智能系统学报
CAAI TRANSACTIONS ON INTELLIGENT SYSTEMS

# 基于卷积神经网络的“拱猪”博弈算法

吴立成, 吴启飞, 钟宏鸣, 王世尧, 李霞丽

引用本文:
吴立成,吴启飞,钟宏鸣,王世尧,李霞丽. 基于卷积神经网络的“拱猪”博弈算法[J]. 智能系统学报, 2023, 18(4): 775–782.
WU Licheng,WU Qifei,ZHONG Hongming,WANG Shiyao,LI Xiali. Algorithm for “Hearts” game based on convolutional neural network[J]. *CAAI Transactions on Intelligent Systems*, 2023, 18(4): 775–782.

在线阅读 View online: https://dx.doi.org/10.11992/tis.202203030

**DOI**: 10.11992/tis.202203030
**网络出版地址**: https://kns.cnki.net/kcms/detail/23.1538.tp.20230322.1520.005.html

# 基于卷积神经网络的“拱猪”博弈算法

吴立成，吴启飞，钟宏鸣，王世尧，李霞丽
（中央民族大学 信息工程学院，北京 100081）

**摘　要**：“拱猪”又称“华牌”，是一款极具特点的牌类游戏，属于非完备信息博弈，由亮牌和出牌 2 个阶段组成，整个游戏过程具有极强的反转性。为了研究“拱猪”计算机博弈算法，本文提出了一种基于深度学习的“拱猪”博弈算法，包含亮牌和出牌 2 个神经网络，分别用于亮牌和出牌阶段。亮牌和出牌网络均采用卷积神经网络（convolutional neural network, CNN）来构建，根据功能特点分别设计为不同的网络结构。采用 11 000 局人类高级玩家的真实牌谱按比例生成训练数据和测试数据，对 2 个 CNN 网络进行了训练、测试和分析。结果表明，亮牌和出牌网络分别达到了 88.4% 和 71.4% 的准确率。对亮牌和出牌的一些具体例子进行的分析表明，本文算法能够产生合理的亮牌和出牌策略。

**关键词**：人工智能；非完备信息博弈；深度学习；卷积神经网络；拱猪；华牌；亮牌；出牌
**中图分类号**：TP183；G892　　**文献标志码**：A　　**文章编号**：1673−4785(2023)04−0775−08

**中文引用格式**：吴立成，吴启飞，钟宏鸣，等. 基于卷积神经网络的“拱猪”博弈算法 [J]. 智能系统学报，2023，18(4)：775–782.
**英文引用格式**：WU Licheng, WU Qifei, ZHONG Hongming, et al. Algorithm for “Hearts” game based on convolutional neural network[J]. CAAI transactions on intelligent systems, 2023, 18(4): 775–782.

# Algorithm for “Hearts” game based on convolutional neural network

WU Licheng, WU Qifei, ZHONG Hongming, WANG Shiyao, LI Xiali
(School of Information Engineering, Minzu University of China, Beijing 100081, China)

**Abstract**: “Hearts”, also known as “Chinese card game”, is a very characteristic poker game, which belongs to incomplete information games. It consists of two stages of card showdown and card playing, and there is strong reversality throughout the game. In order to study the computer game algorithm of “Hearts”, this paper proposes a “Hearts” game algorithm based on deep learning, which includes two neural networks, namely, card showdown and card playing, which are used in card showdown and card playing stage respectively. Both the card showdown network and card playing network are constructed by convolutional neural network (CNN), which are designed into different network structures according to their functional characteristics. Two CNN networks are trained, tested, and analyzed by using the real card playing patterns of 11,000 human advanced players to generate training data and test data proportionally. The results show that the accuracy of card showdown and card playing network reaches 88.4% and 71.4% respectively. The analysis of some specific examples of card showdown and card playing shows that the algorithm is able to produce reasonable card showdown and card playing strategies.

**Keywords**: artificial intelligence; game of incomplete information; deep learning; convolutional neural network; Hearts; Chinese card game; card-showing; card-playing

非完备信息博弈是指参与者无法从游戏对局中获得所有的局面信息，因此对其博弈算法研究具有一定的难度，目前已备受关注，成为热门研究之一。德州扑克[1-3]、“斗地主”[4-5]等，尤其是德州扑克的国内外相关研究成果较多。2013 年，王轩等[6-10]在信息表示、函数优化、博弈树搜索、对手建模和风险模型分析等方面取得的成果显著，并在 2013 年世界计算机扑克大赛（annual computer poker competition, ACPC）2 人限注项目竞赛中，取得了第 4 名的好成绩[11]。2015 年，Bowling 等[12]提出改进型虚拟遗憾最小化（counterfactual regret minimization, CFR）CFR+算法，在 2 人限注项目中取得了重大进展，首次成功地破解了该

**收稿日期**：2022−03−17.　**网络出版日期**：2023−03−27.
**基金项目**：国家自然科学基金项目（61773416, 61873291）.
**通信作者**：李霞丽. E-mail: xiaer_li@163.com.

