

# キーワードによる画像検索における画像内容に基づくマイナス検索

吉川 英作<sup>†</sup> 田島 敬史<sup>†</sup>

<sup>†</sup> 京都大学大学院情報学研究科 〒606-8501 京都市左京区吉田本町  
E-mail: †yoshikawa.eisaku@dl.soc.i.kyoto-u.ac.jp, ††tajima@i.kyoto-u.ac.jp

**あらまし** 情報検索におけるマイナス検索とは、検索結果から除外したい解の条件をなんらかの形で指定するものである。例えば、キーワードによる文書検索では、このキーワードに関連するものは除外したいというキーワードを指定できる機能がしばしば提供される。キーワード検索は Web 上の画像検索でも用いられており、同じページ上の周辺に出現するキーワードを用いて画像を検索できる。しかし、画像検索においては、キーワードの不出現に基づくマイナス検索では不要解を十分に除外できない場合がある。そこで、本研究では、ユーザにはキーワードにより検索条件を指定させつつ、キーワードの不出現ではなく、画像内容に基づいて解の除外を行う画像のマイナス検索手法を提案する。提案手法では、クエリ「A -B」の解を、クエリ「A」の解からクエリ「A B」の解となる画像と類似するものを除去することで実現する。実験により、提案手法は Bing 画像検索より平均 0.279 高い precision@10 を実現し、Google 画像検索より平均 0.119 高い precision@10 を実現することを確認した。

**キーワード** 情報検索, 画像検索, マイナス検索

## 1 はじめに

インターネットに存在する情報は大量であり、情報検索により情報を取捨選択することが必要となる場面が多々ある。しかし、近年は情報検索の解集合も極めて巨大になり、何らかの手法で解の絞り込みを行うことが、効率的な情報収集には必要不可欠となっている。

検索の解を絞り込む手法に、検索オプションがある。よく知られた検索オプション例の一つに、マイナス検索が存在する。マイナス検索は除外検索とも呼ばれ、「A -B」なる検索クエリを用いることで、「A」なる検索クエリの解から B に関連する解を除外した解集合を取得できる。情報に溢れた現在にこそ、このマイナス検索を上手く利用し検索結果を絞り込むことは有用であると考えられる。

マイナス検索の有名な例に、「ジャガー -自動車」が存在する。例えば Google で「ジャガー」と検索すると、検索結果には自動車メーカーのジャガーに関するページが多数含まれる。そこで「-自動車」という文言をクエリに追加することで、自動車に関するジャガーを除外した検索結果を入手することができる。

しかし、このマイナス検索は画像検索では十分に機能しない場合がある。例えば Google 画像検索で「キーボード」と検索すると楽器のキーボードではなくパソコンのキーボードで検索結果が占められるが、「キーボード -PC」と検索してもパソコンのキーボードが検索結果に多数残ってしまう。これは画像検索におけるマイナス検索が、テキスト情報のみを利用して行っていることが原因であると考えられる。すなわち、「A -B」なる検索クエリによるマイナス検索は、「A」なる検索クエリによる検索の結果となる画像が含まれる web ページ内に B というワードが存在するかを検出し、その結果を基にフィルタリングを行うため、キーボードの例のように削除しきれない画像が頻出す

る恐れがある。特に、現代は SNS やマイクロブログといったテキスト情報の少ない web ページや画像がインターネット上に多数存在するため、例えばテキスト情報のみを利用する画像のマイナス検索手法には限界があると考えられる。

本研究では、テキスト情報のみを用いるのではなく、画像内容を用いてマイナス検索を行う手法を実現し、評価を行った。この提案手法は、「A -B」なる検索クエリで除去されるべき画像は、「A B」なる検索クエリの解と類似しているという仮定を基に考案した。すなわち、「A」なる検索クエリの解集合から「A B」なる検索クエリの解と類似するものを除去することで、「A -B」なる検索クエリを実現する手法である。ただし、本手法は「A -B -C」のような、除外する項目が複数のクエリには現時点では対応していない。

## 2 関連研究

本章では、本研究に関連する研究を紹介する。本研究は画像検索に関する研究であり、また検索結果の絞り込みや条件の詳細化を行う研究でもある。現在に至るまでの画像検索の研究を 2.1 節で、検索結果の絞り込みや条件の詳細化に関する研究を 2.2 節で紹介する。

### 2.1 画像検索

画像検索、特にコンテンツベース画像検索 (Content-based image retrieval (CBIR)) [1] の研究は、過去盛んに行われてきた [2] [3]。CBIR は画像検索手法の一つで、画像のメタデータ (キーワード、タグ、説明等) ではなく画像をそのコンテンツに基づいて検索するものである。

この CBIR の対義語として、コンセプトベース画像検索 [4] が存在する。これは説明ベースまたはテキストベースの画像検索とも呼ばれ、画像のメタデータを用いテキストベースの検索

を実行する検索手法である。既存のメジャーな検索エンジンは、画像のマイナス検索の除外機構にこのテキストベースの手法が用いられている [5]。本研究の提案手法では、この除外機構にコンテンツベースの手法を利用する。

CBIR の代表例に、VisualRank [6] が存在する。VisualRank は、コンテンツベースのリランキングを低い計算コストで実行するシステムである。画像間の視覚的類似性の分布から、画像集合の中に複数の視覚的テーマとその相対的強度を発見することで、画像をランキングする。

### 2.1.1 画像の自動タグ付け・画像キャプション生成

画像の自動タグ付けやキャプション生成の研究も、盛んに行われている [7] [8]。基本的に、検索システムは画像に付随するテキストを用いて画像検索を実現する。画像が出現するページのテキストも用いられるが、画像を説明するタグやキャプションも当然利用される。しかし、すべての画像に対してタグやキャプションが存在するわけではない。そうした画像に対して、画像の自動タグ付けや画像キャプション生成は有用である。

画像の自動タグ付けは、入力された画像に対しその画像に関連する単語を自動生成する機構である。また、画像キャプション生成は、入力された画像に対しその画像の内容を説明する文章を生成する機構である。一般にタグはキャプションよりも文章量を圧縮できるが、タグのみでは各単語の関連性が失われてしまう欠点も存在する。

これらの手法を用い画像の内容をテキストで表現することで、ある程度は画像のマイナス検索が実現できると考える。しかし、この手法ですべての考えるマイナス検索に対応することは、実用上不可能であるとも考える。マイナス検索に耐えうるタグやキャプションには、画像に登場する物体や画像の状況について記述される必要があり、また場合によってはアーティストや年代といったメタ情報も持たねばならない。これらの情報をすべて含んだタグやキャプションを考えると、実現できたとしてもそれは極めて大きなデータになり、検索システムという実行時間や実行空間に制限のあるサービスには適切ではない。

