手軽に実現！！ビッグデータ分析

~新製品Big Data Module~

(株)数理システム
bigdata-info@msi.co.jp

MATHEMATICAL SYSTEMS INC.

目次

MATHEMATICAL SYSTEMS INC.

ー デモンストレーション ー

ECサイトのログを例に

- インポート
- 集計
- オンラインロジスティック回帰
- VMSアイコンへの接続

MATHEMATICAL SYSTEMS INC.

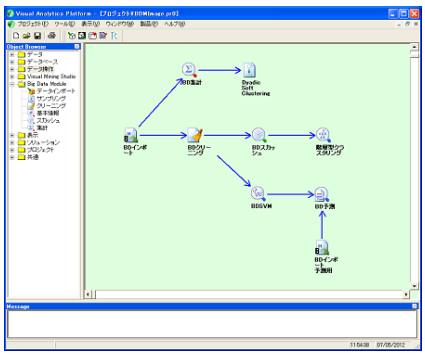
ー 製品紹介 ー

MATHEMATICAL SYSTEMS INC.

製品紹介

Big Data Moduleとは

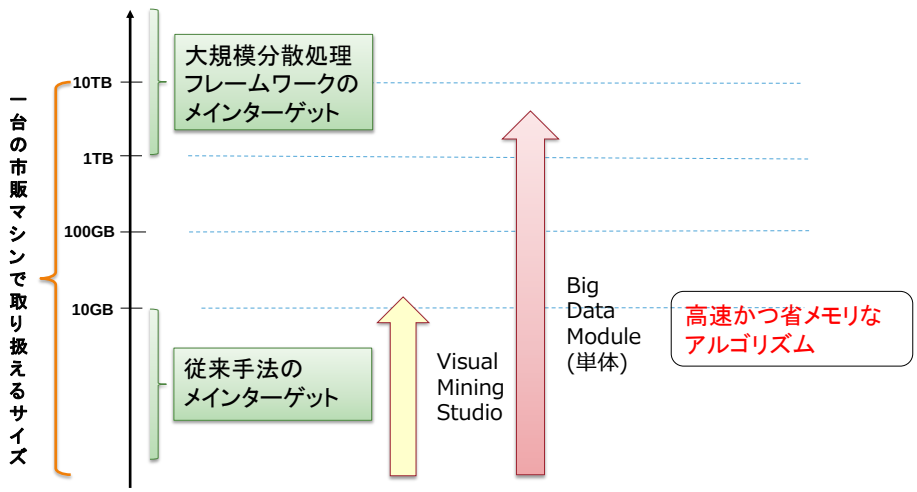
- 数理システム自社開発
- 簡単な操作で高度なビッグデータ分析が可能
- 特殊な分析専用マシンは不要
  市販のマシンを1台用意すれば
  それだけで分析が実行可能



MATHEMATICAL SYSTEMS INC.

製品紹介

データサイズ



MATHEMATICAL SYSTEMS INC.

製品紹介

データサイズ

アンケートデータ
日本全国１億人に、１００の質問
100M*100*4B=40GB（＝0.04TB）

買い物リスト(一年分)
１億人が、一日＝１回、一度＝１０品
100M*365*10*４B=1.5TB

品質管理センサーデータ（１００台、秒単位、１年分）
100*365*24*60*60*8B=25GB(=0.025TB)

