F-021

Twitter とテレビ視聴履歴を用いた個人への時事情報推薦システムの構築
Development of the Topical Information Offering System Using Twitter and TV

世良 拓也† 吉村 枝里子‡ 土屋 誠司‡ 渡部 広一‡
Takuya Sera Eriko Yoshimura Seiji Tsuchiya Hirokazu Watabe

1. はじめに

インターネットは近年急速に発展・普及しており，容易に気象情報やニュース記事などの時事情報を取得することができるようになった．しかし，Web 上の時事情報は日々大量に更新されているため，ユーザの興味がある時事情報のみを取得することは困難になっている．そこで，効率的に時事情報を取得する手段として，興味がある時事情報をコンピュータに推測、提供させることが考えられる．

Twitter[1]での発言(ツイート)には，ユーザの興味・嗜好に関する情報が含まれていると考えられる．また，テレビでは多数の番組が同時に放送されており，ユーザは見たいものを主に視聴していると考えられる．よって，ツイートとテレビの視聴履歴を収集し，分析することによってユーザ個人の嗜好の傾向を取得することができると考えられる．

本研究では Twitter でのツイートとテレビ視聴履歴を用いて，個人の嗜好を考慮した時事情報の推薦を行うシステムを構築する．

2. 関連技術

本研究では，嗜好情報を取得する際にオートフィードバック，嗜好情報と時事情報の特徴的な語との関連の強さを定量的に表す手法である関連度計算方式を使用する．

2.1 概念ベースと関連度計算方式

概念ベース[2]とは，複数の国語辞書や新聞などから機械的に構築した知識ベースである．ある語を概念とし，概念の意味特徴を表す語(属性)とその重要さを表す数値(重み)の対の集合によって定義している．概念数は約 8 万 7 千個存在する．

関連度計算方式とは，ある 2 つの概念間の関連の強さを定量的に表現する手法である．関連度は 0.0 から 1.0 までの実数値で表現され，関連が強いほど大きくなる．

2.2 オートフィードバック

オートフィードバック[3]とは，概念ベースにおける未定義語の意味的特長を表わす属性(単語)とその重要性を表わす重みの組を，Web を用いて取得する手法である．

2.3 Web-IDF

Web-IDF とは，Web 上にある文書のみを用いて索引語の特定性を考慮する手法であり，式 1 で定義される．

$$Web-IDF(t) = log_2 \frac{N}{df(t)} + 1 \quad (1)$$

なお，N を Google[4]が保有している日本語ページ数，$df(t)$を索引語 t の Google で検索を行った時のヒット件数としている．

3. 時事情報推薦システム

まず，Twitter でのツイートとテレビ視聴履歴を解析し，個人の嗜好情報を取得し，カテゴリ分類を行う．Web から時事情報を取得し，その中から特徴的な単語(以下，話題語とする)を選択する．その後，嗜好情報と話題語との重要度を算出し，高い順に時事情報を出力する．

本研究で扱う時事情報は，Web サイトに存在しているニュース記事の見出し文とする．提案システムの概要図を図 1 に示す．

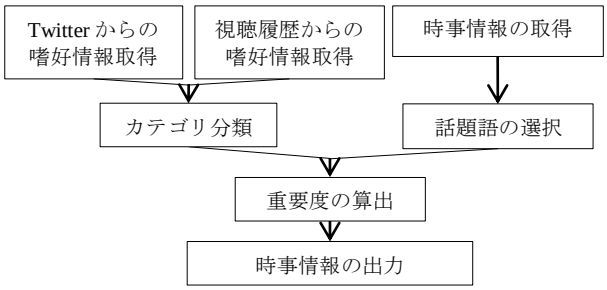

図 1　提案システム概要図

3.1 嗜好情報の取得

Twitter でのツイートと視聴したテレビ番組の紹介文から形態素解析によって名詞句を取得し，オートフィードバック結果を嗜好情報として扱う．取得した嗜好情報はグループ分けを行って統合し，グループごとに扱う．

3.1.1Twitter からの嗜好情報の取得

ユーザの Twitter におけるツイートを取得する．ツイートに含まれる形容詞に対して，あらかじめ用意したグループを決定する形容詞と表記一致を行い，印象ごとにツイートを 3 つのグループに分ける．グループ×となったツイートは，嗜好としてふさわしくないと考え排除する．表 1 にグループを示す．また，表記一致に使用した形容詞の一例を表 2 に示す．

表 1　ツイートのグループ分け

グループ◎	良い印象のツイート
グループ○	どちらでもない印象のツイート
グループ×	悪い印象のツイート

表 2　表記一致に使用した形容詞の一例

良い印象	かわいい，うれしい，よい，うまい…
悪い印象	わるい，うざい，きたない，ひどい…

取得したツイートから形態素解析によって名詞句を取得し，オートフィードバック結果を嗜好情報とする．

† 同志社大学大学院理工学研究科
Graduate School of Science and Engineering, Doshisha University
‡ 同志社大学　理工学部
Faculty of Science and Engineering , Doshisha University

