基于改进的 DDPG 机场机位分配算法研究

顾存昕，周洪涛

（华中科技大学人工智能与自动化学院，武汉 430074）

摘要：综合考虑机场的多个约束条件，以最大出港靠桥的航班数作为优化目标建立相应的数学模型，并将其转化成马尔可夫决策过程模型。设计环境的状态空间和智能体的动作空间，将大规模的离散动作空间通过构建特征的方式转变为连续动作空间，提出基于 *K* 最近邻（*K* nearest neighbor，KNN）和深度确定性策略梯度（deep deterministic policy gradient，DDPG）的机位分配算法，即 DDPG_KNN. 以乌鲁木齐地窝堡国际机场的实际航班数据进行仿真实验来验证模型的有效性，所改进的算法能够提高机位资源的利用率。在对比实验中，DDPG_KNN 的效果优于遗传算法。

关键词：人工智能；机位分配；深度强化学习；DDPG

中图分类号：TP18；V351　　**文献标识码：**A　　**文章编号：**1674-2850(2021)02-0187-15

Research on airport gate assignment based on improved DDPG algorithm

GU Cunxin, ZHOU Hongtao

(*School of Artificial Intelligence and Automation*, *Huazhong University of Science and Technology*, *Wuhan* 430074, *China*)

Abstract: Considering the multiple constraints of the airport comprehensively, a mathematical model is established by taking the maximum number of flights approaching the airport as the optimization objective, and the model is transformed into a Markov decision process model. The state space of the environment and the action space of the agent are designed, and the large-scale discrete action space is transformed into a continuous action space by the way of constructing features. An airport gate assignment algorithm based on *K* nearest neighbor (KNN) and deep deterministic policy gradient (DDPG) is proposed, namely DDPG_KNN. The actual flight data of Urumqi Diwopu international airport is used for simulation experiments to verify the effectiveness of the model, which can improve the utilization rate of gate resources. In the comparison experiment, the effect of DDPG_KNN is better than that of genetic algorithm.

Key words: artificial intelligence; airport gate assignment; deep reinforcement learning; DDPG

0 引言

自 21 世纪以来，航空运输因其舒适快捷的特点，逐渐成为备受大家青睐的交通运输方式。机位是机场运行重要资源之一，提高机位的利用率是改善机场运行效率、减少机场运营成本的重要方式。乌鲁木齐地窝堡国际机场为本文的基础研究对象，其旅客吞吐量、货邮吞吐量和起降架次近年来逐渐攀升，但由于土地、资金、气候、早期规划等因素，导致出现机位资源短缺的问题。机场需要通过软件算法根据每日航班情况进行机位分配，进而提高机位的利用率。基于此实际背景，本文展开基于人工智能技术的机位分配问题研究。

机位分配问题指在已知航班、机场资源、业务规则等信息后，充分利用机位资源，将航班分配到一个

作者简介：顾存昕（1996—），男，硕士研究生，主要研究方向：深度学习、推荐系统

通信联系人：周洪涛，副教授，主要研究方向：运筹与优化、系统建模. E-mail: zht730@qq.com

或多个停机位上，使得航班在预计时间正常起降。机位分配问题属于 NP-hard 问题[1]。机位分配求解的方法主要包括数学规划、专家系统和仿真模拟，采用强化学习解决机位分配问题才刚刚起步。2014 年，AOUN 等[2]基于马尔可夫决策过程来解决不确定条件下的机位分配问题，使用概率形式表达随机的飞机操作参数；并考虑飞机会发生延误等随机事件，建立了基于多智能体马尔可夫决策过程的分配模型，为工作人员提供一个鲁棒优先的解决方案[3]。2018 年，AOUN 等[4]继续针对机位分配出现的随机性情况，设计了一种时变多智能体马尔可夫决策过程模型，该模型在每个时间序列中提供一个鲁棒的优先解决方案。2019 年，许永磊[5]在预分配研究中使用遗传算法和 SARSA 相结合来解决局部最优解之间高度离散化的问题，使用二维表存储每个状态的适应度函数。2019 年，赵家明[6]探索基于策略梯度算法的机位分配算法，并设计了算法演示系统。2020 年，程博[7]提出以最大化靠桥数、最大化满足飞机机位的偏好和合理化同一机位连续航班间的间隔为优化目标，设计了基于蒙特卡洛搜索的强化学习算法并与启发式算法进行了比较。