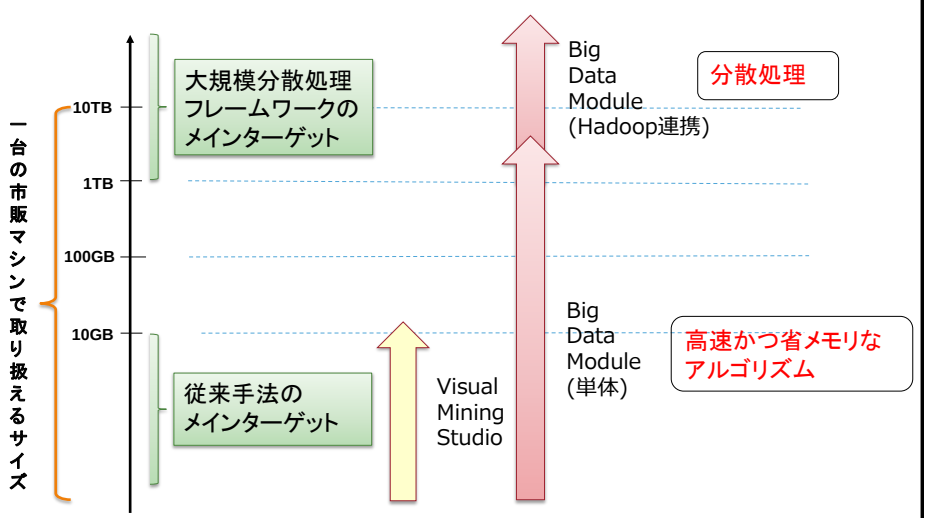
企業評判分析、Web書込み（１ヵ月分）
１00万人、1日＝10 Tweets、１Tweet＝１４０文字
1M*30*１０*140*2B=84GB

……

MATHEMATICAL SYSTEMS INC.

製品紹介

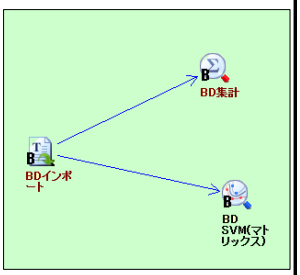
データサイズ



MATHEMATICAL SYSTEMS INC.

Big Data Module の主な特長

- GUIベースの簡単操作
  アイコンを矢印でつなぐだけの簡単操作で、
  高度な分析機能を使用可能



- 並列処理
  並列処理をアイコンベースで簡単に実行可能
  マルチコアをフルに生かして分析を実行

- Visual Mining Studio との連携
  Visual Mining Studio の機能と組み合わせることで、
  Big Data Module の分析結果をさらに掘り下げて、
  多様な分析を行うことが可能

Big Data Module の主な特長

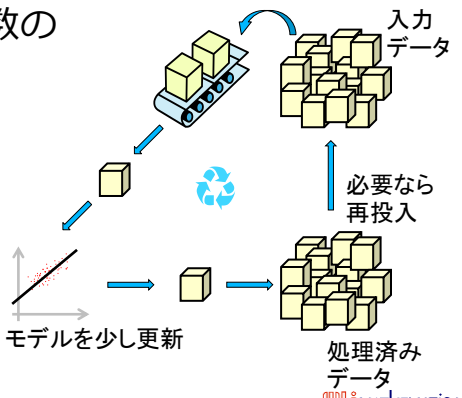
- ビッグデータ分析アルゴリズム
  オンラインアルゴリズムなどの
  ビッグデータに適した分析アルゴリズムを搭載

- Hadoop との連携
  Hadoop上でのデータハンドリングをGUIから簡単に実行
  結果を取得してVAPの分析と接続可能

オンラインアルゴリズム

- データを一つずつ読み込み、モデルを逐次更新
- データをためず、必要最低限のメモリ使用量
- 計算時間は処理データ数の線形オーダー

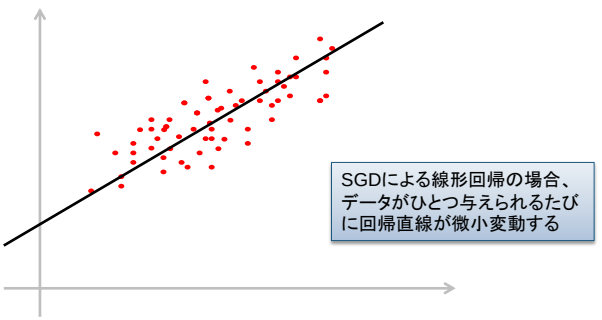
※「オンライン」はチューリングマシンの理論に由来し、「ネットワークにつなぐ」という意味ではありません。



MATHEMATICAL SYSTEMS INC.

