

平成18年 自主課題研究

ニューラルネットワークにおける花粉の飛散量の予測

0408060411 小林 透

現在、ニューラルネットワークといえばパターン認識やデータマイニングに応用されているが、その原理や構造がどうなっているかについて、今回の自主課題研究を通して実際にそれらを学び、実装したプログラムを使い理解を深めることを目的とする。
最終的な目標として「ニューラルネットワークにより、花粉の飛散量を予測する」ことを研究課題として設定する。

ニューラルネットワークとは、脳機能を担っている神経細胞（ニューロン）を計算機上のシミュレーションによって表現しようとする数学モデルである。

花粉の飛散量の予測

平均気温・降水量・日照時間・夏日（25℃を超える日）・降水の無い日（降水が1mm未満の日）・毎年花粉の飛散量のデータを用い花粉の飛散量を予測する。
花粉の飛散量の特徴として次の3つを取り入れてみる。

飛散量は前年の夏から秋の気候による	→前年の7月～11月のデータを用いる
花粉の飛散には2, 3年の周期がある	→その年と過去3年の花粉の飛散量を入力する
実際の飛散量は降水が無い日のみのものである	→2月～5月の降水の無い日との比で求める。

テスト結果（多い・普通・少ないの3段階評価）

1.10年分・花粉の周期あり	2.10年分・花粉の周期なし
・学習したデータに対する正答率 100%	学習したデータに対する正答率 100%
・正答率 21.4%	完全に一致している正答率 42.9%
	ほぼ一致しているものの正答率 57.1%
3.10年分・花粉の周期あり・異なった正規化	4.12年を学習・1年をテスト・花粉の周期なし
学習したデータに対する正答率 100%	学習したデータに対する正答率 100%
正答率 21.4%	正答率 30.7%

考察

1.の正答率は約20%しかない。これは単純に予測結果が「少ない」になるように重みが偏ってしまっていると考えられ、その原因が花粉の飛散量のデータの正規化の仕方にあると考えられ、それを元に行ったのが3.であり、その結果からも分かるように、このニューラルネットワークにおいて、花粉の周期のデータが予測を困難にしているということを示しているのが2.の結果である。以上のことから、使えるデータを全て用いて学習とテストをすることとし、行ったのが4.であり、正答率から言えば決して良くは無いが、その間違え方をみると「多い」ところを「少ない」といったように、遠いところの答えを出して間違えるということが少なかったのも、惜しい答えまでは出せるようにはなっている。

これらの結果は単純に考えると、入力データ数を変えることで結果がかなり変わることから入力データが少ないことがもっとも大きな要因だと考えられる。

