2023 年 2 月 7 日

加藤侑也

ニューラルネットワークを用いた離散的シークエンスの予測に関する研究

松香研究室 4 年 加藤侑也 19L1026A

---

1. 序論

離散的シークエンスのデータは，離散型のデータを時系列で集めたもので， 人間や物事の行動を予測することに多く使用されている。その 1 つとして，人々の投票行動が考えられる。この投票行動に焦点を当てることで，人間の思考の仕組みを理解することを本研究の目的とした。具体的には，離散的シークエンスとして，サザエさんとめざましジャンケンを用いて離散型のデータの分析を行った。

先行研究においては，サザエさんのデータでは， derodero (2018) がニューラルネットワークを用い 58%の正答率を得ることを示した。everytitle (2018) では，同様の手法であるのの，モデルを複雑化することにより（多くの変数を用い），正答率の向上は見られなかった。めざましジャンケンでは，1 週間単位の期間で M.Tanaka (2023) が，特定の期間ではあるものの最高で 50%の正答率を示した。

本研究は，特定の指標を用いて投票行動を行う人に焦点を当て，投票行動者の思考や挙動を予測することが可能であるかを検討した。投票行動者をジャンケン作成者とし，ニューラルネットワークを用いて 6 つの分析を行い，作成者が出す手の予測精度の検討を行った。

2. データの前処理

R で解析を行った。またジャンケンに応じて変数を変更した。

・サザエさんジャンケン 分析 1 から分析 5 まで使用。

年，月，日，週，遡り回数(直近 n 回の手)，各手の出ていない週数 を用いた。

・めざましジャンケン 分析 5 から分析 6 まで使用

年，月，日，曜日，当日回数，当日に出された手を用いた。

3. <u>分析 1 (中間層次元数)</u>

分析 1 は，プログラムに用いた中間層の次元数によって生じる正答率の影響の考察を目的とした。中間層の次元数は分析の実行時間に大きな影響を与えるため，最適な次元数を選出するための検証を行った。

- 方法

変数は上記の変数(n=4)を全て使用した。データはサザエさんを使用した。データ分割は 75%を訓練データ，25%をテストデータに割り振った。比較対象は次元数で 3 次元から 18 次元まで 3 次元ずつだった。この時，遡り回数は 4 回とした。

- 結果・考察

図 1 は中間層次元数による正答率の変化を示した。3 次元が最も中央値が高く 55%だった。最大値や最小値は 6 次元が高い。一方，15 次元以上の高次元は正答率が低かった。

このことから，低次元でも相対的に高い予測が行えることが示された。一方，高次元は過剰一般化が起き，訓練データに適合し過ぎたため，予測的中率が下がったと考えられた。

4. 分析 2 (遡り回数)

分析 2 は，直近回数の結果のみを変数として用い，予測した場合，直近回数の変動が正答率に与える影響の考察を目的とした。先行研究は，直近 3 回の結果から予測したが，回数を増やすことで参考回数が増えるため，正答率が上昇すると考え，分析 2 を行った。

この分析では直近回数を遡り回数と表現した。

- 方法

変数は遡り回数のみにした。また，中間層は 3 次元で固定した。

- 結果・考察

図 2 は遡り回数による正答率の変化を示した。遡り回数が 4 回までは正答率の中央値，最大値，最小値が全て上昇した。4 回の中央値は 55%だった。7 回以上は全項目で減少した。さらに，遡り回数が 1 回の場合は全項目で 33%を割り込み，チャンスレベル未満だった。

分析 1 と 2 で，分析 1 は多くの変数(遡り 4 回含む)+中間層 3 次元，分析 2 は遡り 4 回+中間層 3 次元だったが，3 項目とも差がなかった。他の変数が予測に影響を与えないことが示された。

