

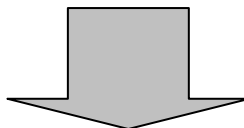
異種混合学習技術

大量データ中に混在する
全く異なるタイプのパターンや規則性を自動発見

NEC
情報・ナレッジ研究所
2012年6月

この技術が開発された背景

1. データマイニングは、データ中の隠れたパターンや規則性を抽出して予測や異常検出に応用する技術。
2. 色々なパターンや規則性に従うデータが、区別なく収集されてしまっている場合は、専門家がパターン切り替えタイミングの仮説をたてて、データを分割整理しマイニングしていた。
3. 正しい仮説を人間が立てるのは困難で、予測や異常検出も上手くできず、ビッグデータを有効活用する障害に。

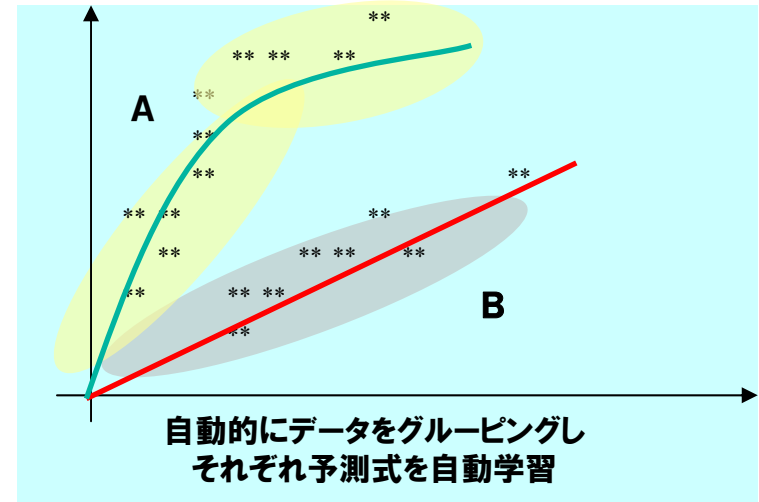
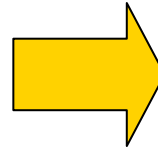
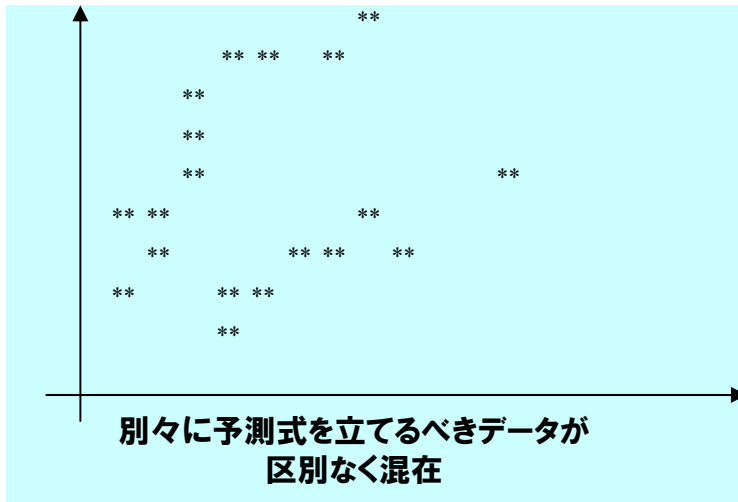


混合型データを自動分離分析する異種混合学習技術

- メリット
- ①データ自動分割：
雑然と収集された元データを分析用に自動分割。
 - ②混在パターンの分離抽出：
分割データをマイニングし、元データで混在していたパターンや規則性も分離抽出。
 - ③正確・スキルレス：
専門家による仮説設定に依存せず、正確な分析を実行。

混合型データを自動分離分析する異種混合学習技術

分けて分析すべきデータが混在している場合でも、人間が想定していない要因を考慮して、自動的に分離し機械学習を行う原理を実現



1)自動的にデータを分割

データ中のパターンや傾向、法則のマイニング精度が高くなるように、自動的に元データを分割。分割の個数や、何を手掛かりに分割すればよいか分からない場合でも、新情報量基準に基づいて、高速に適切な分割を実行。

