

## 重み付き木構造カーネルによるショートコメント分類

総合情報研究科 情報科学専攻 椎名 広光、武田 昌大 (椎名研究室)

**Keywords:** 自然言語処理、文章分類、木構造カーネル、Twitter、Wikipedia

### 1. 研究目的

近年、TwitterやFacebook等のWebサービスの発展により、地域性や流行性のある様々な情報が豊富に手に入るようになった。自然言語処理では、これらWebサービスを通じて発信される膨大な文章情報から有用な情報を如何に自動的かつ高精度に抽出するかという研究が重要な課題となっている。特に、Twitterのようなコメント情報においては、そのコメントのみから得られる素性が限定的であるため、他の外部リソースを活用して、意味情報の拡張を行う研究が盛んに行われている。そこで、本研究では、大規模百科事典であるWikipediaのカテゴリ構造に着目し、Twitterのコメントをカテゴリ構造に変換することで、コメントに対して効果的な文章分類手法を提案する。さらに、提案手法を観光情報の推薦システムへ応用することで、実用面での有効性を示す。

### 2. 文章分類手法の概要

本研究では、Tweetなどのショートコメントを重み付き木構造情報に変換した後、サポートベクターマシンで観光情報であるか否かの分類を行っている (Fig.1)。

分類処理の流れを以下に示す。

- ① ナイーブベイズを用いて、TwitterのコメントがWikipediaのどのカテゴリに所属するか推定する。
- ② Wikipediaの階層構造の各カテゴリからルートまでの部分パスを合成して木構造を作る。また、木構造のノードの重要度をTF-IDFを用いて重み付けする。
- ③ サポートベクターマシンを用いて、木構造を対象にした分類を行う。
- ④ 分類した結果、観光情報に関連するTweetを用いて、検索を行う。

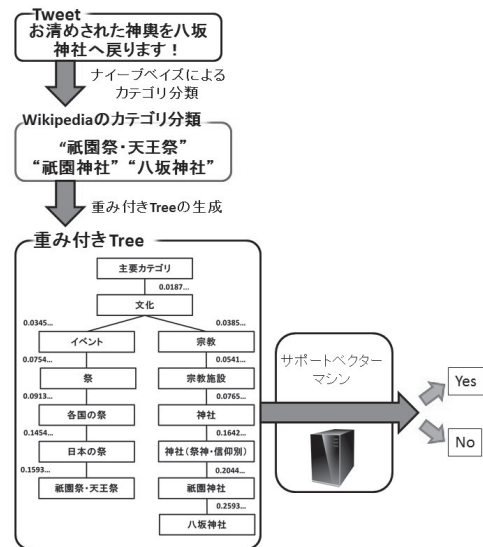


Fig.1 提案手法の処理過程

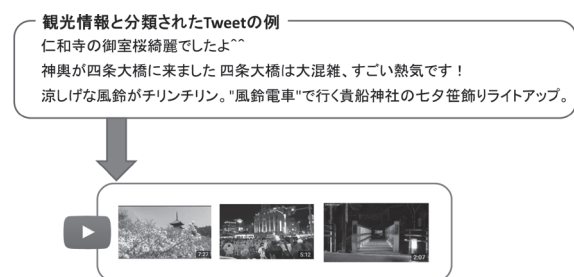


Fig.2 システムの実行画面

### 3. 旅行・観光情報の動画検索システム

提案手法を応用することで、Twitterのコメントから観光情報として有用なコメントを識別し、そのコメントと関連するYouTubeの動画をユーザに提供する旅行・観光情報の動画検索システムを作成した (Fig.2)。

### 4. 応用先と今後の課題

本研究は、コメントをWikipediaの豊富なカテゴリ情報を用いて木構造化していることから、観光情報だけでなく、幅広いジャンルにおけるコメント分類を少ない手間で行えると考えられる。今後分類精度向上のためのデータ構造とその付加情報の精査が必要である。