

[研究論文]

相関ルールを用いた太陽光発電電力の 予測ルール抽出手法

杉村博¹・林敏²・森武昭^{1,2}

1 創造工学部 ホームエレクトロニクス開発学科

2 工学部 電気電子工学専攻

Discovering Prediction Rules of Photovoltaic Power Generation based on Association Rule Mining

Hiroshi SUGIMURA¹, Bin RIN², Takeaki MORI^{1,2}

Abstract

This paper proposes a method, by using datamining approach, for knowledge discovery from historical usage data of a photovoltaic power generation apparatus, which is a central element of a home energy management system (HEMS). The discovered knowledge is intended to use as a basic principle to save energy in a home with HEMS. Each home has different situations of apparatuses and of preferences, thus it is necessary taking into account such differences in practical methods of energy saving. In order to achieve this goal, the obtained knowledge has enough readability and understandability for non-expert person. For this reason, we point out that traditional approaches mostly based on artificial neural networks are inadequate in HEMS environment. In this paper we show how the proposed method makes a solution for the problem.

Keywords: photovoltaic, prediction, data mining, time-series data, association rule

1. はじめに

将来のエネルギー状況のために、無尽蔵でクリーンな自然エネルギーの有効活用が重要となっている。とくに、太陽光発電設備の生産量及び導入量はともに増加を続けており、今後さらにその普及が進むと考えられる。太陽光発電設備で得られる太陽光エネルギーは日射量、パネル面の温度等により大きく変動するために太陽光発電設備の大量導入は、電力系統に対する需要が不安定化する懸念もある。この太陽光発電設備の出力変動を緩和して安定的な電力を得るためには、蓄電池などの電力貯蔵装置を併設する必要があるが、導入コストや設置エリアの確保といった問題から、できる限り小さな蓄電池を有効活用する必要がある。

エネルギー管理のためには HEMS, BEMS が注目されている。これらのシステムの現状は、各設備がネットワ

ークや信号線によって接続されて電力の見える化や遠隔操作による電源の集中コントロールが行われるまでにとどまっている。BEMS においては、それだけでもビル規模に伴って大きな節電効果もあるが、HEMS においてはさらにきめ細やかで、住む人間の負担にならない手法が求められている。そこで本研究では、HEMS への適用を考慮し、電力の変動を含めた太陽光発電の発電電力を予測する手法について開発する。

太陽光発電の予測のために、日射量を予測する研究が多数ある¹⁾。カメラを用いて雲の状態を画像解析して太陽光発電の出力変動を予測する手法があるが²⁾、室外に常設することを鑑みるとこの手法はカメラのメンテナンスコストが大きい。HEMS という家庭用設備のためには温度計や風速計といった原始的なセンサデバイスを用いた手法の開発が必要である。

天気予報と天候変化パターンを用いた日射予測手法³⁾

が提案されている。この手法は気象庁から発表された天気予報から 14 パターンの天気変化パターンを作成し、日射量を予測する手法であるが、この方法は 1 時間平均の日射量を 1 時間ごとと予測するため、蓄電池の運用計画のための予測手法としては予測期間が短い。

他にも、季節や地域の特徴を取り入れてモデルを作ることによって日射量を予測する方法が開発されている^{4,5,6,7}。太陽光発電は日射量に直接の影響を大きく受けるため、最終的な発電量はモジュール温度や木や鉄塔の影等にも影響されるため、日射量ばかりが正確に予測できるだけでは不十分である。また、これらの手法は比較的短期間の電力を予測している。本研究では蓄電池の運用計画を立てられる程度の時間を予測する必要があるが、このためには少なくとも昼までには明日の 24 時間分の発電電力の変動予測を得られる必要があると考えているが、上述した手法では 3 時間程度の予測しかできないため、もう少し長期的な電力予測を行う手法が必要である。

人工知能技術を応用した太陽光発電設備の電力予測技術として、ニューラルネットワークを用いた太陽光発電設備の 24 時間先発電電力予測方法⁸が提案されている。日射量曲線の予測に基づいた予測日積算日射量の誤差低減方法についてもニューラルネットワークを用いる方法が提案されている⁹。このように長期的な予測技術も開発されてはいるものの、ニューラルネットワークを用いたモデルは最終的な知識構造を人間が理解することは困難であり、この予測モデルによって出力された予測ルールを使って HEMS 制御プログラムへ応用する事は困難である。さらにこの手法は、住人による予測ルールのカスタマイズも行えない。