## 2.2 検索システム

マイナス検索と聞くと、最初に連想されるのは boolean 検索であろう。これは複数のキーワードを演算子により結合しクエリを作成する手法で、主に AND 演算と OR 演算が行われる。どちらも現行のメジャーな検索エンジンでは利用することができ、一般にスペースで単語を区切ることで AND 演算が、検索オプションの OR 検索を用いることで OR 演算が実行される。この boolean 検索の演算子の中に NOT 演算が存在し、こちらも現行のメジャーな検索エンジンではマイナス検索や除外検索として利用されている。boolean 検索の NOT 演算としてのマイナス検索の研究は過去に行われていた [9] が、画像検索におけるマイナス検索の研究は我々の知る限りは行われていない。

マイナス検索と類似する研究として、検索結果の絞り込みや条件の詳細化が考えられる。

### 2.2.1 クエリ拡張

文書検索において、ユーザが自身の求めている文書を適切に

指定するクエリキーワードが思い当たらない場合がある。そういったユーザに対し、元のクエリにキーワードを追加して検索性能を向上させる方法を、クエリ拡張と呼ぶ [10] [11] [12] [13] [14]。クエリ拡張は、当該ユーザの入力クエリや他ユーザの過去の検索履歴を利用して行われる。ユーザの意図と関連するキーワードを推薦することにより、ユーザは簡単なステップを踏むことで意図した検索結果に近い解を得ることができる。

クエリ拡張はマイナス検索と異なり対象を除外するわけではないが、例えば「キーボード」なるクエリを使用するユーザに対し『楽器』というキーワードを追加するよう推薦することにより、ユーザの検索意図に沿った画像を出力するよう誘導することが可能である。

ユーザが自身で入力した検索クエリを修正する状況を考える。ユーザがこうした操作を行うのは、元の検索クエリでは十分な検索結果が得られなかったからだと考えられる。すなわち、このユーザは不要な検索結果を見ながら検索クエリの修正を行っているため、直感的にはそれらの解が除去されるようなマイナス検索を行うことが自然だと思われる。しかし、同じ状況でクエリ拡張機能を利用し追加クエリの提案を行うことで、意図した検索結果に近い解を、マイナス検索を使わずに得ることが可能となる。

ただし、すべてのマイナス検索クエリがクエリ拡張によって実現可能となるわけではない。例えば、「日本 春 風景 -桜」なる検索クエリは、日本の春の風景を表現している画像を検索する際に「日本 春 風景」で検索すると桜の画像で上位が占められてしまうので、桜以外の日本の春の風景の画像を探すために「-桜」をつけたものと推測されるが、桜以外の日本の春の風景は菜の花や藤など様々なものがあるので、それら全体を表すクエリはクエリ拡張によって実現することは不可能である。

### 2.2.2 検索結果のクラスタリング

文書検索において、検索結果の表示は単に関連度の高いものから順に並べるだけで十分であるとは限らない。検索結果の表示画面やユーザが確認できるランキングの範囲は限られているため、似た内容の文書を上位に出さないことで検索結果に多様性を与えることがユーザの利便性に繋がる場合がある。

この考えを実現する 1 つの手法が検索結果のクラスタリング [15] である。検索結果の文書集合のうち類似文書をまとめて表示することで、検索内容に多様性が生じ、ユーザは検索意図に近い文書やそのクラスタを見つけやすくなる。

この手法もクエリ拡張同様マイナス検索のように特定の検索結果を除外するわけではない。しかし、2.2.3 項で述べる適合性フィードバック等と組み合わせることで、検索結果の除外が実現できると考えられる。

### 2.2.3 適合性フィードバック

適合性フィードバックはレバンスフィードバックとも呼ばれる手法で、一度検索結果をユーザに評価してもらい、その評価を元に検索に関するパラメータの重み等を修正し再度検索を行う手法である [16] [17]。この評価は何度も繰り返すことが可能であり、ユーザが評価する度に検索システムはユーザの検索意図を学習する。

適合性フィードバックはユーザの評価を入力として受け取る必要があるため、視覚的に評価しやすい画像検索に対して用いられることが多々ある。また、適合性フィードバックをクラスタリング等の検索結果を多様化させる手法と組み合わせることで画像のマイナス検索が実現できると考えられる。例えば、クラスタリングされた画像検索結果のうち特定のクラスタをユーザに選択させることで、そのクラスタに関連する画像を除外するマイナス検索が可能になる。

ただし、この手法は対話型の検索システムであることが前提であり、クエリのみで完結することが不可能な検索手法である。プログラム上の1ステップに組み込む場合など、一方向的かつクエリのみによる検索を行いたい状況は必ず存在するが、そういった場面でこの手法を用いることはできない。

### 3 検索意図によるマイナス検索クエリの分類

提案手法を説明する前に、まず、画像のマイナス検索にはどのようなクエリ意図のものがあるか分類を行った。

本章では、3.1-3.4節にて4通りの検索意図に基づくマイナス検索クエリの分類を紹介し、3.5節で各検索意図によるマイナス検索結果の適不適判定について説明する。

#### 3.1 多義語の分類クエリ

多義語の分類クエリは、多義語の画像のうち、ある特定の意味に関連する画像を消去する意図で用いられるクエリである。例えば「ジャガー」と検索するとGoogle画像検索ではジャガー社の自動車ばかり出現するが、もちろんジャガーには動物のジャガーが存在し、またギターや人名など、ジャガーの名を冠する対象は数多く存在する。動物のジャガーやフェンダー社のギターを調べたい状況で「ジャガー -自動車」なる検索クエリが用いられる。

#### 3.2 B を含まないクエリ

B を含まないクエリは、ある要素を画像内に含まない画像のみを表示する意図で用いられるクエリである。例えば「京都大学 時計台」と検索すると、クスノキを画面に含む特定アングルの画像が上位を占めてしまうが、もちろん時計台を写真に納める画角は他にも数多く存在する。クスノキを含まない画角の時計台の画像を調べたい状況で「京都大学 時計台 -クスノキ」なる検索クエリが用いられる。

#### 3.3 B が主題でないクエリ

B が主題でないクエリは、ある要素が主題ではない画像のみを表示する意図で用いられるクエリである。例えば「日本 春 風景」と検索すると桜の画像ばかりが上位に登場するが、日本の春の風景はもちろん桜だけではない。他の日本の春の風景、例えば菜の花や藤棚、雪解けなどを調べたい状況で「日本 春 風景 -桜」なる検索クエリが用いられる。ただし、この場合は3.2節のB を含まないクエリとは異なり、桜が主題でなければ含まれても問題ない。

クエリ	検索結果にほしいもの	除外してほしいもの
「ジャガー -自動車」		
「京都大学 時計台 -クスノキ」		
「日本 春 風景 -桜」		

図 1: 検索クエリの意図に基づく、適合解と不適合解の分類の例

### 3.4 メタ情報による削除クエリ

メタ情報による削除クエリは、画像に直接は表れない情報による削除を行う意図で用いられるクエリである。例えば「ダビデ像」と検索するとミケランジェロが作成した有名な彫像や、それをモチーフにしたものが多数出現するが、ダビデは彫像における一つのテーマであり、ミケランジェロ以外にもダビデ像を作成した芸術家は数多く存在する。そこでミケランジェロ以外が作成するダビデ像の画像を調べたい場合に「ダビデ像 -ミケランジェロ」なる検索クエリが用いられる。

