

麻雀における複数のニューラルネットワークの教師あり学習と対戦による評価

1225127 藤田 竜貴 【高度プログラミング研究室】

Supervised Learning of Multiple Neural Networks for Mahjong and Their Evaluation by Competition.

1225127 FUJITA, Ryuki 【High-Level Programming Lab.】

1 はじめに

麻雀においてディープラーニングを用いたコンピュータプレイヤーの実装が近年なされてきている。これまで多くの研究ではコンピュータプレイヤーの評価に判定データとの着手一致率が用いられており、GaoらはConvolutional Neural Network (CNN) を用いて 68.8% [1], 著者らの先行研究では Densely-connected Multi-Layer Perceptron (DenseMLP) を用いて 71.01% [2] という結果であった。しかし、着手一致率（もしくは学習における loss の値）が良いことが必ずしも実際の対戦で強いことと結びつかず、対戦させた結果がどうなるかは大きな疑問と言われてきた。

そこで本稿では、まず、新たな入力セットを提案し MLP, DenseMLP, CNN を用いて各モデルに教師あり学習させた結果を示す。そして、教師あり学習によって得られたプレイヤー同士でルールを簡易化した対戦実験を行なった結果を示す。

2 ニューラルネットワークの構造

図1に実験に用いた DenseMLP の構造図を、図2に CNN を用いたニューラルネットワークの概念図を、図3に図2中の畳み込み部分の構造を示す。

図1は中間層が4層の場合の図である。各矢印の始点の層と終点の層は各ユニットが密に接続していることを示し、矢印の終点が重なっている部分はそれぞれの出力が結合されていることを示している。図2はニューラルネットワークが萬子, 筒子, 索子, 字牌, その他の5種類の入力を受け取り、萬子, 筒子, 索子, 字牌の情報については3層の畳み込み層で処理を行い、その後、全ての特徴量は結合され MLP の入力として渡されることを示している。萬子, 筒子, 索子についてはそれぞれの牌の色を変換しても手牌の価値はほぼ等価となり、対称性があると言える。よってこれらの情報の処理にはフィルタを共有することで学習効率の向上を図る。

3 教師あり学習

学習は、ニューラルネットワークがある局面において訓練データと同じ打牌を行うネットワークを獲得する方針で学習を行う。訓練データ, テストデータにはネッ

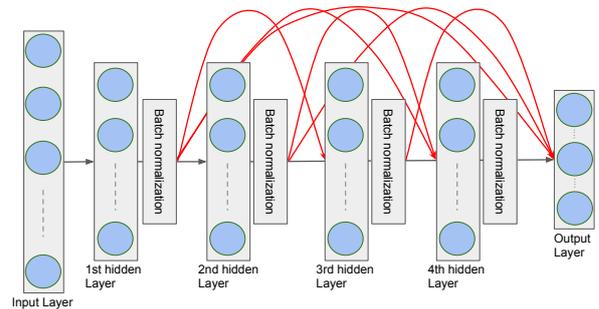


図1 DenseMLP の構造

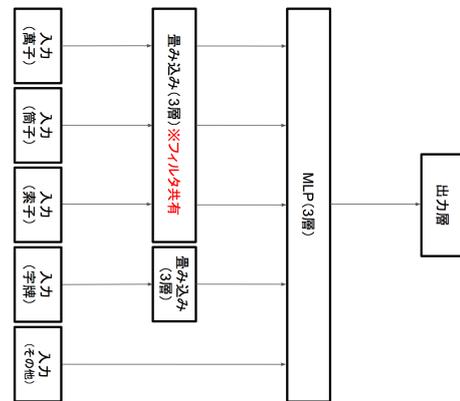


図2 CNN を用いたモデルの概念図

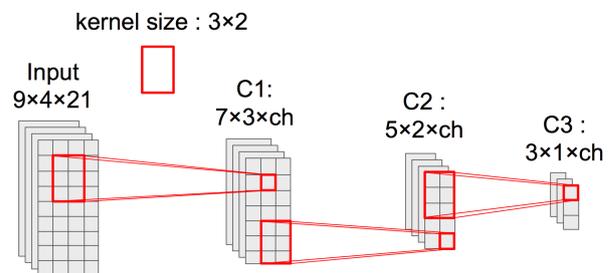


図3 畳み込み部分

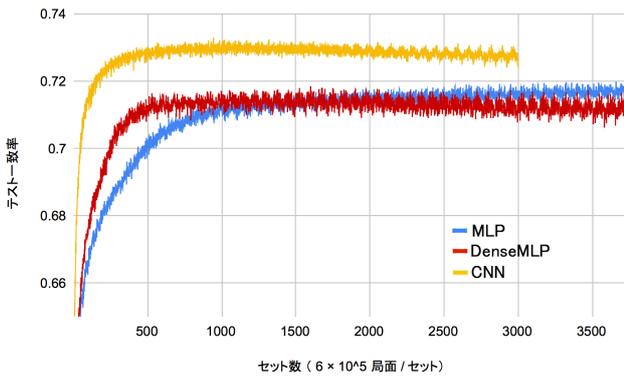


図 4 各モデルのテスト一致率

表 1 各モデルの最高テスト一致率と最低 loss 値

モデル	最高一致率	最低 loss 値
MLP (5 層)	72.00%	0.74560
DenseMLP (5 層)	71.85%	0.74974
CNN (ch 数 200)	73.29%	0.71443

