

カルマンフィルタを用いた記憶に基づく推論のデータベース圧縮

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所

記憶に基づく推論 (MBR: Memory-Based Reasoning) はルールを用いず、事例をそのままの形式で保持することで知識ベースを構築する。そのため知識獲得のボトルネック問題の解消の一手法として位置付けられている。しかしながら、MBR では質問事例と最も近い事例を知識ベースから探索し、その回答部をそのまま推論結果とするため、精度の高い推論を実現するためには大量のデータを必要とする。そこで、本研究では、時系列データを対象にした MBR においてカルマンフィルタを用いて、MBR の知識ベースの圧縮を行う。本手法を KOMI チャートデータに適用した結果、推論能力を保持したまま事例ベースの圧縮が可能であることが確認された。

Data Base Compression of Memory-Based Reasoning by Kalman Filter

Naotake Watabe, Tatsunori Matsui & Toshio Okamoto

Graduate School of Information Systems, The University of Electro-Communications

Memory-Based Reasoning (MBR) is a reasoning method for classification problem, in MBR knowledge base is constructed without rules for reasoning, only by keeping many cases represented by original vector form of features. In MBR, to keep accuracy of reasoning, a huge of cases are required as a MBR knowledge base, so it is a serious problem for MBR method. In this study, we try to compress the size of knowledge base of MBR by using the Kalman Filter, especially in case of applying time series data. As a result of some experiments, our compression method can keep the accuracy of MBR reasoning was shown.

1. はじめに

記憶に基づく推論 (MBR: Memory-Based Reasoning) は、記憶に該当する大量の訓練事例から質問事例に近いものを解として導出する推論手法である [1][2]。MBR は知識ベースの構築に専門家を必要としない、開発期間が短い、トレーニングが不要である、動作理解が可能であり、推論結果が説明可能、1 つの事例からでも推論可能、等の長所をもつ。しかしながら、MBR では訓練事例の中で質問事例と最も近いものの回答部をそのまま推論結果とするため、精度の高い推論を実現するためには大量のデータを必要とする。そのため、システムリソースへの負担が大きく、検索速度が悪化するということが問題と考えられる。

本研究では、大量のデータが必要であるという MBR の短所を補った推論システムの構築を目標とする。すなわち、データの増加は抑えるが、推論能力は保持することが可能なシステムの構築を目指す。そのために、本研究では MBR 知識データの中にある時系列的な性質を持つデータに着目する。時系列的な MBR 知識データについて、過去のデータから現在のデータを予測することで類似する事例を 1 つのクラスターとして扱い、データの増加を抑制する。なお、本研究における時系列的なデータは、ケアプラン作成の手助けとなる KOMI (Kanai Original Modern Instrument) チャート (正式名称は生活過程評価チャート) のデータを対象とする。

2. KOMI チャート [3][4][5]

KOMI チャートはケア対象者の状態を理解するのに有効なツールである。KOMI チャートから対象者が持つ生活上の問題を解決していくための糸口を発見し、ケアプラン作成を支援することを目的としている。KOMI チャートは「生活過程アセスメント項目」として下記の 3 分野、15 項目で構成されている。

- ・ 第 1 分野 (生命の維持過程に直接影響する分野)
 - ①呼吸する ②食べる ③排泄する
 - ④動く ⑤眠る
- ・ 第 2 分野 (周囲の人々とかかわりの質に影響する分野)
 - ①身体を清潔に保つ ②衣服の着脱と清潔
 - ③身だしなみを整える ④伝える・会話する
 - ⑤性にかかわること
- ・ 第 3 分野 (「社会過程」とつながり、その人らしい生活を実現するのに影響する分野)
 - ①役割 (有用感) を持つ
 - ②変化を創り出す ③生活における小管理
 - ④家計 (金銭) を管理する ⑤健康を管理する

この項目が全て自力で営まれたとき、人間は健康で尊厳のある生き方が保たれていると考えられる。各項目は対象者の「認識面」と「行動面」を評価する細目が 5 項目ずつある。「認識面」「行動面」ともにバランスが取れていれば、健康度は高いと判断できる。KOMI

