

気象データに基づく CombNET-II を用いた霧発生予測

Fog Forecasting by CombNET-II

ANTO Satriyo Nugroho 黒柳 奨 岩田 彰

ANTO Satriyo Nugroho, Susumu KUROYANAGI and Akira IWATA

名古屋工業大学 電気情報工学科

Department of Electrical and Computer Engineering Nagoya Institute of Technology

1. はじめに

パターン予測を行うためにさまざまな手法が用いられているが、最近ニューラルネットワークを用いる手法が多くみられ様々なモデルが提案されてきた。ニューラルネットワークは統計的な手段の一つであり、汎化性が高いという特徴をもっている。今回、我々は櫛型のニューラルネット CombNET-II[1]を用いた予測システムを提案し、実際の予測問題に適用した。ここで予測課題として、気象データに基づく霧発生の有無を予測するものである。

2. 実験準備

本実験で用いる気象データは気象庁・数値予報課の提供したものである。オリジナルデータは、北海道の新千歳航空測候所による、航空機運搬を支援するための定時地上気象観測データである。本来このデータは、英文混じりのファジイな電文であるが、気象庁・数値予報課が解読プログラムを開発し、数値データに変換されたものである。このデータは 26 項目から構成されており、それぞれ時間の情報と気温、気圧などの気象情報から成り立つ。このうち「視程」と「天気」の項目によって、データを「霧発生」と「霧発生しない」に分類する。そして、これらデータを CombNET-II に学習データとして与え、霧の発生の有無を出力するネットワークを構築する。データに対する前処理として、「年」、「月」、「日」、「時刻」、「視程」と「天気」のデータを削除し、その他の各項目をそれぞれの最小値と最大値によって、-1 から+1 まで正規化する。なお、気象データ中に観測されなかったものが多く含まれているが、このデータを強制的に 0 に変更することによって学習への影響を最小化できる。

この条件に基づき気象データのクラス分けが行われる。さらにより良い予測結果を得るために、学習データとして予測時刻の前 3 時刻の情報を含めて学習を行なった。

3. STEM ネットワークによるドミナントクラスの分割

今回用いた気象データにおいては、「霧発生」データの数に比べて、「霧発生しない」データ

の数が非常に多い。このようなデータを 3 層の階層型ネットワークで学習することは可能ではあるが、収束するまで非常に時間がかかる[2]。このような問題に対して、「霧発生しない」データをあらかじめ何らかの方法で分割し、この問題を多くのサブ問題に置き換えた方がより学習が容易になると考えられる。そこで、我々は CombNET-II を適用することによって、この問題を多くのサブ問題に変換し、学習しやすい環境を作る。CombNET-II は 2 つのネットワークから構成されており、それぞれ自己増殖形アルゴリズムを適用するベクトル量子化用の Stem Network と BackPropagation 学習則を適用する階層型の Branch Network である。前段の StemNetwork で、「霧発生しない」のデータをベクトル量子化し、得られた各サブグループを「霧発生」のデータと組み合わせ、Branch Network に学習させる。各 Branch Network の学習は独立であるため並列に学習することが可能である。さらに、StemNetwork における分割によって得られた「霧発生しない」データのサイズを小さくおさえられるため、「霧発生」のデータとデータサイズがほぼ均等となり、バランスのとれた学習データが得られる。

4. 実験方法

前処理を施した気象データをネットワークに学習させる。実験で用いるパラメータは以下に示す。

学習用のデータ :1988,1990,1991,1992,1993
年のデータ (86195 個)

テスト用のデータ :1994 年のデータ
(17509 個)

データの次元数 :80(=20+3*20:1 時間 30 分
前
までの情報を含める)

Stem Neuron 数 :18

Branch Network の

Input Layer Neuron :80

Hidden Layer Neuron :50

Output Layer Neuron :2(霧発生の有無に
相当)