本文主要研究采用深度强化学习求解机位分配问题。针对乌鲁木齐地窝堡国际机场现阶段航班数量增加而机位资源有限的问题，对其高峰时段的机位分配进行研究。该研究对象在包含机场通用约束条件下，强调尽可能地提高出港靠桥率，由此作为最终实现目标，通过比较遗传算法和多个深度强化学习模型后，给出一个求解机位分配的可迁移的深度学习模型。

1　问题描述

1.1　机位情况

本文研究对象乌鲁木齐地窝堡国际机场有 T1、T2、T3 三座航站楼，图 1 所示为该机场俯视图。地窝堡国际机场的机位可分为近机位、I 型远机位和 II 型远机位。每个机位都有各自的序号和大小。

1）近机位：靠近航站楼的停机位，航站楼的廊桥可直接连到飞机客舱门，旅客通过廊桥快速上下飞机。其特点是，从近机位上下飞机旅客的满意度最高。

2）I 型远机位：本文规定的一种远机位，不靠近指廊，旅客需要乘摆渡车往返航站楼。其特点是，在 I 型远机位的飞机可被牵引车拖至就近的近机位。

3）II 型远机位：常见的远机位，不靠近指廊，旅客需要乘摆渡车往返航站楼。其特点是，停在 II 型远机位的飞机不再移动。

图 1　地窝堡国际机场俯视图

Fig. 1　Top view of Urumqi Diwopu international airport

1.2 机位的业务逻辑

T1、T2、T3 航站楼的飞机类型有些许不同，在分配机位的过程需要考虑飞机类型与机位相匹配的问题。根据飞机着陆入口速度，飞机可分为 A～E 类，不同类型的飞机需要停在合适的机位上，部分飞机虽然是相同类型，但由于翼展和机身长度不同，也不一定能停在相同的机位上。

T1 航站楼对应的机位负责 11 家航空公司的飞机停放，待分配的航班数量范围在 8～10 架，中小型飞机为主，小飞机多为机型 E190 和 E195，无国际航班，有 5 个近机位和 8 个 I 型远机位以及部分 II 型远机位。

T2 航站楼对应的机位负责 16 家航空公司的飞机停放，待分配的航班数量范围在 30～45 架，各类飞机均有，无国际航班，有少量 VIP 和公务航班，有 9 个近机位和 16 个 I 型远机位以及部分 II 型远机位。

T3 航站楼对应的机位主要负责南航、国际、厦航、北航的飞机停放，待分配的航班数量范围在 40～50 架，各类飞机均有，有国际航班，有定期的 VIP 和公务航班，有 20 个近机位和 10 个常规的 I 型远机位以及部分 II 型远机位。各航站楼的机位资源如表 1 所示。

表 1 航站楼机位资源

Tab. 1 Gate resources in the terminals

航站楼	近机位(#)	I 型远机位(#)	II 型远机位(#)
T1	1/2/3/4/5	101/102/103/104/105/106/107/108	162/163/164/165/166/167/168/169/170
T2	8/9/10/11/12/13/14/15/16	71/72/73/74/75/76/77/78/79/99/100/105/106/107/108/109	110/111/112/113/141/142/143/144/145/146/147/148/149/150/151/152/153
T3	17/18/19/20/23/25/28/29/30/32/33/34/38/39/40/41/42/44/45/47	24/31/43/46/48/49/50/51/52/53/除冰坪	54/55/56/57/58/171/172/173/174/175/176/177/178/179/180/181

1.3 航班调度流程

需要分配的航班包括从当日下午 17 点至次日 12 点前进港且在次日 12 点前出港的航班。每一个航班所分配的机位包括进港机位、停留机位和出港机位，登机口因为资源充足，不在算法优化范围内。地窝堡国际机场工作人员对固定时间段内的计划航班进行分配是有必要的。这些在 17 点至次日 12 点前进港且在 12 点前出港的航班数量在 100 架左右，数量随季节等因素变动，航班数量较多，而且均在次日上午出港，如果不合理安排将会降低机场的运行效率，如将次日较晚出发的飞机安排在近机位，将严重影响整个上午乃至全天的靠桥率。其余时间的机位分配其实是遵循着插空、就近、方便保障等原则，真正对全天靠桥率有较大影响的是上述高峰时段的机位分配问题。

