

概念体系に基づく Fish Eye マッチングと

1U-6

仮説推論を用いたユーザの視点の抽出

飯島 光晴 高間 康史 伊庭 斉志 石塚 満

東京大学工学部電子情報工学科

1 はじめに

昨今の情報環境の整備・拡大により、必要な情報の収集が容易になりつつある反面、その情報量が人間の処理能力を超えてしまう、いわゆる「情報過多」の問題が生じている。

我々の提案する Fish Eye マッチング[1]は、ベクトル空間モデルを拡張したものであり、ユーザの興味・視点を反映した特徴ベクトルの動的な生成・マッチングを行う。

この時、単語を縮退・選択する際の単位として、EDR 電子化辞書の概念体系辞書中の概念エントリ（意味グループ）を用いる。

本研究では仮説推論を用いて意味グループを求める手法について提案する。

2 意味グループの抽出

2.1 Fish Eye マッチング

文献検索の代表的な方法であるベクトル空間法では、単語軸の直交性が仮定されるため、単語間に存在する意味的な類似性は考慮されない。

Fish Eye マッチングはEDR 電子化辞書[2]の概念体系に基づき、概念単位で単語を縮退し、特徴とすることでこの問題を解決することができる。

2.2 EDR 電子化辞書の概念体系

EDR 電子化辞書の概念体系辞書には図1の例に示すような木構造¹で概念が記録されている。図の

枝にあたる上位・下位の概念のペアを一つのレコードとしている。Fish Eye マッチングでは類似した意味を持つ単語群を概念単位でまとめる。このようにして得られた単語のグループを意味グループと呼ぶ。

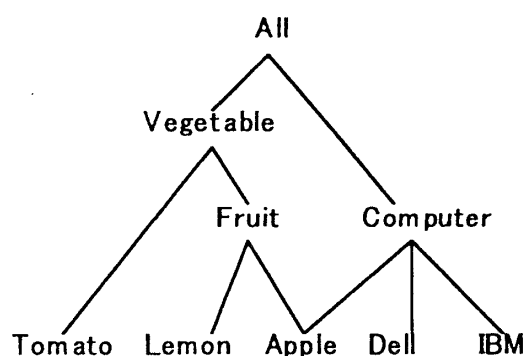


図1 EDR 電子化辞書の概念体系の例

2.3 意味グループの抽出

従来、意味グループの抽出にはグリーディな方法がとられてきた。

まず文書データからベクトル空間法と同様に文書の基本特徴ベクトルを取り出す。特徴ベクトルの軸となる単語につける重みは TFIDF をもとに適合フィードバック法と同様の式で求める。

重みの大きい単語を含む上位概念から、その概念が下位に持つ単語の重みの平均値を指標に、その値が高いものを意味グループとして採用する。

この方法の欠点は、一度重みの大きい単語群から意味グループが選ばれると、たとえ他の意味で用いられている単語が木構造でその意味グループの下位に位置していても、その単語が他の意味で使われていることが全く考慮されなくなるという点である。

Hypothetical Reasoning-Based Extraction of User's Viewpoints in Fisheye Matching. M.Iijima, Y.Takama, H.Iba, M.Ishizuka. Dept. of Information & Communication Engineering, School of Engineering, Univ. of Tokyo.

¹ 多重継承を許すので、厳密には木構造とは言えない

この問題を解決するためには、よりグローバル、つまり概念体系の全体を見渡した上で最良の探索を行う方法を用いて、意味グループを抽出することが必要だと考える。

3 仮説推論

3.1 仮説推論とは

本研究ではグローバルな探索法として仮説推論[3]を利用する。

仮説推論とは、真か偽か不明な事柄を仮説として推論を進め、矛盾なくうまく与えられた問題が解決できたり、説明が達成されれば、その仮説が正しいとする形式の推論である。

扱う知識ベースとして、常に成り立つ背景知識の集合 Σ と、矛盾を生じる可能性のある仮説の集合 H をわけて記述する。また仮説と背景知識との間で矛盾が生じないように無矛盾性制約も与える。

$$\left\{ \begin{array}{l} \Sigma \cup h \vdash G \\ (\Sigma \text{ と } h \text{ から } G \text{ が証明できる;} \\ \vdash \text{ は演繹的証明を示す論理記号}) \\ \Sigma \cup h \text{ は無矛盾} \end{array} \right.$$

3.2 概念体系の背景知識への変換

意味グループ抽出のアルゴリズムとして仮説推論を用いるメリットとして、

1. 概念体系は背景知識として利用しやすいこと
2. 知識獲得のボトルネックが解決されることが挙げられる。

そこで概念体系の背景知識と仮説への変換方法を考える。

- ゴール : 単語
- 中間ノード: 概念ノード
- 仮説: 1-意味グループとして採用すること (図2の FXx)
2-階層を登ること (図2の Hx)
- 無矛盾性制約: 上位・下位にある仮説が同時に選ばれることはない

図2の中で楕円で囲まれているのが仮説であり、上位・下位の関係にある仮説は同時に選ばれないようにする。(図中の矢印で結ばれた仮説が意味

グループとして同時には採用されない。)

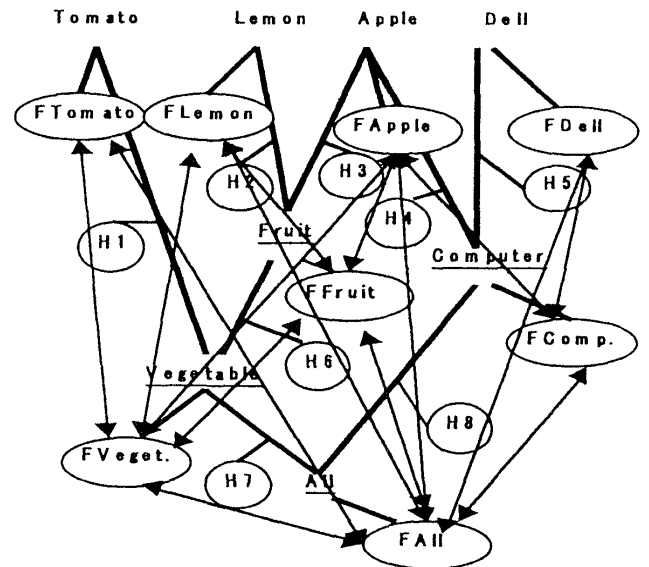


図2 概念体系の背景知識への変換例

3.2 具体的な仮説推論の導入

重みの設定方法はいくつか考えられるが、以下の条件を満たすものでなくてはならない。

- 重要な単語を多く含んでいること
- ユーザの視点を抽出するのにふさわしいだけの具体性を持っている (=抽象的になりすぎない) こと

4 むすび

現時点では、システムを開発し、実験を行っている。階層を上る仮説につける重みによって、大きく結果が変わることが分かっており、理論的な重みの設定方法も検討中である。

参考文献

- [1]高間 康史:概念体系を用いた Fish Eye マッチングによる視点を考慮した文書整理支援機能の実現, 電子情報通信学会誌(1998)
- [2]日本電子化辞書研究所: EDR 電子化辞書 1.5 版 仕様説明書(1996)
- [3]石塚 満:知識の表現と高速推論, 丸善(1996)