

画像類似度に基づくリンク解析を用いた 画像ランキング手法の比較検討

山根 遥香[†] 豊田 正史[‡]

[†] 東京大学情報理工学系研究科 [‡] 東京大学生産技術研究所

E-mail: [†] yamane@tkl.iis.u-tokyo.ac.jp, [‡] toyoda@tkl.iis.u-tokyo.ac.jp

あらまし Jingらが提案した VisualRank は画像類似度に基づく無向グラフを形成し、リンク解析を用いて画像をランキングする手法である。VisualRank は商品画像の検索において有効性が確認されているが、企業や商品のロゴマークなど多くの画像と部分的に類似しやすい画像に高いスコアが与えられる問題が指摘されている。本論文では、画像類似度のグラフを有向化することにより、ランキングを改善できることを示し、画像類似度の有向グラフを用いた directed-VisualRank と、それに HITS を適用した VisualHITS を提案する。これらの手法の比較実験により提案手法の優位性を示す。

キーワード VisualRank, directed-VisualRank, VisualHITS

1. はじめに

Web上の文書量は google の公式ブログの発表によると、現在 1 兆を超えており、検索エンジンの役割は益々、重要度を増している。当初テキストのみであった検索も、Web 上のコンテンツの多様化により、画像、動画、音声などテキスト以外において必要とされるようになってきている。中でも画像検索は多くの検索エンジンにおいて実装されているものの、多くの場合、画像をランキングするのに、画像特徴を使用せず画像の周辺のテキスト情報を使用している。そのため複数の概念を持つクエリを入力した場合、概念の混合した画像が検索結果として表示されるといった問題点がある。

Jingらは、画像間の類似度を表すグラフに対し、リンク解析を適用して画像をランキングする手法である VisualRank [1][2]を提案した。類似度の算出には画像の拡大縮小、回転等に対して頑健な SIFT 特徴量[3][4]を用い、ランキングには Web 上のページの重要度を算出する PageRank[5]を用いているのが特徴である。VisualRank は、商品画像検索に関して評価が行われ、よい精度を示しているが、企業や商品のロゴマークなど明確な形状を持ち、多数の画像の一部として現れる画像が高くランクされる問題点がある。本論文ではこの VisualRank の問題点が、画像間の類似度を有向化することにより改善されることを示し、画像類似度の有向グラフに基づく 2 種類のリンク解析アルゴリズムである directed-VisualRank と VisualHITS を提案する。また、Google 画像検索から取得した画像を用いてこれらの手法の比較実験を行う。

以降では、まず、第 2 章において関連研究について述べる。第 3 章では、今回提案する手法の基となる Jing らの VisualRank について述べ、有向類似度を用いた VisualRank である directed-VisualRank と、有向類似度

グラフに HITS[7]を適用した VisualHITS を提案する。第 4 章では今回行った実験結果を示し、最後に第 5 章で本稿のまとめについて述べる。

2. 関連研究

VisualRank を改善する手法は、これまでにいくつか提案されているが、画像類似度に対称性を導入し、複数のリンク解析手法について比較検討されたものは少ない。

Multiclass VisualRank[8]は類似度のグラフを normalized cut 手法を用いてクラスタリングし、クラスタ毎に VisualRank を適用する手法である。画像のカテゴリ毎のランキングを提示することで多様な画像が得られる。

また、ContextRank[9]は、画像中の visual word をノードとし、画像内及び画像間における visual word 同士の関連を算出してグラフ化する。Visual word のグラフについて PageRank を適用し、画像中の visual word のスコアを合計してランキングを行う。

Zhangらは Descriptive Visual Word 及 Discriptive Visual Phrase と呼ばれる画像類似度の算出手法を提案している[10]。まず、カテゴリに分類された画像データベースから各カテゴリをよく表す visual word を抽出する。同じ画像において共起する visual word のペアを visual phrase として抽出し、これらを用いた VisualRank を用いてランキングを行う。

