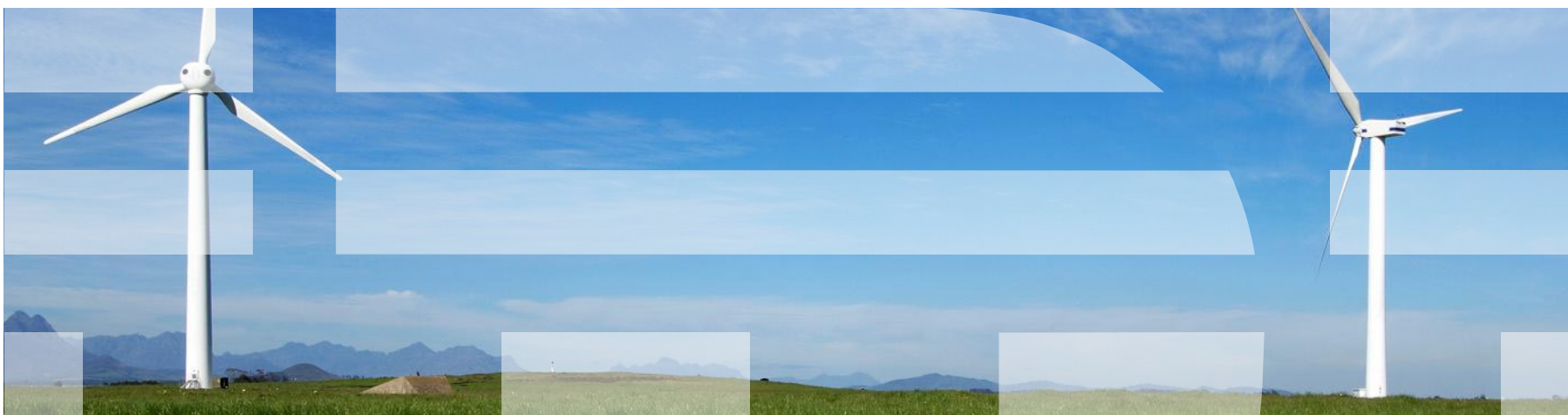


数理解析技術のビジネス応用

IBM東京基礎研究所 数理科学担当
担当部長 井手 剛

数理解析技術(analytics)は、クラウド事業などと並ぶIBMの最重点投資領域のひとつである。本発表では、数理解析技術についてのIBMの最近のビジネス戦略を概観し、IBM東京基礎研究所での最近の応用事例、フォーカス領域等について述べる。

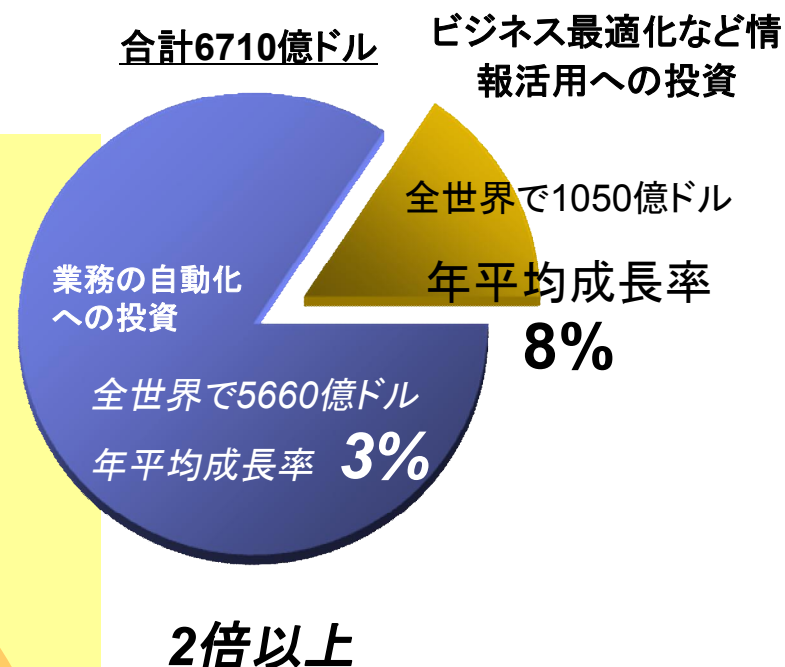
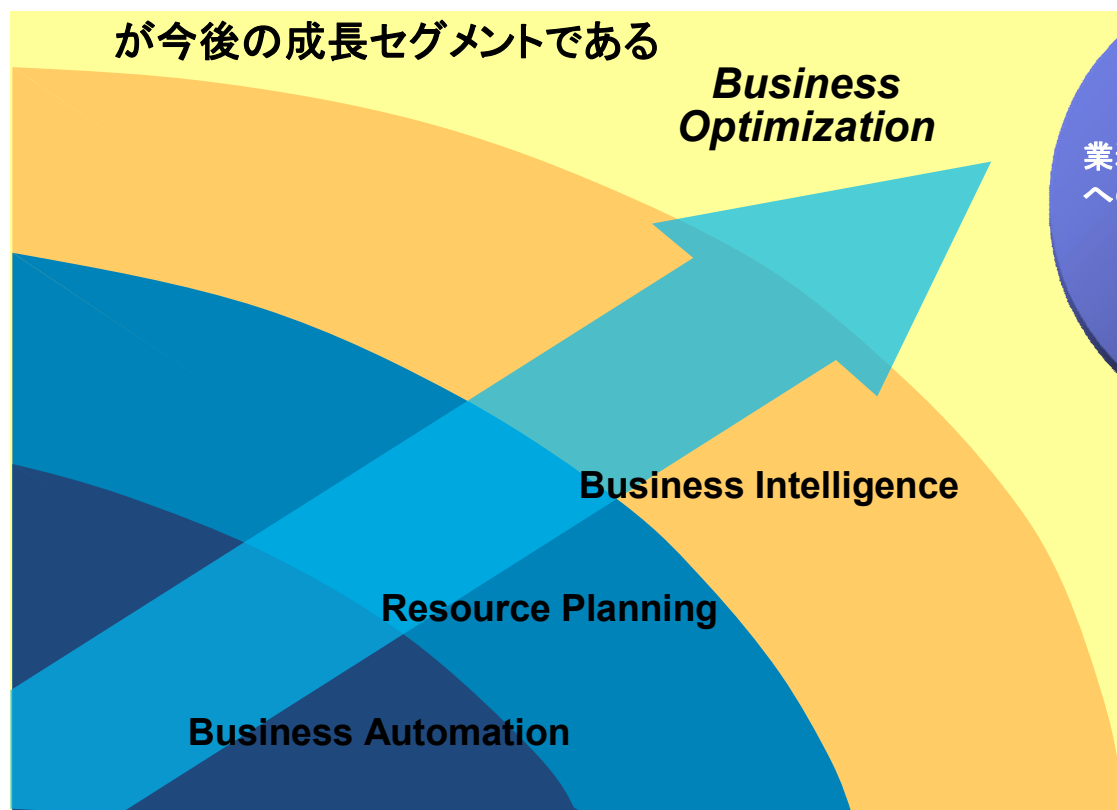


目次

- **イントロダクション: Smarter Planet と Analytics**
- 機械学習と新しいビジネス
- 私たちが解いている問題の例
- 最近私たちが注目している領域
- 技術紹介: 構造学習とその応用
- まとめ

データマイニングへの投資は、IT技術の成熟化に伴う歴史的必然です

- 比較的単純な業務の自動化を目的としたIT投資は漸減傾向
- 情報活用を目的とした積極的なIT投資が今後の成長セグメントである

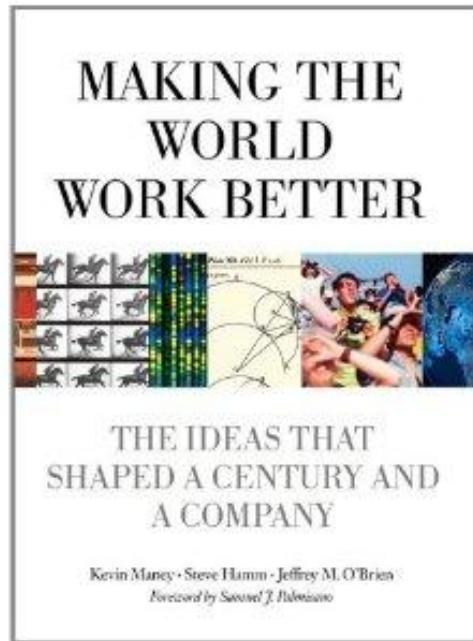


Smarter Planet とは、次世代社会基盤の構築をIT主導で進めるためのIBMの戦略的ビジョンです



- ITインフラの成熟は客観的に明らか
- 高付加価値のビジネスは、従来のIT基盤の上で行われる何らかのインテリジェントな処理
- IT基盤と伝統的な社会基盤を、IT主導で融合することで巨大なビジネスインパクトを目指す
- IT主導=Analytics主導

Smarter Planet のマニフェスト —— IBM創立100周年記念本



- Kevin Hamm, Steve O'Brien, and Jeffrey Maney, "Making the World Work Better: The Ideas That Shaped a Century and a Company," IBM Press, 2011.
- 書評: 知の専制、[「世界をより良いものへと変えていく」](#)

- いわゆる「IT革命」の次に来るべき社会変革のビジョンが Smarter Planet に他ならない
- そのためのコアが数理解析技術
- IBMはかつてはPhysical Layerの覇者であったが、時代と共に、ソフトウェア、サービス、と興味を中心を上位層に移してきた
- IBM Researchを活用した高度な解析技術を用いるサービスを強力に展開している

BAO (business analytics & optimization) は、数理解析技術をベースにしたサービスを行う専門のコンサルタント部隊です

- 誰でもできるような業務の自動化はすでにコモディティ化している
- 高度な解析技術に基づいて、ビジネスに有用な知見を与えることが求められている
- SPSS、CPLEX等のツールを元に解析技術のトレーニングを受けたコンサルタント部隊を育成
- 典型的な数理解析の問題はBAOで処理する
- 定式化されていない問題、困難な問題は研究部門と共同で解決に当たる

