

# 分かりやすい2分割を繰り返し 返す分類方法「決定木」

牧野 浩二

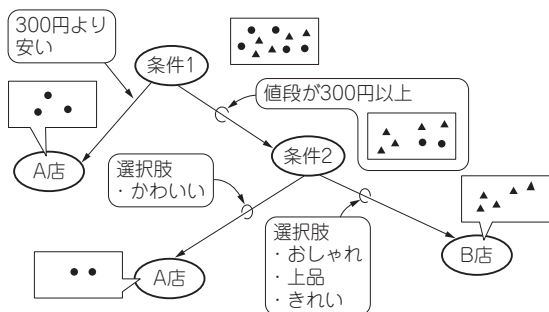


図1 2つに分けることを繰り返して最終的に分類する方法「決定木」

決定木で分類を行います。決定木とは与えられたデータを2つに分けることを繰り返し、最終的にデータを有意差のあるグループに分けるものです。結構、いろいろなところで使われている分類方法です。

例として、ここでもお札の分類を行います。画像処理ではなく、お札の数カ所の明るさ(色の濃淡)で分類します。

## 分類方法「決定木」とは？

### ● 2分割を繰り返して分類

決定木とは、木構造を用いた学習手法の1つです。2つの洋菓子店を例にとり、図1を見ながらイメージをつかみましょう。

2種類のデータを図1では丸と三角で表しています。例えば、2つの洋菓子店(A店とB店)のケーキのデータだとしましょう。データは値段と見た目で構成されているとします。

それをある基準で2つに分けます。例えば、値段が300円よりも高いか安いかで分けると2つに分かれます。左に分かれたものは丸だけ分離できました。こちら側はこれで分類が終わりです。

さらに、ケーキ・データには見た目のデータとして、かわいい、おしゃれ、上品、きれいという4つの選択枝があったとします。そのうちのかわいいが選択され

ているかどうかで分けたとしましょう。この例では、混ざり合った2つのデータがうまく分類できました。

決定木は、このように幾つかの基準で、2つに分けているうちに、うまく分ける条件を導き出すことができます。決定木のキー・ポイントは「この2つに分ける条件を自動的に見つけてくれる点」です。

また、決定木は本物の木をさかさまにしたような形をしているところから木構造を用いた学習方法と呼ばれています。この例では木っぽくないですが、もっと多くのデータを使ってたくさん条件分けをすると図2のように木を逆さまにしたような形となっています。

そういえば木構造を持つ学習にはクラスタ分析がありましたね。なお、図2はRというソフトウェアを使って作りました。これを作るためのコマンドは最後に紹介します。

### ● 他の分類アルゴリズムとの違い…条件が人間にとって分かりやすい

決定木の大きな特徴は、どのような形式のデータでも自動的にうまく分類できる条件を、人間にとって分かりやすい形で示せる点です。他の分類手法では、多くの場合は直接的な原因を示してはけません。

例えば決定木では、重さが1kg以上ならばAグループ、そうでなければBグループなどのように、具体的な分類条件が得られます。そして、それがどの程度重要なのかも、決定木を図示すると一見して分かるようになっていきます。こうした点は他の分類法にない特徴です。

さらに決定木は、値段や長さなどの数値のデータだけでなく、「かわいい」とか「おいしい」、「楽しい」などの言葉が入っていてもうまく分類できる点に特徴があります。

ただし、決定木は過学習(オーバフィッティング)と呼ばれる学習データに特化した分類条件になることが、他の手法に比べて起きやすい点に注意する必要があります。