更にクエリと画像の関係を考慮して重みづけを行う手法も提案されている[11]。

また、VisualRank の応用としては、画像とテキスト間の信憑性分析手法[12]が提案されている。これは、クエリに関連する画像セットとテキストセットから対

応する正解を抽出するが，正解画像の決定に VisualRank を用いている．

3. 提案手法

3.1. VisualRank

まず Jing らの画像類似度グラフに基づくランキング手法である VisualRank について述べる．この研究は画像内容検索(CBIR, content-based image retrieval)に分類され，近年増えてきている研究分野である．VisualRank では特徴量として局所特徴量である SIFT 特徴量を用いる．これは web 上の画像は無作為であり，違うカメラ，焦点距離，構図などの画像を含むため変形にロバストな特徴量の使用が望まれるためである．SIFT によって得られた特徴点のうち画像間で対応する点の割合を画像間の類似度とし，PageRank[5]を適用する．PageRank は Web ページの重要性を測るアルゴリズムである．良質なページからリンクされているページもまた良質であるという考え方を用いており，実際には以下の式で計算される．行列の要素 a_{ij} が i から j へのリンクがある時は 1, そうでない場合は 0 である隣接行列 A を定義する．ユーザがリンクのないページへランダムで移動する Random jump[13]を考慮するとき, A を行で正規化した行列 A^* を用いて PageRank(PR)は以下のように定義される．

$$PR = dA^* \cdot PR + (1 - d)p, \quad \text{where } p = \left[\frac{1}{n} \right]_{n-1} \quad (1)$$

d は減衰係数, p は random jump vector である． d は慣習的に 0.8 以上の値が用いられる．この行列要素を先ほど SIFT で計算した特徴量に置き換える．画像の類似度で対称隣接行列 S を定義したとき行列要素 S_{uv} は画像 u と v の間の共通する SIFT 特徴点の割合であり，それぞれの特徴点の集合を U, V とした時 $|U \cap V| / \{(|V| + |U|) / 2\}$ で定義される． S を画像間の類似度を表す行列とすると， S を正規化した行列 S^* を用いて VisualRank(IR)は以下のように定義される．

$$IR = dS^* \cdot IR + (1 - d)p, \quad \text{where } p = \left[\frac{1}{n} \right]_{n-1} \quad (2)$$

この手法は，商品検索に関して評価実験が行われており，検索結果の改善が見られたと述べられている．しかし，ロゴやブラウザのパネルなど，明確な形状をもち多数の画像の一部として現れる画像のスコアが高くなるといった問題点がある．

3.2. Directed-VisualRank

本節では，VisualRank における画像類似度の定義を有向類似度に修正した directed-VisualRank を提案する．VisualRank の行列 S は対称隣接行列であり方向性を有

- Non-directed similarity

$$s_{u,v} = \frac{|U \cap V|}{|U + V| / 2}$$



- Directed similarity

$$s_{u,v} = \frac{|U \cap V|}{|U|}$$



$$s_{u,v} = \frac{|U \cap V|}{|V|}$$



図 1 類似度の定義

さないが，directed-VisualRank では，図 1 に示すように，画像類似度を有向化する．画像 u, v の特徴点の集合を U, V とし， u から v への類似度を $|U \cap V| / |U|$ ， v から u への類似度を $|U \cap V| / |V|$ と定義する．例えばロゴの画像とそれを含む製品の画像の類似度を考えるとき，方向を考えない無向類似度では，類似度は方向によらず同一である．しかし，有向の類似度ではロゴ画像からロゴを含む製品の画像への類似度は大きくなり，ロゴを含む画像からロゴ画像への類似度は小さくなる．そのため，ロゴ画像よりロゴを含む製品の画像の方が重要と判断されることになる (図 1)．Directed-VisualRank は，この有向類似度を用いた類似度行列を VisualRank の S と置き換えて算出する．

3.3. VisualHITS

本節では Web のリンク解析アルゴリズムである HITS[7]を画像類似度グラフに適用した VisualHITS を提案する．まず web 上のページのランキングアルゴリズムである HITS について述べる．良質なページにリンクされているページもまた良質であるというコンセプトは PageRank と同様であるが，HITS では多くの良いページへのリンクを有しているページをハブ(Hub)と呼び，多くの良いページからリンクを受けているページをオーソリティ(Authority)と呼ぶ．各ページのハブスコア(h),オーソリティスコア(a)は行列 S を用い以下のように定義される．

$$\vec{a} = S^T \vec{h} \quad (3)$$

$$\vec{h} = S \vec{a} \quad (4)$$

表 1 実験で用いたクエリ

建築物	eiffel tower, colosseum, sagrada familia
動物	dog, cat
製品	iphone, dell computer, Nintendo wii system
食べ物	Curry, pasta, beef steak
絵画	Monalisa, picasso painting
複数の意味を持つクエリ	Jaguar, mcdonald's