Japan [変更]

検索

ホーム ソリューション サービス 製品 サポート & ダウンロード My IBM

ようこそ [ログイン] [登録]

「BAO」が導く新たな知見。 膨大な情報を経営に生かす秘訣とは？

散在する多様な情報からビジネス・チャンスを発掘。
予測的分析に基づく業務最適化で、さらなる成長を牽引します。

はじめに BAOでビジネス・チャンスを ソリューション お客様導入事例

BAOソリューション体系
ビジネスの課題にきめ細かく対応させ、BAOサービスを利用しやすしました。
ソリューション体系図を表示する

情報活用診断
膨大なデータをもっと有効活用したいが、どこから手をつけたらいいのかわからないというお客様に最適です。情報活用に関するお客様の現状を調査・診断して、あるべき姿と現実のギャップを明らかにし、優先着手領域を特定します。
・情報活用の成熟度診断
・BIアプリケーション診断

クイック・アセスメント・サービス
「ソリューションの本格導入の前に効果を検証したい」というお客様のご要望に対して、短期間に効果を検証します。
・KPI定義による見える化構想支援
・テキストマイニングを活用したお客様の声活用
・異常値の効率的な検知
・統計アプローチに基づく在庫最適
・価格最適化

BAOソリューション

スマートな経営の可視化
・ビジネス・パフォーマンスを高める経営の可視化
・企業の活動評価指標の最適化

スマートな顧客分析
・お客様の声活用
・インターネット上の顧客行動の分析
・販促費用の最適化
・販売価格の最適化

スマートな生産と物流
・配送経路の最適化
・生産計画と製造ラインのスケジューリング
・動的在庫の最適化
・製造現場と経営をつなぐマネジメント・コックピット

スマートな品質管理・異常検知
・異常値の検知
・映像・画像の自動分類と検索

スマートなコンテンツ管理
・ワンソース・マルチユースを実現するドキュメント作成

フォームでお問い合わせ
まずはお気軽にご相談ください。
入力フォーム

IBM Smart Analytics System
スマートなビジネスを加速する分析アプリケーション
詳細はこちら

最先端のビジネス・アナリティクス

数理解析技術を経営判断に活用すること



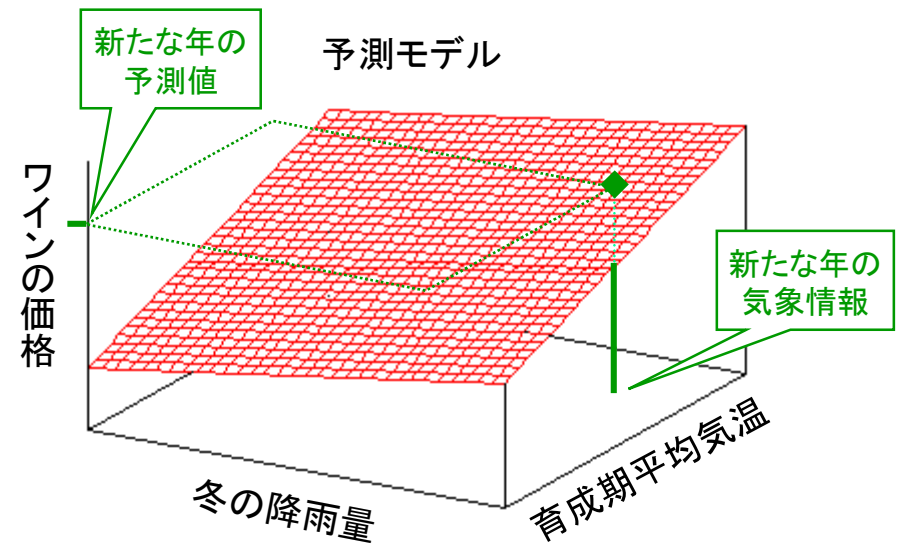
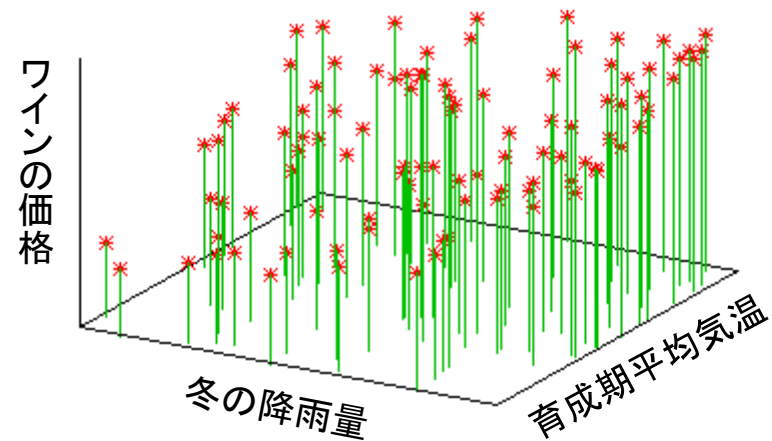
イアン・エアーズ著
文藝春秋 (2007)

- 将来のワイン価値を製造年の気象情報から予測するモデル

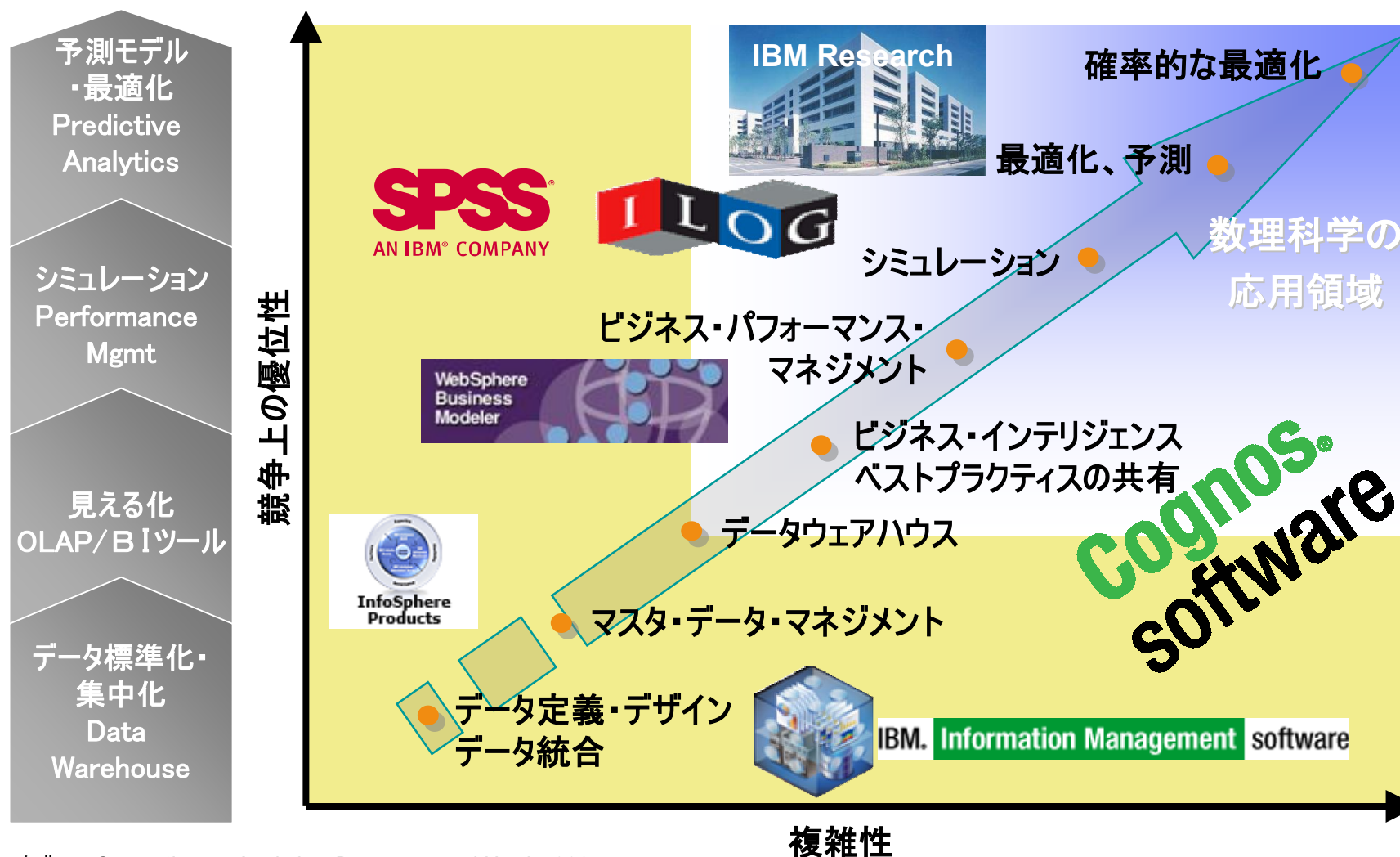
$$\text{ワインの質} = 12.145 + 0.00117 \times \text{冬の降雨量} + 0.0614 \times \text{育成期平均気温} - 0.00386 \times \text{収穫期降雨量}$$

