

曖昧な情報要求に対する分類と推薦を用いた検索支援

和田 洋祐[†] 相場 亮[†]

芝浦工業大学大学院 工学研究科 電気電子情報工学専攻[†]

1. はじめに

現在、WWW(World Wide Web)上には膨大で多様な Web 文書が存在する。目的の Web 文書を探すために我々は Google[1]のような検索システムを用いている。ユーザが情報要求をクエリとして表現し、検索システムに入力することで、クエリに関連する Web 文書をランキングにしたリストを検索結果として入手できるが、望む結果を得るためにはキーワードを正確に入力する必要がある、これが検索を困難にする一つの原因になっている。

その一方で Web がより身近な情報源になるにつれ、検索システムは日常生活やビジネスの場面で問題解決、意思決定にかかわるようになってきた。このような場面でのユーザの情報要求は曖昧であることが多く、検索システムに入力されるクエリのキーワードは、少なく複数の解釈が可能な曖昧なものの割合が増えてきている。このようなキーワードで検索すると膨大で多様な Web 文書が混ざった検索結果が出力されるので、ユーザは試行錯誤を繰り返し、要求を明確にすると共にキーワードを修正していく必要がある[2]。

この問題に対して適合性フィードバック[3]を用いた手法が考えられる。しかし、適合性フィードバックは検索結果をただちに評価する必要があり、ユーザの情報要求が曖昧な段階では評価を下すことが難しく、要求が途中で変化したことにも対応できない。

本研究では曖昧さを詳細化していくペースをユーザに合わせることができるとある検索手法を検討する。

2. 適合性フィードバック

ユーザが検索結果の Web 文書の適合・不適合の

判定を行い、その情報に基づいてクエリを修正、拡張することで検索精度を向上させる手法を適合性フィードバック[1]という。

システムはユーザに検索結果の上位の Web 文書から適当な数の文書を提示し、ユーザはその Web 文書が自身の要求に適合しているかを判定する。これにより適合文書と不適合文書に分けることができ、システムは適合文書からは正のフィードバックを、不適合文書からは負のフィードバックを得ることにより、クエリを修正する。

しかし、ユーザの要求が曖昧である時は検索結果の適合性判定を求めることは性急であり、この時点での判定は間違いを含むことも考えられる。また、ユーザの要求が詳細化するにつれて、判定の基準が変化することも考慮する必要がある。

3. 提案手法

情報要求が曖昧であるときの検索の問題点として、以下の 2 点が考えられる。

- (1) 入力されたクエリは曖昧で、検索結果には複数の話題が混ざっているため効率的な閲覧ができない
- (2) 試行錯誤の初期段階では Web 文書の適合性判断は明確ではない。また要求が詳細化したときに変化している場合がある

そこで本研究では、検索結果を話題ごとに分割するために文書クラスタリング[4]を導入し、文書クラスタリングにより作られたクラスタを探索するユーザの行動をナビゲーションするエージェントの導入を検討する。

3. 1. 文書クラスタリング

文書集合に含まれる異なった文書間の類似度を基準に、内容的に均質なくつかの群(クラスタ)に分ける手法を文書クラスタリングという。文書間の類似度計算をする際には、文書を

Retrieval Support uses Clustering and Recommendation to Vague Information need

[†]Yosuke Wada, Akira Aiba, Department of Electrical Engineering and Computer Science, Graduate School of Engineering, Shibaura Institute of Technology

索引語の重みから構成されるベクトルで表現し、2つのベクトルの余弦を求めることが多い。

クラスタリングによって検索結果をグループ化して表示することで効率的な閲覧が期待できる。そのためには各クラスタがどのような内容の集合であるかを示すラベル付が大きな影響を持っている。

