

分類視点の学習機構を持つ情報自動分類システム

宮崎 哲夫 田中 栄治 古城 則道

学習情報通信システム研究所

E-mail:{miya,eiji,kojo}@srl.co.jp

あらまし 膨大な電子化された情報の効率的な利用のために、文書の自動分類についての研究が多くなされている。本稿では、意味表現と自己組織化アルゴリズムを用いて、ユーザの閲覧文書を逐次的に分類し、かつ、ユーザの分類例から分類視点を学習する機構も備えた文書の自動分類システムを提案する。本システムは、2ステップの処理機構で構成されている。第1のステップは、分類精度の向上と分類粒度の制御を目的とし、特異値分解による意味空間生成を行なう。第2のステップは、ユーザの多様な分類視点への適応機能を持つ自己組織化クラスタリングを行なう。

キーワード 情報検索, 文書分類, 特異値分解, 分類視点, 意味空間, ニューラルネットワーク

Information Auto-Classifying System with a View Tuning Mechanism

Tetsuo MIYAZAKI, Eiji TANAKA and Norimichi KOJO

Software Research Laboratory

E-mail:{miya,eiji,kojo}@srl.co.jp

Abstract Many methods for document classification have been researched with the aim of efficient extraction of useful information from a huge resource. We propose an auto-classifying system of documents, which can learn user's view point from classification examples. The system is composed of two step processings. The 1st processing makes a semantic space through the singular value decomposition, which contributes to an improvement of classification accuracy and a control of classification grain size. The 2nd processing performs a self-organized clustering, which can adapt to various user's view point.

Keywords Information Retrieval, Document Classification, Singular Value Decomposition, View Point, Semantic Space, Neural Network

1 はじめに

近年のインターネットの爆発的普及とともに、誰もが世界中の情報へ容易にアクセスできるようになってきている。その膨大な情報量は情報の洪水を生み出し、ユーザが真に必要な情報の取得を困難なものにしているのも事実である。ユーザにとって価値のある情報の発見には、情報の体系的な整理が必要である。体系的に整理された情報は、知識獲得を容易にするばかりではなく、新たな発想のきっかけを与えてくれる。そのため、文書分類についての

研究は情報検索や情報フィルタリングの分野で数多く成されている [15, 16, 17].

文書分類の手法の中に、文書の意味内容を出現単語の頻度情報に基づいて表現するベクトル空間モデルがある [18]. また、ベクトル空間法の発展に意味空間法がある。関連する単語の一次結合により、意味内容に寄与する直交基底群を抽出する手法である [5, 7, 8, 9, 10, 11, 12]. 寄与率の小さい基底をノイズ成分として除くことにより、あいまいさや冗長性の問題を解決し表現の精度向上を図ることを目的としている。

ベクトル表現された文書に対し、近年良く使われる分類手法として、自己組織化ニューラルネットワークを用いたクラスタリングがある [13, 14, 19, 20]. 自己組織化クラスタリングは、教師なし学習のアルゴリズムであり、高速ではあるが修正学習のメカニズムを持たないため、分類の精度を制御できない [1, 2]. このため、多層ニューラルネットワーク構造を持たせて、誤差逆伝搬学習と組合せる試みも提案されている.

ところで、文書の分類は、何らかの分類の視点を定めて、始めて実現されるものである [3, 4]. 分類の視点は人に依っても、利用する場面に依っても異なる. これまで研究された文書分類の多くは、ユーザによる分類視点の違いが考慮されていない.

本稿では、意味空間と自己組織化クラスタリングを用いて、分類視点を分類例から学習できる文書の自動分類システムを提案する. このため、文書の単語表現と自己組織化プロセスの間に意味空間に基づく中間表現を導入し、ユーザの多様な分類視点に適応できる柔軟な機構を実現する.

2 文書自動分類システムの構成

文書の自動分類におけるポイントは2つある. 1つは文書をいかに表現するかであり、2つめは類似度とクラスタリングアルゴリズムの定義である.

