

概要

情報システム工学科 4年 035 土屋寛貴

研究テーマ

統計処理ソフトRを用いた含有栄養成分による魚介類の分類

研究内容

統計処理ソフトRのサポートベクターマシンとNaive Bayseを用いて含有栄養成分によって魚介類の分類をどのくらいの精度で出来るのかを試みた。

研究方法

解析用データに魚介類100gにおける栄養成分表を用いた。データは下記のホームページから抽出し、それらに分類に用いるラベルとして生物種の門、綱、目まで入力し、csvファイルにした。下の画像はcsvファイルにする前の加工後のデータの一部である。縦169行横36列のデータになった。

<http://www.uol18.com/eiyou/>

	界	門	綱	目	たんぱく質(%)
◆アイナメ(100g)	動物界	脊索動物門	硬骨魚綱	カサゴ目	19.1

Rに上記の解析用のデータを入れてksvmという関数とNaivebayesという関数を用いて学習させ、実際の生物種の分類をどの程度の精度で出来るかを試みた。上記のようなデータの自分以外の行のデータを学習データとして用いて分類した。どちらの関数の方が的中率がいいかを調べた。

精度

	門(4種類)	綱(10種類)	目(35種類)
ksvm	93%	86%	43%
Naivebayes	86%	82%	38%

考察

門、綱、目というふうにラベルの種類が増えていけば行くほど精度が落ちて行くのは予想のとおりだった。目のところで一気に種類が増えるのでそこで精度がかなり落ちた。また目にはスズキ目という種類があり、それが全体の3分の1を占めており、ksvmでは予測するときに数が多いものを予測に出したが、という性質があるらしいので、スズキ目は予測がすべて当たっていたが、全然関係ないところにもスズキ目が予想としてかなりたくさん出てきており、それが精度を下げた大きな原因だと思われる。ksvmとNaivebayesではksvmの方がすべて精度が良かったが、Naivebayesの方は目でスズキ目が過剰に出ることはなく、スズキ目以外の箇所はksvmより精度が良かったのだが、今度は逆にスズキ目があまり当たらなくて、結果ksvmより精度が落ちてしまったようであった。

感想

Rというソフトは今回はじめて触れてみたが、分類や回帰がすごく簡単にできて、とても便利だと思った。自主課題とは関係なく個人的にいろいろ分類や回帰を使って予測させたりしてみるのも面白そうだった。