

SOMを用いたデジタルサイネージ掲載コンテンツの可視化支援

赤堀優志[†] 近藤貴裕[†] 渡邊貴之[†]

静岡県立大学大学院経営情報イノベーション研究科[†]

1. はじめに

近年、デジタルサイネージは、新しいデジタル産業として急成長し注目を集めている。デジタルサイネージはネットワークに接続したディスプレイなどの表示機器により情報を提供するものであり、場所や時間などの条件に合わせたコンテンツの提供が可能であるため、広告媒体として様々な業界が参入を始めている。また、カメラやスマートフォンアプリ連動などユーザの需要に合わせたコンテンツを提供したりするなど、新たなサービスも広がっている。

我々はこれまでに大学向けのデジタルサイネージシステムを開発し、本学で運用してきた[1]。デジタルサイネージのコンテンツの一つとして、SNSと連動し、利用者からの記事投稿も行える機能を実装して本学に関する情報を提供している。しかし実際には、利用者の関心が低下し、少数の特定ユーザによる投稿記事のみを掲載することになってしまっているのが現状である。そこで、本学に関係している記事を、管理者が手作業で探し出して掲載することを試みた。しかし、記事を掲載するにあたり、デジタルサイネージに掲載しても問題がないかどうかの可否を、管理者がひとつひとつ投稿記事の確認を行っているため、作業量が膨大になってしまうという問題がある。

本研究では、SNS 連動型デジタルサイネージを対象として、SNS に投稿された記事から、デジタルサイネージのコンテンツへの掲載の可否や分類を行う管理者の作業を支援するために、自己組織化マップ(Self-Organizing Map : SOM)[2]を用いた可視化支援について検討する。SOMによって分類、可視化することにより、投稿記事の分析および、サイネージ管理者の人為的なミスや負担を軽減し、管理者作業の効率向上を目的とする。対象とするSNSとしてマイクロブログサービスである Twitter を想定し、システム構築を行い、その有用性について考察する。

2. システムの概要

システムのフローチャートを図1に示す。まず、これまでに本学のデジタルサイネージに掲載済みのツイートデータを形態素解析して SOM により分類する。次に、Twitter の全ツイートデータの中から新たな掲載候補のツイートを、管理者のキーワード入力によって検索し、形態素解析して一覧表示する。その後、掲載済みのツイートで学習させた SOM に照合することで掲載候補記事がどのような内容な記事なのか分類し、学習を繰り返していく。SOM による可視化によりサイネージ管理者が容易に掲載候補記事の掲載可否の判定を支援する。

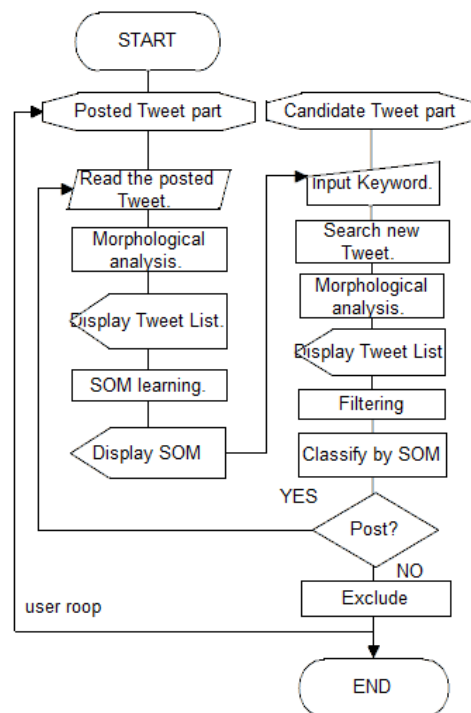


図.1 システムフローチャート

2.1 掲載済みツイートの取得

Twitter データの取得には TwitterAPI を使用する[3]。投稿済みのツイートデータは、デジタルサイネージの Twitter アカウントの User Timeline から取得する。User Timeline には、これまでに投稿したツイートとリツイートのみが反映されている。その後、取得した掲載済みツイートの形態素解析を行い、アプリケーションウィンドウの左半分に一覧表示する。

An assistance function for digital signage based on content visualization using SOM.