→「1989年ものは今世紀最高のヴィンテージで、翌年1990年ものはそれ以上」といった予測が、製造時点で可能に。

回帰分析

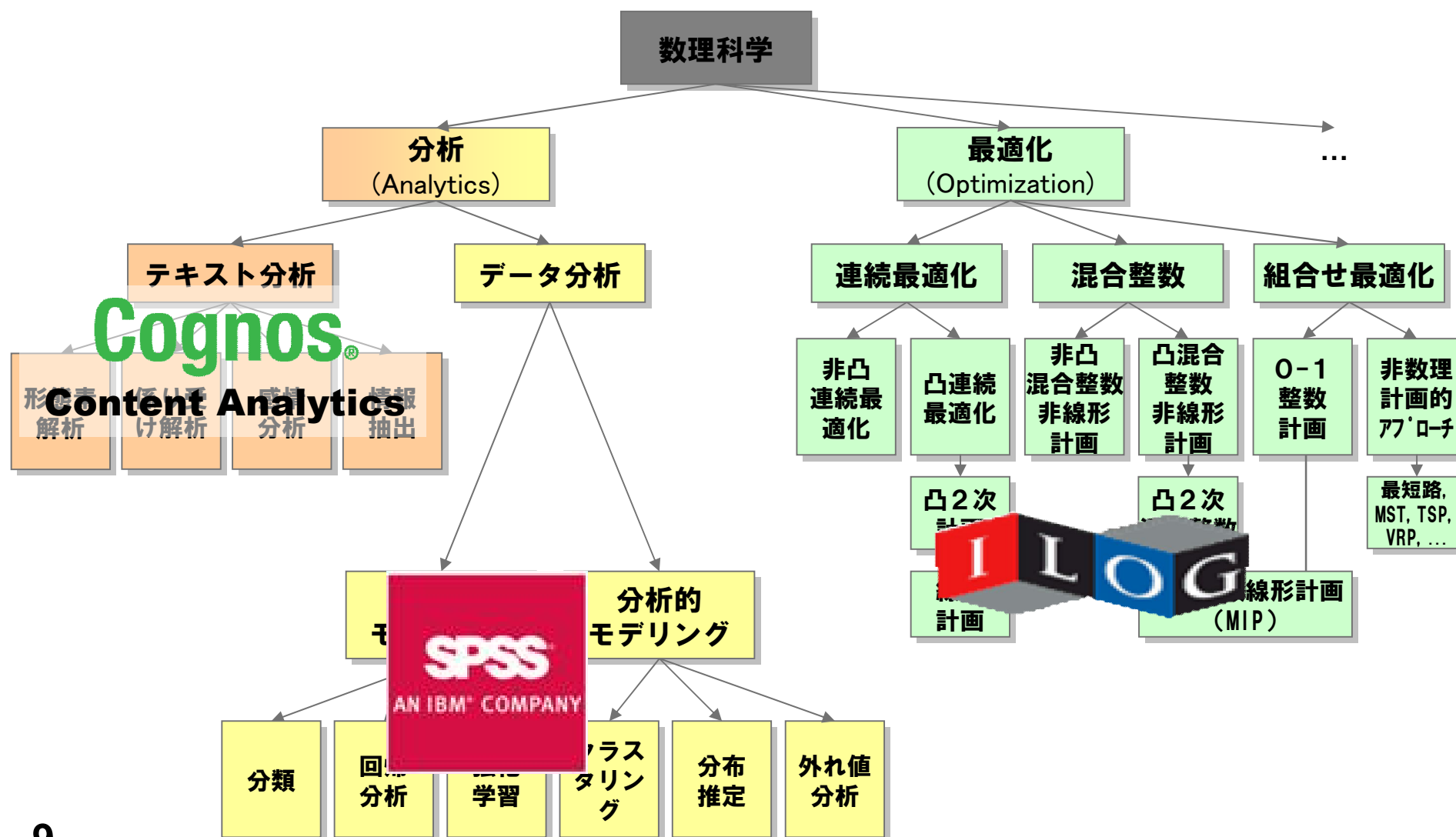


BAOサービスの実行のため、IBMは製品の品揃えを系統的に充実させています



出典： Competing on Analytics, Davenport and Harris, 2007

(ご参考) 数理解析技術に関するIBMのソフトウェア製品:
テキストマイニング、データマイニング、数理最適化



目次

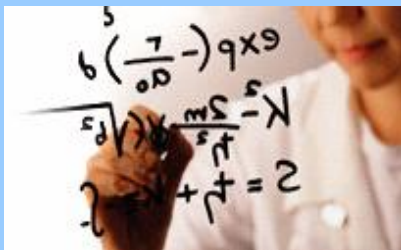
- イン트로ダクション: Smarter Planet と Analytics

- 機械学習と新しいビジネス

- 私たちが解いている問題の例
- 最近私たちが注目している領域
- 技術紹介: 構造学習とその応用
- まとめ

IBM基礎研究部門のストラテジー・エリア

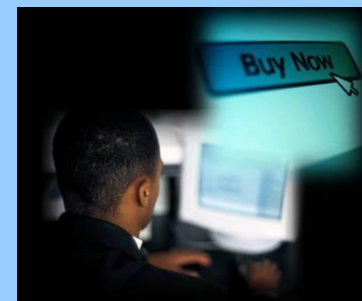
数理科学



インダストリー・ソリューション



サービス



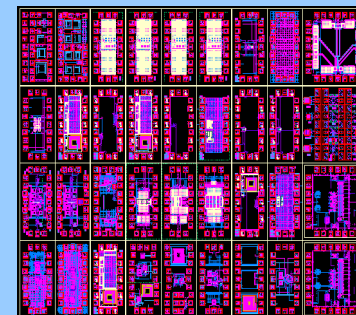
ソフトウェア



システム



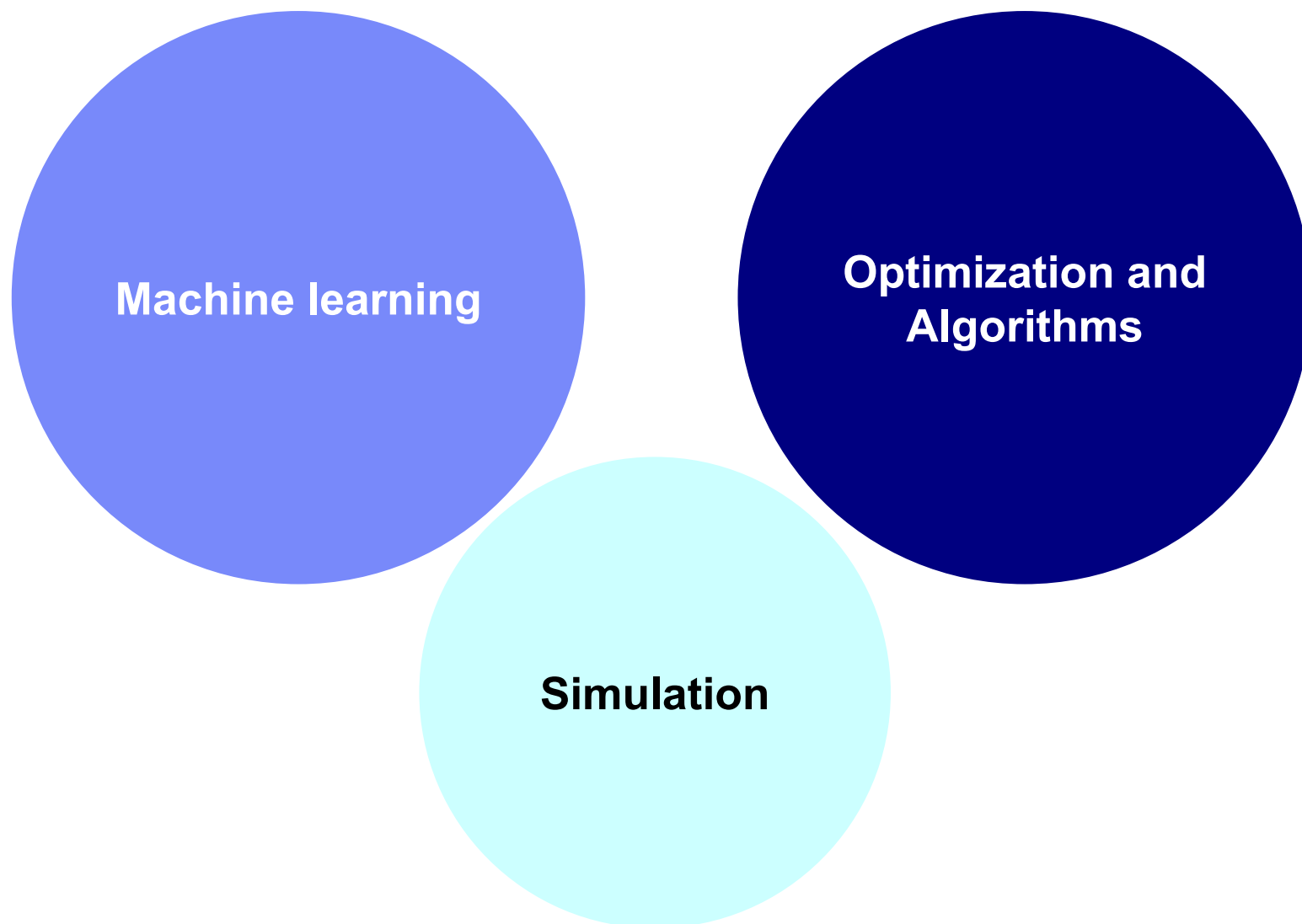
テクノロジー



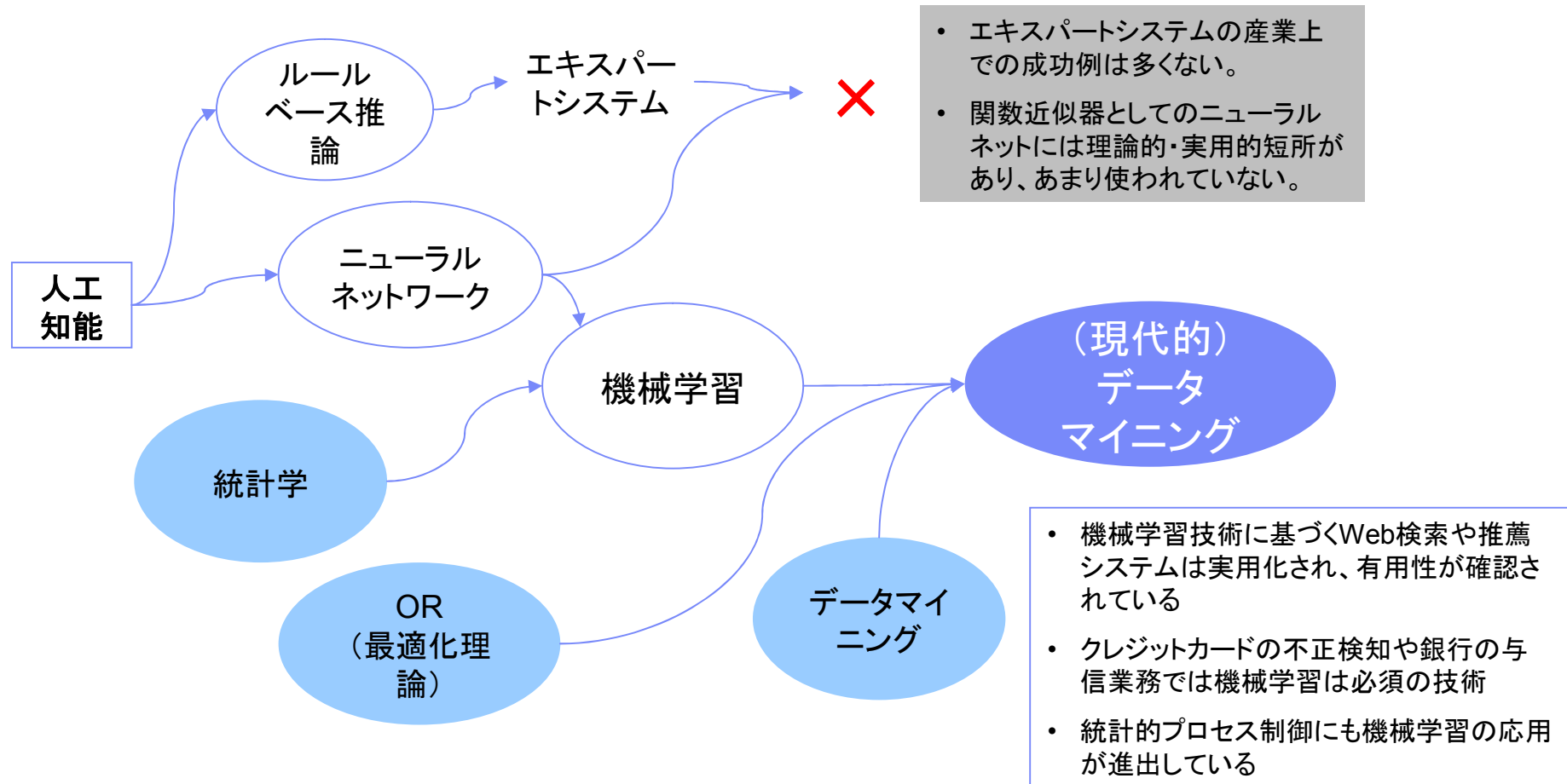
基礎科学研究



IBM東京基礎研究所の数理科学グループの専門分野



ここ10年の機械学習技術の急速な発展に伴い、その産業応用が進展しています



機械学習は新しい市場を創造しつつある

■ 新しいビジネスモデル

－ 書籍販売の推薦システム

- 商品数 × 顧客数 からなる巨大な行列の分解をすることで行える
 - ✓ 行列分解 = 低ランク近似
 - ✓ 少数の類型に商品と顧客を分ける
 - ✓ 特異値分解、非負行列分解など

－ 検索と広告ビジネス

- 引用、被引用を表すネットワークの最大固有ベクトルを求める

■ この例から分かること

- － 商品の推薦、というような、これまでは玄人の技だと思われていたものが自動化でできるようになってきた。つまり、Analytics技術が新しい市場を創造した