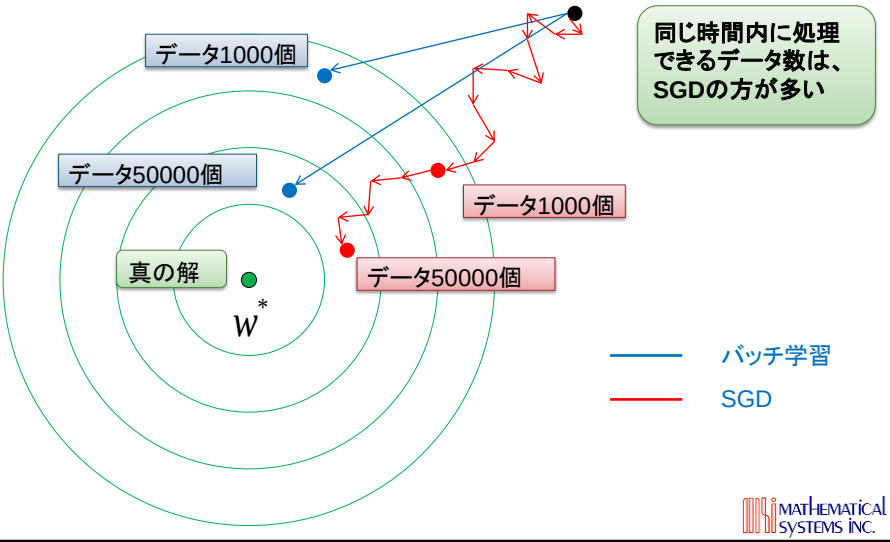
オンラインアルゴリズムの基本エンジンSGD

- Stochastic Gradient Descent
  - 確率最適化問題に対するオンラインな計算手法
    - 線形回帰・ロジスティック回帰・SVM・k-means・行列分解．．．
  - Big Data Module には、最先端のSGD実装を搭載



MATHEMATICAL SYSTEMS INC.

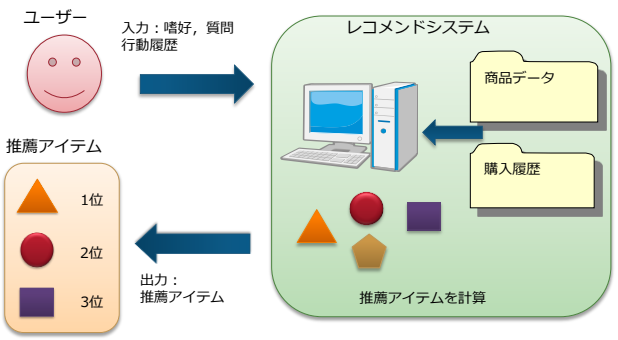
SGDの処理速度



MATHEMATICAL SYSTEMS INC.

SGDの威力

- Netflix Prize(2006/10~2009/7)
  - レコメンデーションの精度コンテスト
  - 1億件のレンタル履歴データ
  - 100万ドルの賞金
  - SGDを駆使したモデルが優勝

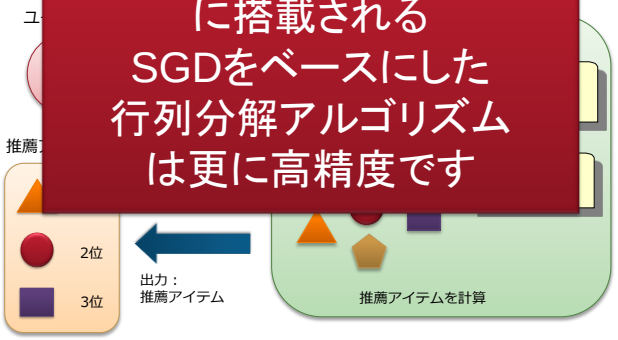


MATHEMATICAL SYSTEMS INC.

