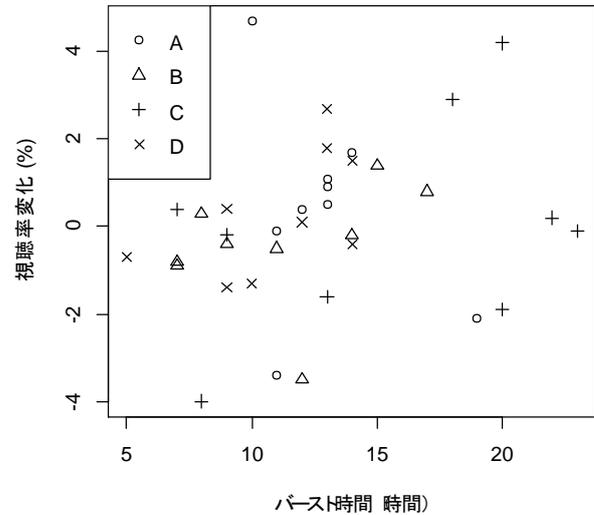


図3 バースト時間と視聴率変化の関係

謝辞 本研究において株式会社ルーター様には、データ収集の協力と有意義なディスカッションをいただきました。ここに謝意を表します。また、本研究は JSPS 科研費 JP16K03944, JP17K13809 の助成を受けたものです。

参考文献

- [1] 大竹洋平, 鈴木良弥 “マイクロブログのバースト現象に着目したユーザのクラスタリングおよび可視化” 言語処理学会第21回年次大会 発表論文集 pp. 226-229 (2015).
- [2] 水沼 友宏 「Twitter におけるバーストの検出と生起要因に関する分析」. 筑波大学大学院図書館情報メディア研究科修士論文 (2015).
- [3] 前川浩基, 中原孝信, 羽室行信, “テレビ番組視聴時における twitter 投稿のバースト検知と情報配信の可能性,” 第 27 回人工知能学会講演論文集, 4 pages (2013).
- [4] 中原孝信, 前川浩基, 羽室行信, “テレビ番組視聴時における Twitter 投稿からのトピック検知,” オペレーション・リサーチ, Vol.58, No.8, pp.442-448 (2013).
- [5] 深沢知明, 高島真之介, 羽山徹彰 “Twitter データを用いたテレビ番組のイベント検出に関する研究,” 情報処理学会第 77 回全国大会論文集, pp. 1-547-548 (2015).
- [6] Kleinberg, J. “Bursty and Hierarchical Structure in Streams,” *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 91-101 (2002).
- [7] MeCab, <https://www.mlab.im.dendai.ac.jp/~yamada/ir/MorphologicalAnalyzer/MeCab.html> (2017年11月20日アクセス)
- [8] Baeza-Yates, R. A. and Ribeiro-Neto, B. A., *Modern Information Retrieval: the Concepts and Technology behind Search* (2nd Edition). Addison-Wesley (2011).
- [9] L. Breiman, “Random Forests”. *Machine Learning*, Vol. 45, No. 1, pp. 5-32 (2001).

Analysis between Burst Phenomena and their Contents on Twitter — Case Study of TV Drama —

Yusuke SATO^{†1} Kohei OTAKE^{†2} Takashi NAMATAME^{†2}

Abstract: In this study, we treat TV dramas and analyze how these TV dramas become hot topics. Analyzing data is Twitter data, then we focus on burst phenomenon which is increases the number of tweet and retweet rapidly. The objective TV dramas were broadcast each week in three-month run of a television series. We could observe tweet and retweet data even while not broadcast time, moreover we obtained some bursts phenomena by analyzing. Furthermore, we pick up some frequent appeared term from the contents of tweet using natural language processing, then we analyze the relationship between burst and terms. Through these analyses, we can solve the burst phenomena in market and expect to understand the evaluation of TV program.

Keywords: TV Drama, Burst Phenomena, Twitter, Natural Language Processing, Random Forest

^{†1} Graduate School of Chuo University
^{†2} Chuo University (Correspondence Author: otake@indsys.chuo-u.ac.jp)
Submitted: 30/11/2017
Accepted: 01/03/2018