以下是工作人员对于一个航班的进港、出港机位的大致要求，即基本调度原则。

1）进港机位：除需要满足一些特殊规则外，如果航班需要在 II 型远机位上进行大车（测试引擎）或者航班一定要靠桥停留，进港机位优先是近机位，其次是 I 型远机位（可拖拽），若进港在 I 型远机位（可拖拽），则出港机位一定是近机位，最后是 II 型远机位，若进港在 II 型远机位，则出港机位一定是 II 型远机位。

2）出港机位：能靠桥尽量靠桥，只有从近机位和 II 型远机位出港这两种情况，出港优先在近机位。

1.4 机位分配情况分析

每个航班需要分配进港和出港 2 个机位，以下梳理机位分配情况：

1）进港机位和出港机位均为近机位；

2）进港机位和出港机位均为 II 型远机位；

3）进港机位在 I 型远机位，出港机位在近机位。

2 模型构建

2.1 符号定义

以下将进行机场机位分配的数学模型构建，表2所示为数学模型中用到的数学符号。

表2 符号表

Tab. 2 Symbols

符号	描述
$F=\{m\mid 1,2,\cdots,M\}$	航班集合 F，航班编号 m，航班数量 M
$C=\{n\mid 1,2,\cdots,N\}$	机位集合 C，机位编号 n，机位数量 N，且 $N>M$
d_{mn}	被分配到机位 n 的航班 m 的预计出港时刻
s_{mn}	被分配到机位 n 的航班 m 的进港时刻
D_{mn}	0～1变量，如果航班 m 的出港机位能够分配到机位 n 则为1，不满足分配条件为0，$m\in F,n\in C$
S_{mn}	0～1变量，如果航班 m 的进港机位能够分配到机位 n 则为1，不满足分配条件为0，$m\in F,n\in C$
Type_n	枚举型变量，如果机位 n 为近机位，$\text{Type}_n=0$；如果 n 为I型远机位，$\text{Type}_n=1$；如果 n 为II型远机位，$\text{Type}_n=2$
$\text{dist}(x_1,x_2)$	两个机位 $x_1,x_2\in C$ 的滑行道距离
AType_m	航班 m 的机型
U_n	机位 n 可停放的机型集合
ADemand_m	枚举型变量，航班 m 的业务需求
Demand_n	机位 n 可处理的业务集合
FSequence_n	求解后按时间先后排列的停留在机位 n 的航班列表，数量为 Z_n
index_{mn}	在机位 n 的航班列表 FSequence_n 中航班 m 的序号，即 $S_{mn}=D_{mn}=1$
Neighbor_n	机位 n 的临近机位集合
NearAirside_n	T3航站楼中，机位 n 的相邻指廊间的机位集合

2.2 目标函数

出港靠桥的航班数最大：

$$\max\sum_{m=1}^{M}\sum_{n=1}^{N}D_{mn}[\text{Type}_n=0], \tag{1}$$

其中，[]为艾佛森括号，方括号内的条件满足则为1，否则为0. $\sum_{n=1}^{N}D_{mn}[\text{Type}_n=0]$ 表示对于一个航班 m 的出港机位分配中，符合该航班 m 的近机位的数量。

2.3 约束条件

1）航班的近机位唯一性：一个航班最多能停靠一个近机位

$$\begin{gathered}\sum_{j=1}^{N}S_{mj}[\text{Type}_j=0]\in\{0,1\},\\ \sum_{k=1}^{N}D_{mk}[\text{Type}_k=0]\in\{0,1\},\\ j=k.\end{gathered} \tag{2}$$

2）航班的远机位唯一性：一个航班最多能停靠一个远机位

$$\sum_{j=1}^{N}S_{mj}[\text{Type}_j=1\vee\text{Type}_j=2]\in\{0,1\},$$

$$\sum_{k=1}^{N} D_{mk}[\text{Type}_k = 1 \vee \text{Type}_k = 2] \in \{0,1\}, \quad j = k. \tag{3}$$

