

Computational Challenges of Massive Data Sets キーワード集

boundary condition	【境界条件】シミュレーションモデルが表現する対象の境界領域の諸量の値。
data assimilation:	【データ同化】シミュレーションなどの数値モデルによる対象状態の時間発展更新と、装置からの部分的な観測量に基づく状態補正の二つを適切に組み合わせる作業。
data mining:	【データマイニング】大量データからの有用情報の発見を目的とする研究分野。データに内在する妥当で新規性があり、有用となる可能性がある理解可能なパターンの工夫した特定。
data set:	【データ集合】例集合として表されるデータ。例えば、10個の検査値で表された患者100人分のデータ。
dependent variables:	【従属変数】時間的にも、空間的にも相互に関連している変数。正確に言えば、どちらの場合もありえる。どちらにも関連していない変数どうしは、独立であるという。
dynamic model:	【動的モデル】システムが時間とともに変容するモデル。dynamical model は力学モデルで、違う概念。
evaluation criteria:	【評価規準】パターンの良さを評価する方法。
example:	【例】値のベクトルであり1つの対象物を表す。事例、タプル、レコードともいう。
ground-based measurement system:	【地上観測網】
initial condition	【初期条件】シミュレーション変数の初期値。
interestingness:	【興味深さ】人間が抱く興味深さに関する評価指標。
Kalman filter:	【カルマンフィルタ】離散時刻の場合を考える。時刻 $t-1$ までの過去全ての観測データにもとづく状態ベクトルの確率分布と、新たに時刻 t のデータを観測したもとの状態ベクトルの確率分布の間の関係を司る漸化式。線形ガウス状態空間モデルの場合はすべての分布がガウス分布になるので、漸化式は具体的には状態ベクトルの平均値ベクトルと、分散共分散行列の時間更新の漸化式となる。カルマンが定式化したのでこの名前がついている。
large-scale measurement system:	複数の人工衛星や、地上の多数の観測網(ネットワーク)などによる全地球の環境をモニターするシステム。あるいは、特定の興味(観測)対象を組織的に同時複数量で計測しデータを蓄積獲得するシステムのこと。
leverage value of information:	【情報価値の向上】スピーカーの造語であり直感的な意味で用いている。例えば、患者データを病気診断手続きに変換することは通常、情報価値を向上することになる。
linear Gaussian state space model:	【線形ガウス状態空間モデル】通常、状態空間モデルと呼ばれる。システムモデル、観測モデルがともに線形かつ、システムモデルにおける擾乱項であるシステムノイズおよび観測モデルのノイズである観測ノイズがともにガウス(正規)分布に従うと仮定するモデルのこと。
machine learning:	【機械学習】学習という行為を通してその性能を向上するソフトウェアに関する研究分野。
Monte Carlo mixture Kalman filter:	【モンテカルロ混合カルマンフィルタ】複数のカルマンフィルタによる状態更新を、確率的に切り替え統合するような、統合化された状態更新アルゴリズム。
non-Gaussian disturbance behavior:	【非ガウスの擾乱】ガウス分布でない分布、つまり非ガウス分布から発生したと仮定する(想像できる)擾乱項。ガウス分布は、標準偏差の3倍、4倍ともなると、その実現値の現出は極めてまれで、膨大なサンプル数でもほとんどゼロに近いが、一方非ガウス分布はある一定の有限値をもつ。この性質を使って、たまにおこる現象の現出を形式的にでも確率モデルとして表すときに、非ガウスノイズが利用される。近年のファイナンス工学でもこの性質を使ってさまざまモデルが提案されている。
numerical simulation model:	通常シミュレーションと呼ばれる、時刻 $t-1$ から時刻 t への計算機上での状態更新を行う実体形式。
observation model:	【観測モデル】measurement model とも呼ばれる。状態ベクトルから、具体的データの観測にいたる道筋、状態ベクトルとデータベクトルとの関係性を表現する関数。
oceanography:	【海洋学】
pattern:	【パターン】自由変数を含む一般的な表記(statement)。

rule:	【ルール】前提部が成立すれば結論部が成立するという形式のパターン。前提部と結論部は、例に関する命題で表される。論理学のルールとは異なり、当てはまらない例があってもよい。例えば、検査値Aが30以上であれば診断Bは陽性など。
search:	【探索】複数個の候補を調べることによる問題解決法。この場合、可能なルールペアを調べ、有望そうなものを表示すること。
simultaneous estimation:	【同時推定】データに基づいて未知母集団に関する複数パラメタの値を定めること。
sound pruning:	【健全な枝刈り】探索において、計算結果を変えことなく候補の一部を調べずに済ますこと。
state vector	【状態ベクトル】システムの状態を記述する変数で構成されるベクトル量。データ同化の場合は、シミュレーション変数及び諸パラメータなどをすべて並べたベクトル量。
stochastic variable:	【確率変数】時間、空間的状态発展(更新)に、確率的な要因を含む変数。
strong nonlinear:	【強非線形】システム、観測モデルが非線形の場合、それらを線形化近似し、カルマンフィルタの考えを適用する状態更新の計算手法が、拡張カルマンフィルタと呼ばれる。このような、モデルをあらかじめ局所線形化するような近似解法が役に立たないほど非線形性が強いこと。
strong pattern:	【強いパターン】比較的多くの例について正確に成立するパターン。
subjective evaluation:	【主観的評価】データ以外の情報を用いて評価すること。領域専門家が、その領域における知識も用いて評価する場合などが当てはまる。
system model:	【システムモデル】update model と呼ばれることもある。状態ベクトルの、その時間更新を表現する実体形式。
weak pattern:	【弱いパターン】比較的少ない例について(不正確に)成立するパターン。