

3ZE-05 MBRのためのカルマンフィルタを用いた時系列データの圧縮と予測 ～KOMIチャートを題材として～

渡部尚武, 松居辰則, 岡本敏雄

電気通信大学大学院 情報システム学研究科

1. 研究の背景と目的

記憶に基づく推論(MBR:Memory-Based Reasoning)とは、対象世界を特徴量によりベクトル表現し、ルールベースを構築することなく推論エンジンを構築する手法である。MBRは「知的活動は、ルールではなく過去の経験的記憶を中心にして行われる」という仮説に基づく推論モデルであり、問題と回答のペアを大量にデータとして蓄えておき、新たな質問にはデータベース内から最も類似した事例を検索し、その類似事例の回答をそのまま質問の回答とするのが基本的な動作である。これまでに、英単語の発音予測¹⁾、天気予測²⁾などに応用されている。

しかし、推論の精度は知識の量に大きく左右されるので、精度の高い推論をするためには膨大な知識を蓄えておく必要がある。そのため、データの増大によるシステムリソースの消費、検索効率の悪化が問題と考えられる。

そこで、本研究は上記のMBRの問題を解消したMBRを用いた推論システムの作成を目的とする。

2. 実現方法

本研究では、データの増大を防ぐために、MBR知識の中の時間的に推移している知識に着目した。

時系列的な性質を持つデータであれば、既存のデータからその次の時点のデータを予測したり、逆に過去のデータを推定したりすることが可能である。それにより、データ量の増大を防ぐことができると考えた。

そのための予測方法として、カルマンフィルタを用いることにした。

カルマンフィルタとは、時系列データの推定に用いられる確率統計的な手法であり、観測するシステムを状態空間モデルにモデル化し、その状態を推定することにより時系列データの予測や平滑化を行う逐次的な計算アルゴリズムである。

MBR時系列データにカルマンフィルタを適用させることによって、時系列データを予測し、不必要な過去のデータを削除するというシステムの構築を行っている。

また、研究の題材として、KOMIチャートの観測データを用いることにした。KOMIチャート(生活過程評価チャート)とは、観察時点でのケア対象者の状態を円形チャート化したものであり、ケアが必要な対象者の「生

活過程」(暮らし全般)に目を向け、一目でその人のその時の生活状況が読みとれるように開発されている³⁾。

本研究では、KOMIチャートのデータをMBR知識とし、その中の同一人物の長期にわたる観測データを時系列データとして扱っている。

KOMIチャートを図1に示す。

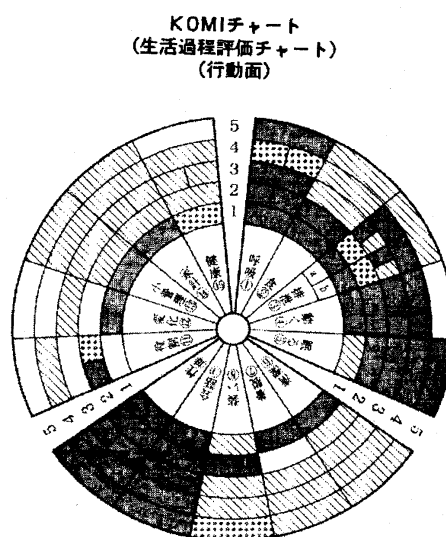


図 1. KOMIチャート記入例

3. システム構成

本システムは、以下のような3つの大きな部位に分けられる。本システムの構成を図5に示す。

3. 1. KOMIチャートデータ登録機構

既成のKOMIチャートのデータや、新規に作成したデータをMBR知識として蓄積する機能である。

本システムの知識表現を以下に示す。

KOMIチャートの78質問項目をそれぞれ1フィールドとし、その中に、質問に対する答え(1~5の5段階)を格納する。回答部として、このチャートから分かることをTEXTで格納する。そして、対象者の名前、性別、生年月日を格納する。これらのデータを用い、次に説明する推論、予測を行う。

Compressing and predicting method for time series data used Kalman Filter for Memory-Based Reasoning

Naotake Watabe, Tatsunori Matsui & Toshio Okamoto

Graduate School of Information Systems The University of Electro-Communications

1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585 Japan

3. 2. MBR 推論機構

各フィールドの重み付けの計算には、条件付き確率をもとにした特徴重み付け方法である CCF (Cross-Category feature importance)^[4]を用いた (図2).

$$w(F, c) = \sum_{c=1}^{N_c} P(c|F)^2$$

F : 事例の特徴
 c : 訓練事例の属する回答 カテゴリ
 N_c : 回答カテゴリの数

図2. 特徴重み付け手法 (CCF)

