自然言語処理 123－9
（1998．1．19）

分類パターンを用いた文書データの自動分類法

徳田 克己　塩見 隆一　青山 昇一　柿ヶ原 康二
松下電器産業（株）マルチメディア開発センター

キーボードに不慣れなユーザーにとって、メニュー選択による検索は、有効な検索手法である。しかし、ネットワークなどを通じて得られるフロー型データに対して、リアルタイムに適切なメニューを作成することは難しい。我々は、分類視点に対して項目とキーワードで定義した分類パターンを多数用いて、文書分類を行なう手法を提案する。本手法では、(1) 文書データ中の重要語を用いて各分類パターンに文書データを対応付け、(2) 3つの指標 (対応文書データ数、有効分類項目数、文書データの分散度) からなる優先度を用いて適切な分類パターンを選択し、検索メニューを作成する。シソーラス、WEBページ上の分類カテゴリ、本の目次情報、新聞の紙面分類などを利用して分類パターンを作成し、新聞記事を分類するメニュー検索システムを試作した。試作したシステムを用いて、前述した優先度の評価を行った。

A Method of Clustering Documents Using Classification Patterns

Katsumi TOKUDA　Takakazu SHIOMI　Shoichi AOYAMA
Kouji KAKIGAHARA

Multimedia Development Center, Matsushita Electric Industrial Co., Ltd.

Retrieval by menu is one of the most effective methods for users that don't always type by keyboard. However, it is difficult to make a menu from flow data through a network in real time. We suggest a method of clustering documents using classification patterns which consist of some pairs of a classification item and keywords. In this method, (1) System connects words that appear frequently in the document and keywords in the classification patterns. (2) System makes a menu by selecting a suitable classification pattern using the priority based on 3 indexes (the number of distributed documents, the number of effective classification items and the distribution of documents). We evaluate the priority on a menu retrieval system using classifcation patterns, which consist of thesaurus, classification categories on Web pages, table of contents of some books and categories on newspapers.

1　はじめに

近年、WWWを代表とするネットワーク上の大量の電子データを個人が取り扱えるようになり、この中から必要な情報を獲得するためのデータ検索技術がより重要になってきている。有力なデータ検索手法として、キーワード検索がある。しかし、キーワード検索には、情報機器に不慣れな初心者ユーザーにとって以下の様な問題点がある。

1. 適切なキーワードを思いつくのが難しい。
2. 文字入力そのものが難しい。

これらの問題点を解消するもう1つの検索方法にメニュー検索がある。メニュー検索では、データを一覧表示し、ユーザーはその中の情報を選択するだけでよい。しかしながら、ネットワーク上のデータのようなフロー型データに対して、即時にメニューを作成することは容易ではない。

そこで、我々は分類視点に対して項目とキーワードで定義した分類パターンを多数用いて、文書分類

を行なう手法を提案する。

本手法を用いることにより、データ検索のための階層的なメニューを作成することができ、ユーザーはメニュー検索によって容易に情報を得ることができる。本稿では、

1. 分類パターンを用いた文書の自動分類手法
2. 本手法を用いた情報検索システム
3. 分類パターンの作成
4. 評価実験

について報告する。

関連研究

文章データの自動分類は、大きく分けて2つの方法に大別できる。1つは、クラスタリングアルゴリズム [1, 2] を用いて分類するものであり、、もう1つは、予め定義したカテゴリーに対して割り付ける方法 [3, 4, 5] である。

前者は、不特定のデータに対応可能であるが、分類されたグループの意味を表現するのが難しく、ユーザーがグループの選択を行なうのが困難とされている。

後者の一般的な手法は、カテゴリーの代表的な文書ベクトルを用意し、対象文書との距離を用いて分類を行なう。この手法ではユーザーにとって、わかりやすい分類グループにデータを分類できるが、不特定のデータに対して分類を行なう場合、大量のカテゴリーを必要とし、これらのカテゴリーを用意することが困難であるという問題点がある。

一方、大量のカテゴリーを用意するため既存のシソーラスを利用して分類し検索メニューを作成する手法 [6, 7] が提案されている。しかし、シソーラスで作成されたメニューの評価実験の結果、検索のためのキーワードをシソーラス中から探索してしまっており、本来の検索メニューの役割を果たし切れていていないことが報告されている。また、シソーラスと目次を組み合わせた検索メニュー作成手法 [8]、キーワード検索結果を目次を利用して提示する手法 [9] も提案されている。