项目所存在的制胜策略，但仍然无法解决超大规模的博弈问题。2018 年，Brown 等[13] 采用有限深度优先的方法进行搜索，所构建的智能体打败了先前版本的人工智能（artificial intelligence, AI）程序。2019 年，Noam[14-15] 使用自博弈的方法来训练智能体，这与训练 AlphaGo Zero、AlphaZero 的方法类似，所构建的智能体 Pluribus[3] 在六人无限注德州扑克项目中打败了人类高手。2020 年，张小川等[16] 设计了基于上限置信区间算法的决策模型回报函数来进行决策更新，并提出一种动态结合深度 Q 网络和 Sarsa 的算法来提高模型的学习效率，所构建的智能体在 2019 年全国大学生计算机博弈竞赛中表现优异，获得了一等奖。2021 年，彭丽蓉等[17] 采用 AC 自动机（aho-corasick, AC）算法，引入专家知识来预训练网络参数，所构建的智能体与其他版本的德州扑克智能体进行了对弈，结果表明每局的平均收益都在 1 个大盲注以上。2022 年，张蒙等[18] 针对对手建模，设计了一种包含智能体离线训练和在线博弈 2 个阶段的集成框架，该框架在面对动态对手策略时，智能体的水平较之前方法有所提升。Zhou 等[19] 通过考虑其他玩家手牌的可能范围来降低策略的可利用性，所构建的 DecisionHoldem 智能体公开战胜了最强的单挑无限德州扑克智能体 Slumbot 以及 Deepstack 的高级复制智能体 Openstack。

还有一些文献研究了“斗地主”非完备信息博弈问题。2018 年，Li 等[20] 将深度学习的方法运用到“斗地主”扑克牌中，完成了对玩家单个牌张的预测。2019 年，You 等[21] 提出组合 Q 学习（combinational Q-learning, CQL）算法，解决了多种组合出牌方式的困难。Jiang 等[22] 针对解决其他玩家出牌方式和策略无法知晓的问题，将使用人类玩家真实的对弈牌局信息训练好的网络模型提供给其他玩家进行决策。彭敫文等[4] 提出基于规则的手牌拆分算法，并采用蒙特卡洛（Monte-Carlo, MC）方法来选择收益最大的节点作为最佳决策，该方法能够较好地实现“斗地主”自我博弈。2020 年，彭敫文等[23] 又将蒙特卡洛搜索树方法和卷积神经网络算法相结合来研究“斗地主”的出牌策略，由该算法所构建的智能体在与其他目前已存在的“斗地主”策略的智能体对弈中，能够在胜率上取得较为明显的优势。同年，徐方婧等[5] 使用自我博弈收集得到的牌局信息来学习“斗地主”策略，采用基于权重的方式来克服训练数据分布不均匀的问题，该模型在与真人对弈中，取得了较高的胜率[5]。2021 年，Zha 等[24] 提出了一种深度蒙特卡洛（deep Monte-Carlo, DMC）方法，即利用深度神经网络、动作编码和并行行为体对传统的蒙特卡洛方法进行改进，所构建的智能体 DouZero 在 BotZone 平台上战胜了所有的“斗地主”AI。2022 年，郭荣城等[25]运用 Alpha-Beta 剪枝算法来解决“斗地主”残局问题，所构建的智能体在“欢乐斗地主”小程序的双人明牌残局对弈模式下，进行了多次模拟测试，取得了全胜战绩。Yang 等[26] 采用了一种完美−训练−不完美−执行框架，智能体可利用全局信息来指导策略训练，实验证明，其构建的智能体 PerfectDou 击败了所有的“斗地主”AI，其性能达到最优。

“拱猪”是一款在全世界华人圈内十分受欢迎的纸牌类游戏，属于非完备信息博弈。目前关于“拱猪”的研究还未见相关文献。虽然德州扑克 AI 已经可以战胜人类专业选手，“斗地主”AI 也逐渐接近人类高手水平，但它们对算力的要求较高，没有足够强大的硬件资源是无法实现的，除此之外，它们所采用的博弈算法也无法直接应用于“拱猪”。因此本文提出了一种基于卷积神经网络的“拱猪”博弈算法，将牌谱中人类高级玩家的亮牌和出牌动作视为正确的标注，通过有监督学习的方式，从真实对战数据中学习到人类玩家在亮牌和出牌时所采取策略。

## 1　游戏规则、牌的表示及算法流程

### 1.1　游戏规则

“拱猪”参与者人数为 4 人，去除大小王的 52 张牌分为“分牌”和“无分牌”2 类，有分值的牌张谓分牌，其余则为无分牌。所有分牌及其相对应的分值、可以进行亮牌动作的牌张以及所有分牌的分值在亮牌动作前后的变化情况参见中国华牌竞赛规则[27]。

游戏开始时每人一张轮流发牌，然后按亮牌、出牌的顺序分阶段进行。亮牌阶段，即开始出牌前，玩家可以选择将手中的黑桃 Q、方块 J、梅花 10 和红桃 A 亮出来，或者不亮。除了玩家手里仅有一张该花色的牌张之外，其余情况下，亮牌阶段中被亮出的牌张在该花色的第 1 轮出牌中不允许打出。