文書表現は文書の特徴ベクトルの作成であり、表現の基底となる単語の選定と抽出法、および、標準単語を基底とする各文書の数量化が問題となる. 本稿では、標準単語は予め与えられているものとし、特異値分解 [5, 6, 7, 12] による意味空間を導入するとにより、表現の精度向上と意味的制御可能な表現基底の検討を行なう.

クラスタリングアルゴリズムには逐次的な自己組織化手法 [13] を採用する. ただし、加えて、意味空間を積極的に利用し、ユーザとシステムの対話による分類視点の調整機能を持たせる.

提案するシステムの構成を図 1 に示す.

システムは意味空間生成エンジンと視点調整の付加機能を持つ分類エンジンで構成される. 意味空間生成エンジンは、分類対象のドメインの典型的文書集合をサンプルに使用して特異値分解による意味空間及び文書の意味空間への変換行列の作成処理を実行する. 分類エンジンは、新規文書から

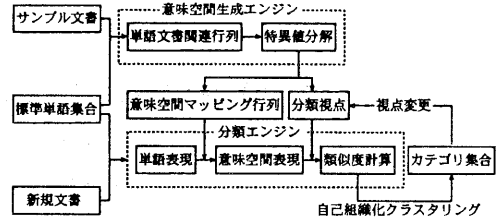


図 1: システム構成

語表現の生成、意味空間表現へのマッピング、そして自己組織化クラスタリングによるカテゴリ決定処理を行う. 分類視点調整処理は、ユーザが分類エンジンが作成した分類に対して修正を施したときに起動し、ユーザの与えた分類を実現するように分類視点の適応学習を行う.

2.1 意味空間と意味空間表現

文書のベクトル表現の作成のために、単語や意味を一段抽象化した意味空間を導入する [8, 9, 10, 11, 21].

対象領域の文書からサンプル文書集合を抽出し、標準単語集合を使って各サンプル文書にキーワードを割り当て、単語-文書関連行列 R を作成する.

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1w} \\ r_{21} & r_{22} & \cdots & r_{2w} \\ \vdots & \vdots & \ddots & \vdots \\ r_{d1} & r_{d2} & \cdots & r_{dw} \end{bmatrix}$$

ここで、 r_{ki} は文書 k に単語 i が出現する回数を表す. すなわち、行列 R の行成分が文書の単語表現、列成分が単語の文書表現になっている.

意味空間の作成には、特異値分解が使われる [5, 6]. ここで、単語-文書関連行列 R から作られる

$$\text{単語の連想行列} = R^t R$$

$$\text{文書の連想行列} = R R^t$$

を考える. 添字の t は転置行列を表す. 特異値分解によると、文書-単語関連行列 R は、連想行列の固有値問題の解を使って

$$R = D^t \Lambda^{\frac{1}{2}} W$$

のように展開される。行列 Λ は、連想行列 $R^t R$ （あるいは RR^t ）の非零固有値が作る対角行列

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix}$$

であり、 r が固有値の個数である。 D と W は、それぞれ、 RR^t と $R^t R$ の固有ベクトルから作られる $r \times d$ と $r \times w$ の行列である。

$$D = \begin{matrix} D_1 \cdots D_d \\ E_1 \\ \vdots \\ E_r \end{matrix} \begin{bmatrix} r \times d \end{bmatrix} \quad W = \begin{matrix} W_1 \cdots W_d \\ E_1 \\ \vdots \\ E_r \end{matrix} \begin{bmatrix} r \times w \end{bmatrix}$$

また、 D と W を構成する行ベクトルは、同一の固有ベクトルの文書空間表現及び単語空間表現である。もちろん、 D と W の行ベクトル集合はそれぞれ正規直交系をなす。

$$DD^t = WW^t = I_r$$

ここで、 I_r は $r \times r$ の単位行列である。さらに、 D に正規直交ベクトル $\{\bar{d}_{r+1}, \dots, \bar{d}_d\}$ 、 W に正規直交ベクトル $\{\bar{w}_{r+1}, \dots, \bar{w}_w\}$ を付け加えると

$$D^t D = I_d, \quad W^t W = I_w$$

という正規直交関係が得られる。

特異値分解で得られる量は、固有ベクトルと固有値の二つである。固有ベクトルは単語（文書）が作る空間内の単語（文書）の連想行列に対する不変集合に相当する正規直交基底である。また、固有値はサンプル文書集合の内容表現に対する各基底の寄与率を表す。したがって、この値は文書や単語の表現における各基底の重みに相当し表現の視点を意味する。本システムでは、文書間の類似度の算定にこの概念を積極的に活用し、分類視点の制御に利用する。