本稿で提案する手法は、上記シソーラスや目次を利用した検索メニュー作成手法を発展させたものであり、シソーラスの部分や本の目次などあらゆる分類のための階層構造を利用できる枠組を提供し、より柔軟な検索メニューを作成できるものである。

また、我々が提案する分類パターンと類似の概念として文献 [10] では、検索語拡張のための「多様分類情報」を定義している。

2 分類パターンを用いた文書の自動分類

2.1 分類パターン

分類パターンは、分類視点に対して分類項目とキーワードの組が複数定義される。表1は分類視点「プロ野球」の例である。3つの分類項目「セリーグ」「パリーグ」「大リーグ」が定義されている。また、データを各分類項目に分類するためのキーワードがそれぞれ定義されている。1つの分類項目に対してキーワードは複数定義できる。

分類視点：プロ野球

分類項目	キーワード
セリーグ	スワローズ、ベイスターズ、カープ ジャイアンツ、タイガース、ドラゴンズ
パリーグ	ライオンズ、ブルーウエーブ、バファローズ ホークス、マリーンズ、ファイターズ
大リーグ	ドジャース、エンジェルス、メッツ

表 1: 分類パターンの例

分類パターンは、以下を利用して作成することができる。

- シソーラス（の一部分）
- WEBページ上の分類カテゴリーの利用
- 本の目次情報
- 新聞の紙面の分類

また、人手により作成することも可能である。これは、サンプル文書が必要な従来の文書ベクトルによるカテゴリー作成では不可能なことである。

2.2 メニュー作成手法の概要

以下の手順で分類パターンを用いて文書データを分類する。

1. 分類対象の文書からは主題を代表するような単語を抽出する。
2. 抽出された重要語を用いて分類対象の文書を分類パターンに対応付けする。
3. 適切な分類パターンを選択し、分類パターンの分類項目一覧を提示する。
4. ユーザーが分類項目の選択を行なったら、分類対象を分類項目に対応する文書に絞り込む。
5. ユーザーの指示があれば、分類対象の文書を再度分類する（2へ）。

従来までの単一のシソーラスの分類では、1つの視点で分類を詳細化するだけであったが、本手法では、分類項目が選択された都度、再分類を行なうた

め、複数の分類視点による分類が可能である。具体的には、最初に分類視点「経済」で分類された1項目を選択した後、分類視点「政治」で分類されることも可能である。

上記を実現するため、シソーラスを利用する場合は、できるだけシソーラスを分解して分類パターンに流用することとする。

2.3 文書データと分類パターンとの対応付け

文書と分類パターンの対応付けは、文書の主題を表すような重要語を抽出し、分類パターン中の分類項目に定義された単語と照合を行なうことによって行なう。抽出した重要語が複数の分類パターン中の単語に対応する場合は、その両方と対応付けする。また、文書の重要語はユーザの検索要求によって異なることが考えられるので、文書中から複数の重要語を抽出し、分類パターンと対応付けることとする。

文書中の重要語抽出には、文献[7]の結果から、文書内単語頻度TF [12]を用いた。

2.4 分類パターンの選択

分類対象の文書データを適切に分類している分類パターンを選択するため以下の3つの指標を使用することとした。

1 対応文書データ数

各分類パターンに対応する文書データ数。対応文書数が多い分類パターンほど適切な分類パターンである。

2 有効分類項目数

有効分類項目は、分類パターンの中で対応文書データがある分類項目のことである。分類項目の一覧表示では、対応文書数0の分類項目は表示しない。有効分類項目が一目でわかる項目数になっている分類パターンは、適切な分類パターンであり、有効分類項目数が少い場合や多い場合は不適切な分類パターンである。