### 3.5 各クエリ意図に基づく適合解と不適合解

3.1-3.4節では4通りの意図に基づくマイナス検索クエリの分類を行った。これらの検索クエリの意図が異なると、マイナス検索を用いて除去されるべき画像も異なってくる。

図1の下から1段目と2段目を比較する。最も下の段の検索クエリはBが主題でないクエリであり、中央の画像は桜が入っているが主題とはいえないため除去されないことが望ましい。一方で、下から2段目の検索クエリはBを含まないクエリであり、中央の画像はクスノキが入っているため除外されることが望ましい。

以下の章では、「A -B」なる検索クエリに対し、マイナス検索クエリの意図を考慮して除外されないことが望ましい画像を適合解、除外されることが望ましい画像を不適合解と呼ぶ。ただし、そもそも「A」なる検索クエリの検索結果として望ましくない画像は、マイナス検索クエリの意図に関係なく非想定解と定義する。

## 4 提案手法

本章では、提案する画像内容を用いた画像のマイナス検索の実現法を解説する。4.1-4.4節にて提案手法の各ステップの内容を記述し、4.5節にて提案手法の詳細と、提案手法の調整可能なパラメータや各ステップの計算の定義に関する他の選択肢について解説する。なお、4.1-4.4節では手順のみを説明し、各ステップにおいてその手法を用いた理由等は4.5節で解説する。5.5節の性能評価は、4.1-4.4節で使用したパラメータや計算の定義に基づいて行われる。性能評価で用いたパラメータや計算の定義は、5.2-5.4節での事前実験による評価を行い確定した。なお、5.2-5.4節での事前実験で用いたその他のパラ

表 1: 各変数の定義

記号	意味
$A$	「A」なる検索クエリの解集合
$a$	画像集合 $A$ の要素
$B$	「A B」なる検索クエリの解集合
$b$	画像集合 $B$ の要素
$v_x$	画像 $x$ の特徴ベクトル
$v_x^i$	$v_x$ の $i$ 次元の値
$l$	$v_x$ の次数
$\text{rank}_X(x)$	画像集合 $X$ のランキングにおける画像 $x$ の順位
$L_a$	$a$ と $B$ の距離
$L_A$	$\{L_a \mid a \in A\}$
$\theta$	フィルタリングに用いる閾値

メータや計算の定義は 4.5 節で紹介する。

簡単のため、以下では「A -B」なる検索クエリのマイナス検索を実現することを考える。ここで、 $A$  は単一の単語に限らず複数の単語からなる文字列であってもよいが、 $B$  は単一の単語あるいは単一の文に制限する。本手法は、除外する項目が一つの単語、あるいは一つの文である場合にのみ対応している。すなわち「A -B -C」のようなクエリには未対応である。

提案手法の解説のため、表 1 のように各変数を定義する。 $B$  は「B」なる検索クエリの解集合ではなく、「A B」なる検索クエリの解集合であることに注意されたい。

#### 4.1 特徴ベクトルの計算

各画像  $x \in A \cup B$  を ImageNet で学習済みの VGG16 [18] に入力し、FC2 層からの出力である 4096 次元ベクトルを画像の特徴ベクトル  $v_x$  として保存する。

#### 4.2 距離の計算

各画像  $a \in A$  に対し  $L_a$  を、4.1 節で計算した特徴ベクトルを基に

$$L_a = \min_{b \in B} \|v_a - v_b\|$$

で計算する。なお、ここでのベクトルのノルムは L4 ノルムを用い、

$$\|v_a - v_b\| = \left\{ \sum_{i=1}^l (v_a^i - v_b^i)^4 \right\}^{\frac{1}{4}}$$

なる計算を行う。

#### 4.3 閾値の決定

$t \in L_A$  に対し、

$$S = \{a \in A \mid L_a \leq t\}$$

$$T = \{a \in A \mid L_a > t\}$$

とおく。各  $t$  における  $S$  と  $T$  の分離度  $\text{sep}(S, T)$  を最大化させる  $t$ 、すなわち

$$\theta = \underset{t \in L_A, |S| \geq 10, |T| \geq 10}{\text{argmax}} \text{sep}(S, T)$$

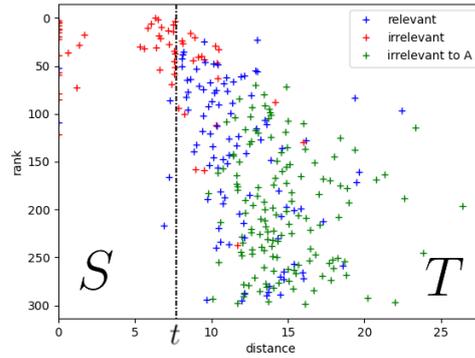


図 2: 距離  $t \in L_A$  と集合  $S, T$  の関係

を閾値と定める。

ここで、互いに素な画像集合  $X, Y$  に対し分離度  $\text{sep}(X, Y)$  は

$$\text{sep}(X, Y) = \frac{|X||Y|}{|X| + |Y|} \frac{\|\bar{v}_X - \bar{v}_Y\|^2}{\sum_{x \in X} \|v_x - \bar{v}_X\|^2 + \sum_{y \in Y} \|v_y - \bar{v}_Y\|^2}$$

と定める。ただし、

$$\bar{v}_X = \frac{1}{|X|} \sum_{x \in X} v_x$$

$$\bar{v}_Y = \frac{1}{|Y|} \sum_{y \in Y} v_y$$

である。

#### 4.4 フィルタリング

4.3 節の閾値  $\theta$  を基に  $A$  のうち  $L_a \leq \theta$  なる要素を除外する。すなわち、解集合は  $O = \{a \in A \mid L_a > \theta\}$  であり、 $o \in O$  は  $\text{rank}_A(o)$  が小さい順に出力する。

#### 4.5 提案手法の各ステップの詳細

本節では、提案手法の調整可能なパラメータや各ステップの計算の定義に関する他の選択肢について解説する。また、提案手法の各ステップによって画像のマイナス検索が実現可能となる理由についても記述する。これらの調整や選択が可能となる要素は、5.2-5.4 節での事前実験によって比較し、最適なパラメータおよび定義を推測する。

##### 4.5.1 CNN の選択

提案手法の特徴ベクトルの計算には、画像認識 Convolutional Neural Network (CNN) を利用した。画像認識 CNN の出力層付近のベクトルには画像の意味内容が表現されていると考えられており、本提案手法では画像の特徴ベクトル抽出に学習済み画像認識 CNN モデルを用いた。

5.2-5.4 節での事前実験では CNN として VGG16 [18] と ResNet-50 [19] を利用した。また、いずれの CNN も ImageNet による学習済みモデルを利用した。VGG16 は FC2 層からの出力である 4096 次元ベクトルを、ResNet-50 は avg.pool 層からの出力である 2048 次元ベクトルを、画像の特徴ベクトルとして利用する。