特異値分解による意味空間の導入の利点は、近似解として表現空間のシステムティックな縮約法を提供してくれることにある。すなわち、単語と文書の表現空間を大きい固有値に対応する固有ベクトルが張る空間に絞ることによって、もとの文書 D_1, \dots, D_d や単語 W_1, \dots, W_w は縮約された空間のベクトルで表現される。この処方では、実データに多分に含まれるノイズやあいまいさを取り除きデー

タに含まれる本質的な因子を段階的に抽出する手続きを与える。さらに、特異値分解が与える解は、単語と文書のコンシステントな表現になっている。

新規文書に対する意味空間表現は、先ず単語の出現頻度から文書の単語スペクトルを作り、次に先の特異値分解で作った単語の意味空間表現を代入すると得られる。すなわち、文書の意味空間表現は、単語の意味空間表現を単語の出現回数で重み付けた和により求められる。

具体的には、単語の意味ベクトルと文書の意味ベクトルの関係

$$\Lambda^{\frac{1}{2}} D = W R^t$$

の列ベクトル成分から

$$\Lambda^{\frac{1}{2}} (\bar{s}_i)^t = W (\bar{d}_i)^t$$

を得る。ここで、 \bar{s}_i は新規文書の意味空間表現、 \bar{d}_i は新規文書の w 次元単語スペクトル（単語-文書関連行列 R の行ベクトルと同じく単語空間内のベクトル）である。従って、新規文書の意味空間表現は

$$(\bar{s}_i)^t = \Lambda^{-\frac{1}{2}} W (\bar{d}_i)^t$$

のように得られる。この式から、行列 $\Lambda^{-\frac{1}{2}} W$ はもとの単語空間から意味空間への変換行列の役割を果たしていることが分かる。

2.2 自己組織化ニューラルネットワークによる文書分類

ベクトル表現された文書は、適切な類似度の定義をすると、類似性をもとに体系的に分類することができる。本システムでは、分類アルゴリズムにあらかじめ分類カテゴリを持たないクラスタリング手法を用いる。また、ユーザによるカテゴリの生成や変更が随時可能な対話的な機構にするため、逐次的な構築アルゴリズムでありしかもクラスタ数が可変であるSCONN[13]のアルゴリズムを採用する。SCONNは、コホーネンの自己組織化ニューラルネットワーク[14]にツリー状のカテゴリ生成メカニズムを付加したものである。

文書の入力から、所属カテゴリの決定までの流れを図2に示す。

今、カテゴリが複数個あったとする。各カテゴリ X の代表ベクトル \bar{c}_x とカテゴリ内分散ベクトル

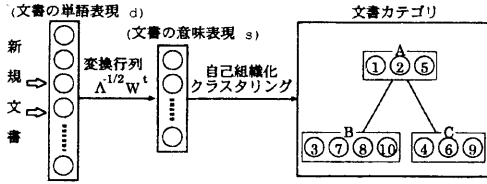


図 2: カテゴリ決定までの流れ

$\bar{\sigma}_x$ は

$$\begin{aligned} \bar{c}_x &= (c_{x_1}, c_{x_2}, \dots, c_{x_m}) \\ \bar{\sigma}_x &= (\sigma_{x_1}^2, \sigma_{x_2}^2, \dots, \sigma_{x_m}^2) \end{aligned}$$

で表される。ここで、各成分値はこのカテゴリに属する文書を使って、

$$\begin{aligned} c_{x_j} &= \frac{1}{n_x} \sum_{k=1}^{n_x} s_{k_j} \\ \sigma_{x_j}^2 &= \frac{1}{n_x} \sum_{k=1}^{n_x} (s_{k_j} - c_{x_j})^2 \end{aligned}$$

