

質問型能動学習による インターネット検索エンジンに関する研究

能登研究室

鈴木信裕 (36040)

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。本研究では、ロボット型検索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による能動学習を使った検索エンジンの構築手法を提案する。

2 ロボット型検索エンジン

現状のロボット型検索エンジンの検索方法では、ユーザが検索エンジンにキーワードを与えて、検索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、検索エンジンはキーワードが合っているにもかかわらず違う意味のページまで探してしまうという欠点がある。

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、カテゴリがあればどんな情報が得られるのかが推測しやすい。

検索エンジンには、キーワードと選択したカテゴリを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、カテゴリに優先順位を付けて整理する必要がある。

また、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、カテゴリを一度検索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。

3.3 質問による能動学習の流れ

本研究では、カテゴリに順位を付けるために、質問による能動学習を使う。この学習方法では、学習を繰り返すことにより、与えられたキーワードに対するカテゴリが最もよく使われる順に並び、検索しやすく、情報を手に入れやすい。カテゴリ作成段階では、システムが適合度を使ってカテゴリに順位を付け、ユーザに出力(質問)する。ユーザはその中から一つ選ぶことでそのカテゴリの適合度を更新し、カテゴリの順位を学習しなおす。

カテゴリ作成段階の学習手順を図2と以下に示す。

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。

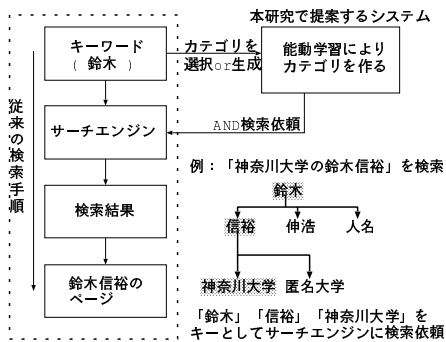


図1: 検索の流れ

3 システムの設計

本研究で提案するシステムでは、(1) ロボット型検索エンジンにカテゴリを作る。(2) 与えられたキーワード一つ一つにカテゴリを作る。(3) ユーザがカテゴリを作る。(4) 質問による能動学習を使ってカテゴリの質を確保する、ということを目指す。

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る。本研究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

本研究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

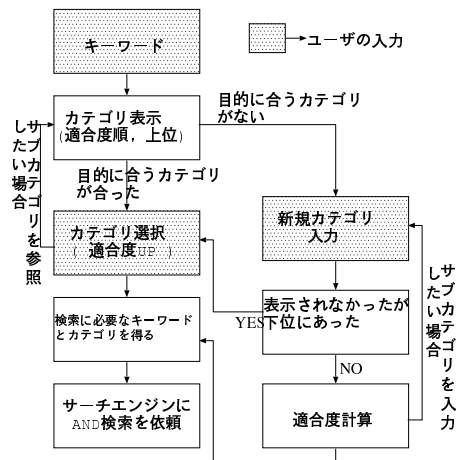


図2: 学習の流れ

4 おわりに

本研究では、現在のロボット型検索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。