

分散環境における最適データソースの複数選択 およびその統合に関する研究

山田和孝 武内裕美 山内康一郎 石井直宏
名古屋工業大学知能情報システム学科

1 はじめに

近年のインターネットの普及により、ユーザは全世界の様々な情報をインターネットを通じて得ることが可能となった。しかし、その情報量は膨大であり、これらの情報からユーザにとって有益である情報のみを検索するには多くの時間と労力が必要である。

また、提供されるサービスも WWW, FTP, NetNews, Mail など、特徴、用途の異なるものが用意されており、ユーザはそのサービスごとにクライアントプログラムを使い分ける必要がある。

故に、現在のインターネットサービスを利用するにあたり、ユーザの負担は決して小さなものではない。

そこで、マルチエージェント技術と Mixture of Experts[1] の統合手法を用いて、自動的にユーザが必要とする情報を得るのに最適なデータソースを複数予測し、それらからの複数の検索結果を統合してユーザに提示するシステムを提案する。

2 システム

2.1 システムの概要

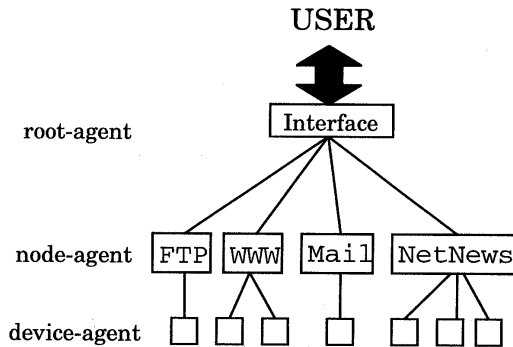


図 1: システム構造

本研究では、UNIX のファイルシステムのような木構造の階層型ネットワークを持つシステムを構築する。(図 1)

このネットワークの根、節、葉に相当する部分はエージェントによって構成される。

Multiple selections of data sources and their integration in the distributed environments

Kazutaka Yamada, Hiromi Takeuchi
Kouichirou Yamauchi, Naohiro Ishii
Dept. of Intelligence and Computer Science,
Nagoya Institute of Technology
Gokiso-cho, Showa-ku, Nagoya 466-8555

device-agent これらは WWW などの検索を行ない、検索結果を統合するのに最適な形式に変換する。デバイスエージェントはそれ自身が検索エンジンを持つものと、他の優秀な検索エンジンに検索処理を任せ、自分自身は検索結果の処理のみを行なうものがある。

node-agent ユーザからのクエリーに対し、その検索に対して最適と予想される下位層のエージェントに検索を依頼し、その結果を統合する。この予測はユーザが必要とするデータを得られたか否かによって学習し、よりの確かな予測を行なうことができるようになる。

root-agent ノードエージェントの機能の他に、ユーザとシステムの間立ち、ユーザにとってわかりやすい検索結果を提供する。

これらのエージェントを協調動作させることにより、効率的な検索を行なう。

2.2 最適なデータソースの選択方法

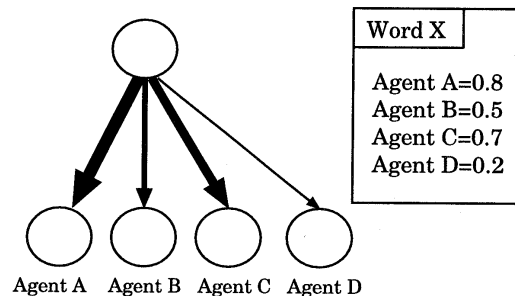


図 2: ある単語 X に対するエージェントの選択をデータ化

各エージェントは、それぞれのクエリーごとに図 2 のようなデータテーブルを保持している。このデータテーブルは検索対象の単語をインデックスに、その次の層のエージェントからその単語について良い検索結果が得られる確率を表す値（以後、検索期待値と呼称する）を保持している。各エージェントは、後述の学習により、単語毎にテーブルを蓄えていく。検索時には検索したい単語がインデックスとして登録されたデータテーブルをデータベース集合から探し、そのテーブルに記述された検索期待値に従い検索の指示をエージェントに送る。

その際、検索を行っても良い結果が返ってくる可能性が低いエージェントに対しては検索指示を行わないようにすることで、検索の効率化を図る。また、ノードエージェントは検索結果を統合する。

