

# 記憶想起支援のためのキーワード抽出手法の改善

渡部 秀太<sup>†</sup> 橋本 悠佑<sup>†</sup> 村上 晴美<sup>†</sup>

大阪市立大学大学院創造都市研究科<sup>†</sup>

## 1. はじめに

過去の記憶は時間が経つほど想起が困難であり、情報過多は日常生活において大きな問題となっている。そのような中、近年ではPIMやライフログへの関心が高まってきている。松本らはカレンダー、Twitterの履歴情報と撮影した写真の枚数を用いてタグクラウドを生成し、個人の記憶の想起を支援するシステムを開発した[1, 2]が、評価実験ではタグクラウドには改善の余地があることが分かった[2]。

本研究では、先行研究[1]のキーワードランキング手法を改善することによりタグクラウド及び記憶想起支援システムの改善を目指す。

本稿の構成は以下のとおりである。2節で先行研究、3節で改善手法、4節で実験について述べる。5節では改良したプロトタイプシステムについて述べる。

## 2. 先行研究

先行研究[1]では、ユーザーが日々利用するカレンダーとTwitterの履歴情報からキーワードを抽出し、タグクラウドで可視化を行う。キーワードのランキングは、印象的な日に関するキーワードに重みを付与するために、撮影した写真枚数に基づいている。

### 2.1 キーワード抽出手法

カレンダーとTwitterの履歴情報を、MeCabを用いて形態素解析を行い、不要語(例:「明日」、「今日」など)処理を行った後、名詞のみを抽出する。連続した名詞は複合名詞として結合する。抽出した語は日時情報とともに履歴構造と呼ぶファイルに格納する。

### 2.2 キーワードランキング手法

$$CalW(t) = \sum_{t \in HS} (1 + C_{photo}(G_{date}(t))) \quad (1)$$

$$TwiA(t) = \frac{1 + C_{photo}(G_{date}(t))}{C_{tweet}(G_{date}(t))} \quad (2)$$

$$TwiB(t) = C_{photo}(G_{time-ba}(t)) \quad (3)$$

$$TwiW(t) = \sum_{t \in HS} (TwiA(t) + TwiB(t)) \quad (4)$$

$$Weight(t) = CalW(t) + TwiW(t) \quad (5)$$

式(1), (4)は、それぞれカレンダー、Twitterのキーワードの重み付けであり、式(5)で合算する。 $G_{date}(t)$ は語  $t$  の日付情報を取得する。 $C_{photo}(R)$ は日付  $R$  における写真の枚数を取得する。 $C_{tweet}(R)$ は日付  $R$  におけるツイート数を取得し、 $G_{time-ba}(t)$ は語  $t$  の前後1時間の日時情報を取得する。 $HS$ はキーワードと日時情報を格納しているファイルを意味する。全ての語の重みを比べ、上位30位までの語をタグクラウドに表示する。

## 3. 改善手法

先行研究では、写真枚数が多い日のキーワードは重要でなくても重みが高くなり、写真を撮影していない日のキーワードは重要でも重みが低くなるという問題点がある。そこで本研究では、新たな重みを付与することでキーワードランキングを改善する。

### 3.1 新たな重みの種類

重要な語として以下の3種類を検討する。

1. カレンダーとTwitterの両方に出現する語 (情報源)
  2. 固有名詞・地域に関する語 (地域)
  3. 固有名詞・人名に関する語 (人名)
- 固有名詞に関してはMeCab (IPA辞書) を利用して判定する。

### 3.2 改善手法の計算方法

新たな重みとして先行研究の上位30件の平均点を加える。新たな重みの計算方法の例を図1に示す。

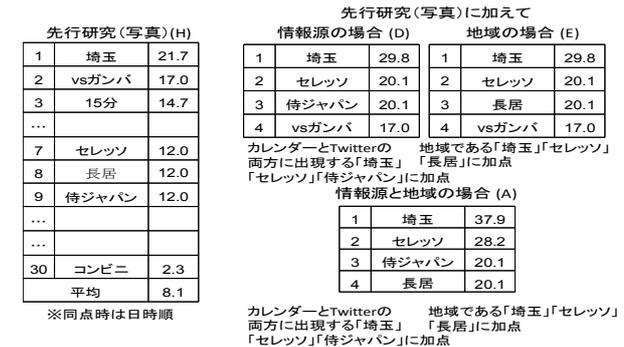


図1: 改善手法の計算例

先行研究の上位30件の平均点8.1を加点する。例えば、情報源の場合、カレンダーとTwitterの両方に「埼玉」「セレッソ」「待ジャパン」が出現するので両方に加点する。地域の場合は地域である「埼玉」「セレッソ」「長居」に加点する。情報源と地域の場合は「埼玉」「セレッソ」に $8.1 \times 2$ 回=16.2点、「待ジャパン」「長居」に8.1点を加点する。なお図中のA等の記号については4節で述べる。

