

写真の枚数を用いた記憶の想起支援

松本 若樹<sup>†</sup> 松浦 翔<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

人の記憶とは不確かなものであり、過去の記憶は時間が経てば経つほど思い出せないものである。一方、インターネットの普及により個人が扱う情報は大幅に増え続けており、これらの情報全てを人が管理するのは非常に困難である。

このような背景もあり、近年住所録やスケジュールなどといった個人情報を管理する PIM や、個人の生活を記録するライフログの研究が多く存在する。

インターネットの普及によって、人々は web 上に日々の活動を記入するようになった。例えば、手帳に書いていた予定を Google Calendar に記入したり、身近で起こった出来事を SNS で報告したりする。また、デジタルカメラやカメラ付き携帯の普及、さらに記憶媒体の大容量化や低価格化に伴い、誰でも簡単に写真を撮り、それらを大量に保存できるようになった。写真は記録のために撮られることが多く、記憶を想起する上で重要なデータであると考えた。

本研究では、web 上にあげられた情報 (Google Calendar, Twitter) と写真の枚数に着目し、記憶の想起を支援するシステムを構築した。

本稿の構成は以下のとおりである。2 節で提案手法、3 節で実験について述べる。4 節では本研究と関連研究の比較を行う。

2. 提案手法

記憶の想起を支援するために、Google Calendar と Twitter からテキストと日時のデータを抽出し、タグクラウドを用いて可視化を行う。

本研究の特徴として、ユーザが撮影した写真の枚数に着目し、よく写真を撮ったより印象的な日の記憶を想起させる。図 1 は提案手法で生成したタグクラウドである。このタグクラウドは第一著者の 1 か月間のデータから生成されたものである。この期間に伊勢と福岡に行ってお

り、これらに関連する語が大きく表示されている。



図 1: 提案手法で生成したタグクラウド

2.1 キーワード抽出アルゴリズム

Google Calendar のイベントタイトルと Twitter のツイートを、Mecab を用いて形態素解析を行い、不要語 (例: 「明日」, 「今日」など 11 語) 除去後、名詞を抽出する。連続した名詞を結合する。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 タグ選択アルゴリズム

タグクラウドにおけるタグ選択アルゴリズムは一般的に、単語の出現頻度  $f$  やコンテンツの人気度によって計算される。本研究では、単語の頻度に写真の枚数を考慮する新たな計算手法を考案した。語の重みづけは情報源によって異なり、下記に示す。

(1) カレンダー

カレンダーの語  $t$  における語の重みを式 1 に示す。

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

式 1 において、 $G_{date}(t)$  は語  $t$  の日付情報を取得し、 $C_{photo}(R)$  は範囲  $R$  における写真の枚数を取得する関数である。また、 $HS$  はキーワードと日時情報を格納している履歴構造ファイルを意味する。カレンダーの語は写真の枚数が増えるほど、重みが大きくなる。

(2) Twitter

Twitter の語  $t$  における語の重みを下記に示す。

Supporting Human Recollection Using the Number of Photos  
<sup>†</sup> Masaki Matsumoto, Sho Matsuura, Harumi Murakami,  
 Graduate School for Creative Cities, Osaka City University

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

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

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

式 2 において、 $C_{tweet}(R)$ は範囲  $R$  におけるツイート数を取得する関数である。 $TwiA(t)$ は一日の写真の枚数を一日のツイート数で割ったものを重みとしている。これは、ツイート数の増加に比例して増える記憶の想起に役に立たない語の重みの増加をおさえるためである。

式 3 において、 $G_{time-ba}(t)$ は語  $t$  の前後 1 時間（合計 2 時間）の日時情報を取得する関数である。 $TwiB(t)$ はツイート時間の前後 1 時間の写真の枚数を重みとしている。Twitter はリアルタイムに情報を発信できるツールであるため、ツイート時間と近い時間に撮影された写真はそのツイートと関連が高いと考えている。

