

社会分析を目的としたCGM画像クラスタリング手法に関する一検討

上條 哲也[†] 豊田 正史^{††}

[†] 東京大学大学院 情報理工学系研究科 〒113-0033 東京都文京区本郷 7-3-1

^{††} 東京大学生産技術研究所 〒153-8504 東京都目黒区駒場 4-6-1

E-mail: †{kamijo,toyoda}@tkl.iis.u-tokyo.ac.jp

あらまし CGMにおいてはイベントやニュースなどその時々話題が数多く取り上げられ、社会分析のための重要な情報源となっている。CGM上には話題に関連する画像が多数存在し、それらを話題中の事象に基づいて分類し組織化することができれば、人物や商品などの画像数および内容の変化、事象の継続期間を視覚的に把握することができる。本論文では、CGM画像を、画像の周辺テキスト、局所特徴量、および時間情報を用いてクラスタリングしランキングする手法を提案する。本手法は検索クエリから抽出した画像をランキングした結果に基づき、適合度の高い画像クラスタを抽出する。提案手法の有効性を示すため、5年間にわたって収集されたブログアーカイブを用いた評価実験を行う。

キーワード 画像クラスタリング、CGM、画像ランキング

1. はじめに

Web上にブログなどのCGM(Consumer Generated Media)が登場し、ユーザーは誰もが自由に商品やニュースに対する本音や意見の書き込みが可能となった。ユーザーが記述するブログの記事にはイベントやニュース、商品などその時々話題やイベントが多数取り上げられ、マーケティング調査や社会分析のための情報源となっている。CGM上には話題に関連する画像が多数存在する。ブログなどにアップロードされた画像はユーザーが愛着を持って撮影したものが多く、CMや商品に対する宣伝や、ニュースや広告などの効果を測定するための重要な情報となりうる。

本論文では、画像を用いて宣伝効果や話題の遷移の調査を可能にするため、ブログの画像を話題やイベントごとにクラスタリングし、ランキングして提示する手法を提案する。ブログ画像を分類すると以下のような分析が可能となる。

- 商品の画像が含まれるクラスタにおける画像数やその商品画像の変化から、人気の度合いやデザインの変化を把握できる。
- ユーザーが注目しているニュースがどのような映像で放映されていたかが分かる。
- イベントや集会などに参加した人の撮った写真からそれらの様子を視覚的に把握できる。

ブログ上のテキストを用いてキーワードの変遷や宣伝効果を統計分析する研究は行われているが、画像データを用いた研究はあまり行われていない。本手法は我々が5年間にわたって収

集したブログアーカイブを用い、ユーザーから与えられたクエリを用いて記事と画像を抽出したうえで、画像をテキスト情報、画像情報および時間情報を用いてを分類し、適合度の高い画像クラスタを抽出する。クラスタリング手法は階層クラスタリングを用いて行う。各クラスタ内の画像について、画像間の類似度を基にしたグラフ上のランダムウォークにより重みを算出し、画像クラスタのランキングを行う。本手法の有効性を示すため、テレビCMや商品、観光名所、社会問題に関するクエリを用いて、評価実験を行う。クラスタリングの精度は人手で分類した正解セットを用いてF値で評価し、画像クラスタのランキングの精度はNDCGを用いて評価を行う。

以降では2章で関連研究、3章で画像のクラスタリング手法を述べ、4章で画像クラスタのランキング手法を述べる。5章で評価実験を述べ、6章で本稿のまとめと今後の課題について述べる。

2. 関連研究

Web上に膨大な画像データがあふれ、これらを組織化することが急務となっている。画像を組織化するための研究として、メタデータを用いた手法、クラスタリング手法、再ランキング手法など様々な先行研究があり、これらの手法を用いた研究をいくつか紹介する。

Yahooの画像検索エンジンであるFlickr [2] ではユーザーが誰もが自由にテキストデータを付加させることができるタグが存在する。また画像には写真の撮られた位置情報を記録するジオタグや時間情報などのメタデータが存在し、これらを利

用した研究 [1], [10], [12], [16], [17], [11] が盛んに行われている。Liu [1] らは Flickr [2] 上のタグを画像の内容と関連の高い順に並べる Tag Ranking 手法を提案している。Crandall [11] らは位置情報が付与された画像を画像情報、テキスト情報、時間情報を学習し写真の取られた位置を推定する手法を提案している。

