Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186-5140
Volume 7, Number 1, pages 38–40, January 2018

Deep Neural Network to Predict NDA Graduates Refusing to Become SDF Officers

Kazuhiro Seiwa, Keisuke Iwai, Takashi Mastubara, and Takakazu Kurokawa
Dept. of Computer Science, National Defense Academy of Japan
1-10-20 Hashirimizu, Yokosuka 239-8686, Japan
{em55036, xiwai, xmatubara, xkuro}@nda.ac.jp

***Abstract*— This paper proposes feed forward neural network architecture to predict NDA graduates refusing to become Self Defense Force officers on National Defense Academy of Japan. Using the actual student data including 24 items such as course grades and club activities, our deep neural network recognizes the pattern of the students. It will resolve the problem on personnel affairs. Through neural network architecture has some hyper parameters which need adjustment for each task, its simulation results showed the optimum parameters by several verifications. The result of this research shows it becomes possible to predict NDA graduates refusing to become SDF officers with an accuracy of 93.0%, and its F-measure equals to 0.31.**

Keywords*— *neural networks, random forests, support vector machine, machine learning.

I. はじめに

防衛大学校は，陸・海・空各自衛隊の幹部自衛官となる者を教育する防衛省の機関である．防衛大学校の卒業生は，陸上・海上・航空自衛隊の幹部候補生学校へ進学し，自衛隊の幹部候補生となる．現在，防衛大学校においては，民間企業や他の官公庁に再就職するため，学校を卒業しても幹部自衛官となることを辞退する者（以下「任官辞退者」）がおり，防衛計画における人員確保の難点で問題となっている．実際，2017 年には卒業学生 380 人中，8.4 %にあたる 32 人が任官辞退している．

学校が有する学生のデータは，任官辞退するかとの相関性が明らかな要素を含んでおらず，統計的に計算し予測モデルを作るのは非常に難しい．一方，脳の神経回路の仕組みを模したモデルであるディープニューラルネットワーク (以下「DNN」) の技術が，画像や統計などの多次元量で線形分離不可能な問題を解決する手法として，様々な分野で用いられている．DNN は多層のニューラルネットワークであり，この手法を用いた機械学習を深層学習という [1].

我々はどの学生が任官辞退するか否かを予測し，その学生に対して注意及び指導する必要がある．そこで，本研究では，多層構造を持つ DNN により学生の任官辞退を推測する解法を提案し，任官辞退が見込まれる学生の早期発見及び対処が可能となることを目指している．

また，DNN 以外の非線形モデルの回帰分析のランダムフォレスト及びサポートベクターマシンを用いて任官辞退者を予測し，DNN との予測精度の比較も行う．

II. 機械学習

ここでは，本研究で用いる機械学習の代表的 3 手法（DNN，ランダムフォレスト，サポートベクターマシン）について述べる．

A. DNN

画像・音声・自然言語を対象とする分野では，DNN の有効性が確かめられ，広く認知されることとなった [2, 3, 4]．しかし，それらの分野以外においても，データマイニングの分野においてそれを活用しようとする研究がいくつも存在する．例えば，情報推薦技術においては，DNN を用いて消費者が好むファッションブランドを推薦する手法が提案されている [5]．これは，消費者の「性別」「年齢」「身長」「よく着るブランド」を入力し学習することにより，その消費者が好きなブランドを予測し，それを提示するものである．

本研究で用いる DNN は，図 1 で示されるような入力層から一方向にのみネットワークをたどり，有向閉路を持たない順伝播型ニューラルネットワークである [1]．順伝播型ニューラルネットワークは，入力層，中間層，出力層で構成される．図 1 の中の○で示された各ユニットは，式 (1) のように入力値 x と重み w の掛け合わせとバイアス b の総和 u を算出する．この u を活性化関数 $z(u)$ に代入した出力値が，そのユニットの出力値となる．出力層のユニットの出力値 $f(u)$ が最終的なネットワークの出力となる．w は学習を繰り返す度に，目標値 d に近づくよう誤差逆伝播法 [6] と呼ばれる計算手法により更新される．これは一般的に「教師あり学習」と言われる．

本研究では，この DNN を C 言語により実装した．

$$u = x_1 w_1 + x_2 w_2 + \cdots + x_n w_n + b \tag{1}$$

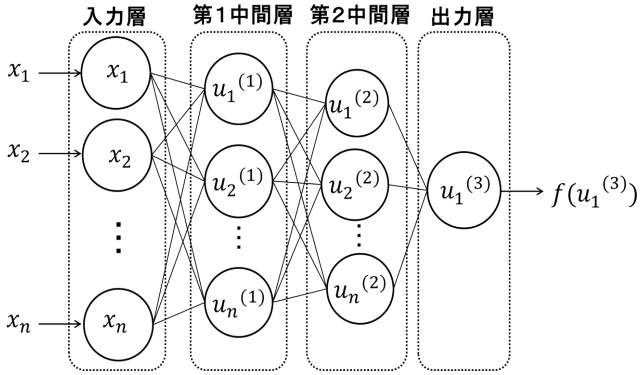

図 1　4 層の順伝播型ニューラルネットワーク

B. ランダムフォレスト

ランダムフォレスト [7] は，非線形モデルの回帰分析の 1 つであり，分類・回帰・クラスタリングに用いられる．多数の弱分類器として機能する決定木の出力を集約して予測を行う，アンサンブル学習法の 1 つである．

