

自己組織化マップと検索エンジンを用いた Web 文書の分類ビュー機構

波多野 賢治[†] 佐野 綾 一[†]
段 一 為[†] 田中 克己[†]

本論文では、Kohonen の自己組織化マップ (Self-Organizing Map) と既存の検索エンジンを用いた、Web 文書分類ビュー機構について述べる。提案する分類ビュー機構は、検索エンジンを用いて収集した Web 文書をユーザの視点や目的に応じて効果的に自動分類することを目的としている。現在、Web 文書に対する検索手段として様々な検索エンジンが提供されている。これらの検索エンジンの検索結果は通常膨大な量であり、検索された文書の一つ一つ閲覧していくことはユーザにとって大変な労力となる。よって、こういったシステムを補完する目的で、収集された Web 文書をユーザの視点や目的に応じて自動分類する機能が必要である。また、Web 文書の中には、一つ的话题をリンクでつながったいくつかの Web 文書で表現している場合があり、そういった文書群は現在の検索エンジンの検索結果には現れない場合が多い。本論文では、(1) 収集した Web 文書を自己組織化マップを用いて動的に分類し、(2) 得られた分類マップに対してユーザの視点を反映するための操作を対話的に施すことで、分類マップを段階的に修正できる機構を開発した。さらに、収集した Web 文書の分類を行う上で基本となる情報単位を個々の Web 文書とするのみならず、関連する Web 文書群を単位とした分類を行うための方式を提案している。本研究におけるユーザビューとは、ユーザの Web 文書の分類意図や検索目的をシステムに対して反映させる機構を意味している。

A Classification View Mechanism for Web Documents based on Self-Organizing Maps and Search Engines

KENJI HATANO,[†] RYOUICHI SANO,[†] YIWEI DUAN[†]
and KATSUMI TANAKA[†]

In this paper, we propose an effective classification view mechanism for hypertext data such as web documents based on Kohonen's Self-Organizing Map (SOM) and search engines. Web documents collected by search engines are automatically classified by SOM and the obtained SOMs are incrementally modified according to the user's point of view and their purposes. At present, various search engines are provided to retrieve web documents. When we use search engines to retrieve web documents, we get many answers as ever before, so we have a lot of labor to examine each web document. Therefore, in order to make up for search engines, we need a function to classify web document corresponding to the user's point of view and their purposes. Furthermore, we cannot retrieve pertinent web documents by conventional search engines, when a specific topic is often described by more than one web document. To solve these problems, we exploited a content-based clustering system for web documents. In this system, web documents are automatically clustered by their *feature vectors* produced from web documents or minimal subgraphs consisting of multiple web documents, and their overview maps are dynamically generated by SOM. Furthermore, we propose a method by which an obtained SOM is modified by user's interaction such as feedback operations. It is important how to reflect the aim of classification and the purpose of retrieval to this system. In our research, we intend to solve these problems by providing a view mechanism in which a basic unit of web document are determined by users and relevance feedback operations enables the generation of an overview map which reflects user needs.

1. はじめに

World Wide Web (WWW) の発展は我々に多様な情報を提供しており、近年の情報化社会においては必要不可欠なものとなってきている。しかし、WWW

[†] 神戸大学大学院自然科学研究科
Graduate School of Science and Technology,
Kobe University

で公開される情報量は圧倒的で、我々が必要とする情報を効率よく検索することを困難にしている。通常は必要とする情報を収集するために AltaVista¹⁾ に代表されるロボット型検索エンジンや Yahoo!²⁾ に代表されるディレクトリ型検索エンジンを利用してユーザがキーワードを入力し、現在数千万以上とも言われる Web 文書に対してフィルタリングを行うことで必要な情報に該当するページを得ている。この手法は WWW と DBMS を連動させたシステムの一つであり、膨大な量の情報に対して問い合わせを行い、その結果を表示することができる点で評価できる。しかし、この手法を利用しても問い合わせの結果は依然膨大であり、結果的に検索結果を一つ一つ閲覧していかねばならず大変な労力を要する。さらに、WWW は、膨大な量の情報資源からなるハイパーテキスト構造を構成しており、それぞれの情報資源はリンクにより結び付けられ Hypertext Markup Language (HTML) により半構造化されている。したがって、リンクでつながった複数の文書で 1 つの話題を表現する場合も考えられる。このとき従来型の検索エンジンは、各文書ごとにキーワード索引を生成するために、ユーザの問い合わせキーワード全てにマッチする文書のみを答えとして返すため、1 つの話題を複数の文書で表現したような文書群は現在の検索エンジンの問い合わせ結果には現れない場合が多い。

これらの問題点を解決するために、従来から Web 文書群全体の把握や曖昧検索、分類作業の支援などを目的として、Web 文書群の集約的な構造を把握する研究が数多く行われている^{2),7),15)}。また、大量の文書群の自動分類等にニューラルネットワーク技術を採用入れるといった研究^{10),14),24)}も行われている。しかし、いずれの場合においても収集した文書群の分類にユーザの視点や分類目的を反映させるような機構や、分類する文書の単位を選択できるような機構については十分な研究がなされているとは言い難い。

そこで本論文では、ロボット型の検索エンジンと、ニューラルネットワークの一種である自己組織化マップ (Self-Organizing Map, 以下 SOM と略す)¹¹⁾ による文書群の動的な分類・検索システム⁸⁾を統合し、ユーザの多様な検索目的を支援した Web 文書の分類ビュー機構を提案する。

本論文で提案する手法は、ユーザの多様な分類・検索意図をインタラクティブに支援するユーザビューを提供するものである。ここで挙げるユーザビューとは以下のようなものである。

- ユーザの視点を取り入れた Web 文書の分類

従来のシステム⁸⁾には、ユーザ側の意図や興味を分類マップに反映することができないという問題点があった。そこで、ユーザ自身が分類結果を把握しやすいように、システムが返す分類結果に対してフィードバックを行い、ユーザの視点を分類結果に反映させることでユーザ独自のビューを生成する。

- 分類対象である Web 文書の基本単位の決定
検索エンジンを利用した Web 文書の検索は、ユーザがあらかじめ明確な検索目標を持っているわけではないため、関連のある Web 文書群を検索の基本単位と考えたほうがよい場合がある。そこで、Web 文書群を一種のグラフ構造とみなして¹⁹⁾、問い合わせに対する Web 文書の極小部分グラフを生成し、これを基本単位として Web 文書の分類・検索を行う方式を提案する。この場合も、Web 文書の情報単位をユーザ自身が選択できるという点で、一種のユーザビューと考えることができる。

以下、第2章では基本的事項として Kohonen の自己組織化マップの原理と伝統的手法である適合フィードバックについて、第3章では Web 文書の分類のための情報単位とその特徴ベクトルの生成について、第4章では対話による分類マップの段階的修正について、第5章では本システムの実行例およびその評価について述べ、第6章で本研究の結論および問題点について述べる。

2. 基本的事項

2.1 自己組織化マップ (Self-Organizing Map: SOM)