- － ビジネスモデルからしてBig Dataの処理が必然になってきた

(ご参考)現代の機械学習の2つのキーワード: 非線形性とスパース性

今日の後半部分での話題

非線形性

- 入力と出力の間の非線形的な関係を柔軟かつ簡単に取り込む
- 技術的キーワード
 - カーネル法

スパース性

- 入力と出力の間の関係を、より少ないパラメーターや変数で表す
- 技術的キーワード
 - L1正則化、Lasso
 - 関連度自動決定 (ARD: automated relevance determination)

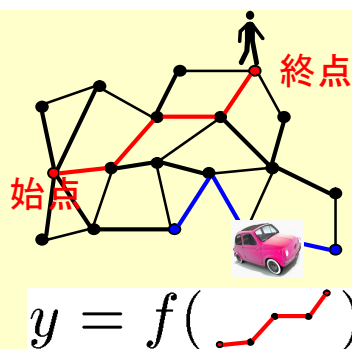
目次

- イン트로ダクション: Smarter Planet と Analytics
- 機械学習と新しいビジネス
- 私たちが解いている問題の例
- 最近私たちが注目している領域
- 技術紹介: 構造学習とその応用
- まとめ

私たちが解いている問題の例(1/3)

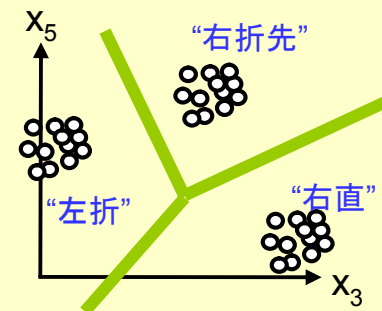
トラジェクトリ解析

- 自動車や人間の移動の軌跡を解析
- マーケティング等に有用な情報を抽出



自動車の事故様態判別

- センサーデータを基に、危険状態を自動分類
- より安全なナビゲーションを可能に

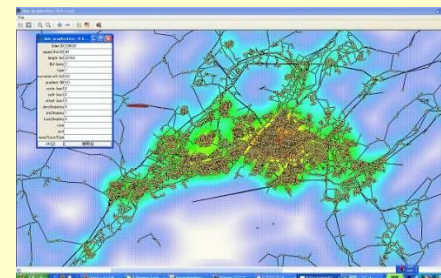


自動車の最適設計パラメーター決定支援

- 過去の数値実験の傾向から、車両強度の予測モデルを作成
- 次に調べるべき設計パラメーターを提示する

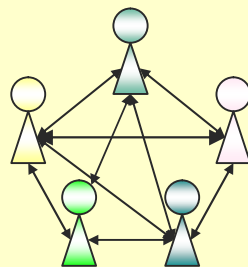
大規模交通シミュレーション

- IBMの並列化計算基盤を駆使した超大規模計算



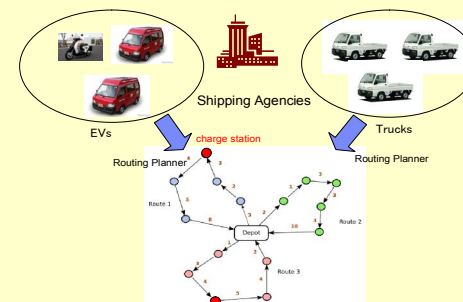
マーケットシミュレーション

- 広告の伝播プロセスをエージェントシミュレーション
- 広告戦略の可否を検証



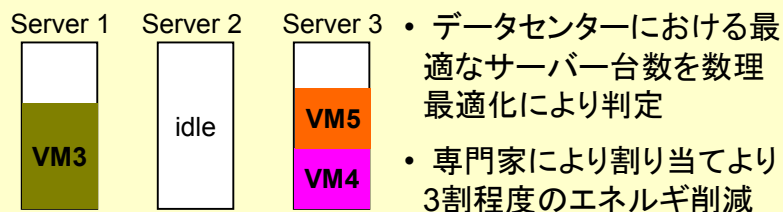
配送経路最適化

- 電気自動車のバッテリーに関する制約を正しく取り込んで配送経路を最適化

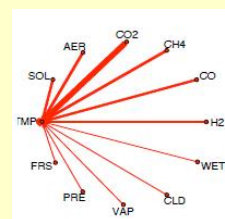


私たちが解いている問題の例(2/3)