Leuken [7] らは Flickr 上の画像を Folding, Maxmin, Reciprocal Election 手法を用いてクラスタリングする手法を提案している。これらの手法は k-means や階層クラスタリングと比べアルゴリズムが単純であるが、計算コストが軽いという特徴がある。Kennedy [13] らは建物や風景などのランドマーク画像をクラスタリングしそのクラスをランキングする手法を提案している。クラスタリング手法は色や形状といった大域特徴量 [8], [9] を使い、K-means クラスタリングを行う。クラスターのランキングは多数のユーザーが存在するクラスは重要であるというユーザー情報、クラスター内の画像類似度が高ければ重要であるという画像情報、クラスター内の画像の時間情報が集中していれば重要であるという時間情報を用いている。

周辺テキストと画像特徴量を使った再ランキングを研究している論文がある。Cui [15] らは Flickr から画像およびその画像に付与されたタイトル、タグ、コメント、ユーザー情報を抽出し、画像の再ランキング手法を提案している。提案手法では画像情報、テキスト情報およびユーザー情報をノードとし、FIG (Feature Interaction Graph) を構築し、マルコフランダムフィールドを用いて画像類似度を算出している。ランキングの精度はテキスト情報、画像情報、ユーザー情報のどちらか一方を用いる場合より、全情報を用いた場合が最も高かった。Yao [14] らは Bing の画像検索から画像、周辺テキストおよび画像とテキストのランキング情報を抽出し、画像の再ランキング手法を提案している。画像の類似度は周辺テキストの頻度 (tf) および SIFT の画像特徴量を使用している。画像のランキングはテキスト情報を基に定義したランダムウォーク、画像情報を基に定義したランダムウォークを融合しランキングデータを算出している。

CGM では画像データを用いた研究はあまりされていない。CGM 上のデータはその時々話題が順次アップロードされ、変化が激しいことから画像の組織化が難しくなっている。

3. ブログ画像のクラスタリング手法

ブログ画像のクラスタリング手法は画像間の類似度行列から算出できる階層クラスタリングを用いる。クラスタリングに必要な画像類似度は画像情報、テキスト情報、時間情報を用いる。画像情報は SIFT を利用した局所特徴量を用い、テキスト情報は画像に属するブログ記事の前後 3 行の文を抽出し形態

素解析したキーワードを重みとした特徴量を用いる。また、時間が離れている画像は関連性が低下すると仮定し、両特徴量に時間の減衰係数を導入する。以降ではクエリ q としたとき、抽出したブログの記事集合を $B(q) = (b_1, \dots, b_n)$ と表し、抽出したブログ記事の画像集合を $X(q) = (x_1, \dots, x_m)$ と表す。また、画像 x に付与されたキーワードを $x = (w_1, \dots, w_i)$ と表す。

3.1 画像類似度

クラスタリングを行う際の画像類似度としては画像情報、テキスト情報および時間情報を用いた。出現した時間が近く、類似度の高いものに重みを与えるため、画像の出現時間が離れたものについては類似度にペナルティを付加する。

画像情報としては SIFT を用いた局所特徴量を利用しており、画像 x_i の特徴点の集合を A 、画像 x_j の特徴点の集合を B としたときの類似度を以下のように定義する。

$$sift(A, B) = \frac{m(A, B)}{(|A| + |B|)/2} \cdot e^{-\alpha t} \quad (1)$$

ただし、 $m(A, B)$ はマッチした特徴点の数を表す。 α は画像情報の時間の減衰係数を表し、 t は画像 x_i と画像 x_j の時間の差とする。

テキスト情報は画像に属するブログ記事の前後 3 行の単語を形態素解析した単語を重みとして定義する。画像 x に付与された単語 w_i の重みは

$$tfidf(w_i) = tf(w_i) * \log \frac{|X(q)|}{df(w_i)} \quad (2)$$

とする。 $|X(q)|$ はクエリ q から得られた画像数、 $tf(w_i)$ は画像 x の中に属する単語 w_i の出現頻度、 $df(w_i)$ は画像集合 $X(q)$ の中で単語 w_i を含む画像数を表す。

