

主観観測モデル理論とそのフレームの情報検索への応用

3Q-1

持田信治

三菱重工業株式会社

矢鳴虎夫

九州工業大学

1.はじめに

現在、設計部門に於いて一層の知的生産性の向上が求められている。そのためには情報処理能力の向上を図らねばならない。求められる情報処理の1つに文書検索がある。ところが最近、電子化された文書が増え、文書量そのものが大幅に増えてきている。従ってその全てに目を通すことは出来ない。よって今後、自動的に文書を選択して、欲しい文書だけを選択する機能が必要となる。この文書を選択する行為に於いて、"おもしろさ"や"読み易さ"等、人の感覚が重要である。そこで本研究では、主観観測モデル理論を用いた文書の自動分類を試みた

2.文書自動分類機能の実現

文書の自動分類機能を実現するためには以下の課題を解決する必要がある。
 (1)人の知覚に合う分類法を決定すること。
 (2)データを分類するための座標軸を決定すること。
 (3)主観観測により文書を評価、分類すること。

以前の研究により人の知覚に基づく座標軸の設定方法としてベキ関数を利用する事が有効であることが解っている。

そこで本研究では次の課題である主観観測による文書の評価、分類について研究を行った。

今回、文書の分類手法に主観観測モデル理論に於けるフレームの考え方をを用いた。

フレームの考え方では主観による重み設定の後、文章の特徴フレームを求め、次に標準フレームとの間で類似度を求める。従って主観の設定を統一感覚的にできる利点がある。

一方、単に主観観測ベクトルを設定する方法では観測結果を正規化する必要があり、主観を感覚のままに設定することは不可能であった。

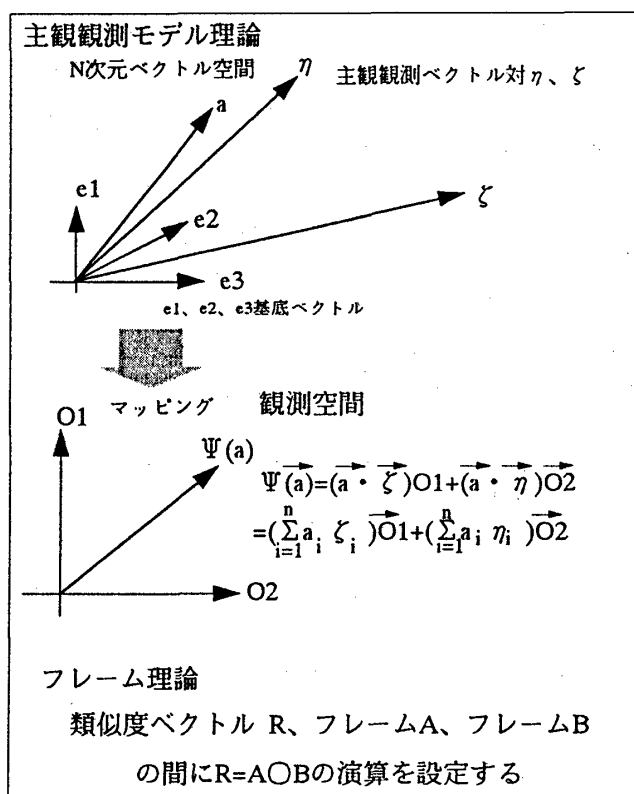
3.主観観測モデル理論とフレームの考え方

各文書の特徴値として主観による観測値を用いることとする。また観測の手法として主観観測モデル理論を用いる。

主観観測モデル理論はN次元のベクトル空間上に配置された対象を観測空間にマッピングする(図1参照)。このとき対象は主観により観測されることになる。通常観測空間は2次元である。それは2次元が最も認識しやすいからである。そして観測空間のO1、O2は2次元の直交基底ベクトルである。

そしてフレームの考え方により各文書の特徴フレームを設定する。次に文書の特徴フレームと標準フレームとの類似度により文書を分類する。

図1 主観観測モデル理論



An Application of Subjective Observation and Frame Theory to Information Retrieval

Sinzi Mochida

MITSUBISHI HEAVY INDUSTRIES,LTD.