図 1　中間層次元数による正答率の変化

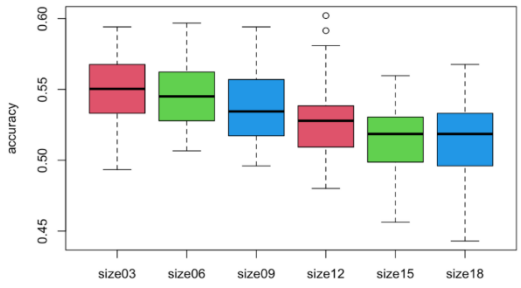
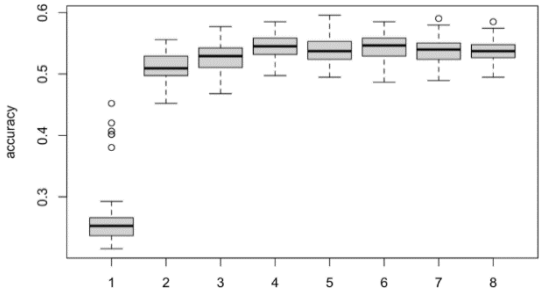


図 2　遡り回数による正答率の変化



５．分析 3(傾向変化)

分析 3 は，一定の区間を訓練データとして用い，他の区間に出された手を予測した際の正答率を各年，各区切りで比較した。

- 方法

訓練データを 1 年，2 年間とし，テストデータを他の 1 年，2 年間として，各期間に対して分析を行った。変数は分析 1 と同様で，中間層は 3 次元とした。主に，次の区間の予測的中率を比較対象とした。また，全部の期間を予測した際の精度を平均正答率とした。

- 結果・考察

平均正答率で，1992 年から 1998 年が 40%後半，1999 年から 2007 年で 40%前半，2008 年から 2019 年で 50%前後，2020 年から 2021 年が 45%だった。正答率の変動から 4 つの期間

に分けられた。直近の期間で傾向が変わっており，既知の予測では対応出来ないと考えた。

6. 分析 4(近年の傾向)

分析 3 は，2020 年以降の正答率が他年と比べ減少した。データ範囲を狭くすることで，現期間の予測精度が高まると想定し分析 4 を行なった。データ範囲を 2020 年以降に限定し分析 1 と 2 と同様の分析を行った。

- 結果・考察

図 3 は，データを 2020 年以降の期間にした場合，中間層次元数による正答率の変化を示した。中間層は 3 次元が最も良く，遡り回数は 3 回まで的中率が上昇した。的中率は 50%付近であり，分析 3 の正答率を上回った。傾向が変化した場合，新データを使用することで正答率が上がると考えた。

7. 分析 5(めざましジャンケン)

分析 5 はデータをめざましジャンケンに変更して行った。同一日に複数回実施されるなどサザエさんと異なり，違う傾向・周期がある可能性を想定し分析を行なった。

本分析では，めざましジャンケンにデータを変更し，分析 1 と 2 と同様の分析を行った。めざましジャンケンの傾向と，サザエさんの傾向の比較を目的とした。変数は同一とした。

- 結果・考察

図 4 は，データをめざましジャンケンに変更した場合，中間層次元数による正答率の変化を示した。また，図 5 は，めざましジャンケンの場合，遡り回数による正答率の変化を示した。分析 1 で的中率は 3 次元が最も良く，44%だった。2 番目に 12 次元の正答率が高かった。分析 2 では遡り回数は 4 回付近が最も良かった。正答率は 45%付近だった。

結果より，めざましジャンケンの正答率は，サザエさんよりも低かった。また，遡り回数 1 回の場合，チャンスレベルよりも高い正答率が得られた。

図 3 2020 年以降の次元数 正答率変化

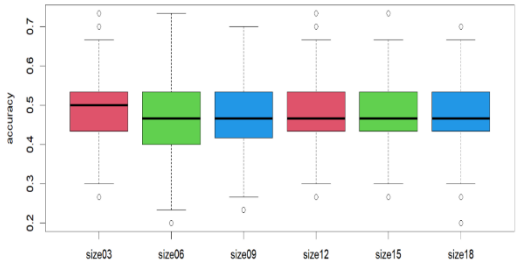
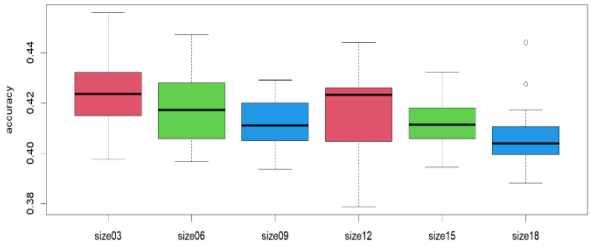


