

色・形状・テクスチャに基づく画像検索の自動評価と多様化

富樫 陸^{1,2,a)} 藤田 澄男^{1,b)} 酒井 哲也^{2,c)}

概要: 本論文では、色や形をはじめとするプリミティブな画像内要素が重要となる e コマースなどにおける画像検索のために、人手による適合性判定なしに多値適合性を測ることのできるフレームワークを提案する。この論文で提案する手法では、色、形、テクスチャに関する低次元特徴量によって記述された各画像が特徴空間における潜在的な検索意図から生成されていると捉えることによるシンプルな確率的生成モデルを導入することで、この上にカテゴリ内の多値適合性を定義し、nDCG を計測することを可能にする。提案手法により自動計算した nDCG とユーザの好みとの一致度を大規模クラウドソーシングによって計測し、評価の妥当性を示す。更に、同じモデルの上で D-nDCG や ERR-IA のような多様化指標の理想的リストも自然に導出できることを示し、これを用いた検索結果の多様化についてもいくらか議論を行う。

キーワード: 類似画像検索, e コマース, 多値適合性, 多様化

Automatic Evaluation and Diversification of Ranked Product Images based on Colour, Shape, and Texture

TOGASHI RIKU^{1,2,a)} FUJITA SUMIO^{1,b)} SAKAI TETSUYA^{2,c)}

Abstract: We consider the problem of automatically evaluating the search result quality for image-to-image product search applications where each product image is iconic (i.e., the image clearly shows what the product looks like and basically nothing else) and is accompanied by a product category information. We propose a simple probabilistic model that assumes that an image is generated from latent intents that each represent primitive image factors such as colour, shape, and texture. This enables us to automatically compute a graded relevance score for each retrieved image, and thereby to compute graded relevance evaluation measures such as nDCG (normalised Discounted Cumulative Gain) without relying on human relevance assessments. We evaluate how the proposed measure, which we call I-nDCG, aligns with human preference judgments through a large-scale crowdsourcing experiment. Moreover, we show that image search result diversification methods can naturally be derived based on our model by following the approaches of existing diversity evaluation measures, namely, D-nDCG and intent-aware measures.

Keywords: content-based image retrieval, e-commerce, graded relevance, search result diversification

1. はじめに

類似画像検索は、特に e コマースにおける商品検索に

おいて重要なアプリケーションである。実際、Microsoft Bing^{*1}, Pinterest[1], Alibaba[2] などが一斉に研究結果をアプリケーションとしてリリースした。

これらの背景には物体検出技術と物体特徴抽出技術の成熟がある。2013 年から 2016 年にかけて、物体検出技術は急速に発展した [3][4][5]。物体検出技術を用いることで、

¹ ヤフー株式会社
Yahoo Japan Corporation
² 早稲田大学
Waseda University
^{a)} rtogashi@yahoo-corp.jp
^{b)} sufujita@yahoo-corp.jp
^{c)} tetsuyasakai@acm.org

^{*1} <https://blogs.bing.com/search-quality-insights/2017-06/beyond-text-queries-searching-with-bing-visual-search>

類似商品検索におけるクエリとドキュメントは背景の影響が少なく、物体が大きく中心に写っているような画像に変換することが可能になる。物体が大きくはっきりと背景と区別がつくように写った画像を“iconic な画像”と Berg ら [8] は呼んでいる。この論文では、それにならって、十分物体の輪郭にフィットした bounding box によって囲まれた領域を切り出したような物体画像を iconic な画像と呼ぶことにする。物体検出技術の成熟によって、クエリも検索対象も iconic な画像になりつつある一方で、スマートフォン上のインタラクティブな UI によって、物体領域を任意で切り出すといった UI も一般的になってきていることが iconic クエリの増加をより促している。

しかし、物体（商品）画像検索には、評価において大きな課題がある。まず一つに、大規模な物体画像検索向けデータセットが存在しないことが挙げられる。大規模な画像データセットは Imagenet[9], MS COCO[10] などが挙げられるが、物体領域ごとのランキングは提供されていない。ランキングを評価するための適合性判定を人手で行うのは極めて高価であり、公開されているものは少ない。