事例：データセンターの省エネルギー



事例：地球温暖化の原因推定

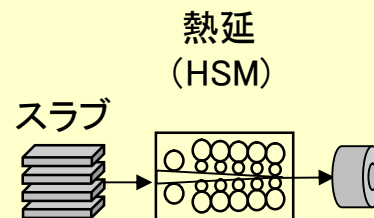


- 異常な温度上昇と、他の環境要因との依存関係を気象データから自動学習
- CO₂との因果関係が見出された

事例：プリンタの状態監視

- 感光体の劣化具合をセンサーデータから判定
- ジャムの前の正確な早期交換を可能に

事例：製鋼工程でのホットコイルの分類



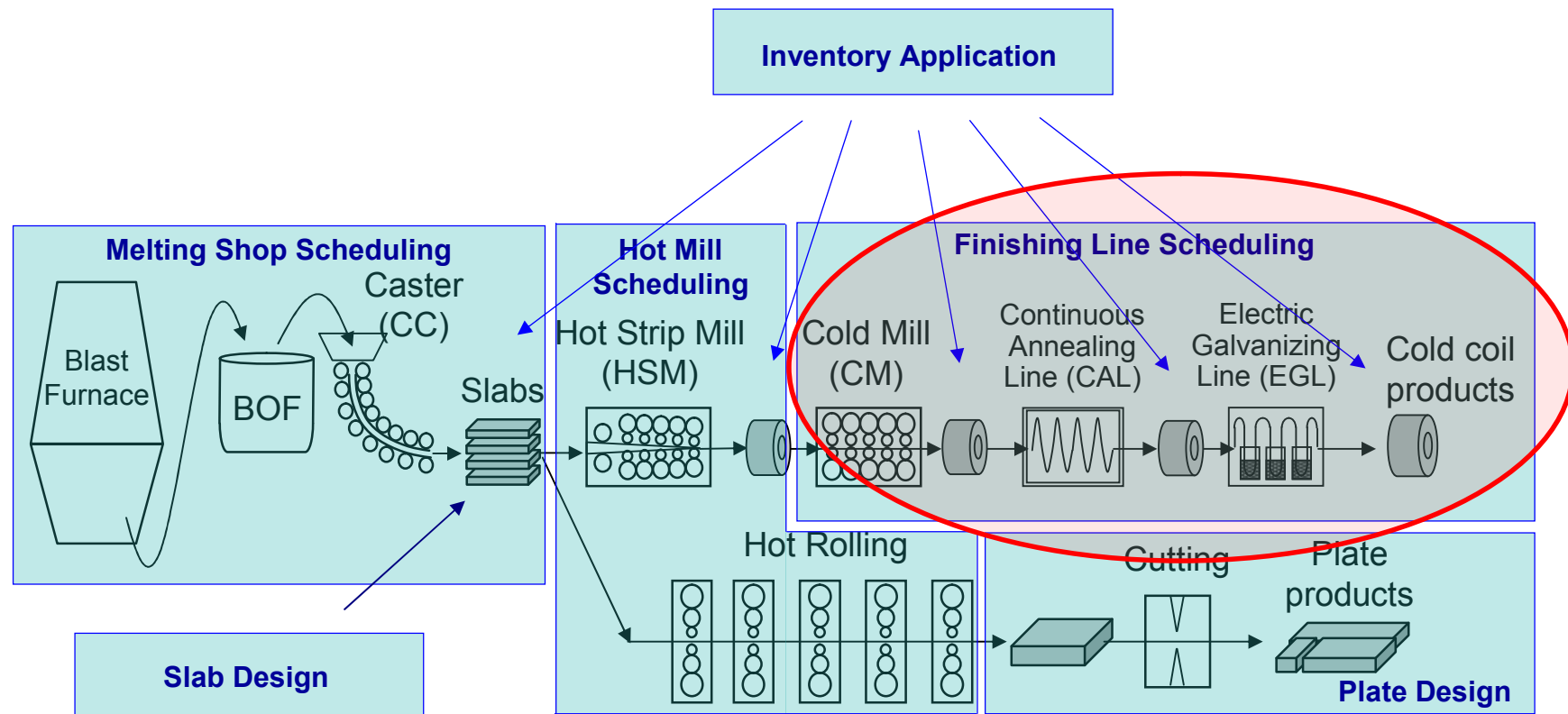
- ホットコイルの良・不良を自動判定
- 人の目で判別が困難な分類に成功

Deep QA – クイズ番組でチャンピオンを目指す

2009年、IBMは、「ジョパディー」というアメリカのクイズ番組に挑戦することを宣言しました。最先端の自然言語処理技術と高度な数理解析技術に基づいて、任意に与えられた質問文に高速に答えを返します



私たちが解いている問題の例(3/3): 鉄鋼生産最適化は、OR技術の産業応用の典型例です



目次

- イン트로ダクション: Smarter Planet と Analytics
- 機械学習と新しいビジネス
- 私たちが解いている問題の例
- 最近私たちが注目している領域
- 技術紹介: 構造学習とその応用
- まとめ

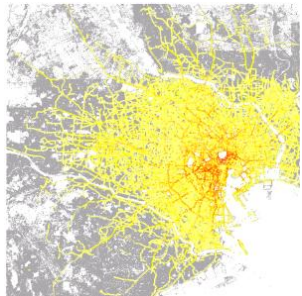
複雑に相互作用するシステムの最適化

- 人や物が複雑に相互作用するシステムを多様なシナリオにおいて評価するには、しばしばエージェント・シミュレーションが唯一の現実的な選択肢となります。シミュレーション結果はモデルのミクロなパラメータの値に強く依存しますが、ミクロなパラメータは直接観測不可能であることが多く、マクロな観測量からミクロなパラメータを推定する技術を実現しました。また、単にシミュレーションを異なる条件で反復するのではシステムを最適に制御するための条件を算出することは計算量的に困難であり、実用的な最適化のための高度な数理科学技術を研究しています。

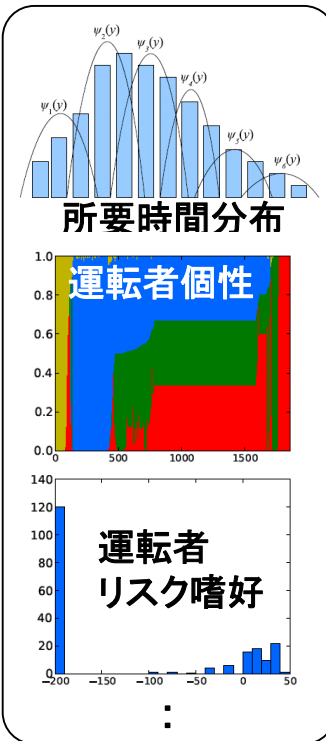
精確なシミュレーション・モデル構築

複雑に相互作用するシステムの最適化

プローブカーデータ

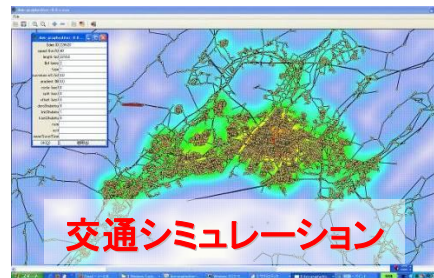


マクロ観測量から
ミクロパラメータ
推定



パラメータ

結果

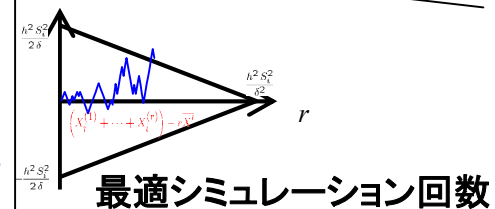


パラメータ
修正

制御条件

シミュレーション結果回帰

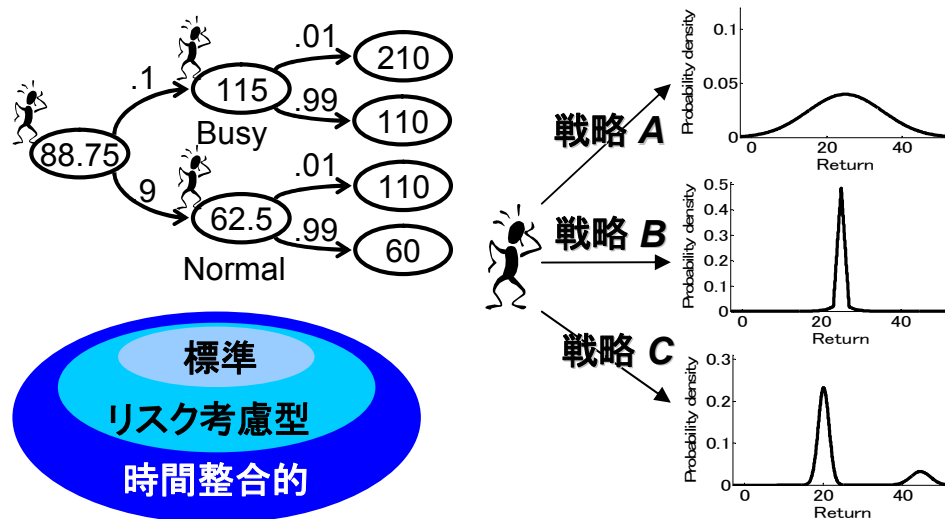
制御条件
最適化



意思決定の科学

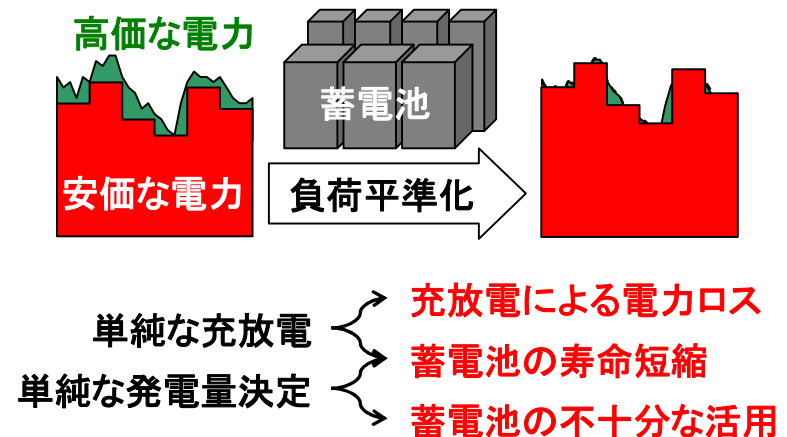
- 人の意思決定がしばしば合理的でないことが行動経済学や認知心理学によっても指摘されています。ビジネスにおいては不確実な情報に基づいて意思決定をする状況が多々ありますが、まさにそのような状況において、人は非合理的な意思決定をしてしまいます。ビジネスにおける意思決定を合理的かつ最適なものにするために、数理科学に基づく最適化技術を開発しました。また、人が介在するシステムの理解や最適化には、現実の人の意思決定を反映する行動モデルが必要となりますが、そのような行動モデルをデータから推定するための数理科学の手法も研究しています。

多様な合理的意思決定を支える マルコフ決定過程・強化学習

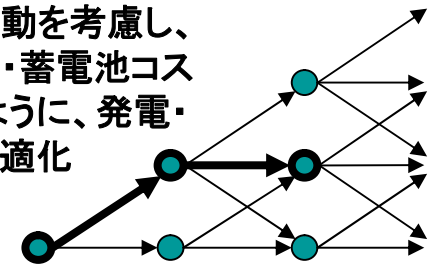


不確実な情報に基づく意思決定を最適化するマルコフ決定過程・強化学習技術において、考慮できる合理的なリスク嗜好が限られているという課題を解決しました

応用例: 発電・充放電計画の最適化



不確定な需要変動を考慮し、中長期的な電力・蓄電池コストを最小にするように、発電・充放電計画を最適化



センサーデータ解析

- 製造業などで取得されるセンサーデータは、強いノイズ、複雑な相関構造、など解析困難な特徴を持ちます。最新の機械学習の技術を用いてこれらの困難を突破し、実世界において価値を生み出します

