

# Web テキストを利用した画像クラスタリングシステム

90132109 谷内田研究室 永田晶子

## 1. はじめに

我々がインターネットを通してアクセスできる画像データの量は年々増加しており、必要な画像を効率良く得るための画像検索技術が極めて重要になってきている。そこで本研究では、画像検索結果をクラスタリングし、ラベル付きクラスタの形で提示するシステムを提案する。

## 2. システムの概要

本システムの処理の流れとしては、まず画像収集部によって画像群の収集を行い、次に画像ラベリング部で収集された各画像に複数のラベルを付与する。その後、画像クラスタリング部により先に付与したラベルに基づき画像のクラスタリングを行い、最後に、クラスタリング結果出力部により結果の出力を行う。以下に各部の詳細を示す。

### 2.1 画像収集部

Google イメージ検索により、WWW から検索キーワードに関する画像を含むHTMLテキストと画像ファイルを取得する。

### 2.2 画像ラベリング部

HTML テキストの重要文を用いた画像ラベリングシステム [相良 04]を用いて、取得した画像に複数のラベルを付与する。

### 2.3 クラスタリング部

同じラベルが付いた画像を、重複を許可してまとめ、クラスタを作成する。そして、2つの評価モジュールでクラスタのラベルに評価値を与え、検索キーワードと関連の深いラベルが付いたクラスタを選択する。

#### 2.3.1 複合語による評価モジュール

検索キーワードとラベルの複合語が存在するラベルは、検索キーワードに深く関連する。そこで複合語が存在するか否かを、複合語を検索キーワードとした場合の検索エンジンの hit 数で評価し、この値が 1000 を越えたラベルを採用する。

$$phrase(keyword, label) = \max_{i=1,2} hit(P_i) \quad (1)$$

ただし、 $hit(P_1)$ は複合語"keyword.label"の hit 数、 $hit(P_2)$ は複合語"label.keyword"の hit 数である。

#### 2.3.2 共起性による評価モジュール

ラベルと検索キーワードの条件付き共起確率を検索エンジンの hit 数で近似した評価値を導入する。この値が高ければラベルと検索キーワードの関連性が深いといえる。

$$connection(keyword, label) = \frac{hit(keyword \square label)}{hit(label)} \quad (2)$$

phrase で採用したラベルと評価値 connection の上位ラベルを合わせて、最終出力ラベルとする。

### 2.4 クラスタリング結果出力部

クラスタリング結果出力部では前節の評価値により決定したラベルとそのラベルを持つ画像クラスタをブラウザに出力する。出力インタフェースを図 1 に示す。

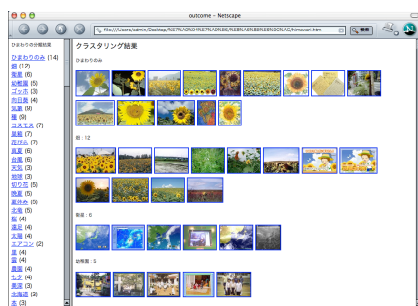


図 1 出力インタフェース

## 3. 実験

クラスタリングやラベル提示の、画像検索における有効性を検証するため、評価実験を行った。実験方法は Google イメージ検索で用いられている画像提示、本システムによって作成したラベルなしクラスタでの画像提示、本システムによって作成したラベル付きクラスタでの画像提示の 3 システムで検索キーワードに関する画像集合を提示し、その中から課題の条件に一致する画像を見つけるのに要する時間を測定する。これを被験者 18 名に対して、表 1 に示す 9 つの課題について行った。

表 1 検索キーワードと課題

検索キーワード	課題
映画	映画の広告・ポスター
ケーキ	チョコレートケーキ
ジュース	りんごジュース
癒し	癒し空間
ファッション	服がメインでうつつている画像
阪神	阪神タイガースの選手
ひまわり	花のひまわりの絵
宇宙	宇宙空間に地球がある画像
携帯電話	カメラ付き携帯電話

### 3.1 結果と考察

図 2 に画像を見つけるまでの平均所要時間を示す。ラベルなしクラスタの結果の方が Google イメージ検索の結果より全体的に平均所要時間が短く、良い結果が得られた。これは、検索結果をクラスタ表示にすることによって全画像を見る必要がなくなり、クラスタ内の画像を 2, 3 個見て、目標画像に関連がないクラスタを飛ばすことができるためであると考えられる。また、ラベル付きクラスタの結果がラベルなしクラスタの結果よりも平均所要時間が短く、全体的に良い結果が得られたのは、クラスタにラベルがついている場合、目的の画像と明らかに関連が無いラベルがついているクラスタは、そのクラスタ内の画像を見る必要がなくなり、その分時間が短縮されたためであると考えられる。これらの実験結果によりクラスタリングとラベルの有効性が示せたといえる。

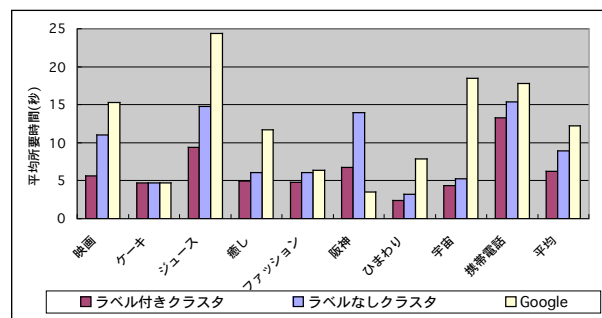


図 2 画像を見つけるまでの平均所要時間

## 4. まとめ

本研究では、Web ページ中の画像をクラスタリングしてラベル付きで提示するシステムを提案し、その有効性を検証した。本システムでは検索キーワードとクラスタのラベルの関連性の深さを測る事で、検索キーワードに対して適切なラベルの付いたクラスタのみを提示する事ができる。今後の課題として、出力インタフェースの改善が挙げられる。