

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

長島悠貴† 工藤直哉† 伊藤裕二†1 小坂弘史†2 湯瀬裕昭† 渡邊貴之†
静岡県立大学† (株)メディア・ミックス静岡†1 スカパーJSAT(株)†2

1. はじめに

我々は、W3C の標準化ドラフト[1]や国内標準仕様案[2]をもとに、大学内で使用するデジタルサイネージの開発を行ってきた。開発したシステムは、静岡県立大学において 2015 年 8 月から本格稼働し、現在 2 キャンパス計 18 台のデジタルサイネージが運用されている。

また、デジタルサイネージは地震などの災害時に災害情報を提供する媒体としても注目されている。しかし実際の災害時には、停電や地上のインターネット回線が切断され、情報提供が難しくなることが想定される。そこで、我々はソーラーパネルと衛星回線を組み合わせることにより、停電や回線不通といったトラブルが発生した場合でもデジタルサイネージを継続運用する「サイネージステーション構想」を提案し文献[3]において衛星回線を使用した場合でも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は、災害時における情報伝達の手段として有望であり、公共性の高い駅(鉄道や道の駅)や空港などへの設置を進めている。2016 年 7 月には岩手県遠野市の道の駅遠野「風の丘」にサイネージステーションを設置した。設置したサイネージステーションの画面例を図 1 に示す。

我々の開発しているシステムでは、Twitter [4]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いる。この機能では、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、サイネージにツイートが掲載される。この機能を活用すれば、平時には観光やイベントの情報を掲載し、災害時には被災状況や避難情報を掲載することができる。CGM を活用することで、管理者側でコンテンツを作成するコストを削減できる。しかし、既存の CGM 連携機能では、サイネージ管理者が手動で表示したいツイートの選別を行っており、今

後システムの設置箇所が増加するに従い、ツイート選別にかかる手間の増大が課題となっている。

本研究では、これらの課題を解決するために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機能を実装し、これを評価する。

2. 関連研究

我々は文献[5]において、サイネージに掲載するツイートの候補を検索するために、過去に掲載したツイートに含まれる単語の共起尺度を利用した検索キーワード選定機能の開発を行い、管理者の負担を軽減する方法を検討した。しかし、検索キーワードを絞り込みすぎると掲載すべきツイートの取りこぼしが発生し、逆に絞り込みが甘いと関係のないツイートまで大量に抽出してしまうため、その調整が非常に困難である。

一方で、サイネージに掲載するツイートを発信するユーザは特定のアカウントにほぼ限られている。例えば大学であれば、広報部門のアカウントや、クラブサークルのアカウントなどである。そのため、それらのアカウントをあらかじめフォローしておくことで、タイムラインに流れてきたツイートをリツイートするかしないかで掲載の可否を判定することができる。

文献[6]では、ツイートのハッシュタグを学習したページアンフィilterによる分類器を用意し、ハッシュタグのついていないツイートのタグを推定する手法について述べている。本研究では、ハッシュタグを分類するのではなく、リツイートをするか否かを分類するためにページアンフィilterを活用する。

3. 投稿自動判別システム

まず、サイネージ管理者の Twitter アカウントを用いて、自動判別対象者のアカウントを登録したリストを作成しておく。

次に、アカウントごとに管理者が過去に掲載したツイートについて MeCab[7]を用いて形態素解析を行い、単語とその出現数を集計する。同

Auto-Detect System of Posted Articles in Digital Signage Having CGM Linkage Function

† Yuuki Nagashima, Naoya Kudo, Hiroaki Yuze, Takayuki Watanabe, School of Management and Information, University of Shizuoka.

†1 Yuji Ito, Media Mix Shizuoka Co., Ltd.

†2 Hiroshi Kosaka, SKY Perfect JSAT Corporation.

