

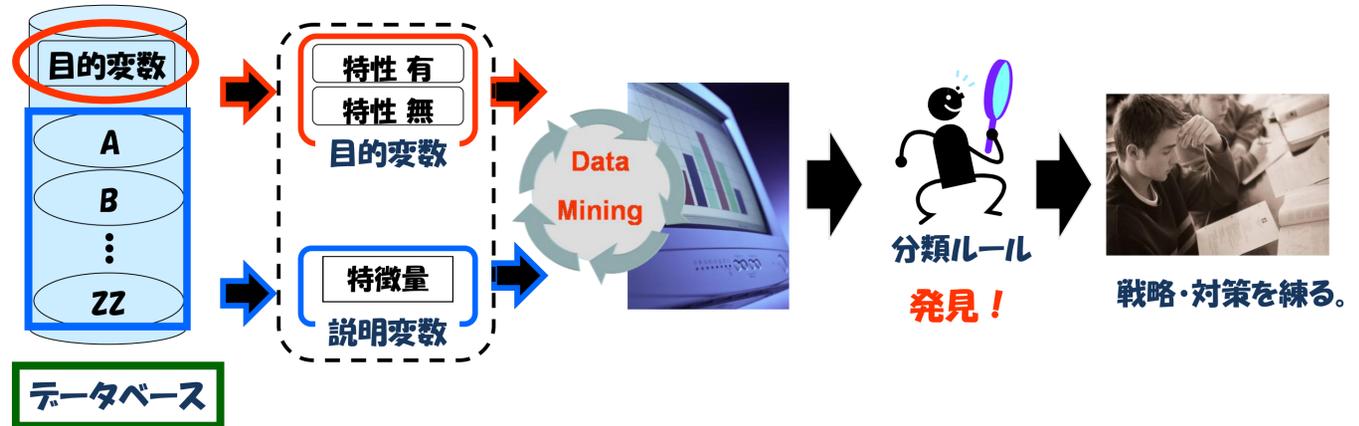
# bump hunting の研究

## 研究目的

巨大データベースの中から、あるターゲットを探し出すことを目的とする。例えば、優良顧客。通常、分類手法が用いられるが、それが困難な場合には、分類することを止めて、代わりにターゲット濃度の高い部分を探し出すことを考える。これをbump huntingと呼ぶ。

## データマイニング

多くの特徴を持つターゲットから、意味のある特徴だけを抜き出して、ターゲットを探す。



## bump hunting

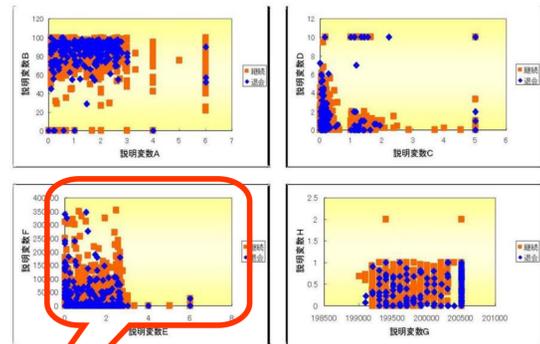
説明変数を無作為に二つ選出。二次元グラフ化。青・赤の点が複雑に重なっている。 **分類困難!**

## データ

データ総数 約100万件

説明変数 約100個

計算量削減のため、説明変数の数を足し合わせ・掛け合わせを行う。模擬データとデータ数を絞った二つのデータを用いる。

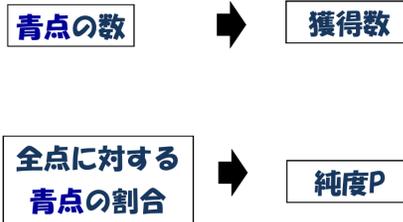
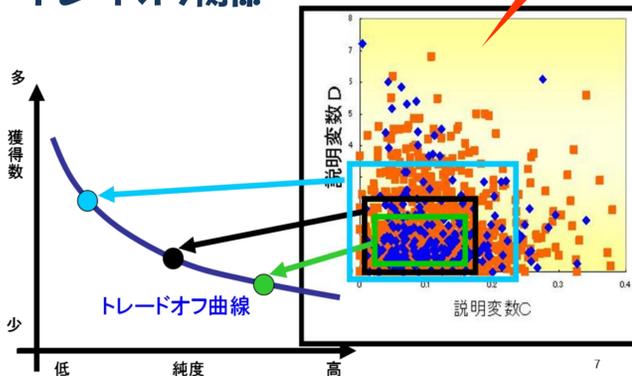


## bump領域

青点密集領域

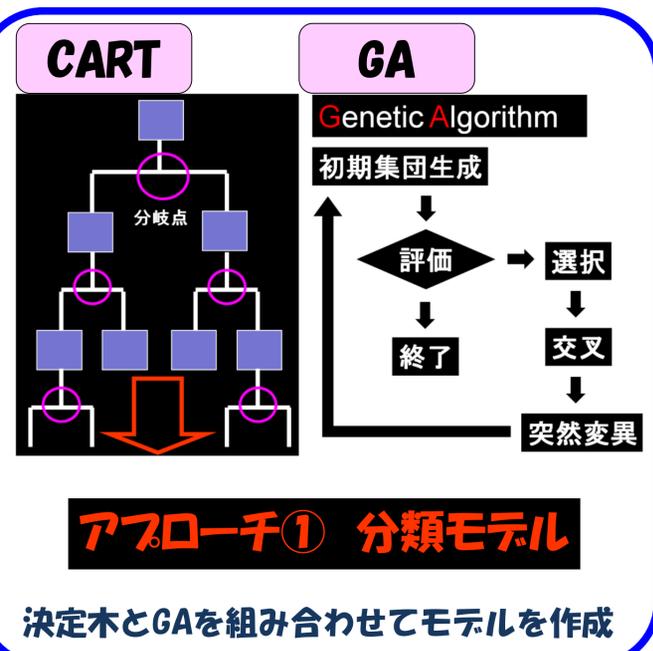
拡大

トレードオフ関係



**目標** ターゲット濃度が高いところで、多くのターゲット数を確保する。

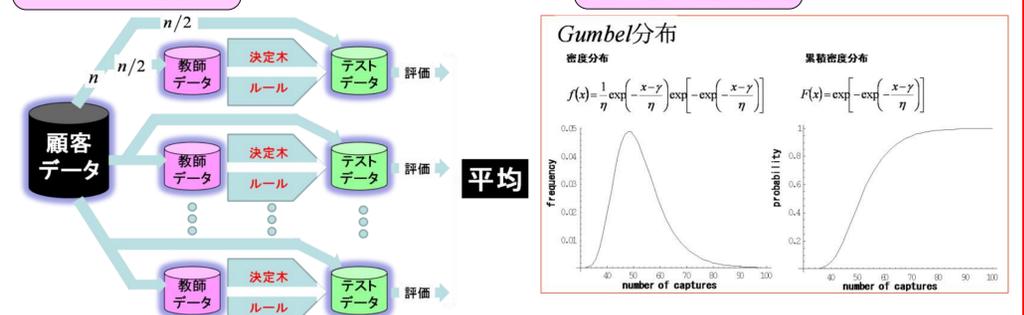
## 研究概要



BHO法と極値分布を用いて、分類モデルが本当に正しいかについて検証を行う。

## BHO法

## 極値分布



## アプローチ② 分類モデルの信頼性

極値分布と信頼度推定を用いて分類モデルの妥当性の検証

## アプローチ③ アクション

抜き出されたターゲットに対してアクションを起こす。