

人工知能 グループ16

1 メンバー全員の学生番号 & 名前

165732B 親川弥月
145770F 秋田海人
165750A 伊波若奈
155749B 清家 拓海

2 2つのアルゴリズムにはどんな特長があるかを簡単に解説

2.1 自己組織化マップ

例えば最初のランダムに選ばれた (r, g, b) (=赤、緑、青) の数値の 1 0 0 色のパレットに、青色のマスを選んでその周辺の 8 マスを青色にするというプログラムを設定する。

一見ランダムに色が動いているだけだが、だんだんと色の分類が始まっている。つまり、似た色が近くに集まってくる。たとえば、ある領域に赤や青、緑っぽい色が集まっているようになる。このように、自然に「色」の分類が生まれてくる。計算ステップには、どこにも「分類せよ」という指令は入っていない。単に、特定のルールに従って計算しているだけだ。それにもかかわらず、結果として、自然に色のカテゴリー分けを始めてしまう。このように教えていないのに自然に秩序が生まれてくることを「自己組織化」と言う。

この分類アルゴリズムは、開発した人の名前を取って、「コホネン（コホーネン）の自己組織化マップ（SOM）」と呼ばれている。

自己組織化のプロセスは、ヒトがものを学習していく様子にすごく似ている。たとえば、幼児が言語を習得するときには、「言葉を学べ」とだれから教えこまれなくても、ただ言葉に暴露されるだけで、自然に言語を身につけていく。親や周囲の人の会話を聞くだけで、自発的に単語の意味や使い方を分類していく。これと同じように、僕らはものごとをだれから習うこともなく分類したり理解したりできる。そういう学習方法を「教師なし学習」と言う。「コホーネンの自己組織化マップ」は教師なし学習の代表例である。

2.2 決定木アルゴリズム

決定木の学習は、構造を学習するアルゴリズムのうちで最も簡単な部類のものではあるが、実用的に良く用いられて、うまくいっている方式の 1 つである。

決定木の目的は、属性とその値の組 { 属性 1=値 1, ..., 属性 n=値 n } によって表現されたデータをいくつかのクラスと呼ぶものに分類することである

目的変数に属する確率を複数の説明変数の組み合わせで算出する方法。

イメージは以下で、Yes/No などの条件に属するかどうかで確率を算出する。

決定木の学習とは、具体的な判断事例から決定木を生成することである。事例は入力={ 属性 1=値 1, ..., 属性 n=値 n}:出力=Yes/No のような形式のデータで、入力の属性値が具体的にどのような値のときに出力が Yes なのか No なのかを指定する。

このようなデータの集まりを、学習アルゴリズムに対する訓練例という。学習アルゴリズムは、訓練例を入力として受け取ると、その訓練例をなるべく良く再現できるような決定木を出力するものである。

3 どのように組み合わせるを使うか？

組み合わせるを使う方法として半教師あり学習があります。一部にラベルが付与されているデータセットに対して、教師あり学習と教師なし学習を組み合わせる手法です。

- ・教師ありデータを見える化・可視化するとき
- ・回帰モデル・クラス分類モデルを構築するとき
- ・モデルの適用領域を設定するとき

参考 URL: <http://univprof.com/archives/16-06-09-3767887.html>

4 どのようなデータを対象に 3 を実行するか？

4.1 目的

車のボディタイプを SOM を用いて分類を行い。決定木により車のボディタイプのルールを決定づけることを目的とする。

4.2 データ収集は可能か？可能であれば何処から？

手作業になるのですが下記の URL から情報を得てデータを作ります。

車のボディタイプの url

<http://car-moby.jp/235617#c4>

乗車人数

<https://www.navikuru.jp/articles/knowledges/37/>

排気量 (燃費)

<https://www.car-erabi.com/18.html>

4.3 データの前処理は必要か？

今回は私たちがデータを作ったので前処理を必要としない。

4.4 実験・検証

今回用いるデータは図 1 になります。

図 1 を見る限り、ミニバンとセダンが近くになるのではないかと思います。また、オープンカーのみ屋根がないので、遠くに位置すると考えられる。

	セダン	クーペ	コンパクトカー	オープンカー	ミニバン	SUV	軽自動車
5人以上	1	0	1	0	1	1	0
4人以下	0	1	0	1	0	0	1
2ドア	0	1	1	1	0	1	0
4ドア	1	0	0	0	1	0	1
屋根あり	1	1	1	0	1	1	1
屋根無し	0	0	0	1	0	0	0
アウトドア向け	0	0	0	0	1	1	0

図 1: データセット

SOM を用いて分類を行なった。

それが図 2 となる。

SOM の実行結果より、私たちが予想した通りミニバンとセダンが近くに分類された。また、オープンカーは遠くに位置されることがわかる。

クーペのみ他の車と隣接しているのでどれとも似ていることがわかった。

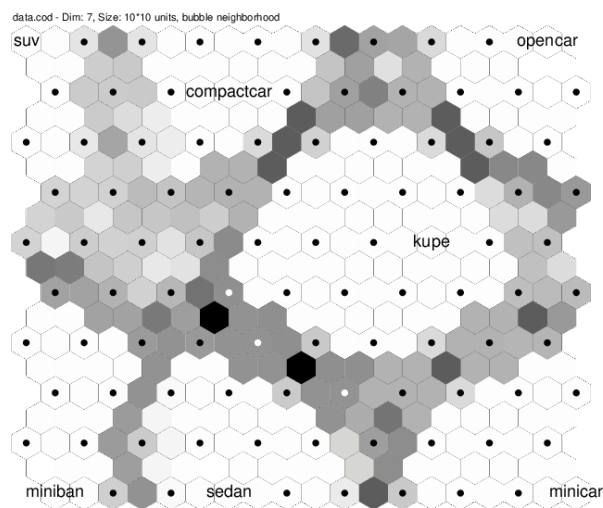


図 2: SOM

今回決定木を試すまでには至らなかったが、データセットを用意する大変さや SOM の分類がどんなものか理解を深めることができました。