在出牌阶段，首轮一般是由初始手牌中含有梅花 2 的玩家先出，并且每次只允许出一张牌。下家根据当前手牌情况，选择一张与本轮次首位出牌玩家花色相同的牌张进行出牌，若没有，则可以选择垫一张不同于其花色的牌张，在 4 个玩家都依次出完牌后，本轮次的所有分牌都将会由牌张最大的玩家收集得到，垫不同于首家花色的牌张视为最小，同一花色牌张的大小关系为：

A 为最大, 2 为最小。首轮结束后, 每一轮都会按照以上一轮牌张最大的玩家先出的规则依次进行出牌, 直至游戏结束。

游戏结束时计算分数, 每位玩家需先各自计算出原始分数, 再计算每位玩家的最终分数, 具体算法为: 某玩家最后得分等于该玩家原始分数的 3 倍减去其他 3 个玩家原始分数之和。如果最终得分为正, 则为赢家; 如果得分为 0 或负, 则为输家。这样算得的 4 家得分之和将正好为 0。

### 1.2　牌的表示

本文用 1×52 的数组表示 52 张牌, 花色在数组中的排列顺序为: 黑桃、红桃、方块、梅花; 不同花色相对应的 13 张牌在数组中存储的位置也不同, 按照 A, 2, 3, …, K 的顺序依次排列。例如黑桃 Q 对应数组的下标为 11, 红桃 J 对应数组的下标为 23。全部牌张在数组中的对应下标如表 1 所示。

**表 1　各个花色牌张对应数组下标**

**Table 1　Corresponding array subscript of each suit card**

| 花色 | 数组下标 |
|---|---|
| 黑桃(S) | 0~12 |
| 红桃(H) | 13~25 |
| 方块(D) | 26~38 |
| 梅花(C) | 39~51 |

### 1.3　算法流程框架

“拱猪”共分为亮牌和出牌 2 个阶段, 每个阶段设计不同的卷积神经网络(convolutional neural network, CNN)结构来构建模型。“拱猪”算法流程框架如图 1 所示。

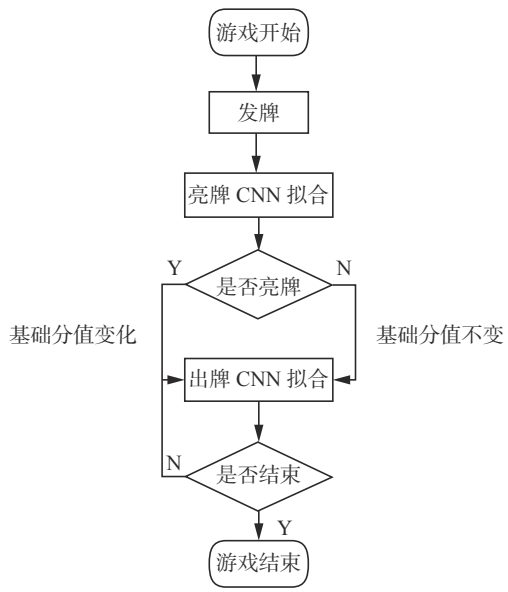


**图 1　“拱猪”算法流程框架**

**Fig. 1　“Hearts” algorithm process framework**

## 2　亮牌算法

### 2.1　亮牌类别表示

“拱猪”共有 4 张可以亮的牌, 本文用一个 1×4 的数组 $[x_0, x_1, x_2, x_3]$, 依次表示梅花 10、方块 J、黑桃 Q 和红桃 A 的亮牌情况。若某张牌被玩家亮牌了, 则将其对应的数组元素设置为 1, 否则设置为 0。

“拱猪”共有 16 种亮牌类别, 可以用序号 0~15 表示, 类别序号的计算方法为

$$K_{\text{ind}} = x_0 \cdot 2^3 + x_1 \cdot 2^2 + x_2 \cdot 2^1 + x_3 \cdot 2^0 \tag{1}$$

如数组 [1, 0, 0, 0] 表示只有梅花 10 进行了亮牌, 其亮牌类别序号为 8。

### 2.2　亮牌网络设计

#### 2.2.1　数据集

每局牌中 4 位玩家都可分别决定自己的亮牌类型, 因此 11 000 局人类高级玩家真实牌谱共可得到 44 000 条亮牌实验数据。本实验将亮牌数据划分为训练集和测试集, 划分比例为 4∶1。

#### 2.2.2　网络输入与输出

每位玩家只能根据自己的初始手牌进行亮牌决策, 因此亮牌阶段的输入信息为单个玩家的 13 张初始手牌。本文将玩家初始手牌信息用 1×52 的数组表示, 数组的相应元素为 1 表示有此牌, 为 0 则表示无。因此, 每个初始手牌的数组中有 13 个 1, 其他为 0。为了便于进行卷积操作, 将 1×52 的数组顺次转化成 4×13 的矩阵, 即 CNN 的输入信息为 4×13 的矩阵。