設計パラメータ最適化

自動車組立工の作業
ミス監視

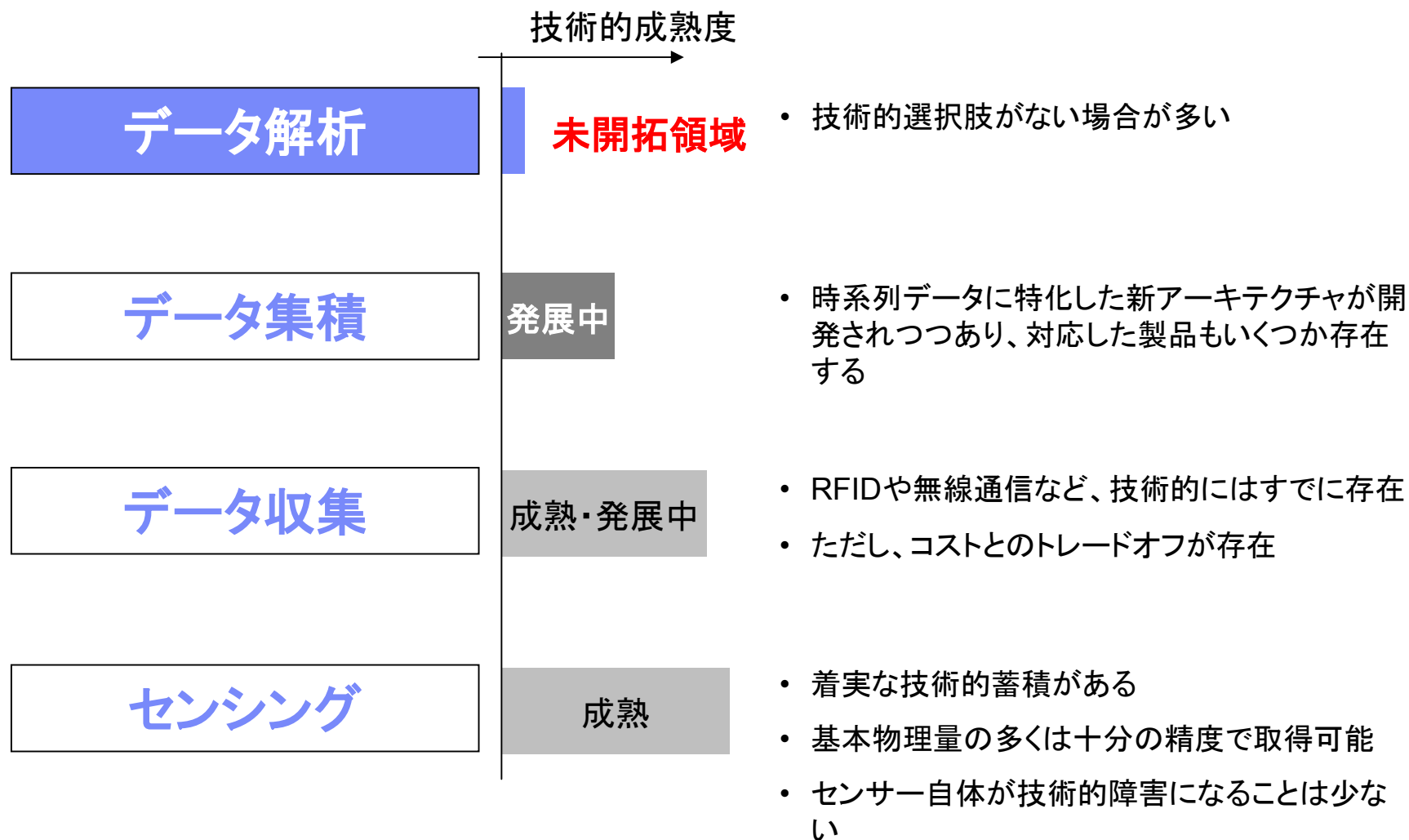
鉄道車両状態監視

自動車生産ラインの品質監視

電池劣化解析

発電設備状態監視

センサーデータの活用はまだ進んでいないのが現状です



なぜセンサーデータは活用されていないのでしょうか

巨大なデータサイズ

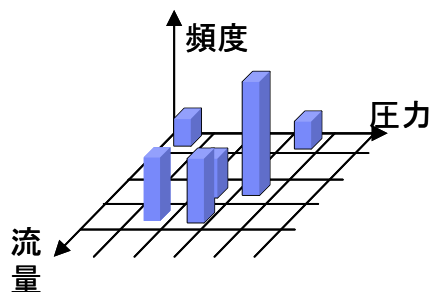
伝統的な関係DBは、物理センサーからの巨大な時系列データの扱いに最適化されてはいるわけではありません

数百のセンサーからの時系列データを数日間蓄積するだけで、そのサイズは容易にテラバイト程度となります。

クエリ処理技術の欠如

簡単な集計をしようと思っても、素朴な方法で作られたDBでは、処理に時間がかかりすぎるのが通例です

例えば、ある圧力と流量の範囲に入る頻度を知りたい、というような素朴なクエリであっても、素朴に設計されたDBでは、処理に相当の時間がかかってしまいます。



解析技術の欠如

売り上げデータなどと異なり、そもそもどういう「クエリ」をすれば不具合解析が可能なのか自明で花ありません

不具合解析のためには、素朴な集計を越えた高度な解析技術が必要です。現状ではそのような実用的機能を提供している時系列DB製品は存在しません。

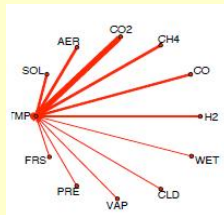
目次

- イン트로ダクション: Smarter Planet と Analytics
- 機械学習と新しいビジネス
- 私たちが解いている問題の例
- 最近私たちが注目している領域
- 技術紹介: 構造学習とその応用
- まとめ

変数間の隠れた関係を推定する技術は、ビジネスデータ解析において重要です

- これら二つの話題の裏には共通の機械学習の技術が使われています
- それは「多くの変数の間の隠れた関係を推定する」という技術です

地球温暖化の原因推定



- 異常な温度上昇と、他の環境要因との依存関係を気象データから自動学習
- CO₂との因果関係が見出された

“Spatial-temporal causal modeling for climate change attribution,” A. C. Lozano et al., Proc. KDD, 2009.

センサー系の自動検査システム

- 自動車設計開発時の測定系に自動異常検知システムを導入
- 人の目で判別が困難な異常の検知により、開発効率を向上

“Proximity-Based Anomaly Detection using Sparse Structure Learning,” T. Ide et al. Proc. SDM, 2009.

「構造学習」とは変数同士の隠れた関係を見出す技術です

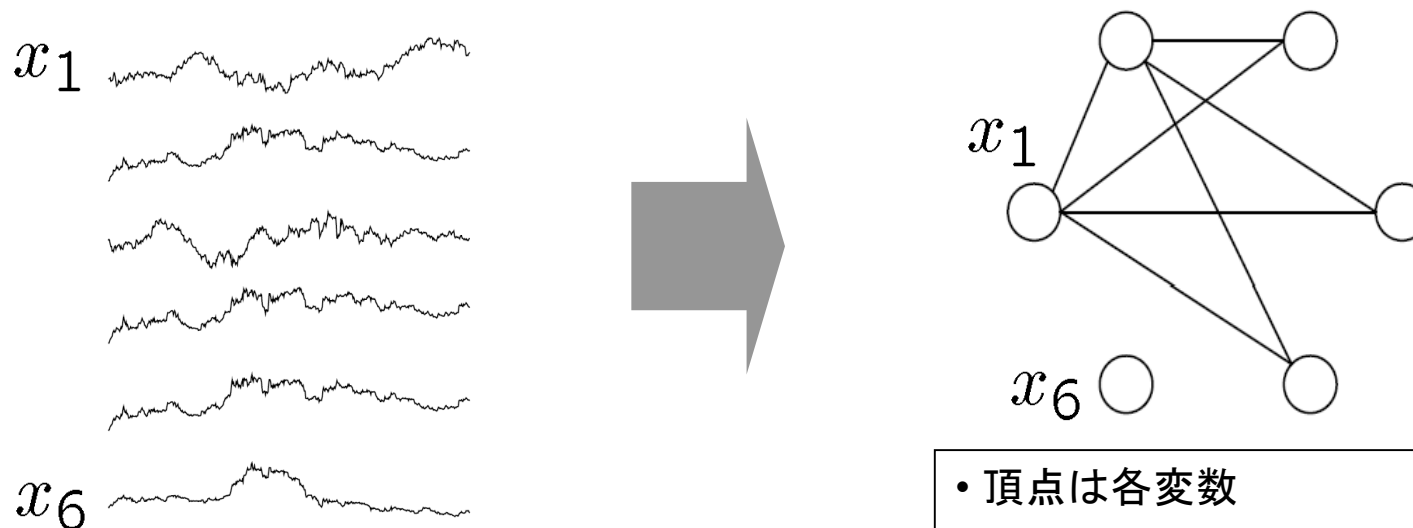
- 構造学習ができれば...
 - 異常気象に、どういう気象観測データが寄与しているか分かる
 - 売り上げ台数に、どういう経済指標／KPI／etc. が関係しているか分かる

- しかし一般に、変数同士の構造を推定するのは困難です
 - 多少のノイズにより結果が左右されがちです
 - c.f. 風と桶屋
 - 変数の数が多くなると構造推定が非常に難しくなります
 - 従来技術(「共分散構造解析」)は深刻な実用上の困難を抱えていました

