

可変なカテゴリ構造を用いた文書検索支援手法の実験的評価

岩崎正秀 仲川こころ 高田喜朗 関浩之

奈良先端科学技術大学院大学 情報科学研究科

WWW で提供される情報が加速度的に増加するに伴い、より多くのユーザが WWW 検索サービスを必要とするようになりつつある。そのため、必要な情報を探し出すために専門的な知識や技術を必要としないような、使いやすい検索方法の実現が重要になってくる。既存の検索サービスであるディレクトリサービスは、データベース中の文書を分類したカテゴリ階層構造を提供するため、分類が明示されていて使いやすい。しかし、カテゴリ構造が固定されているため、ユーザの要求に柔軟に対応できない。そこで本研究では、検索のたびに、その時のユーザの意図に近い小規模で可変なカテゴリ階層構造を提供する手法を提案する。また、この手法を実装した試作システムを用いて行った評価実験について述べる。

An Experimental Evaluation of a Document Retrieval Support Method Based on Flexible Category Structure

Masahide Iwasaki Kokoro Nakagawa Yoshiaki Takata Hiroyuki Seki

Graduate School of Information Science
Nara Institute of Science and Technology

An existing WWW retrieval service using a category structure of WWW documents is easy to use because of its understandability. However, such a category structure has often too much volume to search information without effort. Furthermore, a structure prepared for a retrieval system cannot always reflect user's intention of search. In this paper, we propose a method for supporting WWW retrieval, in which a user can construct a small and flexible category structure without special knowledge and technique. An empirical evaluation using a retrieval system based on the proposed method is also presented.

1 はじめに

World Wide Web (WWW) で提供される情報が大幅に増大しつつある現在、WWW に対する様々な検索サービスが多くの利用者にとって不可欠なものとなっている。WWW 検索サービスの利用者の多くは検索を専門としない一般のユーザであり、そのようなユーザでも容易に目的の情報を得られるようなサービスが求められている。しかし現状では、いずれのサービスを用いても目的の情報を得ることが困難な場合がしばしばある。

そこで本研究では、検索毎に適切なカテゴリ構造を提供する WWW 検索支援手法を提案し、この手法を実装したシステムを用いて評価実験を行う。

1.1 現在の WWW 検索サービス

現在の WWW 検索サービスには、キーワード検索サービス型とディレクトリサービス型の 2 種類がある。

キーワード検索サービスは、ユーザにキーワードまたはキーワードの組合せ(検索式)を入力してもらい、入力語を多く含む文書を出力するものである。

一方ディレクトリサービスは、データベース中の文書を分類して登録したカテゴリ階層構造を提供するものである(図 1 (a))。目的に合わせてカテゴリを選択するだけで自然な絞り込みが行えるため、ユーザの負担が少な

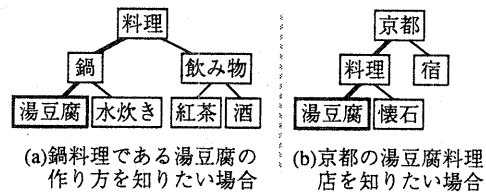


図 1: カテゴリ階層構造の例

い検索手法である。

検索結果を評価する基準として、再現率と適合率がよく用いられる。再現率は、データベースに存在する全ての正解に対して、結果に取り出すことのできた正解の割合である。適合率は、取り出した結果に対して、そこに含まれる正解の割合である。精度の良い検索とは、再現率と適合率が共に高い結果を出力するものということができる。

現状では、キーワード検索サービス、ディレクトリサービスのいずれの手法を用いても、必ずしも精度の良い検索を容易に行えるとは限らない。この原因として以下のような問題点が考えられる。

- (1) 適切なキーワードを考えるのが困難。データベースからうまく正解だけを抽出することができるような

入力語を考えることは、一般に難しい。

- (2) ディレクトリサービスが提供するカテゴリ構造は巨大。常にカテゴリ構造の一部しか見えていないユーザは、カテゴリの選択を行う際に迷いや誤解を生じやすい。
- (3) 分類の仕方は様々であり、適切なカテゴリ構造は検索毎に異なる。提供されるカテゴリ構造は常に固定されているが、ユーザが期待する構造がそれとは異なる場合もある。例えば「湯豆腐」というカテゴリに対して、図 1(a) のように配置したい場合や、(b) のように配置したい場合など、検索毎に様々な場合が考えられる。固定されたカテゴリ構造では全ての要求に対処できない。