テキスト情報を用いた類似度は式 (2) の tfidf を単語の重みとしたコサイン類似度を利用しており、画像 x_i と画像 x_j の類似度を以下のように定義する。

$$text(x_i, x_j) = \cos(x_i, x_j) \cdot e^{-\beta t} \quad (3)$$

β はテキスト情報の時間の減衰係数を表し、 t は画像 x_i と画像 x_j の時間の差とする。

両特徴量は以下のように線形結合し、画像のランキングおよびクラスタリングのための類似度として算出する。画像 x_i と画像 x_j の類似度を以下のように定義する。 γ はテキスト情報と画像情報を結合するための混合率を表し (0,1) のパラメータとする。

$$s(x_i, x_j) = \gamma \cdot sift(x_i, x_j) + (1 - \gamma) \cdot text(x_i, x_j) \quad (4)$$

3.2 画像クラスタリング

本研究では画像の類似度行列からクラスタリングを行う階層

クラスタリング手法を用い、検索クエリから得られた画像を分割する。階層クラスタリング手法は M 個の画像が与えられたとき、1 個の画像だけを含む M 個の画像クラスタがある初期状態を作り、距離が最も近い二つの画像クラスタを逐次的に併合する手法である。以降では得られた N 個の画像クラスタを $C = (c_1, \dots, c_N)$ とする。ここで c_i は 1 つ以上の画像を含む画像クラスタを表す。本研究では、二つの画像クラスタ間の距離を、画像クラスタ c_i に属する画像 x_i と、画像クラスタ c_j に属する画像 x_j としたとき、 x_i と x_j の最小距離を画像クラスタ c_i と c_j との距離とする最短距離法 (Single linkage method)、 x_i と x_j の最長距離を画像クラスタ c_i と c_j との距離とする最長距離法 (Complete linkage method)、全ての x_i と x_j の組み合わせに対し類似度を求め、その平均を距離とする群平均法 (Group average method) の三つの手法を用いて比較する。

4. 画像クラスタのランキング

ブログ画像にはアフィリエイトを目的とした商品画像などが多くノイズが多く含まれている。そのため 3 節のクラスタリング結果から適合度の高い画像クラスタを抽出する。はじめに画像の重みを定義するため、3.1 節で定義した画像類似度を基にしたグラフのランダムウォーク [3] [6] を行う。次に画像クラスタ内に存在する画像の重みを利用し、画像クラスタのランキングを行う。

4.1 画像ランキング

ブログ画像のランキングは、[6] で定義される Visual Rank (VR) を改良し、画像類似度を 3.1 節で紹介した画像類似度に置き換えた Visual Text Rank (VTR) を用いて算出する。提案手法では、画像をノード、エッジの重みを画像類似度としたグラフを構築し、ランダムウォークを用いて画像の重みを算出する。ランダムウォークにより得られる画像の重みは以下のように定義する。

$$VTR = dS * VTR + (1 - d)R \quad (5)$$

VTR は各画像の重みを表すベクトルである。類似度行列 S は画像類似度を列で正規化した行列である。 d は (0,1) の値をとる減衰係数を表し、 R はランダムジャンプベクトルである。ランダムジャンプベクトルは画像 x_i に属する記事 B とした時、テキストの特徴を用いて以下のように定義する。

$$R_{x_i} = \frac{1}{|x_i|} \sum_{w \in b} \frac{tf(w) * idf(w)}{\log d(w, x_i)} \quad (6)$$

$|x_i|$ は画像 x_i に属する単語数を表す。 $tf(w)$ は画像に属する単語 w の出現頻度を表し、 $idf(w)$ は 5.1 節で紹介するブログアーカイブ全体の単語 w の idf を表す。 $d(w, x_i)$ は画像 x_i を

原点とし、原点から形態素解析で得られた単語 w までの距離を表す。これはアップロードされた画像に近い単語ほど重要度が高いとするためである。

4.2 画像クラスタのランキング

画像クラスタのランキングは 4.1 節で紹介した Visual Text Rank (VTR)、およびクラスタに属する画像類似度を用いる。VTR を利用した画像クラスタのスコアは、クラスタに属する画像のスコアの総和とする。VTR によって算出された画像 x の重みを $VTR(x)$ としたとき、画像クラスタのスコアを以下のように定義する。

$$C_{score}(c_i) = \sum_{x \in c_i} \log(VTR(x) + 1) \quad (7)$$

また、式 (7) を応用し、クラスタに属する画像類似度を用いる手法を提案する。画像クラスタに属する画像間の類似度が高い画像クラスタは重要な画像クラスタであると考えられる。そのため画像クラスタのスコアを、画像クラスタに属する画像間の類似度の平均値を導入し以下のように定義した。

$$C_{score}^+(c_i) = \delta \cdot \frac{C_{score}(c_i)}{\sum_{c_j \in C} C_{score}(c_j)} + (1 - \delta) \cdot \frac{\mu(c_i)}{\sum_{c_k \in C} \mu(c_k)} \quad (8)$$