(a)VisualRank



(b)directed-VisualRank



(c)VisualHITS

図 2 「dell computer」 の上位 5 枚の画像

VisualHITS では、ページを画像に置き換え、行列 S の要素は directed-VisualRank と同様に有向類似度を用いる。高いオーソリティスコアを持つ画像を重要な画像であると見なし、結果として出力する。

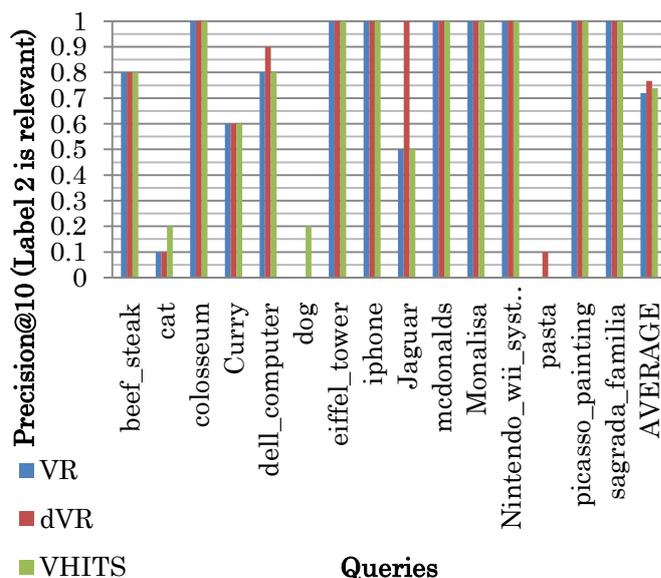
4. 実験

第 3 章で述べた VisualRank, direct-VisualRank, VisualHITS について以下の比較実験を行った。

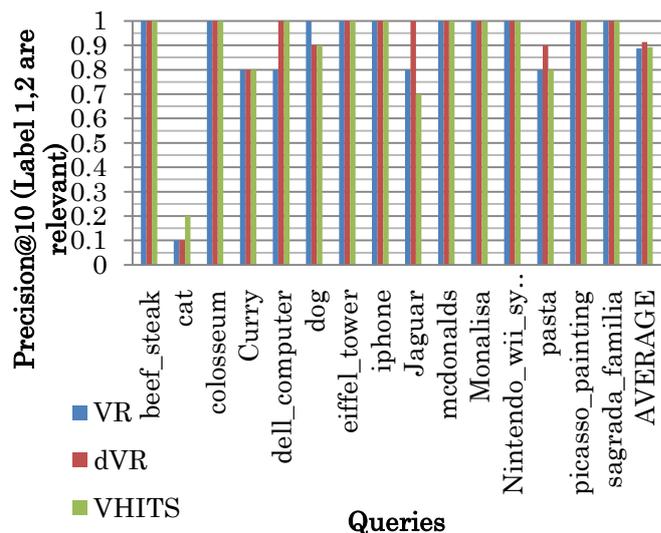
4.1. 実験方法

まず、表 1 に示す 6 種類 15 個のクエリを用いて、Google image search で検索を行い、検索結果上位 200 枚の画像を取得した。取得した画像セットそれぞれに対して VisualRank, direct-VisualRank, VisualHITS を適用し画像の再ランキングを行った。

上位 10 枚の画像について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った。評価の基準は 2 名の評価者の議論に



(a)2(非常に関連している)と判断された画像のみを正解画像としたとき



(b)2(非常に関連している)および 1(関連している)と判断された画像を正解画像としたとき

図 3 上位 10 件の適合率

表 2 上位 10 件の平均適合率

	VisualRank	Directed-VisualRank	VisualHITS
(a) Score 2 is relevant	0.72	0.77	0.74
(b) Score 2 and 1 are relevant	0.80	0.84	0.82

よって定めた。例えば、dog, cat に関しては、犬や猫の実物の写真を 2, 犬や猫のイラストやそれをモチーフとした商品などを 1 と判定した。

評価には上位 10 件の適合率を用いる。評価 2 の画像のみを正解とした場合、および評価 2, 1 を共に正解とした場合の適合率を算出し、比較を行う。

4.2. 結果と考察

再ランキングした画像の例として「dell computer」の上位 5 枚の画像を図 2 に示す。VisualRank ではリストの 3 位にロゴがランクされているが、directed-VisualRank と VisualHITS ではコンピュータを主体とした画像がランクされており、ロゴのランキングが低くなっているのが分かる。