ニューラルネットワークの一種である SOM は教師無し競合強化学習モデルである。出力層の各セルが層の中で位置を持つという点が他の学習モデルと異なる。データに隠されているトポロジカルな構造を学習アルゴリズムにより発見し、通常 2 次元空間で表示するという特徴を持っているため、特徴のよく似たデータ同士は出力マップ上の近い位置に配置されるようになっている。生成されたマップはそれぞれのデータの位置関係によって、類似しているデータかどうかが直観的に理解しやすいという点からシステムの視覚化に利用できる。SOM には様々な種類があるが、ここでは最も基本的なものについて述べる。

SOM で用いられるネットワークは、セルを 2 次元に六角格子状に配置したものである。それぞれのセル i はセルの特徴ベクトル $\mathbf{m}_i(t) \in R^n$ (R は実数) を持っており (t は時間を表し、 $\mathbf{m}_i(0)$ は適切な方法で

初期化されている), これらのセルの特徴ベクトルを, 入力である特徴ベクトル $\boldsymbol{x}_j \in R^n (j = 1, 2, \dots, d)$ に選択的に近づけることによって学習は進行する. このとき, SOM では入力となる特徴ベクトルに一番近いパターンを持つ出力セルおよびその近傍のセルの集合のみが入力ベクトルに近づくことができるようなアルゴリズムをとる.

2.2 適合フィードバック (Relevance Feedback)

情報検索システムを構築する上での大きな問題の 1 つは, ユーザがどのような検索意図を持っているときにどのような問い合わせがなされるのか, あらかじめ予測できないため, ユーザが入力した問い合わせに対して検索意図に沿った検索結果が返ってくる保証がないという点である.

このような状況を解決するには, システムがユーザの視点や分類目的をインタラクティブに支援すべきである. このインタラクティブな情報検索の伝統的な手法の 1 つに適合フィードバック (Relevance Feedback)¹⁶⁾ という手法が存在する. この手法は初期の検索結果に対してユーザが適合または不適合と判断した検索結果を基にシステムが新たな問い合わせを生成し, 検索結果にフィードバックをかけながら問い合わせを繰り返すことで, 徐々にユーザの情報要求に近づけていくというものである. 近年の検索エンジンにはこのような機能が付加されたものも存在する^{1), 5)}.

本論文における適合フィードバックとは, 伝統的な適合フィードバックのような検索結果に対してのフィードバックではない. むしろ検索エンジンの検索結果を自動分類した分類マップに対するフィードバックのことであり, 分類マップに対してユーザの分類視点や分類データの基本単位の決定などの分類目的をインタラクティブに反映させるために利用するものである.

3. Web 文書の分類のための特徴ベクトル生成

我々は, 従来より SOM を文書データベースのブラウジングツールおよび問い合わせインターフェースと見なし, 文書の自動的な構造化や分類を支援するシステムを開発してきた⁸⁾. しかし, このシステムでは,

- マップ上のキーワード同士の関連を見ながらの検索だけでは, 目的の情報にたどり着くまでに時間がかかる.
- ユーザ側の意図や興味をマップに反映することができないため, ユーザ本位のインターフェースを提供できない.

といった問題点があった.

このような問題点を解決するために, システムに対してユーザのインタラクションを取り入れる研究が始まっている. Kohonen の研究グループでは, 文書中に現れる単語のカテゴリ分類を行った上で文書の分類に SOM を利用し, カテゴリ分類の際に単語を選択したり, データをブラウズする際にマップ上でデータの詳細表示を行うといったユーザの操作がシステムに反映されるような工夫がなされている^{10), 14)}. また, 江口らは伝統的な適合フィードバックを利用し, ユーザの興味の変化を検出しシステムがユーザの検索目標に適應できるような方式を提案しており²³⁾, Xerox Palo Alto Research Center でも収集した情報に対してユーザのインタラクションを分類に反映することができる Scatter/Gather を開発している⁹⁾.

これに対し, 我々はこれらの研究では行われていない収集した情報の分類ビューを利用し, 情報の分類傾向を把握しながら分類ビューに対してユーザのインタラクションをサポートする機能を実現することで問題点を解決する. すなわち分類・検索対象となる Web 文書の情報単位の決定やシステムによって返された分類結果に対して, ユーザの視点や分類目的をシステムにインタラクティブに反映させる機能を持った Web 文書分類ビュー機構を実現する. 本システムの全体構成を図 1 に示す.

3.1 Web 文書の特徴ベクトルの生成

SOM を利用した検索エンジンの検索結果に対して情報の組織化を行うには, 検索エンジンに対する問い合わせの結果から各 Web 文書の Uniform Resource Locators (URL) を抜き出し, それにしたがって Web 文書自体を解析し特徴を抽出しなければならない. 本論文では, 検索エンジンによって検索された各 Web 文書に対して単語抽出を行い, これを基に各 Web 文書の情報単位ごとに特徴ベクトルを構成する. すなわち, 特徴ベクトルの基底は Web 文書から抽出された単語を表しベクトルの各成分はそれらの重要度を表す. 以下に, 特徴ベクトルの生成過程を示す.

(1) 検索エンジンの検索結果の獲得

通常の実験エンジンで検索を行う場合と同様にユーザは Web ブラウザ上でキーワード k_1, k_2, \dots, k_m を入力する. つまり, 検索エンジンに対して検索質問 $q = k_1 \wedge k_2 \wedge \dots \wedge k_m (m \geq 1)$ を問い合わせると, 検索結果である URL が返され, それを利用して検索された Web 文書のソースを得る.

(2) Web 文書からの単語の抽出

位として検索が可能となる。また、Web 文書の分類・検索の単位を Web 文書単体か極小部分グラフかを定めるのはユーザであるため、分類・検索するデータの単位に関する一種のユーザビューの生成がなされていると考えられ、図1に示した Web 文書分類ビュー機構に組み込むことが可能である。以下に、その概要を示す。

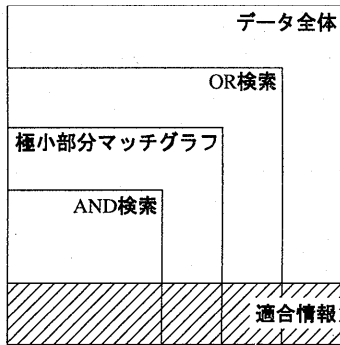


図2 検索法の違いによる検索可能な範囲

Fig. 2 Retrieval Results depend on Difference of Search Methods

3.2.1 極小部分マッチンググラフ

本小節では、極小部分マッチンググラフの定義を行う。問い合わせを行う Web 文書のデータベースを

$$DB = \{U_1, U_2, \dots, U_l\}, \quad (l \geq 1)$$

と定義する。ただし、各 $U_i (i = 1, \dots, l)$ は各 Web 文書をノード、文書間をつなぐリンクを有向枝とする互いに素な連結グラフであり、 $U_j \cap U_k = \phi (j, k \in \{1, \dots, l\}, j \neq k)$ となる。

