

学内模擬試験を用いた診療放射線技師国家試験の合否予測 - 統計的手法とニューラルネットワークの比較 -

相子真介^{†1} 渡辺博芳^{†2} 佐々木茂^{†2} 坂本重己^{†3}

帝京大学大学院理工学研究科^{†1} 帝京大学^{†2} 日本医療科学大学^{†3}

1. はじめに

医療系国家試験の合否予測は、教員の経験に基づいて直観的に行われていることが多い。これに対して、客観的な合否予測に基づいた指導ができると効果的であると考えられる。本研究では1つの学校での活用を想定して、学内模擬試験成績から国家試験合否予測手法の確立を目的とする。これまでに、ある学校の学内模擬試験成績および国家試験成績を用いて、マハラノビス距離による判別分析での合否予測を検討した^(1,2)。本稿ではさらに、ロジスティック回帰分析とニューラルネットワークを使った合否予測の検討を行ったので報告する。

2. 対象とする国家試験とデータ

診療放射線技師国家試験は毎年1回実施され、14科目から200問が出題される。合格基準は200点満点で120点以上、かつ0点の科目が1科目以下である。出題形式は五者択一のマークシート方式である。

実験で使うデータは、ある学校における2011年度と2013年度の2年分の学内模擬試験成績と国家試験成績である。学内模擬試験成績は科目毎の得点、国家試験成績は自己採点による総合点と正式な合否が準備されている。各分析で使う属性は「14科目偏差値」と、模擬試験実施時期の指標となる「国家試験までの日数」とする。ただし、国家試験の自己採点による合否判定と実際の合否が異なるデータは分析の対象外とした。なおデータ提供元の学校や学生個人に関する情報が特定できないように配慮してデータを扱っている。

以降の判別分析、ロジスティック回帰分析、ニューラルネットワークは統計ソフトR⁽³⁾を使用した。

3. 判別分析を使った合否予測

これまでの検討⁽²⁾において、判別分析を用いる際には、あらかじめ国家試験の自己採点による総合点を基に複数の群に分けることが有効であることが分かった。そこで、合否予測精度の高かった4群分け(合格群を2群、不合格を2群)で判別分析を行う。

判別の中率の算出は、マハラノビス距離を使った判別分析で群予測を行う。予測した群が合否のどちらに対応しているかを基に合否予測を行い、実際の合否と比較し、正しく予測できた回数をカウントし合否予測精度を求める。以上の処理を他年度の成績を使って当該年度の合否予測を行う。

4. ロジスティック回帰分析を使った合否予測

片方の年度の全データを使って、(1)式からの変数 β_0, \dots, β_n を求める。次に予測対象年度のデータをサンプル毎に(1)式の x_1, \dots, x_n に代入し、国家試験合格の確率 y を求める。この値が0.5以上ならば合格と予測し、0.5未満ならば不合格と予測して、これを実際の合否と比較する。正しく予測できた回数をカウントし、合否予測精度を求める。

$$y = \frac{1}{1 + \exp\{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)\}} \quad \dots(1)$$

5. ニューラルネットワークを使った合否予測

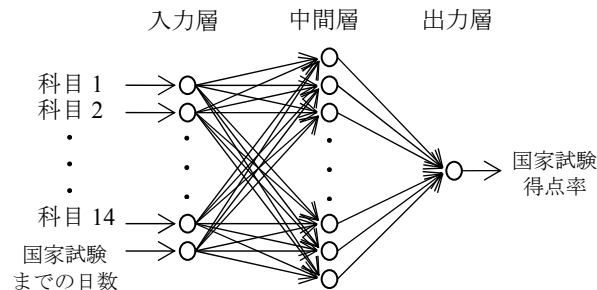


図1 フィードフォワード型モデルを使った合否予測

図1はニューラルネットワークの実験モデルである。モデルの設定パラメータは中間層のユニット数30、最大学習回数1000回、衰退重み0.05とした。14科目成績と国家試験までの日数を入力層に与えると、出力層から国家試験総合点の得点率が出力される。この値が0.6以上ならば合格予測、0.6未満ならば不合格予測とする。

しかし、ニューラルネットワークの出力値は学習状況により変動する。そこで、複数のニューラルネットワークを使った実験を行う。各サンプルにおいて、各ニューラルネットワークから出力された合否予測を、多数決方式により最終的に一つの合否予測として決定する。合格予測数が不合格予測数より大きければ合格予測として決定する。この合否予測

Study on Pass-Fail Prediction for the Radiological Technologist National Examination using Mock Examinations - Comparison between Statistical Methods and Neural Network -

^{†1} Shinsuke Aiko, Graduate School of Science and Engineering Teikyo University