3 文書データの分散度

各有効分類項目に対応する文書データ数が極端に偏っているより均等に近い方が適切な分類パターンである。

この3つの指標を用いて、分類パターンiの優先度P_iを算出する計算式(1)を決定した。

$$P_i = \alpha \cdot \frac{n_i}{max_i n_i} + \beta \cdot f(m_i) + \gamma \cdot \frac{1}{1+\sigma^2} \quad (1)$$

n_iは分類パターンiの対応文書数である。最初の項は最大対応文書数で割ることによって0～1の値に正規化している。

m_iは分類パターンiの有効分類項目数。関数$f(x)$は有効分類項目数xに対する分類パターンの優先度を定義する関数で、実際の値は図 1の値を取ることとした。

σ^2は分散である。最後の項は、分散が0のとき最大値1を取り、分散が大きいほど0に近い値を取る。

計数α, β, γは、上記3つの項に重み付けを行なう係数である。これらの値を変化させ実験を行なった。

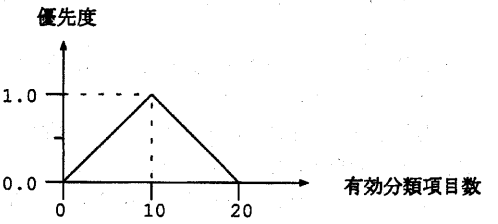

図 1: 有効分類項目数に対する優先度

3 情報検索システム

実際に、分類パターンを用いた文書データの自動分類法を利用した情報検索システムをWindows95上のプログラムとして作成した。図2は、本システムの画面の一部分である。

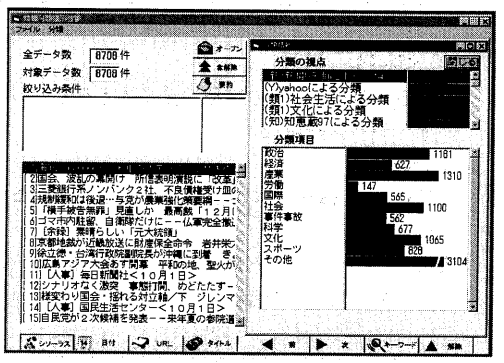

図 2: メニュー検索システムの画面

本システムでは、データ保存時に形態素解析を行ないキーワード抽出及び重要語抽出を行なう。自動分類ボタンが押されると、分類パターンを用いた自動分類を行ない、一番優先度の高い分類パターンの分類項目と対応文書データ数の一覧を表示すると共に、優先度上位10個の分類パターン一覧も表示する。ここで、一覧表示された分類パターンの1つを選択すると、選択された分類パターンの分類項目と対応文書データ数の一覧が表示される。また、一覧

表示されている分類項目が選択されると文書データの絞り込みが行なわれる。絞り込まれた文書データに対して再度自動分類を行ない検索メニューを表示させることもできる。

自動分類以外にもデータを絞り込むためのキーワード検索機能も備えており、キーワード検索で絞り込まれた文書データを自動分類し検索メニューを表示させること、検索メニューで絞り込まれた文書データを対象にキーワード検索を行なうことも可能となっている。

本分類手法は、共通する単語を含む文書データをまとめあげたものにすぎない。このため、メニューの各分類項目に含まれている文書データが類似した文書データの集合になっているとはいい難い。そこで、目的とする文書データを1つ選択すると、類似の文書データを収集し、ランキング表示する補助機能を付ける予定である[13]。ここでは、類似文書データを収集する機能の詳細は省略する。

4 分類パターンの作成

さまざまなベースデータから半自動あるいは手動で382個の分類パターンを作成した。評価のため、作成した分類パターンを以下の3つに類別した。

1 大分類パターン
世の中の事象全般を分類できると思われる分類パターン

2 中分類パターン
ある程度広い分野の事象を分類できると思われる分類パターン

3 小分類パターン
特化した分野の事象を分類できると思われる分類パターン

表2は、ベースデータ別に類別した分類パターン数をまとめたものである。

ベースデータ	分類パターン数			
	大	中	小	合計
シソーラス	1	10	200	211
新聞紙面	1	10	0	11
本の目次	2	23	58	83
UDC	1	8	0	9
Web page	1	11	0	12
その他	0	1	55	56
合計	6	63	313	382

表 2: ベースデータ別／種類別分類パターン数

以下、ベースデータ別に分類パターンを作成した方法を示す。

4.1 シソーラスの利用

分類パターンを抽出するシソーラスとして小学館類語例解辞典[14]を用いた。小学館類語例解辞典は、約25000語を約6000の類語グループに分類し、さらに、これらの類語グループを10の大分類と20の中分類によって200のグループに分類したものである。この中から、211の分類パターンを自動抽出した。図3は、小学館類語例解辞典の構成と抽出した分類パターンの関係を示したものである。分類パターンは、3種類の方法で抽出されている。