次に、このデータベースに対する問い合わせ $Q = k_1 \wedge \dots \wedge k_m (m \geq 1)$ の解として得られる各連結グラフ U の極小部分マッチンググラフ $G = (V, E)$ を以下のように定義する (図3参照)。

- (1) G は U の部分グラフである。
- (2) 各キーワード $k_j (j = 1, \dots, m)$ に対し、 G 内の少なくとも 1 つの節点 (に該当する Web 文書) がこれを含む。
- (3) G のいかなる部分グラフ G' も上記 (1), (2) を満足しない。

3.2.2 極小部分マッチンググラフの評価尺度

極小部分マッチンググラフをデータベースに対する問い合わせの解と考えた場合も、従来の検索方法の場合と同様、解として膨大な量の極小部分マッチンググラフが返される。よって、検索エンジンの場合と同様それら極小部分マッチンググラフのスコアリングを行うための評価式を定義し、極小部分マッチンググラフを評価する必要がある。

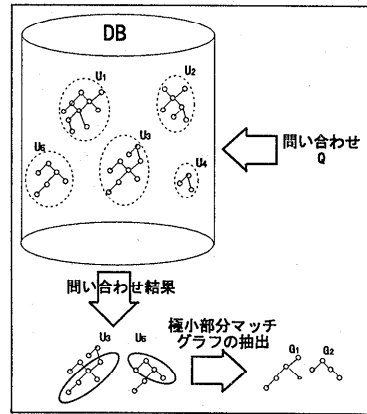


図3 検索モデルの事例

Fig. 3 Example of Our Data Model

ある。ここで、極小部分マッチンググラフを評価する際には、以下の点を考慮する必要がある。

- 極小部分マッチンググラフの「直径」

ここでいう「直径」とは部分グラフの最長パスのノード数を表す。

極小部分マッチンググラフの直径が大きくなると、どのキーワードにもマッチしないノードを内部に含む可能性が高くなり、質問に無関係な Web 文書を多く含むことになる。つまり、このような極小部分マッチンググラフのランク付けを行う必要があると考えられる。

- 各節点のキーワード数

極小部分マッチンググラフの各節点が、質問で指定されたキーワードをどの程度有しているかも重要な尺度となる。

これらを表現するために本研究では評価式として以下の式を提案する。

$$\text{評価尺度1} = \frac{1}{d} \quad (2)$$

$$\text{評価尺度2} = \frac{1}{\sum_{v \in V_i} \frac{m}{n \cdot K(v)}} \quad (3)$$

ただし、 n は極小部分マッチンググラフ中のノード数、 d はグラフの直径、 m は問い合わせ Q のキーワード数、 v は直径上のノード、 V_i は直径上のノード集合、 $K(v)$ は極小部分マッチンググラフの各節点 v における Q にマッチするキーワード数を表す。ただし、式3において、 $K(v) = 0$ の場合は $\frac{m}{n \cdot K(v)} = C (C \geq 1$ である定数) とする。

式2は、グラフの直径のみを考慮した評価式であり、部分グラフの大きさが大きくなればなるほど、どのキーワードにもマッチしないノードを含む確率が高くなるためスコアが悪くなる。また、式3は、グラフの

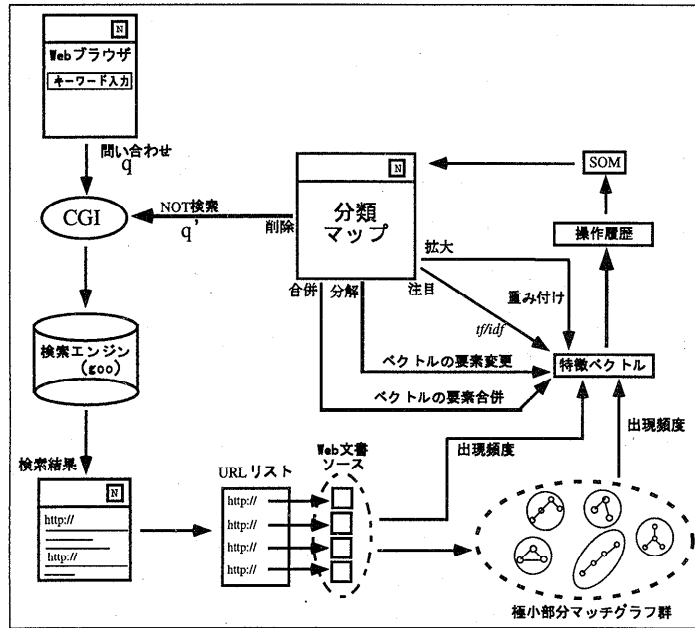


図1 システムの全体構成

Fig. 1 Our Classification View Mechanism for Web Documents

得られた Web 文書のソースから単語を抽出しそれぞれの総出現頻度を求める。このとき、システムの仕様上、総出現頻度の高いものから順に上位 500 の単語 w_1, \dots, w_{500} を特徴ベクトルの基底とする。

(3) Web 文書の特徴ベクトルの生成

Web 文書 D_i における各基底 w_1, \dots, w_{500} の重要度を求め、各文書の特徴ベクトル $F(D_i)$ を生成する。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \frac{f_2^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (1)$$

ここで、 $f_j^i (j = 1, \dots, 500)$ は Web 文書 D_i における単語 w_j の出現頻度、 N_i は D_i 内に現れる単語の総出現頻度を表す。

(4) Web 文書の特徴ベクトルの学習

こうして生成された Web 文書の特徴ベクトル $F(D_i)$ を SOM に入力して学習し、Web 文書を自動分類した分類マップ 3D-SOM を得る。

3.2 Web 文書の情報単位

Web 文書はハイパーテキストの一種であり、互いにリンクによってつながれた構造を形成している。こうした Web 文書の構造的特徴を利用した Web 文書の検索に関する研究が近年行われている。Weiss らは Web 文書中に含まれている単語の類似度と各々のリンクの関係を共に利用した Web 文書の分類を行っている²⁰⁾。しかし、この研究は各 Web 文書を単にハイ

パーテキスト中のノードとして取り扱っているため、検索時にユーザがいくつかのキーワードを用いて問い合わせを行った場合は、それら全てのキーワードを含んだ Web 文書のみが答えとして返ってくる。つまり、問い合わせキーワードのうちいくつかを含んでいる Web 文書や Web 文書制作者がある話題についていくつかの Web 文書にまたがって記述した場合など関連のある文書の検索に問題を生じることとなる。また、Web 文書の検索はユーザがあらかじめ欲している情報について明確にその回答を持っているわけではないため、AND 検索のような絞り込みを行っても新たな検索目標を発見しにくい。

したがって、本論文ではこういった問題を解決するために、SOM に入力する特徴ベクトルを Web 文書単位で生成するのではなく、ある話題を基に Web 文書のハイパーテキスト構造を分割して生成される極小部分グラフを単位として特徴ベクトルを生成し、Web 文書を意味情報単位で分類・検索を行う方法の提案を行う。この提案は図2に示すように AND 検索および OR 検索の中間的な検索*を行うことを目的としている。つまり、従来の AND や OR 検索では検索が不可能であった関連のある Web 文書群を 1 つの情報単