ト麻雀天鳳 [3] の 2015 年鳳凰卓東南戦の牌譜を用いた。最適化アルゴリズムには Adam を使用し、バッチサイズは 5000 とした。学習局面数は 4.5×10^7 局面とし、1 セット 6.0×10^5 局面として、MLP と DenseMLP の場合は 3750 セット、CNN の場合は 3000 セットとした。MLP と DenseMLP については各中間層のユニット数を 512 に固定し、中間層の数を変えて実験を行なった結果、どちらも中間層 5 層の場合で最大テスト一致率と最低 loss 値を記録した。また、CNN は畳み込み部分のチャンネル数を変えて実験を行なった結果、チャンネル数 200 の場合で最大テスト一致率と最低 loss 値を記録した。図 4 にテスト一致率のグラフを、表 1 に最高テスト一致率と最低 loss 値を示す。

4 対戦実験

3 節で示した DenseMLP で最低 loss 値を記録したプレイヤー (Dns) と CNN で最低 loss 値を記録したプレイヤー (CNN) を対戦させた。また、DenseMLP の最低 loss 値と同程度の loss 値を記録した CNN プレイヤ (CNNfit) を用意し、対戦させた。各プレイヤーの loss 値、テスト一致率を表 2 に、それぞれの対戦結果を表 3、表 4 に示す。

本節で行なった対戦実験は、ルールを簡易化して行なった。各プレイヤーはテンパイしたら即リーチを宣言する。また、ポンやチーなどの副露行為はなしとする。それ以外のルールはネット麻雀天鳳 [3] に準拠する。

対戦は 2 種類のプレイヤーを 2 人ずつ用意し、同じプレイヤーを対面同士に設定し、席順の影響を受けないようにした。対戦半荘数は 5 万半荘に設定した。ポイントはネット麻雀天鳳の 7 段配分 (1 位 90pt, 2 位 45pt, 3 位 0pt, 4 位 -135pt) でポイントをつけるように設定した。得点はウマやオカなど設定せず、開始時の得点との差のみを計算している。

表 2 各プレイヤーの loss 値, テスト一致率, 学習セット数

モデル	loss 値	テスト一致率	学習セット数
Dns	0.74974	71.50%	1074
CNNfit	0.74974	71.88%	139
CNN	0.71443	72.90%	1104

表 3 Dns と CNN の対戦結果

	CNN ₁	Dns ₁	CNN ₂	Dns ₂
平均順位	2.4661	2.5302	2.4686	2.5352
平均 pt	2.771	-2.525	2.701	-2.948
平均得点	430.68	-379.8	404.52	-455.39

表 4 Dns と CNNfit の対戦結果

	CNNfit ₁	Dns ₁	CNNfit ₂	Dns ₂
平均順位	2.4894	2.5160	2.4896	2.5050
平均 pt	1.319	-1.918	1.544	-0.946
平均得点	84.37	-149.76	82.19	-16.81

5 まとめ

3 節の実験では、特徴抽出に CNN を用いたプレイヤーが他のプレイヤーより、loss 値、テスト一致率共に良い記録を残し、従来の最高テスト一致率 71.01% [2] を 2.28% 上回る 73.29% を記録した。また、4 節で行なった実験では、表 3 を見ると loss, 一致率共に良い CNN が勝ち越している。しかし、表 4 の実験を見ると、Dns も CNNfit も loss, 一致率の点ではほぼ同等のプレイヤーで、従来の評価方法ではどちらのプレイヤーが良いとも言い難いプレイヤー同士の対戦にも関わらず、CNNfit が勝ち越しているという興味深い結果を得た。考察として、CNN による特徴抽出の方が麻雀のゲームの特性に適した特徴抽出が行えているのではないかと予想する。展望として、この予想の検証が望まれる。

参考文献

- [1] S.Gao, F.Okuya, Y.Kawahara, Y.Tsuruoka : Supervised Learning of Imperfect Information Data in the Game of Mahjong via Deep Convolutional Neural Networks, ゲームプログラミングワークショップ 2017 論文集, pp.163-170 (2017).
- [2] 藤田竜貴, 松崎公紀 : 麻雀の打牌選択学習における Densely Multi Layer Perceptron の有効性と局面から獲得可能な特徴量の有効性, ゲームプログラミングワークショップ 2019 論文集, pp.213-220 (2019).
- [3] オンラインネット麻雀天鳳. <https://tenhou.net/>