[†]Yushi Akahori, Takahiro Kondo, Takayuki Watanabe: Graduate School of Management and Information of Innovation, University of Shizuoka. J13101@u-shizuoka-ken.ac.jp

2.2 掲載候補ツイートの取得

掲載候補ツイートの取得は、Twitter Search API を使用して、管理者が任意のキーワードを入力し取得する。その後、形態素解析を行い、アプリケーションプログラムの右半分に一覧表示する。また、取得した掲載候補ツイートには、手動で不要なツイートを除外することで、ベイジアンフィルタの学習を行う機能を付加した。これにより、検索するたびに掲載候補ツイートの中の不要な記事を自動で除外する。

2.3 形態素解析

形態素解析は、文を形態素単位に分割して、辞書を用いて品詞の情報や意味、活用形などを付加することである。本研究では、形態素解析ツールとして MeCab[4]を使用し、名詞と形容詞を抽出して学習データを作成している。ただし、Twitter 上の文書は流行や一般的な文法とは違った表現などにより適切に解析できないことも多くみられるという欠点には、MeCab の辞書を、はてな[5]と Wikipedia[6]で鍛えることで対応した。

2.4 自己組織化マップ(SOM)による分類

自己組織化マップ(SOM)は、コホネンにより提唱された競合学習型ニューラルネットワークの一種であり、 n 次元のベクトル集団を学習することにより2次元のマップにそれらのベクトルの関係を写像することができる。

本システムでは、形態素解析を行った掲載済みツイートを学習データとして SOM に入力し学習させる。さらに、SOM の重みベクトルに対し、 k -means 法を適用することで k 個のクラスに分類する。その後、形態素解析を行った掲載候補ツイートについて、掲載済みツイートで教師なし学習を行ったノード集合の中から類似度の高いノードを選択し、そのノードが含まれるクラスにおける出現頻度の高い単語群を表示する。

3. 可視化結果

図2に本システムの表示例を示す。表示には JavaScript を用いた WebAPI を使用している。メインウィンドウ画面左側に掲載済みのツイート一覧、画面右側に掲載候補(キーワード検索結果)ツイートを一覧表示している。SOM の実行結果を子ウィンドウとして前面に表示している。SOM 出力画面には、六角形の格子が描かれており、SOM のノードを示している。ツイートまたは、SOM ノードにカーソルを合わせると、対象のツイートがどのクラスに分類されているかがわかるように表示が変化する。

可視化を行ったことにより、掲載候補ツイートのおおよその内容が割り当てられたクラスから予想でき、管理者が掲載候補ツイートの内容を直

感的に把握できるため、管理者の作業量を減らすことができる。また、これまでに掲載してきたツイートの内容の傾向も把握できる。



図2 掲載支援システムの表示例

4. まとめ

本研究では、SNS 連動型デジタルサイネージのコンテンツ投稿支援のために、デジタルサイネージに掲載したツイート、および掲載候補のツイートを SOM により分類し可視化を行った。

デジタルサイネージに掲載される記事は、季節やイベントによって出現単語に偏りが生じると考えられたため、過去の学習データが常に有効であるとは限らない。そのため、今後の展望として、時系列の伴った分類手法を実装し、利用者に有益な情報を見逃さない可視化システムの向上を目指す。

また、今回は連動している SNS の Twitter を対象に支援システムを構築したが、その他のデジタルサイネージに掲載するコンテンツにも似たような課題は残っているため、今後はそれらの掲載支援システムの構築も必要であると考えられる。

参考文献

- [1]. 近藤貴裕, 杉山航平, 大多和均, 渡邊貴之: 大学におけるデジタルサイネージシステムへの CGM 機能の検討と設計開発, 情報通信学会第 75 回全国大会, 2013.3
- [2]. T. コホネン著, 徳高平蔵, 堀尾恵一, 大北正昭, 大藪又茂, 藤村喜久郎訳: 自己組織化マップ, シュプリンガーフェアラーク東京, 2005 年
- [3]. Twitter: <https://twitter.com>
- [4]. MeCab: mecab.sourceforge.jp/
- [5]. はてな: www.hatena.ne.jp
- [6]. ウィキペディア(Wikipedia): ja.wikipedia.org/
- [7]. 野呂和誉, 中川敦志, 大井健太郎, 高瀬治彦, 北英彦, 森田直樹, 林照峯: 記述式小テストの解答の傾向を把握するための分類手法, 2007 PC カンファレンス論文集, 2007, p.39-42
- [8]. 東宏一, 掛谷英紀: 国会議員のツイッター分類とその応用, 言語処理学会第 18 回年次大会発表論文集, 2012.3
- [9]. 大北正昭, 徳高平蔵, 藤村喜久郎, 権田英功編: 自己組織化マップとそのツール, シュプリンガー・ジャパン, 2008.12
- [10]. 中村伊知哉, 石戸奈々子: 日本を動かす次世代メディア デジタルサイネージ戦略電子看板最前線