IBM研究部門は、新しい「スパース構造学習」技術により、いくつかの実問題を解決しています

■ スパース構造学習

- 入力: 実数値の多次元データ
- 出力: つながりを表す重み付きグラフ
 - ノイズによる偽のつながりを排して、本質的な関係だけを表すものでなければならない



- 頂点は各変数
- 辺は変数間の関連
- 2つの頂点間に辺がない=他を与えた時に両者は独立

Graphical Gaussian Model (GGM) におけるグラフの定義: 「精度行列の行列要素がゼロなら辺なし」

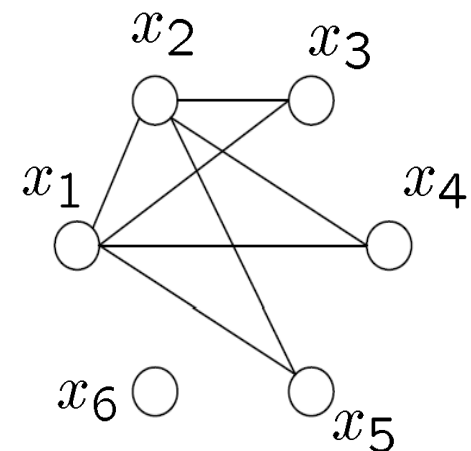
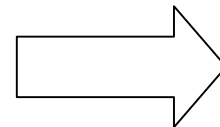
- 精度行列 Λ = 共分散行列 S の逆行列

- 例: $\Lambda_{1,2} = 0$ なら x_1 と x_2 は条件付き独立で、頂点1と2の間には辺はない

– なぜなら exp の部分が因子化されるから: $\mathcal{N}(x|0, \Lambda^{-1}) = \frac{\det(\Lambda)^{1/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}x^\top \Lambda x\right)$

- 例2: 6変数の場合の例

$$\Lambda = \begin{pmatrix} * & * & * & * & * & 0 \\ * & * & * & * & * & 0 \\ * & * & * & 0 & 0 & 0 \\ * & * & 0 & * & 0 & 0 \\ * & * & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 0 & * \end{pmatrix}$$



グラフィカル・ガウシアン・モデルについてのコメント

- 一般に、データは多変量ガウス分布にはまったく従わないが、異常検知の文脈ではGGMは非常に有用
 - 密度推定そのものが必要である場合は多くない。ゴールはたとえば異常検知など別の問題
 - 2次キュムラント以上の高次の統計量は安定した計算が困難で、その解釈も難しい
- GGMでは因果性が表現できない。因果グラフの学習は応用上も非常に重要であるが、ノイズにロバストな計算手法は今後の課題
- ダイナミックスの取り込みは非常に興味深い研究課題だが、模索中
 - c.f.
 - Y. Liu et al., “*Learning dynamic temporal graphs for oil-production equipment monitoring system*,” KDD 2009
 - 今回は、滑走窓の意味でのみ時間変動を考える

疎な精度行列を得るための手法 (2/2):
 L_1 正則化によりスパース性を得る手法が近年発展している。

Covariance selection

Doesn't work for rank-
deficient situations

Graphical Lasso [Friedman et al. 08]

- Can handle rank-deficient situations
- Stable even under collinearities

(Independent) Lasso

[Meinshausen &
Bühlmann 06]

- Can handle rank-deficient situations
- Quite instable under collinearities

GGMの構造学習のための手法: ラプラス事前分布を付したMAP推定を行い精度行列を求める

- 観測モデルが正規分布

$$p_G(\mathbf{x}|\Lambda) = \mathcal{N}(\mathbf{x}|\mathbf{0}, \Lambda^{-1}) = \frac{\det(\Lambda)^{1/2}}{(2\pi)^{M/2}} \exp\left(-\frac{1}{2}\mathbf{x}^\top \Lambda \mathbf{x}\right)$$

- 精度行列 Λ についての事前分布

$$p(\Lambda) = \prod_{i,j=1}^M \frac{\rho}{2} \exp\left(-\rho|\Lambda_{i,j}|\right)$$

- MAP (Maximum A Posteriori) 推定で求める

$$\Lambda^* = \arg \max_{\Lambda} \left\{ \ln p(\Lambda) \prod_{n=1}^N p_G(\mathbf{x}^{(n)}|\Lambda) \right\}$$

GGMの構造学習のための手法: MAP方程式は、 L_1 正規化項付きの最適化問題に帰着される

- 入力: 共分散行列 S

- 平均ゼロ、分散1に標準化したデータが前提

- 普通、ランク落ちしているので逆は存在せず

$$S_{i,j} \equiv \frac{1}{N} \sum_{n=1}^N x_i^{(n)} x_j^{(n)}$$

- 出力: スパースな精度行列 Λ

- 精度行列 = 共分散行列の逆行列

- 方法: L_1 正規化項付きの最尤方程式を解く

$$\Lambda^* = \arg \max_{\Lambda} \left\{ \underbrace{\ln \det \Lambda - \text{tr}(S\Lambda)}_{\text{対数尤度}} - \underbrace{\rho \|\Lambda\|_1}_{\text{正則化項}} \right\}$$

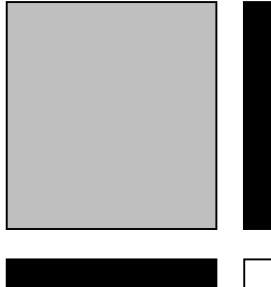
$$\text{対数尤度} \ln \prod_{t=1}^N \mathcal{N}(x^{(t)} | 0, \Lambda^{-1})$$

正則化項

Graphical Lasso algorithm: 各列に着目して、行列についての最適化問題をベクトルに対する問題に直す(ブロック勾配法)

▪ 精度行列を1列(1行)ずつ最適化

– 灰色部分を定数だと思って、青色部分についての最適化問題を導く

$$\Lambda = \begin{pmatrix} L & l \\ l^\top & \lambda \end{pmatrix}$$


– 青色ベクトルについての最適化問題は、 L_1 正則化項付きの2次計画問題になる

• 劣勾配法により効率のよい固定点方程式を導ける(Friedman et al. 2008)

▪ スパースな精度行列を、明示的な逆行列計算なしに求めることができる

– 副産物として、精度行列の逆も(逆行列計算なしに)求まる

• 標本共分散行列Sの修正版のようなもの

(詳しくは: T. Idé et al., “Proximity-Based Anomaly Detection using Sparse Structure Learning,” SDM 2009)

Graphical Lasso algorithm: 各列についての最適化問題は、 いわゆるLassoと同等になる

- 着目する変数が一番最後の列に来るように変数を並びかえる

$$\Lambda = \begin{pmatrix} L & l \\ l^\top & \lambda \end{pmatrix} \quad \Sigma \equiv \Lambda^{-1} = \begin{pmatrix} W & w \\ w^\top & \sigma \end{pmatrix} \quad S = \begin{pmatrix} S^{\setminus i} & s \\ s^\top & s_{i,i} \end{pmatrix}$$

- 素朴に行列をばらしてゆくことにより(予稿参照)、ベクトル $\beta \equiv W^{-1}w$ についての次の最適条件が導かれる

$$\min_{\beta} \left\{ \frac{1}{2} \|W^{\frac{1}{2}}\beta - b\|^2 + \rho \|\beta\|_1 \right\} = 0$$

$$b \equiv W^{-1/2}s$$

L_1 正則化項付きの
線形回帰と同等＝
Lasso

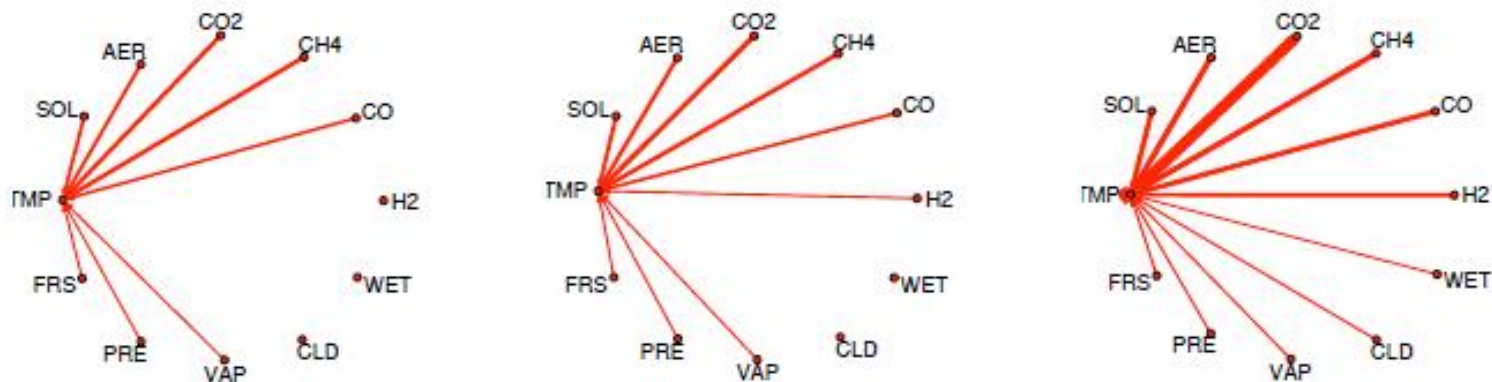
- 結局、精度行列のひとつの列は次のように求まる

$$\lambda = \frac{1}{\sigma - \beta^\top W \beta}, \quad l = -\frac{\beta}{\sigma - \beta^\top W \beta}$$

収束するまですべての
変数について計算
を繰り返す

応用例1: 地球温暖化の原因推定

- 異常高温の発生と、地表面で観測された様々な気象データがどのように影響しているかを調べた
 - TMP: 過去の異常な温度上昇の発生の起こりやすさを表す変数
 - CO2: 二酸化炭素濃度
 - CH4: メタンガス濃度、など
- その結果、CO2がもっともTMPに影響を与えていることが見出された