本研究では人工知能技術を応用したデータ解析技術である時系列データマイニングを用いて、12 時間先から 24 時間分程度の長期的な発電電力の予測ができる手法を開発する。さらに、ここで抽出する知識が HEMS のプログラムへと応用が可能になるために、IF-THEN 型のルールを抽出する。そこで、相関ルールマイニングの手法をベースとした時系列データ予測ルール抽出方法を開発した。

2. システム概要

本システムによって、発見する知識を Fig. 1 に示す。複数センサはそれぞれ各時系列データを格納しており、それぞれの数値の変化（形状）の組み合わせから発電電力の未来状態を予測する。直感的な例としては、湿度が上がり、気圧がさがって行くときに雨、もしくは曇りの予測ができるとすれば、近い将来の大きな発電電力は得られないだろうと予測できる。ここで、蓄電池の運用計画のために「発電する・しない」、「発電量が多い・少ない」といった定性的な予測ではなく、図に示すとおり「どのような動きで発電する」といった値の動きまで含んだ予測を行うことができれば、蓄電池の運用計画が立てられると考えられる。

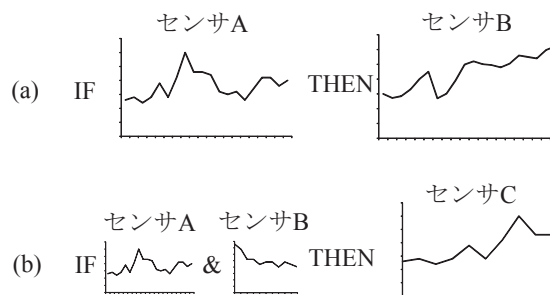


Fig. 1 抽出する知識

Fig. 2 にシステムの概要を示す。太陽光パネルからの発電電力、電圧、電流が時間とともに時系列データベースに記録されている。さらにこの時系列データベースには、気温、湿度、日射量、太陽光パネルの温度といった太陽光発電のコンディションにかかわると思われるセンサデータについても記録している。知識発見装置はここで記録されたデータに基づく相関ルールを抽出する。抽出されたルールはルールデータベースに記録されるとともに、ユーザに対して出力する。

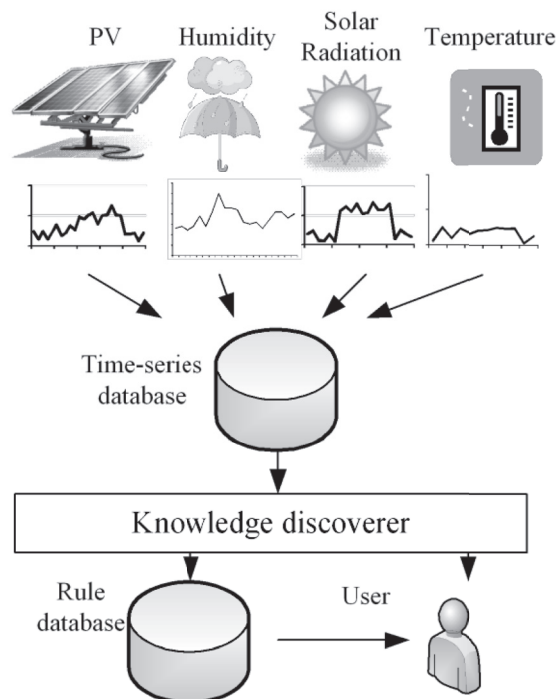


Fig. 2 システム概要

Table 1 に太陽光発電のデータを示す。Index はデータレコードに連番を振っており、データの計測時間を Time 列に格納する。計測データについては、PG kWh は発電電力量、PG kW は発電電力、PG V は発電電圧、PG A は発電電流をそれぞれ示している。

Table 1 時系列データベース

Index	Time	PG kWh	PG kW	PG V	PG A
1	2009/08/16 5:00	0.00	0.00	104.0	0.03
2	2009/08/16 5:05	0.00	0.00	104.0	0.03
3	2009/08/16 5:10	0.00	0.03	104.0	0.95
4	2009/08/16 5:15	0.01	0.07	104.0	1.04
5	2009/08/16 5:20	0.00	0.10	104.0	1.11