1種類目は全体を視点とするものである。分類項目として大分類、各分類項目のキーワードとして各大分類下の中分類の単語を使用する。この分類パターンは1パターンだけ抽出された。

2種類目は各大分類を視点とするものである。分類項目として中分類、各分類項目のキーワードとして各中分類下の類語グループの代表語を使用する。この分類パターンは大分類の個数と同じ10パターンが抽出された。

3種類目は各中分類を視点とするものである。分類項目として類語グループの代表語、各分類項目のキーワードとして類語グループに含まれる単語を使用する。この分類パターンは全中分類の個数と同じ200パターンが抽出された。

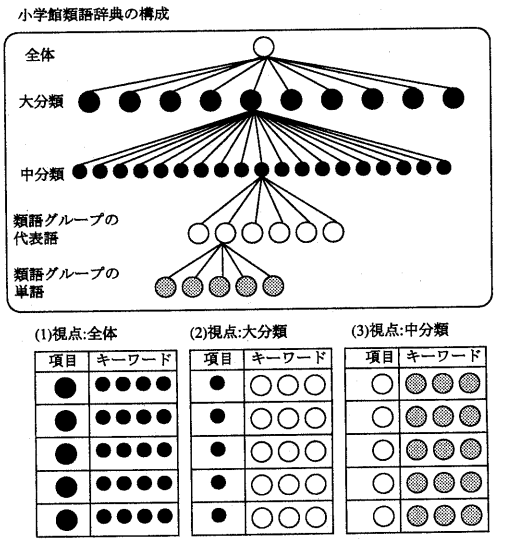

図 3: 小学館類語例解辞典からの分類パターンの抽出

上記の通り抽出された各分類パターンに対して、

人手で若干の修正（類似している分類項目の統合、キーワードの補充）を行ない分類パターンとして使用することにした。

4.2 新聞紙面分類の利用

文献 [16] は新聞記事を大分類、中分類に分類するために分類項目と中分類に対応するキーワードを策定している。このデータから、全体を分類視点とし大分類を分類項目とした分類パターン 1 個と、各大分類を分類視点とし、中分類を分類項目とした分類パターン 10 個を抽出した。文献 [16] は、大分類に直接キーワードを定義していない。そこで、前者の分類パターンでは、大分類下の中分類に定義されているキーワードを統合し、分類項目に対応するキーワードとする。また、後者の分類パターンのキーワードは、中分類下のキーワードを人手で若干修正したものを使用することにした。

4.3 本からの抽出

百科辞典など、対象を体系的にまとめている本の目次も利用できると考えた。具体的には、辞書 [17, 18]、統計資料 [19, 20]、図書目録 [21, 22] を利用した。本の目次から、全体を分類視点とし目次の章を分類項目とした分類パターンを抽出し、さらに目次に項があれば、各目次の章を分類視点とし、目次の項を分類項目とした分類パターンを抽出した。この情報を人手で修正を行ない分類パターンとした。

4.4 UDC の利用

図書を分類するための UDC[23] を分類パターンに流用した。まず、図書分類全体を分類視点とし、UDC の 9 部門[1] を分類項目とした分類パターンを抽出した。次に各部門を分類視点とし、部門下の副部門を分類項目とした分類パターンを抽出した。キーワードは、既に作成した他の分類パターンを参考に人手で付与した。

4.5 Web Page からの抽出

Yahoo[24] のホームページにある Web Page の分類カテゴリーを分類パターンに利用した。まず、全体を分類視点とし、トップページの大きいカテゴリーを分類項目とした分類パターンを抽出した。また各カテゴリーを視点として、カテゴリーのページ内のカテゴリーを分類項目とした分類パターンも抽出した。キーワードは、既に作成した他の分類パターンを参考に人手で付与した。

[1] 部門番号 4 は現在空位

4.6 その他

当社保有の地名辞典から都道府県、国名の分類パターンを、当社保有の仮名漢字変換用辞書の企業名を業種別に分類し、業種別企業名の分類パターンを作成した。

また、プロ野球、Ｊリーグ、温泉、料理などの分類パターンも作成した。

5 評価実験

作成した分類パターンを用いて、以下の 2 点について調べた。

- 分類パターンの評価
  分類パターンが多いと分類のための計算量が多くなる。分類効果が高く、必要かつ十分な分類パターンのセットを用意することが望ましい。不要な分類パターン、不足している分類パターンを評価する。
- 優先度の有効性
  パラメータ α, β, γ の値を適切に設定し、計算式 (1) によって適切な分類パターンを優先的に提示することができるかを評価する。