以上の条件のもとでシミュレーションを行った。予測を行うには、第一層に入力パターンを通し、各 Stem ニューロンの参照ベクトルとの距離が最も近いものから数番目までのニューロンを選び、次に選ばれたニューロンを担当する階層型ネットワークに入力ベクトルを通し、出力層の中で最も高い発火レベルになったものの出力値を調べる。そして、(StemNetwork における参照ベクトルとの距離) α (出力値) の値が最も高くなったものを識別結果として選ぶ ($0 < \alpha < 1$)。 α の値を大きな値に設定すれば、Branch Network の影響が強くなり過ぎて、「霧発生」のニューロンが発火しやすくなり、ドミナントクラスである「霧発生しない」の正解率が下がる。

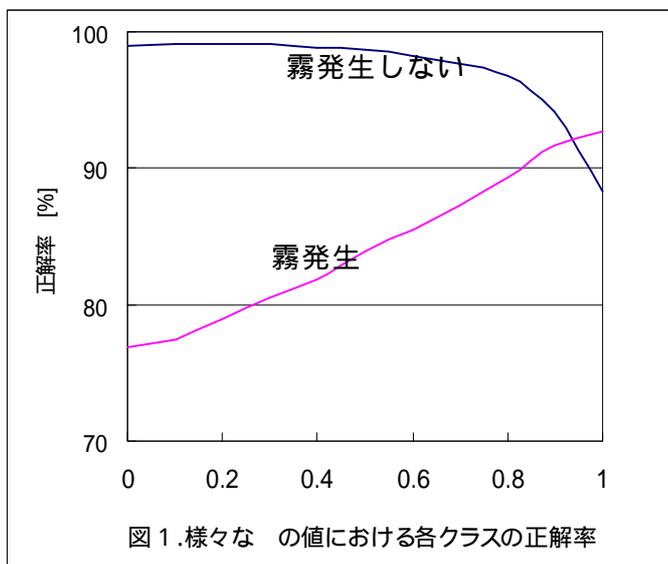
全体として高い正解率を得るために、 α の値を小さい値に設定する。今回比較対象として、3層階層型ネットワークと最近隣法を用いた。階層型ネットワークの構造は

Input Layer Neuron : 80
 Hidden Layer Neuron : 50
 Output Layer Neuron : 2
 学習回数 : 1250

とし、学習則には BackPropagation を用いた。最近隣法はテストパターンをすべての学習データとのユークリッド距離を定めて、最も近いパターンの属するカテゴリを識別結果とする方法である。この方法は高い正解率を得られるが学習データの数が多いため必要な計算量が膨大になる。

5. 実験結果

CombNET-II で得られる各クラスの正解率は α の値によって図 1 のように変化している。



が小さい場合は、霧が発生しない場合の正解率が高いが、霧が発生した場合の正解率が低い。ここで、 α を 0.2 の値を採用する場合は全体の正解率は 98.8%(誤認識パターン数:210)になっている。これに対し、最近隣法の正解率は : 98.7%(誤認識パターン数:232)、階層型ネットワーク : 98.6%(誤認識パターン数:249)である。以上の結果より、CombNET-II が最もよい正解率を得られることが分かる。

CombNET-II は最近隣法に比べて予測するのに必要な時間が非常に短いので、認識システムとして有利である。また、CombNET-II ではもとの問題のサイズを最小化できるので、階層型ネットワークに比べて学習し易い利点を得られる。

6. まとめ

我々は、霧発生予測をするのに CombNET-II を用いた手法を提案し、最近隣法と階層型ネットワークより高い正解率を得られた。さらに、CombNET-II における BranchNetwork の規模が小さいため、学習が容易になる。以上のことより、CombNET-II は予測手段の一つとしての有効な方法と考える。

7. 参考文献

- [1] 堀田健一、岩田彰、松尾啓志、鈴村宣夫, “大規模ニューラルネット”CombNET-II,” 信学論 (D-II), vol.J75-D-II, no. 3, pp.545-553 (1992-03)。
- [2] Rangachari Anand, Kishan G.Mehrotra, Chilukuri K.Mohan, and Sanjay Ranka, “An Improved Algorithm for Neural Network Classification of Imbalanced Training Sets,” IEEE Trans. on Neural Networks, vol.4, no.6, November 1993。