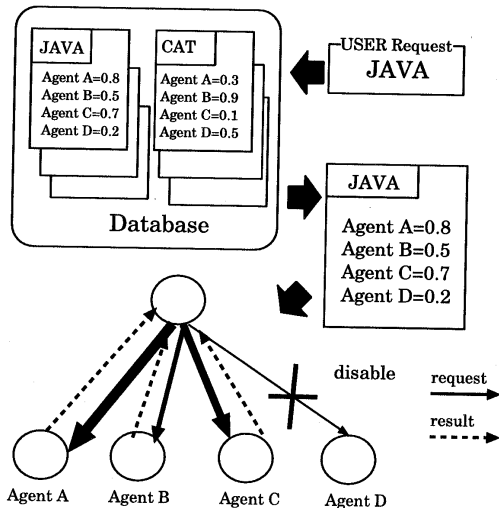


図 3: 検索実行時の動作

2.3 エージェントの学習

学習は前述した単語毎のテーブルを作成、検索期待値を更新することで行っていく。

2.3.1 新規作成

過去に検索例の無い単語に関して検索の依頼がなされた場合、その単語に関するデータテーブルを新規に作成する。

テーブルを作成する際に、すべての単語に対して、全く情報の無い状態からデータテーブルを作成するのは効率が悪い。そのため、すでに類似した単語が登録されている場合に限り、そのデータテーブルをコピーして利用することを考える。この時、単語間類似度を求めるために WordNet のデータベースを用いている。[2]

このデータベースを用いて、新規にデータテーブルを作成する単語と、過去に検索した単語との単語間類似度を求め、その類似度に応じて過去のデータをコピーするか、もしくは新規にデータテーブルを作成するかを決定する。

2.3.2 検索期待値の更新

各単語に対して作成されたテーブルのデータを、実際の検索結果とその評価に応じて更新していくことでより良い検索が行なえるように学習していく。すなわち、良い検索結果を得られるエージェントの評価（検索する可能性）を上げ、逆に良い結果を得られないエージェントの評価を下げる。

提示した検索結果に対するユーザーの評価をもとに、検索期待値の更新値を決定する。

ユーザーが直接評価を下すのはデバイスエージェントに対してのみであり、ノードエージェント、ルートエージェントは下位に属するデバイスエージェントの評価をもとに、検索期待値の更新値を決定する。

ノードエージェントの場合、検索期待値の更新値は

$$\frac{\sum_{i=1}^N u_i x_i}{Na}$$

という式で与えられる。なお

N : このノードエージェントの下位層のエージェント数

x_i : デバイスエージェント d_i のこの単語に関する検索期待値

u_i : d_i の検索結果に対するユーザーの評価

未見ならば 0, 良ならば s , 悪ならば $-s$ (s : 正の定数)

a : 更新値の増減の幅を決定する定数

全てのノードエージェントに共通

である。

2.4 ユーザへの提示

2.4.1 提示形式

検索結果は、ノードエージェントごとにユーザーへ提示される。その際にすべての検索結果を表示するのではなく、各ノードエージェントの下層に位置するデバイスエージェントのなかでソートを行ない、上位数個のデバイスエージェントの検索内容のみ表示する。

3 まとめ

本稿では、分散環境における最適なデータソースの選択と検索結果の統合手法について述べた。

将来的にはインターネット上のサービスの他、UNIX のマニュアルや EPWING 形式の CD-ROM などに対応したデバイスエージェントを用意し、ユーザがデータソースの違いを気にかけることなく、検索を可能とするシステムを構築する予定である。

今後はこれらを実装し、統合手法の改善を行いながら、評価実験を行っていく。

参考文献

[1] M.I. Jordan, R.A. Jacobs, "Modular and hierarchical learning systems", The Handbook of Brain Theory and neural Networks., Cambridge, MA: MIT Press, (1995)

[2] Masanobu Kobayashi, Xiaoyong Du, Naohiro Ishii: A New Measure of Word Similarity Based on Information Content, Proc. of International Symposium on Database, Web, and Cooperative Systems (DWA-COS'99), pp. 85-90, Baden-Baden, Germany (Aug. 1999).