3）同时刻航班唯一性：同一时刻一个机位只有一个航班

$$\sum_{m=1}^{M} S_{mn} \in \{0,1\}, \quad \sum_{m=1}^{M} D_{mn} \in \{0,1\}. \tag{4}$$

4）机型机位大小匹配性：航班所停机位与机型相匹配

$$\text{AType}_m \in U_n, \quad \forall m,n \in \{m,n \mid S_{mn} = 1 \vee D_{mn} = 1\}. \tag{5}$$

5）机型机位业务匹配性：航班的业务需求与机位业务能力匹配

$$\text{ADemand}_m \in \text{Demand}_n, \quad \forall m,n \in \{m,n \mid S_{mn} = 1 \vee D_{mn} = 1\}. \tag{6}$$

6）近机位出港时间约束：出港在同一近机位，时间紧邻的航班出港时间间隔至少 90 min

$$d_{qn} - d_{pn} \geqslant 90, \quad \forall p,q \in \{p,q \mid \text{Type}_n = 0,\ \text{index}_{qn} - \text{index}_{pn} = 1,\ \text{index}_{pn} \geqslant 0,\ \text{index}_{qn} \leqslant Z_n\}. \tag{7}$$

7）部分相邻机位不同航班进出间隔 5 min

$$\begin{gathered} \left|d_{pj} - d_{qk}\right| \geqslant 5, \\ \left|d_{pj} - s_{qk}\right| \geqslant 5, \\ \left|s_{pj} - d_{qk}\right| \geqslant 5, \\ \left|s_{pj} - s_{qk}\right| \geqslant 5, \\ \exists q \in \text{Neighbor}_q. \end{gathered} \tag{8}$$

8）部分指廊不同航班出港时间间隔 10 min

$$\begin{gathered} \left|d_{pj} - d_{qk}\right| \geqslant 10, \\ \left|d_{pj} - s_{qk}\right| \geqslant 10, \\ \left|s_{pj} - d_{qk}\right| \geqslant 10, \\ \left|s_{pj} - s_{qk}\right| \geqslant 10, \\ \exists q \in \text{NearAirside}_q. \end{gathered} \tag{9}$$

9）次日 12 点后的飞机不再分配。

3 马尔可夫决策过程建模

3.1 深度强化学习概述

1）马尔可夫决策过程

马尔可夫决策过程是由$\langle S, A, \boldsymbol{P}, R, \gamma \rangle$构成的一个元组，其中，$\gamma \in [0,1]$为一个衰减因子，$S$为一个有

限状态集，A为一个有限行为集，$\boldsymbol{P}$为集合中基于行为的状态转移概率矩阵：

$$\boldsymbol{P}_{ss'}^{a}=\boldsymbol{E}[R_{t+1}\mid S_t=s,A_t=a]. \tag{10}$$

R为基于状态和行为的奖励函数：

$$R_s^a=\boldsymbol{E}[R_{t+1}\mid S_t=s,A_t=a]. \tag{11}$$

个体在给定状态下从行为集合中选择一个行为的依据称为策略，用π表示。策略π是某一状态下基于行为集合的一个概率分布：

$$\pi(a\mid s)=\boldsymbol{P}[A_t=a\mid S_t=s]. \tag{12}$$

在马尔可夫决策过程中，策略仅通过依靠当前状态就可以产生一个个体的行为，策略描述的是个体行为产生的机制，是不随状态变化而变化的，不论状态是相同或是不同。

2）价值函数

价值函数$v_\pi(s)$是在马尔可夫决策过程下基于策略π的状态价值函数：

$$v_\pi(s)=\varepsilon[G_t\mid S_t=s]=\varepsilon[R_{t+1}+\gamma v_\pi(S_{t+1})\mid S_t=s]. \tag{13}$$