ディレクトリサービスは (1) の問題を軽減しているが、(2), (3) のようなカテゴリ構造に関する問題を解決することはできない。

1.2 研究の目的

本研究では、あまり専門的な調査を目的としない一般的なユーザが、より容易に必要な情報を得られるような検索システムの構築を目指している [6, 7]。

前節の考察に基づき、本研究では、ユーザの検索目的に合わせた以下のような適応的なカテゴリ構造を提供する方法を考える。

- (i) ユーザの検索目的に関連のある部分に限られており小規模である。
- (ii) ユーザの必要とする文書が少数のカテゴリに集中している。
- (iii) ユーザにとっての自然な分類順序に従って構成されている。これにより、目的カテゴリ周辺にもユーザの興味に近いカテゴリが配置される。

検索毎に、それに応じて適切なカテゴリ構造を構築すれば、上の (2), (3) で述べたようなディレクトリサービスの問題を解決できる。さらに、例えば図 1(a) の構造では「湯豆腐」の周辺に「水炊き」や「飲物」、(b) の構造では「懐石」や「宿」というように、目的カテゴリの周辺にもユーザの興味に近いものを配置することができると考えられる。

また、検索の際にユーザに要求する手間をできるだけ抑えることも必要である。具体的には、適合率の高さをあまり要求されない程度のキーワードを入力することと、システムによって提示されるカテゴリから目的に近いものを選択することのみで検索を行えることを目指している。

以上の方針に沿って設計した検索支援システムと、このシステムを用いた評価実験について、以下で説明する。

2 システム設計の方針

2.1 基本方針

提案システムは以下の 2 段階の処理を行う。

- (1) ユーザの興味を反映する文書の集合を大まかに決める。
- (2) その文書集合に依存してカテゴリ構造を作る。

(1) の段階で対象とする文書集合を大まかに限定し、その文書集合の特性を利用して、検索毎の目的に合わせたカテゴリ構造を作る。(2) のカテゴリ構造の構築には、設計者がシステムに予め与える分類観点という予備知識を用いる。

分類観点とは、一揃いのカテゴリの集合である。例えば、カテゴリ「京都」「大阪」「奈良」などからなる分類観点「地域」や、カテゴリ「料理」「読書」「アウトドア」などからなる分類観点「趣味」などが考えられる。ディレクトリサービスでいうと、各分類観点は、カテゴリ階層構造に現れる兄弟ノードの集合 (1 ノードに対する子の集合) に当たる。このような分類観点を複数用意し、それらを組み合わせて適切なカテゴリ構造を構築することを目指している。

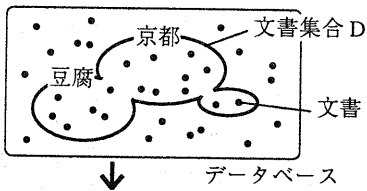
2.2 システムの動作の概要

システムの動作手順は以下ようになる [7] (図 2)。

- (1) ユーザが複数のキーワードを入力する。
 - (2) 文書データベースから、入力キーワードに関連する文書を収集する。この文書集合を D とする。
 - (3) システムに用意した各分類観点 S_j に沿って D を分割する。すなわち、各文書 $d \in D$ について、 d が属するカテゴリ w_{ji} を S_j の要素から一つ選ぶ。 w_{ji} は、 S_j の要素のうち d との類似度が最大のものとする (3.2 節)。
- 分割を行った後、すべての分類観点に対してその適切さを評価する。すなわち、ユーザの問い合わせに対して適切かどうか、あるいはユーザにとって有益かどうかを反映するような評価関数を使って、各分類観点の評価値を決める。この評価関数については 2.3 節で述べる。
- (4) 上記の評価値の高い (複数の) 分類観点と、それぞれの要素であるカテゴリをユーザに提示する。
 - (5) ユーザが、提示されたカテゴリの中から適するものを選ぶ。同時に、そのカテゴリに分類された文書集合 D' を検索結果として検索を終了するか、さらに D' の分割を行うかを選ぶ。
 - (6) ユーザが D' の分割を希望したときは、 D' について (3) 以下を繰り返す。