表 2: 3通りの距離の定義

名称	数式
重心からの距離	$L_a = \ v_a - \bar{v}_b\ $ , $\bar{v}_b = \frac{1}{m} \sum_{b \in B} v_b$
重み付き重心からの距離	$L_a = \ v_a - \bar{v}_b'\ $ $\bar{v}_b' = \sum_{b \in B} \frac{ B  - \text{rank}_B(b) + 1}{ B ( B  + 1)} v_b$
最短点からの距離	$L_a = \min_{b \in B} \ v_a - v_b\ $

#### 4.5.2 距離の定義

「A -B」なる検索クエリの解として望まれるものは、「A」なる検索クエリの解  $A$  のうち「A B」なる検索クエリの解  $B$  に類似する要素を取り除いたものである。各画像  $a \in A$  は3章と同様、「A -B」なる検索クエリの解として適当な画像を適合解、「A」なる検索クエリの解としては適当だが「A -B」なる検索クエリの解としては適当でない画像を不適合解、「A」なる検索クエリの解として適当でない画像を非想定解と3通りに分類される。この分類において、「A B」なる検索クエリの解集合  $B$  に含まれる画像は不適合解であると考えられる。

画像内容の類似度は画像の特徴ベクトル同士の距離で測定可能であるならば、 $a \in A$  が不適合解の場合は  $B$  との距離  $L_a$  が近くなり、 $a \in A$  が適合解か非想定解の場合は  $B$  との距離  $L_a$  が遠くなるのが期待される。そのため、この距離  $L_a$  を基に  $A$  を不適合解の集合と適合解及び非想定解の集合に分割することで、画像の除外を実現する。更に、本提案手法はこの分割を行った後、適合解が含まれる集合を元のランキングが高い画像を上位にランキングするよう検索解を出力することで、非想定解が上位に出力されないよう工夫している。

この各画像  $a \in A$  と画像集合  $B$  との距離  $L_a$  には、様々な定義が考えられる。事前実験では、この距離  $L_a$  の定義を表2のように3通り試した。

重心からの距離や重み付き重心からの距離で用いられるベクトル  $\bar{v}_b$  及び  $\bar{v}_b'$  は、不適合解の画像内容の普遍的な特徴を持つことが期待される。また、最短点からの距離は、不適合解の多くは  $\exists b \in B$  との距離が近く、適合解の多くは  $\forall b \in B$  との距離が遠い場合に良い性能が得られる。

#### 4.5.3 ノルムの定義

ノルムには様々な定義があり、本提案手法ではLP ノルムを用いた。これは、

$$\|x\|_p = \left\{ \sum_{i=1}^l x_i^p \right\}^{\frac{1}{p}}$$

なる式で定義されるノルムである。LP ノルムはユークリッド距離  $L_2$  が最も一般的だが、画像の特徴ベクトル同士の距離を考えた場合、ある一つの次元での差が大きい場合に距離も大きくなる性質が、画像内容の類似度の計算には適切である可能性がある。

#### 4.5.4 画像の枚数

「A」なる検索クエリの画像枚数  $|A|$  及び「A B」なる検索クエリの画像枚数  $|B|$  は様々な設定できる。一般に  $|A|$  や  $|B|$  を多くすれば幅広い画像を含む画像集合を作成できるが、非想定解を多数含んでしまう恐れがある。また、本実験では画像集

合を作成するツールとして Bing Image Search API を利用したが、このAPIの都合で本論文の実験では  $|A|$  及び  $|B|$  の最大値は300とする。

#### 4.5.5 分離度の定義

4.3節での分離度  $\text{sep}(S, T)$  は、大津の二値化[20]における分離度を提案手法に沿うよう修正したものである。

大津の二値化は判別分析法とも呼ばれる手法で、ある集合を自動で二値化する際に用いられる。スカラ値を持つ集合  $Z$  をある閾値  $t$  で2つの集合  $X, Y$  に分けた場合を考える。すなわち、

$$X = \{z \in Z \mid z \leq t\}$$

$$Y = \{z \in Z \mid z > t\}$$

とする。このとき、分離度をクラス内分散  $S_w^2(X, Y)$  とクラス間分散  $S_b^2(X, Y)$  の比

$$\frac{S_b^2(X, Y)}{S_w^2(X, Y)} = \frac{|X||Y|}{|X| + |Y|} \frac{(\bar{x} - \bar{y})^2}{\sum_{x \in X} (x - \bar{x})^2 + \sum_{y \in Y} (y - \bar{y})^2}$$

と計算し、この分離度が最大となる  $t$  を  $Z$  を二値化する閾値と定める。

提案手法の閾値の決定には、この式の分散をベクトルに拡張する。提案手法において、ある閾値  $t$  を用いて集合  $A$  を

$$S = \{a \in A \mid L_a \leq t\}$$

$$T = \{a \in A \mid L_a > t\}$$

と分けた場合を考える。フィルタリングを行う閾値は、 $S$  に不適合解が、 $T$  に適合解が多く含まれていることが望ましい。すなわち、クラス内分散  $S_w^2(S, T)$  が小さく、クラス間分散  $S_b^2(S, T)$  が大きい条件が望ましく、これは分離度  $\frac{S_b^2(S, T)}{S_w^2(S, T)}$  を最大化する問題と一致する。

自然な拡張を行うと、互いに素な画像集合  $X, Y$  に対し、クラス内分散  $S_w^2(X, Y)$  は

$$S_w^2(X, Y) = \frac{1}{|X| + |Y|} \left( \sum_{x \in X} \|v_x - \bar{v}_X\|^2 + \sum_{y \in Y} \|v_y - \bar{v}_Y\|^2 \right)$$

であり、クラス間分散  $S_b^2(X, Y)$  は

$$\begin{aligned} S_b^2(X, Y) &= \frac{|X|}{|X| + |Y|} \|\bar{v}_X - \bar{v}\|^2 + \frac{|Y|}{|X| + |Y|} \|\bar{v}_Y - \bar{v}\|^2 \\ &= \frac{|X||Y|}{(|X| + |Y|)^2} \|\bar{v}_X - \bar{v}_Y\|^2 \end{aligned}$$

となる。よって分離度  $\text{sep}(X, Y)$  は

$$\begin{aligned} \text{sep}(X, Y) &= \frac{S_b^2(X, Y)}{S_w^2(X, Y)} \\ &= \frac{|X||Y|}{|X| + |Y|} \frac{\|\bar{v}_X - \bar{v}_Y\|^2}{\sum_{x \in X} \|v_x - \bar{v}_X\|^2 + \sum_{y \in Y} \|v_y - \bar{v}_Y\|^2} \end{aligned}$$

となる。ここで、

$$\bar{v}_X = \frac{1}{|X|} \sum_{x \in X} v_x$$

$$\bar{v}_Y = \frac{1}{|Y|} \sum_{y \in Y} v_y$$

$$\bar{v} = \frac{1}{|X| + |Y|} \left( \sum_{x \in X} v_x + \sum_{y \in Y} v_y \right)$$