* 極小部分マッチンググラフはデータの基本単位が他の場合と異なり複数の Web 文書群であるが、極小部分マッチンググラフを構成している文書に置き換えて考えると、OR 検索では検索されるが AND 検索では検索できなかった文書と考えることができる。

極小部分 マッチンググラフ	評価尺度1 のスコア	評価尺度2 のスコア
(1) ○ k ₁ k ₂ k ₃	1	1
(2) ○—○ k ₁ k ₂ k ₃	1/2	4/9
(3) ○—○—○ k ₁ k ₂ k ₃	1/3	1/3
(4) ○—○—○ k ₁ k ₂ k ₃	1/3	2/5
(5) ○—○—○ k ₁ k ₂ k ₃	1/3	1/2

図4 評価の一例

Fig. 4 Example of Our Evaluation

直径だけではなくグラフ中の各ノードが問い合わせ Q にどれほどマッチしているかについて考慮した評価式である。それぞれの評価式により図4の例のようにグラフのスコアが計算される。式2の場合、図中の(3), (4), (5)のようにグラフの構造が同じであるが、ノードにマッチするキーワードが異なるような場合でも同じスコアが与えられてしまうという欠点を持つ。そこで、式3は問い合わせ Q に対する各ノードのマッチキーワード数を考慮した。ただし、図中の(4), (5)のようにノイズのノードが含まれている場合も考えられ定数 $C(C \geq 1)$ を定めている。 C はこれらノイズのノードをどう扱うかにより変化させることができ、 C を大きくすればノイズが数多く含まれるとスコアが悪くなる。図4の例では、 $C = 1$ として多少のノイズに対しては無視している。

3.2.3 極小部分マッチンググラフの抽出

極小部分マッチンググラフを抽出する方法を以下に述べる。

- (1) 問い合わせ Q を $Q = k_1 \wedge \dots \wedge k_m$ とする。本モデルでは、答えの Web 文書には全てのキーワード k_1, \dots, k_m が含まれている必要がないため、検索エンジンに対する問い合わせは OR 検索である $q = k_1 \vee \dots \vee k_m$ となる。
- (2) 問い合わせ q により返された Web 文書それぞれのリンク構造を調べ、各連結グラフ U_i を生成する。
- (3) 生成された各連結グラフの親ノードから順に、各ノードのマッチするキーワードを調べ、 U_i の中から極小部分マッチンググラフを抽出する。例えば、図5のように、問い合わせが $Q = k_1 \wedge k_2$ 、部分グラフ U が与えられた場合、(1) ~ (5) のような極小部分マッチンググラフが得られる。

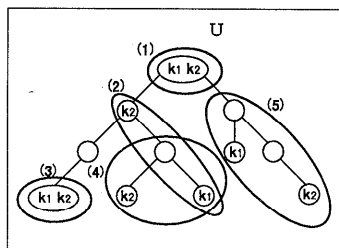


図5 極小部分マッチンググラフの抽出

Fig. 5 Extraction of Minimal Subgraph

3.2.4 極小部分マッチンググラフの自動分類

抽出された極小部分マッチンググラフを Web 文書単体の場合と同様に自動分類するためには、極小部分マッチンググラフの特徴ベクトルを生成しなければならない。以下に、極小部分マッチンググラフ G_i の特徴ベクトルを生成する方法について説明する。

- (1) 極小部分マッチンググラフの各ノードは Web 文書 D_j であるため、3.1節に述べたように各ノードの特徴ベクトル $F(D_j)$ を求める。この $F(D_j)$ の生成方法には 2 通りあり、単純にノード内に出現する単語の出現頻度を基に計算する方法と、 tf/idf 法^{17),18)} を基に計算する方法が考えられる。 tf/idf 法は情報検索の分野で用いられている方法だが、WWW における情報検索においてもその有効性が確認されている²²⁾。
- (2) 本論文では、極小部分マッチンググラフは各ノードの集合であるため、極小部分マッチンググラフの特徴ベクトル $F(G_i)$ も各ノードの特徴ベクトル $F(D_j)$ の和であると考えられる。ただし、各ノードは

- 問い合わせキーワードに対するマッチ率
- グラフの形状 (直径上のノードかどうか)
- グラフ内における親子関係

によって各々重要度が異なっているため、重み $\alpha(D_i)$ を考慮に入れる必要がある。 $\alpha(D_i)$ には、極小部分マッチンググラフを評価するために提案した式3を利用する。

$$\alpha(D_i) = \frac{1}{\sum_{v \in V_i} \frac{m}{n \cdot K(v)}} \quad (4)$$

したがって、極小部分マッチンググラフの特徴ベクトル $F(G_i)$ は以下のように求められる。

$$F(G_i) = \frac{\sum_j \alpha(D_j) \cdot F(D_j)}{n} \quad (5)$$

ここで、 n は極小部分マッチンググラフのノード数である。

こうして得られた極小部分マッチンググラフの特徴ベクトル

ル $F(G_i)$ を Web 文書単体で特徴ベクトルを生成した場合と同様に, SOM の入力として学習させ, 分類マップを生成することで自動分類が可能となる. これらの機能は, ユーザの問い合わせ時に Web 文書単体または極小部分マッチグラフで分類・検索を行うのかを選択するという形で実現する予定である.

3.2.5 極小部分マッチグラフの予備実験

現段階では極小部分マッチグラフを扱うことが可能な分類ビュー機構が未実装であるため, ここでは予備実験を行うことで WWW 情報検索における極小部分マッチグラフの有用性を示す. 予備実験では, 検索エンジン goo を利用し, 3 つの検索キーワードを与えて検索されたデータに対しての評価は適合率と再現率を用いた. さて, 適合率・再現率を算出する際, 全ての Web 文書を閲覧して利用者の意図に合致した Web 文書 (以後この Web 文書を「適合文書」と呼ぶ) を得るのは実質上不可能である. そこで本予備実験では, OR 検索で検索された Web 文書の中からランダムに抽出し分類・検索の母集団を定めてから, その母集団に対して我々が適合文書を定める. この適合文書の割合から OR 検索で検索された Web 文書の中に存在する適合文書を算出し, さらに検索エンジンによって検索された Web 文書, されなかった Web 文書を算出することで, 適合率・再現率を計算する.

