

# SNS 連動型デジタルサイネージのための記事掲載期間自動判定機構

## An automatic judgment mechanism of posting period for SNS linked digital signage

長島悠貴十 伊藤雄二†1 小坂弘史†2 湯瀬裕昭† 渡邊貴之†

Yuki Nagashima Yuji Ito Hiroshi Kosaka Hiroaki Yuze Takayuki Watanabe

### 1. はじめに

デジタルサイネージは公共空間において広く使われるようになり、我々は日常的に多くのデジタルサイネージを目にするようになった。それに伴い、様々な機能を備えた多様なデジタルサイネージが登場している。最近では、東急電鉄が乗り換え案内を検索できるサイネージを駅のホームへ設置を行った。[1]しかし、デジタルサイネージを設置しただけでは視聴者の興味・関心を効果的に引くことはできないため、コンテンツや機能の充実化、広告の工夫などが行われている。このような、デジタルサイネージを日常的に観光などの情報発信に利用し、災害時にもそれを活用するというニーズが高まってきている。

このような背景から、我々は災害時にも利用可能なデジタルサイネージシステムの開発を行ってきた。文献[2]において我々は、「サイネージステーション構想」を利用した安定運用が可能であることを示した。「サイネージステーション構想」とは、ソーラーパネルと衛星回線を利用することによって、停電や回線不通といったトラブルが生じた場合でもサイネージを継続運用するための仕組みである。2016年7月には、岩手県遠野市の道の駅遠野「風の丘」に開発したデジタルサイネージを設置した。その後、東京銀座の岩手県のアンテナショップ「いわて銀河プラザ」、岩手県盛岡市の岩手県の特産品ショップ「らら・いわて盛岡店」にも設置し、設置箇所の拡大を行っている。設置したサイネージの画面例を図1に示す。

サイネージにおいて、ありきたりな広告を表示しているだけでは、効果的に視聴者の興味関心を引くことはできない。そのため、複数のサイネージを使用し、連動するコンテンツを作るなど様々な工夫が行われている。しかし、これらは簡単に作成できるものではない。このように、デジタルサイネージを運用する際の課題としては、コンテンツ作成に手間を要することが挙げられる。近年この問題を解決するために、TwitterなどのSNSへの投稿をデジタルサイネージのコンテンツとして表示するSNS連動型デジタルサイネージが増加している。店舗設置型のSNS連動型デジタルサイネージは、その店舗のSNS投稿だけを掲載すれば良いが、公共空間に設置するSNS連動型デジタルサイネージでは、複数のアカウントの投稿から掲載したい投稿を選別する必要がある。我々の開発しているシステムでは、Twitterに投稿されたツイートを表示するCGM連携機能を搭載しており、SNS連動型デジタルサイネージとなっている。この機能は、サイネージ管理者のTwitterアカウントで掲載したいツイートをリツイートすることでサイネージ上にツイートが掲載されるというものである。ツイートの選別を管理者が行う必要がある場合、デジタルサイネージの設置箇所の増大に伴って管理者に

かかる手間も増大するという課題がある。その課題を解決するために、我々は文献[3]においてページアンフィルターを用いた投稿自動判別システムの実装・評価を行った。これは、サイネージ管理者が過去に掲載したツイートを学習し、自動でツイートの掲載判別を行うという機能である。

サイネージ上にツイートを掲載する作業の自動化対応は行ったが、掲載期間が終わった際にそのツイートの掲載を解除する必要がある。掲載する作業と同様に設置箇所の増加によって管理者にかかる手間の増大が課題となる。

本研究では、その課題を解決するためサイネージ上にツイートを掲載する際に、ツイート内から日付表現を抽出し、掲載期間を判定する機能の実装・評価を行う。

### 2. 関連研究

日付に関わらず、特定の表現を文書の中から抽出しようという研究は数多く行われている。

文献[4]では看護記録から看護現場において有用な情報を抽出するために、形態素解析と正規表現によるパターンマッチングを併用したシステムの開発を行っている。看護記録を分析する際には、患者の状態などの記述の対象だけでなく表現の正確性や適切性も分析の対象となる。それらを分析するためには、形態素解析によって単語を切り出すよりもパターンマッチングを使用するほうが効果的であるとしている。

また、文献[5]においては、形態素解析での数詞の未知語処理として、正規表現ルールを用いた処理によって数詞の抽出を行っている。形態素解析を行う際に、正規表現ルール処理を挟むことによって辞書に登録できない無数のバリエーションがある数詞の判別を可能としている。また、実装の際にUnicodeを用いることによって日本語、中国語、韓国語で共通のルールを定義できることが証明されている。

上記2件の研究では、特定の表現を抽出する際に、形態素解析の処理の過程で正規表現ルールによる抽出を行っている。本研究では、抽出したい表現は日付表現に限られるので、必ずしも形態素解析は必要とされない。そのため、今回は、正規表現ルールを用いることで、表記揺れを広い範囲でカバーすることにより日付表現の抽出を行う。



図1 サイネージ画面例

† 静岡県立大学

†1 (株)メディア・ミックス静岡

†2 スーパーJSAT(株)

### 3. 記事掲載期間自動判定機構

本研究では、道の駅遠野「風の丘」に設置されているサイネージの管理者の Twitter アカウントがフォローしているある行政の広報アカウント  $\alpha$  のツイートを利用する。 $\alpha$  が 2016 年 5 月 31 日から 2017 年 6 月 5 日までに投稿したツイート 516 件に含まれる日付表現をベースに正規表現ルールの作成を行った。 $\alpha$  がリツイートしたツイート及びリプライは除外した。 $\alpha$  の投稿は様々な日付表現を用いている。また 1 件の投稿の中に複数の日付が入っているため、文書内で抽出された日付を比較して掲載期間を判断するという処理においても適していると判断した。

