

## パーソナライズ情報収集支援システムのための Twitter を利用した嗜好抽出の提案

## The User Preference Extraction Based on Twitter for Personalized Information Gathering system

○澤田瑞穂<sup>1</sup>, 戸田健<sup>2</sup>\*Mizuho Sawada<sup>1</sup>, Takeshi Toda<sup>2</sup>

Abstract: Even as the sheer volume of information on the Web continues to expand, not everyone is able to find the information they seek using the conventional search methods available. Here we propose a solution in the form of a system that first assesses a user's circumstances and preferences in real time based on micromedia comments posted to the Web, then infers and recommends information desired by the user. Using content analysis focused on objects and predicates used in micromedia comments posted on Twitter, blogs, and social network services (SNSs), the system builds a user profile by monitoring and storing information relating to the user's circumstances and experiences. Search terms are then extracted based on the user's profile, which provides personalized search and search results.

## 1. はじめに

WEB 上には多様な情報が膨大に蓄積されている。それとともに、ユーザによる WEB 上の情報の利用目的も多様になった。「検索する対象や目的といったゴールが明確である調査」の場合の検索は技術の発達によって容易なものになったが、WEB 上の情報や利用するユーザが多様化した事によって生まれた「ユーザにとって最善の結果の調査、明確なゴールのない新しい情報の発見」<sup>[1]</sup>を目的とする検索は、従来の検索方法では時間と労力がかかってしまう。

現在、WEB 上において様々な情報推薦技術の開発が進んでいる。各種商品購買サイトにおける商品の表示履歴や購入履歴から同様な購入履歴を持つユーザの履歴に残った商品を薦める、「商品推奨サービス」があるが、こうした履歴情報の利用を不安に思うユーザも多く、これらのサービスでは、過去のユーザの所望の情報から提供情報を選択し提供するために、情報の新規性が失われている場合が考えられる。

## 2. 提案システム

これまで我々は、ユーザにとって最善の検索を目的としたパーソナライズ情報収集支援システムを提案してきた<sup>[2]</sup>。提案システムでは、近年普及したソーシャルネットワークサービス(Social Network Service : SNS)や Blog, Twitter 等のマイクロメディアの発言からユーザの嗜好を抽出する。提案システムの概要としては、Twitter で「東京の海好き～」と発言した時、「好き」という述語の目的語である「東京、海」をユ

ーザの嗜好ワードとして蓄積する。このようなユーザプロフィールを蓄積する事で、ユーザに対応した新着情報を収集することが可能とする。また本システムでは、ユーザが気付かなかったような潜在的な興味が推測される事も考えられる。

## 3. 実験

実験の目的は、マイクロメディア上のユーザの発言からユーザプロフィールを再現できるかの検証である。本研究の情報収集支援システムにおいて、情報推薦の基となるユーザプロフィールは重要であり、高い再現率が求められる。

## 3.1 実験方法

本実験を実施する為に、被験者 8 名が Twitter 上で一ヶ月間発言を行い、事前実施した嗜好調査アンケートと比較をする。実施期間については、冬期(2 月中旬～3 月中旬)と夏期(6 月中旬～7 月中旬)に分けて一ヶ月ずつ二期行い、ユーザの嗜好や興味の遷移を測定した。被験者については、Twitter の利用に慣れている大学院生 8 名に自由に Twitter 上で発言してもらった。

## (1) 事前アンケート

被験者 1 人ひとりに[自分の好きな物], [嫌いな物], [興味のある物], [興味の無い物]の単語を挙げてもらい[大好き], [好き], [普通], [嫌い], [大嫌い]の五段階で評価してもらった。

## (2) Twiteer 発言の単語に対する重み付け

重み付けには以下の方法を用いてシステムの挙動を再現した。[大好き]…3, [好き]…2, [普通]…1, [嫌い]