表 3: 事前実験に使用したクエリとその分類

クエリ名	クエリの種類	検索クエリ
daiwa	多義語の分類	ダイワスカーレット -ウマ娘
jaguar	多義語の分類	ジャガー -自動車
key	多義語の分類	キーボード -PC
brazil	B を含まない	ブラジル -サッカー
egypt	B を含まない	エジプト -ピラミッド
KU	B を含まない	京都大学 時計台 -クスノキ
disney	B が主題でない	ディズニー -ミッキー
japan	B が主題でない	日本 春 風景 -桜
panda	B が主題でない	上野動物園 -パンダ

である。

## 5 実験

本手法には 4.5 節で示したように、多くのパラメータや入れ替え可能な手法が存在する。これらの選択可能な要素を絞り込むために、5.2-5.4 節のような事前実験を行った。

事前実験 1 では、様々なパラメータで提案手法の距離の計算までを実行し、 $L_a$  と  $rank_A(a)$  の分布から高い性能の見込めるパラメータを選定した。また、事前実験 2 では、5.2 節の事前実験で絞ったパラメータで提案手法の距離の計算までを実行し、各距離に閾値を設定した場合の適合率や再現率のグラフからパラメータのさらなる絞り込みを試みた。事前実験 3 では、5.2, 5.3 節の事前実験で絞ったパラメータで提案手法の閾値の決定までを実行し、分離度や precision@10 のグラフから最終的なパラメータの選択を行った。

5.5 節で、5.2-5.4 の事前実験を基に定めたパラメータでの提案手法の性能評価を行う。提案手法と比較する既存手法には、Bing 画像検索及び Google 画像検索を利用した。更に、5.6 節では、性能評価実験の結果を踏まえ追加実験を実行した。

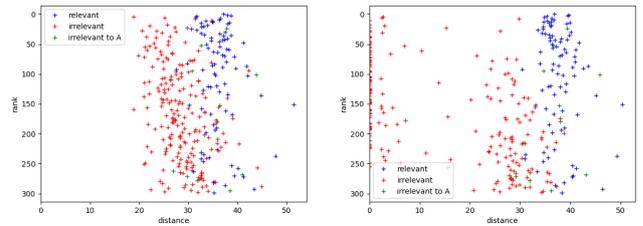
### 5.1 データセット

本実験では日本語のクエリを 32 種類、それらを英訳したクエリを 32 種類、合計 64 種類のマイナス検索クエリを用意した。各クエリ「A -B」に対し Bing Image Search API を利用し画像の収集を行い、「A」なる検索クエリの解 A と「A B」なる検索クエリの解 B を収集し、データセットを作成した。また、画像の収集と同時に各画像集合  $X$  のランキングにおける画像  $x$  の順位  $rank_X(x)$  も保存する。 $|A| = 300$  で収集し、 $|B| = 100, 300$  の 2 通りで収集した。なお、日本語クエリは国コードを jp-JP、英語クエリは国コードを en-US に指定して画像の収集を行った。

また、「A」なる検索クエリの解については、3.5 節の基準に則り適合解、不適合解、非想定解の 3 種類にラベリングも行った。このラベリングを元に提案手法の性能を評価する。

### 5.2 事前実験 1: 散布図の作成

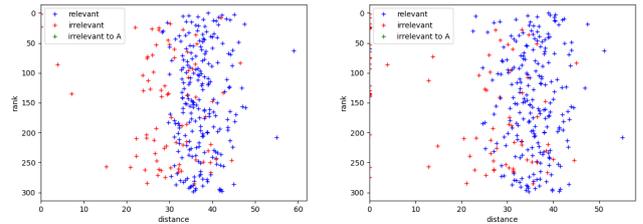
4.5 節の通り、本提案手法には調整可能な要素が多数存在する。故に、まずは提案手法を部分的に実行しその結果を分析することで、一部のパラメータを確定する事前実験を行った。



(a) 重心からの距離

(b) 最近点からの距離

図 3: 事前実験 1 散布図を距離の定義で比較 (daiwa クエリ, L2 ノルム,  $|B| = 300$ , ResNet-50)



(a)  $|B| = 100$

(b)  $|B| = 300$

図 4: 事前実験 1 散布図を  $|B|$  の数値で比較 (panda クエリ, 最近点からの距離, L2 ノルム, ResNet-50)

本節の実験では、表 3 の 9 種類のクエリに対し、提案手法の距離の計算までを行い、図 3 や図 4 のような散布図を作成する。各図の + は画像  $a \in A$  を表し、青色は適合解、赤色は不適合解、そして緑色は非想定解を表す。また、横軸が各画像  $a$  の距離  $L_a$  を、縦軸が各画像  $a$  の元のランキング  $rank_A(a)$  を表す。

提案手法では、この散布図のある距離を閾値とし、その閾値より距離の大きな画像、すなわち閾値より右側の画像を  $rank_A(a)$  の小さい順に出力する。本節の事前実験では、適合解の集合と不適合解の集合がうまく分割される縦線が引ける散布図が描画されるパラメータを調べる。

距離の定義は 4.5.2 項で述べた 3 通りの方法を試す。また、ノルムの取り方は L2, L4, L8 ノルムの 3 通りを試す。画像の特徴ベクトル抽出には VGG16 と ResNet-50 の両方を用い、 $|A| = 300$  に固定し  $|B| = 100, 300$  の 2 通りで実験を行った。すなわち各クエリに対して 36 通りのパラメータで提案手法を実行する。

この事前実験により、複数の事実が判明した。

まず、重心からの距離と重み付き重心からの距離は、どちらも似た散布図を描くことがわかった。すなわち、 $\{v_b \mid b \in B\}$  の重心と重み付き重心は、クエリによらず近い値をとる。

また、重心からの距離や重み付き重心からの距離よりも、最近点からの距離の方が適合解と不適合解を分離できることがわかった。図 3 は、距離の定義以外の条件を同じにした場合の散布図である。重心からの距離の場合一定の距離未満のノードは存在しないが、最近点からの距離の場合は距離 0 をとるノードも複数存在する。このことから、特徴ベクトルの重心を取る操作では、画像内容の普遍的な特徴をもつベクトルの作成は不可

能だとわかった。

次に、最短点からの距離の場合、 $|B| = 100$  よりも  $|B| = 300$  の方が適合解と不適合解をよりよく分離できることもわかった。 $|B|$  の数値を増やすと、一部の不適合解の距離は小さくなる。図 4 の例では、 $|B| = 300$  の図の縦軸に多数の不適合解が集まっている。これは、 $B$  の要素数を増やすことで、不適合解  $x \in A$  と類似する画像を  $B$  に含むことができ、 $L_x$  を小さくすることが可能になるからだと考えられる。

同様に、特徴ベクトル抽出に利用する CNN や LP ノルムの選択による散布図の変化も確認できた。ただし、これらの最適な組合せについて、本節の実験で結論づけることはできなかった。

以上のことから、提案手法において距離の定義は最短点からの距離が、画像の枚数は  $|B| = 300$  が適切であるとわかる。

### 5.3 事前実験 2：階段状グラフの作成

5.2 節の事前実験で距離の定義及び  $|B|$  の数値を決定したため、本節の事前実験ではそれらのパラメータは固定し、その他のパラメータを変更した際の比較を行った。