検索システムから得られた Web 文書リストを基にクラスタリングする場合には各文書のタイトルとスペニットに含まれる語を用いる

3. 2. エージェント・ナビゲーション

エージェントが行うナビゲーション中の行動としては以下の二つを検討している

- (1) ユーザの探索行動の観察
- (2) ユーザの検索意図問い合わせと、それに基づく新規クラスタの作成と推薦

以下で、現在構想中のシステムの大まかな流れ(図 4)を説明すると共に、システム内でエージェントが果たす役割についても説明する。

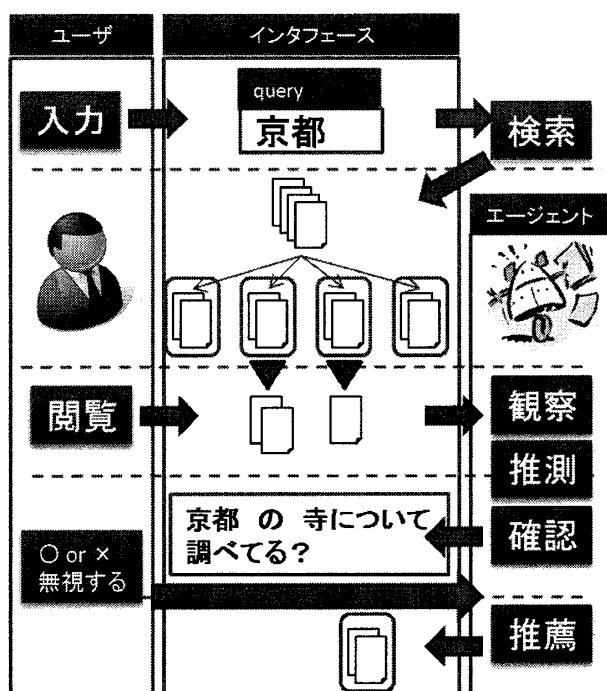


図 4：システムの大まかな流れ

- (1) ユーザが入力したクエリを既存の検索システムを用いて検索し、その結果をクラスタリングしてユーザに提示する
- (2) ユーザがクラスタを選択すると、分類され

ている文書のタイトルと概要を見ることが出来る

- (3) ユーザが文書を選択することで、文書の全文を見ることが出来る
- (4) エージェントはユーザが選択したクラスタの閲覧時間とユーザがどの文書の詳細を閲覧したかなどといったユーザのこのクラスタへの関心度の強さの指標となる情報を収集する
- (5) エージェントはこれらの情報からユーザが現在注目していると思われる話題の候補を特定する
- (6) エージェントは分析の結果をユーザに示し、ユーザの関心があるかどうかを確認する
- (7) ユーザはエージェントからの問いかけに対し肯定あるいは否定による回答を行うか、または無視することができ、エージェントは回答によりこの話題の推薦を行うか否かを決定する
- (8) 否定的回答の場合にはエージェントは次候補を検討し、その結果によって次候補の話題を推薦するか、あるいはユーザのさらなる反応を待つ

4. まとめと今後の研究方針

本稿では曖昧な情報要求しか持っていないユーザが検索システムから情報を得ることが難しいことを指摘した。また、従来の解決手法の適合性フィードバックもユーザの情報要求が曖昧だと正しく機能しないことを指摘した。

これらを考慮した解決のアプローチとして文書クラスタリングとエージェントによるナビゲーションによる検索支援を提案した。

今後はエージェントを用いたインタラクションに関する研究の調査とユーザ検索意図の分析方法のアイデアを詳細に詰め、実際にシステムを構築し有効性の検証を行う予定である。

参考文献

- [1] <http://www.google.co.jp>
- [2] 是津耕司, Web サーチ技術の現状と今後の展望-Web2.0 時代に向けて-, 人工知能学会誌 21 巻 4 号 2006
- [3] 徳永健伸, 情報検索と言語処理, 東京大学出版会 1999
- [4] 岸田和明, 文書クラスタリングの技法: 文献レビュー, Library and Information Science No. 49 2003