であり、 n_x は所属する文書数である。

新しい文書 i が入力されると、まず単語空間表現 \bar{d}_i が作られ、続いて意味空間表現 \bar{s}_i に変換される。この文書の所属は、重み付き標準ユークリッド距離

$$d = \sum_{j=1}^m \frac{\widetilde{w}_j^2 (s_{ij} - c_{x_j})^2}{\sigma_{x_j}^2}$$

を用いた類似度 [1, 2]

$$\text{Sim} = e^{-\frac{\gamma d}{2}}$$

が最大となるカテゴリに決定される。ただし、

$$\bar{X} = \frac{X}{\sum_{j=1}^m w_j^2}$$

である。重み係数 w_j^2 は文書分類における視点を表す。また、 γ は任意の正定数である。

最大類似度が或るしきい値 ϵ 以上ならば、この文書はそのカテゴリに属すると決定され、カテゴリの代表ベクトルとメンバー数を

$$\begin{aligned} n'_x &= n_x + 1 \\ \bar{c}'_x &= \frac{n_x \bar{c}_x + \bar{s}_i}{n_x + 1} \end{aligned}$$

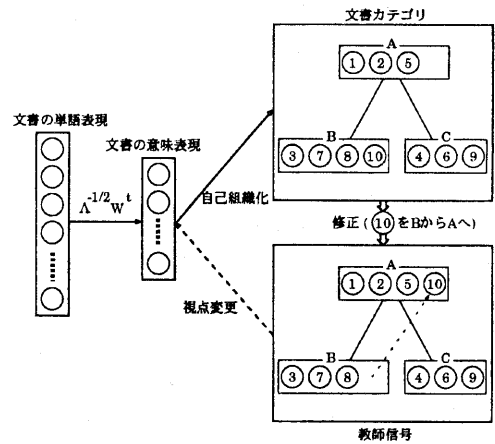


図 3: 分類視点の教示による視点の変更

のように更新する。他方、最大類似度が ϵ 未満ならば、そのカテゴリの子として、新しいカテゴリ

$$\begin{aligned} \bar{c}_{new} &= \bar{s}_i \\ n_{new} &= 1 \end{aligned}$$

を生成する。

3 分類例の教示による視点の変更

分類の視点は、人によって異なる。また、同じ人でも使う場面によっても異なる。従って、分類エンジンに必要な機能は分類視点を変更できることである。文書の自動分類システムにより作成された分類に対して、ユーザが同意せず図 3 のように修正を施した場合を考える。

この例のように、システムの作った分類に対して、ユーザが修正を加えたものはユーザによるシステムへの分類例の教示と考えることができる。

修正後の分類における各カテゴリの代表ベクトルは

$$\begin{aligned} \bar{c}'_A &= \frac{\bar{s}_1 + \bar{s}_2 + \bar{s}_5 + \bar{s}_{10}}{4} \\ \bar{c}'_B &= \frac{\bar{s}_3 + \bar{s}_7 + \bar{s}_8}{3} \\ \bar{c}'_C &= \frac{\bar{s}_4 + \bar{s}_6 + \bar{s}_9}{3} \end{aligned}$$

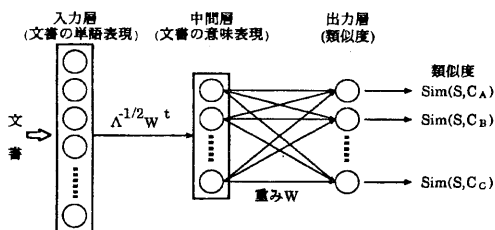


図 4: 3 層ニューラルネットワークの模式図

である。文書 1 から文書 10 の各文書のベクトルとこれらのカテゴリベクトルとの類似度に基づく仕分けが、ユーザの指定した先の分類結果を実現しなければならない。

ここで、われわれのシステムのカテゴリ機構を図 4 のような 3 階層のニューラルネットワークとして捉える。入力層は、単語表現された文書が入力し出力される。各ノードは単語に対応する。中間層は、単語表現を変換した意味表現が入力し出力される。各ノードは意味表現の基底に対応する。出力層のノードは、分類先のカテゴリに対応し、ノード情報として中間層のノード数と同じ次元の代表ベクトル \vec{c}_x を持つ。出力層の各ノードからは、自分の持つカテゴリ代表ベクトルと中間層から出力された意味表現との類似度の値が出力され、自己組織化の処理によりカテゴリの選定あるいは生成がなされる。