(1) キーワードを入力 豆腐 京都 食べ物

(2) 関連文書をある程度広く収集

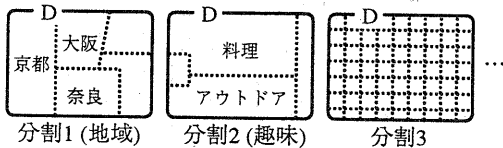


データベース

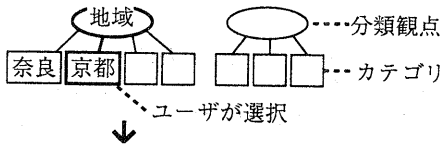
(3) 分類観点に沿って D を分割

地域: {京都, 大阪, 奈良, ...}
趣味: {料理, 読書, アウトドア, ...}

分類観点の集合



(4) 評価値の高い分類観点を提示
(5) ユーザが適するカテゴリを選択



(6) 選択されたカテゴリに対応する文書集合について, (3)~(4)を繰り返す

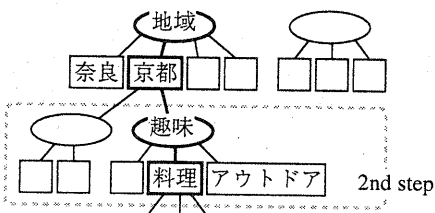


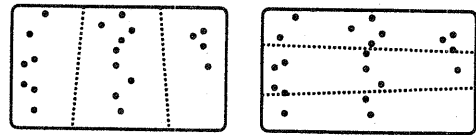
図 2: システムの動作

2.3 分類観点の評価

分類観点の評価値を決めるための評価基準として, 分類の明確さとエントロピーという 2 種類の評価基準を考えた. これらの比較は, 4 節の評価実験によって行う.

2.3.1 分類の明確さに基づく評価

提案システムでは, 分類観点の適切さを, 分類の明確さという評価基準に基づいて評価する. これは, その分類観点が文書集合 D を分類するのにどの程度適しているかを表すひとつの尺度である. 分類が明確であるとは, 文



(a) 分類が明確

(b) 明確でない

図 3: 分類の明確さ

書 d が属すべきカテゴリが明確である, すなわち, d とカテゴリ w_{ji} との類似度が d と他のカテゴリとの類似度よりも十分高いということである (図 3 (a)). 分類が明確な分類観点がユーザの検索意図を必ずしも反映するとは限らないが, 対象集合のもっともらしい分類という点で, ユーザにとって有益な情報と考えられる.

類似度の形式的定義は 2.4 節で述べる.

2.3.2 エントロピーに基づく評価

一方これとは別に, エントロピーを使って分類観点の適切さを評価することも可能である. これは, 1 回の選択操作の効率が良い方が, 検索の際の絞り込み回数 (ユーザへの質問回数) が少なくなり, ユーザにとって使いやすいという考えに基づいた基準である. つまり, 図 4 (a) のようにほぼ均等に分類された場合は, 1 回の選択で検索の対象となる文書数を数分の一に減らすことができる. しかし, (b) のように偏った分類が行われた場合は文書数をうまく減らすことができない場合もあり, その場合は何回もカテゴリ選択を行う必要が生じるため, ユーザの負担が増える.

そこで, 以下の式で与えられるエントロピーに基づいて各分類観点 S_j を評価することを考える.

$$H(S_j) = - \sum_{i=1}^{k_j} P_i \log P_i. \quad (1)$$

ただし,

k_j : S_j に属するカテゴリの数,

$$P_i = \frac{D_{ji} \text{の要素数}}{\text{全文書数}},$$

D_{ji} : カテゴリ w_{ji} に属する文書集合.

つまり, エントロピーが大きい分類観点は文書集合をほぼ均等に分割しており, 逆にエントロピーが小さい分類観点は偏りがある分割をしているということである.

2.4 ベクトル空間モデルによる文書の表現

文書と単語の関連, および文書同士の関連を測る手法として知られている, ベクトル空間モデル [10] について説明する. このモデルでは, 各文書の特徴は, 各単語との関連度を並べたベクトルで表現されるとする. つまり, システムが扱う単語の集合を

$$W = \{w_1, w_2, \dots, w_n\}$$

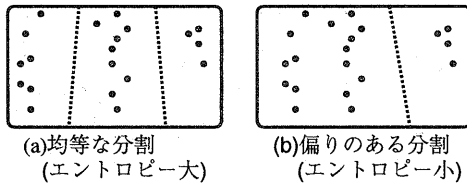


図 4: エントロピー

とすると、文書 d の特徴ベクトルは

$$c(d) = (c_{w_1,d}, c_{w_2,d}, \dots, c_{w_n,d})$$

の形で表される。ただし $c_{w_i,d}$ は、単語 w_i と文書 d との関連度である。

文書と単語との関連度は、文書中での単語の出現回数と定義する。あるいは、文書集合全体での出現回数が少ない単語により高い重みを与えたり、また単語の出現回数を文書の長さで正規化する定義なども考えられる [1]。

ロボットプログラムを使って文書を収集している検索サービスの多くでは、収集した各文書 d_j について各単語 w_i の出現回数 c_{w_i,d_j} を計測し、 c_{w_i,d_j} の集合からなるデータベースを検索対象としている [3, 8]。今回のシステムでも、システムが扱う全文書 d に対して、特徴ベクトル $c(d)$ をデータベースに保持するようにした。