5.1 データ

実験を行なう分類対象データとして、「CD 毎日新聞 94 版」[25] に収録された新聞記事データを用いた。記事データから 3 つの記事セットを作成した。表 3 は記事セットをまとめたものである。

セット	内容	記事数
S1	20 記事/1 日で 1 年分	7220
S2	4 月の全記事	9061
S3	10 月の全記事	8708

表 3: 評価用新聞記事データ

5.2 計算式の係数

以降の実験で用いる計算式 (1) の係数 α, β, γ のパターンを 3 通り決定した。表 4 は係数のパターンをまとめたものである。p_1 は、対応する文書データ数だけで適切な分類パターンを決定する。p_2 は、p_1 に分類項目数を考慮して適切な分類パターンを決定する。p_3 は、さらに文書データの分散を考慮して適切な分類パターンを決定する。

5.3 大分類パターンの評価

分類対象データの量が多く、世の中の事象全般が含まれている時、用意した大分類パターンを用いて分類し、検索メニューを作成するのが望ましいと考

係数パターン	α の値	β の値	γ の値
p_1	1	0	0
p_2	1	1	0
p_3	5	5	1

表 4: 計算式 (1) の係数パターン

えられる。このような結果を得られることを確認するため、用意した記事のテストに対して計算式 (1) の係数のパターンを変化させながら自動分類を試みた。

記事	S1			S2			S3		
係数	p_1	p_2	p_3	p_1	p_2	p_3	p_1	p_2	p_3
新聞	1	1	1	1	1	1	1	1	1
知恵蔵	2	2	2	2	2	2	2	2	2
Yahoo	3	5	5	3	5	5	4	5	5
UDC	4	3	3	4	3	3	3	3	3
Imidas	5	4	4	5	4	4	5	4	4
類語	9	6	6	9	6	6	9	6	6

表 5: 記事・パラメータセットによる分類パターンの優先順位

表 5 から、記事セットやパラメータ設定にあまり左右されず、大分類パターンが優先順位の上位を占めていることがわかる。特に 1 位、2 位は不変である。

そこで、大分類パターンは品質の高い 1 つの分類パターンだけで十分であると考えて、これら 6 つの大分類パターンを統合して 1 つの大分類パターンを作成した。また、大分類パターン同様、中分類パターン 63 個の中で類似した分類パターンが多く、これらを 28 個の中分類パターンに統合した。

5.4 中分類パターン、小分類パターンの評価

分類パターンの質と優先度の有効性を調べるため、絞り込まれた文書データに対する自動分類結果を被験者に主観で評価して貰う実験を行なった。

文書データの絞り込みは、以下の 2 通りを行なった。

1. 記事データ全体を自動分類すると統合された大分類パターンにより分類される。この分類の各分類項目を選択することによってデータを絞り込む。
2. 100 ～ 900 程度の記事に含まれるキーワードをランダムに 30 個抽出し、そのキーワードを用いてキーワード検索を行ないデータを絞り込む。

記事データは S1 を用い、パラメータは 3 種類を用いて自動分類を行なった。自動分類結果を被験者に提示し、上位 5 つの分類パターンについて「良い」「どちらかといえば良い」「どちらかといえば悪い」「悪い」の 4 段階での主観評価を依頼した。被験者はキーワード検索に慣れており、システムの利用時に戸惑うことはなかった。

絞り込み方法 1 の被験者数は 2 名で、10 個の大分類項目による絞り込み後の文書データに対して表示される分類パターンの評価を依頼した。

絞り込み方法 2 の被験者数は 10 名で、30 個のキーワードを 5 セットに分割し、1 セットにつき 2 人でキーワードによる絞り込み後の文書データに対して表示される分類パターンの評価を依頼した。