Philbin ら [11][12] などの従来の画像検索（特定物体認識）のデータセットにおいては、同じランドマークの写っている画像というカテゴリの一致でもって適合とし、むしろ見た目的な要素は特徴量の不変性によって吸収するような手法が開発されてきた。この時用いられる評価尺度は MAP (Mean Average Precision) などであり、多値適合性を考慮していない。これは e コマースにおける見た目に関して詳細なランキングが求められる類似画像検索と極めて異なる問題であるといえる。

本論文では iconic 画像における検索性能の評価にフォーカスし、規格化された MPEG7 Visual Descriptors[13] とその特徴空間における確率的生成モデルを導入することで、その上で自動的に得られる多値適合性を定義し、nDCG に相当するランキング指標を人手による適合性判定なしに計算するフレームワークを提案する。更に、提案手法で計測される nDCG とユーザの好みとの一致性を計測し、それを示す。

一方で、画像検索のもう一つの大きな問題として、検索意図の曖昧性がある。例えばある物体を含んだ画像がクエリとして与えられた時、ユーザにとって画像中の物体に色に興味があるのか、物体の形に興味があるのかによって情報要求が異なる。しかし、検索意図は画像のみからでは不明であり、このような場合において情報要求を満たすためには検索結果の多様化が一つのアプローチとなる。

提案手法では、前述の画像間の多値適合性だけでなく、情報検索の分野において従来開発されてきた多様化を考慮したランキング指標を適用するために必要となる検索意図毎の多値適合性とトピックのみが与えられた時の検索意図の事前確率も同じ確率的生成モデルの一部として自然に定

式化する。

2. 提案手法

iconic な画像に対しては、背景や画像内の局所性の影響はかなり少ないと考えていい。一方、iconic な画像に対しては MPEG7 Visual Descriptors のような低次特徴が見た目を記述する上で十分高性能である。このことを利用して、既存の分類用カテゴリの適合性判定付きデータセットにおける画像検索のカテゴリ内の視覚的な多値適合度を評価することができるようにするフレームワークを提案する。

2.1 問題設定

本論文では、検索対象である商品画像のデータセット $D = \{d_j | 1 \leq j \leq n\}$ の各画像に、e-コマースサービスにおける商品カテゴリ、例えば、“レディース ブーツ”、“レディーススニーカー”などが、一つ付与されているものとする。これらのカテゴリを、情報検索の評価実験における検索要求の表現であるトピックとみなして、 $T = \{t_j | 1 \leq j \leq m\}$ とする。画像検索における適合性は、従来、クエリ画像と検索済み画像が共通のトピック (カテゴリ) を持つことをもって適合であるとしてきた [11][12]。提案方式は、詳細なサブトピックなど追加のアノテーション作業なしに、この粗い適合性判定を、画像素性の観点から、よりきめ細かい段階的な判定を反映した評価が可能にすることを目的とする。

2.2 低次特徴量

提案手法は規格化されている MPEG-7 Visual Descriptors[13] の中から色、形、テクスチャに関する三つの低次特徴量を採用した。

- Scalable Color
- Region Shape (ART)
- Homoginuis Texture

2.3 I-nDCG

直感的に、ある 2 つの画像 q, d が先述の低次の画像特徴量の距離においてある閾値 $\theta (> 0)$ 以下であれば適合とすることが考えられる。つまり、以下のようにある特徴量における適合度 $r(q, d)$ を定義する。

$$r(q, d) = \begin{cases} 1 & (|f(q) - f(d)|_2^2 < \theta) \\ 0 & otherwise \end{cases} \quad (1)$$

ここで、 q はクエリ画像、 d はドキュメント画像、 f は特徴抽出である。多値適合性のためにももちろん複数の閾値があってもよい。

一方で、レディースロングブーツのような黒いものが圧倒的多数を占めているような、ユーザが詳細な色の違いを区別するようなカテゴリと多様な色のものが含まれるレ

ディーススニーカーのようなカテゴリでは、その閾値は異なるべきである。そこで、カテゴリ毎、特徴毎に距離の閾値を与えることが考えられるが、これにもまだ問題がある。実際の e コマースにおけるカテゴリはレディースロングブーツやレディーススニーカーという粗い粒度であり、実際にはカテゴリ内に複数の集団が存在する。その場合、集団毎の閾値を考慮するのが望ましいが、それらをハイパーパラメータとして管理するのは極めて煩雑である。また、この式 1 の定義においては、理論的に、その特徴空間においてクエリ周りにデータが球上に偏り無く存在することを仮定していることになる。これは特徴次元が一切の相関をもっていないことを仮定していることと等しい。