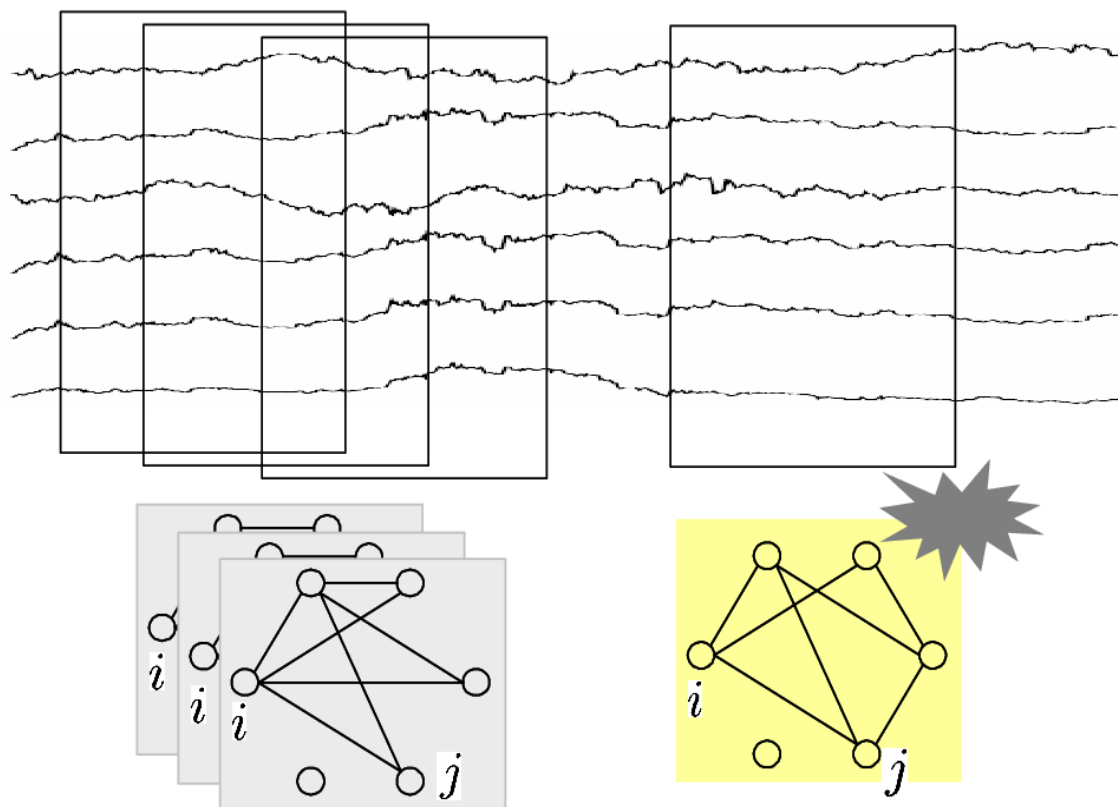
高い
(疎)

因果グラフのスパースさの度合い

低い
(密)

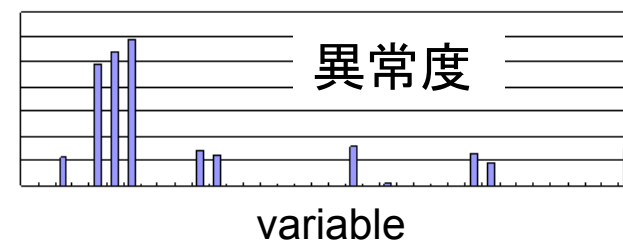
応用例2: センサー系の検証(1/2)

変数同士の依存関係の重大な変化を異常として検知します

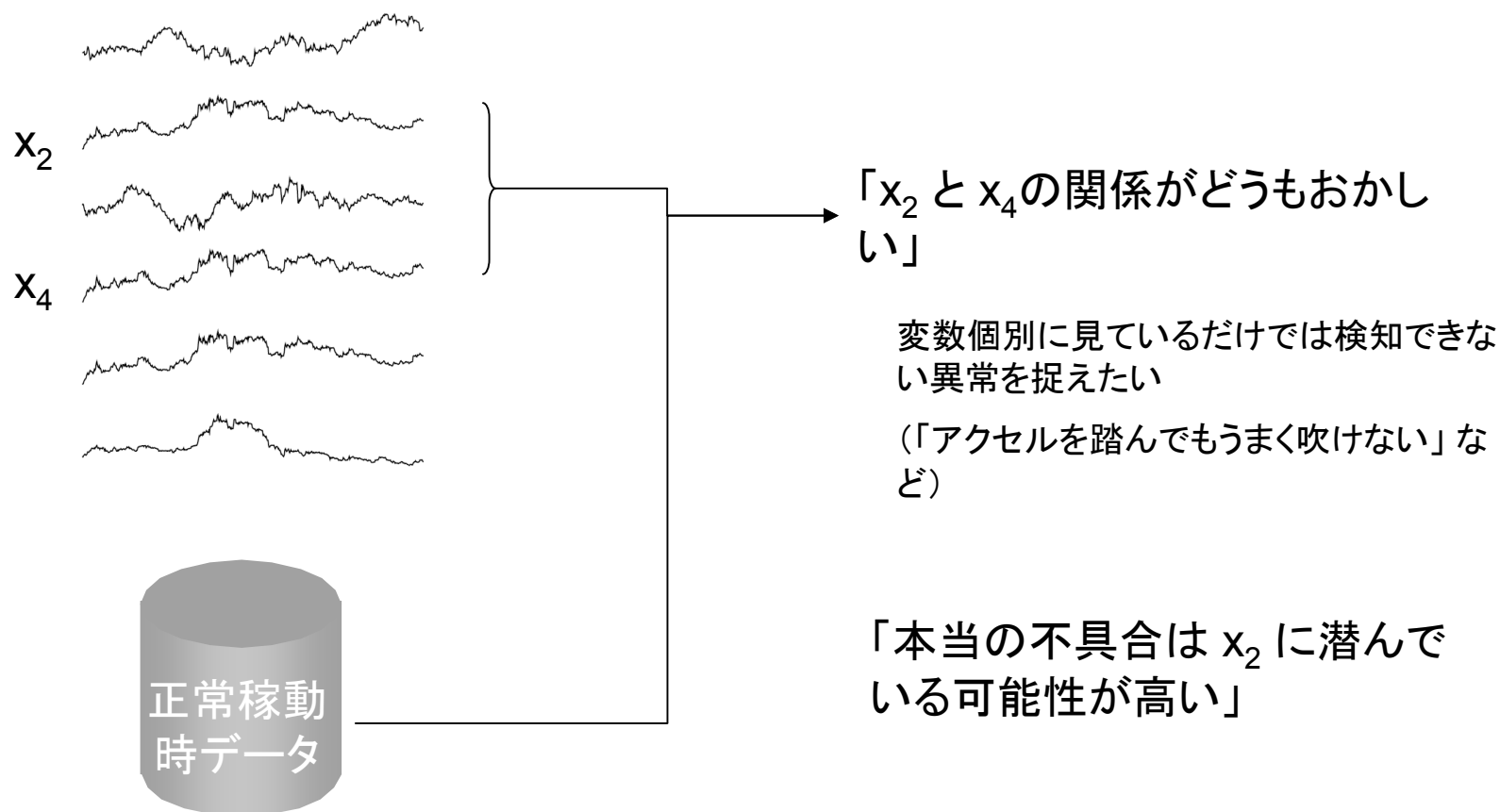


(1) スパース構造学習

(2) 相関異常度の
スコアリング



(参考) 変数同士の「関係の崩れ」を検出したい。正常時のデータを元にして、個々の変数の相関異常度を計算したい



応用例2: センサー系の検証(2/2)

熟練者にも発見が難しい異常を明瞭に検出しました

- 実際の機械系の実際のセンサーエラーに対して異常度を計算した

- 主たるエラーは配線ミス

- 人の目では見つけるのが困難
- センサー自体は生きているため

配線ミスが含まれる
データの一部

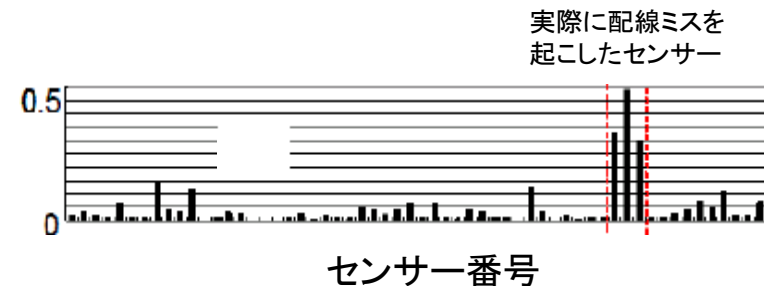
正常時のデータ
の一部

- 右図: 異常度の計算例

- 加速度センサーの3軸が入れ替わっている

- 人間の目では検知が難しかった例
- 非常に明瞭に異常センサーを特定している

配線ミス(人間の目ではほぼわからない)



目次

- イン트로ダクション: Smarter Planet と Analytics
- 機械学習と新しいビジネス
- 私たちが解いている問題の例
- 最近私たちが注目している領域
- 技術紹介: 構造学習とその応用
- まとめ

まとめ

- 数理解析技術は次世代の社会基盤の中心を担うであろう最重要技術
- IBMは数理解析技術に真剣に取り組んでいる
- 機械学習の技術はここ10年で大幅に進歩した。解析技術が市場を牽引する分野すらある
- サービス工学のためには、複雑系の解析、人間系の解析が必須。IBM東京基礎研究所の数理解析チームは、それらの領域にフォーカスしている
- 構造学習は、機械学習の進歩を象徴する「スパース化」の技法により最近大幅に進歩を見た。産業応用が進んでいる

どうもありがとうございました