文書同士の類似度が高いとは、2ベクトルのなす角が小さいと定義する。すなわち 2ベクトル u, v の類似度 $sim(u, v)$ を

$$sim(u, v) = \frac{u \cdot v}{|u| |v|}$$

とする。

3 実装

3.1 システムの構成

試作したシステムは、ユーザによる検索が行われる前に予め実行しておくデータベース構築部と、ユーザの問い合わせに応じる問い合わせ処理部からなる (図 5)。

図 5 は、各プログラムとその入出力ファイルの関係を簡単に表している。一番上にあるのが検索対象となる文書 (HTML ファイルなど) の集合である。WWW 全域を検索対象とする検索サービスを実現するには、ロボットプログラムなどを用いて広く WWW 文書を収集する必要があるが、今回は実験用として学内の WWW 文書や新聞記事 CD-ROM などを検索対象とした (3.3 節)。

本システムは、全文キーワード検索システムである Freya[2] を拡張する形で実現した。文字列をキーとして、それを含む文書を検索するためのデータ構造である索引 (lex, idx, map の 3 ファイル) を、Freya を用いて作成する。また、システムが扱う単語の集合 (2.4 節の W)

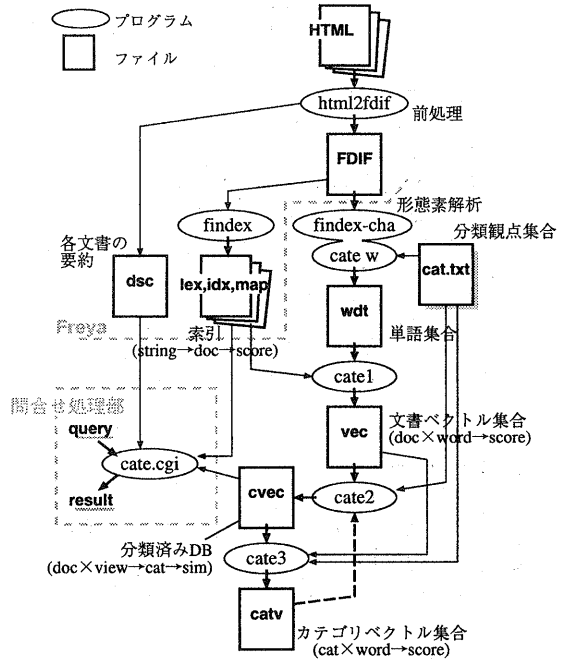


図 5: 試作システムの構成

を決めるため、日本語形態素解析システム茶釜 [5] に同梱の茶釜ライブラリを用いて全文書を形態素解析する。

問い合わせ処理部は CGI プログラムとして実装した。

3.2 処理手順

O. 準備

2.4 節のとおり、システムが扱う単語の集合 W 、文書 d の特徴ベクトル $c(d)$ 、および、特徴ベクトル間の類似度 $sim(u, v)$ を定義する。 W の要素の選び方は次節で述べる。また、システムが扱う全文書集合を D_0 とする。集合 D の要素数を $|D|$ と書く。

システム設計者が与える分類観点の集合を

$$S = \{S_j \mid 1 \leq j \leq m\}$$

とする。ただし、分類観点 $S_j = (l_j, W_j)$ は、分類観点名 l_j とカテゴリ名の集合 $W_j \subseteq W$ の 2 字組である。 $k_j = |W_j|$, $W_j = \{w_{j1}, \dots, w_{jk_j}\}$ とする。

S_j による文書集合 D の分割 (D_{j1}, \dots, D_{jk_j}) を、

$$D_{ji} = \{d \in D \mid sim(c(w_{ji}), c(d)) = \max_{1 \leq h \leq k_j} sim(c(w_{jh}), c(d))\}$$

と定義する。 $sim(c(w_{jh}), c(d))$ を最大にする h が複数ある場合は、任意の 1 個を選んで $d \in D_{jh}$ とする。このように決めた d が属するカテゴリを $w_{jmax_j(d)}$ とする。

ここで、 $c(w_{ji})$ はカテゴリ w_{ji} の特徴ベクトルである。これを w_{ji} のカテゴリベクトルと呼ぶ。 $c(w_{ji})$ の初期値を、語 w_{ji} に対応する成分が1で他の成分がすべて0の n 次元ベクトルと定義する。この初期値を使った分割は、カテゴリ名との関連度に基づいて各文書の属するカテゴリを決めることに相当する。