上記の問題に対して、本論文では、確率的生成モデルを導入することを提案する。つまり、各特徴によって記述されたカテゴリ内の画像が成すいくつかのクラスタという潜在変数から特徴空間上の画像が生成されていると捉える。すなわち、以下のような正規分布の重み付き和 (GMM: Gaussian Mixture Model[14]) によってモデル化される。

$$p_j(d|t) = \sum_i \pi_i \mathcal{N}(f_j(d)|\mu_i, \Sigma_i) \quad (2)$$

ここで、 i はカテゴリ内のクラスタを表す。 \mathcal{F} を低次元特徴量の集合として、 f_j は特徴量 $j \in \mathcal{F}$ のために用いられた特徴抽出を指している。 Σ_i はクラスタ i まわりにおける局所的な距離計量であると考えられることができるので、これによってカテゴリ内のクラスタと与えられた画像との距離がよりデータ全体を考慮したものになる。つまり、GMM によるソフトなクラスタリングはカテゴリ内のデータの分布を考慮したクラスタ毎の閾値の獲得のような働きをしている。一般に、閾値は特徴量の各次元に関しても異なると考えられるため、各コンポーネントの共分散によってその周辺における次元毎の計量を得ることは自然である。

得られたモデル $p_j(x)$ を用いて、以下のような多値適合度を定める。ある特徴量 $j(\in \mathcal{F})$ において得られたクラスタの集合を \mathcal{I}_j とすると、提案する指標における適合度 $r_{I-nDCG}(q, d)$ を人手による適合性判定なしに計算することが可能である。

$$p_j(i|d) = \frac{\pi_i \mathcal{N}(f_j(d)|\mu_i, \Sigma_i)}{\sum_{i \in \mathcal{I}_j} \pi_i \mathcal{N}(f_j(d)|\mu_i, \Sigma_i)} \quad (3)$$

$$\Theta(p) = \begin{cases} 1 & (p > \epsilon) \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$r_{I-nDCG}(q, d) = \sum_{j \in \mathcal{F}} \sum_{i \in \mathcal{I}_j} \Theta(p_j(i|q)) \Theta(p_j(i|d)) \quad (5)$$

ϵ は確率に対する閾値である。GMM の負担率 $p_j(i|d)$ は 0 から 1 に近い値をとりやすく、ピーキーであるため、実際には ϵ は 0.1 などに設定する。これを適合度として従来の nDCG にならってポジションごとのディスカウント式を適用したものを I-nDCG と呼ぶ。

$$g(r) = 2^{r_{I-nDCG}(q, d_r)} - 1 \quad (6)$$

$$I-nDCG@k = \frac{\sum_{r=1}^k g(r) / \log(r+1)}{\sum_{r=1}^k g^*(r) / \log(r+1)} \quad (7)$$

3. DATA

3.1 E-commerce Shoes Dataset

今回、論文に必要な実験をするにあたって使用するデータセットは、大規模な e コマースサイトにおいて集まった 78 カテゴリの靴関連商品の画像から高品質のものを取り出したものである。

3.2 Statistics of the Dataset

本論文の実験に用いたデータセットは、78 カテゴリの 129,755 枚の靴の画像からなる。カテゴリ毎のデータ数の最大は 18,518、最小は 124、平均は 1,663、標準偏差 2,781 となる。

3.3 データセットのクリーニング

本論文では e コマースサイトに商品データが入稿された時点で画像が iconic といえるような画像を選びだすことでデータセットを構築した。具体的には、6 項目の基準でデータセットとは個別に分離して用意したデータに対して人手によるアノテーションを行った。基準は (1) 人が写っていない (2) 背景が白い (3) 文字が写っていない (4) 装飾がない (5) 商品数が一つ (6) 商品のサイズが適切になる。これによって得られたサブデータセットに対して、Convolutional Neural Network[15] による分類モデルを各項目毎に個別に作成した。それらのモデルによってデータセット全体に対してスコアを付与し、スコア平均を閾値処理することで高品質な画像であると分類されたもののみをデータセットとした。また、画像同士は perceptual hash[16] による重複除去を行っている。