チャートの記入例を、図1に示す。

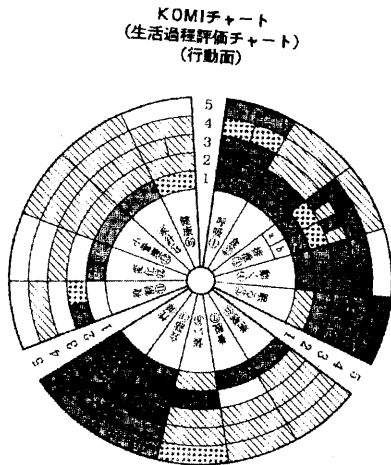


図1 KOMIチャートの記入例

3. システム構成

本研究では、カルマンフィルタによる予測機構を取り入れたMBRによる推論システムを、対象KOMIチャートから何が分かるかを回答するシステムに実装する。本システムのシステム構成図を図2に示す。本システムは、知識データ登録機構、推論機構、予測機構の3機構から構成される。

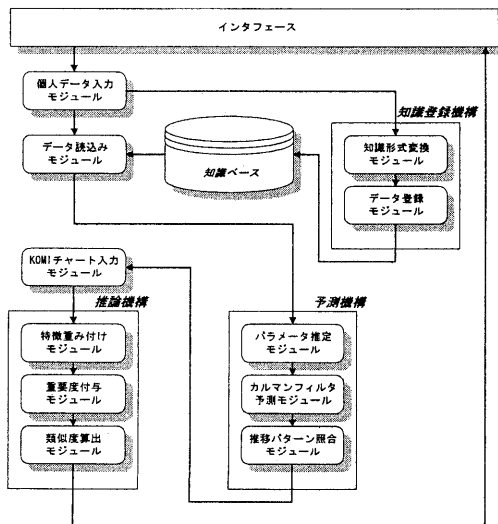


図2 システム構成

知識データ登録機構は、インタフェースの指示に従って既存のKOMIチャートのデータを入力し、それを本システムの知識形式に置き換えて知識ベースに登録する機構である。推論機構は、KOMIチャートデータから、そのKOMI

チャートから判定できる内容を推論する機構である。推論はMBRにより行われる。MBRの特徴重み付け手法としてCCF (Cross-Category Feature importance) を用いている。その後、KOMIチャートの特徴を考慮した重要度を特徴ごとの重みに反映させ、最終的な特徴の重みを算出する。この特徴重みを用いて質問データとの類似度を算出し、もっとも類似度の高いものを推論結果として出力する。

予測機構は、カルマンフィルタ[6][7]を用いて、過去のデータから1期先のKOMIチャートの形を予測する機構である。まず、多変量ARモデルのパラメータを推定する。その手法として、ユールウォーカー法を用いている。その後、そのARモデルを状態空間モデルに置き換え、カルマンフィルタによりKOMIチャートのデータを予測する。予測されたKOMIチャートは、推論時の質問データ入力に反映される。すなわち、予測された値を質問入力時の既定値としておくことにより、入力の手間を軽減する。また、過去の時系列データの変化のパターンを記憶しておくことにより、カルマンフィルタの予測精度の向上を目指し、そのデータを参照することにより予測結果の自己評価を行える機能も有する。

4. MBRの知識表現

MBRの知識は、問題部としてKOMIチャートの各項目を1つの特徴をして全ての項目を並べ、その後回答部としてKOMIチャートから判定可能な内容を格納する。最後に、名前、性別、生年月日、作成日時を格納するという形式をとる。問題部の並び順は、重要な特徴ほど前に配置されるようにしてある。すなわち、第1分野、第2分野、第3分野の順に並べてあり、さらに、各項目の1番目の質問を最初に来るように並べてある。問題部の各データは、5. 本人ができる、4. 判別できない(要観察要項)、3. 身内の援助でまかなわれている、2. 専門家の援助がはいっている、1. 本人はできない、という点数付けをして格納している。知識形式を図3に示す。

		問題部					回答部		
		特徴1	特徴2	...	特徴i	...	特徴l		
事例1		x_{11}	x_{12}	x_{13}	...	x_{1i}	...	x_{1l}	y_1
...									
事例i		x_{i1}	x_{i2}	x_{i3}	...	x_{ij}	...	x_{il}	y_i
...									
事例n		x_{n1}	x_{n2}	x_{n3}	...	x_{ni}	...	x_{nl}	y_n

図3 MBR知識形式

特徴数を l (本システムでは $l=78$)、蓄積された事例数を m 、知識ベースを $MBRrecordSet(i, j)$ と表した場合、事例は

$$MBRrecordSet(i, j) = (x_j, y_i)$$

$$(i=1,2,\dots,n \quad j=1,2,\dots,l)$$

と表現することができる。

5. 特徴重み付け手法と類似度算出方法

KOMI チャートは、熟練すると一目で対象者が現在どのような状況にあるのかを知ることができる。それには、円形であるという KOMI チャートの特徴が大きく影響している。KOMI チャートの形の上での特徴として、