ここで δ は (0,1) のパラメータとする。 $\mu(c_i)$ は画像クラスタ c_i に属する画像間の類似度の平均値とする。

5. 評価実験

5.1 データセット

本研究の有効性を示すため、表 1 で示した検索クエリを用いて実験を行う。マーケティングに関連するクエリとして「アフラック」、「ソフトバンク」、「オトナグリコ」、「マスキングテープ」、「キットカット」、「ヒルサイドテラス」、を用い、政治、社会問題、外交に関するクエリとして「鳩山由紀夫」、「上海万博」、「派遣村」、「尖閣諸島」、「村木厚子」を用いた。ブログデータとしては我々が 2006 年から収集しているブログアーカイブを利用する。ブログ記事集合は 2006 年 2 月から約 100 万の RSS, ATOM フィールドを毎日収集したもので、その中から検索クエリに対応する記事を取得し、記事中に含まれる画像を抽出した。

5.2 評価手法

5.2.1 画像類似度の評価

画像類似度の評価は人手でラベル付けした正解データを用いて行った。クエリに関する画像をサンプリングし、画像に関連の深いと考えられる K 個の話題に分類した。分類したラベルを $L = (l_1, \dots, l_K)$ とする。ただし、各ラベルには 1 つ以上の画像が含まれており、ラベルの重複は許していない。ラベル数およびラベル付けした画像数は表 1 にまとめた。評価方法は各手法

表 1 評価に用いたクエリの詳細

クエリ	記事数	画像数	期間	評価に用いたラベル数	ラベル付けした画像数
ソフトバンク AND お父さん AND CM	3504	2591	2007/6~ 2009/12	12	91
アフラック AND CM OR まねきねこダック	1909	1324	2009/8~ 2009/12	10	136
おとなグリコ OR オトナグリコ OR "otona grico"	862	487	2007/9~ 2009/12	9	77
マスキングテープ	456	980	2005/8~ 2010/06	8	64
ヒルサイドテラス	1502	3670	2005/5~ 2010/07	9	65
派遣村	16497	6542	2008/12~ 2010/06	11	95
鳩山由紀夫	116190	33688	2005/9~ 2010/06	13	202
キットカット	32209	35626	2005/1~ 2010/05	11	199
村木厚子	3164	834	2008/12~ 2010/12	8	49
尖閣諸島	44952	19323	2005/2~ 2011/01	15	289
上海万博	23578	24415	2005/5~ 2011/01	19	237

表 2 時間減衰を考慮した画像類似度の NDCG の平均値

	テキスト情報	画像情報
半減期 1 カ月	0.8174	0.8306
半減期 2 カ月	0.8188	0.8307
半減期 3 カ月	0.8193	0.8307
半減期 4 カ月	0.8193	0.8306
半減期 5 カ月	0.8191	0.8305
半減期 6 カ月	0.8190	0.8305
半減期 7 カ月	0.8186	0.8304
半減期 8 カ月	0.8185	0.8303
なし	0.8156	0.8281

で算出した画像のペアを類似度でランキングし、二つの画像に同じラベルが付いていれば 1、そうでなければ 0 とし、以下のように定義した NDCG [4] を算出する。

$$N_n = Z_n \sum_{i=1}^n \frac{2^{r(i)} - 1}{\log(1 + i)} \quad (9)$$

$r(i)$ は i 番目の類似度のスコアであり、 Z_n は全体が 1 となるように正規化した値である。

表 2 は式 (1)、式 (3) で定義した時間減衰を考慮した画像情報およびテキスト情報の類似度の NDCG である。各特徴量から時間減衰を導入することで、わずかに NDCG が上昇した。平均値では NDCG の変化が小さいが、「派遣村」、「鳩山由紀夫」、「村木厚子」など社会問題や政治に関するクエリでは、時間減衰を考慮することで 1% 程度 NDCG が上昇した。また、時間減衰を導入することで NDCG が低下した話題はなかった。以降の画像類似度ではテキスト情報の半減期を 3 カ月、画像情報の半減期を 3 カ月として実験を行った。