各クエリに対する適合率を図 3 に示す。(a)は評価 2 のみを正解としたとき、(b)は評価 2,1 を正解としたときのものである。「cat」、「dell computer」、「dog」、「Jaguar」、「pasta」のクエリについて提案手法による適合率の向上が見られる。例えば、「Jaguar」では、車と動物のジャガーの写真を正解としていたが VisualRank では車の Jaguar のロゴ及びミシンの Jaguar が上位にランクされており、directed-VisualRank では、これらが低くランクされることにより適合率が向上した。「dell computer」では図 2 に示した効果により directed-VisualRank, VisualHITS 双方において適合率が向上している。また、「pasta」においても白い背景で単純な形状を持つ不適合画像が VisualRank では高くランクされており、これが directed-VisualRank では低くランクされている。

各手法における平均適合率を表 2 に示した。(a)では、directed-VisualRank が 0.77 と最高の平均適合率を示しており、VisualRank と比べて 5% 向上している。また、VisualHITS は directed-VisualRank には及ばないもののわずかながら適合率の向上が見られる。この傾向は、(b)でも同様であり、directed-VisualRank は VisualRank を 4% 上回り、最高の平均適合率を示している。

5. まとめと今後の課題

本論文では、画像類似度に基づく無向グラフを形成し、リンク解析を用いて画像をランキングする手法である VisualRank の問題点を示し、画像類似度の定義を有向化することによって改善を試みた。有向化した類似度を用いた画像のグラフに対し、2 種類のリンク解析アルゴリズム(PageRank 及び HITS)を適用した手法である、directed-VisualRank 及び VisualHITS を提案し、VisualRank との比較検討を行った。Google image search から取得した画像の最ランキング結果について上位 10 件の適合率を比較した結果、directed-VisualRank,

VisualHITS では、VisualRank において適合率悪化の原因となっていた画像のランクが低下し、適合率が改善されることが確認された。

今後の課題としては、提案手法の大規模実験、アプリケーションの開発等が挙げられる。例えば、現在、本提案手法をブログアーカイブ上の画像に適用し、時期毎の画像ランキングを行うアプリケーションを開発しており、これにより画像のトレンドの変遷が追跡可能となる。

参考文献

- [1] Y. Jing and S. Baluja, "PageRank for Product Image Search," Proc. of the 17th international conference on World Wide Web (WWW'08), pp 307-316, 2008.
- [2] Y. Jing and S. Baluja, "VisualRank Applying PageRank to Large-Scale Image Search," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.30, no.11, pp.1877-1890, 2008.
- [3] D. G. Lowe, "Object recognition from local scale-invariant features," Proc. of the 7th international Conference on Computer Vision, pp.1150-1157, 1999.
- [4] D. G. Lowe. "Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol.60, no.2, pp.91-110, 2004.
- [5] Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd, "The PageRank Citation Ranking: Bringing Order to the Web", 1998.
- [6] Taher H. Haveliwala, "Efficient Computation of PageRank", Stanford Technical Report, 1999.
- [7] J. M. Kleinberg. Kleinberg. "Authoritative sources in a hyperlinked environment." In Proceedings of ACM-SIAM Symposium on Discrete Algorithms, 1998. Also appears as IBM Research Report RJ10076(91892). www.cs.cornell.edu/home/kleinber/auth.ps.
- [8] M. Ambai and Y. Yoshida, "Multiclass VisualRank: VisualRanking Method in Clustered Subsets Based on Visual Features", SIGIR2009.
- [9] S. Wang, Q. Huang, S. Jiang, L. Qin and Q. Tin, "Visual ContextRank for Web Image Re-ranking", ACM Multimedia 2009: 657-660.
- [10] S. Zhang, Q. Tian, G. Hua, Q. Huang and S. Li, "Descriptive Visual Words and Visual Phrases for Image Applications", ACM Multimedia 2009.
- [11] L. Wang, L. Yang and X. Tian, "Query Aware Visual Similarity Propagation for Image Search Ranking", ACM Multimedia 2009.
- [12] 山本雄介, 田中克己, "Web 上の画像. テキスト対の信憑性分析," 第一回データ工学と情報マネジメントに関するフォーラム(DEIM 2009), A5-5.
- [13] M. R. Henzinger, A. Heydon, M. Mitzenmacher, and M. Najork. "On near-uniform URL sampling", In WWW9, Amsterdam, May 2000. www9.org/w9cdrom/88/88.html.