様に掲載しなかったツイートに含まれる単語とその出現数を集計する。

集計した単語とその出現数を用いて、ページアンフィルターの分類器の学習を行う。ページアンフィルターとは、文中に含まれる単語は全て独立であると仮定して、文書のフィルタリングを行う分類器である。学習では、サイネージ管理者がツイートを掲載した確率 $P(H_1)$ と、掲載しなかった確率 $P(H_2)$ を計算しておく。また、掲載したツイートに単語 d_i が現れる確率である $P(d_i|H_1)$ と、掲載しなかったツイートに単語 d_i が現れる確率 $P(d_i|H_2)$ を計算しておく。

学習した分類器を用いて、新たなツイートの掲載の可否を判定する際には、ツイートに含まれる単語 $D = \{d_1, d_2, \dots, d_n\}$ に対して、ツイートを掲載する確率 $P(H_1|D)$ と、掲載しない確率 $P(H_2|D)$ を式(1)(2)により求める。

$$P(H_1|D) = \frac{P(D|H_1)P(H_1)}{P(D)} \quad (1)$$

$$P(H_2|D) = \frac{P(D|H_2)P(H_2)}{P(D)} \quad (2)$$

$P(H_1|D) > P(H_2|D)$ が成り立つとき、ツイートは掲載と判定される。また、ゼロ頻度問題を避けるため、ラプラススムージングを適用している。

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 アカウントを用いて、ある観光施設のユーザアカウント A について自動判別機能の評価を行った。約5ヶ月間に渡ってサイネージ管理者が選別したツイートを学習させ、学習後から1週間にわたるツイートについて、自動判定を行う。自動判定結果と、管理者が実際に手動で判定した結果の二者を比較し、評価指標に基づき評価を行う。評価指標として、適合率と再現率、およびF値を求めることで評価を行う。

まず学習データとして、2016年7月18日から2016年12月9日までにユーザアカウントAが投稿したツイート186件を学習させた。これを期間Aとする。その後の一週間における新たなツイート21件について自動判別と手動判別の比較を行い、評価指標を計算した結果を表1に示す。結果からF値は0.7となった。次に、学習データとして用いる期間を1週間延長し、この期間内にユーザアカウントAが投稿したツイート207件を学習させた。これを期間Bとする。その後の新たなツイートについて自動判別と手動判別の比較を行い、評価指標を計算した結果を表1に示す。

表1から学習データ数を増加させるほど、F値が上昇する結果が得られた。

5. まとめ

本研究では、デジタルサイネージに掲載するツイートの選別にかかる手間の増大に対応するために、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機能を実装し、評価を行った。今後は、平時には観光情報を、災害時には被災状況や避難情報を掲載できるように、平時と災害発生時でフィルターを切り替える機能の実装を検討したい。

参考文献

- [1]. World Wide Web Consortium, "Web-based Signage UseCasesandRequirements", <https://www.w3.org/2016/websigns/ucr/>
- [2]. 総務省, "デジタルサイネージ標準システム相互運用ガイドライン", http://www.soumu.go.jp/main_content/000398550.pdf
- [3]. 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.
- [4]. Twitter, <https://twitter.com/>
- [5]. 近藤貴裕, 赤堀優志, 渡邊貴之, "Twitter 連動型サイネージにおけるコンテンツ掲示支援機能とその評価", 情報処理学会第77回全国大会, 2015年3月.
- [6]. 竹中姫子, 古宮嘉那子, 小谷善行, "ページアンフィルターを用いたTwitterにおけるツイートのハッシュタグ分類", 情報処理学会研究報告, Vol. 2011-IFAT-102 No. 1, 2011年3月.
- [7]. MeCab, <http://taku910.github.io/mecab/>

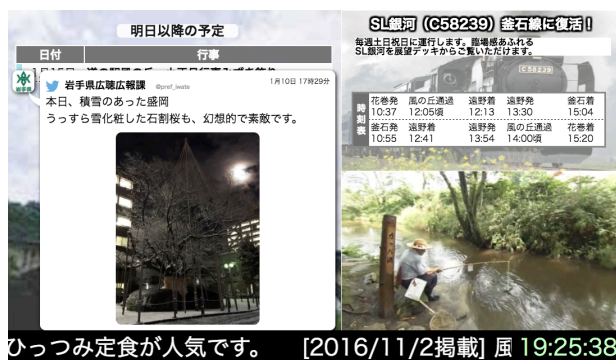


図1 サイネージステーションの画面例

表1 実験結果

	適合率	再現率	F 値
期間 A	0. 6667	0. 7369	0. 7000
期間 B	0. 8333	0. 8333	0. 8333