ユーザの分類にシステムのカテゴリを適応させるには、本システムでは 3 つのパラメタの調整が考えられる。それらは、類似度計算式に現れる重み、中間層のノード数、入力層と中間層の間の結合係数である。本稿では、視点の違いから生じる分類の違いに着目して、前二者の学習アルゴリズムを検討する。すなわち、文書分類における視点を表す重み係数 w_i の修正、意味ノードの追加である。

3.1 分類視点の逐次学習

カテゴリへの分類は入力文書とカテゴリの類似度をもとになされるのであるから、カテゴリノードの出力を類似度ベクトル

$$\vec{z}_i = (z_{iA}, z_{iB}, \dots, z_{iZ})$$

$$z_{ix} = \text{Sim}(\vec{s}_i, \vec{c}_x)$$

として、ニューラルネットワーク的なアプローチにより、分類視点を分類例から導くことが考えられる。ここで、各カテゴリの代表ベクトルは、ユーザによる修正後の分類にもとづくものとする。

カテゴリノードの出力を類似度ベクトルと考えると、教師データもまた、類似度ベクトル

$$\vec{z}_i^T = (z_{iA}^T, z_{iB}^T, \dots, z_{iZ}^T)$$

の形で与えなければならない。例えば、文書 \vec{s}_{10} はカテゴリ A に属するという教示の場合は

$$\vec{z}_{10}^T = (1, 0, 0, \dots, 0)$$

が教師データになる。

ここで、文書 i に対する評価関数を

$$V = \sum_{x=A}^Z \frac{1}{2} (z_{ix}^T - z_{ix})^2$$

で定義すると、この値を最小にするようなパラメタ $\{w_j\}$ を見つけることが学習となる。

修正学習に対して最急降下法を用いると、パラメタ w_k の修正量は

$$\delta w_k = -\alpha \frac{dV}{dw_k}$$

である。 α は正の微小数である。ここで、先に定義した重み付き標準ユークリッド距離に基づく類似度を用いると

$$\delta w_k = -\alpha \gamma \widetilde{w}_k \sum_{x=A}^Z (z_{ix}^T - z_{ix}) z_{ix} \left(\frac{(s_{ik} - c_{kx})^2}{\sigma_{kx}^2} - \sum_{j=1}^m \widetilde{w}_j^2 \frac{(s_{ij} - C_{xj})^2}{\sigma_{xj}^2} \right)$$

と導かれる。

すべての文書について一括して修正する式は、

$$\delta w_k = -\alpha \gamma \widetilde{w}_k \sum_{i=1}^N \sum_{x=A}^Z (z_{ix}^T - z_{ix}) z_{ix} \left(\frac{(s_{ik} - c_{kx})^2}{\sigma_{kx}^2} - \sum_{j=1}^m \widetilde{w}_j^2 \frac{(s_{ij} - C_{xj})^2}{\sigma_{xj}^2} \right)$$

となる。 N は分類対象の文書数である。

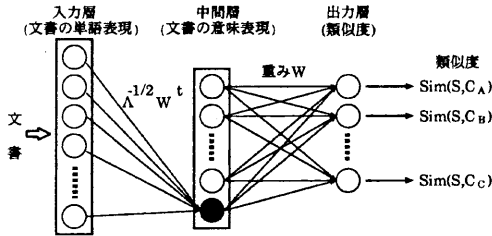


図 5: 中間層のノードの追加による修正

3.2 意味空間の次元数の変更による調整

意味空間の次元数は、文書表現の記述の詳細度に関係している。従って、ユーザの細かい要求に対応する上で、意味空間の次元すなわち中間層のノードを増やすのが有効である。次元が増えると、すべての文書の意味空間表現、従って、カテゴリ代表ベクトルも変更を受ける。もともとある成分についてはベクトルの正規化の因子による変更を受ける。文書間の類似度も変わり、全文書の再分類が必要となる。中間層のノードを $(m+1)$ 個に増やしたときの、修正学習の手順は以下のようなものである。中間層のノードは、特異値分解で得られた r 個の固有ベクトルの中の、まだ意味表現に使用されていない $(r-m)$ 個の中の一つを選ぶ。選ばれた固有ベクトルの単語表現と固有値を、入力層と中間層を結ぶ変換行列に追加する。