- ・円の中心に近いほどシリアスな問題である。
- ・15項目を大別して3分野に分けることにより、障害のある分野が明確になる。

が挙げられる。これらの特徴を MBR に反映させるために、重み付けにこの特徴を反映させた。

特徴重み付け手法には、条件付き確率をもとにした特徴重み付け手法である CCF を用いることにした。CCF では次式で特徴重みを算出する。

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

ただし、 F は事例の特徴、 c は訓練事例の属する回答カテゴリ、 N_c は回答カテゴリの数を表す。したがって、本システムで用いる重みは以下の式で計算される。特徴 j の重みを w_j とすると、

$$w_j = \sum_{k=1}^5 \left(1 / \sum_{i=1}^n f_k(x_{ij}) \right)^2$$

ただし、 $f_k(x_{ij}) = \begin{cases} 1 & (x_{ij}=k) \\ 0 & (\text{otherwise}) \end{cases}$ である。本システム

では事例数だけ回答カテゴリが存在するという特徴があるので、各回答カテゴリの存在確率はその特徴の頻度の逆数になる。また、本システムでは、特徴は 1, 2, 3, 4, 5 のいずれかの値を取ることにしてあるので、それぞれの頻度を求めて回答カテゴリの存在確率を求め、それらの逆数の 2 乗和をとっている。さらに、KOMI チャートの形の反映をこの重み付け部で行っている。

類似度算出方法には、SUM を用いている。ここでは、KOMI チャートの特性をから各評価値間の距離を考え、その距離を基に類似の程度を 0 から 1 の間で数値化している。

$$\text{Similarity}(u,v) = \sum_{j=1}^l (W_j \cdot \text{Matrix}(x_{ij}, v_j))$$

ここで、 u_j は対象訓練事例、 v は質問事例、 l は特徴数、 x_{ij} は対象訓練事例 u_j の j 番目の特徴の値、 W_j は特徴 j の重み値、 $\text{Matrix}(x_{ij}, v_j)$ は特徴間の距離を表す。

6. 予測機構

6.1 KOMI チャートのモデル化

カルマンフィルタにより予測するために、KOMI チャートの「行動面」は全 78 項目の質問があるため、78 変数の多変量 AR モデルであると考える。

$$X_n = \{x_{n1}, x_{n2}, \dots, x_{n78}\}^t$$

が 78 変数事系列データの場合

$$X_n = \sum_{i=1}^m A_i X_{n-i} + v_n \quad (78 \text{ 変量 AR モデル})$$

$$A_i = \begin{bmatrix} a_i(1,1) & a_i(1,2) & \dots & a_i(1,78) \\ \vdots & \vdots & \ddots & \vdots \\ a_i(78,1) & \dots & \dots & a_i(78,78) \end{bmatrix}$$

m は AR モデルの次元である。この多変量 AR モデルのパラメータを、ユールウォーカー法によって求める。それにより、次元数、AR 係数の分散共分散行列を算出する。

6.2 カルマンフィルタの適用

カルマンフィルタに適用するため状態空間モデルに変換する。78 変量 AR モデルを状態空間モデルに変換すると、以下ようになる。

$$\text{状態ベクトル: } z_n = (X_n, X_{n-1}, \dots, X_{n-m+1})^t$$

$$X_n = (x_{n1}, x_{n2}, \dots, x_{n78})^t$$

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

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

$$X_n = Hz_n \quad (\text{観測モデル})$$

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

その変化の次の事例を使用して、予測された結果と比較する。直前の事例から得られた評価基準事例と変化のパターンから得られた評価基準事例が異なった場合、変化のパターンを用いた方を優先して比較、評価を行う。変化のパターンを用いた方が 2 つ前までの事例を用いて最適なものを選出していると考えられるからである。

7. 実験結果と評価

7.1 実験データ

評価実験を行うため、まず以下のようなデータを知識ベースとして蓄積した。

- ・訓練事例：過去に実際に作成された KOMI チャートデータ特徴スロット 78 個、130 事例。
- ・質問事例：推論精度を実験するために作成したテスト事例。特徴数は訓練事例に同じ、20 事例。

7.2 実験結果と評価

- (1) KOMI チャートに特化の有用性

KOMI チャートへの特化点を反映させた場合、全ての候補において類似度が高くなった。第2候補以下の類似度がどのテスト事例でも大きく上昇している。すなわちこれは、工夫を反映させないときよりも質問事例に対する推論精度が良いという結果を表している。