文書集合 D に対する分類観点 S_j の評価値 $e_D(S_j)$ を以下のように定義する。分類の明確さは、文書とそれが属するカテゴリとの類似度の平均と定義する。すなわち分類の明確さで評価する場合は、

$$e_D(S_j) = \sum_{d \in D} \text{sim}(c(w_{j \max_j(d)}), c(d)) / |D|.$$

また、エントロピーで評価する場合は、式(1)の右辺を $e_D(S_j)$ の定義とする。

A. データベース構築部

- (1) 単語集合の抽出 (cate w). (a) 形態素解析によって抽出された語、および、(b) 分類観点集合 S に現れるカテゴリ名からなる集合を単語集合 W とする。(a) について、ここでは形態素解析で得られた語のうち日本語の名詞・動詞・形容詞のみ W に含めた。
- (2) 文書ベクトルの構築 (cate1). 索引 (lex, idx, map) を、文書 d_j と語 $w_i \in W$ の組から関連度 c_{w_i, d_j} を得られるようなデータ構造に変換する。
- (3) 第1回の分類 (cate2). カテゴリベクトルの初期値を使って、各分類観点 S_j による全文書集合 D_0 の分割 (D_{j1}, \dots, D_{jk_j}) を求める。
- (4) カテゴリベクトルの再定義と第2回の分類 (cate3, cate2). (3) の分類では、カテゴリベクトルが離散的なため、どのカテゴリとも関連度が0の文書が多くなるなどの問題がある。そこで、カテゴリと関連する文書の平均特徴ベクトルをカテゴリベクトルとすることで、分類を平滑化する。
上で求めた各分割に基づいて、カテゴリベクトル $c(w_{ji})$ を、集合 D_{ji} の要素の平均特徴ベクトルと定義しなおす。すなわち

$$c(w_{ji}) = \sum_{d \in D_{ji}} c(d) / |D_{ji}|.$$

このカテゴリベクトルに基づいて、各分類観点 S_j による D_0 の分割を再計算する。

B. 問い合わせ処理部 (cate.cgi)

ユーザによるキーワード入力、および、カテゴリの選択が行われるたびに、以下を行う。

- (5) 入力キーワードに対する文書集合 D を抽出。索引を使って OR 検索を行い、文書集合を抽出する。この集合を D とする。

- (6) 指定されたカテゴリに属する文書集合 D' を抽出。キーワード入力後、ユーザが選択したカテゴリの系列を

$$\text{path} = w_{j_1 i_1}, \dots, w_{j_k i_k}$$

とすると、

$$D' = \bigcap_{h=1}^k D_{j_h i_h}.$$

- (7) D' に対する各分類観点の評価値を計算。 D' に対する分類観点 S_j の評価値 $e_{D'}(S_j)$ を計算する。
- (8) 評価値の降順に分類観点を出力、または指定されたカテゴリに属する文書の一覧を出力。ユーザの指示に従って、評価値の高い順に分類観点の一覧、または、 D' の要素の一覧 (dsc ファイルに保存されている、URL、タイトル、本文の一部など) を出力する。

3.3 検索対象文書集合

後述の実験では BMIR-J2 テストコレクション [4] を検索対象として用いた (5,080 記事)。これは、(社) 情報処理学会・データベースシステム研究会が、新情報処理開発機構との共同作業により、毎日新聞 CD-ROM'94 データ版を基に構築した情報検索システム評価用テストコレクションである。

3.4 分類観点集合

既存のディレクトリサービスで提供されている複数のカテゴリ構造を参考に、分類観点集合を作成した (分類観点 68 個、延べカテゴリ数 1,723)。

3.5 システムの諸元

本システムは、DEC (コンパック) 社の AlphaStation 255/300 (主記憶 128MB) および DIGITAL Personal Workstation 600au (同 768MB) 上に、C++ 言語を使って実装した。前者を主に問い合わせ処理、後者をデータベース構築に使用している。

形態素解析以降のデータベース構築時間は、概ね文書数と延べカテゴリ数の積に比例する。600au を使って CD-毎日新聞'94 年版の全文書約 10 万をデータベース化した場合、この部分に約 3 時間かかる。

問い合わせ処理の実行時間は、概ね初期キーワード検索で抽出された文書数に比例する。CD-毎日新聞'94 の全文書を抽出した場合、AlphaStation 255/300 では分類観点一覧が表示されるまでに約 25 秒かかる。数千文書程度までなら数秒以下である。

4 評価実験

本研究の目的は、検索のたびにユーザの検索意図や興味に沿ったカテゴリ構造を提供することにより、より容易に精度の良い検索を行えるようにすることである。

表 1: 実験に用いた検索課題

検索課題	正解数
農薬	26
飲料品	60
減税	302
核兵器	98
教育産業	17
株価動向	37
安売りをを行う流通行者	36
映画	21
東南アジアから日本への輸出	59
女性の雇用問題	80