それに加え、KOMI チャートの特徴を表現するために、さらなる重み付けを行っている。KOMI チャートは一目で対象者のそのときの状態が見て取れるように円形をしており、円の中心にある項目ほどシリアスな項目を表し、また、生活レベルを表す3つの分野を配置している。

したがって、円の中心ほど、生命活動に直結するものほど、重みが大きくなるようにしてある。

また、一般的な MBR 知識表現のように2値表現をしていないので、返答値の距離をとることによりより柔軟な推論を行えるようにしている。

3. 3. カルマンフィルタ予測機構^{[5][6][7]}

KOMIチャートというドメインを多変量自己回帰モデル(多変量 AR モデル) (図3) と捉え、カルマンフィルタが適用できるように状態空間モデルで表現する (図4)。次に、予測するために必要な過去のデータ数を計算し、それ以外のデータは削除する。そして、予測や推定をすることによってデータの削減を計る。

多変量 AR モデルのパラメータ推定にはユールウォーカー法を用いている。

予測機構には、時系列の変化のパターンを記録しておく、それを参照して予測精度の向上を計り、予測の自己評価をする機能を持たせる。

この予測結果は、推論するための質問データを入力するときにも利用する。データベースに同一対象者の過去の KOMI チャートが存在すれば、そのデータを用いて予測し、質問データ入力時にデフォルト値としてあらかじめ予測値を入れておくことにより入力の手間を軽減する。

4. まとめ

以上のような構成で、現在システムを構築中である。膨大な量の知識データが必要な MBR であるが、本研究のように時系列データに着目すれば、知識の削減が可能であると考えられる。少なくとも、データベースに格納されている対象者に対して数個の過去時系列データが存在すれば、それ以外は削除可能であると考えられる。

知識の削減により推論精度の低下が考えられるが、予測、推定の精度が向上することにより最小限に押さえられるのではないかと期待している。

$$\text{多変量時系列: } x_n = (x_n(1), \dots, x_n(l))'$$

$$\text{多変量自己回帰モデル: } x_n = \sum_{m=1}^M A_m x_{n-m} + v_n$$

図3. 多変量 AR モデル

$$\text{状態ベクトル: } x_n = (Y_n, Y_{n-1}, \dots, Y_{n-m+1})'$$

$$Y_n = (y_n^1, y_n^2, \dots, y_n^l)'$$

$$x_n = Fx_{n-1} + Gv_n \quad (\text{システムモデル})$$

$$F = \begin{bmatrix} A_1 & A_2 & \dots & A_m \\ 1 & & & 0 \\ & & & \dots \\ 0 & & & 1 & 0 \end{bmatrix}, G = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

$$Y_n = Hx_n \quad (\text{観測モデル})$$

$$H = [1 \quad 0 \quad \dots \quad 0]$$

図4. KOMI チャートの状態空間モデル

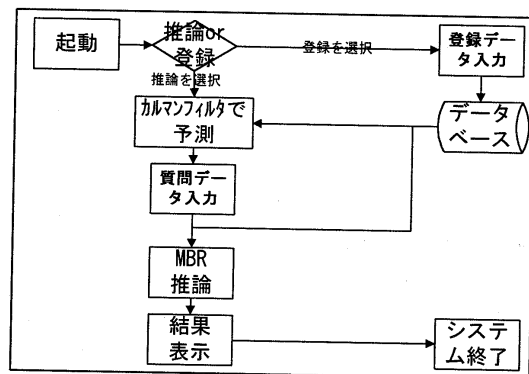


図5. システム構成

参考文献

- [1]金井一薫他:「生活過程評価チャート」をベースにした分野別チャートの開発, 学校法人日本社会事業大学, 1997
- [2]Stanfill, C. and Waltz, D.: Toward Memory-Based Reasoning, Comm.of ACM, Vol.29, No.12, 1986
- [3]毛利隆夫, 田中英彦: 記憶に基づく推論による天気予測, 人工知能学会誌 Vol.10 No.5, 1995
- [4]Creedy, R.H., Masand, B.M., Smith, S.J. and Waltz, D.L.: Trading MIPS And Memory for Knowledge Engineering, Commun, ACM, Vol.35, No.8, pp.48-63, 1992
- [5]北川源四郎: FORTRAN77 時系列解析プログラミング, 岩波書店, 1993
- [6]有本卓: カルマン・フィルター, 産業図書, 1977
- [7]片山徹: 応用カルマンフィルタ, 朝倉書店, 1983