C. サポートベクターマシン

サポートベクターマシン [8] もまた，非線形モデルの回帰分析の 1 つであり，高次元の空間への非線形変換とその空間でのカーネルトリックという方法により，複雑な識別に対応可能である．マージン最大化を取り入れることで，少ない訓練データでも高い汎化性能が得られる．

本研究では，ランダムフォレスト及びサポートベクターマシンの検証を，R 言語（version 3.1.0）を用いて行った．R 言語は，オープンソース・フリーソフトウェアの統計解析向けのプログラミング言語及びその開発実行環境である．

III. 任官辞退者予測 DNN

ここでは，本研究で任官辞退者を予測するために用いたデータセット，DNN の設定及び予測精度の評価方法について述べる．

A. 学生のデータセット

本研究では，2014 年から 2017 年の 4 年間分のデータセットを用いる．データセットの内訳を表 1 に示す．1 人分のデータには，表 2 で示されるように所属，家族構成，学科成績などの 25 項目が含まれる． この訓練データは，各学生が任官辞退したかどうかの結果を含む．

表 2 の項目 1 の「所属」は，学生が所属する大隊・中隊・小隊の区分である．項目 2 の「要員」は，陸・海・空のいずれの要員かを示す．項目 3 の「専攻」は，14 学科ある学生の専攻学科である．項目 17 の「YG 性格検査」と項目 18 の「クレペリン検査」は，いずれも心理検査の結果である．

表 1　学生のデータセット内訳

	学生の人数
任官辞退	143
任官	1,579
合計	1,722

表 2　一人当たりの学生データが有する項目

No.	項目	No.	項目
1	所属	14	1 年次適性
2	要員	15	2 年次適性
3	専攻	16	3 年次適性
4	出身県	17	YG 性格検査
5	家族構成	18	クレペリン検査
6	浪人年数	19	航空適性
7	留年年数	20	所属運動部
8	1 年次学科成績	21	所属文化部
9	2 年次学科成績	22	所属同好会
10	3 年次学科成績	23	健康状態
11	1 年次訓練成績	24	体力級
12	2 年次訓練成績	25	任官・任官辞退
13	3 年次訓練成績		

B. DNN の構成

DNN では，パラメータを以下のように設定した．

- N_{in}: 入力層のユニット数
- N_n: 第 n 中間層のユニット数
- N_{out}：出力層のユニット数
- L：中間層数
- z：中間層の活性化関数
- f：出力層の活性化関数
- E：損失関数

ここで N_{in} は，学習データの項目から入力パラメータを作成した結果，259 の入力値となったため $N_{in} = 259$ に固定した．N_{out} は，出力値が 1（任官辞退）か 0（任官辞退しない）とした．活性化関数 z 及び f については，従来シグモイド関数が多く用いられて来たが，近年では，Yann Lecun らにより ReLU(rectified linear unit) が良いとされる報告がある [9]．そのため，今回の評価では両方の関数を比較対象とした．実際の値 d_n と予測された値 y_n の誤差を求める関数である損失関数 E については，出力値が 2 値分類となるため，式 (2) に示す最尤推定を使用した．

$$E(w) = -\sum_{n=1}^{N}(d_n \log y_n + (1 - d_n)\log(1 - y_n)) \quad (2)$$

C. 予測精度の評価

本研究では，10 分割交差検証を行った．つまり，1,722 人を 10 分割し，そのうち 1 つ（172 または 173 人分）をテストデータとし，残りを学習データとして検証を 10 回行い，予測精度を確認する．

DNN，ランダムフォレスト及び重回帰分析の予測精度は，F 値 及び正解率によって評価する．任官辞退するか否かの予測は 2 値分類問題であり，機械学習の予測結果と実際の結果に基づき表 3 のような混合行列により区分する．任官辞退した人数が任官した人数よりかなり少ない割合であるため，任官すると予測し実際任官した割合が大きくなるのは自然であり，正解率も高くなるのは明らかである. そのため，任官辞退者を正しく予測できているかを判断するには，F 値 に着目しなければならない.

F 値 は，*presicion* 及び *recall* から求められる．*precision* は，任官辞退すると予測された人数のうち，実際に任官辞退した人数の割合であり，*recall* は，実際に任官辞退した人数のうち，任官辞退すると予測された人数の割合である．F 値，*precision*，*recall* は式 (3)(4)(5) により求められる．

正解率は，10 分割交差検証により得られた，10 回分の予測結果の平均により求める．

表 3　混合行列

		実際	
		任官辞退	任官
予測	任官辞退	TP	FP
	任官	FN	TN

$$accuracy = \frac{TP + FP}{TP + FP + TN + FN} \quad (3)$$

$$prediction = \frac{TP}{TP + NP} \quad (4)$$

$$F\ 値 = \frac{2 \times\ accuracy \times\ prediction}{accuracy + prediction} \quad (5)$$