図 1 は式 (4) で定義した混合率 γ を 0 から 1 まで変化させたときの全クエリの NDCG の平均値である。時間減衰を考慮したとき各特徴量を混合したときの NDCG の平均値と、時間減衰を考慮しないとき各特徴量を混合したときの NDCG の平均値を算出した。 $\gamma = 1$ では、テキスト情報のみを用いたときの NDCG であり、 $\gamma = 0$ では、画像情報のみを用いたときの

NDCG である。テキスト情報と画像情報を混合した場合、時間減衰を導入することで導入しない場合と比較して、0.5% ほど NDCG が上昇した。表 2 および図 1 の結果から時間減衰を導入することで NDCG が上昇することが分かった。また、テキスト情報と画像情報の両特徴量を混合することにより全ての話題で NDCG が上昇した。 $\gamma = 0.3$ において最も高い NDCG が得られ、各特徴量を単独で用いた場合より 4% 程度 NDCG の平均値が上昇した。以降では $\gamma = 0.3$ を用いる。

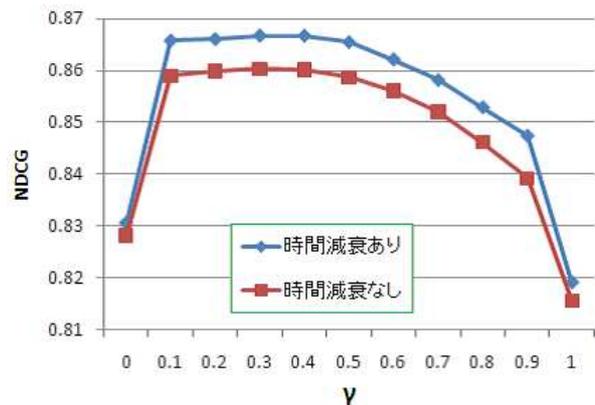


図 1 画像情報とテキスト情報の混合率 γ による NDCG の変化

5.2.2 クラスタリングの評価

クラスタリング結果の評価は、ラベル付けしたクエリについて、適合率 (precision)、再現率 (recall)、F 値について評価を行った。適合率および再現率は以下の定義を用いる。

$$Precision = \frac{TP}{TP + FP} \quad (10)$$

$$Recall = \frac{TP}{TP + FN} \quad (11)$$

ここで TP (True Positive) とは、同じ画像クラスに属する画像のペアのうち、両方の画像がどちらも同じラベルに属する画像ペアの総数とする。FP (False Positive) とは、同じ画像クラスに属する画像ペアのうち、ラベルが異なる画像ペアの総

表 3 Average を用いた各話題のクラスタリングの精度

	Precision	Recall	F
アフラック	0.353	0.200	0.256
ソフトバンク	0.578	0.272	0.370
オトナグリコ	0.647	0.567	0.605
マスキングテープ	0.283	0.136	0.184
ヒルサイドテラス	0.423	0.651	0.513
派遣村	0.451	0.352	0.395
鳩山由紀夫	0.552	0.512	0.531
キットカット	0.529	0.583	0.575
村木厚子	0.223	0.179	0.198
尖閣諸島	0.539	0.220	0.312
上海万博	0.511	0.202	0.290
平均	0.463	0.352	0.384

数とする。FN(False Negative) とは、同じラベルに属する画像ペアのうち、画像クラスタが異なる画像ペアの総数とする。 F 値は適合率・再現率の調和平均であり、以下のように定義する。

$$F = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

表 3、表 4 および表 5 に Average, Single, Complete の各手法を用いたときのクラスタリング結果の適合率、再現率および F 値を示す。各結果は F 値が最大となるような類似度でデンドログラムを切断したときの値である。各話題の F 値の平均値では、Average を利用したときのクラスタリングが最も高い F 値を得た。話題別では、「アフラック」、「ソフトバンク」では Single を用いたクラスタリングで最も高い F 値が得られた。「村木厚子」では Complete を用いたクラスタリングで最も高い F 値が得られ、その他の話題では Average を用いたクラスタリングで最も高い F 値が得られた。Single を用いたクラスタリング手法では一つの巨大な画像クラスタを作る傾向があり、安定した F 値が得られなかった。また Complete では、画像クラスタ C_i に属する画像 x_i と画像クラスタ C_j に属する画像 x_j のうち、 x_i と x_j のペアで一つでも類似度が 0 となると、 C_i と C_j は結合しなくなるため、高い F 値が得られなかった。このため本研究での階層クラスタリングは安定した F 値が得られた Average を用いる。