Table 2 に、温度、湿度といった外部に設置したセンサのデータベースを示す。一定間隔で取得したデータは日時とともに記録される。

Table 2 環境センサデータベース

Index	Time	Temp	Hum	Wind	Rain
1	2009/08/16 5:00	23.3	81	0.0	0.0
2	2009/08/16 5:30	23.4	81	0.0	0.0
3	2009/08/16 6:00	23.7	80	0.4	0.0
4	2009/08/16 6:30	24.1	79	0.4	0.0
5	2009/08/16 7:00	24.6	76	0.9	0.0

Fig. 3 に知識発見装置のアルゴリズム全体の動きを示す。相関ルールマイニングを行うために、3 種類の内部機能が動作する。データベース内部には複数のセンサからなる時系列データが格納されているため、代表的な形状発見の部分に関しては各センサごと独立して実行する。まずは分割機能によって時系列データを 0 時から 23 時 55 分で切りとることで日毎に分割する。これによって獲得した全ての部分時系列データを用いてクラスタリングを行い、センサごとの代表的な数値変動の形状を獲得する。クラスタリングではデータの距離を定義する必要があり、通常はユークリッド距離を用いて行うが、時系列データの性質から Dynamic Time Warping (DTW) 距離に基づいて計算を行う。これにより獲得したクラスターの重心が、そのセンサの代表的な数値変化の形状とみなすことができる。

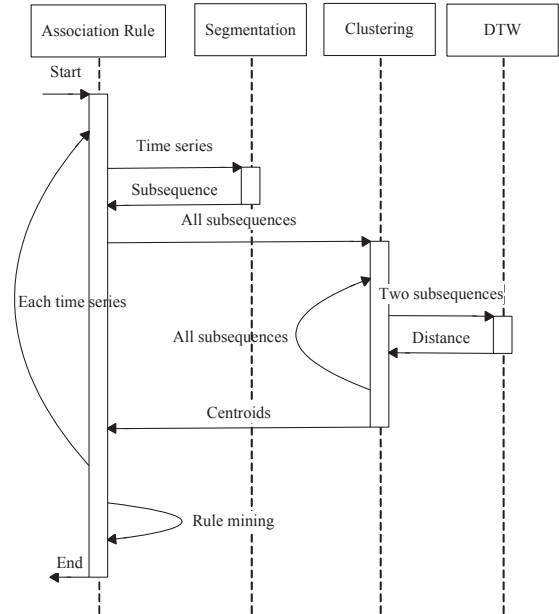


Fig. 3 ルール発見アルゴリズム

以上の操作を全てのセンサの時系列データに関して行うことで、すべてのセンサの代表的な形状を獲得する。最後に、この代表的な形状の出現確率を基にして相関ルールマイニングを行い、数値変化の関係性について IF-THEN 形式の知識を抽出する。これらの詳細な機能について、次節から説明する。

3. ルールの抽出

部分時系列データを網羅的に獲得するためにスライドウィンドウによって切り出す。この手法を用いた場合には得られる部分時系列データの種類の総数は膨大になり、それらの各出現頻度もゼロに近くなるため、切り出した全ての部分時系列データから幾つかの代表的な部分時系列データを発見する必要がある。時系列データの類似度または距離を計算することでグループ分けを行い、その中心値となるデータを代表時系列データとして取り出す。代表的な部分時系列データの発見とグループ分けのためにクラスタリングを行う。

クラスタリングには k -means¹⁰⁾を用いる。 k -means は最適なセントロイドの位置を探索しながらグループ分けを行う手法で、今回のグループの代表的な形状を抽出する目的にマッチしており計算量の短縮が見込める。 k -means は、分割の良さの評価関数を定めて、評価関数を最適にする分割を探索する。評価基準は部分集合 G_i に含まれる各データ v と G_i のセントロイド (重心) g_i の間の距離である。重心を $g_i (i = 1, \dots, k)$, $D(P, Q)$ をデータ P とデータ Q の間の距離を測定する距離関数としたとき、評価基準は次の式(1)で定義される。

