

内閣府 総合科学技術・イノベーション会議
エネルギー・環境WG様向け

データキュレーションサービスのご案内



2017年03月15日
富士通株式会社

基本概要

富士通のデータサイエンティスト：データキュレーター

業界に先駆け、2011年にデータ分析を専門とする組織を立ち上げました。



フレームワーク

200件を超えるPJから
分析FWを蓄積



分析ノウハウ

数百アルゴリズム×
数億パラメータを駆使

| | | | |
|--------|---------|-------------|-----------|
| 統計量 | SVM | ランダムフォレスト | 交差検定 |
| 系列パターン | バギング | ナイーブベイズ | ホールドアウト |
| 項目結合 | ブースティング | ニューラルネットワーク | 適合性判定 |
| 空間分類 | 決定木 | 自己組織化マップ | F Measure |

分析環境

HPC技術を活用した
大規模計算環境



マシンパワーを使って、大量のデータから集合知を得る

データから集合知を得る

人がコンピューターにルールを教える（プログラミングする）のではなく、コンピューターがデータからルールを見つけ出します。

データから強い将棋ソフトを作る

これまでの将棋ソフト

将棋をよく知っている人
有名な強名人、プロ

超死者の固定観念、先入観、
主観などにより、パラメーター設定
約 **500** パラメータ

アマチュア有段者レベル

機械学習を用いた将棋ソフト (2015年 #Borisica)

プロの棋譜 6万局
各駒の価値 歩 角
駒と駒の位置関係

「簡易評価関数」の最適なパラメータを自動学習
約 **1億** パラメータ

プロレベルの棋力を実現

FUJITSU CONFIDENTIAL Copyright © 2016 FUJITSU LIMITED

未活用のデータから疾病リスク予測

2万6000人、過去3年分のデータをもとにした推論結果（社内実証実験）

富士通独自の総合判定方式

一般的な判定項目
HbA1c 空腹時血糖

+

健康診断データ

| | | | | |
|---------|-----------------|-------------|----------|-----|
| 血糖値(空腹) | HDLコレステロール | 血圧 | 白血球数 | |
| 心臓病リスク | 尿酸値 | γ-GT(γ-GTP) | GOT(AST) | MCV |
| HbA1c | BUN | CRP(ALI) | 尿酸値 | 腎臓 |
| 中性脂肪 | HDL-C | LDLコレステロール | コレステロール | |
| →21項目 | 血中尿酸 (ハネアロ) (値) | その他... | | |

レセプトデータ
診察/治療/入院 薬名

予測

糖尿病になる可能性あり

糖尿病になる可能性なし

実績

糖尿病になっていない

糖尿病になった
高血糖
手戻

糖尿病になっていない

糖尿病になった

FUJITSU CONFIDENTIAL Copyright © 2016 FUJITSU LIMITED

マシンパワーにより、
プロフェッショナルと同等なルールを発見することを目指します

データから強い将棋ソフトを作る

これまでの将棋ソフト

将棋をよく知っている人

将棋の強い人、プロ



開発者の固定観念、先入観、
主観などにより、パラメーター設定

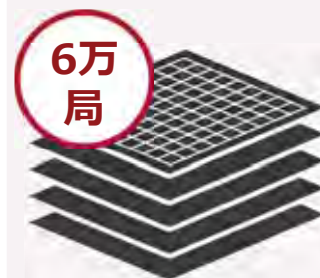
約 **500** パラメータ



アマチュア有段者レベル

機械学習を用いた将棋ソフト (2005年 ※Bonanza~)

