

## TF-IDF 法による協調行動情報に対する重み付け

藤田 征士<sup>†</sup> 高丸 尚教<sup>†</sup>

<sup>†</sup> 中部大学大学院工学研究科情報工学専攻

### 1 はじめに

人の行動には、それぞれ意味がある。行動単体だけではなく、その連なりに意味が生まれる。例えば、教師が学生にその分野での確実に行うであろう行動を教えるように、ある分野の中での特徴的な意味を抽出することができれば、その分野に新規参入した人でも、教師と同じように思考できることが期待される。

そこで、我々は、単語や文書の関係性を調べる TF-IDF 法を行動情報に対して適用することで、意味論的解釈をすすめる手法を提案する。

### 2 文章系と行動系

文章において、単語そのものにも、その連なりである sentence にも意味論的解釈ができる訳ではなく、段落レベルまでの文章群から集合知を得ることができるよう、行動においても、行動要素そのものだけでは実行行為者の意図は他者から読み取ることはできず、その連なりに意味を見いだすことであると考えている。ただし、一般に文章では単純な「繰り返し」は行われず、表面上異なる文章群を用いることで「繰り返し」を表すのに対し、行動空間においては、まさにそのままでの「繰り返し」が頻発し、そのことが「迷い」という重要な意味論的解釈を生み出している点が、文章系と行動系との大きな違いであると考えられる。

そこで、我々は基盤となる情報アセンブリとしてのユーザーインターフェースとしてサッカーボール型意味論的インターフェースを開発してきた。次節において簡単に紹介する。これは、静止画像、動画像などのマルチメディア情報表示基盤となっているだけでなく、複数の情報活動者の行動を記録し、その活動を解析することができる。このツールに本研究で提案する協調行動情報の解析結果を加え、例えば、教師行動に対する生徒の行動学習活動にフィードバックすることを念頭に研究を推進する。

By using TF-IDF method, weighting of information on the cooperative action space.

Masashi FUJITA<sup>†</sup> and Hisanori TAKAMARU<sup>†</sup>

<sup>†</sup>Graduate School of Engineering, Chubu University

Kasugai-shi, 487-8501, Japan

fujita@ss.cs.chubu.ac.jp

### 3 サッカーボール型意味論的インターフェース

サッカーボール型意味論的インターフェース (以下 SSI) とは、切頂二十面体の面に要素 (画像や映像) を貼り付けることで脳内関連 DB を構築するものである。図 1 のように、五員環に貼り付けた要素の周囲に、関連した要素を貼り付けていくことで、直接的な関連から間接的な関連を見つける可能である。そして、要素

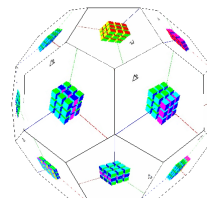


図 1: サッカーボール型セマンティックインターフェースの追加や削除などの一連の行動を表現することで思考を表現できる。

SSI で可能な行動は、4 つある。これらの行動を扱う際に、画像・映像データの判別と貼り付ける位置の判別が必要になる。これらのデータ番号や位置の番号の情報を 1 つのビット列として扱い、その数値をタグとする。これらの行動群でログを構成する。このタグを文書の単語とすることで TF-IDF 法が適用可能となる。

### 4 提案手法

本研究では、文書における情報検索の手法を行動情報に適用する初歩として、TF-IDF 法を行動情報に適用する。具体的には、3 章で説明したログの行動情報の連なりから、抽出された意味論を TF-IDF 法によって解析する。意味論の抽出法については、ユーザーの行動情報と意味論の教師行動情報との類似度によって意味論を判別して抽出する。

### 5 評価実験

今回の実験には、何かに対しての脳内関連 DB をいくつかが作成しなくてはならない。そこで、本実験では、8 人の被験者にあらかじめ用意された「紙」に関する要素 (画像) を 10 個与えて、脳内関連 DB を構築さ

せ、ユーザー別の行動情報を作成する。この行動情報にTF-IDF法を適用する。

また、意味論の抽出法について、教師行動情報に対してどのような意味論的ばらつきが起きるか検証する。その手法は、教師行動を解説しながら見せることで意味論を理解させ、その教師行動を同じ意味論の中で行動させることで、同一の意味論における行動情報のばらつきを確認する。