$$\sum_{i=1}^k \sum_{v \in G_i} D(v, g_i) \quad (1)$$

スライドウィンドウによって機械的に切り出した部分時系列データに対し、評価関数としてユークリッド距離を用いて k-means を行った時に、意味がないセントロイドが出力されてしまう問題がある¹¹⁾。そのため、機械的な部分時系列の切り出しという枠組みの中で、この問題を解決する方法も提案されているが¹²⁾、ここではスライドウィンドウのサイズとスライド幅を単純に固定するのではなく、分野依存の知識に基づいて決定することで解決する。文献(13)は対象領域の知識を使ってデータの変化に注目したスライドウィンドウのサイズとスライド幅の決定を行い、不等間隔となるスライドウィンドウを用いることで意味のあるクラスタリングを行って意味のあるクラスタを抽出する方法を提案している。本論文でも数値の変化に基づいた不等間隔の切り出しを行う。実際には各センサデータのポイント数を固定せずに、0 時から 23 時 59 分までのデータを 1 日分として切り出す。このため、各データのサイズは等間隔にならず、これによって得られた部分時系列データから抽出されたセントロイドが、単純なsin波にはならないことを確認している。k-means において距離関数 $D(P, Q)$ は一般的に属性ベクトル間のユークリッド距離で求められるが、ユークリッド距離はシーケンスの各要素を独立して比較するため、長さの異なるシーケンスのペアの距離を計算できない。そこで距離測定関数 $D(P, Q)$ は Dynamic Time Warping (DTW) を用いる。DTW¹⁴⁾は 2 つの時系列データの距離を最小化するように時間軸を伸長させる変換処理で、DTW 距離は動的計画法に基づいて計算される。Fig. 4 にユークリッド距離と DTW 距離の違いを示す。Dinamic Time Warping では中央の同形を認識して時間のずれとして距離を求める。

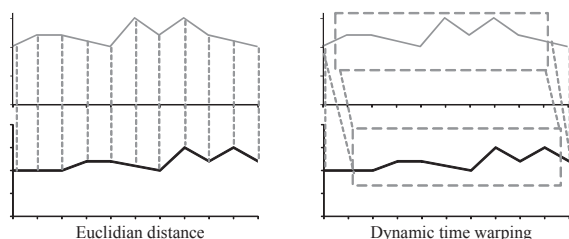


Fig. 4 距離計算の違い

長さ i のシーケンス $P = (p_1, \dots, p_i)$ と長さ j のシーケンス $Q = (q_1, \dots, q_j)$ との DTW 距離 $D(P, Q)$ は、下記の式(2)で定義される。

$$D(P, Q) = f(p_i, q_j) \min \begin{cases} f(p_i, q_{j-1}) + s \\ f(p_{i-1}, q_j) + t \\ f(p_{i-1}, q_{j-1}) + u \end{cases} \quad (2)$$

この時、 s と t は時間軸に沿ってシーケンスを伸縮するコストを示す、そして、 u は値の違いによる距離コストを示す。

次に、クラスタリングされた形状を基にして、相関ルールを作成する。最も一般的な相関ルールは Agrawal が提案した、アイテム集合をトランザクションとし、その同時生起についてユーザが設定した最少支持度と最少確信度の制約を同時に満たすような全ての相関関係を導き出すもので、代表的な手法に Apriori アルゴリズムがある¹⁵⁾。この手法は全てのアイテムを等価に扱う。この手法を改良して、ルールの帰結部をクラスという特別な属性に固定した相関ルールを取り出すことで、クラス識別の要因を説明するルールのみを取り出すようにした手法をクラス相関ルールという¹⁶⁾。

本研究では、太陽光発電の状態（電圧、電流、電力など）の予測が目的であるため、クラス相関ルールを用いることによって、たとえば明日の電力の予測を説明するルールを取り出すことができると考えられる。Fig. 5 に相関ルールを獲得するためのトレーニングデータを示す。

Sensor A	Sensor B	Sensor C	PV Power
2013.5.10	2013.5.10	2013.5.10	2013.5.11
2013.5.11	2013.5.11	2013.5.11	2013.5.12
2013.5.12	2013.5.12	2013.5.12	2013.5.13
2013.5.13	2013.5.13	2013.5.13	2013.5.14

IF part THEN part

Fig. 4 training data

Fig. 5 に示すように、センサーA, B, C と太陽光発電電力の数値変動は前述の方法によってそれぞれ日毎にクラスタ化されており、さらに太陽光発電電力だけが次の日のクラスタとして一行分のデータをつくる。この一行分のデータをトランザクションと呼ぶ。センサーA, B, C から得られた値変動をルールの条件部、太陽光発電電力をルールの帰結部、つまりクラスとする。