2)高精度に複雑なパターンを学習

分割を適切に行うことにより、まったく異なるタイプのパターンや傾向であっても、高精度に学習。これにより、従来の技術では極めて分析が難しかった混合型データに対しても、効率的に高精度なデータマイニングが可能。

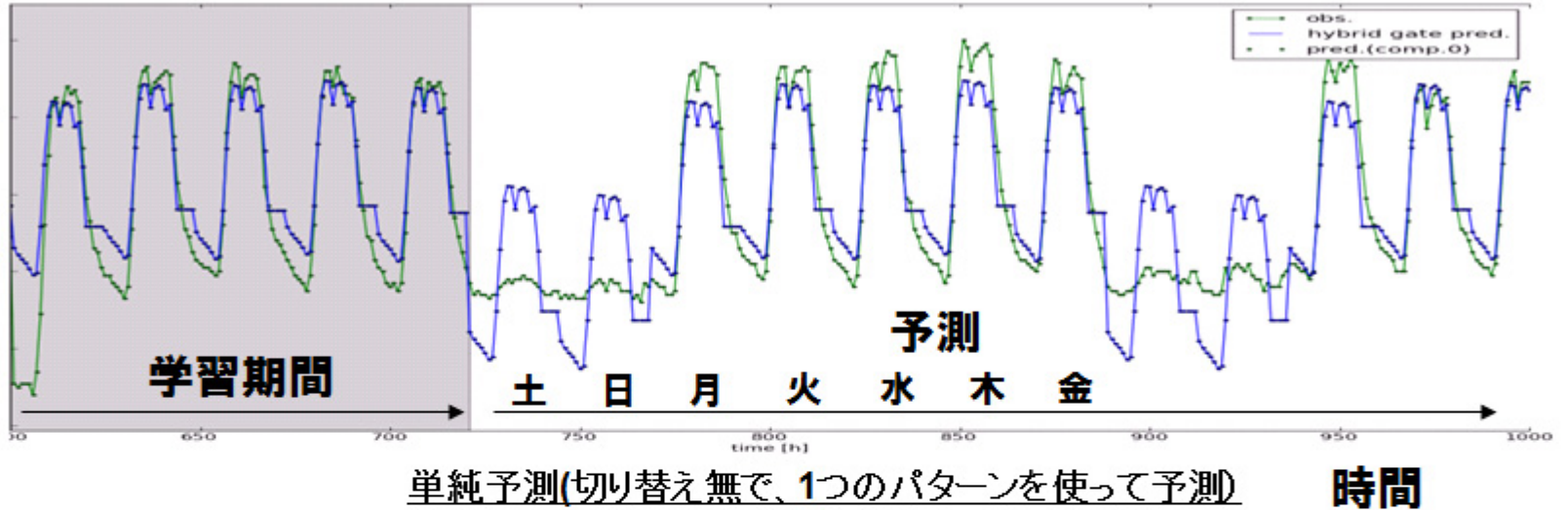
3)様々なタスクに適用可能

高精度に学習されたデータ中のパターンは、データの山に埋もれた隠れた知識の発見に役立つだけでなく、将来のデータ値の予測をする式として用いれば、高精度な予測システムとして利用可能。逆にパターンから外れるデータを抽出することで、高精度な異常検出システムとしても利用可能。さらに、見つけたパターンを制約条件として、その下での最適化問題を自動的に解くことなどにより、様々なシステムの最適制御にも用いることが可能。