^{†2} Hiroyoshi Watanabe, Shigeru Sasaki, Teikyo University

^{†3} Shigemitsu Sakamoto, Nihon Institute of Medical Science

と実際の合否を比較して、正しく予測できた回数をカウントし合否予測精度を求める。

図2はニューラルネットワークの数による予測精度の比較結果を示している。この結果から、ニューラルネットワークが1個では、合否予測精度が非常に低いことがわかる。出力値が合否を分ける境界付近の値であると予測精度が不安定になり低下するからだと考えられる。ニューラルネットワークを10個にすると9.0~13.8%と大幅に精度が高くなり、さらに20個にすると0.5~0.6%高くなった。20個を超えると値はほとんど変化しなかった。この傾向から、20個のニューラルネットワークを使った合否予測をすることにした。

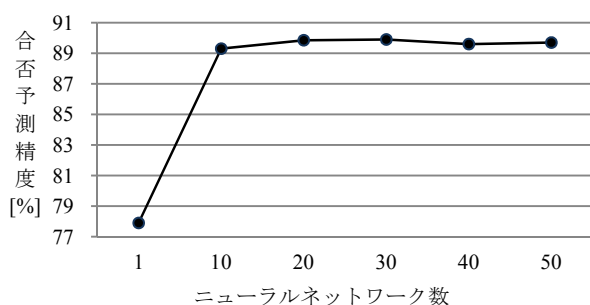


図2 ニューラルネットワークの数による合否予測精度

6. 経験的知識の利用

国家試験受験を指導する教員が経験的に受験者の合否を予測していた視点に基づいて、判定精度の高い判定ルールを定義を試みた。教員がこの条件を満たせば合格するであろうと予測していた条件を基に、以下の判定ルールを抽出した。

- (1) 第*i*回模擬試験以前の模擬試験総合得点が120点(合格点)以上の数が n_i 以上ならば合格と予測する。
- (2) 模擬試験における科目*j*の得点が1回でも x_j 以上であれば合格と予測する。

データを分析して、判定ルールでほぼ正確に合格判定ができるように、模擬試験の回数*i*毎の n_i および科目*j*毎の合格点 x_j を決定した。

これらの判定ルールを各手法と組み合わせて使用する。まず、予測対象のサンプルが判定ルール(1)または(2)のいずれかを一つ以上満たしていれば合格と予測する。判定ルールで合格と判定できなかったサンプルのみを、各手法によって予測する。

7. 分析手法による実験結果

分析手法別の実験結果を表1に示す。各分析手法での合否予測精度は、4(2+2)群分けによる判別分析(DA)87.9~89.4%、ロジスティック回帰分析(LR)89.6~90.2%、20個のニューラルネットワークモデル(NN)89.2~90.0%となった。

さらに、3つの手法に経験的知識により定義した

判定ルールを適用すると、判別分析で0.5ポイント、ロジスティック回帰分析で0.2ポイント向上している。判定ルールによる効果は、判定ルール無しで予測精度が低い場合に対して高かった。

表1 分析手法による実験結果 [%]

年度	DA	LR	NN	R+ DA	R+ LR	R+ NN
2011 →2013	87.9	90.2	89.2	88.4	90.2	89.2
2013 →2011	89.4	89.6	90.0	89.4	89.8	90.0

DA：判別分析4群(合格2群と不合格2群)分け

LR：ロジスティック回帰分析

NN：ニューラルネットワーク20個のモデル

R+：判定ルール適用

2011→2013は2011年度を基に2013年度を予測

8. 考察

模擬試験の14科目偏差値と、国家試験までの日数を属性として使った場合、どの手法を用いても合否予測精度に大差はなかった。さらに経験的知識を用いて定義した判定ルールを適用した結果、どの手法においても予測精度が90%近い精度となった。以上の結果から、本研究で対象とした2年分のデータセットに対する合否予測精度は90%前後に限界があると考えられる。これらの手法の中で最も安定して高い精度が得られたのは判定ルールを用いたロジスティック回帰分析で89.8~90.2%あった。

9. おわりに

本研究では2年分のデータを使い学内模擬試験成績から国家試験合否予測の手法の検討を行った結果、どの手法でも90%近い合否予測精度が得られた。

今後の課題は、多年の成績データを使って同様の検証を行うこと、年度毎の国家試験難易度を予測して組込む手法を検討すること、また合否予測結果を学生や指導者にどのような形で提供するかを検討することがあげられる。

参考文献

- (1) 相子真介, 佐々木茂, 渡辺博芳, 坂本重己: “学内模擬試験を用いた判別分析による診療放射線技師国家試験の合否予測の検討”, 情報処理学会研究報告, Vol.2014-CLE-13, No.7, pp.1-6 (2014).
- (2) 相子真介, 佐々木茂, 渡辺博芳, 坂本重己: “学内模擬試験を用いた判別分析による診療放射線技師国家試験の合否予測 -2年分の成績データを用いた検討-”, 教育システム情報学会第39回全国大会講演論文集, D4-3, pp.379-380 (2014).
- (3) 金明哲: “Rによるデータサイエンス データ解析の基礎から最新手法まで”, 森北出版株式会社, pp.148-255 (2007).