SGDの威力

- Netflix Prize(2006/10~2009/7)
  - レコメンデーションの精度コンテスト
  - 1億件のレンタル履歴データ
  - 100万ド
  - SGDを駆

Big Data Module
に搭載される
SGDをベースにした
行列分解アルゴリズム
は更に高精度です

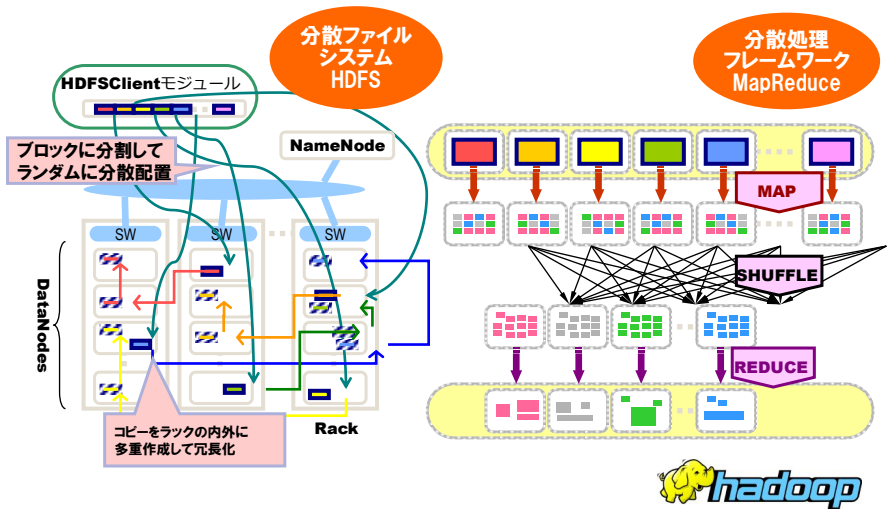


Big Data Module の主な特長

- ビッグデータ分析アルゴリズム
  オンラインアルゴリズムなどの
  ビッグデータに適した分析アルゴリズムを搭載

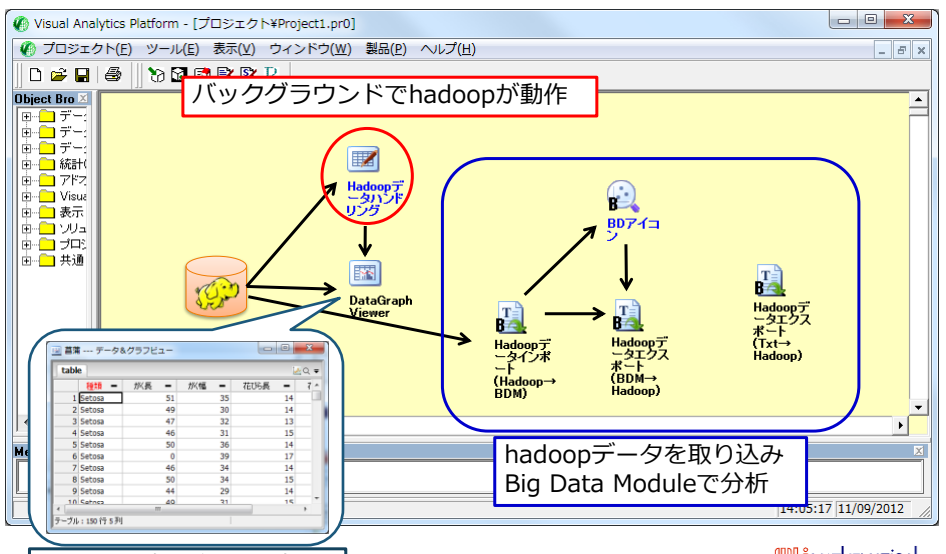
- Hadoop との連携
  Hadoop上でのデータハンドリングをGUIから簡単に実行
  結果を取得してVAPの分析と接続可能

hadoop



MATHEMATICAL SYSTEMS INC.

hadoop連携イメージ



MATHEMATICAL SYSTEMS INC.

Big Data Module の機能一覧

- 基本機能
  - インポート
  - クリーニング
  - サンプリング
  - 基本情報
  - 集計
  - データハンドリング **新機能**
- マイニング機能
  - オンライン線形回帰
  - オンラインロジスティック回帰
  - サポートベクターマシン(SVM)
  - スカッシング
  - オンライン K-means **新機能**
  - オンライン行列分解 **新機能**
- その他
  - hadoop連携 **新機能**

MATHEMATICAL SYSTEMS INC.

