

データマイニングを用いた太陽光発電の発電電力予測 A Prediction Method for Generating Power of PV System using Data Mining

林 敏†
Bin Rin

杉村 博‡
Hiroshi Sugimura

森 武昭††
Takeaki Mori

1. はじめに

本稿では、データマイニングを用いた太陽光発電の発電電力予測を提案する。二酸化炭素の排出量削減を目的として、太陽光発電をはじめとする再生可能エネルギー機器の導入が進められている。しかしながら、太陽光発電の発電電力は気象条件に大きく作用されるため、発電電力が安定しないという問題がある。太陽光発電を有効に活用するために、発電電力を事前に予測する必要がある。そこで、データマイニングの技術を応用し、過去データベースから有効なパターンを発見することで、未来の発電電力を予測する手法を提案する。

2. 提案手法

図1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、その時の気温、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されるセンサデータについても記録している。知識発見装置は図2に示す手順によって、後述する技術を用いることによって、最終的に時系列データを用いた相関ルールを抽出する。抽出された相関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。

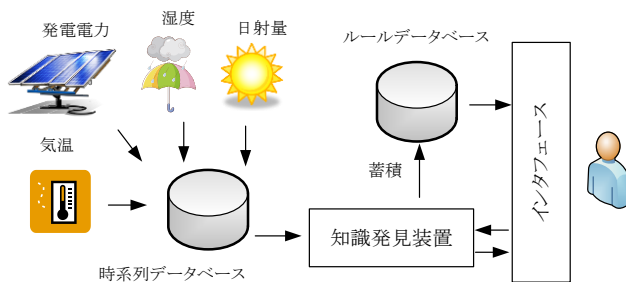


図1. システムの概要

本システムによって、発見する知識を図3に示す。いくつかの種類から集められた時系列データの形状的な変化に基づいた相関ルールを抽出することにより、そのセンサ間の関係性について発見する。

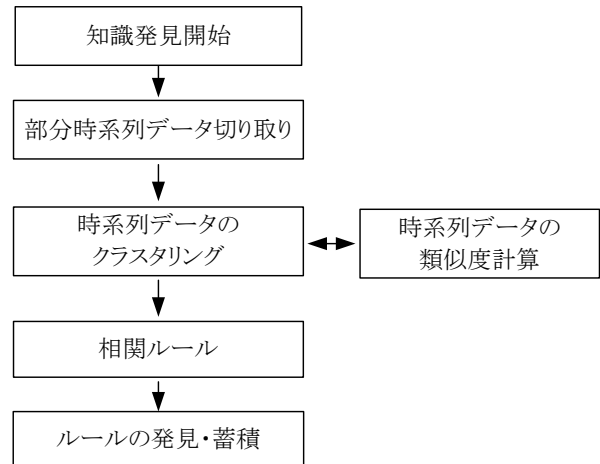


図2. 知識発見の流れ

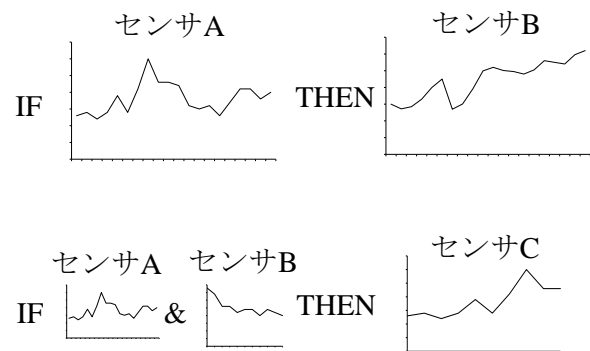


図3. 発見する知識

2.1 クラスタリング

本論文では非階層的クラスタリングのK-meansを適用する。この手法は、分割の良さの評価関数を定め、その評価関数を最適にする分割を探索する。評価基準は $x_i^{(j)}$ を含む部分集合 C_i のそれぞれのデータ x_i とクラスタのセントロイド(重心) C_j の間の距離である。重心を C_1, \dots, C_k , $D(p, q)$ は、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、評価基準を $Err(X)$ は次の式で定義される。

$$Err(X) = \sum_{j=1}^k \sum_{i=1}^N D(x_i^{(j)}, c_j)$$

本研究において、各データは時系列データであるため、クラスタリングの距離計算 $D(p, q)$ にはDynamic Time Warpingを用いた。

2.2 Dynamic Time Warping

Dynamic Time Warping(DTW)は事前に定義されたコストパラメータを用いて二つのシーケンスの距離を計算する。違う長さを持つ i, j の二つのシーケンスのために、距離 $g(i, j)$

† 神奈川工科大学大学院 工学研究科 電気電子情報専攻
Kanagawa Institute of Technology, Graduate School of Engineering, Department of Electrical and Electronic Engineering

‡ 神奈川工科大学 創造工学部 ホームエレクトロニクス開発学科
Kanagawa Institute of Technology, Faculty of Creative Engineering, Department of Home Electronics

は下記の計算式で定義される。このとき、 $i-1$ と $j-1$ は最終値を削除したシーケンスである。

$$g(i, j) = \min \begin{cases} g(i, j-1) + q \\ g(i-1, j) + r \\ g(i-1, j-1) + s \end{cases}$$

q と r は時間軸に沿って伸縮を行うコストを表している、そして s は値の違いのコストを示している。より類似のペアは小さい値となり、それらが完全に一致するときにゼロとなる。