5.2.3 画像ランキングの評価

画像ランキングの評価として、評価対象とする画像をそれぞれ的话题からランキングした上位 100 枚の画像を選択した。クエリとの関連性を、2(非常に関連している),1(関連している),0(関連していない)の 3 段階のスコアに分類し人手で評価を行い、NDCG を算出する。

図 2 は式 (5) で定義した Visual Text Rank(VTR) の減衰係数 d を 0.45 から 0.85 まで変化させたときの全クエリの

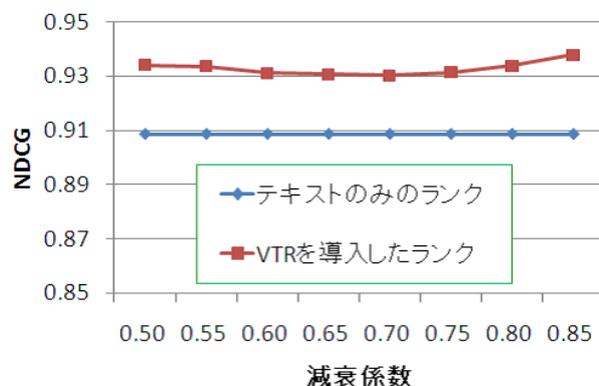
表 4 Single を用いた各話題のクラスタリングの精度

	Precision	Recall	F
アフラック	0.233	0.294	0.260
ソフトバンク	0.524	0.321	0.398
オトナグリコ	0.773	0.387	0.516
マスキングテープ	0.889	0.022	0.044
ヒルサイドテラス	0.813	0.163	0.271
派遣村	0.391	0.342	0.365
鳩山由紀夫	0.004	0.469	0.009
キットカット	0.001	0.986	0.001
村木厚子	0.250	0.221	0.235
尖閣諸島	0.094	0.495	0.158
上海万博	0.245	0.210	0.226
平均	0.383	0.355	0.226

表 5 Complete を用いた各話題のクラスタリングの精度

	Precision	Recall	F
アフラック	0.683	0.113	0.194
ソフトバンク	0.742	0.215	0.333
オトナグリコ	0.824	0.152	0.256
マスキングテープ	0.667	0.041	0.077
ヒルサイドテラス	0.448	0.220	0.295
派遣村	0.796	0.239	0.368
鳩山由紀夫	0.508	0.275	0.357
キットカット	0.383	0.246	0.300
村木厚子	0.400	0.198	0.265
尖閣諸島	0.725	0.075	0.137
上海万博	0.497	0.116	0.188
平均	0.607	0.172	0.252

NDCG@100 の平均値である。テキストのみのランキングとは画像のスコアを式 (6) で定義したときのランキングである。図 2 の結果から、VTR を導入することによりテキストのみのランキングと比較して NDCG が 2 % 程度上昇した。しかし、減衰係数 d を変化させたときの NDCG の変化は少なかった。本研究ではわずかに高い NDCG が得られた減衰係数 $d = 0.85$ を用いる。

図 2 減衰係数 d を 0.5 から 0.85 まで変化させたときの VTR における NDCG

5.2.4 画像クラスタのランキングの評価

画像クラスタのランキング評価ではクエリに対する適合画像の割合を画像クラスタのスコアとし、NDCG を算出する。それぞれの話題から上位 10 個の画像クラスタの適合画像を人手で評価した。

図 3 は全話題の NDCG の平均値をとったときの、各手法のクラスタランキングの結果を示す。横軸は式 (8) で定義した結合パラメータ δ であり、 $\delta = 1$ では VTR のみを利用したときのランキングとなる。Baseline としては画像クラスタに属する画像数の大きい順にランキングしたものである。VTR とは式 (7) を画像クラスタのスコアとしてランキングしたものである。VTR+SIM とは式 (8) を画像クラスタのスコアとしてランキングしたものである。Baseline と VTR を比較した場合、VTR では 5 %ほど高い NDCG が得られた。VTR と VTR+SIM を比較した場合、 $\delta = 0.8$ のとき VTR+SIM では 1 %ほど高い NDCG が得られた。しかし、VTR+SIM では δ が小さくなるにつれ NDCG が低下した。これは画像クラスタに属する類似度の平均の重みを強くした場合、不適合なコピー画像のみを含む画像クラスタが上位にランキングされ、NDCG が低下したためと考えられる。