Torao Yanaru

Kyusyu Institute of Technology

4. 文書の特徴値設定

本件では各文書の特徴値として"おもしろさ"や"読み易さ"を重視した文書の分類方法に関して説明する。

以下に本件での文書自動分類手順を示す。

- (1)各文書の特徴ベクトルS を求める。
- (2)文書検索フレームA を設定する。
- (3)文書検索標準フレームBを設定する。
- (4)フレームA とフレームBの類似度により文書の自動分類を行う。

文書の特徴ベクトルS の設定を以下の通りとする。

$$S=(s_1, s_2, s_3, s_4)$$

各パラメータの意味は以下の通り

- s1=第1キーワードのおもしろくて簡単な感覚
- s2=第1キーワードの新しくて重要な感覚
- s3=文書中の各文の平均長
- s4=文書の作成時期

s1とs2は第一キーワードをキーワード対感覚マップを用いて主観観測することにより得た。

キーワードは文書中から出現頻度の高い順に第一キーワードと第二キーワードとした。またキーワード対感覚マップは100人を対象に行ったアンケートを元に作成した(表1参照)。

アンケートは代表的な対象物を幾つかの感覚に沿って数値を当てはめさせるものとした。

数値の範囲は 10^5 から 10^{-5} の範囲とした。

次に読み易さとは文書の平均長を用いた。

例えば人の呼吸は10回/分程度であり、成人が1分間に読む文字数は平均600~800字程度と言われている、一方人は1呼吸に読める文章が最も理解しやすいと言われている。従って60~80文字程度が最も読みやすい文章長であるといえる。

表1 アンケート結果

	おもしろさ	簡単さ	新しさ	重要性
マイクロプロセッサ	6.74	2249	207	22508
パソコン	690	1	90	1
EWS	146	8	1	2320
微分法	6.5	260	0.31	271

5. 文書の自動分類

当部門で作成された技術資料を本手法により分類した。対象資料は94年11月から95年4月までに作成された40件である。本実験では文書管理に関する資料の特徴値を標準フレームとした。実験では第1キーワードと第2キーワードを用いたフレームを求め次にその特徴フレームを設定して類似度を求めた。以下に結果を示す(図2参照)。

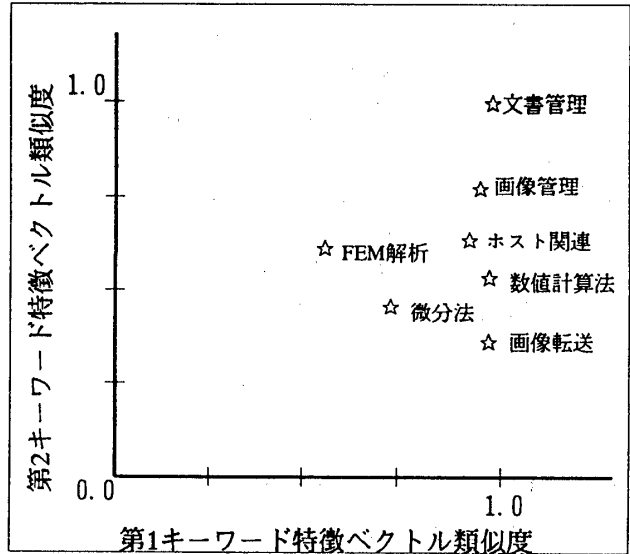


図3 観測結果

4. むすび

本研究により人の主観を尺度にした資料分類方法の実現方針が明らかになった。そして主観に基づいた文書分類機能実現の可能性が高いことが解った。今後、文書の特徴ベクトルの次元数を増やすことにより更に精度の良い分類が期待される。

今後の課題は文書の特徴ベクトルの決定方法、主観の設定方法そして文書文類後の評価方法である。また本手法の実用化のためには主観の設定と結果表示が簡単に対話的に出来ることが必要である。

参考文献

- 1) 矢鳴虎夫、吉田香
フレームに基づくパターンのベクトル化に関する主観理論 九州工業大学 H5年度 卒業論文 1994
- 2)リンゼイノーマン 情報処理心理学入門 1、2 サイエンス社 1983
- 3)大山正、東様編 認知心理学講座1、2 東京大学出版協会 1993
- 4)渡部 様編 多変量解析法入門 基礎編 福村出版 1993