1 : 日大理工・院 (前)・電気 2 : 日大理工・教員・電気

Table1. Example of ranking by the subject

実験 I 期(冬期)				
事前アンケート			ユーザ発言	
順位	嗜好度	単語	重み	単語
1	5	サカナクション	18	サカナクション
2	5	茅ヶ崎	8	茅ヶ崎
3	4	スノーボード	8	スノーボード
4	4	ビーフシチュー	8	ビーフシチュー
5	4	アイスクリーム	2	アイスクリーム
6	3	雪だるま	2	雪だるま
7	3	ダーツ	2	ダーツ
8	2	もつ鍋	0	卓球
9	2	満員電車	0	時計
10	1	定期試験	0	海
11	1	しいたけ	-1	もつ鍋
12	0	卓球	-1	定期試験
13	0	時計	-1	満員電車
14	0	海	-2	しいたけ

実験 II 期(夏期)				
事前アンケート			ユーザ発言	
順位	嗜好度	単語	重み	単語
1	5	サカナクション	18	ダーツ
2	5	ダーツ	8	サカナクション
3	5	卓球	8	卓球
4	4	時計	2	海
5	4	海	2	時計
6	1	もつ鍋	0	もつ鍋
7	1	しいたけ	0	茅ヶ崎
8	0	茅ヶ崎	0	スノーボード
9	0	スノーボード	0	ビーフシチュー
10	0	ビーフシチュー	0	アイスクリーム
11	0	アイスクリーム	0	雪だるま
12	0	雪だるま	0	定期試験
13	0	定期試験	0	満員電車
14	0	満員電車	-6	しいたけ

…-2, [大嫌い]…-3 と述語に点を与える. 発言中に上記述語群が含まれていたら, 目的語となる単語に述語に準ずる点数を与える. 一単語に対し, 点数の総数と出現回数を掛け合わせたものを重みの値とした.

### 3.2 実験結果

#### 3.2.1 嗜好の遷移の結果

嗜好の遷移については時期に関わらず嗜好の変化の無いもの, 時期によって嗜好に変化のあるものの二つのパターンに分けられる事が分かった. この傾向は被験者全員に見られた.

#### 3.2.2 ランキング結果の分析

ランキング結果の分析は, 次式で表されるランキングの再現率及び分散で評価する<sup>[3]</sup>.

$$Recall = \frac{|T_i^x \cap \mathfrak{S} L_i^L|}{|T_i^x|} \times 100[\%] \quad (1)$$

$$Variance = \frac{\sum_{i=1}^N |R_i^T - R_i^L|}{N} \quad (2)$$

但し, ランキングの再現率(1)式において,  $|T_i^x|$ はアンケートに出現した単語の総数,  $\mathfrak{S} L_i^L$ は Twitter 上の発言に出現した単語の写像  $L_i^L$ の像であり, ユーザ発言のランキングの全アイテムを示す. またランキングの分散(2)式において,  $N$ はランキングの総数,  $R_i^T$ は事前アンケートでのランキング,  $R_i^L$ は Twitter による重みのラ

ンキングである. これらの分析の結果, 第一期, 第二期ともに被験者全てに 100%の再現率が得られた. またランキングの分散については, 第一期では 1.7, 第二期では 0.8 のであった.

### 3.3 考察

実験結果より, Twitter 上の発言からほぼ正確に被験者の嗜好・興味を再現でき, 正確なユーザプロフィールに沿った情報推薦が出来ると考えられる.

## 4. まとめと今後

本実験結果から提案システムの効果が期待できることがわかった. 今後, 本提案システムの有用性を検討する.

## 5. 参考文献

- [1] J. Krones: "Exploring The Shift In Search Behaviors With Microsoft's Jacquelyn Krones" Search Engine land, <http://searchengineland.com/exploring-the-shift-in-search-behaviors-with-microsofts-jacquelyn-krones-85750> (2011)
- [2] Mizuho Sawada and Takeshi Toda : "Personalized Information Gathering System Based on Micromedia", Proc. First IEEE Int'l Conf. on Internet Operating Systems and New Applications (ICIOS-2012), Irvine, California, Dec. 2012.
- [3] 土方嘉徳:「特集 利用者の好みをとらえ活かす-情報抽出技術の最前線-1.嗜好抽出と情報推薦技術」, IPSJ Magazine, Vol48, No.9, Sep. 2007