表1 分類・検索単位の違いによる適合率と再現率

Table 1 Precision and Recall Ratios depend on Difference of the Basic Unit of Data Clustering

検索対象	文書数	適合率 (%)	再現率 (%)
Web 文書 (AND)	9	11.11	6.605×10^{-4}
極小部分グラフ	294	21.43	4.161
Web 文書 (OR)	20531	7.373	100.0

表1における「文書数」とは検索エンジンによって検索された Web 文書数を表している. これらの値を見れば分かるように, 分類・検索の単位を Web 文書単体にするよりは極小部分マッチグラフを利用した方が, AND 検索を行う場合に比べ検索されるデータ数が多くかつ OR 検索のように膨大なデータ数とはならない. また, 対象データを関連のある Web 文書群である極小部分マッチグラフとしているため, AND 検索では検索することができなかったような関連のある Web 文書も同時に検索することが可能となっている. つまり, AND 検索および OR 検索の中間的な検索を行うという当初の目的を達成することができた. しかし, この条件下で分類精度を評価を行っても実質的な評価とはなりえない. なぜなら, 評価対象のデータ母集団

が通常の AND 検索や OR 検索は Web 文書単体であるのに対し極小部分マッチグラフの場合は Web 文書群であるため, 表1における極小部分マッチグラフの評価はグラフ中に 1 つでも適合情報にマッチしている Web 文書があれば適合したグラフであると評価しており, 評価の場が異なるからである. また, 関連ある Web 文書群には, 本来, 図6のようにトピックに関する並列構造や階層構造を持ったものが考えられるため, これらの構造を考慮して極小部分マッチグラフの評価を行う必要もあると考えられるからである.

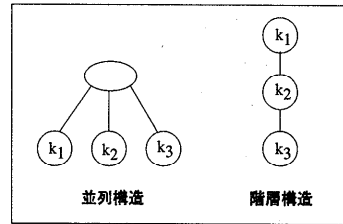


図6 極小部分グラフの構造

Fig. 6 Structures of Minimal Subgraph

4. ユーザとの対話による分類マップの段階的修正

SOM は, 入力データ群をトポロジカルマッピングの性質により 2 次元空間に表示することができるため, 分類マップを生成するには大変有効な手段であると考えられる. そこで, 本論文では SOM に Virtual Reality Modeling Language (VRML) ベースの 3 次元インターフェースを採用し, ユーザがマップ中のセル (円柱) を選択することでそのセルに分類された Web 文書の URL とセルが持っている重要な 3 つのキーワードを Web ブラウザで参照できるように設計している. また, 分類結果に対して適合フィードバックの機能を実現することで, 検索された Web 文書の分類をユーザ自身で変更することが可能となっているため, それらの分類状況を把握しやすいようになっている. 以下にユーザインターフェースにおける各機能を説明する.

● マップの 3 次元表示

マップの出力セルは円柱になっており, その高さはそのセルに分類された Web 文書の数に比例している. したがって, 分類マップの外観から分類されている Web 文書の数を把握できる.

● 各出力セルのキーワード

各出力セルの特徴ベクトルの成分中, 最も大きな値を持つ成分, すなわちその出力セルのベクトル

の成分の中で最も値の大きな単語をそのセルのキーワードとする。同様に、成分が 2 番目に高いもの、3 番目に高いものを選び各セルを 3 つのキーワードでラベルづけを行う。以下、これらのキーワードをそれぞれ第 1 キーワード、第 2 キーワード、第 3 キーワードと呼ぶ。

- マップの領域分割

生成された分類マップは、各セルにつけられた第 1 キーワードを利用して色分けされる。SOM の性質上、第 1 キーワードが等しいセルはマップ中の近い位置に配置されるため、分類マップは第 1 キーワードによる領域形成がなされる。

- マップの詳細表示

第 1 キーワードによって色分けされた領域を選択すると、その領域をさらに細かく表現した詳細マップが表示される。この詳細マップは第 2 キーワードを利用して色分けされたものであり、関連した Web 文書を検索する場合に利用できる。同様に、この詳細マップをさらに詳細に見ることも可能である。

- 適合フィードバック機構

得られた分類マップに対して、ユーザの意図や興味を反映させる機能は、ユーザが欲している情報をフィードバックを行うことで絞り込み、効率的に検索する上で大変重要な機能である。Scatter/Gather⁹⁾においてはこの機能を利用して検索の高速化に役立っている。本システムにおいてもこれにない、以下のような適合フィードバック機構を実現している。ただし、以下では文書群の特徴ベクトルを Web 文書単体から生成された $F(D_i)$ として説明している、3.2 節で前述した極小部分マッチグラフから生成された $F(G_i)$ も利用できる。

(1) 領域の拡大

ユーザがある領域 K に関する Web 文書を欲している場合その領域を指定すると、その領域内の Web 文書のうち特徴ベクトルの成分 $\frac{f_i^i}{N_i}$ が最も大きいものを探し出し、その文書が分類されているセル c を中心にすべての $F(D_i)$ の成分 $\frac{f_i^i}{N_i}$ に対して同心円状に重み付けの再計算を行うものである。具体的には以下のように再計算が行われる。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \dots, \alpha(r) \cdot \frac{f_j^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (6)$$

$$\alpha(r) = \frac{\alpha(r-1)+1}{2}, \quad r \geq 1,$$

$$\alpha(0) = 5$$

ここで、 r はセル c から D_i が含まれているセルまでの距離 (半径) を表す。この特徴ベクトルを用いて再度 SOM によって学習させることで、キーワード K の重要度が高い Web 文書を分類マップ上に多く浮かび上がらせることができる。

(2) 領域の削除

ユーザがマップ上のある領域 K は全く必要ないと判断した場合、その領域を指定するとキーワード K を含んだ Web 文書自体を検索対象から削除する、すなわちユーザビューが検索エンジンに対して NOT 検索を行う機能を実現している。この場合、ユーザは意識することなく新たな質問 q' を生成し検索エンジンに対して問い合わせができるようになっており、新たな分類マップを得ることができる。

$$q' = k_1 \wedge k_2 \wedge \dots \wedge k_m \wedge \neg K$$

(3) 領域の合併

分類マップ上の複数の領域 (キーワード) $K = \{w_l | l \in \{1, \dots, 500\}\}$ を 1 つに合併したい場合に用いる方法である。例えばキーワードに “red” や “white” など色に関するキーワードが数多く出ている場合、これらを “color” などと 1 つのキーワードにまとめて見やすいマップを再構成することを目的としている。

$$F(D_i) = \left(\frac{f_1^i}{N_i}, \dots, \sum_l \frac{f_{w_l}^i}{N_i}, \dots, \frac{f_{500}^i}{N_i} \right) \quad (7)$$

(4) 領域の注目

ユーザがマップ上のある領域群に注目したい場合、それらの領域内にある Web 文書だけで特徴ベクトルを再生成し、分類マップを再構成するものである。その際に、単純に単語の出現頻度を利用した特徴ベクトルを生成するのではなく、*tf/idf* 法を用いる。実際に特徴ベクトルを生成する際には、3.1 節と同様に行うが、重要度の大きさは出現頻度を基に計算する場合とは異なり *tf/idf* 法を基に重要度が算出される。値の大きい順に上位 500 の単語 w'_1, \dots, w'_{500} を抽出した場合、Web 文書 $D_i (i = 1, \dots, M)$ の

tf および idf は

$$tf_j^i = \frac{f_j^i}{N_i}, \quad idf_j^i = \log\left(\frac{M}{df_j^i}\right)$$

