

自己組織化手法を用いた情報の組織化とユーザ適応機構<sup>†</sup>

1 S-4

○ 田中 栄治 宮崎 哲夫 古城 則道<sup>‡</sup>  
学習情報通信システム研究所<sup>§</sup>

## はじめに

近年のインターネットの爆発的普及は、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。このような膨大な情報源からの知識獲得には、情報の整理 / 体系化が必要であり、そこで文書の自動分類が役立つ。

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、様々な手法が提案されてきているが、本稿では知識記述としては浅いが汎用的な手法であるキーワードの出現情報を利用した文書のベクトル表現に基づく手法を採る。

本研究では、3層からなるニューラルネットワーク構造の文書分類システムを構築し、自己組織化クラスタリングで用いる文書の中間表現自体を修正することにより、ユーザの分類に適応させる学習メカニズムを提案する。

## システム概要

システムの構成を Fig. 1 に示す。

システムはユーザの教示に適応するネットワーク学習機構の付加機能を持つ分類エンジンで構成される。分類エンジンは、新規文書からキーワード（標準単語）の有無をもとに単語表現を生成、中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。カテゴリ決定は、逐次的な構築アルゴリズムであり、しかもカテゴリ数が可変の自己組織化ニューラルネットワークを用いる [1]。これにより、ユーザによるカテゴリの生成や変更が随時可能である対話的な機構を実現できる。

修正学習機能は、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中間表現生成の適応学習を行う。本システムの中間表現は明確な意味を持たないが、その学習機能はユーザ自身の集めた文書と分類教示を通して対話により自分用の意味空間を作り上げていく事が可能としている。

<sup>†</sup>Self-Organized Information Classification with User-Adaptive Mechanism

<sup>‡</sup>Eiji TANAKA, Tetsuo MIYAZAKI, Norimichi KOJO

<sup>§</sup>Software Research Laboratory

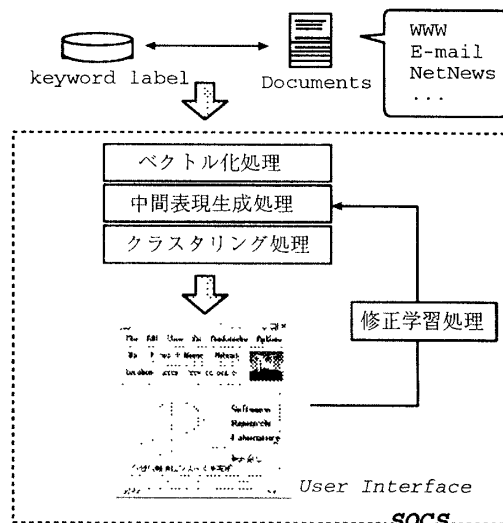


Fig. 1 システム構成

本システムは、基本的には文書のベクトル表現があれば、自己組織化クラスタリングにより、実行される。中間表現の役割は、ユーザの要求に適応できる柔軟性をシステムに提供するとともに、次元を縮約することにより計算時間の短縮を図ることにある。

## 分類例の教示による修正学習

ここで、システムの与えた分類に対して、ユーザが Fig. 2 のように変更を施した場合を考える。

ユーザの変更は、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。システムの分類機能をユーザの変更に適応させる機構を考えるため、我々のシステムを3階層のニューラルネットワークとして捉える。

このシステムで、システムの与えた分類に対するユーザの修正に適応させる機構として、類似度計算における重みの修正、中間層のノード数の変更、そして、中間表現の修正が考えられるが、本稿では、中間表現の修正による学習を検討する。中間表現の修正は、単語表現から中間表現へ変換する変換行列の修正を促す。この変換行列の修正学習は、誤差逆伝播の手法により定式化できる。

