

概要

テーマ：統計処理ソフトRを用いた含有栄養成分による食品の分類

313 渡部起平

研究内容

研究内容は統計処理ソフトRを用いて含有栄養成分によって食品の分類をどのくらいの精度で出来るのかを試みる。

研究方法

<http://kcal.emo-web.com/>に掲載されている食品別栄養成分表をexcelでまとめcsvファイルに変換をおこなった。データは縦96行横25列の行列となり11の属性に分類することにした。項目として21の栄養素の含有量をもとに分類をおこなった。

Rによる分類方法

具体的にはRに上記の解析用のデータを入れてksvmという関数を用いて分類させ、実際の分類をどの程度の精度で出来るかを試し、そこにどのような傾向があるかを調べた。

結果

RBF 線形カーネル

10分割交差検定 63.12% 83.12%

leave-one-out cross-validation 57.85% 85.13%

考察

今回の実験ではデフォルトのRBFカーネルよりも線形カーネルの方が大幅に正答率が高かった。これはデータによって最適なカーネルが異なるという原因から起きた結果であった。またこれらの誤差率出力より、肉類、種実類は誤りが多い結果を得られた。詳細を調べてみると特徴ベクトルがほかのデータに近い値であったこともあるが、なにより個体数が多いクラスラベルに分類されやすいことが分かった。今回は使用したデータが少なかったため各個体数の差が結果に顕著に表れてしまった。実験データをさらに膨大なものにすれば精度は上がるのではないかと考える。

総括

今回の実験では、使用したデータが小規模なものであったことにより望むような結果を得ることができず非常に残念であった。しかし、この結果から得られた教訓を生かし、今後データ処理から価値ある情報を生み出していきたい。