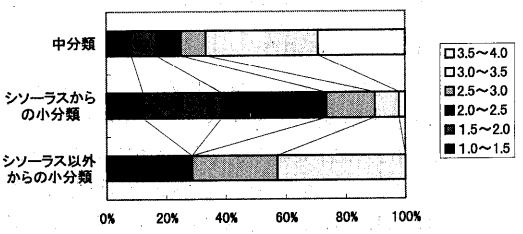

図 4: 中分類、小分類パターンの評価

図 4 は、今回の評価実験で実際に使用された 74 個の中分類・小分類パターンの評価平均を「悪い」の 1.0 から「良い」の 4.0 まで 0.5 刻みの 6 段階でグラフ化したものである。なお、大分類の評価平均は 3.2 であった。全 342 パターン中 75 パターン[2] が今回の実験で利用された。

パラメータ	表示 1	表示 1-2	表示 1-3	相関係数
$p1$	48%	63%	78%	0.356
$p2$	50%	80%	88%	0.400
$p3$	55%	85%	93%	0.478

表 6: 評価順位 1 位の分類パターンの表示順位

表 6 は、被験者による評価順位 1 位の分類パターンと優先度に基づく表示順位の関係をまとめたものである。評価順位 1 位の分類パターンが表示順位 1 位になっている比率、1 ～ 2 位に入っている比率、1

[2] 大分類パターンを含む

～3位に入っている比率、及び評価順位と表示順位の間の相関係数をパラメータ設定 $p1, p2, p3$ についてそれぞれ調べた。

6 考察

6.1 分類パターンの評価について

全342パターン中75パターンが利用されているが、これは7220件の新聞記事を分類するのに、342パターンは冗長であり、不要な分類パターンの存在を示していると考えられる。

また、図4を参照すると、評価値2.5を境として評価結果をOKとNGに分けた場合、平均では58%の分類パターンがOK評価を得ていることが分かる。これは特に高い数字ではないが、その理由は小分類パターンに関する評価が分かれているためである。

シソーラスから抽出した小分類パターンは73%がNG評価を受けているのに対し、それ以外の分類パターンは逆に70%以上がOK評価を得ている。この理由は以下の様に考えられる。

1. 分類パターンの不足
   適切な分類パターンがないため、不適切な分類パターンが表示された。(例) 人間の体
2. 不適切なキーワード
   一般的なキーワード(人、時など)が複数の分類パターンで定義されていたため、不適切な分類パターンが表示された。
3. 不適切な分類項目名
   分類項目名が抽象的で、各項目に分類されている文書データの内容を想像することが困難だった。(例) 男女、時間

上記の1～2は分類パターン自身の問題であり、分類パターン作成時に留意すべき点である。3は分類パターンとして不要であり、削除する必要がある。

シソーラスは論理的な整合性を重視しているため、抽象的な分類項目名があり、単純に分解して小分類パターンとして利用するには問題があるといえる。

シソーラス以外から作成した小分類パターンは一般に良い評価を得ている。これは「Jリーグ」の様に具体的で詳細な分類パターンであったため被験者にとって分かりやすかったと考えられる。シソーラスから作成した小分類パターンでも、具体的で詳細な分類パターンは良い評価を得ていた。

しかしながら、分類パターンが具体的かつ詳細であればあるほど、適切な分類パターンがない場合に次善の候補として表示されると被験者は違和感を感じる。このため、小分類パターンの優先度計算には具体性や詳細性の評価を導入して、候補とならないようにする必要がある。

6.2 優先度の有効性について

表6によると、評価順位1位の分類パターンが実際に上位(1～3位)の候補として提示される割合は、パラメータ設定 $p1, p2, p3$ の順に大きくなっていることがわかる。また、分類パターンの評価順位と表示順位の間の相関係数を計算すると、$p1$, $p2$, $p3$ の順に相関が強くなっていた。

このことから、人間の評価に近い順で分類パターンを表示するためには、分類パターンの優先度計算式(1)を構成する**対応文書データ数、有効分類項目数、文書データの分散度**の各項を組み合わせて利用することが有効であることが確認できた。

7 まとめ

1. 分類パターンを用いた文書データの自動分類法とこれを用いたメニュー検索システムを試作した。
2. 大分類、中分類パターンと比較して小分類パターンは具体的で詳細なものが好まれるが、分類パターンが具体的かつ詳細であればあるほど、適切な分類パターンがない場合に次善の候補として表示されると被験者は違和感を感じる。このため、小分類パターンの優先度計算に具体性や詳細性の評価を導入する必要がある。
3. 人間の評価に近い順で分類パターンを表示するためには、3つの指標(対応文書データ数、有効分類項目数、文書データの分散度)を組み合わせて分類パターンの優先度を定義することが有効であることを確認した。