E をアイテム集合とすると、 E を含むトランザクションの要素数は $\sigma(E)$ とする。 T をデータベース中の全トランザクションとした時、共通要素を含まないアイテム集合 X と Y 、 $X \cap Y = \emptyset$ に対して、相関ルール $X \rightarrow Y$ の支持度 s 、確信度 c は次の式(3)、(4)で定義される。

$$s = \frac{\sigma(X \cup Y)}{T} \quad (3)$$

$$c = \frac{\sigma(X \cup Y)}{\sigma(X)} \quad (4)$$

本研究では、全ての相関ルールを作成した後に、右辺がクラスとなっているルールのみを抽出することによってクラス相関ルールを導きだした。これが予測ルールである。

4. ルールの適用

Fig. 6 にルール適用のためのアルゴリズムを示す。全センサーデータについて、センサーデータを1日分切り取り、ルールに使用したセンサに関するすべてのクラスタとDTWを用いて比較する。

最少距離のクラスタをそのセンサーデータのクラスタとみなす。次に、そのクラスタ集合がどの相関ルールに当てはまるかをチェックする。この時、複数の相関ルールとマッチする場合があるが、もっとも多くのクラスタが相関ルールの条件部のクラスタとマッチしたものを最終的に選択する。このようにして選択した相関ルールのTHEN部分を予測結果とする。

```

; 全センサーデータについて下記を行い、クラスタを決定する。
Loop ( Sensor data  $d' \in \text{All sensors}$  )
; センサーデータを1日分切り取る
Subsequence  $seq = \text{Segmentation}(d', \text{day})$ ;
; ルールに使用したセンサに関するすべてのクラスタと比較
Loop (  $cluster_i \in \text{All clusters for the sensor}$  )
 $sim_i = \text{DTW}(seq, cluster_i)$ 
End
; 最少距離のクラスタを選択する
Select a cluster  $C_j$  which is minimum  $sim_i$ ;
End
; ルールの内、最大マッチするルールを選択する
 $R = \text{Maximum matching}(C_j, \text{Association Rules})$ 
; 選択したルールのTHEN部分を予測結果とする
Return ( Then part of  $R$  )

```

Fig. 6 相関ルールによる予測アルゴリズム

5. 実験

実際の家屋に取り付けた太陽光パネルとセンサを用いて実験を行った。Fig. 7 は実際のセンサの取り付け状況の写真である。一般住宅の屋根の上に太陽光発電パネルを設置し、環境測定用のセンサとしてそのパネル設置位置の一角に気温、気圧、雨量、湿度、風速、風向、日射センサを取り付けた。また、パネル裏面に熱電対をとりつけてパネルの温度を測定している。分電盤では一次側と二次側において電流、電圧、電力、電力量を計測している。



Fig. 7 環境センサーの取り付け状態

Fig. 8 は計測環境の写真である。各センサから獲得したデータはロガーを用いて計測し、パソコンに集約して蓄積している。



Fig. 8 ロギング環境

使用したデータは2009年8月から2012年4月までの期間で、1398日分を用いた。環境測定用のセンサは30分おきに計測し、およそ4万件の記録となった。パネルの温度は10分おきに計測し、およそ14万件の記録となった。分電盤のセンサは5分おきに計測し、およそ29万件の記録となった。

相関ルールの実験結果をTable 3に示す。最少支持度を0.01、最少確信度を0.3として実験する。ここで、用いるセンサーデータを「日射量」「日射量&気温」「日射量&気温&湿度」の3タイプで抽出されたルール数を求めた。

Table 3. センサの増加によるルール数の増加

使用センサ	ルール数
日射量	10
日射量, 気温	322
日射量&気温&湿度	714

実際に抽出されたルールの例をFig. 9示す。抽出されたルールは最少確信度0.3の元で発見されているため、全

ルールの予測精度は 30%以上を確保できる。

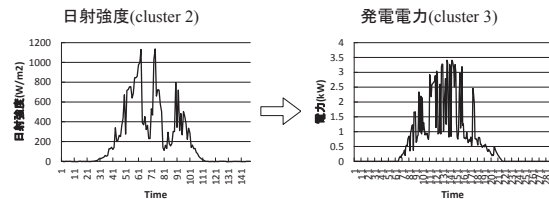


Fig. 9 実際に抽出された相関ルールの例

6. 終わりに

従来の 3 時間程度の予測ではなく、12 時間前から 24 時間分といった長期的な予測を行うことは、HEMS において家庭用蓄電池を有効に活用する技術として重要である。さらに、ここで得られる予測モデルが実際に HEMS 上で実行可能な形式になっており、ユーザがルールの意図を理解してカスタマイズ可能なことが重要である。そこで本研究では、従来のような高精度な予測はできなくとも、比較的長期的な予測が行えて、抽出されたモデルがシステムへ簡単に導入可能な IF-THEN 形式で表現されるモデルを抽出するために、時系列データマイニング技術を用いて太陽光発電電力の予測を行う技術を開発した。BEMS, HEMS といったシステムが注目されている中で、このような技術が今後の HEMS には重要になってくると考えられる。