Improvement of Keyword Weighting Method for Human Memory Recollection  
<sup>†</sup> Shuta Watanabe, Yusuke Hashimoto, Harumi Murakami, Graduate School for Creative Cities, Osaka City University

## 4. 実験

改善手法の有効性を評価するために実験を行った。被験者は20代の男性7名である。被験者は全て、Twitterを使用し、予定をメモ帳やカレンダーやスマートフォンのアプリなどで管理しており、スマートフォンやカメラなどで写真を撮影している。扱うデータは、実験月の一か月前から一年前（十二か月前）までの間の一か月前である。

新しい重みを組み合わせた改善手法（A～G）、写真のみを用いた先行研究（H）、写真を使わない単純頻度[1]の比較を行う。組み合わせの種類を表1に示す。

表 1：手法の種類

	A	B	C	D	E	F	G	H	I
情報源	○	○		○			○		
地域	○		○		○		○		
人名		○	○			○	○		
写真	○	○	○	○	○	○	○	○	

### 4.1 実験1

手法A～Iの重み付けアルゴリズムの比較を行った。各手法から上位30個の語を抽出し、被験者に提示する。「記載されている単語は記憶の想起に役立ちますか」と質問し、これらの語が記憶の想起に有効かを5段階で評価させた。

表 2：実験1の結果

	A	B	C	D	E	F	G	H	I
平均	2.47	2.03	2.48	2.70	2.64	2.59	2.49	2.69	2.69

結果を表2に示す。手法Dがわずかに先行研究(H)を上回った。

### 4.2 実験2

手法A～Iで作成したタグクラウドの比較を行った。「全てのタグクラウドの中でどれが一番記憶の想起に役立つと思いますか」という質問において、被験者7人のうち5人がDを選択し、1人がBを、1人がCを選択した。

「地名や人名に関するキーワードは記憶の想起に役立つ」というコメントが複数あった。また、「その月に行ったイベントが多く含まれている方が記憶想起しやすい」という意見があった。

### 4.3 考察

実験結果から新たな重みが先行研究のタグクラウドを改善することが示唆された。現時点では先行研究に「複数の情報源に出現する語」のみ重みを加える手法が最も良かった。地域と人名は記憶の想起に有効であるが、ゴミも増えるために単純に重みを付与すればよいわけではないことも示唆された。

## 5. プロトタイプ

改善したランキング（ここではD）を用いて、記憶の想起を支援する2種類のプロトタイプシステムを実装した。図2に改善したタグクラウドのシステム、図3にネットワークのシステムのネットワーク部を示す。タグクラウドシステムでは、重みはタグの大きさで表現し、日時順に並べている。複数の情報源に出現した語は赤、カレンダーのみは緑、Twitterのみは青色としている。

ネットワークのシステムではタグクラウドと同じキーワードを履歴構造中の共起に基づきネットワークで視覚化している。また、改善したランキング上位3件のタグクラウドを見出しとして付与している。



図 2：改善したタグクラウドのシステム

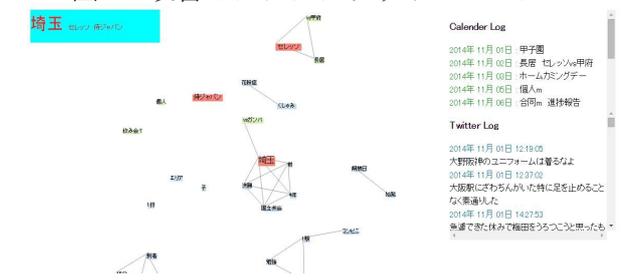


図 3：ネットワークのシステム

## 6. おわりに

記憶想起支援のためのキーワードランキング手法の改善を行い、システムの改良を行った。今後も被験者を増やして実験を継続予定である。

### 参考文献

- [1] Matsumoto, M., Matsuura, S., Mitsuhashi, K., Muarkami, H.: Supporting Human Recollection of the Impressive Events, Using the Number of Photos, *ICAART 2014 - Proceedings of the 6th International Conference on Agents and Artificial Intelligence*, Vol.1, pp.538-543 (2014).
- [2] 松浦 翔, 松本若樹, 村上晴美, タグクラウドを用いた記憶の想起支援, 情報処理学会第77回全国大会講演論文集 (2015).