4. EXPERIMENT: I-nDCG

4.1 EXPERIMENTAL SETUP

I-nDCG がカテゴリ内のランキング評価のために妥当であることを検証する。あるクエリに対して MPEG7 Visual Descriptors を用いた十分チューニングされた検索システムによって検索を行う。この検索結果の n 位から $n+9$ 位を切り出してトップをクエリそのものとしたものを一つの検索結果と見なして複数取り出し、提案手法による nDCG を計測する。切り出した検索結果の件数は一つのクエリあたり最大 2 つまで取り出した。これによって得られた検索結果のうち、I-nDCG が 0.4 以下 (low)、0.4 から 0.6 の間 (mid)、0.6 以上 (high) の三群に分け、これらの中から二つの検索結果を並置して計 3 通りワーカーに見せた際の preference との一致性を計測する。884 のクエリそれぞれについて重複度 10 以上 30 以下のワーカーによ

て投票が行われた。よって、884のクエリそれぞれについて low-mid, low-high, mid-high の三通りで投票数 n と m ($10 \leq n + m \leq 30$) が得られることになる。参加したワーカーは 1,521 人、投票数は 60,350 になる。

4.2 Results

あるクエリに関する low の得票率を、あるクエリの low の検索結果がワーカーに提示された回数のうち、これらが票を得た (mid or high より prefer された) 回数の割合とする。図 1 は side-by-side のユーザ投票における、各スコアの検索結果の得票率の箱ひげ図である。また、表 1 は I-nDCG において low, mid, high の 3 つのシステムに関して対応あり TukeyHSD による統計検定を行った結果である。効果量は Hedge's g によって計測している。これらの結果から、それぞれの組み合わせに関して統計的に有意に識別可能である。

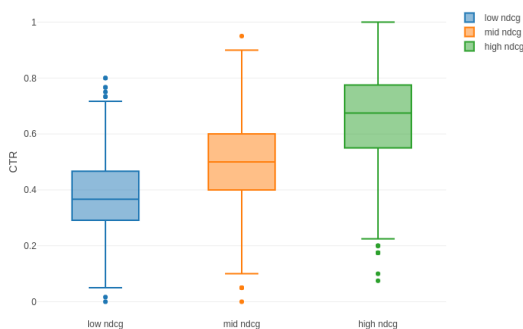


図 1 side-by-side 比較における得票率

表 1 side-by-side 比較投票に関する対応あり TukeyHSD による統計検定 (99%信頼区間) の結果

inst.	diff	lower	upper	p-value	effect size
mid-low	.118	.0927	.143	<2e-16	.649
high-low	.270	.244	.295	<2e-16	1.48
high-mid	.152	.127	.177	<2e-16	.836

4.3 考察

I-nDCG が iconic 画像データセットに関して評価指標として有効なことを示した。図 2 は I-nDCG の評価に用いた検索結果と同じ方法での作成した検索結果の例である。赤い枠で囲われた画像がクエリで、右に続く画像が検索結果である。5 行目はスニーカーカテゴリーの検索結果であるが、カテゴリー内の色のバリエーションが多いため、中スコアの検索結果上位にピンクの靴が色に関して適合となっている。一方で、6 行目の男性用革靴の検索結果では、低スコアの検索結果もそれなりに類似しているものが含まれているように見える。しかし、男性用革靴は色や形の分散が極めて小さいため、提案手法はより詳細な評価指標として

適応することで多値の指標として機能して、検索結果の品質を識別している。

5. 多様化

提案するフレームワークでは I-nDCG だけでなく、従来の多様化指標を最大化しているとみなすことのできる多様化ランキング手法を導出することが可能である。この章では二つの多様化指標の理想的リストの獲得とみなせる多様化手法を説明する。

まず、GMM によって得られたカテゴリ内各特徴におけるクラスタを検索意図と見なすことを考える。つまり、各特徴量で記述された画像が、検索意図という潜在変数から生成されているとみなす。このとき、トピック t について各多様化指標で必要となるのは以下のものである。