のように計算される。ただし、 f_j^i ($j = 1, \dots, 500$) は Web 文書 D_i に含まれている単語 w_j^i の出現頻度、 df_j^i は単語 w_j^i を含む Web 文書 D_i の数を表す。したがって、 D_i の特徴ベクトルは以下ようになる。

$$\mathbf{F}(D_i) = (f_1^i, f_2^i, \dots, f_{500}^i), \quad (8)$$

$$f_j^i = tf_j^i \cdot idf_j^i$$

この方式を採用した理由は、分類マップ全体を概観する際には、単純に出現頻度による特徴ベクトルを利用することがよく、目的の情報を絞り込んでいくという際には tf/idf 法を利用することがよいことが実験からわかっているためである⁸⁾。よって、この機能はユーザの欲している文書を絞り込む際に有効な手段であると考えられることができる。

(5) 領域の分解

Web 文書は世界中のサイトに散らばっていることを考えると、言語の問題が生じてくる。つまり、ユーザになじみのない単語が分類マップ上に現れることも考えられる。この方法は、そういった意味不明な単語によるマップの領域形成を防ぐために用いられるものである。この場合も、重要度の高い単語が変化するため上位 500 の単語 w_1^i, \dots, w_{500}^i を抽出しなおす必要がある。

$$\mathbf{F}(D_i) = \left(\frac{f_1^i}{N_i}, \frac{f_2^i}{N_i}, \dots, \frac{f_{500}^i}{N_i}\right) \quad (9)$$

また、過去に行ったフィードバック操作の履歴を利用することで、同類の問い合わせに対して過去の適合フィードバック操作を利用した分類ビューもユーザに提供でき、ユーザ独自のビューを生成することに応用できる。

5. Web 文書分類ビュー機構の実行例とその評価

5.1 分類ビュー機構の実行例

実際に 20×20 の大きさのマップを用いて分類マップの生成を行った。本論文では、Web 文書を検索する検索エンジンとして **goo**⁶⁾ を利用した。検索エンジンに対して “worldcup” という検索キーワードで得られた検索結果から、検索エンジンの検索結果の高スコア順に 400 件を抽出し、3.1 節の方式に従って Web

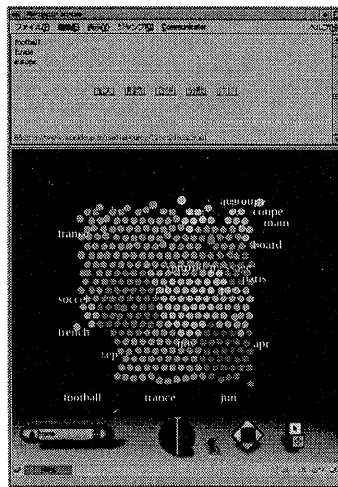


図7 分類マップ概観

Fig. 7 Overview Map for Web Documents

文書単体から特徴ベクトルを生成を行っている。特徴ベクトルの学習には Kohonen の研究グループで開発された **SOM.PAK**¹³⁾ を改良し、米・シリコングラフィックス社のワークステーション、Origin 200 上で学習させると、セル数 400 の分類マップは、ユーザが問い合わせを発行してから 5 分程度で生成された^{*}。

生成された分類マップの様子を図7に示す。この図を見ればわかるようにマップは、“world”, “football”, “france”, “soccer”, “description”, “apr”, “coupe”, “french”, “communication”, “paris”, “jun”, “english”, “ski”, “korea”, “main”, “oct”, “dec”, “nov”, “sep” のキーワードによって、19 の領域に分割されている。

4章で述べたようなユーザインターフェースの機能により、ユーザはこの 3 次元マップを検索エンジンの検索結果のブラウジングツールおよび問い合わせツールとして利用できる。分類マップ上のセルを選択するとすると、3 つのキーワード (図7の場合は、“football”, “finale”, “europe”), 適合フィードバックへのリンク、およびセルに振り分けられた URL のデータが Web ブラウザに表示される。また、マップ上に表示されているラベルを選択するとその領域の詳細マップを見ることができる。ここで、適合フィードバック機能が選択されると、マップ上のどの領域 (キーワード) に対して適合フィードバックを行うかの選択画面 (図8) が表示され、それらを選択することでそれぞれの領域に対して適合フィードバックが行われるようになっている。

^{*} 検索エンジンが検索する対象を海外のサイトとした場合の数値。対象を国内のサイトとした場合はさらに日本語の品詞分解を行う必要があるため 10 分程度の時間を要する。

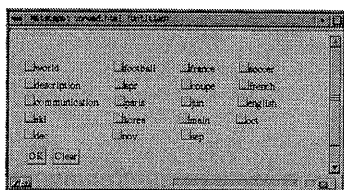


図8 キーワード選択画面

Fig. 8 Selection of Keywords

以下、それぞれの適合フィードバックの実行例を示す。

(1) 領域の拡大

図7に対して、“france”というキーワードで、「拡大」の適合フィードバックを行った。マップ上で“france”を第1のキーワード、第2のキーワード、第3のキーワードとして分類されているWeb文書数の変化は表2のようになった。表を見ればわかるように、“france”に関連する文書をマップ上に数多く浮かび上がらせることが可能である。

表2 拡大によるマップ上のキーワード数の変化

Table 2 Number of Keywords in each level 3D-SOM (“enlarge” Operation)

	拡大前(個)	拡大後(個)
第1キーワード	27	32
第2キーワード	15	17
第3キーワード	1	9
合計	43	58

(2) 領域の削除

図7に対して、“ski”というキーワードで、「削除」の適合フィードバックを行ってみた結果、図9のようなマップが生成された。マップは、“description”, “dec”, “jun”, “sep”, “nov”, “apr”, “french”, “soccer”, “jul”, “prev”, “paris”, “sports”, “world”, “message”, “board”, “france”, “main”, “usa”, “coupe”, “aus”, “groupe”のキーワードによって、21の領域に分割され、“ski”というキーワードがマップ上には表れず、また、検索エンジンに対してNOT検索を行ったため検索対象となるデータ群の中にも“ski”に関連した文献は見当たらなかった。

(3) 領域の合併

「合併」は、複数のキーワードの特徴ベクトルの成分を加えて、ある1つのキーワードに置き換える機能である。最初に生成された図7のマップ上の、“apr”, “jun”, “oct”, “dec”, “nov”,

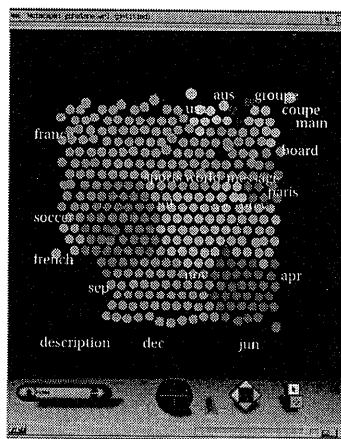


図9 削除の適合フィードバック

Fig. 9 “delete” Operation

“sep”のキーワードを“month”というキーワードに置き換えて、特徴ベクトルを再生成しSOMに学習させてみたところ、“description”, “french”, “month”, “soccer”, “korea”, “paris”, “ital”, “usa”, “main”, “world”, “green”, “france”, “groupe”, “communication”, “coupe”, “ski”, “english”のキーワードによって、17の領域に分割された。図10のマップにおいて、合併された“month”という領域がマップの右下の4分の1を占める大きな領域を作っているのが分かる。