上記の  $TwiA(t)$ と  $TwiB(t)$ を足し、Twitter の語の重み  $TwiW(t)$ を算出する(式 4)。

最終的な語の重みを式 5 に示す。

$$Weight(t) = CalW(t) + TwiW(t). \quad (3)$$

### 2.3 タグクラウド表示

上位 30 語のタグクラウドを出現日時順に表示する。語の重みによって大きさを変える。二つの情報源に出現した語は赤、カレンダーのみに出現した語はオレンジ、Twitter のみに出現した語はうすい紫とした。

## 3. 実験

被験者 5 名に対して、1 か月前の記憶の想起における実験を行った。

本実験では、タグ選択アルゴリズムを評価するために、提案手法と比較手法における比較実験を行った。比較手法では単語の出現頻度  $f$  を用いている(式 6)。

$$Com(t) = \sum_{t \in HS} f(t). \quad (6)$$

二つの手法から上位各 30 個の語を抽出し、マージして重複を省いて文字コード順にソートした状態で被験者に提示する。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。結果を表 1 に、被験者毎のデータセットを表 2 に示す。

P30 と C30（数字は各手法で計算された重みの上位単語数を表す）の結果から提案手法が  $f$  を用いた手法よりも、記憶の想起という点において有効であることがわかる。

ただし、被験者 1, 3 に関して差がほとんどなく、被験者 4 に関しては結果が同じとなってい

る。これらの結果に関して、被験者 1, 3, 4 はデータ量が少なかったことが原因の一つとして考えられる。例えば、結果の悪い被験者 3 に関して、二つの手法から取得されたキーワードの内、27 個が同じものであった。一方で結果の良かった被験者 2 においては、重複キーワード数が 10 個であった。

表 1：実験 1 の結果

	P5	C5	P10	C10	P30	C30
被験者 1	4.0	3.3	3.2	3.8	3.4	3.3
被験者 2	4.2	3.2	3.1	2.7	3.0	2.5
被験者 3	4.2	4.2	4.2	2.7	2.4	2.3
被験者 4	2.6	3.8	2.6	3.2	2.6	2.6
被験者 5	4.0	4.0	4.2	2.9	3.9	2.4
平均	3.8	3.7	3.5	3.1	3.1	2.6

(注)P は提案手法、C は比較手法を表す。

表 2：データセット

	キーワード数	カレンダーのみ	Twitterのみ	写真
被験者 1	69	13	57	36
被験者 2	467	8	462	73
被験者 3	33	7	28	19
被験者 4	65	25	50	41
被験者 5	748	15	735	32
平均	276.4	13.6	266.4	40.2

## 4. 関連研究

個人の記憶の想起を支援するために、ユーザのサービスの利用履歴を用いた研究が行われている。Murakami ら[1]は多様な情報源から抽出したキーワードをネットワーク図で可視化している。本研究では情報源を二つに絞り、タグクラウドを用いて可視化を行っている。

タグクラウドに関する研究は多くある[2, 3]が、我々が調べたところ、個人の記憶の可視化にタグクラウドを用いた研究はなかった。

## 5. おわりに

ユーザが撮影した写真の枚数に着目し、タグクラウドにおけるタグ選択アルゴリズムを新たに考案した。被験者 5 名による評価実験では、世の中で多く使われている  $f$  を用いた手法よりも記憶の想起という観点で良い結果が得られた。現在評価実験を継続中である。

## 参考文献

- [1] Murakami, H. and Mitsuhashi, K.: A System for Creating User's Knowledge Space from Various Information Usages to Support Human Recollection, *International Journal of Advancements in Computing Technology*, Vol.4, No.22, pp.496-508 (2011).
- [2] Rivadeneira, A. W., Gruen, D. M., Muller, M. J. and Millen, D. R.: Getting Our Head in the Clouds: Toward Evaluation Studies of Tagclouds, *Proc. CHI'07*, pp.995-998, ACM (2007).
- [3] Venetis, P., Koutrika, G. and Garcia-Molina, H.: On the Selection of Tags for Tag Clouds, *Proc. WSDM'11*, pp.835-844, ACM (2011).