- t に対する j 個目の特徴量における各検索意図の確率 $p_j(i|t)$
- 各画像ドキュメント d の j 個目の特徴量における検索意図 i に対する適合度 $r_j(d, i)$

今、各特徴空間で得られた GMM に基づいて、 $p_j(i|t)$ と $r_j(d, i)$ は以下のように表される。

$$p_j(i|t) = \pi_i \quad (8)$$

$$r_j(d, i) = p_j(i|d) = \frac{\pi_i \mathcal{N}(f_j(d)|\mu_i, \Sigma_i)}{\sum_{i \in \mathcal{I}_j} \pi_i \mathcal{N}(f_j(d)|\mu_i, \Sigma_i)} \quad (9)$$

5.1 I-D-nDCG の理想的リストに基づく多様化

ある画像が与えられたときに、各検索意図に対する適合度が GMM によって得られる。これを並べたものは検索意図を考慮した画像特徴量とみなすことができる。以下では、これを意図意識型特徴と呼ぶ。

$$F_{IA}(q) = (r_{col}(q, i_{col0}) \cdots r_{shp}(q, i_{shp0}) \cdots r_{tex}(q, i_{tex0}) \cdots)^T (10)$$

この特徴量の各要素は確率として求められており、この特徴量の比較を二つの特徴の内積で行った場合、2つの画像がいくつかの検索意図を共有しているかを測ることができる。

$$Sim_{IA}(q, d) = F_{IA}(q)^T F_{IA}(d) \quad (11)$$

一方で、この意図意識型特徴による検索の上位 K 件は D-nDCG[6] の理想的リストと対応する。D-nDCG は検索意図毎の多値適合度を扱うことのできる多様化指標である。今、I-nDCG が式 7 で表されるとする。この利得 $g(r)$ を以下のグローバル利得で置換えたものが D-nDCG として提案されている。

$$GG(r) = \sum_i p(i|q) g_i(r) \quad (12)$$

D-nDCG におけるクエリとドキュメントの適合度の定義より、



図 2 左から I-nDCG において高スコア (0.6 以上) の検索結果, 中スコア (0.4 以上 0.6 以下) の検索結果, 低スコア (0.4 以下) の検索結果. 赤い枠で囲われた画像がクエリである.

$$g_i(r) = p(i|d_r) \quad (13)$$

$$r_{I-D-nDCG}(q, d) = \sum_i p(i|q)p(i|d) \quad (14)$$

$$= \frac{1}{|\mathcal{F}|} \sum_{j \in \mathcal{F}} \sum_i r_j(q, i) r_j(d, i) \quad (15)$$

$$= \frac{1}{|\mathcal{F}|} Sim_{IA}(q, d) \quad (16)$$

5.2 I-ERR-IA の理想的リストに基づく多様化

Chapelle ら [18] に従うと, クエリのトピックが t であるとき, t に対する検索意図 i の確率 $p(i|t)$ を用いて, 検索意図毎にそれぞれの ERR を計算して確率でもって凸結合をとることで, IA-metrics は定義される.

$$ERR-IA@k = \sum_i p(i|t) ERR@k(i) \quad (17)$$

Agrawal ら [17], LEMMA 1 より IA metrics の下での検索意図の多様化は NP 困難であるため, 大規模データセットで最適解を見つけるのは一般に難しい. しかし, Agrawal ら [17] で提案されている貪欲法的なアプローチによる劣モジュラ最適化の二次近似 (IA-select) によって高品質な疑似理想的リストを得ることができることが知られている. すなわち, GMM で得られたドキュメントの各検索意図に関する適合度を用いて, IA-select を適用することで ERR-IA における疑似理想的リストを得られる. 本論文ではこれを多様化手法として提案し, 評価を行う.

6. 実験

6.1 予備実験

今回のような e コマース向けの詳細な画像検索データセットに関する多様性が有効であるかに関して, クラウドソーシングによって一部確認した. 240 人のワーカーに I-ERR-IA に基づく多様化を行った検索結果と単純に特徴 MPEG7 Visual Descriptors によって検索を行った検索結果を side-by-side で見せて, どちらが便利かを選ばせた. その際に, どのような理由でそれを選んだかについて

自由記述欄を設け, 意見を集計した. その際に得られた記述の中で検索結果の多様性を好むようなコメントを一度でも行ったユーザは 240 人中 42 名であった.