D. 評価結果

活性化関数 z，f 及び固定値をとらないパラメータとなる L，N_n を変更して検証を行い，最も予測精度が高くなるようなパラメータを求めた．

まず，z，f を，それぞれシグモイド関数を用いた場合と ReLU を用いた場合とを比較する．このときの他のパラメータは，$L = 1$，$N_1 = 150$ に固定した．結果は図 2 に示すように，シグモイド関数がより予測精度が高くなることを確認出来た．

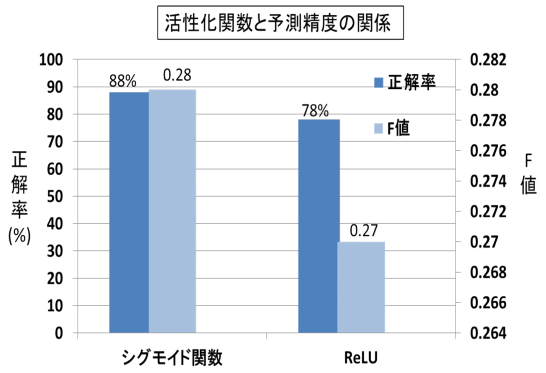

図 2　活性化関数と予測精度の関係

次に，L を，1 から 4 まで変更し，比較する．それぞれの L の設定において中間層のユニット数 N_n を各層いずれも同数として N_n を 100 から 300 まで 50 刻みで変更した場合を比較する．結果は図 3 に示されるように，$L = 2$，中間層のユニット数を各層いずれも 250 に設定した場合において，最も高い予測精度が得られた．

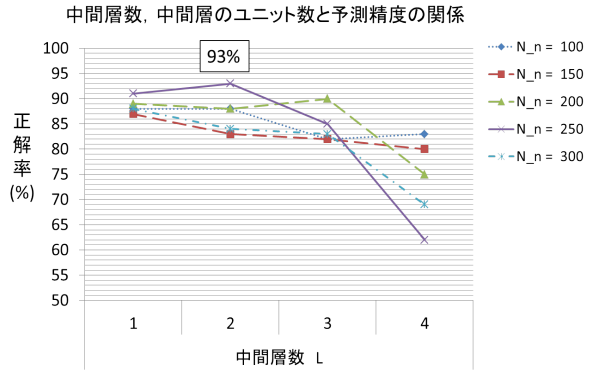

図 3　中間層数，中間層のユニット数と予測精度の関係

E. 他手法との比較

最終的に，最も予測精度の高かった DNN の結果と，ランダムフォレスト及びサポートベクターマシンの予測精度の比較を，表 4 に示す．また，DNN で最も予測精度が高くなったときの予測結果を表 5 に示す．

正解率で見た場合，3 手法とも正解率が 90%を上回る結果が得られたが，DNN はランダムフォレスト及びサポートベクターマシンに比較し，若干ながら上回った．F 値 で見た場合，DNN が最も高い結果となった．しかし，F 値については，任官者を任官辞退すると予測してしまう割合が大きく，まだ改善の余地があると考える．

表 4　3 手法による任官辞退者の予測精度

機械学習の手法	F 値	正解率
DNN	0.31	93.0%
ランダムフォレスト	0.29	90.6%
サポートベクターマシン	0.23	90.6%

表 5　任官辞退者予測 DNN の予測結果

		実際	
		任官辞退	任官
予測	任官辞退	32	33
	任官	111	1,546

IV. まとめ

防衛大学校の学生のデータを入力することで，その学生が任官辞退するか否かを予測する任官辞退者予測 DNN を提案し，その最適なパラメータを設定して予測精度を検証した．

本研究の結果，DNN の予測結果は F 値 が 0.31 であり，他の手法に比べて高くなったものの，まだ実用上不十分であると考える．F 値 が低い要因となる事項を探求し，予測精度の向上のために，データセット及び学習のアルゴリズムの再検討を要する．

参考文献

[1] 岡谷貴之：深層学習，講談社，pp.1-54, 2015.

[2] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” arXiv, 2014.

[3] F. Seide, G. Li, D. Yu, “Conversational Speech Transcription Using Context-Dependent Deep Neural Networks,” proc. Interspeech, pp.437-440, 2011.

[4] Y. Dauphin, A. Fan, M. Auli, David Grangier, “Language Modeling with Gated Convolutional Networks,” arXiv:1612.08083v1 [cs.CL], 2016.

[5] 脇田夕嘉, 奥健太, 川越恭二, “深層学習を用いたファッションブランドの推薦システムに向けて,” DEIM Forum, C7-6, 2016.

[6] D. Rumelhart, G. Hinton, R. Williams, “Learning representations by back-propagating errors,” Nature, 323, pp.533-536, 1986.

[7] L. Breiman, “Random forests,” Machine Learning, vol. 45, no. 1, pp. 5-32, 2001.

[8] 山下浩，田中茂：サポートベクターマシンとその応用 www.msi.co.jp/vmstudio/materials/svm.pdf, 2001.

[9] Y. Lecun, Y. Bengio, G. Hinton, “Deep learning,” Nature, 521, pp.436-444, 2015.