价值函数$q_\pi(s,a)$是在马尔可夫决策过程下基于策略π的行为价值函数：

$$q_\pi(s,a)=\varepsilon[G_t\mid S_t=s,A_t=a]=\varepsilon[R_{t+1}+\gamma q_\pi(S_{t+1},A_{t+1})\mid S_t=s,A_t=a]. \tag{14}$$

一个状态的价值可以用该状态下的所有行为价值来表达：

$$v_\pi(s)=\sum_{a\in A}\pi(a\mid s)q_\pi(s,a). \tag{15}$$

解决强化学习问题是要学习一个最优的策略让个体在与环境交互过程中获得始终比其他策略都要多的收获，最优策略用π_*表示。策略π优于$\pi'(\pi\geqslant\pi')$，意味着对于有限状态集里的任意一个状态s，不等式$v_\pi(s)\geqslant v_{\pi'}(s)$成立。

最优行为价值函数是所有策略下产生的众多行为价值函数中的最大者：

$$q_*(s,a)=\max_\pi q_\pi(s,a). \tag{16}$$

存在结论，最优策略下的行为价值函数均等同于最优行为价值函数：

$$q_{\pi_*}(s,a)=q_*(s,a). \tag{17}$$

最优策略可以通过最大化最优行为价值函数$q_*(s,a)$来获得：

$$\pi_*(a\mid s)=\begin{cases}1, & a=\arg\max_{a\in A} q_*(s,a),\\ 0, & \text{其他}。\end{cases} \tag{18}$$

公式（18）表示，在最优行为价值函数已知时，在某一状态s下，对于行为集里的每一个行为a将对应一个最优行为价值函数$q_*(s,a)$，最优策略在面对每一个状态时将总是选择能够带来最优行为价值的行为，即求解强化学习问题转变为求解最优行为价值函数问题。

3.2 马尔可夫决策过程

若已知马尔可夫决策过程中的所有东西，那么可以不用在环境中做出动作便可直接求解，称为规划。预测是已知一个马尔可夫决策（或奖励）过程和一个策略π，求解基于该策略的价值函数v_π. 控制是已

知一个马尔可夫决策过程，求解最优价值函数 v_* 和最优策略 π_*.

调研发现机位分配人员在处理待分配的航班时，按出港顺序进行机位选择，在考虑这些航班的分配时，本文将这些航班看成离散的时间序列，并进行分配决策。马尔可夫决策过程建模提供一个数学体系结构模型[8]，该机位分配问题可以转化为马尔可夫决策过程。求解机位分配的过程是在不基于模型的条件下，通过个体的学习优化价值函数，同时改善自身行为的策略以最大化获得累积奖励的过程，即不基于模型的控制过程。本节先进行机位分配的马尔可夫决策建模，对模型的状态空间、动作空间和奖励函数进行设计。

3.2.1 状态空间设计

当问题规模较大，不再使用字典之类的查表方式来存储状态或行为的价值，而是引入适当的参数，选取恰当的描述状态的特征，通过构建一定的函数来近似计算得到状态或行为价值。这种设计的好处是不需要存储每一个状态或行为价值的数据，而只需要存储参数和函数设计。理论上任何函数都可以被用作近似价值函数，实际选择何种近似函数需要根据问题的特点，比如较常用的近似函数有线性函数组合、神经网络、决策树、傅里叶变换等。这些近似函数在强化学习中主要任务是对一个状态进行恰当的特征表示。

在设计近似方法前，首先需要表示环境的状态空间，状态空间表示机位分配过程中所需要的资源，对于一个智能体，最需要的信息是机位利用情况和航班信息，下文将介绍两种资源的表示方法。

1）机位资源图

机位资源图，表示在一个决策下，即航班选择机位过程中，机位资源的使用情况。已知机场有 N 个机位，待分配航班为 M 个，即决策次数为 M，待分配的航班时间范围从当日 17 点至次日 12 点，记为时间 T，则资源图的大小是 $N\times T$，经过 M 次的决策后完成机位分配。

当 $N=10$ 且 $T=20\times t$ 时，表示资源图的大小是 $10\times 20\times t$，t 表示一个单位时间，单位为分钟或小时，如图 2 所示，横向表示机位的个数，从 1#～10#，纵向表示时间块，每个时间块为 t，有 20 个时间块，若每个时间块表示的时间 $t=1$ h，那么纵向就表示 20 h 的时间跨度，以 17 点作为起始时间，那么该图的时间跨度在 17 点～次日 13 点。图 2 中有颜色覆盖的区域表示该机位在这个时间跨度内被占用，例如蓝色区域表示 3#机位在 19 点～次日 4 点被占用，其他颜色同理。

2）航班资源图

航班资源图，表示一个决策内容，即一个航班的状态。已知机场有 N 个机位，时间范围在当日 17 点至次日 12 点，记为时间 T，则资源图的大小是 $N\times T$.