6.2 ユーザ実験

提案手法における I-D-nDCG の理想的リストと ERR-IA の疑似理想的リストによって得られる二つの検索結果の比較を行った. 各カテゴリから 10 個, 計 780 個のクエリに対してそれぞれの手法でもって検索結果を作成し, 二つの検索結果を side-by-side でワーカーに見せて, どちらかを選ぶタスクを行わせる. 一つの検索結果対に対して重複度 10 で投票を行わせた. ワーカーに実際の検索者の挙動を厳密に行わせることは難しいため, 三つの指示を与えてその観点でもってどちらがより使いやすいかについていずれかの検索結果に投票を行わせた. 指示は (a) クエリにより似ているものを探している場合 (b) クエリの色, 形, テクスチャのいずれかが気に入らない場合 (c) 様々な観点で商品を探したい場合の 3 つからなる. 合計 1,710 人のワーカーが参加した.

6.3 結果

この章では I-D-nDCG の理想的リストと I-ERR-IA の疑似理想的リストの比較を行う. I-D-nDCG の理想的リストと I-ERR-IA の疑似理想的リストの各指示ごとのユーザの好みとして一つの検索結果対に関して 10 人のユーザによる投票率を計測した. その投票率に関する箱ひげ図が図 3 である. 全ての指示に関して投票の平均値は I-D-nDCG の理想的リストの方が高いことがわかる. その投票数をクエリ毎に関する対応あり対応あり t 検定によって統計検定を行ったの結果が表 2 である. 全ての指示においてユーザは I-D-nDCG の理想的リストをより好んだが, 指示が多様性を好むものに変わるにつれてある程度 I-ERR-IA の疑似理想的リストが選ばれるようになることがわかった.

あるトピック t を持つクエリ q が与えられた時, 表示件数 K のもとで最適な半順序集合 $S = \{d_k | 1 \leq k \leq K, d_k \in \mathcal{D}\}$

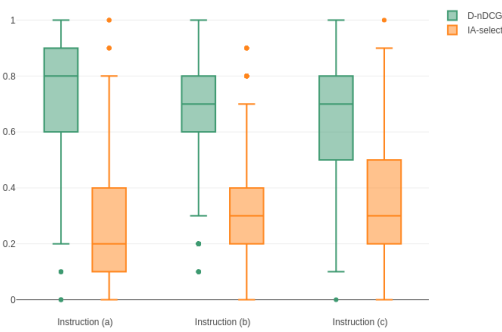


図 3 指示毎の手法における投票率

表 2 各指示 (a), (b), (c) における side-by-side 投票に関する I-D-nDCG の理想的リストの得票率に対する I-ERR-IA の得票率の対応あり t 検定 (99%信頼区間) の結果。

inst.(ia-dndcgr)	diff	lower	upper	p-value	effect size
(a)	-.435	-.470	-.400	<2e-16	-1.64
(b)	-.365	-.401	-.329	<2e-16	-1.29
(c)	-.307	-.342	-.273	<2e-16	-1.16

を求めることが問題となる。

6.4 考察

文書検索のクエリが一般に情報が不十分であるがゆえに曖昧であるのに対して、類似画像検索においては情報過多であるがゆえに曖昧である場合がある。一つは、クエリ画像の一部にユーザは興味があり他の部分は除きたいというケースがある。例えばユーザは画像に写る靴の色に興味があって、形は異なるものが欲しい場合にもこれをクエリとして使いたいとする。この時システムから見ると、クエリ画像が持つ検索意図のうちでどれに興味があってどれを無視したいかという部分において曖昧である。これは文書検索における曖昧クエリの問題と極めて異なる。そのため、クエリから検索意図が取り出されるとき、それらの検索意図とそうでない検索意図は別に扱う必要がある。検索意図に関する再現率を最大化する方法ではそれを考慮せずランキングしてしまうため、過度な多様化を行ってしまうことが考えられる。I-D-nDCG と I-ERR-IA の間にあるクエリの検索意図を考慮に入れるかという違いは、ある程度今回結果に反映されたと考えられる。