亮牌神经网络的输出层由 16 个神经元依次对应输出 16 种亮牌类别的概率, 最终可选出概率最大的类别进行亮牌。多种类别的概率同为最大时, 随机选择一种。

#### 2.2.3　网络结构

亮牌网络共 14 层。第 1 层为输入层, 输入信息为 4×13 的矩阵; 1~2 卷积层的卷积核个数为 32, 卷积层后为 1 个 ReLU 层, 经由第 1 个大小为 2×2 的 Max-pooling 层作用后, 变换为 2×6 的矩阵; 3~4 卷积层的卷积核个数为 64, 卷积层后为 1 个 ReLU 层; 5~6 卷积层的卷积核个数为 128, 卷积层后为 1 个 ReLU 层, 再经由第 2 个大小为 2×2 的 Max-pooling 层作用后, 变换为 1×3 的矩阵, Max-pooling 层后为 1 个 Dropout 层, 随机丢弃值设置为 0.2。1~2 卷积层的卷积核大小为 3×3, 3~6 卷积层的卷积核大小为 2×2, 所有卷积核步长均为 1, 所有 padding 均采用 same 模式。最后 1 层为全连接层, 采用 Softmax 函数进行亮牌动作分

类，根据每个亮牌动作的概率来进行亮牌决策。亮牌模块的 CNN 网络模型结构如图 2 所示。

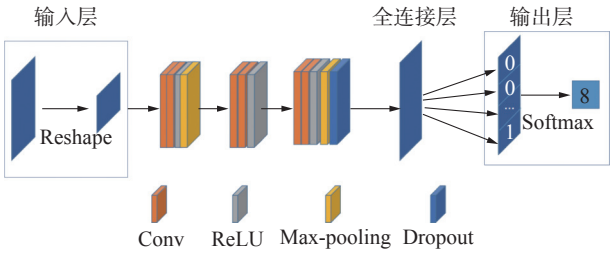


**图 2　亮牌 CNN 模型网络结构**

**Fig. 2　CNN mode network structure of card-showing**

### 2.2.4　评价指标和损失函数

准确率计算则以网络的输出是否与牌谱中人类玩家亮牌相一致为标准，准确率越高表明模型越能很好地学习到人类玩家的亮牌决策。亮牌共有 16 种动作，属于多分类问题，因此亮牌模型选择采用多分类损失函数。

亮牌共有 16 种动作，属于多分类问题，因此亮牌模型选择采用多分类损失函数。亮牌模型共需要 16 个输出向量值，每个类别的输出向量值经过 Softmax 函数转化后，其值可表示为模型对该类别的预测概率，且满足亮牌模型输出的 16 个预测概率值的和等于 1，Softmax 函数为

$$\mathrm{Softmax}(z)_i=\frac{\mathrm{e}^{z_i}}{\sum_{j=0}^{N-1}\mathrm{e}^{z_j}},i=0,1,\cdots,N-1 \tag{2}$$

式中：$z_i$ 为各个类别原始输出向量值；$N$ 为标签的类别数，对于亮牌分类问题其值为 16。

采用如下交叉熵损失函数：

$$-\log(P(y)) \tag{3}$$

式中：$y$ 为真实标签，亮牌类别分别对应着 $z_0$，$z_1$，…，$z_N$。

将式(2)代入式(3)中，得到：

$$-\log\mathrm{Softmax}(z)_i=-\left(z_i-\log\sum_{j=0}^{N-1}\mathrm{e}^{z_j}\right) \tag{4}$$

如果实际的对应类别输出值为 $z_{11}$，则损失函数为

$$-\log\mathrm{Softmax}(z)_{11}=-\left(z_{11}-\log\sum_{j=0}^{N-1}\mathrm{e}^{z_j}\right) \tag{5}$$

## 2.3　亮牌实验与分析

### 2.3.1　训练效果

经过 40 次的调参训练，训练轮数 epoch 设置为 50，准确率和损失值都处于收敛状态；优化器选择 Adam，即动态地调整学习率；输出为 16 种亮牌决策，即损失函数选择的是多分类损失函数；调用 ReduceLROnPlateau 函数优化学习率；单轮批量 batch_size 设置为 128，一次输入 32 局人类高级玩家牌谱的初始手牌信息，满足不能过大或过小的原则。其超参数设置如表 2 所示。

**表 2　亮牌超参数设置**

**Table 2　Card-showing hyperparameter settings**

| 参数名 | 数值 |
|---|---|
| 训练轮数 | 50 |
| 单轮批量 | 128 |
| 损失函数 | 多分类损失函数 |
| 衰减因子 | 0.5 |
| 最小学习率 | 0 |
| 优化器 | Adam |

亮牌模型的训练效果如图 3 所示。由图 3 可知，亮牌训练集上的准确率处于逐步上升的状态，在 epoch 值为 50 时，其准确率达到最高值 91.8%；亮牌测试集上的准确率则逐步达到平稳状态，在 epoch 值为 30 时，其准确率达到最高值 88.4%。

