所要時間のためのガウス過程を用いた Multi-Task Deep Neural Network により制限時間を考慮する自動テスト構成

Automated test assembly for time limitation using Multi-Task Deep Neural Network as Gaussian Process for prediction of item response time

石川 文弥[*1]，渕本 壱真[*1]，植野 真臣[*1]
Fumiya ISHIKAWA[*1], Kazuma FUCHIMOTO[*1], Maomi UENO[*1]
[*1] 電気通信大学
[*1]The University of Electro-Communications
Email: ishikawa@ai.lab.uec.ac.jp

あらまし：e テスティングでは，異なる問題項目で構成されるにもかかわらず，同一精度の測定を実現できる．e テスティングでは，可能な限り多くのテストが必要である．しかし，現在最も多くのテストを構成できる手法は制限時間を考慮しておらず，受検者の能力推定値の推定誤差が増加する．本研究では，Multi-Task Deep Neural Network による受検者の各項目への所要時間予測手法を開発し，制限時間のための自動テスト構成手法を提案する．提案手法は 95%以上の受検者が制限時間内に全ての項目を回答でき，受検者の能力推定値の推定誤差が他手法より小さく等質なテストを生成できた．
キーワード：**e テスティング**，**深層学習**，**マルチタスク学習**，**整数計画法**

1. はじめに

e テスティングとは，異なる問題項目で構成されるにもかかわらず，同一精度の測定を実現できるテストである．e テスティングでは複数回受検が可能となるなど多くの利点がある．

e テスティングでは，可能な限り多くのテストが必要である．Fuchimoto et al.は，現在最も多くのテストを構成可能な最大クリーク問題と整数計画法を用いた二段階並列探索手法(1)を提案している．しかし，この手法はテストの制約時間を考慮していないため，受検者が時間内に全問題項目を解けず，受検者の能力値推定値の推定誤差が増加する．

この問題を解決するために，本研究では，Multi-Task Deep Neural Network(2)を用いて正誤反応と所要時間を同時に予測する手法を提案する．さらに，本研究では提案手法をガウス過程(3)による各予測値の不確かさが得られるモデルに拡張する．この手法を用いて，95%以上の受検者が制限時間内に全問題項目を解き終える制約式を提案する．その結果，提案手法は受検者の能力推定値の推定誤差が他手法よりも小さく，95%以上の受検者が制限時間内に全ての項目を回答できるテストを生成できた．

2. 所要時間予測のための Multi-Task DNN

本研究では，Multi-Task Deep Neural Network(以降，MTDNN)から着想を得て，受検者の各項目への正誤反応と所要時間を同時に予測する手法を提案する．図 1 は,提案手法の概要図である．

入力ベクトル$\boldsymbol{x}$は入力層$\boldsymbol{L}_{input}$を通り独立したE個のフィードフォワード・ニューラルネットワーク(FF・NN)層に入力される．

$$\boldsymbol{L}_{input} = leaky_relu(\boldsymbol{W}_{\boldsymbol{L}_{input}}\boldsymbol{x} + \boldsymbol{b}_{\boldsymbol{L}_{input}}),$$

$$\boldsymbol{L}_1^e = leaky_relu\left(\boldsymbol{W}_{\boldsymbol{L}_1^e}\boldsymbol{L}_{input} + \boldsymbol{b}_{\boldsymbol{L}_1^e}\right), (e = \{1, 2, \ \ldots, E\})$$

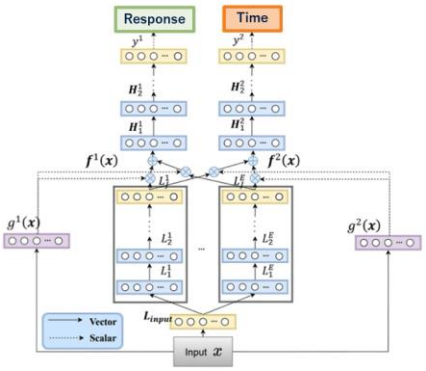

図 1. 所要時間予測のための Multi-Task DNN

$$\boldsymbol{L}_j^e = leaky_relu(\boldsymbol{W}_{\boldsymbol{L}_j^e}\boldsymbol{L}_{j-1}^e + \boldsymbol{b}_{\boldsymbol{L}_j^e}), (j = \{2, 3, \ \ldots, J\}),$$

$$leaky_relu(z) = \begin{cases} z, & z > 0 \\ \gamma z, & z \leq 0 \end{cases}.$$