今後は、発電電力の予測だけでなく、負荷側についての予測手法を検討し、その両者を組み合わせて実際の蓄電池運用計画を導き出す手法についても開発する予定である。

参考文献

[1] 大関崇・加藤丈佳・荻本和彦, 「太陽光発電の発電量予測の現状と課題」, 電気学会研究会資料. MES, メタボリズム社会・環境システム研究会, Vol. 2011, No. 1, pp.19-24 (2011)

[2] 山本茂広・片木威・朴在植・橋本武・橋本岳, 「雲の画像解析による太陽光発電の出力変動予測に関する基礎的研究」, 電気学会論文誌. B, Vol. 119, No. 8, pp.909-915 (1999)

[3] 嶋田尊衛・黒川浩助: 「天気予報と天気変化パターンを用いた日射予測」, 電気学会論文誌 B (電力・エネルギー部門誌), Vol. 127, No. 11, pp.1219-1225 (2007)

[4] 山本茂広・角和芳・西川榮一・橋本武: 「太陽光発電電力予測を用いた太陽光・ディーゼルハイブリッド発電システム運用法」, 電気学会論文誌 B, Vol. 124, No. 4, pp.521-530 (2004)

[5] 工藤満・竹内章・野崎洋介・遠藤久仁・角田二郎: 「エネルギーネットワークにおける太陽光発電予測技術」, 電気学会論文誌 B, Vol. 127, No. 7, pp.847-853 (2007)

[6] 鈴木孝宣・後藤悠主・寺園隆宏・若尾真治・大関崇: 「Just-In-Time Modeling に基づく日射量予測手法の開発」, 電気学会論文誌 B, Vol. 131, No. 11, pp.912-919 (2011)

[7] 桶真一郎・村田浩之・東山伸・滝川浩史: 「全国各地の日日射量曲線予測のための気候区代表モデル法」, 電気学会論文誌 B, Vol. 128, No. 1, pp.151--156 (2008)

[8] 奥那篤史・千住智信・舟橋俊久・関根秀臣: 「ニューラルネットワークを用いた太陽光発電設備の 24 時間先発電電力予測」, 電気学会論文誌 B (電力・エネルギー部門誌), Vol. 128, No. 1, pp. 33-39 (2008)

[9] 桶真一郎・村田浩之・滝川浩史・榊原建樹: 「日日射量曲線の予測に基づいた予測日積算日射量の誤差低減」, 太陽エネルギー, Vol. 33, No. 2, pp.53-58 (2007)

[10] T. Kanungo, D. Mount, N. Netanyahu, C. Piatko, R. Silverman, and A. Wu: "An Efficient k-means Clustering Algorithm: Analysis and Implementation", *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, Vol. 24, No. 7, pp. 881-892 (2002)

[11] Eamonn Keogh and Jessica Lin and Wagner Truppel: "Clustering of Time Series Subsequences is Meaningless: Implications for Previous and Future Research", *Proceedings of the Third IEEE International Conference on Data Mining (ICDM'03)*, pp. 115-122 (2003)

[12] 藤巻 遼平・広瀬 俊亮・中田貴之: 「部分時系列クラスタリングの周波数解析と位相整列処理の評価」, 人工知能学会論文誌, Vol. 25, No. 4, pp. 540-548 (2010)

[13] 大崎 美穂・阿部 秀尚・北口 真也・横井 英人・山口 高平: 「不等間隔な標準化と量子化を用いた時系列パターン抽出手法の開発と評価」, 人工知能学会知識ベースシステム研究会, Vol. 69, pp. 39-43 (2005)

[14] D. J. Berndt, and J. Clifford: "Using Dynamic Time Warping to Find Patterns in Time Series", *In Proceedings of KDD-94: AAAI Workshop on Knowledge Discovery in Databases*. pp. 359-370 (1994)

[15] Rakesh Agrawal, Tomasz Imielinski and Arun Swami: Mining association rules between sets of items in large databases, *ACM SIGMOD Record*, Vol. 22, No. 2, pp. 207-216 (1993)

[16] F. Coene, P. Leng and L. Zhang: "Threshold Tuning for Improved Classification Association Rule Mining", *Proceedings of Pacific-Asia Conf. In Knowledge Discovery and Data Mining (PAKDD05)*, (2005)