当 $N=10$ 且 $T=20\times t$ 时，表示资源图的大小是 $10\times 20\times t$，t 表示一个单位时间，单位为分钟或小时，与图 2 的横纵向表示相同，图 3 横向表示 10 个机位，纵向表示 17 点～次日 13 点的时间跨度。图中所覆盖的机位包括 2#、3#、6#、9#机位，表示该航班可以停在这 4 个机位，并且航班进港时间是 18 点，起飞时间是次日 8 点，这样既包括了业务规则，同时也包括了航班的起降信息。

3.2.2 动作空间设计

每一个航班由进港机位、出港机位表示，则二元组 $\langle j,k\rangle$，$j,k\in C$ 表示一个航班分配情况，j 表示进港机位，k 表示出港机位。二元组 $\langle j,k\rangle$，$j,k\in C$ 构成决策的动作空间，即智能体可选动作集合。

假设 II 型远机位用“D”表示，那么航班分配到 II 型远机位是一个动作 $\langle D,D\rangle$。除将航班分配到

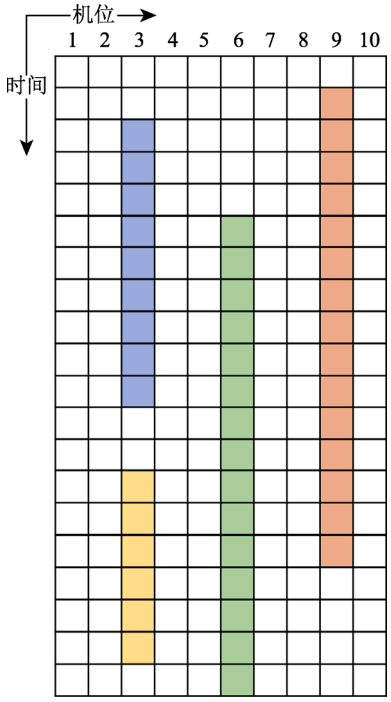

图2 机位资源图

Fig. 2 Resource map of gates

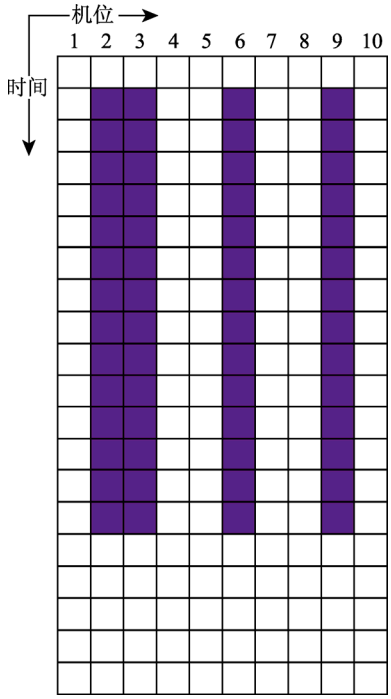

图3 航班资源图

Fig. 3 Resource map of flights

II型远机位外，其他的分配结果中出港机位均为近机位，停留机位为I型远机位或近机位。已知近机位有34个，I型远机位有30个。

3.2.3 奖励函数设计

针对一个特定的问题如何设计奖励函数反馈都是耐人寻味的。反馈设计不好一般会导致网路不收敛，结果不优或者智能体根本没有体会到需要学习的东西。一般来说，反馈设计都遵循着越简单越好的原则。设计的反馈其实表现的仅是设计者基于自身观测结果的期望，而并非是完全正确的定义，这个结果需要被放在环境的上下文里进行考虑。