ー Big Data Module 応用事例紹介 ー

予測モデル作成
レコメンデーション

MATHEMATICAL SYSTEMS INC.

予測モデル作成

- 売上予測
- 優良顧客判別

- 株価予測
- 与信管理

- 電力需要予測
- スマートグリッド

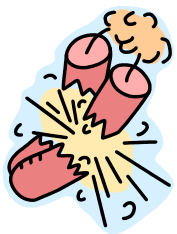
- 年々積み重なるデータ
- 技術の進歩で取得できる情報の種類も量も増加
  → ビッグデータ化
- データが豊富になったんだから予測精度も上がるはず

MATHEMATICAL SYSTEMS INC.

大規模データを用いた未来予測

従来の予測モデル

- 従来の学習アルゴリズムをビッグデータに適用しようとすると…
  データ数、説明変数の数が多すぎる！
  メモリ使用量の爆発的な増加
  計算時間の爆発的な増加
  ⇒ **計算不可能**

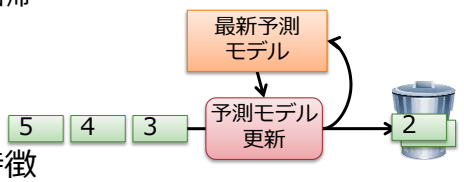


- 従来の学習アルゴリズムをビッグデータに適用することは
  現実的に不可能
- 対応方法
  アルゴリズム的な工夫

MATHEMATICAL SYSTEMS INC.

オンラインマイニングアルゴリズム

- Big Data ModuleはSGDを使用した分析アイコンを搭載
  - オンライン線形回帰
  - オンラインロジスティック回帰
  - オンラインSVM



- SGDを使った分析機能の特徴

  データサイズに依存せず**省メモリ**

  計算時間はデータサイズの**線形オーダー**

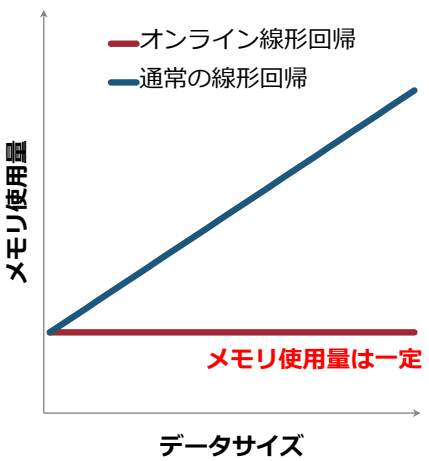
  ⇒ ビッグデータと非常に相性が良い！

メモリ使用量

- 通常のアルゴリズムでは、データサイズに応じてメモリ使用量が増加

  → 計算不可能に

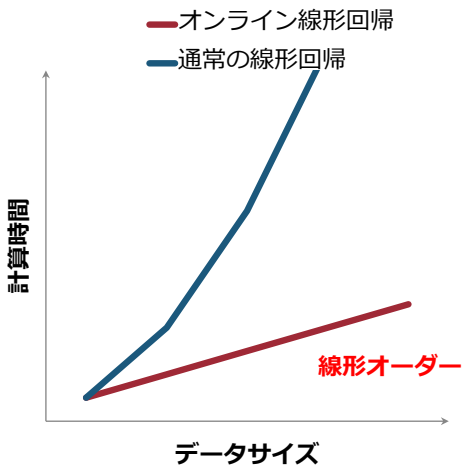
- オンライン線形回帰では、データサイズが増加しても、メモリ使用量は一定



計算時間

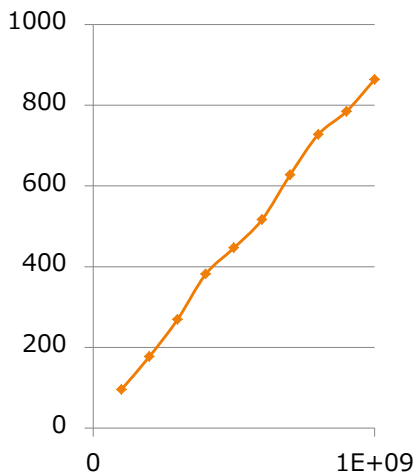
- 通常のアルゴリズムでは、データサイズに応じて計算時間が爆発的に増加
- オンライン線形回帰では、データ量の線形オーダーの時間で計算が終わる

計算時間=α×(データサイズ)



MATHEMATICAL SYSTEMS INC.