ここで，$\boldsymbol{W}$は NN の重みパタメータ，$\boldsymbol{b}$はNN のバイアスパラメータを表す．ただし，γはチューニングパラメータである．

その後，以下のゲート付きネットワークを用いてタスクに応じて各下層 FF・NN の結果に異なる重みを割り当てられ，タスクkを予測するための独立した上層 FF・NN に入力される．

$$g^k(\boldsymbol{x}) = softmax\left(\boldsymbol{W}_{g^k}\boldsymbol{x} + \boldsymbol{b}_{g^k}\right),$$

$$f^k(\boldsymbol{x}) = \sum_{e=1}^{E} g^k(\boldsymbol{x})\, \boldsymbol{L}_J^e,$$

$$\boldsymbol{H}_1^k = leaky_relu(\boldsymbol{W}_{H_1^k} f^k(\boldsymbol{x}) + \boldsymbol{b}_{H_1^k}),$$

$$\boldsymbol{H}_m^k = leaky_relu(\boldsymbol{W}_{H_m^k}\boldsymbol{H}_{m-1}^k + \boldsymbol{b}_{H_m^k}), (m = \{2, 3, \ \ldots, M\}).$$

最後に，タスク固有の予測値は以下から得られる．

$$y^k = leaky_relu(\boldsymbol{W}_{y^k}\boldsymbol{H}_M^k + \boldsymbol{b}_{y^k}).$$

ただし，$k=1$は正誤反応の予測タスク，k = 2は所要時間の予測タスクである．さらに，本研究では予測所要時間の不確かさも考慮するために，出力層をガウス過程に変更した．これにより，出力で以下の予測値の分布が得られるようになる．

$$f^{y^k} \sim GP\left(0,\ K(\boldsymbol{H}_M^k, \boldsymbol{H}_M^k)\right),$$

$$k(x, x') = \sigma^2 \exp(-\omega(x - x')^2/2).$$

この手法を MTDNN-GP と呼ぶ．

3.　所要時間を考慮した自動並行テスト構成

本研究では，提案手法の予測所要時間を用いて二段階並索手法の整数計画法に以下の制限時間の制約を追加する．提案手法(MTDNN)，提案手法(MTDNN as GP)はそれぞれ以下の通りである．

$$LB_{TL} \le \sum_{i=1}^{n} t_{hi} x_i \le UB_{TL}.$$
$$LB_{TL} \le \sum_{i=1}^{n} (t_{hi} - d_{t_{hi}}) x_i,$$
$$\sum_{i=1}^{n} (t_{hi} + std_{t_{hi}}) x_i \le UB_{TL}.$$

ここで，LB_{TL}とUB_{TL}はそれぞれテストの所要時間の下限と上限を表し，t_{hi}は$h(=1,2,\dots,H)$番目のカテゴリの能力値を持つ受検者が項目iに要する予測所要時間を表す．また，$std_{t_{hi}}$はh番目のカテゴリの能力値を持つ受検者が項目iに要する所要時間の予測分布の標準偏差を表している．これにより，提案手法は全カテゴリ内の能力値を持つ受検者が，制限時間内に全問題項目を解き終えるテストを生成できる．

4.　比較実験

4.1 所要時間予測手法の精度比較

ここでは，提案手法と従来手法(4) (5)の所要時間の予測精度を比較した．比較実験には，本学における入学時 CBT 学力調査試験の結果を用いた．科目は数学，アイテムバンクの問題数は 100 問，受検者数は 666 人である．各手法による予測結果を表 2 に示す．ただし，RMSE は受検者iの項目jへの実所要時間t_{ij}と予測所要時間 $\hat{t}_{ij}$を用いて以下の式で表す．

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(t_{ij} - \hat{t}_{ij}\right)^2}.$$