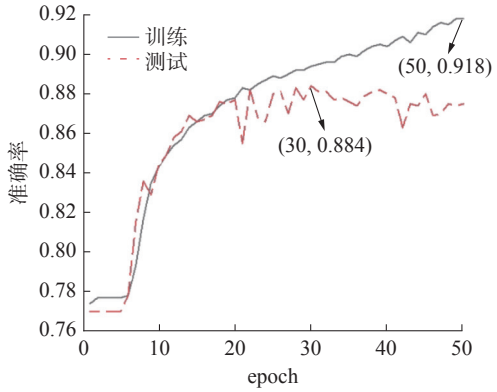


**图 3　亮牌训练效果**

**Fig. 3　Training effect of showing cards**

### 2.3.2　示例分析

保存模型，将测试集数据输入模型中，比较模型输出的亮牌结果与原数据的类别标注是否一致，模型亮牌结果与原数据标注类别如图 4 所示。

标注数据：[0,0,0,0]

预测结果：3，即 [0,0,1,1]

(a) 玩家初始手牌示例 2

标注数据：[0,0,0,0]

预测结果：5，即 [0,1,0,1]

(b) 玩家初始手牌示例 1

**图 4　亮牌结果**

**Fig. 4　Examples of show result**

示例 1 玩家 4 种花色牌张皆有, 红桃牌张数略多于其他花色且有红桃 A, 另外可以亮的牌张还有黑桃 Q 和方块 J, 属于初始手牌局面较为复杂的情况。红桃牌张虽较多且有大数值牌, 但是也有红桃 5 和红桃 2 小数值牌, 收“全红”的难度很大, 除此之外, 方块牌张较少, 方块 J 有着很大的机率被其他玩家“圈羊”。模型预测输出的亮牌结果是 5, 即选择将方块 J 和红桃 A 都亮出, 这与原始数据标注的结果是不一致的。分析可知, 模型预测结果没有人类玩家真实对局标注的结果好。

示例 2 玩家黑桃和红桃牌张占绝大多数, 有黑桃 Q 且小数值黑桃牌张较多, 有红桃 A 且大数值红桃较多, 属于初始手牌局面较为简单的情况。在其他玩家出梅花或方块花色牌张时, 可以将黑桃 Q 和红桃 A 迅速打出, 使其他玩家得到更多的负分。模型预测输出的亮牌结果是 3, 即选择将黑桃 Q 和红桃 A 都亮出, 这与原始数据标注的结果是不一致的。分析可知, 模型预测结果要比人类玩家真实对局标注的结果好。

由以上预测错误示例的分析可知, CNN 模型有些预测结果要比人类玩家的亮牌策略要好, 但是在对牌面信息较复杂的初始手牌时, CNN 模型的预测结果还不是很理想。

## 3　出牌算法

### 3.1　出牌类别表示

“拱猪”共有 52 张牌且每次只能出一张牌, 因此玩家出牌可有 52 种类别, 可用 1×52 的数组表示, 形式为 $[x_0, x_1, \cdots, x_{51}]$, 相应元素为 1 表示出此牌, 为 0 则不出。牌张和数组下标的对应关系见表 1。出牌类别可以用所出牌的对应数组下标来表示。例如出黑桃 Q, 就将黑桃 Q 对应的数组下标作为类别序号, 即 11。

### 3.2　出牌网络设计

#### 3.2.1　数据集

每次出牌都可以是一个样本数据, 但本实验采用的“拱猪”牌谱在所有分牌都已出完的情况下, 就直接结束牌局, 因此 11 000 局人类高级玩家真实牌谱共可得到 497 020 条实验数据。本实验将出牌数据划分为训练集和测试集, 划分比例为 4:1。

#### 3.2.2　网络输入与输出

本文将当前轮次玩家的手牌、当前轮次还在其他玩家手里的牌、各个玩家的牌局亮牌信息、当前轮次其他 3 个玩家的出牌、其他 12 轮各个玩家的出牌和当前轮次各个玩家已经收集得到的牌, 分别用 1×52 的数组表示, 其中, 未知轮次的玩家出牌信息均以 0 填充, 即输入信息为 61×52 的矩阵。输入信息矩阵的具体含义描述如表 3 所示。

**表 3　出牌 CNN 输入结构**
**Table 3　Play CNN input structure**

| 行数 | 含义 |
|---|---|
| 1 | 当前轮次玩家的手牌 |
| 2 | 当前轮次还在其他玩家手里的牌 |
| 3~6 | 各个玩家的牌局亮牌信息 |
| 7~9 | 当前轮次其他3个玩家的出牌 |
| 10~57 | 其他12轮各个玩家的出牌 |
| 58~61 | 当前轮次各个玩家已经收集得到的牌 |