オンラインロジスティック回帰の計算時間



行数	計算時間(秒)
1億	95
2億	177
3億	269
4億	381
5億	446
6億	516
7億	627
8億	727
9億	784
10億	863

MATHEMATICAL SYSTEMS INC.

レコメンデーション

商品、記事、サービスなどのアイテムを推薦する機能。

- ECサイト
  amazon、iTunes、
  アパレル、オンライン書店等の
  おすすめ商品
- SNS サイト
  知り合い候補を列挙
- ニュースサイト
  関連記事

その他、さまざまな箇所で導入されている技術

MATHEMATICAL SYSTEMS INC.

協調フィルタリング

ユーザーの嗜好データを用いた推薦方法。

1．嗜好が似たユーザーが高得点を付けたアイテムを推薦。

	映画A	映画B	映画C	映画D	…
ユーザー１	-	5	4	5	
ユーザー４	3	5	4	-	
…					…

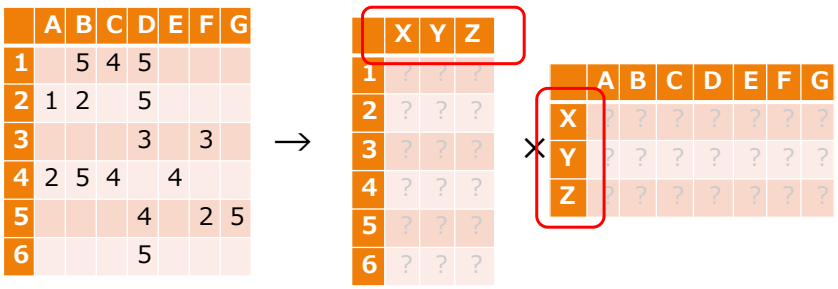
おすすめ！！

- 必要な技術
  １．ユーザーの類似度計算
  ２．類似ユーザーの探索
  （コンテンツベースの方法と同様の技術。）

MATHEMATICAL SYSTEMS INC.