3.1.2 テレビ視聴履歴からの嗜好情報の取得

ユーザから，3 つのグループに分けてテレビの視聴データを収集する．表 3 にグループを示す．

表 3　テレビ視聴履歴のグループ分け

グループ◎	集中して視聴していた(録画含む)
グループ○	視聴したかったができなかった
グループ△	なんとなく視聴していた

視聴データとして得られた番組のタイトル・番組紹介文から形態素解析によって名詞句を取得し，オートフィードバック結果を嗜好情報とする．

3.2 嗜好情報のカテゴリ分類

情報源として利用している Web サイトは，あらかじめニュースをカテゴリ分類している．そこで嗜好情報をカテゴリ分類する．使用したカテゴリを表 4 に示す．

表 4　使用したカテゴリ

政治	社会	国際	経済	スポーツ	音楽	芸能

嗜好情報は，カテゴリそれぞれとの関連度を算出し，関連度が閾値である 0.08 以上のカテゴリ上位 3 つに分類する．例を表 5 に示す．

表 5　カテゴリ分類の例(カテゴリ一部省略)

		カテゴリ				
		社会	国際	経済	スポーツ	音楽
嗜好情報	高校野球	0.185	0.120	0.088	0.589	0.216

「高校野球」は関連度 0.08 以上のカテゴリが存在するため，上位 3 つに分類される．

3.3 時事情報の取得

時事情報を新聞社等の Web サイトから取得する．

取得する際に，取得元のカテゴリに時事情報を分類する．カテゴリは嗜好情報の分類と同じものを使用する．

3.4 話題語の選択

取得した時事情報から形態素解析により名詞句を抽出し，Web-IDF 値が閾値である 3.0 以上の名詞句を話題語として取得する．例を表 6 に示す．

表 6　話題語の例

時事情報	話題語
ツーリングで不明の女性，救助	ツーリング

3.5 重要度の算出

同一カテゴリに分類された嗜好情報と時事情報の話題語の関連度をそれぞれ算出する．算出した関連度に対し，グループごとの重みを掛け合わせ，その値の総和を話題語の重要度とする．グループごとの重みは，◎を 2.0，○を 1.0，△を 0.8 としている．重要度の算出の例を図 2 に示す．

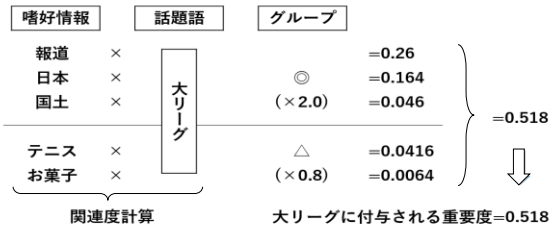

図 2　重要度算出の例

3.6 時事情報の出力

まず話題語を用いて時事情報に重要度を付与する．時事情報中に含まれている話題語を表記一致により検索し，その話題語の重要度を時事情報に付与する．

その後，重要度が高い順に時事情報を出力する．出力結果の例を表 7 に示す．

表 7　出力結果例

カテゴリ	時事情報	重要度
芸能	千原ジュニア:過去の女性と対面で衝撃…	45.3556

4.評価

2015 年 10 月 23 日における時事情報 451 件を使用した．また，3 名の被験者(20 代，男性，大学生)から収集した，17 日から 23 日の 7 日間のテレビ視聴履歴および，23 日における最新の Twitter のツイート 200 件を用いて評価を行なった．

システムより被験者ごとに提案された時事情報の，上位 30 件における興味がある時事情報の割合を正解率として，表 8 に示す．

表 8　評価結果

	被験者 A	被験者 B	被験者 C	平均
提案システムの正解率	50.5%	50.0%	60.0%	53.3%
全記事中の興味がある時事情報の割合	30.1%	32.2%	50.1%	37.5%

5 考察

平均すると，37.5%の割合で興味がある時事情報の中から，53.3%の精度で興味がある時事情報を推薦することができた．

嗜好情報が多く分類されたカテゴリでは，そのカテゴリの時事情報の話題語の重要度が高くなった．その結果，そのカテゴリの時事情報が上位に出力されやすくなった．

6 おわりに

本研究では，ユーザの Twitter でのツイートとテレビ視聴履歴を用いて嗜好を考慮し，ユーザが興味を持つと考えられる時事情報を推薦するシステムを構築した．その結果，37.5%の割合で興味がある時事情報の中から，平均で53.3%の精度で興味がある時事情報を推薦することができた．

謝辞

本研究の一部は，JSPS 科研費 16K00311 の助成を受けて行ったものです．

参考文献

[1] “Twitter”，https://twitter.com/
[2] 奥村紀之，土屋誠司，渡部広一，河岡司，”概念間の関連度計算のための大規模概念ベースの構築”，自然言語処理，Vol14, No.5，pp.41-64，2007
[3] 辻泰希，渡部広一，河岡司，”www を用いた概念ベースにない新概念およびその属性獲得手法”，人工知能学会全国大会，2D1-01，2003
[4]“Google”，http://www.google.co.jp/