出牌神经网络的输出层由 52 个神经元依次对应输出 52 张牌的出牌概率, 选概率最大的牌出。多张牌的概率同为最大时, 随机选择一张。

#### 3.2.3　网络结构

出牌模块的网络结构共有 46 层。第 1 层为输入层, 输入信息为 61×52 的矩阵; 1~4 卷积层的卷积核个数为 32, 经由第 1 个大小为 2×2 的 Max-pooling 层作用后, 变换为 30×26 的矩阵; 5~8 卷积层的卷积核个数为 64, 再经由第 2 个大小为 2×2 的 Max-pooling 层作用后, 变换为 15×13 的矩阵; 9~12 卷积层的卷积核个数为 128, 然后经由第 3 个大小为 2×2 的 Max-pooling 层作用后, 变换为 7×6 的矩阵; 13~16 卷积层的卷积核个数为 256, 最后经由第 4 个 2×2 的 Max-pooling 层作用后, 变为 3×3 的矩阵。每个卷积层后都接 1 个 ReLU 层, 每个 Max-pooling 层后为 1 个批归一化(batch normalization, BN)层和 1 个 Dropout 层, 随机丢弃值设置为 0.25。1~12 卷积层卷积核大小均为 5×5, 13~16 卷积层的卷积核大小均为 2×2, 所有卷积核步长均为 1, 所有 padding 均采用 same 模式。最后 1 层为全连接层, 采用 Softmax 函数进行出牌动作分类, 根据每个出牌动作的概率来进行出牌决策。出牌 CNN 网络结构如图 5 所示。

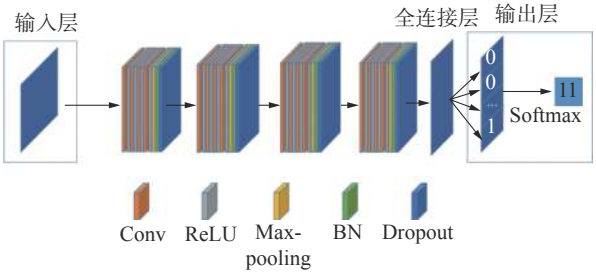


**图 5　出牌 CNN 模型网络结构**
**Fig. 5　CNN mode network structure of card-playing**

#### 3.2.4　评价指标和损失函数

准确率计算以网络的输出是否与牌谱中人类玩家出牌相一致为标准，准确率越高表明模型越能很好地学习到人类玩家的出牌决策。

出牌共有 52 种动作，属于多分类问题，即出牌模型仍选择采用多分类损失函数，公式参见 2.2.4 节，其中由于出牌有 52 种类别，$N$ 值为 52。

### 3.3　出牌实验与分析

#### 3.3.1　训练效果

经过 60 次的调参训练，确定训练轮数 epoch 为 50，准确率和损失值都处于收敛状态；优化器选择 Adam，即动态地调整学习率；输出为 52 种亮牌决策，损失函数选择的是多分类损失函数；调用 ReduceLROnPlateau 函数优化学习率；单轮批量 batch_size 设置为 440，即单轮批量输入为 2 个轮次的出牌信息。其超参数设置如表 4 所示。

**表 4　出牌超参数设置**

**Table 4　Card-playing hyperparameter settings**

| 参数名 | 数值 |
|---|---|
| 训练轮数 | 50 |
| 单轮批量 | 440 |
| 损失函数 | 多分类损失函数 |
| 缩放因子 | 0.5 |
| 最小学习率 | 0 |
| 优化器 | Adam |

出牌模型训练效果如图 6 所示。由图 6 可知，出牌训练集上的准确率处于逐步上升的状态，在 epoch 值为 50 时，其准确率达到最高值 85.6%；出牌测试集上的准确率则逐步达到平稳状态，在 epoch 值为 23 时，其准确率达到最高值 71.4%。

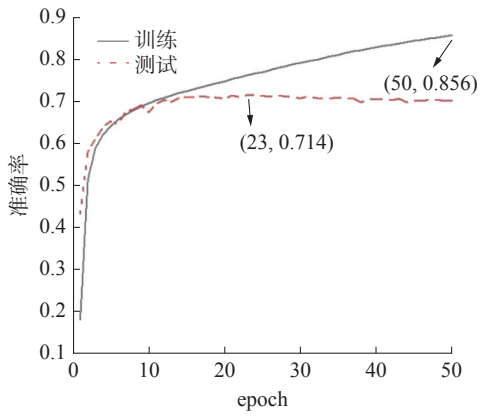


**图 6　出牌训练结果**

**Fig. 6　Training of paluing cards**

#### 3.3.2　示例分析

保存模型，将测试集数据输入模型中，比较模型输出的出牌结果与原数据的类别标注是否一致，模型出牌结果与原数据标注类别如图 7 所示。

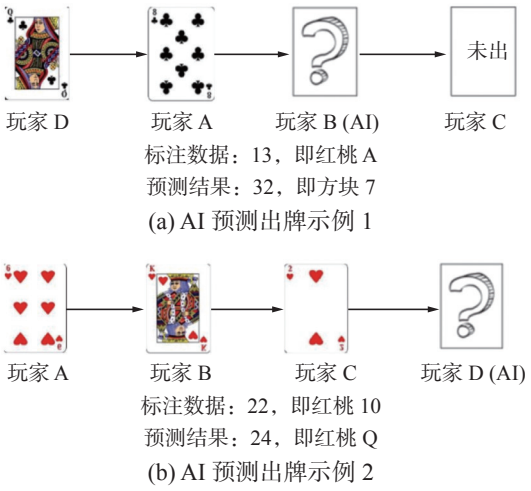


**图 7　出牌结果示例**

**Fig. 7　Examples of play result**

AI 预测出牌示例 1 中，玩家 B 为 AI 玩家，玩家 A、C 和 D 均为人类玩家。此轮次的玩家出牌顺序为 D→A→B→C，属于较为复杂的牌局状态。人类玩家 D 和 A 依次打出梅花 Q 和梅花 8。因为玩家 B 没有任何梅花花色的牌张，在已经得到红桃 K 和梅花 10 的情况下，为了避免得到更多的负分，应该选择将红桃 A 打出。AI 预测出牌为方块 7，这与原始数据标注的结果不一致。分析可知，模型预测结果没有人类玩家真实对局标注的结果好。