表 2 より，提案手法は従来手法よりも RMSE が小さく，所要時間の予測精度を向上できた．

表 2　各手法における予測所要時間の RMSE(秒)

Linden 2007	岸田 2024	MTDNN	MTDNN-GP
163.60	113.16	110.11	109.24

4.2 自動並行テスト構成手法の比較

ここでは，提案手法と従来手法(2) (6)により構成したテストを受検者にランダムに割り当て，テストの各項目に対する実データの所要時間から回答をシミュレーションした．ただし，実データがない場合は，階層モデル(4)の推定値を使用した．また，各受検者は，合計所要時間が制限時間（60 分）を超えた時点で回答終了とした．

テスト構成条件は，所要時間の上限が 60 分，下限値が 36 分とした．さらに，本研究では，受検者全体の 95%が回答終了するテストを構成するために，能力値のカテゴリは標準正規分布の下位 95%の値である-1.645 から 2.000 を五等分した値に設定した．

表 3 は，各受検者が受検したテストの所要時間（分）の平均（Avg. time）と標準偏差（SD.time），制限時間内に回答終了した受検者の割合（Rate），各受検者の能力推定値の平均推定誤差（Avg.SE）を示す．表 3 より，制限時間を考慮していない HMCAPIP 法を除く全ての手法で制限時間内に回答終了した受検者の割合を 95%以上にすることができた．さらに，Linden MIP は全受検者が制限時間内に回答終了できたが，テストの平均所要時間が短く，受検者の能力推定値の平均誤差は一番大きかった．一方で，提案手法は能力推定値の平均推定誤差を一番小さくできた．

図 1 は，左の図が経過時間ごとの回答終了した受検者数を表し，右の図が受検者の能力推定値とその推定誤差の散布図を表す．図 1 より，提案手法(MTDNN-GP)は提案手法(MTDNN)よりも能力推定値の推定誤差は悪化したが，テストの所要時間の散らばりを抑え，制限時間内に回答終了した受検者の割合を向上できた．

表 3　各手法によるテストの受検結果

手法	Avg.Time	SD.time	Rate (%)	Avg.SE
Linden MIP	**28.9**	**3.66**	**100**	0.39
HMCAPIP	50.8	7.49	82.1	0.16
MTDNN	47.6	6.37	97.0	**0.14**
MTDNN-GP	44.1	5.87	98.3	0.18

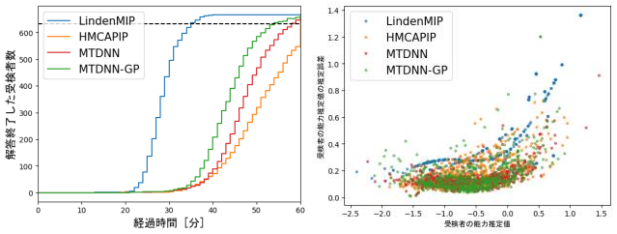

図 1. 経過時間ごとの回答終了した受検者数と受検者の能力推定値の推定誤差

5.　むすび

今後は，適応型テストへの拡張を考える．

参考文献

(1) Fuchimoto, K., Ishii, T., and Ueno, M.: Hy brid maximum clique algorithm using parallel integer programming for uniform test assembly, IEEE Transac tions on Learning Technologies, Vol. 15, No. 2, pp. 252 264 (2022)

(2) Ma, Jiaqi, et al. "Modeling task relationships in multi-task learning with multi-gate mixture-of-experts." *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*. 2018.

(3) Damianou, Andreas, and Neil D. Lawrence. "Deep gaussian processes." *Artificial intelligence and statistics*. PMLR, 2013.

(4) Linden, Wim J. van der. :“A Hierarchical Framework for Modeling Speed and Accuracy on Test Items.”, *Psychometrika* 72 (3): 287–308. (2007)

(5) 岸田若葉，堤瑛美子 and 植野真臣．″解答時間予測のための Deep-IRT.″ 人工知能学会全国大会論文集 (2024)

(6) van der Linden, Wim J. "Test design and speededness." *Journal of Educational Measurement* 48.1 (2011): 44-60.