$$\Lambda^{\frac{1}{2}} W^t \rightarrow \Lambda_{m+1}^{\frac{1}{2}} W_{m+1}^t$$

各文書の単語表現から、上記の変換行列を使って、意味表現を作りなおす。意味空間の次元を増やすこの処理は、もとの部分については何の変更もなく、文書の意味表現に $(r+1)$ 次元目が追加される。

ユーザの指定した分類例に従い、各カテゴリの代表ベクトルを所属している文書の意味表現から計算する。

各文書の意味表現と各カテゴリの代表ベクトルをもとに、 $(m+1)$ 個の視点パラメタ $\{w_j\}$ を先に述べた手法により決定する。

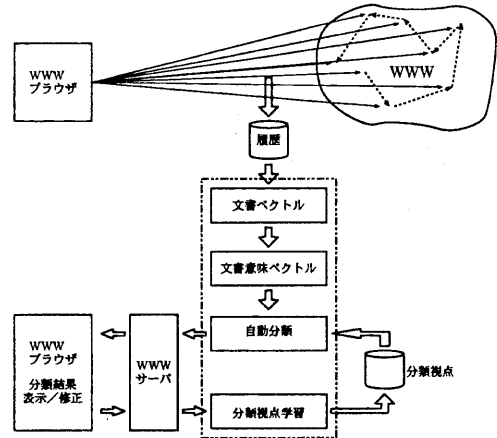


図 6: プロトタイプシステム

4 プロトタイプシステム

本研究で構築したプロトタイプシステムの構成を図 6 に示す。本プロトタイプシステムは、ユーザが WWW 上に存在するページを閲覧した履歴をもとに、それらのページを分類する。その後、ユーザがシステムが行なった分類を修正し、その修正結果をシステムに渡すことにより、システムはユーザの分類視点を学習おこなう。以降の分類に対し、学習した分類視点を考慮して分類する。ブラウザには、Netscape を想定し、分類結果の表示および修正には Java アプレットを使用して、GUI インターフェースを提供している。

処理の流れは、

- (1) ユーザが WWW ページを閲覧した履歴をもとに、標準単語の出現状況で各ページの文書ベクトルを作成する。
- (2) 文書ベクトルを意味空間へマッピングし、文書意味ベクトルを作成する。
- (3) 文書意味ベクトルを入力とした自己組織化ニューラルネットワークにより分類を行なう。
- (4) 分類結果を WWW ブラウザに引渡し、ユーザに分類結果を提示する。ユーザが本システムの分類結果に同意しない場合は、分類結果を修正することができる。

- (5) ユーザが分類結果を修正した場合には、システムの持っている分類視点とユーザの持っている分類視点が異なっていると考えられるので、分類視点を表している重み付けデータをユーザの意図に合うように修正する。
- (6) 次回の分類には、新しい分類視点(重み付けデータ)を用い、分類を行なう。

のようになる。

5 まとめ

本稿では、意味空間と自己組織化クラスタリングを用いて、ユーザの分類視点の学習メカニズムを備える文書の自動分類システムについて報告した。

システムの構成は、文書表現の抽象化としての特異値分解による意味空間生成機構、および、ユーザの分類視点への適応機能を持つ自己組織化クラスタリング機構の2ステップの処理機構からなる。

意味空間の導入は、文書表現空間のシステムティックな縮約法を与えるとともに、空間次元数の意味的な制御を可能にしている。自己組織化クラスタリングでは、分類例の教示と学習という形のシステムとユーザによる対話的な分類システム構築環境を実現している。

今後の課題として、文書からの特徴単語の自動抽出法の検討、単語の追加に対応できる機構の検討が挙げられる。単語の追加は、単語-文書関連行列の作り直しを必要とし、意味表現を作るための対角化の再計算を必要とする。また、ユーザの指定した分類が視点の調整のみでは実現しないことも考えられる。このような状況への対応として、意味表現の変更、さかのぼると単語表現から意味表現への変換行列の調整により実現できる。この機構は現在検討中である [22]。