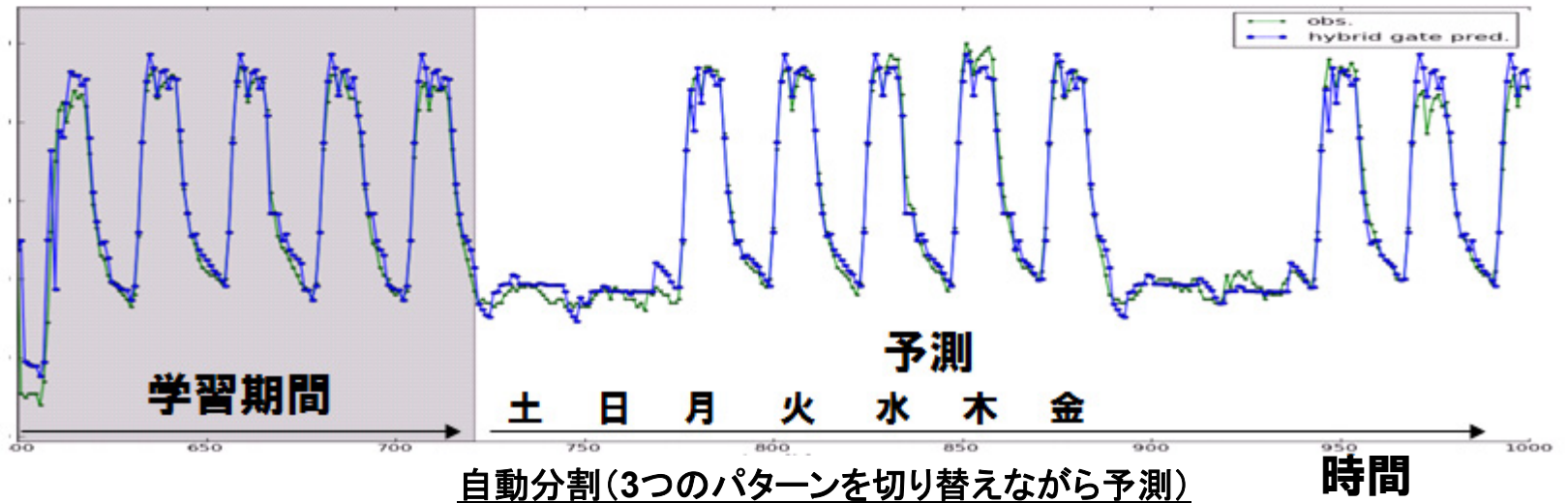
実データでの効果比較1(単純予測 vs 今回の技術)