7. 結論

本論文では、検出用データセットの上に類似画像検索の評価指標を構築する方法を提案した。これによって、データセットに付与されたカテゴリにおいて、より詳細なカテゴリ内のランキング評価が可能になった。また、I-nDCG とユーザの好みとの一致性を大規模なクラウドソーシングによって確認し、統計的優越性を確認した。一方で、GMM で得られたコンポーネントを検索意図と見なすことで多様化を行うことが可能であることも同時に示し、二つの

多様化の手法を導出した。I-D-nDCG の理想的リストと I-ERR-IA の擬似的リストを side-by-side でユーザに投票させたところ、I-D-nDCG の理想的リストをユーザはより好んだ。

参考文献

- [1] Andrew Zhai, Dmitry Kislyuk, Yushi Jing, Michael Feng, Eric Tzeng, Jeff Donahue, Yue Li Du, Trevor Darrell: *Visual Discovery at Pinterest*, KDD (2017)
- [2] YuPing Hu, Hua Yin, Dezhi Han, Fei Yu: *The Application of Similar Image Retrieval in Electronic Commerce*, The Scientific World Journal (2014)
- [3] Girshick, Ross and Donahue, Jeff and Darrell, Trevor and Malik, Jitendra: *Rich feature hierarchies for accurate object detection and semantic segmentation*, CVPR (2014)
- [4] Ren, Shaoqing and He, Kaiming and Girshick, Ross and Sun, Jian: *Faster R-CNN: Towards real-time object detection with region proposal networks*, NIPS (2015)
- [5] Liu, Wei and Anguelov, Dragomir and Erhan, Dumitru and Szegedy, Christian and Reed, Scott and Fu, Cheng-Yang and Berg, Alexander C: *Ssd: Single shot multibox detector*, ECCV (2016)
- [6] Sakai, Tetsuya and Song, Ruihua: *Evaluating diversified search results using per-intent graded relevance*, SIGIR (2011)
- [7] Li, Piji and Zhang, Lei and Ma, Jun: *Dual-ranking for web image retrieval*, CIVR (2010)
- [8] Berg, Tamara L and Berg, Alexander C: *Finding iconic images*, CVPR (2009)
- [9] Deng, Jia and Dong, Wei and Socher, Richard and Li, Li-Jia and Li, Kai and Fei-Fei, Li: *Imagenet: A large-scale hierarchical image database*, CVPR (2009)
- [10] Lin, Tsung-Yi and Maire, Michael and Belongie, Serge and Hays, James and Perona, Pietro and Ramanan, Deva and Dollár, Piotr and Zitnick, C Lawrence: *Microsoft coco: Common objects in context*, ECCV (2014)
- [11] Philbin, J. and Chum, O. and Isard, M. and Sivic, J. and Zisserman, A.: *Object Retrieval with Large Vocabularies and Fast Spatial Matching* CVPR (2007)
- [12] Philbin, J. and Chum, O. and Isard, M. and Sivic, J. and Zisserman, A.: *Lost in Quantization: Improving Particular Object Retrieval in Large Scale Image Databases*, CVPR (2008)
- [13] Manjunath, Bangalore S and Salembier, Philippe and Sikora, Thomas: *Introduction to MPEG-7: multimedia content description interface*, IEEE Transactions on circuits and systems for video technology (2002)
- [14] Reynolds, Douglas: *Gaussian mixture models*, Encyclopedia of biometrics (2015)
- [15] Krizhevsky, Alex and Sutskever, Ilya and Hinton, Geoffrey E: *Imagenet classification with deep convolutional neural networks*, NIPS (2012)
- [16] Zauner, Christoph: *Implementation and benchmarking of perceptual image hash functions*, 2010
- [17] Agrawal, Rakesh and Gollapudi, Sreenivas and Halverson, Alan and Ieong, Samuel: *Diversifying search results*, WSDM (2009)
- [18] Chapelle, Olivier and Ji, Shihao and Liao, Ciya and Velipasaoglu, Emre and Lai, Larry and Wu, Su-Lin: *Intent-based diversification of web search results: metrics and algorithms*, Information Retrieval (2011)