今後、分類パターン抽出基準や優先度計算方法の改良を行ない、よりよい分類提示ができるようにしていく予定である。

謝辞

小学館類語例解辞典の電子化データの使用を許可して下さった株式会社小学館に感謝致します。

参考文献

[1] 宮崎哲夫, 田中栄治, 古城則道：”文書の意味空間へのマッピング,” 第53回情報処理学会全国大会講演論文集, 3-167～168, 1996.

[2] 有田英一, 安井照昌, 津高新一郎：”単語集合の自動構造化機能を持つ「情報散策」方式,” 電子情報通信学会・信学技法, NLC95-17, pp. 69-74 1995.

[3] 上田隆也，大谷紀子，伊藤史郎，柴田昇吾，池田裕治：" フロー情報収集・活用のための知的検索システムＦｉｔ（１）（２）（３），" 第５３回情報処理学会全国大会講演論文集，3-183 ～ 188, 1996.
[4] 森本由紀子，間瀬久雄，辻洋，絹川博之：" 新聞記事自動分類システム構築の検討と評価，" 第５３回情報処理学会全国大会講演論文集，3-205 ～ 206, 1996.
[5] 西野文人：" テキスト分類のためのカテゴリ割り付け戦略，" 情報処理学会研究会報告，Vol.NL 106-3, pp. 13-18, 1995.
[6] 千田恭子，篠原靖志，坂内広蔵：" 汎用シソーラスを利用した検索用の索引メニュー構成法，" 情報処理学会研究会報告，Vol.NL 111-4, pp. 21-26, 1996.
[7] 塩見隆一，徳田克己，青山昇一，柿ヶ原康二：" シソーラスを用いた文書データの自動分類法，" 情報処理学会研究会報告，Vol.NL 117-14, pp. 99-104, 1997.
[8] 千田恭子：" シソーラス「分類語彙表」の検索メニューへの適用，" 言語処理学会第３回年次大会発表論文集，Vol.3, pp. 537-540, 1997.
[9] 塩見隆一，徳田克己，青山昇一，柿ヶ原康二他：" 電子マニュアルにおける検索機能の評価，" 第５３回情報処理学会全国大会講演論文集，3-148 ～ 149, 1997.
[10] 下畑光夫，坂本仁：" 多様分類情報による検索語拡張，" 情報処理学会研究会報告，Vol.NL 115-19, pp. 135-140, 1996.
[11] Ben Shneiderman（東、井関訳）：" ユーザー・インターフェースの設計，" 日経ＢＰ, pp.93-98, 1988.
[12] G.Salton: "Automatic Text Processing," Addison Wesley, 1989.
[13] 野本昌子，野口直彦：" 文書構造と共起表現を用いた文書ランキング手法，" 第５２回情報処理学会全国大会講演論文集，4-202 ～ 203, 1996.
[14] 小学館辞典編集部：" 類語例解辞典，" 小学館，1994.
[15] 国立国語研究所（編）：" 分類語彙表，" 秀英出版，1964
[16] データベース振興センター：" 新聞記事分類キーワードの標準モデル構築と自動付与に関する調査研究，" データベース振興センター,1996
[17] 大谷洋行編：" 知恵蔵 1997." 朝日新聞社,1997
[18] 鈴木力編：" 情報知識 imidas 1997." 集英社,1997
[19] 日本能率協会総合研究所：" ビジネス調査資料総覧 1996," 日本能率協会総合研究所,1996
[20] 日本能率協会総合研究所：" ビジネス調査資料総覧 キーワードインデックス 1996," 日本能率協会総合研究所,1996
[21] 日本理学書籍目録刊行会：" 日本理学書総目録９５年度版，" 日本理学書籍目録刊行会,1995
[22] 工業書目録刊行会：" 電気電子工学書目録 1997," 工業書目録刊行会,1997
[23] 中村幸雄訳：" ＵＤＣの使い方，" 情報科学技術協会,1994
[24] http://www.yahoo.co.jp/
[25] 「CD 毎日新聞 94 版」