(2) 知識ベース削減による推論精度への影響

系列データを予測することにより削減した知識ベースを用いて推論を行った結果においては、削減を行わない場合と比較して僅かに類似度が小さくなっている。ただし、これは事例数が少なくなったために、条件付き確率をもとにした特徴重み付け手法である CCF の重み値が小さくなったためによる影響であると考えられる。すなわち、時系列データを削減することによる知識ベースの削減の、MBR 推論精度に及ぼす影響は僅かであると言える。したがって、知識ベースを削減しても、同程度の推論精度を保持することができたと考えられる。

(3) 予測精度

訓練事例中、4 つ以上の時系列データが存在する対象者は 11 人であったので、この 11 人のデータを用いた実験の結果を述べる。予測結果を質問事例、実際に出力されるべき事例を訓練事例と見なして、MBR の計算に用いた類似度を計算した。その結果、

- | | |
|---------|---------|
| ①45.107 | ②47.268 |
| ③47.126 | ④44.167 |
| ⑤52.330 | ⑥45.232 |
| ⑦50.616 | ⑧49.938 |
| ⑨42.808 | ⑩46.712 |
| ⑪46.025 | |

という類似度を得た。ただし、この類似度を計算するための重み値は前項の推論の実験と同一である。したがって、(2) の類似度と同じスケールである。(2) の第2候補の類似度の平均は 47.220、第3候補の類似度の平均は 45.708 であり、本実験で求めた類似度の平均は 47.029 である。よって、(2) の第2候補と同等の類似度が算出されたと言える。以上より、本システムの予測精度は高いと考えられる。

(4) 知識データの削減についての実験結果

訓練事例は 130 事例である。この中の時系列データのうち、AR モデルの次数を計算して削減可能と思われるものを削減した結果、15 事例削減されて 115 事例になった。本実験での削減率は 88% であった。これは、期待よりも低い数値であったが、その理由として、時系列的なデータがあまり集まらず、削減できないデータが多く蓄積されているということが挙げられる。多くの同一対象の時系列データが蓄積されれば、削減率の上昇が期待される。

8. まとめ

本研究は、カルマンフィルタにより時系列的な性質を持つ MBR データを予測し、類似事例を 1 つのカテゴリにまとめていくことにより知識ベースの増加を抑えるシステムの構築を目指した。そして、知識ベース

の大きさを抑え、かつ推論力は同等のシステムを目指した。その評価のため、推論に関する実験、予測に関する実験、データ削減に関する評価実験を行い、本システムの有用性を実証した。評価実験の結果、推論能力は同等で知識ベースの大きさを抑えるシステムが構築されたと考えられる。データ削減率に関しては期待したほどにはならなかった。しかし、これは本システムの事例数は 130 事例と少ないためであり、時系列データの数が増加することにより向上が望める。一方、推論と予測に関しては良い結果が得られたといえる。以上より、研究目的は達成されたと言える。

本システムの問題点として実行速度が挙げられる。データ追加時、予測時に毎回 AR モデルのパラメータを推定しているため、長時間計算に費やされる。ただし、本システムには 130 事例しか蓄積されていなく、MBR 推論の処理速度は十分満足できるものであった。一般には訓練事例が多くなればなるほど MBR 推論の処理速度は悪化する。知識ベースが大きくなればなるほど推論速度が遅くなり、それを回避するため知識ベースを削減すると、削減処理で遅くなる。しかし、個人の過去データは予測により削減されるため、全体の事例数が増えれば増えるほど、削減による実行速度上昇が認められると期待できる。

引用文献・参考文献

- [1] Stanfill, C. and Waltz, D. : Toward Memory-Based Reasoning, Comm. of ACM, Vol.29, No.12, 1986.
- [2] Creecy, R. H., Masand, B. M., Smith, S. J. and Waltz, D. L. : Trading MIPS And Memory for Knowledge Engineering, Comm., ACM, Vol.35, No.8, pp.48-63, 1992.
- [3] 金井一薫編著：KOMI チャート—日常ケアの実践を導く方法論，現代社，1996.
- [4] 金井一薫：ケアの原形論—看護と福祉の接点とその本質，現代社，1998.
- [5] 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.
- [6] 有本卓：カルマン・フィルタ，産業図書，1977.
- [7] 片山徹：応用カルマンフィルタ，朝倉書店，1983.