本実験では、日本語クエリ 32 種類に対し、提案手法の距離の計算までを行い、各距離に閾値を設定した場合の適合率、再現率、削除率、precision@10 について、図 5 のような階段状グラフを作成する。各図は緑線が適合率、青線が再現率、赤線が削除率、黒線が precision@10 であり、横軸が閾値の距離、縦軸がスコアを表す。削除率は距離  $t$  に閾値を設定された場合、 $A$  の集合のうちどれだけの割合がフィルタリングされるかを表すスコアで  $\frac{|\{a \in A | L_a \leq t\}|}{|A|}$  なる式で計算される。基本的には、precision@10 の最大値及びそのときの削除率を基に、各パラメータにおける提案手法の性能を評価した。

画像の特徴ベクトル抽出には VGG16 と ResNet-50 の両方を用い、ノルムは  $L2, L4, L8$  ノルムの 3 通りで実験を行った。すなわち各クエリに対して 6 通りのパラメータで提案手法を実行する。

この事前実験により、複数の事実が判明した。

図 5 は「東京大学 - 安田講堂」なる検索クエリで作成した階段状グラフである。削除率が極端に高い部分を除き、削除率に対する適合率と再現率の推移はいずれのグラフでも似通っている。このように、各クエリにおいて、削除率に対する適合率と再現率の推移は、LP ノルムによる大きな影響を確認できなかった。また、一部のクエリを除き、削除率に対する適合率と再現率の推移は、用いる CNN による大きな影響も確認できなかった。

しかし、precision@10 の振る舞いは用いるパラメータによって異なっている。図 5 を確認すると、precision@10 の最大値はパラメータの組み合わせによって異なり、このクエリにおいては CNN は ResNet-50 を用い、ノルムは  $L4$  ノルムあるいは  $L8$  ノルムを用いる手法が最適であると考えられる。

32 種類のクエリの削除率に対する precision@10 のスコアの振る舞いを確認すると、CNN による性能差が認められたクエリは表 4 の 7 種類存在し、7 種類中 VGG16 がより優れているクエリは 5 件、ResNet-50 がより過ぎれているクエリは 2 件で

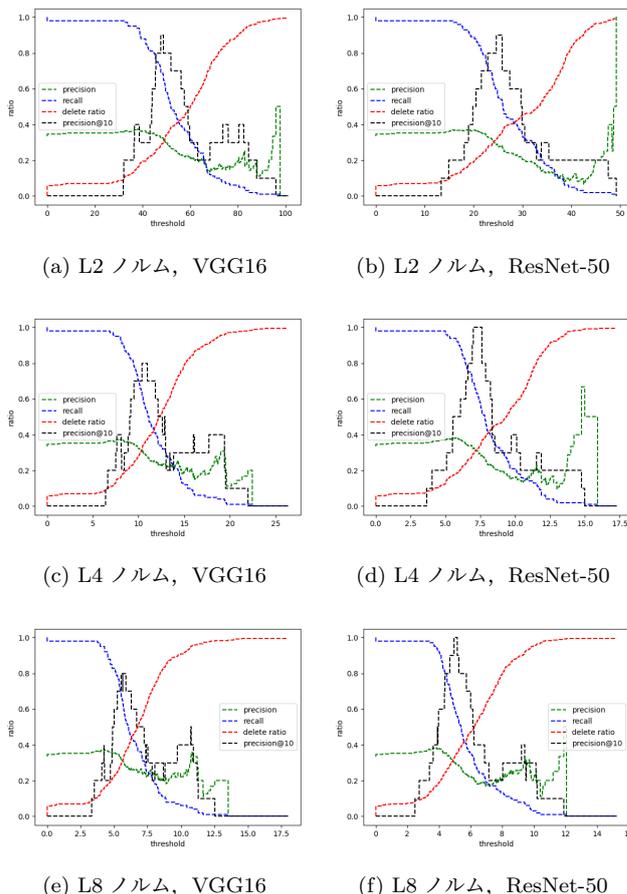


図 5: 事前実験 2 階段状グラフを LP ノルム及び CNN の選択で比較 (UT クエリ)

表 4: 事前実験 2 CNN の選択により precision@10 の最大値が変化するクエリ

クエリ名	VGG16	ResNet-50
expo	✓	
mouse	✓	
olympic	✓	
tokyo	✓	
torii		✓
UT		✓
venus	✓	

ある。また、LP ノルムによる性能差が認められたクエリは 6 種類存在し、 $L4$  ノルムはこの 6 種類全てにおいて高い性能を実現できた。

以上の結果、ベクトルのノルムには  $L4$  ノルムを採用するのが最も適切であると導かれる。一方で、どちらの CNN が優れているかはクエリによって様々であり、本設の事前実験で確定することはできない。

また、高い precision@10 の解集合を出力することが提案手法では不可能なクエリの存在も確認できた。図 6 上側の precision@10 のスコアからわかるように、この二つのクエリに対してはいかなる閾値をもってしても、高い precision@10 を出力することは不可能である。precision@10 を高くすることが不可

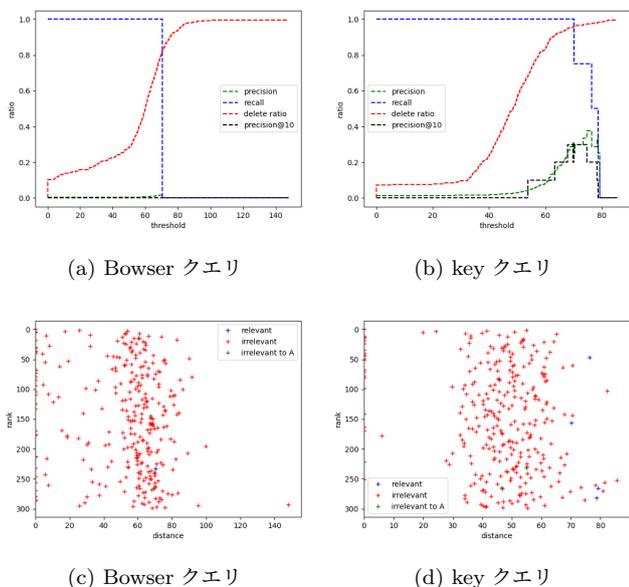


図 6: 事前実験 2 提案手法では高い precision@10 を実現できないクエリのグラフ (L2 ノルム, VGG16)

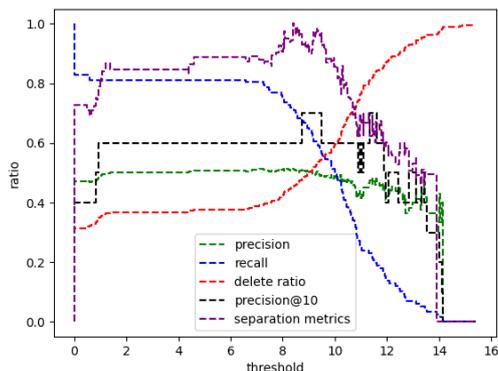


図 7: 事前実験 3 階段状グラフの例 (japan クエリ, VGG16)

能である原因は元の画像集合  $A$  にある. 図 6 下側のように, 元の検索解に適合解が極端に少ない場合, 本論文の提案手法は十分な性能を発揮できない.

