

WWW における能動型学習サーチエンジンの構築*

5 R-1

鈴木 信裕†

能登 正人†

神奈川大学工学部‡

1 はじめに

現在、世界中でインターネットが使われ、その中でも特に WWW(World Wide Web) は最も使われているサービスである。インターネット上には非常に多くのホームページが存在し、自分の目的に合ったページを探すのは困難である。ユーザの要求に応じ、本当に欲しい情報を得るため、膨大なデータを整理し、情報検索を効率良く行う必要がある。現在では種々のサーチエンジンが提案されており、さまざまな手法により情報検索を支援している。

本稿では、ロボット型サーチエンジンでユーザが望むような情報を容易に検索でき、無駄な検索を省けるような質問型能動学習付きサーチエンジンの構築手法を提案する。

2 現在のサーチエンジン

現在、インターネット上の情報検索のために、主に2種類のサーチエンジンが使われている。ディレクトリ型とロボット型で、それぞれ以下のような特徴がある。

● ディレクトリ型

サーチエンジンの管理者が情報を管理するため、情報はあまり多くない。しかし、カテゴリに分かれているため、ある程度検索が容易である。

● ロボット型

WWW ロボットと呼ばれるプログラムが自動的にインターネットから情報ページを取得するため、情報量は非常に多い。しかし、本当に欲しい情報を探すのが困難である。

現状のロボット型サーチエンジンの検索方法では、ユーザがサーチエンジンにキーワードを与えて、サーチエンジンは与えられたキーワードにマッチするホームペー

ジの URL を出力する(図1の点線内部)。しかし、ユーザや時と場合によってキーワードの意味は異なるのが普通であるのに、サーチエンジンはキーワードが合っているでも違う意味のページまで探してしまうという欠点がある[2]。

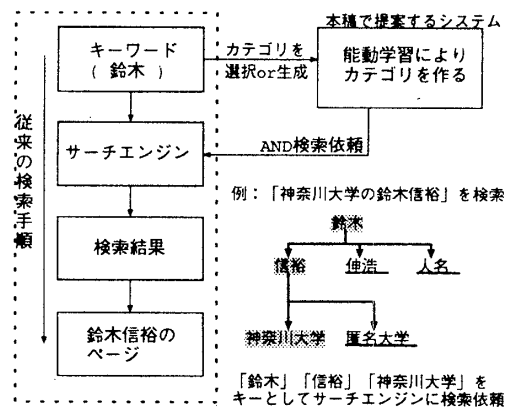


図1: 学習の流れ

3 システムの設計

本稿で提案するシステムでは以下のことを目的とする。

- ロボット型サーチエンジンにカテゴリを作る。
- 与えられたキーワード一つ一つにカテゴリを作る。
- ユーザがカテゴリを作る。
- 質問による能動学習を使ってカテゴリの質を確保する。

3.1 能動学習

能動学習(Active Learning)とは、学習者が自発的に学習を進めていく学習形態を言う。代表的なものに「質問による学習」と「強化学習」がある。前者は、学習者が今までの学習結果から質問点を選んで質問することに

*Design of Search Engine on WWW by Active Learning

†Nobuhiro Suzuki and Masato Noto

‡Faculty of Engineering, Kanagawa University

より効率的に学習を行う。後者は、環境内を動き回り、行動規則を獲得していくロボットなどのモデル化であり、自らとる行動に対し環境から与えられる報酬をもとに行動規則を獲得する [1]。本稿の目的では前者を使った方が最適であると考え、質問による学習を使う。

3.2 検索手順

従来の検索手順 (図 1 の点線内部) では、与えられたキーワードを含むページが内容に関係なく表示されてしまう。一方、本稿で提案するシステムでは、キーワードが与えられた時にそのキーワードに対するカテゴリを返し、ユーザはそれを選んで選択範囲を絞ることで検索の手間を大幅に軽減することができる。また、カテゴリを見ることで与えたキーワードでどのような情報が得られるのかが推測しやすい。

カテゴリを作るのはユーザ自身であり、最後にキーワードと選択したカテゴリをサーチエンジンに渡しホームページを検索 (AND 検索) させる (図 1)。しかし、カテゴリが増えすぎたり不必要なカテゴリが出来る可能性があるため、質問による能動学習を行いカテゴリデータを整理する。

3.3 質問による能動学習の流れ

キーワードが与えられた時、サーチエンジンが提示するカテゴリの中にユーザに有用なカテゴリがあるかどうか聞き、なければ新しくカテゴリを作るようにし、あればそのカテゴリを辿る。

任意なカテゴリの必要性を判断するために適合度を導入し、カテゴリに順位を付ける。適合度は、カテゴリの順位付けのために使われ、与えられたキーワードのカテゴリがうまくサーチエンジンで AND 検索できるか、ユーザが望むようなものかの目安となる。適合度の求め方としては、カテゴリを一度サーチエンジンに渡して返ってきた URL のヒット数とクライアントの利用率を基準に求める。当然、URL のヒット数が「0」の時は適合度が 0 となるため、カテゴリとして登録しない。

カテゴリを作る際の学習手順の概略を図 2 と以下に示す [3]。

1. 学習したカテゴリを (適合度が高い順に上位のみ) 表示してユーザにカテゴリを選んでもらう。
2. ユーザからの入力により適合度を計算しなおす。
3. 適合度からカテゴリの順位を更新する。

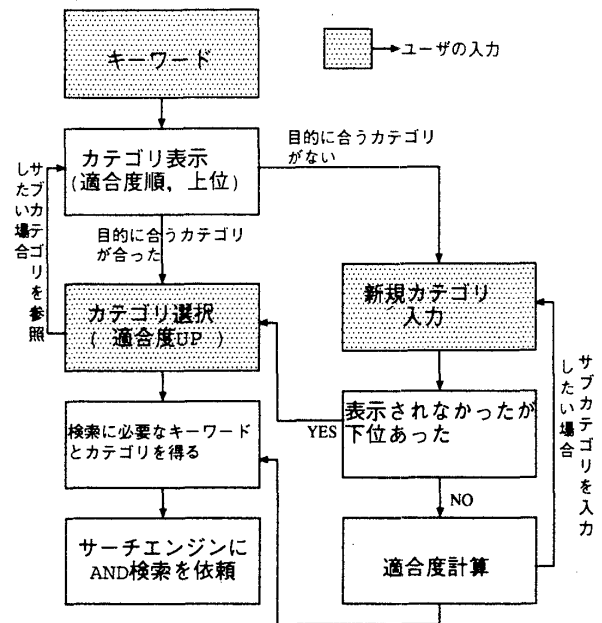


図 2: 学習の流れ

4 おわりに

本稿では、現在のロボット型サーチエンジンで検索効率をあげるために動的にカテゴリを設定し、質問による能動学習システムを導入することを提案した。今後は、プロトタイプシステムを作成し、検索効率について考察する予定である。もし、このサーチエンジンがうまく動作すればインターネット上での情報検索の負担がある程度軽減されると思われる。

参考文献

- [1] 安倍直樹, 中村篤祥: “能動学習に関する計算論的学習理論の研究”, 情報処理学会誌, Vol.37, No.7, pp.575-581(1997).
- [2] 仲川こころ, 高田喜郎, 関浩之: “可変なカテゴリ構造を用いた WWW 検索支援方法”, 情報処理学会第 57 回全国大会, 3 分冊, pp.153-154(1998).
- [3] Gerhard PAASS(訳:麻生英樹): “予測とモデル選択のための質問選択”, 情報処理学会誌, Vol.37, No.7, pp.562-568(1997).