行列分解

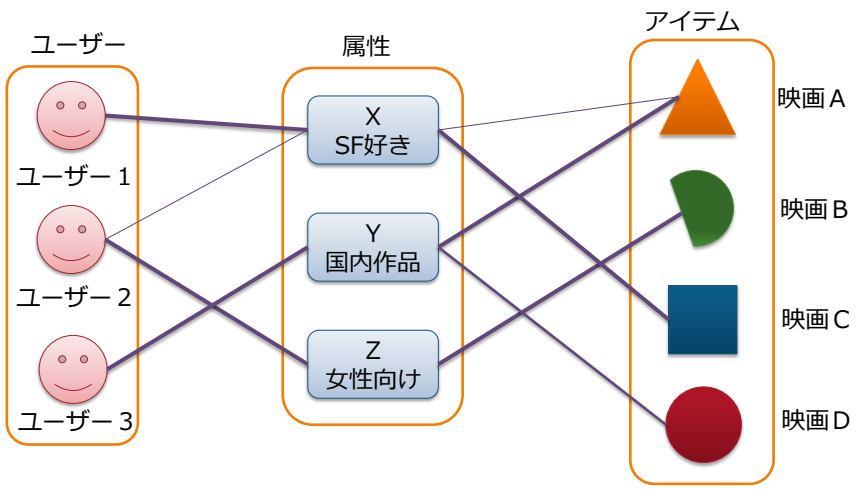
ユーザーと映画の関係を表す行列を
2つの行列の積に分解



新たに表れたX、Y、Zが属性を表す

行列分解を用いた協調フィルタリング

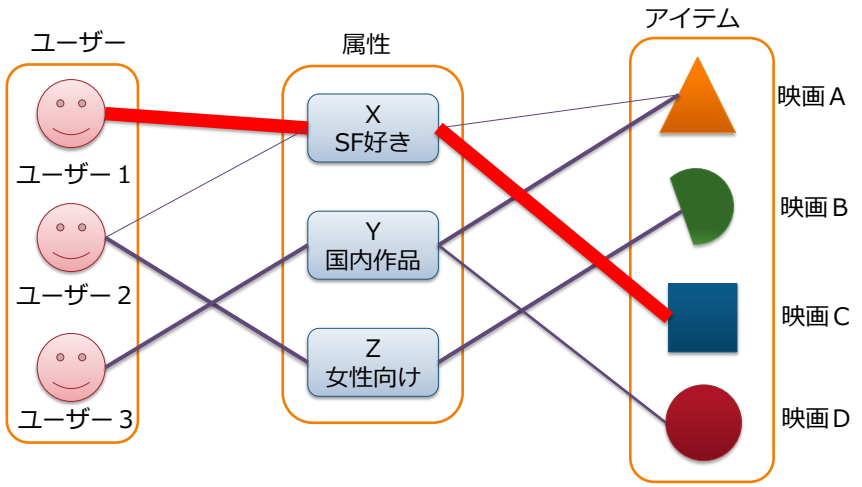
行列分解を用いた協調フィルタリング



レコメンデーション

行列分解を用いた協調フィルタリング

行列分解を用いた協調フィルタリング

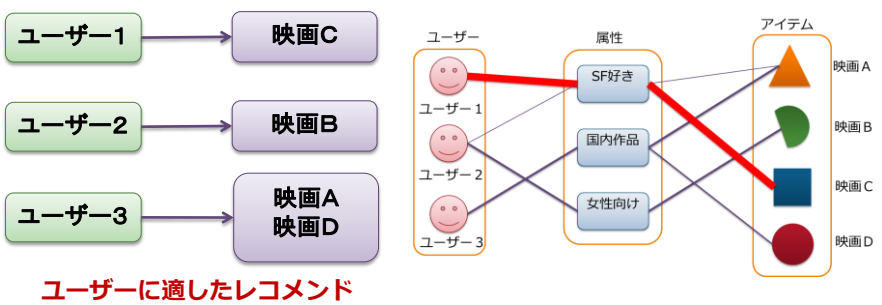


MATHEMATICAL SYSTEMS INC.

レコメンデーション

行列分解を用いた協調フィルタリング

- 各ユーザー・各アイテムと属性との関係性を導く。
- この関係を用いる事でユーザーに適したレコメンド。
- 他ユーザー情報の活用により意外性のあるレコメンド。



ユーザーに適したレコメンド

- Big Data Module のオンライン行列分解アルゴリズムにより、効率的な協調フィルタリングが可能。

MATHEMATICAL SYSTEMS INC.

Big Data Module 導入の流れ

ご要望を御伺いしてお見積り・仕様提案致します．分析のお手伝いや数理的なコンサルティングもご提供可能でございます．

個別相談会 → システム構築お見積り

↓ 試用版でお試し

↓ ソフトウェアの導入

↓ システム構築

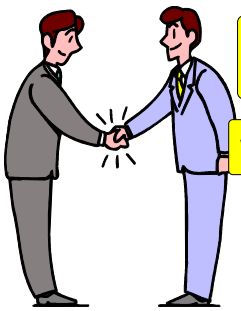
MATHEMATICAL SYSTEMS INC.

本日はありがとうございました

★数理システムがお手伝いできること

※自社展開／他社へ向けてサービス展開は不問

- ご相談・ヒアリング 解決のためのご提案
- 分析のコンサルテーションのご提供
- ソフトウェアのご提供



- 既存のシステムに分析モジュールの組み込みご相談
- 他社に向けての共同でのご提案

bigdata-info@msi.co.jp

何でもご相談下さい!

MATHEMATICAL SYSTEMS INC.