对于机位分配决策过程来说，分配到近机位的情况是要优于分配到I型远机位的情况，也优于分配到II型远机位的情况，奖励函数设计便基于停留机位分配结果，若停留机位分配到近机位则反馈为正数，若停留机位分配到II型远机位则反馈为负数，分配到I型远机位则反馈介于前两者之间。这样设计的好处是可以直观地表达出港靠桥次数最大的优化目标，将这个优化目标融合进奖励函数。

4 算法设计

4.1 基于KNN的大规模动作处理

机位分配的动作空间属于大规模动作空间，当动作过多时，神经网络输出节点增加，大大增加参数存储内存，本文使用KNN将机位分配的离散动作空间转化为连续动作空间。根据DDPG基本原理可知，该算法常用于连续动作空间问题求解，将KNN与DDPG结合可以降低机位分配在大规模动作空间求解难度。以下将介绍KNN的基本算法。

KNN属于分类算法，通过计算不同特征值之间的距离进行分类。大致思路为样本的类别与特征空间中的k个最相似的样本的大多数类别相同。算法通过计算样本特征之间的距离来作为样本间的非相似性

指标，一般距离使用欧氏距离或曼哈顿距离。

*k*d 树的构建方式：从 m 个样本的 n 维特征中，分别计算 n 维特征每一维的数值方差，用方差最大的那一维特征作为 *k*d 树的根节点，那一维特征的中位数作为样本的划分点，将所有样本根据该特征的值划分到左右两个子树中，对左右子树采用同样方式找到方差最大的特征作为子树根节点，依次递归下去直到完成 *k*d 树，时间复杂度由 $O(nm)$ 降为 $O(n\log m)$.

4.2 DDPG 算法基础

对于连续动作控制空间，Q-Learning、深度 Q 网络（deep Q network，DQN）算法是没有办法处理的。在连续动作的场景下，神经网络输出一个具体的数值，用 $\mu_\theta(s_t)$ 表示确定性的策略，该策略输入一个状态 s，神经网络参数计算得到的是一个动作。神经网络要输出连续动作，一般可以在输出层加上激活函数 tanh，输出范围固定在[–1, 1]之间，再按实际动作范围做缩放。

针对连续动作空间，借鉴 DQN 的思路，Google DeepMind[9]引入经验回放和双网络的方法，提出 DDPG. DDPG 保留了 DQN 的 Q 网络的网络结构，加上一个策略网络，即 Actor 策略网络和 Critic 价值网络。算法流程[9]如下。

算法：DDPG 算法

输入： γ , τ

输出： 优化后的 θ_{critic} , θ_{actor}

初始化： Actor 策略网络 $\mu(s\,|\,\theta_{\text{actor}})$，Critic 价值网络 $Q(s,a\,|\,\theta_{\text{critic}})$，目标策略网络 μ'，目标价值网络 Q'，经验池 ***D***

重复 增加全局步数达到预设上界:

初始化一个随机噪声 N

循环 $t=1$ 到 T

在探索噪声 N 下生成动作 $a_t=\mu(s_t\,|\,\theta_{\text{actor}})+N_t$

执行 a_t

获得环境的状态 s_{t+1}，反馈 r_{t+1}

将$(s_t,a_t,r_{t+1},s_{t+1})$ 储存到经验池 ***D*** 中，随机采样 M 组

设 $y_i=r_{i+1}+\gamma Q'(s_{i+1},\mu'(s_{i+1}\,|\,\theta'_{\text{actor}})\,|\,\theta'_{\text{critic}})$

通过最小化损失函数来更新价值网络参数：

$$L=\sum_i(y_i-Q(s_i,a_i\,|\,\theta_{\text{critic}}))^2/M$$

通过采样的策略梯度更新策略网络参数：

$$\nabla_{\theta_{\text{actor}}}J\approx\sum_i\nabla_a Q(s,a\,|\,\theta_{\text{critic}})|_{s=s_t,a=\mu(s_i)}\,\nabla_{\theta_{\text{actor}}}\mu(s\,|\,\theta_{\text{actor}})|_{s_i}/M$$