AI 预测出牌示例 2 中，玩家 D 为 AI 玩家，玩家 A、B 和 C 均为人类玩家。此轮次的玩家出牌顺序为 A→B→C→D，属于较为简单的牌局状态。人类玩家 A、B 和 C 依次打出红桃 6、红桃 K 和红桃 2。因为玩家 D 手中还有红桃 4、红桃 10 和红桃 Q，在玩家 B 已经出红桃 K 的情况下，为了避免得到更多的负分，玩家 D 应该将负分值较大的红桃 Q 打出。AI 预测出牌为红桃 Q，这与原始数据标注的结果不一致。分析可知，模型预测结果要比人类玩家真实对局标注的结果好。

综上分析，CNN 模型在“拱猪”出牌上的运用是具有可行性的，“拱猪”AI 具有基本的出牌策略，有些预测结果要比人类玩家的亮牌策略好，但是在一些较为复杂的牌局状态下，AI 还没有很好地学习到人类玩家的出牌策略特征，存在着出牌策略不合适的问题。

## 4　结束语

本文提出了一种基于卷积神经网络的“拱猪”博弈算法，采用 11 000 局人类高级玩家的真实牌谱，分别对其亮牌和出牌的动作进行标注，通过

有监督的方式去学习人类高级玩家的亮牌和出牌策略, 实验证明该模型在“拱猪”博弈算法研究上取得了不错的效果, “拱猪”AI 具有一定的亮牌和出牌能力, 但该模型在面对复杂的初始牌型和牌局状态时, 预测结果不是很理想, 下一步研究将会分析初始手牌的复杂性并改进 CNN 网络结构, 或者使用强化学习的方法, 增加训练数据量, AI 通过自对弈的方式去学习亮牌和出牌策略, 解决在面对初始手牌和牌局状态较为复杂时, CNN 模型预测效果不好的难题。

## 参考文献:


[1] BLAIR A, SAFFIDINE A. AI surpasses humans at six-player poker[J]. Science, 2019, 365(6456): 864–865.

[2] MORAVČÍK M, SCHMID M, BURCH N, et al. Deepstack: expert-level artificial intelligence in heads-up no-limit poker[J]. Science, 2017, 356(6337): 508–513.

[3] BROWN N, SANDHOLM T. Superhuman AI for heads-up no-limit poker: Libratus beats top professionals[J]. Science, 2018, 359(6374): 418–424.

[4] 彭啟文, 王以松, 于小民, 等. 基于手牌拆分的“斗地主”蒙特卡洛树搜索 [J]. 南京师大学报(自然科学版), 2019, 42(3): 107–114.
PENG Qiwen, WANG Yisong, YU Xiaomin, et al. Monte Carlo tree search for “Doudizhu” based on hand splitting[J]. Journal of Nanjing Normal University (natural science edition), 2019, 42(3): 107–114.

[5] 徐方婧, 魏鲲鹏, 王以松, 等. 基于卷积神经网络的“斗地主”策略 [J]. 计算机与现代化, 2020(11): 28–32.
XU Fangjing, WEI Kunpeng, WANG Yisong, et al. “Doudizhu” strategy based on convolutional neural networks[J]. Computer and modernization, 2020(11): 28–32.

[6] 马骁, 王轩, 王晓龙. 一类非完备信息博弈的信息模型 [J]. 计算机研究与发展, 2010, 47(12): 2100–2109.
MA Xiao, WANG Xuan, WANG Xiaolong. Information model for a class of incomplete information games[J]. Computer research and development, 2010, 47(12): 2100–2109.

[7] 王轩, 许朝阳. 时序差分在非完备信息博弈中的应用 [C]//中国机器博弈学术研讨会. 重庆: 重庆工学院学报, 2007: 16–22.
WANG Xuan, XU Chaoyang. The application of temporal difference in incomplete information games [C]//China Machine Game Academic Symposium. Chongqing: Journal of Chongqing Institute of Technology, 2007: 16–22.

[8] ZHANG Jiajia, WANG Xuan, YANG Ling, et al. Analysis of UCT algorithm policies in imperfect information game[C]//2012 IEEE 2nd International Conference on Cloud Computing and Intelligence Systems. Piscataway: IEEE, 2013: 132–137.

[9] ZHANG Jiajia. Building opponent model in imperfect information board games[J]. TELKOMNIKA Indonesian journal of electrical engineering, 2014, 12(3): 1975–1986.

[10] ZHANG Jiajia, WANG Xuan. Using modified UCT algorithm basing on risk estimation methods in imperfect information games[J]. International journal of multimedia and ubiquitous engineering, 2014, 9(10): 23–32.