ビル電力の予測……時間帯や外気温などの値と電力消費量の動き間の関係性パターンは様々な要因で切り替わっていく

電力消費

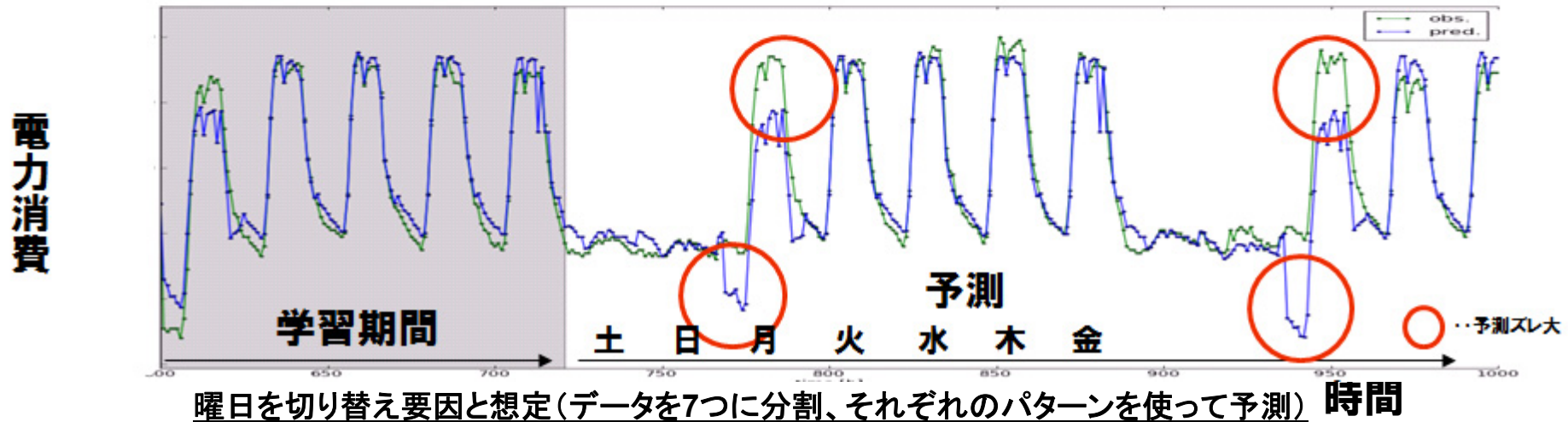
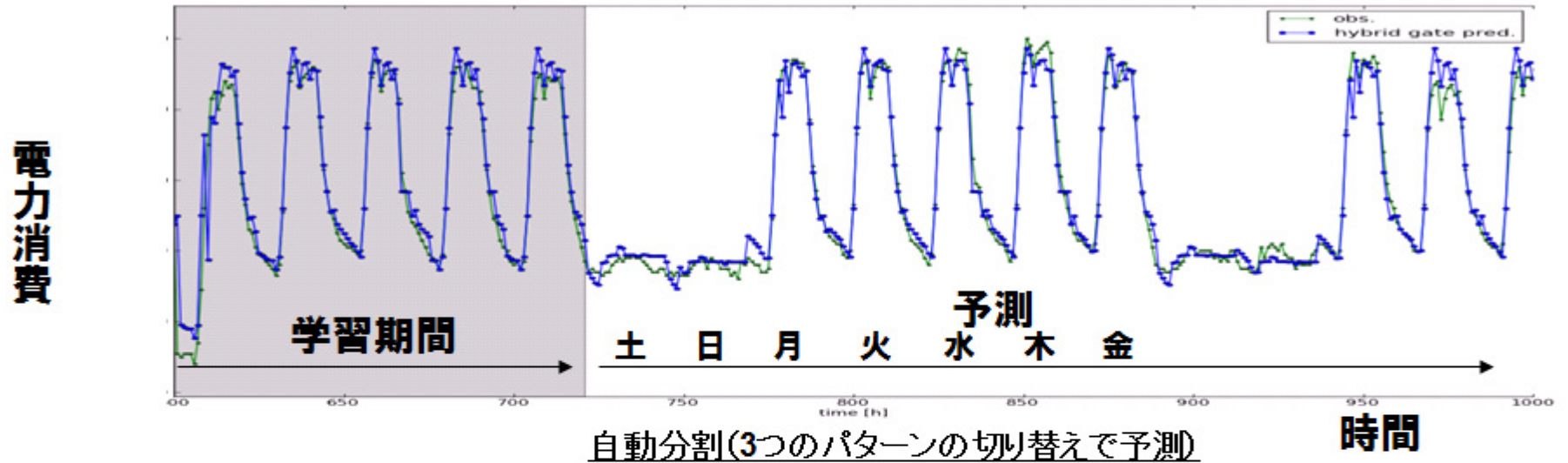


電力消費



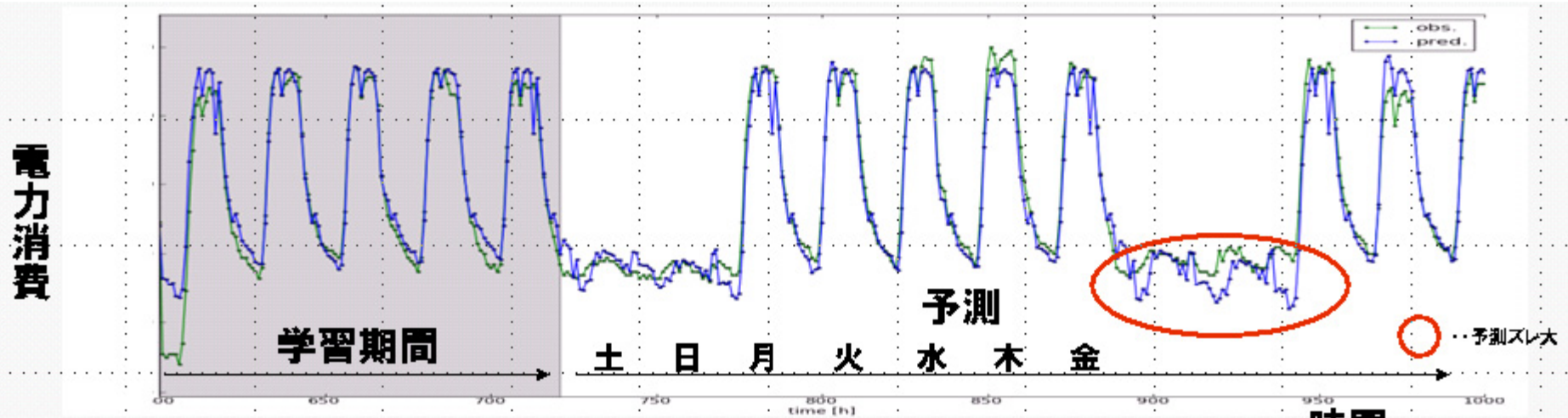
実データでの効果比較2(人が要因仮定 vs 今回の技術)