そこで今回は、次の (1), (2) の比較実験を行った。ここで、実験に用いたシステムを次のように定義する。

- システム A: 分類観点の評価に分類の明確さを用いたシステム (提案手法 1)。
- システム B: 分類観点の評価にエントロピーを用いたシステム (提案手法 2)。
- システム C: クラスタリングによる分類 (4.1 節) を行うシステム。

また比較の基準として検索の精度を用いた。ここで検索の精度とは、ユーザの行動パターンを仮定したときに得られる再現率と適合率とする。これは、1.2 節の目的 (ii) に対応する。

- (1) 提案手法を含む 3 種類の分類法を用いた場合の検索精度の比較 (4.2 節)。
提案手法による検索の精度を調べる。この実験では、システム A, B, C の 3 システムについて、検索課題を用いて再現率と適合率の関係を求め、各システム間で検索精度 (同再現率に対する適合率) の比較を行う。
- (2) 分類観点の各評価基準における、評価値と検索精度の相関関係 (4.3 節)。
ここでは、高い検索精度の得られる分類観点がユーザにとって有益と考えているので、分類観点の評価値が検索精度を反映していることが望ましい。そこで、この実験では、システム A, B それぞれについて検索課題毎に、適合率による順位と評価値による順位との間に正の相関関係があるかどうかを調べる。また、そのときの相関係数を調べる。

実験には、3.3 節で述べた BMIR-J2 文書集合を用いた。また、今回は被験者を用いず、仮定した行動パターンを機械的に実行して実験を行った。検索課題には、BMIR-J2 テストコレクションに用意されている検索課題のうち表 1 に挙げた 10 課題を利用した。

4.1 クラスタリング

提案手法では、予め与えられた分類観点という知識を用いて分類するという形をとっている。しかし、このような知識を用いる代わりに、統計的分類手法 (クラスタリング) を用いる方法も考えられる [9]。そこで、提案手法との比較のため、クラスタリングを用いたシステムも実験対象とした。

今回使用したクラスタリングのアルゴリズムは K-means 法 [11] と呼ばれるもので、以下の手順で実行される。

- Step1.** K 個のクラスタの重心 (代表値) となる特徴ベクトルを適当に決める。
- Step2.** 全ての文書を、どのクラスタの重心に一番近いか (類似度が高いか) によって K 個のクラスタに分類する。
- Step3.** 各クラスタについて、それに属すると判定されたパターンの平均値を求め、新しいクラスタの重心とする。
- Step4.** 全てのクラスタの重心が変化しなくなるまで、Step2, 3 を繰り返す。

今回の実験では、 $K = 10$ とした。また、クラスタの重心が必ずしも収束しないことがあるため、100 回反復した時点で収束していなければ、その時点の結果を出力するものとした。

4.2 評価実験 (1)

実験手順

評価実験 (1) の手順について、システム A を用いて分類を行った場合を例に説明する。

- (1) 検索課題から思いつく適当なキーワードを入力する。
- (2) 分類の明確さが最大の分類観点を選ぶ。
- (3) 適合率の高いカテゴリを選ぶ。
選んだカテゴリの文書数がしきい値 Th 以下であれば、それに属する文書を出力文書集合に加える。文書数が Th より多いときは、再分類を行った上で再帰的に (2) 以下を繰り返す。
- (4) 出力文書集合 (文書数が Th 以下になったカテゴリに属する文書の集合) の再現率が目標再現率以上になるまで (2), (3) を繰り返す。目標再現率以上になったら、この時点での出力文書集合の再現率と適合率を求める。
- (5) (1) から (4) の手順を、目標再現率を 0.1 から 1.0 まで 0.1 ずつ増やしながら行う。

ここでの再現率、適合率はそれぞれ、(1) の初期検索の結果における値を 1 とした場合の値である。

システム B の場合は、(2) でエントロピー最大の分類観点を選んで同様の作業を行った。システム C の場合は、

(2)の作業を行わずに、(3)で適合率の高いクラスタを選んで実験を行った。

しきい値 Th は一画面で閲覧するのに負担が少ないと思われる文書数の上限で、今回の実験では 30 とした。

結果

評価実験 (1) によって得られた再現率と適合率の関係をグラフにすると図 6 のようになった。ただし、図の Clearness, Entropy, Clustering はそれぞれシステム A, B, C に対応する。グラフの横軸と縦軸は、それぞれ再現率と適合率である。一般に、再現率が高いときでも適合率が高ければ、すなわち曲線がグラフの右上に近いほど、精度の良い検索手法であるといえることができる。

図 6 を見ると、提案手法であるシステム A, B はシステム C より精度において上位にある。各検索課題毎の検索精度を調べた場合も、多くの課題でシステム A または B の方が優れていた。よって、システム A, B はシステム C に対し同等以上の精度を持つと言える。