Baseline、VTR および $\delta = 0.8$ を用いたときの VTR+SIM を話題別に NDCG を算出した結果を図 4 に示す。検索クエリ「鳩山由紀夫」では、Baseline と比較して VTR または VTR+SIM において NDCG が大幅に改善された。VTR では多数の画像と高い類似度で結ばれている画像は重要な画像となる。テキスト情報と画像情報の両方の類似度が高い画像を多く含むクラスタはランクが上がり、テキスト情報のみの類似度が高い画像クラスタはランクが下がる。Baseline では画像クラスタに含まれる画像の類似度が高くない場合でも、クラスタに含まれる画像数が多ければ上位にランク付けされる。検索クエリ「鳩山由紀夫」では、記事の内容ではクエリと関連している場合でもアスキーアートや漫画など関連のない画像が挿入される記事が多数存在する。そのためテキスト情報のみの類似度が高く画像数の多いクラスタが生成され、高い NDCG が得られなかった。

また「キットカット」、「上海万博」でも同様に VTR または VTR+SIM を用いることで NDCG が改善された。一方、「アブラック」、「ソフトバンク」、「オトナグリコ」では各手法による NDCG の変化が小さかった。これは CM や商品などの画像に対して的確に説明した文が多く存在し、ノイズが少なかったためであると考えられる。NDCG の平均では、VTR と画像クラスタ内の類似度を考慮した手法である VTR+SIM において、NDCG が最も高く精度が得られた。

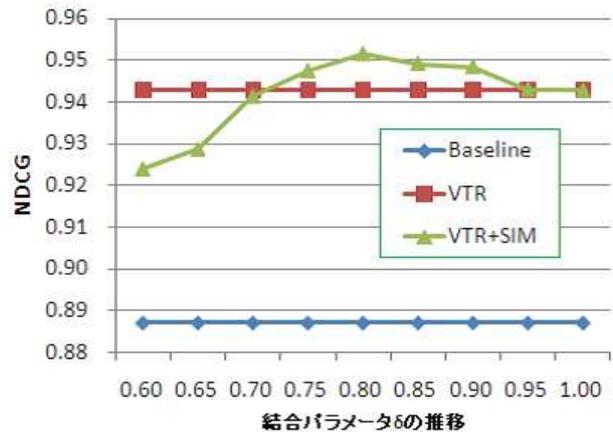


図 3 VTR(Visual Text Rank) と画像クラスタに属する類似度を用いたときの画像クラスタのランキングの NDCG

6. まとめと今後の課題

与えられたクエリに関する記事と画像を抽出したうえで、画像のクラスタリング手法およびクラスタのランキング手法を提案した。画像類似度としてテキスト情報、画像情報および時間情報を用いた。算出された画像類似度から階層クラスタリングを用いてクラスタリングし、ランダムウォークを用いて算出した画像の重み、画像クラスタに属する類似度の平均を利用して画像クラスタのランキングを行った。

評価実験より、テキスト情報、画像情報、時間情報を用いることで、類似度の精度が上昇することを示した。本手法の類似度を用いた階層クラスタリングでは Single, Complete, Average を利用してクラスタリングを行った結果、Average が最も高い F 値が得られることが分かった。また、ランダムウォークを用いて算出した画像の重みと、画像クラスタに属する類似度の平均を利用して、適合率の高い画像クラスタが抽出できることを示した。

今後の課題として、クラスタリング結果の時系列上の可視化することにより、イベントなどの事象の出現期間を視覚的に把握できる。また、各画像クラスタの内容を表すラベルを自動生成する手法を検討している。

文 献

- [1] D.Liu, X.Hua, L.Yang, M.Wang, H.Zhang, "Tag Ranking", WWW 2009 MADRID!, Track: Rich Media / Session: Tagging and Clustering
- [2] Flickr. <http://www.flickr.com/>
- [3] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", 1998.
- [4] K. Jarvelin and J. Kekalainen. "Cumulated Gain-Based Evaluation of IR Techniques", In ACM Transactions on Information System, 2002.