[11] GINSBERG M L. GIB: imperfect information in a computationally challenging game[J]. Journal of artificial intelligence research, 2001, 14: 303–358.

[12] BOWLING M, BURCH N, JOHANSON M, et al. Computer science. Heads-up limit hold'em poker is solved[J]. Science, 2015, 347(6218): 145–149.

[13] BROWN N, SANDHOLM T, AMOS B. Depth-limited solving for imperfect-information games[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. New York: ACM, 2018: 7674–7685.

[14] SILVER D, SCHRITTWIESER J, SIMONYAN K, et al. Mastering the game of go without human knowledge[J]. Nature, 2017, 550(7676): 354–359.

[15] SILVER D, HUBERT T, SCHRITTWIESER J, et al. A general reinforcement learning algorithm that masters chess, shogi, and go through self-play[J]. Science, 2018, 362(6419): 1140–1144.

[16] 李轶. 德州扑克计算机博弈智能决策模型研究 [D]. 重庆: 重庆理工大学, 2020.
LI Yi. Research on intelligent decision model of texas Hold'em computer game[D]. Chongqing: Chongqing University of Technology Graduation Thesis, 2020.

[17] 李轶, 彭丽蓉, 杜松, 等. 一种德州扑克博弈的决策模型 [J]. 软件导刊, 2021, 20(5): 16–19.
LI Yi, PENG Lirong, DU Song, et al. A decision model for texas Hold'em game[J]. Software guide, 2021, 20(5): 16–19.

[18] 张蒙, 李凯, 吴哲, 等. 一种针对德州扑克 AI 的对手建模与策略集成框架 [J]. 自动化学报, 2022, 48(4): 1004–1017.
ZHANG Meng, LI Kai, WU Zhe, et al. An opponent modeling and strategy integration framework for Texas Hold'em AI[J]. Chinese journal of automation, 2022, 48(4): 1004–1017.

[19] ZHOU Qibin, BAI Dongdong, ZHANG Junge, et al. De-

cisionHoldem: safe depth-limited solving with diverse opponents for imperfect-information games[EB/OL]. (2022−01−27)[2022−03−17]. https://arxiv.org/abs/2201.11580.

[20] LI Saisai, LI Shuqin, DING Meng, et al. Research on fight the landlords' single card guessing based on deep learning[M]. Cham: Springer International Publishing, 2018: 363−372.

[21] YOU Yang, LI Liangwei, GUO Baisong, et al. Combinational Q-learning for Dou Di Zhu[EB/OL]. (2019−2−19)[2022−05−19]. https://arxiv.org/pdf/1901.08925v1.pdf.

[22] JIANG Qiqi, LI Kuangzheng, DU Boyao, et al. DeltaDou: expert-level doudizhu AI through self-play[C]//Proceedings of the 28th International Joint Conference on Artificial Intelligence. Hawaii: AAAI Press, 2019: 1265−1271.

[23] 彭啟文. 基于蒙特卡洛树搜索的“斗地主”研究 [D]. 贵阳: 贵州大学, 2020.
PENG Qiwen. Research on “Doudizhu” based on Monte Carlo tree search [D]. Guizhou: Graduation Thesis of Guizhou University, 2020.

[24] ZHA Daochen, XIE Jingru, MA Wenye, et al. DouZero: mastering DouDizhu with self-play deep reinforcement learning[EB/OL]. (2021–06–11)[2022–03–17]. https://arxiv.org/abs/2106.06135.

[25] 郭荣城, 李淑琴, 龚元函, 等. 二打一游戏残局模式下的对弈策略研究 [J]. 智能计算机与应用, 2022, 12(4): 151–158.
GUO Rongcheng, LI Shuqin, GONG Yuanhan, et al. Research on game strategy in two-on-one game endgame mode[J]. Intelligent computer and application, 2022, 12(4): 151–158.

[26] YANG Guan, LIU Minghuan, HONG Weijun, et al. PerfectDou: dominating DouDizhu with perfect information distillation[EB/OL]. (2022–03–30)[2022–03–17]. https://arxiv.org/abs/2203.16406.

[27] 《中国华牌竞赛规则》编写组. 中国华牌竞赛规则 (试行) [M]. 北京: 人民体育出版社, 2009: 2–4.

## 作者简介:

吴立成, 教授, 博士生导师, 国家民委首批中青年英才培养计划, 主要研究方向为智能机器人、计算机博弈、计算语言学。主持国家自然科学基金项目、863 项目等 10 余项, 授权发明专利 4 项, 获教育部科技进步二等奖 1 项、江苏省科技进步三等奖 1 项。发表学术论文 100 余篇, 出版专著 1 部、教材 1 部、译著 1 部。

吴启飞, 硕士研究生, 主要研究方向为计算机博弈。

李霞丽, 教授, 主要研究方向为机器博弈。主持国家自然科学基金面上项目 2 项、国家自然科学基金青年项目 1 项、省部级项目 1 项, 获得北京市高等学校青年英才计划奖励 1 项, 授权发明专利和登记软件著作权 10 余项。发表学术论文近 60 篇。