

自己組織化マップを用いた分類表示による文献検索結果の絞込み支援\*

4W-5

獅子倉 恵美子<sup>†</sup> 松田 郁夫<sup>‡</sup>  
日本工業大学<sup>§</sup>

1. はじめに

検索者が、十分な知識を持たない分野に対して、キーワード検索により文献を収集する場合、検索結果が極端に多くなったり、少なくなることがある。これは、その分野に対する語彙が不足するため、適切なキーワードが選ばれていないことが原因と考えられる。

「自然言語」のように関連する研究が幅広い語をキーワードとしたとき、検索結果が大量となる。絞り込み検索を行う場合、新たな追加キーワードは文献の中から選ばれる。目的とする内容に関連のある文献を大量の検索結果の中から探すことは、文献収集において時間のかかる過程である。

この過程を支援するために、自己組織化マップを用いて大量の検索結果を文献内容により分類表示し、文献閲覧回数を減少させる方法を提案する。

2. 支援の実現

大量の検索結果の中から文献や追加キーワードを見つけることが困難となるのは、検索結果の出現順が原因として挙げられる。

2.1. 従来の検索結果

文献検索の結果の多くは、題目や著者名などがリスト状に表示される。結果が表示される順番に

- ・ 題目の辞書順
- ・ 著者名別
- ・ データベースの登録順

などがある。

表示された検索結果がこれらの順に並べられている場合、文献や追加キーワードを得るためには、結果リストの中から、ひとつずつ求めている内容に関連のある文献を探さなければならない。

しかし、内容に関係なく結果が並べられているのであれば、関連のない分野の文献をいくつも読むことになってしまい、このことが文献の閲覧回数を多くさせ、文献や追加キーワードを探していく上で時間のかかる要因の一つになる。

2.2. 絞り込み検索のための検索結果表示

絞り込み検索のための追加キーワードを得る場合、文献閲覧回数を減少させるためには、すでに関連のないとわかっている分野の文献は、それ以降で見る必要のないようにしたい。そのためには、分野ごとに内容を比較できることが必要である。

検索によって得られた文献の集合の分類表示を行うことで、分野別に求める内容との比較が可能になり、検索結果のリストを順次閲覧するより、文献の閲覧回数を減少させられる[図.1]。

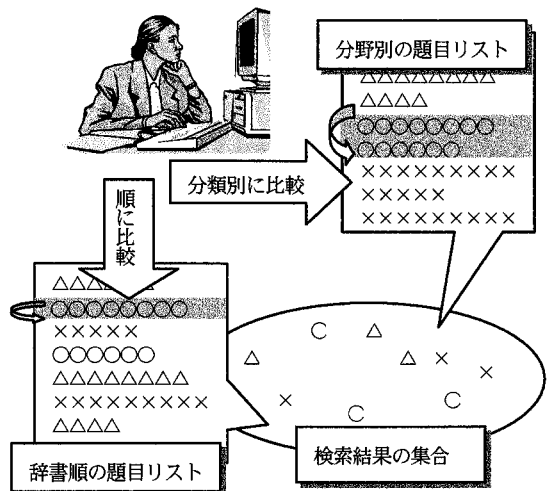


図 1. 文献の分類表示

\*The support system for focusing on document retrieval by showing classified results using Self-Organization Maps.

<sup>†</sup> Emiko SHISHIKURA

<sup>‡</sup> Ikuo MATSUDA

<sup>§</sup> Nippon Institute of Technology, 4-1, Gakuendai, Miyashiro-machi, Minamisaitama-gun, Saitama-ken, 345-8501

### 3. 分類

文献の分類は著者キーワードを特徴ベクトルとして用いた自己組織化マップにより行う。

#### 3.1. 文献間の関連と著者キーワード

文献の著者が選んだキーワードは、その文献の内容を要約したものである。本研究では、ある文献に対して付与された著者キーワードが、もう一方の文献にも付与されているとき、それらの文献は内容に関して関連があるとし、共通するキーワードの数が多ほど内容に強く関連があるとしている。

#### 3.2. 自己組織化マップ

自己組織化マップ(SOM)<sup>[1]</sup>は多次元の情報を、2次元平面に写像することが可能である。SOMには格子状にセルが配置され、各セルは適当な値を持った特徴量を $n$ 次元ベクトルの形でもっている。

SOMは、特徴ベクトルがよく似ているもの同士をマップの近い位置に配置する。

#### 3.3. 自己組織化による結果の分類

検索結果が分類され、文献同士の相互関連が失われないように2次元上に縮約表現が行われれば、必要とする文献やキーワードの収集を、分類されたまとまりごとと比較するだけですむ。

著者キーワードを収集し、特徴ベクトルとすることで、SOMは文献間に関連のあるものを近くに配置する。作成されたマップは、内容によって検索結果を分類したものとなる。

また、文献の配置されたセルの特徴ベクトルを調べることによって、求める分野の著者キーワードを収集することができる。

## 4. システムの構成

以下にシステムの構成[図.2]と各機能の概要を記す。

#### (1) 検索要求

検索者が入力した検索キーワードをもとに、そのキーワードを著者キーワードに含む文献を、該当文献として検索する。

各該当文献の著者キーワードも収集し、このキーワードは自己組織化マップの学習時に、各文献の特徴ベクトルとして用いる。

#### (2) 分類

キーワード検索によって得られた文献を、著者キーワードを元に自己組織化マップを用いて2次元平面に分類する。

#### (3) 表示

得られたマップと、各セルに対応付けられた文献の題目リスト、特徴ベクトルとして用いた著者キーワードを併せて表示する。

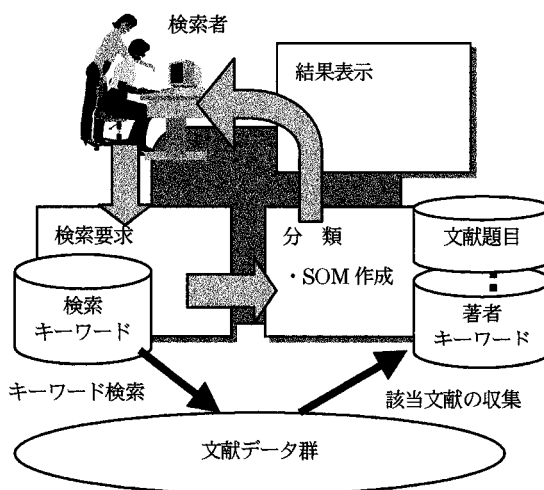


図2. システム構成

## 5. まとめ

検索結果として大量の文献が検索された場合に自己組織化マップを用いて文献間の関連が強いものを近く、弱いものを遠くに、2次元上に表示することで分類を行った。

検索結果を分類表示し文献の内容を比較する回数を減らすことで、絞込み検索のためのキーワード収集の支援を行った。

## 参考文献

- [1] T.コホネン, 徳嵩 他, "自己組織化マップ", シュブリンガー・フェラーク東京, 2000(初版4刷)
- [2] 長尾 他, "マルチメディア情報学 8", 岩波, 2000