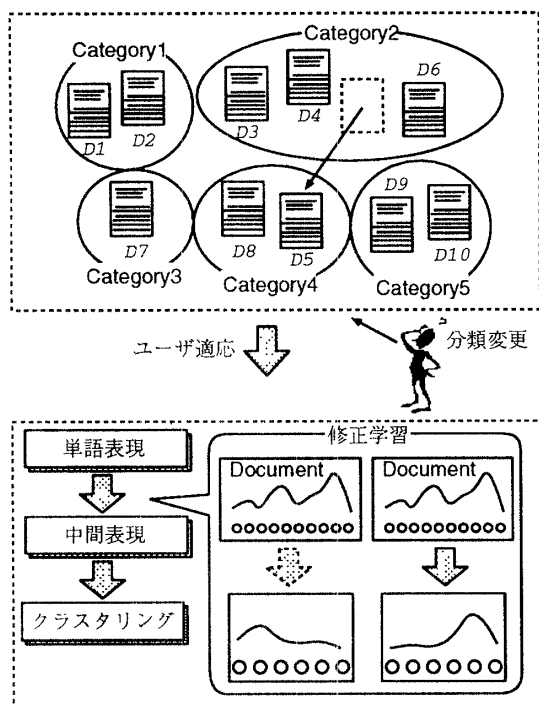


Fig. 2 分類変更とユーザ適応

プロトタイプ構築

Web 上に構築したプロトタイプシステム (SOCS<sup>1</sup>) の概要を Fig. 3 に示す。このプロトタイプシステムでは、ユーザが Web 上で行う Net-Surfing から得た履歴情報を自動的に分類し、情報の組織化を行う機構を実現している。また、サーチエンジン (Web 上の検索システム) とのインターフェース機能も付加されており、サーチエンジンによる検索結果を自動的に収集 / 分類する機能も有する。

このプロトタイプシステムを利用する事により、情報の組織化が自動的に行われ、ユーザが探索 / 検索した Web 上の情報をカテゴリ別に分類する事が可能となる。

検討および考察

本稿では、ユーザの分類変更に対して分類アルゴリズムを適応させる文書自動分類システムの提案を行った。ユーザからの変更依頼に対し、本システムでは文書の中間表現の修正、すなわち、単語表現から中間表現を生成する変換行列の修正で対応する手法を使った。

プロトタイプシステムを用いた分類実験では、ユーザの自由探索 (Net-Surfing) からの情報を用いた Web ドキュメントの組織化を行う事ができた。また、分類

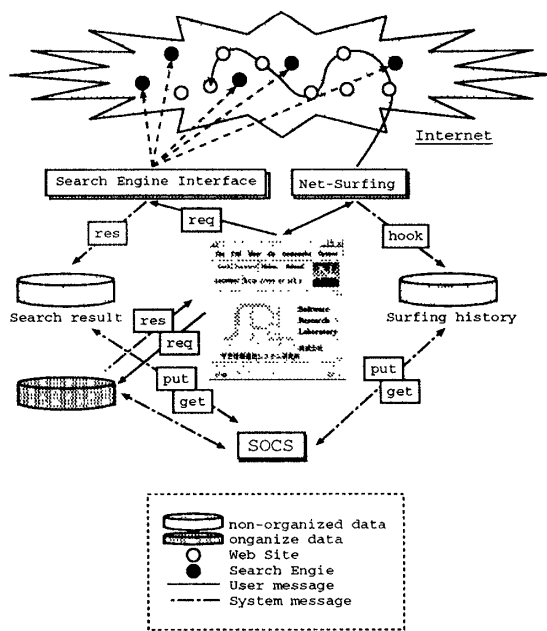


Fig. 3 プロトタイプ (SOCS) の概要

変更機構も正常に動作し、再分類時における変換行列の修正機構が正しく機能している事が確認できた。

今後の課題として、まず単語の自動抽出機構やキーワードラベルの自動生成 / 修正機構の追加が挙げられる。本稿の手法では、キーワードラベルへの依存度が大きく、この機構により、更に精度の高い分類結果を期待する事ができる。

参考文献

- [1] 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).
- [2] 宮崎哲夫, 田中栄治, 古城則道 : 分類視点の学習機能を持つ情報自動分類システム, 情処研報, NL97 (1997).
- [3] 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 (1996).
- [4] 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察, 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).
- [5] 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).
- [6] 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).

<sup>1</sup> Self-Organizing Classification System