人間がパターン切り替え要因を想定して、データ分割・マイニングしてもうまくいかない

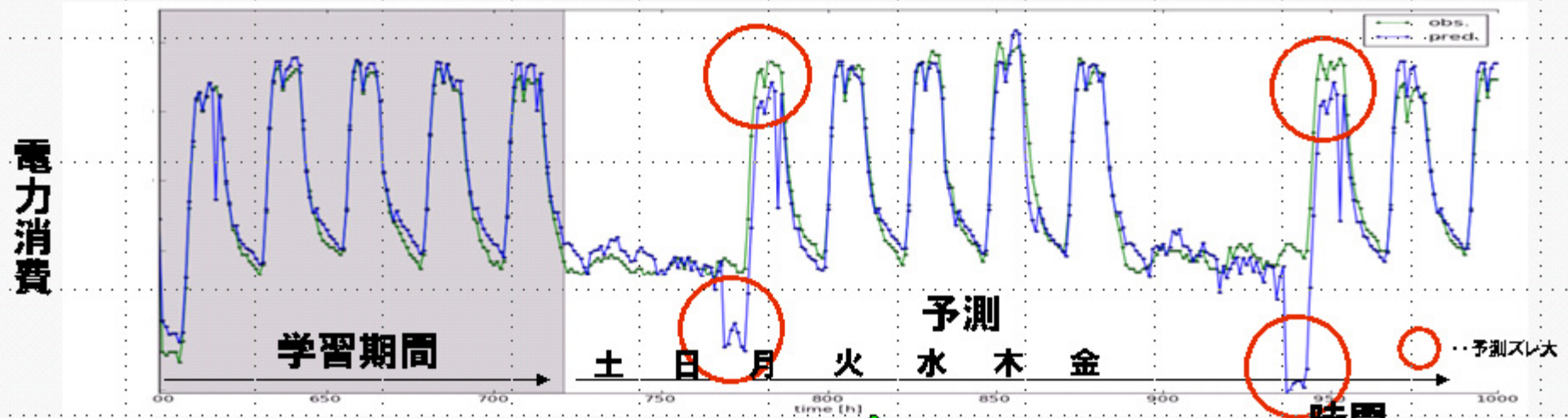


実データでの効果比較3(人が要因仮定 vs 今回の技術)

人間がパターン切り替え要因を想定して、データ分割・マイニングしてもうまくいかない



時間帯を切り替え要因と想定(データを24分割、それぞれのパターンで予測)



曜日×時間帯を要因と想定(データを168分割、それぞれのパターンで予測)

今後の想定適用例

1)ビルなどの電力需要予測

ビルの電力予測に関して、精度の高い需要予測を行うことで、単純なピークカットではなくよりインテリジェントな節電対策が可能。

2)ヘルスケア

多くの人々の通常時の身体活動データから色々なパターンを自動学習し、パターンから外れる異常状態を検出することで、精度の高い健康管理システムを実現。

3)農業

刻々と変化する空気中の成分濃度や土中の物質濃度などを予測し、最適な農作業の内容や時間帯の制御・推薦を行うことが可能。

*上記の例に限らず、これまでは整理・分析が極めて困難だった「雑然と収集されたビッグデータ」に対しても、混在するパターン・規則性を分離抽出し、予測や異常検出、制御等に有効活用することが可能。

Empowered by Innovation

NEC