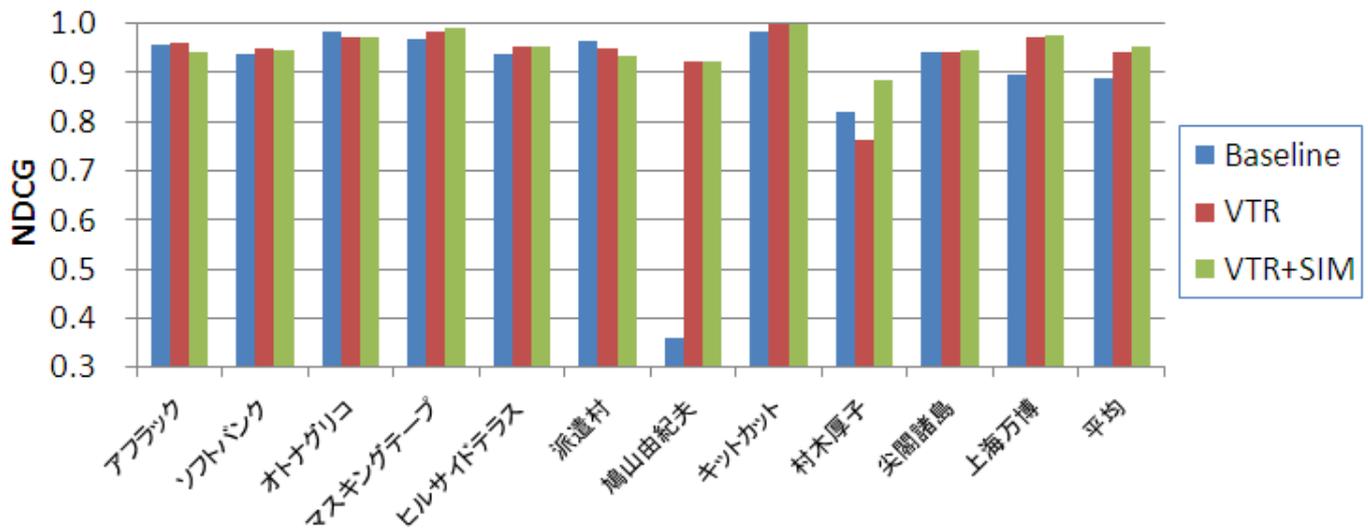


図4 画像クラスターのランキングのNDCGを話題別に表示

- [5] D.G.Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", Accepted for publication in the International Journal of Computer Vision, 2004
- [6] Y.Jing, S.Baluja, "PageRank for Product Image Search", WWW 2008 / Refereed Track: Rich Media, April 21-25, 2008, Beijing, China
- [7] R.H.Leuken, L.Garcia, X.Olivares, R.Zwol, "Visual Diversification Image Search Results", WWW 2009 MADRID! Track: Rich Media / Session: Tagging and Clustering
- [8] M.S.M.Orengo, "Similarity of color images", Proc. SPIE Storage and Retrieval for Image and Video Databases, 2420:381-392,1995.
- [9] B.Manjunath, W.Ma, "Texture features for browsing and retrieval of image data", Pattern Analysis and Machine Intelligence, IEEE Transactions on, 18(8):837-842,1996
- [10] B.Sigurbjornsson, R.Zwol, "Flickr Tag Recommendation based on Collective Knowledge", WWW 2008 / Refereed Track: Rich Media, April 21-25, 2008, Beijing, China
- [11] D.Crandall, L.Backstrom, D.Huttenlocher, J.Kleinberg, "Mapping the World's Photos", WWW 2009 MADRID!, Track: Social Networks and Web 2.0 / Session: Photos and Web 2.0
- [12] L.Wu, L.Yang, N.Yu, X.S.Hua, "Learning to Tag", WWW 2009 MADRID!, Track: Rich Media / Session: Tagging and Clustering
- [13] L.Kennedy, M.Naaman, "Generating Diverse and Representative Image Search Results for Landmarks", WWW2008, April 21-25,2008,Beijing, China.
- [14] T.Yao, T.Mei, C.Ngo, "Co-reranking by Mutual Reinforcement for Image Search", CIVR'10, July 5-7. Xi'an,China.
- [15] B.Cui, A.Tung, C.Zhangm Z.Zhao, "Multiple Feature Fusion for Social Media Applications" SIGMOD'10, June 6-11, 2010, Indianapolis, Indiana, USA.
- [16] R.Zwol, B.Sigurbjornsson, "Faceted Exploration of Image Search Results", WWW 2010, April 26-30,2010,Raleigh, North Carolina, USA.
- [17] Z.Zha, L.Yang, T.Mei. M.Wang, Z.Wang, "Visual Query Suggestion", MM'09, October 19-24, 2009, Beijing, China.