

AI 自習ドリル

第5回 新規データの仕分けが得意で高性能な サポート・ベクタ・マシン

牧野 浩二, 足立 悠

サポート・ベクタ・マシン (Support Vector Machine, SVM) とは、答えの分かっているデータを幾つかのグループに分類するための手法です。これは教師あり学習の一種で、現在用いられている教師あり学習の中で、かなり良い性能を発揮する分類法です。

この原理が分かると、どのようなデータの分類に適した方法なのか分かるだけでなく、得られた結果をうまく評価できるようになります。ぜひ使いこなしたい分類方法の1つです。

1 サポート・ベクタ・マシンでこんなことができる

医療

医療分野では、原因がはっきりとしない分類法の使用は敬遠されがちです。サポート・ベクタ・マシンは、判定基準を調べることは難しいのですが、判定基準を可視化できます。そのため、医療への応用の研究や試みが多く行われています。

● 血液検査の結果から病気の有無

血液検査からは多くの情報(中性脂肪、ヘモグロビン、アミラーゼなど)が得られます。実際に医師はこの情報を基に病気の診断を行っています。サポート・ベクタ・マシンは、これらの情報をうまく組み合わせることで、特定の病気かどうかを判定します。

● 心電図から病気の有無

心電図の波形は病気かどうかの判定に利用されます。連続する時系列データを周波数成分に分解するなどして波形の特徴量を数値化し、サポート・ベクタ・マシンで判定します。

● 薬の化学構造から効果の分類

薬の効能は実際に飲んでみないと分からないのですが、これには大きなリスクが伴います。そこで薬の化学構造を情報として、薬の効果を分類する方法が研究されています。これができると、ある病気に効くと予想される薬の化学構造を判定できるかもしれません。

情報技術

サポート・ベクタ・マシンは、情報の特徴をうまく抽出して分類することが得意です。この特性を利用して検出や判定に利用されています。

● 顔検出

デジカメやスマートフォンで人を写すと、顔の部分に枠が出ることがあります。画像の特徴量をサポート・ベクタ・マシンで仕分け、顔を見つけています。

● 迷惑メールの判定

メール・ソフトウェアによっては、迷惑メールを自動的に仕分けてくれるものがあります。これにもサポート・ベクタ・マシンが利用されています。

コンピュータは単語を数字に置き換えています。迷惑メールによく用いられている単語に相当する数字が含まれていたり、その数字の並びに特定のパターンが含まれていたりすることで判定しています。

● 飲食店推薦システム

口コミサイトにはお店に関する情報がたくさん書きこまれています。上記の迷惑メールの判定と同じように文章を解析し、サポート・ベクタ・マシンで判定し、お店を選ぶシステムがあります。