システム A, B 間については大きな差は見られなかったが、エントロピーを評価基準とするシステム B の方がやや精度が良い傾向にあった。

考察

上記のとおり、提案手法 (システム A, B) はクラスタリングによる分類 (システム C) よりも精度が高かった。これは、提案手法では分類観点という予備知識を使用しているので、クラスタリングに比べ、より文書の内容を反映した分類が行えたことが原因と考えられる。

一方、分類観点の評価基準の違いに関しては、エントロピーを評価基準とした場合 (システム B) の方が分類の明確さをを用いた場合 (システム A) よりも若干優れていた。これは、今回尺度として用いた検索精度について、より細かく分類するほど検索精度を高くしやすいことが原因ではないかと考えられる。すなわち、例えばひとつのカテゴリを出力文書集合に加える代わりに、そのカテゴリをさらに分割し、各部分集合のうち適合率が 0 でないもののみ選択した方が、出力文書集合の再現率は元と同じで適合率は高くなる (低くはならない)。エントロピーは細かい分類ほど値が大きくなるので、結果としてエントロピーの高い分類は適合率も高くしやすいと考えられる。

4.3 評価実験 (2)

実験手順

次に評価実験 (2) の手順について説明する。

- (1) 検索課題から思いつく適当なキーワードを入力する。
- (2) 各分類観点 (68 個) について適合率の高い順にカテゴリを選び、それに属する文書を出力文書集合に加える。これを、出力文書集合の再現率が目標再現率と等しくなるまで続ける。

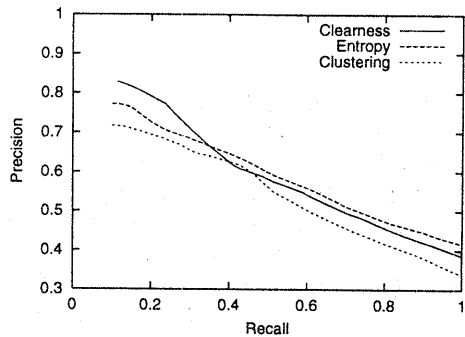


図 6: 各システムの再現率と適合率 (10 課題の平均)

表 2: 評価基準と適合率との相関係数の平均

評価基準	r	課題数	\bar{R}	カテゴリ数
Clearness	0.2	9	0.496	2.2
	0.5	8	0.595	4.1
	0.8	8	0.646	7.3
Entropy	0.2	10	0.596	2.3
	0.5	10	0.675	4.9
	0.8	10	0.684	8.7

*1: r は目標再現率, \bar{R} は課題毎の相関係数の平均。

*2: 評価基準 Clearness, Entropy はそれぞれ、分類の明確さ、エントロピー。

*3: カテゴリ数は、評価値の高い上位 10 個の分類観点について、目標再現率を達成するのに必要なカテゴリ数の平均。

- (3) 出力文書集合の適合率を求める。
- (4) 適合率と評価値の正の相関を検定し、相関がある場合は相関係数を求める。
- (5) (1) から (4) を、それぞれ目標再現率が 0.2, 0.5, 0.8 の場合について行う。

結果

評価実験 (2) で、検索課題を使って分類観点の適合率の順位と評価値の順位との間の正の相関を調べたところ、結果は表 2 のようになった。システム A による分類の場合、10 個の検索課題のうち、8 ないし 9 個の検索課題で正の相関が認められた。システム B による分類の場合は、全検索課題について正の相関があった。

また評価値の高かった上位 10 個の各分類観点について、目標再現率 r を得るのに必要だったカテゴリ数を調べた。表 2 の「カテゴリ数」は、この値を 10 個の検索課題にわたって平均したものである。その結果、システム B を用いた場合の方が、システム A を用いた場合よりもやや多いことが分かった。評価基準 (分類の明確さ、エントロピー) と検索課題 (10 個) を因子として分散分析を行ったところ、 $r = 0.5$ および $r = 0.8$ の場合それぞれについて、評価基準による差があることが言えた。

($p < 0.05$).