本研究で使用する正規表現のパターンの例は図 2 の通りである。この他作成したルールは以下の通りである

1. 年月日表現(-, /, /区切りのものもルール内に加えた)
2. 月日表現(1 と同様)
3. 日付のみの表現
4. 曜日表現
5. 本日・今日・今夜などの当日を示す表現
6. 昨日・一昨日・明日・明後日などの前後の日程を示す表現
7. 今月・来月・先月○日などの表現
8. 期間を表す表現(～, , , のもの)

上記ルールに該当したものを抽出し、年/月/日の形に統一を行う。抽出された日付の翌日を掲載終了日とする。複数の日付表現が見つかった際には、統一した日付を比較して一番最新の日付の翌日を掲載終了日とする。

上記ルールのみであるとして、図 3 のような内容のツイートは判定に用いたいものは、縮切の日付だが、最新の日付として日程の日付が存在するため正しい判定ができない。これを解決するために、ツイート内容を一文区切りにし、同文内に日付と縮切表現が抽出された際には、その日付に重み付けを行い日付の比較を行う。これによって縮切の日付の判定を可能とした。

### 4. 性能評価

本研究では、道の駅遠野「風の丘」に設置されているサイネージの管理者の Twitter アカウントがフォローしている複数のアカウントのツイートに対して以下のように評価を行う。自動判定機構による判定結果と、実際にツイートの内容を確認して人手による判定した結果の 2 者を比較し正答率を求めた。

判定に使用するデータセットは、以下の 4 つのアカウントが投稿したツイートを利用した。データセット A はルールを作るのに使用したアカウント  $\alpha$  のツイート。データセット B は、ある観光関係の宣伝を行っているアカウントのツイート。データセット C は、ある市の情報発信を行っている公式アカウントのツイート。データセット D は、旅館やホテルの宣伝などを行っているユーザーアカウントの投稿したツイート。データセット E は、ある空港の情報発信を行っているユーザーアカウントのツイート。データセットごとのツイートの投稿期間、データ件数、判定結果を表 1 に示す。

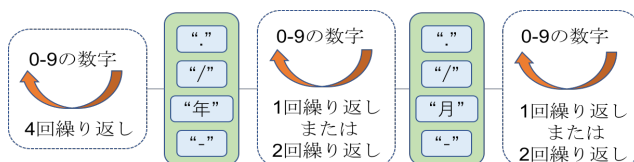


図 2 正規表現パターン

【新緑のふるさと協力隊】③静岡県静岡市  
日程：10/5(木)～9(月・祝)  
内容：秋の収穫祭(稲刈り、もちつき)、町民大運動会参加、豆腐作り、カニ漁見学、キノコ採り、神楽見学、農家民泊  
定員：10名  
締切：9/21(木)

図 3 正しく判別できないツイートの例

各データセットで非常に高い正答率が確認できた。 $\alpha$  の投稿をベースとしたルールで、様々なバリエーションがある日付表現の多くをカバーできたと考えられる。正しく判定できなかったものの例としては、以下のものが挙げられる

1. 岩手-名古屋間を 1 日 4 往復します
2. 10 月 2 日から 3 日間セールを・・・
3. 今日の最低気温は 9.8℃と予想されます
4. 2017 年 6 月 25 日に行われるイベントで・・・
5. 12 月 10 日～1 月 15 日にかけて・・・

文脈的判断が必要だと考えられるもの、投稿の誤字脱字については正規表現ルールでの限界と考えられる。2, 3, 5 のようなものは、ルールを追加していくことで対応していく。

### 5. まとめ

本研究では、デジタルサイネージに掲載したツイートの掲載期間が終わった際に、ツイートの掲載を解除するための自動判定機構の実装・評価を行った。判定基準は、ツイート内から日付表現の抽出を行い、掲載期間を自動で判断させた。今回は正規表現のパターンマッチングによる抽出を行ったが、今後は機械学習などを用いて日付を固有表現として抽出して判定する機能の検討を行いたい。また文章としてではなく、ツイートに添付された画像に掲載期間を判断する材料となる期間が載っている場合もあるため、画像を OCR を用いて解析し掲載期間の判断材料とする機能の追加も検討したい。

### 参考文献

- [1]. サイポ, “武蔵小杉のデジタルサイネージを調査してきた”, <https://www.si-po.jp/post/market/27479.html>
- [2]. 工藤直哉, 山上真由, 伊藤裕二, 小坂弘史, 湯瀬裕昭, 渡邊貴之, “オープンデータとソーシャルメディアを活用した Web-based デジタルサイネージシステムの複数環境における運用と評価”, 情報処理学会第 78 回全国大会, 2016 年 3 月
- [3]. 長島悠貴, 工藤直哉, 伊藤裕二, 小坂弘史, 湯瀬裕昭, 渡邊貴之, “CGM 連携機能を持ったデジタルサイネージにおける投稿自動判別システム”, 情報処理学会第 79 回全国大会, 2017 年 3 月
- [4]. 村松洋, 渡部勇, 大崎千恵子, 小塚和人, “看護記録のテキストマイニング”, 情報処理学会論文誌 データベース Vol.3 No.3, 2010 年 9 月
- [5]. 青木和夫, 中山章弘, 松崎剛士, “形態素解析での数詞の未知語処理”, 情報処理学会第 66 回全国大会, 2004 年 3 月

表 1 判定結果

データセット	投稿期間	データ件数	正答率
A	2016/5/31~2017/06/5	516 件	94.76%
B	2016/3/29~2017/6/5	485 件	96.70%
C	2016/1/4~2017/6/3	447 件	96.42%
D	2017/4/21~2017/6/5	1191 件	96.96%
E	2016/7/1~2017/6/5	423 件	96.92%