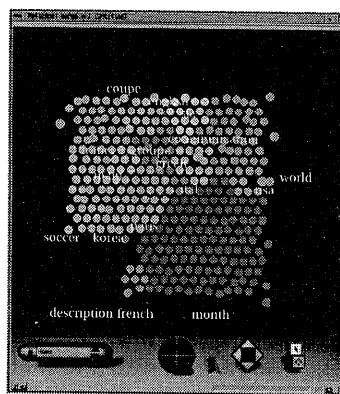


図10 合併の適合フィードバック

Fig. 10 “merge” Operation

(4) 領域の注目

「注目」は、ユーザの欲している情報を最終的に絞り込むために使われる機能である。ここで、出現頻度によって「注目」の機能を実現した場合と *tf/idf* 法によって実現した場合のマップの

分割状況を見てみると、出現頻度の場合には 25 個、*tf/idf* 法を利用した場合は 35 個の領域に分割された。つまり、*tf/idf* 法を利用したほうがそれぞれの文書特有のキーワードにより領域が細かく分類されることで、ユーザが欲する文書を特定する際に非常に有用であることがわかる。

(5) 領域の分解

図 7 において、“coupe” や “main” といったキーワードは、何を表現しているものなのかユーザには非常にわかりにくい。こうした意味不明の単語による領域形成を防ぐための操作であるため「領域の削除」の操作と似ているが、検索エンジンに問い合わせの過程がない点が異なる。

5.2 分類ビュー機構の評価

Web 文書分類ビュー機構の評価を、分類マップ上に分類されたデータの適合率と再現率を用いて行った。極小部分マッチンググラフを用いた分類ビューの評価はシステムが未実装であるためここでは議論しない。さて、本評価実験における適合率と再現率は以下のように定義される。

全ての Web 文書の中から、マップ上において n 個の領域を作っているキーワード $k_i (i = 1, \dots, n)$ を特徴として表す Web 文書の数を $similar(k_i)$ とし、 k_i の領域に分類されている Web 文書の数を $neighbour(k_i)$ で表わすとすると、適合率は、

$$\frac{1}{n} \cdot \sum_{i=1}^n \frac{|neighbour(k_i) \cap similar(k_i)|}{|neighbour(k_i)|}$$

で表わされ、また再現率は、

$$\frac{1}{n} \cdot \sum_{i=1}^n \frac{|neighbour(k_i) \cap similar(k_i)|}{|similar(k_i)|}$$

で表わされる。

今回の様に検索エンジンの検索結果というものは、「情報が適合する、しない」という評価は主観的であり、評価者によって異なるものである。ここでは第一著者の基準でページの内容の評価を行い、 $similar(k_i)$ を定めた。検索エンジンの出力結果に適合フィードバックを施す前と施した後の適合率と再現率の変化を表 3 に示す。

表 3 が示しているように、適合フィードバックを行うことにより適合率および再現率が上昇しているため、検索時に目的の情報にたどり着くまでのユーザの労力が軽減されており、これらのフィードバック機能の有効性が確かめられている。

表 3 適合フィードバックによる適合率と再現率の変化
Table 3 Precision and Recall Ratios depend on Relevance Feedback Operations

	適合率 (%)	再現率 (%)
適合フィードバック前	37.44	34.50
適合フィードバック後	48.82	44.84

しかし、現在のところ以下のような問題点を抱えており、それぞれの問題点に対処していく必要があると思われる。

- 評価の際にフィードバック前後の分類マップの各領域の適合率および再現率を行っているが、フィードバックによる検索対象データ群の変化や、各 Web 文書の特徴ベクトルの変化について考慮していないため、これらの数値の評価がそのままシステムの評価とはなり得ない。つまり、新しい評価尺度を提案すべきである。
- 検索エンジン固有のランキング手法について全く考慮せず、検索対象データ 400 件を選定している。したがって検索エンジンのランキングに影響されないような検索対象データの選定方法を提案すべきである。
- ユーザがシステムに対して問い合わせを発行してから検索結果が返ってくるまで 5 分とユーザのインタラクションを反映した検索システムである Scatter/Gather に比べ実用的ではない。したがって、現在のように問い合わせの度に検索対象データをネットワーク上に収集しに行くようなシステム構成ではなく、Web ロボットを利用してあらかじめデータを収集しデータベースに格納されたデータに対して、問い合わせを行うようなシステムを構築し時間短縮を図るべきである。これにより、Scatter/Gather と同様なインタラクションの機能をサポートする Web 文書の極小部分マッチンググラフの分類ビュー機構の実現が図られる。
- Web 文書の分類に関しては、ユーザのインタラクションを反映できるインタフェースとなっているが、検索に関してはインタラクションをサポートする機能は存在しない。したがって、検索作業に関するフィードバック機能もサポートすべきである。
- 本評価における適合率、再現率の値がそれほど良い値ではない。これに関しては、これまでの研究で SOM の学習パラメータや特徴ベクトル生成アルゴリズムを適切に設定すればよいことが分っている。ただし、これらパラメータの設定は、経験則的な部分があるため最適値を設定するのが困難

であるという問題がある。

6. おわりに

本論文では、ニューラルネットワークの一種である SOM を用いて、ユーザの多様な分類・検索意図をインタラクティブに支援した Web 文書の自動分類ビュー機を実現し、その一部機能について実装を行った。

本分類ビュー機構の利点は、

- 検索エンジンに対する問い合わせの答えが、従来のようにリスト表示ではなく 3 次元分類マップにより提供されるので、ユーザは分類状況全体を把握しながら検索することができる。つまり、分類を行っていくにしたがって、ユーザ自身がどういった情報を要求しているのかを理解しながら検索を行うことが可能となる。
- システムがユーザの分類意図や検索目的をインタラクティブに支援することができるため、ユーザ本位の分類マップを生成することができる。つまり、検索効率の向上を図ることが可能である。
- 検索対象に極小部分マッチンググラフ、すなわちある話題に関する Web 文書群を検索出来るようになれば、従来では検索できなかった Web 文書を検索できる可能性がある。

ことが挙げられる。さらに、今後の課題として以下のような問題が挙げられる。

- 検索エンジンで算出されているスコアの利用および高速化を考慮した、Web 文書の特徴ベクトルの生成および SOM 学習アルゴリズムの改良。また、SOM 学習パラメータの動的最適設定法の確立。
- 適合フィードバック機能による検索対象データ母集団の変化や Web 文書特徴ベクトルの変化を考慮した、本システムの評価法の提案。
- 検索対象データを極小部分マッチンググラフとした、Web 文書分類ビュー機構の実装とその評価。
- 検索の高速化を考慮してロボット型の検索エンジンの機能を本システムに付加し極小部分マッチンググラフを扱うことができる Web 文書分類ビュー機構を実装。
- 検索作業に関するフィードバック操作機能の充実。