以上のことから, 提案手法においてノルムは L4 ノルムが最も汎用的であるとわかる. また, 特徴ベクトル抽出に用いる CNN は, ResNet-50 よりも VGG16 の方が precision@10 の最大値が大きくなるクエリが多い傾向が確認された. 更に, 元の検索解に適合解が極端に少ない場合, 提案手法が十分な性能を発揮できないことも認められた.

#### 5.4 事前実験 3: 閾値の決定機構

5.2 節及び 5.3 節の事前実験では, 距離の定義,  $|B|$  の数値, そしてノルムの確定を行った. そのため, 本節の事前実験ではそれらのパラメータは固定し, 特徴ベクトル抽出に用いる CNN を決定することを考える.

本実験では, 表 3 の 9 種類のクエリ及び表 4 の 7 種類のクエリの合計 16 種類の日本語クエリに対し, 提案手法の閾値の

表 5: 事前実験 3 各クエリの precision@10

	VGG16	ResNet-50
brazil	1.0	1.0
daiwa	0.9	1.0
disney	1.0	1.0
egypt	0.8	0.9
jaguar	1.0	0.9
japan	0.7	0.7
key	0.0	0.0
KU	0.8	0.7
panda	1.0	1.0
expo	0.2	0.2
mouse	0.7	0.7
olympic	0.7	0.7
tokyo	0.6	0.6
torii	0.7	0.6
UT	0.7	0.3
venus	0.7	0.3
平均値	0.72	0.66
0.6 以上	14	12

決定までを行い, 各距離に閾値を設定した場合の適合率, 再現率, 削除率, precision@10, そして分離度について, 図 7 のような階段状グラフを作成する. この図は, 5.3 節の事前実験に, 各閾値での分離度を載せたものである. 緑線が適合率, 青線が再現率, 赤線が削除率, 黒線が precision@10, そして紫線が分離度であり, 横軸が閾値の距離, 縦軸がスコアを表す. ただし, 分離度は最大値が 1 となるよう正規化を行っている. 提案手法の閾値は分離度が最大となる距離と決められるため, その距離での precision@10 と削除率のスコアから, 各種パラメータを評価する.

画像の特徴ベクトル抽出には VGG16 と ResNet-50 の両方を用いる. すなわち各クエリに対して 2 通りのパラメータで提案手法を実行する.

16 種類のクエリを各パラメータで実行した場合の, 分離度が最大となる距離での precision@10 のスコアを表 5 にまとめる.

まずは表 4 の 7 種類のクエリの precision@10, すなわち表 5 の中央下 7 行を確認されたい. 5.3 節の事前実験では, これらのクエリのうち torii 及び UT は ResNet-50 が VGG16 より precision@10 の最大値が高く, その他 5 つのクエリは VGG16 が ResNet-50 より precision@10 の最大値が高かった. しかし表 5 の結果から, torii については VGG16 も ResNet-50 と同程度の precision@10 であり, UT については VGG16 が ResNet-50 よりも高い precision@10 となる. 表 3 のクエリにおいても VGG16 は ResNet-50 と同程度かそれ以上の precision@10 をとるため, 提案手法の性能評価には VGG16 を使うこととした.

以上の実験結果より, 特徴ベクトル抽出には VGG16 を用いることが適切である.

#### 5.5 提案手法の性能評価

5.2-5.4 節の事前実験で確定した各種パラメータを選択した

表 6: 各手法の precision@10 の平均値と MRR

	提案手法	Bing 画像検索	Google 画像検索
precision@10 の平均値	0.680	0.413	0.561
MRR	0.688	0.470	0.601

表 8: 各手法の precision@10

	Bing 画像検索	提案手法 (B)	Google 画像検索	提案手法 (G)
lens	0.2	0.7	0.8	0.1
en.lens	0.0	0.0	0.7	0.9

表 7: precision@10 で評価した提案手法と既存手法の比較 (クエリ数)

	Bing 画像検索	Google 画像検索
提案手法 > 既存手法	40	29
提案手法 = 既存手法	17	18
提案手法 < 既存手法	4	17

提案手法を実行し、既存手法との比較を行った。提案手法は 4 章の 4.1-4.4 節で述べた通りの手順で行われる。性能評価には precision@10 及び MRR (Mean Reciprocal Rank) を用いる。

比較する既存手法には Bing 画像検索及び Google 画像検索を利用した。既存手法の各クエリにおけるマイナス検索の precision@10 及び MRR は、日本語クエリは言語を日本語、地域情報を日本で検索した結果から、英語クエリは言語を English、地域情報を U.S. で検索した結果から測定した。

表 6, 7 に提案手法と既存手法の性能比較をまとめる。ただし、Bing 画像検索の日本語クエリの一部は解が得られなかったため、表 6, 7 での Bing 画像検索との比較では、当該の 3 種類のクエリを取り除いて計算を行った。

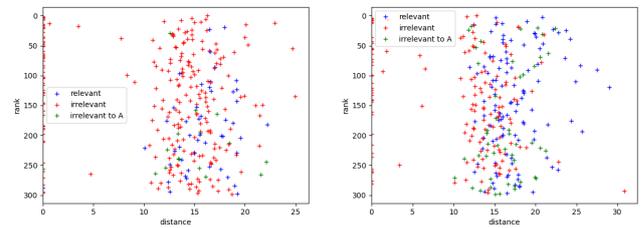
表 6 の通り、提案手法は 2 つの既存手法と比較して高い precision@10 と MRR を実現することがわかる。表 7 から、precision@10 で比較すると、提案手法を用いたマイナス検索は Bing 画像検索のマイナス検索に 61 種類のクエリのうち 40 種類のクエリで勝り 4 種類のクエリで劣る。また、提案手法を用いたマイナス検索は Google 画像検索のマイナス検索に 64 種類のクエリのうち 29 種類のクエリで勝り 17 種類のクエリで劣る。

以上の結果から、提案手法は既存手法のマイナス検索より高い精度のマイナス検索を実現できることが示された。

## 5.6 Google 画像検索結果をデータセットとした提案手法の性能評価

5.5 節の性能評価では、提案手法は既存手法と比較して平均して高い性能を実現できた。しかし、一部のクエリでは Google 画像検索に対して著しく低い性能となった。ただし、それらのクエリの多くは提案手法と同様に Bing 画像検索も Google 画像検索よりも低い性能となっている。本実験はデータセットとして Bing Image Search API を利用しているため、Google と提案手法ではデータセットとして用いている画像が異なり、性能差が出ていると考えられる。そこで、本節では一部のクエリに対し Google 画像検索結果の画像によるデータセットを作成し、提案手法の性能評価を行った。

Bing と異なり Google は多数の検索結果の画像を収集する API が存在しないため、手動で収集を行った。そのため、今回は一つの日本語クエリとその英訳クエリの、合計 2 種類のクエ