## 6 結果

前章で述べた「紙」に関する脳内関連DBを構築させた結果の一部を図2に、同一の意味論内での行動情報のばらつきの特徴的な3パターンを図3に示す。図

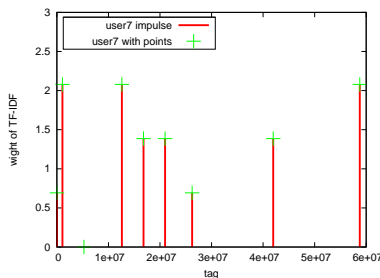
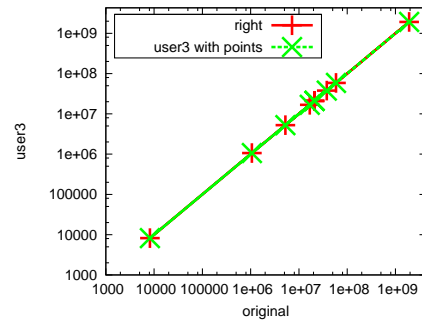


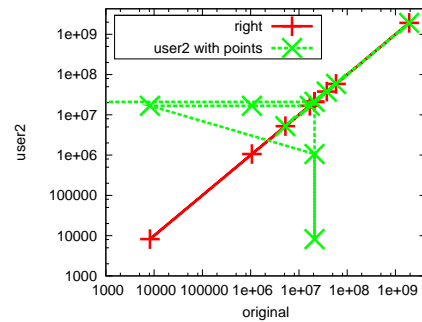
図2: ユーザの脳内関連DBの行動情報重要度グラフ (横軸:tag 縦軸:重要度)

2は、縦軸にTF-IDF法により求められた個々の行動情報に対する重要度を、横軸に行動情報のタグを対応させたグラフである。このグラフを見ると、最大重要度に差があまりついていない。これは、タグデータが少ないために起こる現象である。このことから、ユーザのログデータが大量になければ、特徴的な行動を抽出することができないとわかる。しかし、これではUIとして有用とは言えないので、行動情報一つずつでは、評価不可能であるとわかる。

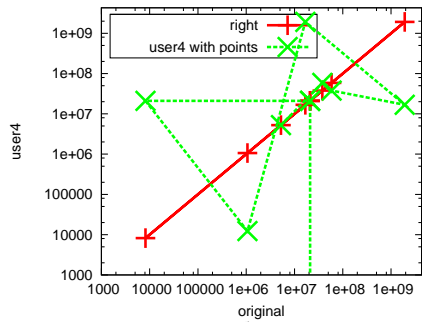
図3は、縦軸に被験者の行動情報のタグを、横軸に教師情報行動のタグを対応させたものである。この3つのグラフは、それぞれ特徴がある。3(a)は、完全に教師行動情報を再現できているパターンであり、3(b)は、水平方向に飛び出している部分は迷いの行動となっているため、この行動情報は迷いが生じたパターンであり、3(c)は、行動情報が斜めに推移している部分は行動順番のずれによるものなので、行動順番が違うパターンである。これらの特徴は、行動情報から見ると違うパターンであるが、意味論的には同義である。そのことから、やり直し動作パターンと行動順番のずれパターンの傾向については、同じ意味論とすることで意味論のカテゴリ化が可能であると考えられる。



(a) 完全一致パターン



(b) やり直し動作パターン



(c) 行動順番のずれパターン

図3: 意味論のばらつき (横軸:教師となる行動のタグ 縦軸:ユーザのタグ)

## 7 まとめ

本研究では、文書系と行動系が「繰り返し」以外対応関係にあることから、行動情報をタグにして情報検索分野の解析であるTF-IDF法を適用することを考案した。しかし、行動情報タグ単体での解析では、重要度を得ることができなかった。そこで、行動情報の意味論を抽出し、その意味論を解析に使用することを考え、意味論を抽出する際、行動情報の意味論はどのようなばらつきでカテゴリ化可能なのかを検討した。

今後の展望としては、実際に行動情報から意味論のカテゴリ化を行い、ログの特徴的な意味を抽出できるか検証する。

## 参考文献

- [1] 北 研二, 津田 和彦, 獅子堀 正幹: "情報検索アルゴリズム", 共立出版 (2002-1)