謝辞 本研究の一部は、日本学術振興会未来開拓学術研究推進事業における研究プロジェクト「マルチメディア・コンテンツの高次処理の研究」および文部省科学研究費重点領域研究(課題番号 08244103)によるものである。ここに記して誠意を表します。

参考文献

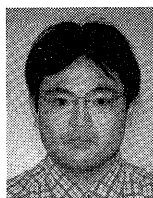
- 1) <http://www.altavista.com/>.
- 2) Botafogo, A. R., Rivlin, E. and Shneiderman, B.: Structural Analysis of Hypertext: Identifying Hierarchies and Useful Metrics, *ACM Transactions on Information Systems*, Vol. 10, No. 2, pp. 142-180 (1992).
- 3) Botafogo, A. R. and Shneiderman, B.: Identifying Aggregates in Hypertext Structures, *Proc. of the 3rd ACM Conference on Hypertext*, ACM, pp. 63-73 (1991).
- 4) Cutting, R. D., Karger, R. D., Pedersen, O. J. and Tukey, V. J.: Scatter/Gather: A cluster-based approach to browsing large document collections, *Proc. of the 15th Annual International ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 318-329 (1992).
- 5) <http://www.excite.com/>.
- 6) <http://www.goo.ne.jp/>.
- 7) Hara, Y., Keller, M. A. and Wiederhold, G.: Implementing Hypertext Database Relationships through Aggregations and Exceptions, *Proc. of the 3rd ACM Conference on Hypertext*, ACM, pp. 75-90 (1991).
- 8) Hatano, K., Qian, Q. and Tanaka, K.: A SOM-Based Information Organizer for Text and Video Data, *Proc. of the 5th International Conference on Database Systems for Advanced Applications (DASFAA'97)*, World Scientific, pp. 205-214 (1997).
- 9) Hearst, A. M. and J., P. O.: Reexamining the Cluster Hypothesis: Scatter/Gather on Retrieval Results, *Proc. of the 19th annual international ACM SIGIR conference on Research and development in information retrieval*, ACM, pp. 76-84 (1996).
- 10) Honkela, T., Kaski, S., Lagus, K. and Kohonen, T.: WEBSOM - Self-Organizing Maps of Document Collections, *Proc. of the Workshop on Self-Organizing Maps (WSOM'97)* (1997).
- 11) Kohonen, T.: The Self-Organizing Map, *Proceedings of the IEEE*, Vol. 78, No. 9, pp. 1464-1480 (1990).
- 12) Kohonen, T.: *Self-Organizing Maps*, Springer, Berlin (1995).
- 13) Kohonen, T., Hynninen, J., Kangas, J. and Laaksonen, J.: SOM.PAK: The self-organizing map program package, Technical Report A31, Helsinki University of Technology, Laboratory of Computer and Information Science (1996).
- 14) Lagus, K., Honkela, T., Kaski, S. and Kohonen, T.: Self-organizing Maps of document col-

- lections: A new approach to interactive exploration, *Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining (KDD-96)*, p. 238-243 (1996).
- 15) Mukherjea, S., Foley, D. J. and Hudson, E. S.: Interactive Clustering for Navigating in Hypermedia Systems, *Proc. of European Conference on Hypertext Technology (ECHT'94)*, ACM, p. p. 136-145 (1994).
- 16) Rocchio, J. J.: *The SMART Retrieval System: Experiments in Automatic Document Processing*, Prentice-Hall, chapter Relevance feedback in information retrieval, pp. 313-323 (1971).
- 17) Salton, G.: *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley (1989).
- 18) Salton, G., Allan, J. and Buckley, C.: Automatic Structuring and Retrieval of Large Text Files, *Communications of the ACM*, Vol. 37, No. 2, pp. 97-108 (1994).
- 19) Tajima, K., Mizuuchi, Y., Kitagawa, M. and Tanaka, K.: Cut as a Querying Unit for WWW, Netnews, and E-mail, *Proc. of the 9th ACM Conference on Hypertext and Hypermedia*, ACM, pp. 235-244 (1998).
- 20) Weiss, R., Vélez, B., Sheldon, A. M., Namprempre, C., Szilagyi, P., Duda, A. and Gifford, K.D.: Hypursuit: A hierarchical network search engine that exploits content - link hypertext clustering, *Proc. of the 7th ACM Conference on Hypertext'96*, ACM, pp. 180-193 (1996).
- 21) <http://www.yahoo.com/>.
- 22) Yuwono, B. and Lee, L.D.: Search and ranking algorithms for locating resources on the World Wide Web, *Proc. of the 12th International Conference on Data Engineering (ICDE'96)*, p. p. 164-171 (1996).
- 23) 江口浩二, 伊藤秀隆, 隈元昭: ユーザへの適応性を考慮した適合フィードバックによる WWW 情報検索, 電気学会論文誌 電子・情報・システム部門誌, Vol. 117-C, No. 11, pp. 1643-1649 (1997).
- 24) 仁木和久, 田中克己: ニューラルネットワーク技術の情報検索への応用, 人工知能学会誌, Vol. 10, No. 1, pp. 1-7 (1995).

(平成 10 年 9 月 20 日受付)

(平成 10 年 12 月 27 日採録)

(担当編集委員 石川 博)



波多野賢治 (学生会員)

1971 年生。1995 年神戸大学工学部計測工学科卒業。1997 年同大学大学院自然科学研究科博士前期課程情報知能工学専攻修了。現在、同大学大学院自然科学研究科博士後期課程情報メディア科学専攻に在学中。情報検索、データベースの研究に従事。



佐野 綾一 (学生会員)

1976 年生。1998 年神戸大学工学部情報知能工学科卒業。現在、同大学大学院自然科学研究科博士前期課程情報知能工学専攻に在学中。情報検索、データベースの研究に従事。



段 一為 (学生会員)

1973 年生。1995 年中国天津大学計算機科学系卒業。現在、神戸大学大学院自然科学研究科博士前期課程情報知能工学専攻に在学中。情報検索、データベースの研究に従事。



田中 克己 (正会員)

1951 年生。1974 年京都大学工学部情報工学科卒業。1976 年同大学大学院修士課程修了。1979 年神戸大学教養部助手。1986 年同大学工学部助教授。1994 年同大学工学部教授 (情報知能工学科)。1995 年同大学大学院自然科学研究科 (知能科学専攻) 専任教授, 現在に至る。工学博士。主にデータベースの研究に従事。現在本会データベースシステム研究会主査。96 年度より通信・放送機構「次世代デジタル映像通信の研究開発」の研究総括責任者, 文部省科研費重点領域研究「分散発展型データベースシステム技術の研究」の研究代表者。神戸マルチメディアインターネット協議会理事。人工知能学会, ACM 等各会員。