考察

分類観点の評価基準として分類の明確さを用いた場合、エントロピーを用いた場合のいずれにおいても、ほとんどの検索課題で適合率と正の相関があることが分かった。

10個の検索課題のうち相関があった課題の数、および平均相関係数 \bar{R} の値については、表2より、評価基準としてエントロピーを用いた場合の方が優れている。これも、評価実験(1)の考察と同様に、エントロピーの高いより細かい分類は適合率も高くしやすいことが原因と考えられる。

逆に、エントロピーの高い分類では、目標再現率を達成するのに必要なカテゴリ数が多くなるはずである。実際、表2より、エントロピーを評価基準とした場合は分類の明確さを用いた場合より必要だったカテゴリ数が多い。同程度の精度の検索を行う場合、選択しなければならないカテゴリ数が少ない方が望ましいので、この点では、エントロピーよりも分類の明確さの方が、ユーザにとっての使いやすさを実現していると考えられる。

5 まとめ

今回の評価実験では、主に精度の良い検索が行えるかどうかに着目して実験を行った。その結果、提案手法による分類を用いた方が、クラスタリングによる分類より精度が高いことが言えた。また、提案手法における分類観点の2つの評価基準については、エントロピーを使用した方が精度がやや高かった。

検索精度については、エントロピーを用いた方が優れていたが、同程度の検索精度を得るために必要なカテゴリ数は多くなっていった。また、上記の実験結果とは別に、評価値の高い分類観点や適合率の高いカテゴリ名を調べたところ、分類の明確さを用いた方が検索課題に対して妥当と思われるものが多かった(図7)。今回の実験は、適合率の高いカテゴリから順に選択する、すなわちユーザが理想的な選択を行うと仮定した上での検索精度を調べている。しかし現実のユーザが検索システムを用いる場合は、分類観点やカテゴリ名から正解の多いカテゴリを推測する必要があるため、それらが検索課題に対して妥当かどうかは重要である。これらのことから、ユーザにとっての使いやすさに関して検索精度以外の要素も考慮すれば、分類の明確さを用いた方が優れているのではないかと考えられる。

しかし、ユーザにとっての使いやすさについては、今回行った実験ではその一部しか評価されていない。実際のユーザに使用される場合は、前述のような分類観点やカテゴリ名の妥当性も特に重要となると思われる。今後は、実際にユーザを使った評価実験を行い、より適切な

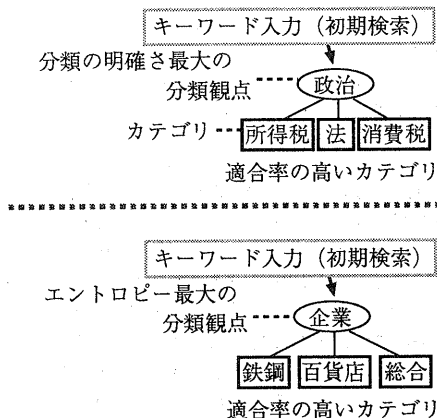


図7: 「減税」に対するカテゴリ構造

分類観点の評価法など提案手法の改良について考察していく予定である。

謝辞

分類観点の評価基準について有益な御助言・御討論を頂きました奈良先端科学技術大学院大学 石井信 助教授に深く感謝致します。

参考文献

- [1] Dreilinger, D. and Howe, A. E.: Experiences with Selecting Search Engines Using Metasearch, *ACM Trans. Information Systems*, 15(3), pp.195-222, July 1997.
- [2] 原田: Freya version 0.92, 1998-6, <http://odin.ingrid.org/freya/>.
- [3] 河野, 長谷川: WWW データ資源に対する重み付き相関ルール導出アルゴリズムの適用, 重点領域研究「高度データベース」松江ワークショップ, 1, pp.90-99, 1996-9.
- [4] 木谷, はか: 日本語情報検索システム評価用テストコレクション BMIR-J2, 情報処理学会研究報告, DBS114, pp.15-22, 1998-1.
- [5] 松本, 北内, 山下, 平野, 今一, 今村: 日本語形態素解析システム【茶釜】 version 1.5 使用説明書, 1997-7.
- [6] 仲川, 高田, 関: 可変なカテゴリ構造を用いた WWW 検索支援方法の提案, 電子情報通信学会第9回データ工学ワークショップ (DEWS'98), 講演番号 22, 1998-3.
- [7] 仲川, 高田, 関: 検索目的を反映したカテゴリ構造に基づく WWW 検索支援, 情報処理学会研究報告, HI82, pp.59-64, 1999-1.
- [8] Pinkerton, B.: Finding What People Want: Experiences with the WebCrawler, *Proc. 2nd WWW Conf.*, 1994.
- [9] Pirolli, P., Shank, P., Hearst, M. and Diehl, C.: Scatter/Gather Browsing Communicates the Topic Structure of a Very Large Text Collection, *Proc. CHI 96*, pp.213-220, April 1996.
- [10] Salton, G., Singhal, A., Buckley, C. and Mitra, M.: Automatic Text Decomposition Using Text Segments and Text Themes, *HyperText '96 Proc.*, pp.53-65, 1996.
- [11] Tou, J. and Gonzalez, R.: Pattern Recognition Principles, Addison-Wesley, Reading, Mass, 1974.