2.3 相関ルール

相関ルール (Association Rule) とは、ある事象 X の下である事象 Y が発生する関係を表し、矢印を用いて $X \rightarrow Y$ と記述する。矢印の左側の X の部分を条件部、 Y の部分を帰結部と呼ぶ。 X が起きたとき必ず Y が起きるのではなく、一定の確率で起きる事を示している。

相関ルールの重要度を表す指標として支持度と確信度がある。支持度は全トランザクションのうち、条件部と帰結部に登場する全てのアイテムを含むトランザクションの割合であり、そのルールが登場する頻度を表すものと言える。確信度は、前提部のアイテムを買う人が帰結部のアイテムを買う確率である。支持度と確信度の双方が高いほど重要なルールで、通常は最低支持度と最低確信度を予め決め、両者をクリアするルールを抽出する。

本研究では、この出現頻度を基にした考えを用いて、上述した時系列データと時系列データに関するルールを抽出する。

3. 実験

本研究ではまず、1日の日射量と発電電力についての相関ルールを知識として抽出する。この2つのデータは直感的に直接的な関係があると思われるため、システム全体の手法として期待される効果の確認ができると思われる。

用いたデータは2009年8月から2012年4月までの、実際の家庭で観測した太陽光発電のデータを使用する。図4のように、部分時系列データは日毎に区切って獲得する。

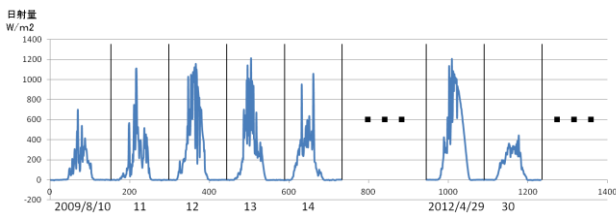


図4. 日射量の部分時系列データ (日毎)

図5に日射量のデータを実際にクラスタリングした結果を示す。 k -meansを用いているため、クラスタからはその代表となるセントロイドを獲得できる。図5はそのセントロイドである。太陽光発電電力についても同様に日毎に区切った部分時系列データを基にしてクラスタリングを行った。

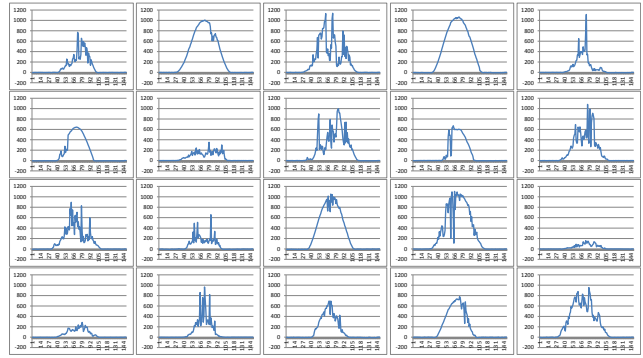


図5. 日射量のクラスタ ($k=20$)

図6に日射量と太陽光発電電力の相関ルールを示す。Apriori アルゴリズムを単純に用いた後で、現実世界における依存関係を考慮して、左辺が日射量のものだけを取り出した。最少支持度は0.001, 最少確信度は0.3としている。右辺の数値は確信度を示している。

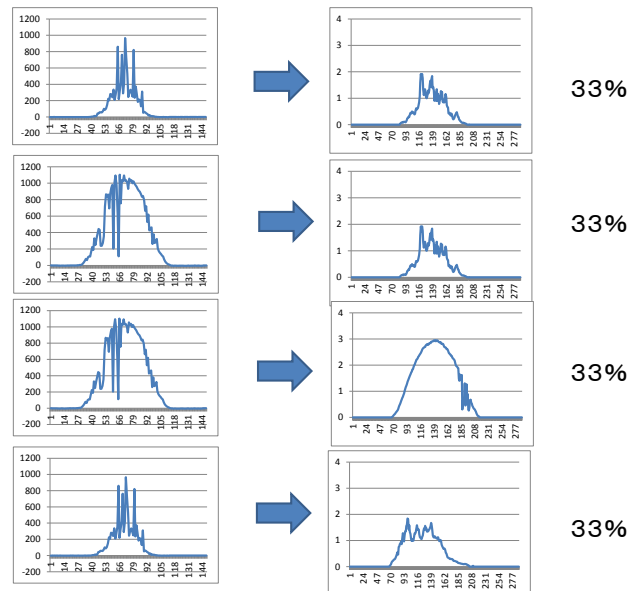


図6. 日射量と太陽光発電電力の相関ルール

4. まとめ

本論文では、データマイニングを用いた太陽光発電の発電電力予測を提案した。発電電力以外のセンサデータを用いることによって、現実世界の状況との関係性を抽出する。そのために、時系列データ同士の相関ルール抽出方法について開発した。

今後の研究は、抽出したルールによって、実際に未来状態を予測してゆくことで、さらなる研究の発展を目指す。

参考文献

- [1] 元田 浩 ほか, データマイニングの基礎, IT Text 情報処理学会編集, オーム社, (2006)
- [2] 杉村博, 松本一教: 背景知識を利用したデータマイニング, 2012年人工知能学会, 2B1-R-3-6, 2012.
- [3] 小野田崇, 坂井美帆, 山田誠二: k -means 法の様々な初期値設定によるクラスタリング結果の実験的比較, 人工知能学会, IJ1-OS9-1, 2011.