プロの棋譜



各駒の価値



駒と駒の位置関係



「局面評価関数」の最適なパラメータを自動学習

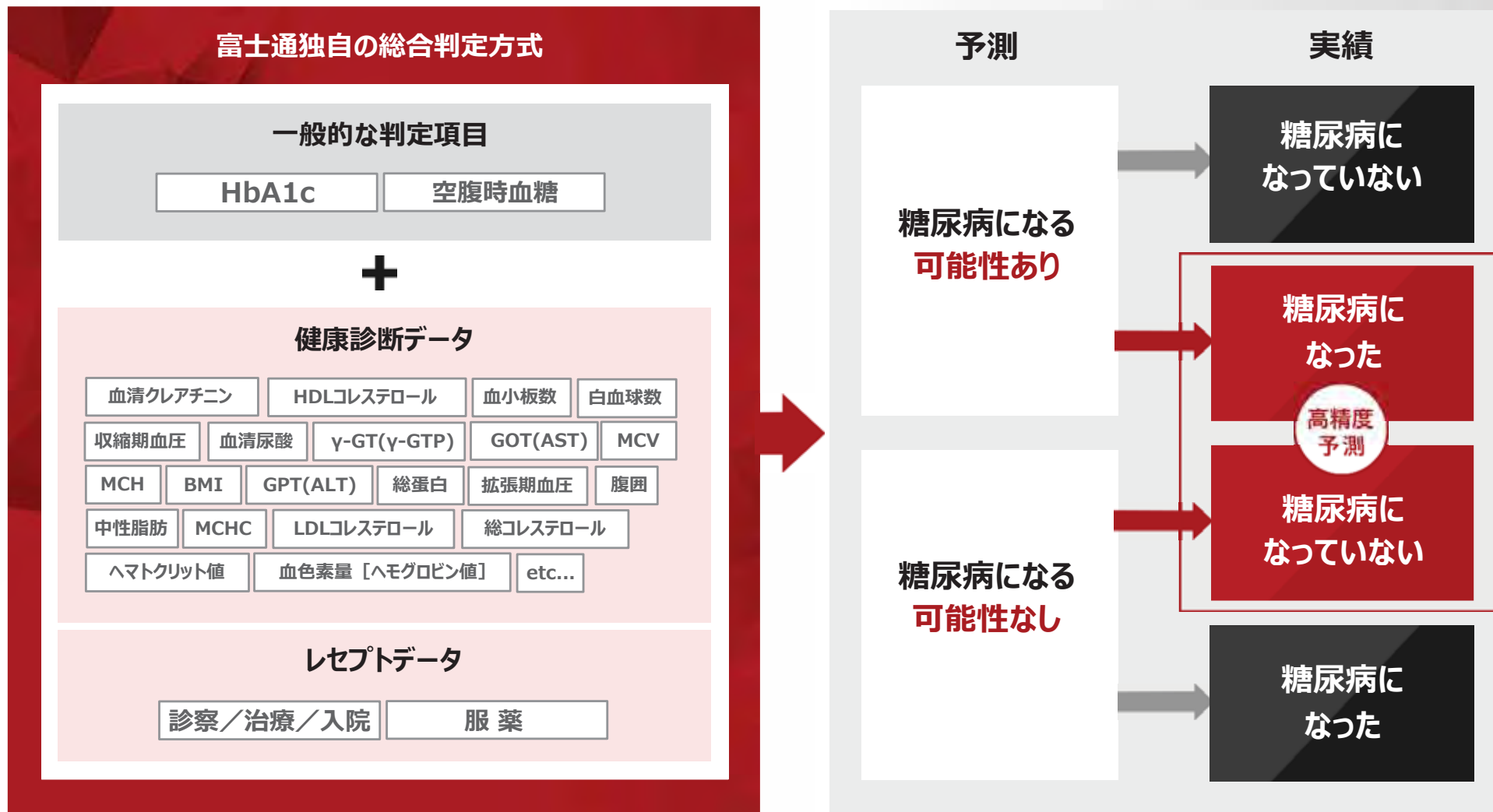
約 **1億** パラメータ

プロレベルの棋力を実現

※「Bonanza」は、保木国仁先生（現 電気通信大学特任助教授）が開発したコンピュータ将棋ソフト

未活用のデータから疾病リスク予測

2万6000人、過去3年分のデータをもとにした推論結果（社内実証実験）



データキュレーションサービスのフロー

データキュレーションサービスは、ビッグデータ分析のプロフェッショナルである「データキュレーター」が、従来のデータ分析手法では解くことが困難な課題を抽出し、課題解決への具体的なアプローチへとつなげていくサービスです。



課題

| | | | |
|------------|-----------|-----------|---------------|
| 業務効率を上げたい | 在庫を最適化したい | 顧客動向を知りたい | 業務上の課題を解決したい |
| 売り上げを伸ばしたい | 新商品を開発したい | リスクを分析したい | 分析精度を今より高くしたい |

お客様保有データの提供



外部データの活用



モデリング



アナリティクス



システムデザイン



お客様の課題をデータで解く
データ活用モデル作成

+



データ解析レポートの提供

富士通の強み