参考文献

- [1] 宮永喜一, 枋内香次, 自己組織化と教師によるネットワークの拘束・高精度学習について, 信学論A, Vol.J78-A, No. 11, pp.1475-1484, November, 1995.
- [2] 大貫和永, 宮永喜一, 枋内香次, 教師学習を考慮した自己組織化ネットワークに関する一考
- [3] 沢田裕司, 大川剛直, 馬場口登, 観点を考慮した連想機構の実現, 情報処理学会論文誌, Vol. 35, No. 5, pp.714-724, May, 1994.
- [4] 笠原要, 松澤和光, 石川勉, 河岡司, 観点に基づく概念間の類似性判別, 情報処理学会論文誌, Vol. 35, No. 3, pp.505-509, Mar, 1994.
- [5] Scott Deerwester, et. al., Indexing by Latent Semantic Analysis, J. of American Society for Information Science, 41(6), 391-407(1990).
- [6] 柳井, 竹内著, UP 応用数学選書:射影行列・一般行列・特異値分解, 東京大学出版会, 1983年.
- [7] Peter W. Foltz, Susan T. Dumais, Personalized Information Delivery Analysis of Information Filtering Methods, COMMUNICATION OF ACM, Vol.3, No.12, pp.51-60, December, 1992.
- [8] 小嶋秀樹, 伊藤昭, 意味空間のスケール変換による動的シソーラスの実現, 情報処理学会自然言語処理研究会研究報告, NL108-13, pp.81-88, July, 1995.
- [9] 吉岡康裕, 本間直人, 石川真澄, 情報検索におけるキーワードの空間表現, 信学技報, NC95-144, pp.219-226, March, 1996.
- [10] 本間直人, 石川真澄, 情報検索におけるキーワードと文献の空間表現, 信学技報, NC95-143, pp.211-218, March, 1996.
- [11] 渡辺正裕, 石川佳治, 吉川正俊, 植村俊亮, ベクトル空間の視覚的探索機能を有する情報検索, 情報処理学会データベースシステム研究会研究報告, DBS109-2, pp.7-12, 1996.
- [12] 宮原隆行, 清水康, 北川高嗣, 意味の数学モデルによる意味的連想検索の高速化アルゴリズム, 情報処理学会データベースシステム研究会研究報告, DBS109-1, pp.1-6, 1996.
- [13] Doo-Il Choi, Sang-Hui Park, Self-Creating and Organizing Neural Networks, IEEE

Transactions on Neural Networks, Vol.5,
No.4, pp.561-575, July, 1994.

- [14] Teuvo Kohonen, Self-Organizing Map, Proceedings of IEEE, Vol.78, No.9, pp.1464-1480, September, 1990.
- [15] 森田昌宏, 速水治夫, 情報フィルタリングシステム-情報洪水への処方箋, 情報処理学会誌, Vol.37, No.8, pp.751-758, August, 1996.
- [16] 住田一男, 三池誠司, 知的情報検索の動向, 人工知能学会誌, Vol.11, No.1, pp.10-16, January, 1996.
- [17] 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適用, 人工知能学会誌, Vol.10, No.1, pp.45-51, January, 1995.
- [18] 伊藤哲郎, 情報検索 (ソフトウェア講座 19), 昭晃堂, 1986.
- [19] 田中 栄治, 宮崎 哲夫, 古城 則道, 文書の組織化手法を用いた学習者モデル構築への一考案, 1996年電子情報通信学会情報・システムソサイエティ大会 D-480 (1996).
- [20] 田中栄治, 宮崎哲夫, 古城則道, 情報検索における文書集合の組織化について, 情報処理学会第53回全国大会 講演論文集, 1T-7 (1996).
- [21] 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, 情報処理学会第53回全国大会 講演論文集, 1T-8 (1996).
- [22] 田中栄治, 宮崎哲夫, 古城則道, 情報探索支援システムの構築 (1), 信学技報, ET96-95, pp.55-62, Dec, 1996.