

タグクラウドを用いた記憶の想起支援

松浦 翔[†] 松本 若樹[‡] 村上 晴美[†]

大阪市立大学大学院創造都市研究科[†]

シーティーシー・テクノロジー株式会社[‡]

1. はじめに

近年、PIM やライフログへの関心が増加し、多くの研究で多種多様な個人の情報を活用するアイデアが提示されている。このような背景の中、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]。

本研究では、[1]の提案手法及びシステムの有効性を、5つの評価実験により評価した。実験1ではタグ選択アルゴリズム、実験2では提案手法で作成したタグクラウド、実験3では時間情報を用いたタグのソート方法の有効性の確認を行った。実験4ではタグの重み付けのパラメータの変更実験、実験5ではシステム全体の使用実験を行った。

本稿の構成は以下のとおりである。2節で[1]の提案手法及びシステムについて、3節で実験について述べる。4節では実験の結果からの考察について述べる。

2. 先行研究

[1]では、記憶の想起を支援するために、Google Calendar (以下カレンダー) 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。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。

2.1 キーワード抽出アルゴリズム

カレンダー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) 除去後、名詞を抽出する。連続した名詞は結合する (複合名詞)。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

一般的に、タグクラウドにおけるタグ選択アルゴリズムは単語の出現頻度 tf やコンテンツの人気度で計算される。[1]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 (式5)。語の重みづけは情報源によって異なる。

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

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

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

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

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

式1, 4ではそれぞれ、カレンダーの語 t の重み、Twitter の語 t の重みを計算する。 $G_{date}(t)$ は語 t の日付情報を取得し、 $C_{photo}(R)$ は日付 R における写真の枚数を取得する関数である。 $C_{tweet}(R)$ 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ファイルを意味する。すべ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. 実験

提案手法及びシステムの有効性を評価するために5つの実験を行った。実験1~4は19歳から26歳の男性10名、実験5は8名を対象としている。被験者は全て、Twitter を使用し、予定をメモ帳やカレンダーなどに記載し、携帯電話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) までの中の一か月分である。

3.1 実験1

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を比較した。比較手法では単語の出現頻度 tf を用いている (式6)。

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

二つの手法から上位各30個の語を抽出し、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。

提案手法では平均2.6、比較手法では平均2.2であり、提案手法の方が良かった。

3.2 実験2

二つの手法から6つのタグクラウド (A-F) を生成し被験者に見せ、記憶の想起に有効だと思うものを選択させた。それぞれのタグクラウドの違いを表1に示す。

Supporting Human Memory using Tag Clouds

[†]Sho Matsuura, Harumi Murakami, Graduate School for Creative Cities, Osaka City University

[‡]Masaki Matsumoto, CTC TECHNOLOGY Corporation

表 1: 6つのタグクラウド

	情報源		重みづけ
A	カレンダー	Twitter	提案手法
B	カレンダー		提案手法
C		Twitter	提案手法
D	カレンダー	Twitter	比較手法
E	カレンダー		比較手法
F		Twitter	比較手法

10人中8人がカレンダー, Twitter, 提案手法の組み合わせで作成されたタグクラウド A を選択した。

3.3 実験 3

異なる 3 つのソート方法 (A: 日時順, B: 文字コード順, C: 単語の重み順) で生成されたタグクラウドを見せ, 記憶の想起に有効だと思うものを選択させた。

10人中9人が日時順で生成された A を選択した。

実験 1-3 より, 提案手法で作成したタグクラウドの有効性を確認した。

3.4 実験 4

カレンダーと Twitter の語の重みの係数を変更 (式 7) して生成されたタグクラウドを見せ, 記憶の想起に有効だと思うものを選択させた。結果を表 2 に示す。

$$Weight(t) = \alpha \times CalW(t) + (1 - \alpha) \times TwiW(t) \quad (7)$$

$$(0 \leq \alpha \leq 1)$$

表 2: カレンダーと Twitter の重みの割合

被験者	1	2	3	4	5
α の値	0.5	0.6	0.5	0.9	0.9
被験者	6	7	8	9	10
α の値	0.5	0.5	0.6	0.2	0.9

α の値の平均は 0.6 であり, 0.2 から 0.9 までのばらつきがあった。この結果から, 被験者によってより良いタグクラウドが得られる α の値は異なることがわかった。

3.5 実験 5

図 1 は第一著者の 2014 年 8 月のデータを用いた記憶想起支援システムの表示例である。第一著者は, この期間に友達と大阪観光をし, 多くの写真を撮影した。図 1 では, このことに関する語 (難波, 海遊館, 串カツなど) が大きく表示されている。カレンダーの語は緑色, Twitter の語は青色, 両方にある語は赤色で表示される。タグクラウド内の語をクリックすると, その語が使われたツイート, 予定を検索することができる。また, システムには α の値を変更することができるスライダーを付けている。

システムを実際に操作させ, 質問紙, インタ

ビュー, 観察による評価を行った。

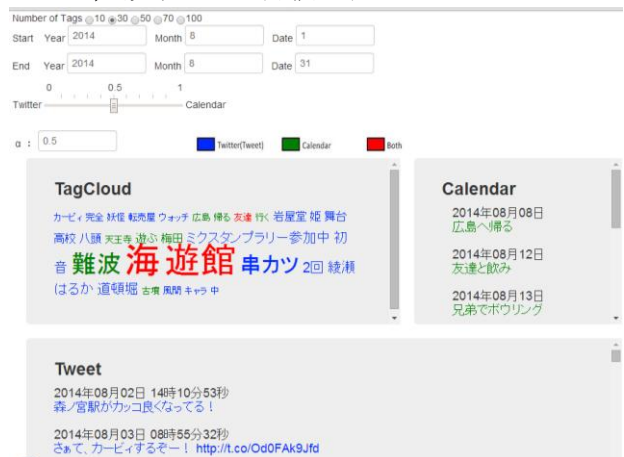


図 1: 記憶想起支援システム

「このシステムが記憶の想起に役に立ちましたか」という質問への 5 段階評価の回答結果を表 3 に示す。

表 3: システムの有効性

被験者	1	2	3	4
評価	5	5	4	4
被験者	5	6	7	8
評価	5	5	5	5

全ての被験者が 5 (非常に役に立った), または 4 (やや役に立った) と回答し, 評価の平均も 4.8 と高かった。この結果から, 本システムは記憶の想起に有効であることがわかった。

4. 考察

実験 1~5 の結果から [1] の提案手法, システムは記憶の想起に有効であることがわかった。しかし, タグクラウドの中にゴミが目立つなどのコメントもあった。カレンダーの予定の量, ツイート数, 写真枚数の偏り方によってタグクラウドの評価が変わるため, それらを踏まえ, より良い重みづけ手法の考案が必要である。

5. おわりに

提案手法とシステム [1] を 5 つの実験により評価し, 記憶の想起に有効であることを確認した。今後はタグクラウド以外の視覚化方法 (先行研究 [2] など) との比較実験を行う予定である。

関連研究

- [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] 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).