(a) 「レンズ-カメラ」

(b) 「lens-camera」

図 8: Google 画像検索結果をデータセットとした提案手法

リのみでの性能評価となった。クエリには、5.5 節の性能評価で Google 画像検索が提案手法や Bing 画像検索より高い性能を実現した、「レンズ-カメラ」なるクエリを用いる。すなわち、「レンズ」なる検索クエリの画像と「レンズ-カメラ」なる検索クエリの画像をそれぞれ 300 枚ずつ収集した。また、同様の操作を、英訳クエリ「lens-camera」に対しても行った。

実行結果をまとめたものが表 8 となる。このように、英語クエリは十分高い precision@10 を実現できたが、日本語クエリは高い precision@10 を実現できなかった。

提案手法が「レンズ-カメラ」をうまく実現できなかった理由は、データセットに適合解が著しく少ないためであると考えられる。図 8 の散布図を確認すると、英語クエリでは多数の適合解が存在し分離も行えているが、日本語クエリには適合解が極めて少なく、また分離も十分に行えていない。また、Google 画像検索結果を確認すると、「レンズ-カメラ」なる検索クエリの解の上位 10 件に存在する適合解 8 枚のうち、「レンズ」なる検索クエリの解の上位 300 枚に含まれるものは 3 枚のみである。すなわち、Google による「A」なる検索クエリの画像検索結果の上位を収集し作成したデータセットは、Google が画像検索「A-B」なる検索クエリを実行する際に用いているデータセットと同じとは言えない。そのため、日本語クエリに対して正当な評価は不可能である。

クエリが少ないため一般化はできないが、英語クエリの性能向上から、以前の実験で十分な性能が実現できなかったクエリに関しては、データセットに適合解が少ないため発生していると考えられる。この問題を解決するために、Google 画像検索結果をデータセットとして提案手法を実行する際は、画像枚数  $|A|$  をより増やすことが必要だと考えられる。

## 6 結 論

本論文では、画像のマイナス検索を行う新たな手法を 4 章で提案し、5.5 節で性能評価を行った。既存手法である Bing 画像検索及び Google 画像検索と比較して、提案手法は precision@10 及び MMR について高い性能を実現することが示された。本論

文での性能評価実験は Bing Image Search API を用いて作成したデータセットを利用して計測したため、似たデータセットを使用していると思われる Bing 画像検索に対して高い性能を実現したことは、提案手法が Bing 画像検索よりも優れた画像のマイナス検索を実現できていることを意味する。また、Google 画像検索との正しい比較を行うために、Google 画像検索が用いていると思われるデータセットを作成し比較することが望まれる。

提案手法は高い性能を実現できたが、さらなる改良が考えられる。

まず、画像の特徴ベクトル抽出には ImageNet で学習済みの VGG16 を用いたが、よりマイナス検索に特化した学習を行えばその分性能は向上すると考えられる。特徴ベクトル抽出に VGG16 以外の CNN を試すことや CNN 以外の手法で画像の特徴ベクトル抽出を行うことも、性能に寄与することが予想される。

また、閾値の決定機構を別のシステムに置き換える改良も考えられる。提案手法では大津の二値化を拡張した機構を利用したが、より最新の技法も試すことも望まれる。

その他にも、提案手法には時間計算量の問題が存在する。提案手法の時間計算量は、 $O(|A||B|l)$  である。VGG16 は  $l = 4096$  であり、 $|A| = 300$ ,  $|B| = 300$  で実行すると、情報検索の実行時間としては少し結果が出るまでに時間がかかりすぎる。そのため、より低次元の特徴ベクトルを利用することや、ボトルネックとなっている距離の計算を軽量化することは重要である。

## 謝 辞

本研究は JSPS 科研費 21H03446 の助成を受けたものです。

## 文 献

- [1] Sahil Jain, Kiranmai Pulaparathi, and Chetan Fulara. Content based image retrieval. *Int. J. Adv. Eng. Glob. Technol.*, Vol. 3, pp. 1251–1258, 2015.
- [2] Ritendra Datta, Dhiraj Joshi, Jia Li, and James Z Wang. Image retrieval: Ideas, influences, and trends of the new age. *ACM Computing Surveys (Csur)*, Vol. 40, No. 2, pp. 1–60, 2008.
- [3] V.N. Gudivada and V.V. Raghavan. Content based image retrieval systems. *Computer*, Vol. 28, No. 9, pp. 18–22, 1995.
- [4] Hsin-Liang Chen and Edie M Rasmussen. Intellectual access to images. 1999.
- [5] Google. Refine web searches, 2023. <https://support.google.com/websearch/answer/2466433>.
- [6] Yushi Jing and Shumeet 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.
- [7] Benjamin Z. Yao, Xiong Yang, Liang Lin, Mun Wai Lee, and Song-Chun Zhu. I2t: Image parsing to text description. *Proceedings of the IEEE*, Vol. 98, No. 8, pp. 1485–1508, 2010.
- [8] Oriol Vinyals, Alexander Toshev, Samy Bengio, and Dumitru Erhan. Show and tell: A neural image caption generator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015.
- [9] Shmuel T. Klein. On the use of negation in boolean ir

- queries. *Information Processing & Management*, Vol. 45, No. 2, pp. 298–311, 2009.
- [10] Joseph John Rocchio Jr. Relevance feedback in information retrieval. *The SMART retrieval system: experiments in automatic document processing*, 1971.
- [11] Efthimis N Efthimiadis. Query expansion. *Annual review of information science and technology (ARIST)*, Vol. 31, pp. 121–87, 1996.
- [12] Makoto P. Kato, Tetsuya Sakai, and Katsumi Tanaka. Structured query suggestion for specialization and parallel movement: Effect on search behaviors. In *Proceedings of the 21st International Conference on World Wide Web, WWW '12*, p. 389–398, New York, NY, USA, 2012. Association for Computing Machinery.
- [13] Jinxi Xu and W. Bruce Croft. Query expansion using local and global document analysis. *SIGIR Forum*, Vol. 51, No. 2, p. 168–175, aug 2017.
- [14] Claudio Carpineto and Giovanni Romano. A survey of automatic query expansion in information retrieval. *ACM Comput. Surv.*, Vol. 44, No. 1, jan 2012.
- [15] Marti A Hearst and Jan O Pedersen. Reexamining the cluster hypothesis: Scatter/gather on retrieval results. In *Proceedings of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, pp. 76–84, 1996.
- [16] Yong Rui, T.S. Huang, M. Ortega, and S. Mehrotra. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 8, No. 5, pp. 644–655, 1998.
- [17] Xiaoou Tang, Ke Liu, Jingyu Cui, Fang Wen, and Xiaogang Wang. Intentsearch: Capturing user intention for one-click internet image search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 34, No. 7, pp. 1342–1353, 2012.
- [18] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2014.
- [19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.
- [20] 大津展之. 判別および最小 2 乗規準に基づく自動しきい値選定法. 電子情報通信学会論文誌 D, Vol. 63, No. 4, pp. 349–356, 1980.