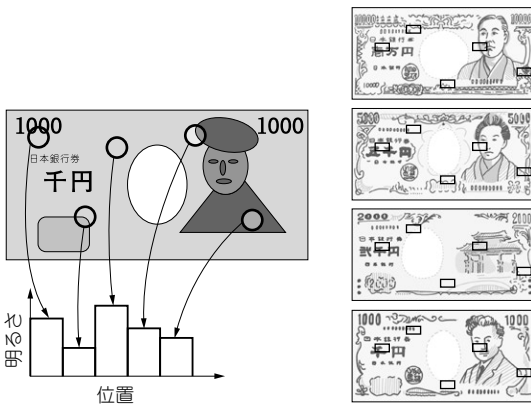


わりとよく使われるタイプは動かしてガッテン! ダウンロード・データあります

# 人工知能アルゴリズム探検隊

## 第29回 分かりやすい2分割を繰り返す分類方法「決定木」

牧野 浩二



(a) お札の5カ所の明るさを測定 (b) 1000円札以外の  
お札も使う

図1 「決定木」という方法を使って測った色の濃淡からお札を分類する

今回は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、いろいろところで使われている分類方法です。

例として、前回に続いてお札の分類を行います。お札は画像処理ではなく、お札の数カ所の明るさ(色の濃淡)で分類します(図1)。

### 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2を見ながらイメージをつかみましょう。

2種類のデータがあるとします。図2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

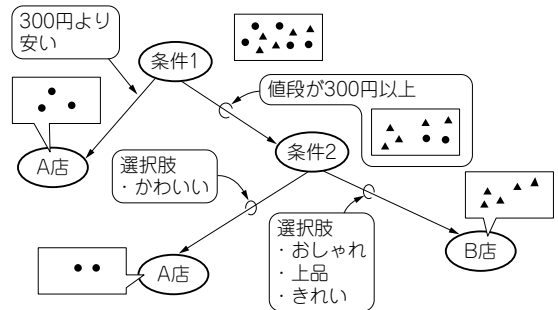


図2 2つに分けることを繰り返して最終的に分類する方法「決定木」

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つの選択肢があったとします。そのうちのかawaii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3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(2017年2月号)がありましたね。なお、図3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