図 4 めざましジャンケンの次元数 正答率変化



2023 年 2 月 7 日
加藤侑也

8.　分析 6(めざましジャンケン 遡り回数 1 回)

分析 5 は，遡り 1 回の場合，チャンスレベルより高い正答率だった。サザエさんじゃんけんと異なる結果を得た。そのため，遡り回数 1 回の分析をより詳細に検討した。

分析 6 は，遡り回数 1 回の場合，中間層次元数の影響を検討した。中間層の次元数による正答率の影響が以前の研究と比べ，異なる可能性があると考えられる。

- 結果・考察

図 6 は，めざましジャンケンで遡り回数 1 回の場合，中間層次元数による正答率の変化を示した。遡り回数 1 回では，18 次元の成績が最も良く，次元を大きくすると，成績が向上した。分析 6 では分析 5 と比較し，遡り回数 1 回×3 次元の正答率が 1.7 ポイント上昇した。新たに導入した変数により中央値が上昇したと考えられる。

めざましジャンケンは，曜日や回戦数など特有の変数が予測に大きい影響を与えていると考えられる。1 日に複数回かつ曜日も複数あり，収録と生放送が混ざるシステムである。よってジャンケンの製作者は，一定のパターンを曜日と回戦ごとに決めておき，手の回数の偏りを減らすことを試みた可能性が考えられる。また収録時に予め手を決定しておく必要性から，事前に用意したパターンを使用することは合理的であるとも考えられる。そのため，遡り回数 1 回は，1 日の中で前回の手を参考にできるため，成績が向上したと考える。

図 5 めざまし 遡り回数 正答率の変化

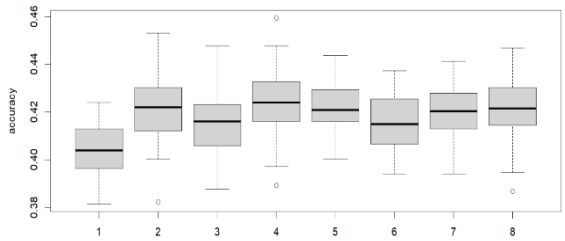
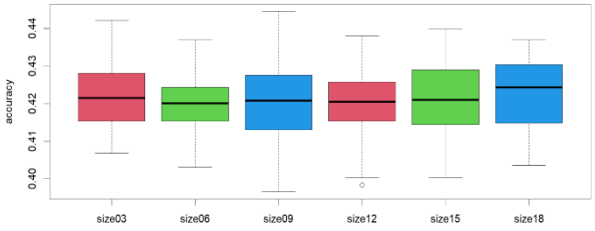


図 6 めざまし遡り 1 回 中間層次元数 正答率



9.今後の展望

まず，ニューラルネットワークで用いた変数の効果を検証する必要がある。サザエさんでは，遡り回数が予測精度に影響を与えた。一方で，めざましは，当日の手と曜日，当日回数が影響を与えたと考える。しかし，他の変数が影響を与え，現変数は二次的に影響を与えたとも考えられる。他の変数を導入することで，現変数の効果を検証する必要があると考える。

また，本研究は，正答率を高めることで，投票行動者の思考を導くと考え分析を行なった。担当者は，前任者の影響や自身の以前の行動を踏まえ投票行動を行うと考えた。これは他分野の行動予測に応用できると考える。人間は，自身の行動を振り返り行動を決め，他者の行動に影響を受け行動しやすい。よって本研究を発展・拡張することで，人間の行動予測の理解が促進することが期待される。