更新目标网络参数：$\theta'_{\text{critic}}=\tau\theta_{\text{critic}}+(1-\tau)\theta'_{\text{critic}}$，$\theta'_{\text{actor}}=\tau\theta_{\text{actor}}+(1-\tau)\theta'_{\text{actor}}$

结束循环

4.3 DDPG_KNN 实验设计

智能体的 Actor 策略网络、目标策略网络、Critic 价值网络、目标价值网络是卷积神经网络（convolutional neural networks，CNN），初始化网络参数并与环境进行交互，智能体的输入是环境状态，用图片去表征环境状态，Actor 策略网络输出的是确定性的动作，用连续数值向量表示，利用 KNN 计算在动作空间中的最近邻动作，计算过程中屏蔽无效动作，得到有效的最近邻动作。添加的噪声是不相关的、均值为 0 的高斯噪声。基于 DDPG_KNN 的机位分配算法框架如图 4 所示。

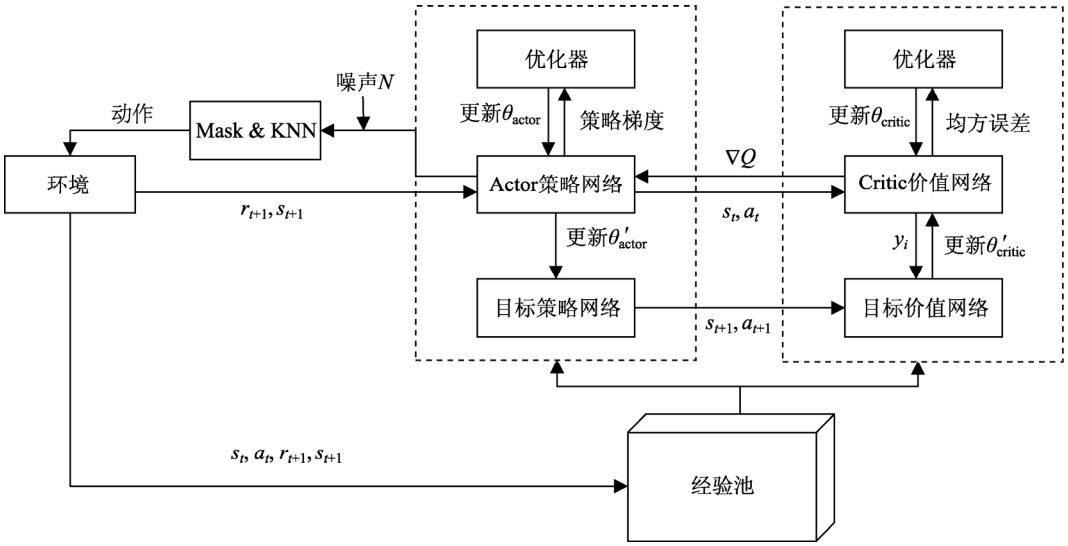

图 4　DDPG_KNN 结构

Fig. 4　Structure of DDPG_KNN

将离散动作空间连续化需要对动作做特征构造，二元组$\langle j,k\rangle$，$j,k\in C$作为离散动作表示进港机位、停留机位、出港机位。选取的特征包括进港机位的 xy 坐标、出港机位的 xy 坐标以及 3 个机位间的拖行距离这 5 个特征，部分动作的特征描述如表 3 所示。对各维特征进行标准化，便得到连续的动作空间。

表 3　动作特征

Tab. 3　Features of action

动作	进港机位 x 坐标	进港机位 y 坐标	出港机位 x 坐标	出港机位 y 坐标	拖行距离
⋮	⋮	⋮	⋮	⋮	⋮
('99', '5')	0.491	0.156	0.413	0.130	0.521
('1', '1')	0.530	0.327	0.502	0.361	0.021
('10', '10')	0.381	0.173	0.344	0.123	0.092
⋮	⋮	⋮	⋮	⋮	⋮

5　实验结果

5.1　DDPG_KNN 的参数配置

由图 5 可知，Actor 策略网络和目标策略网络是分开的，两个网络初始化相同的参数，输出的维度是上一节构造的特征数量，若动作为二元组表示，输出神经节点个数为 5. Critic 价值网络和目标价值网络输出动作价值。

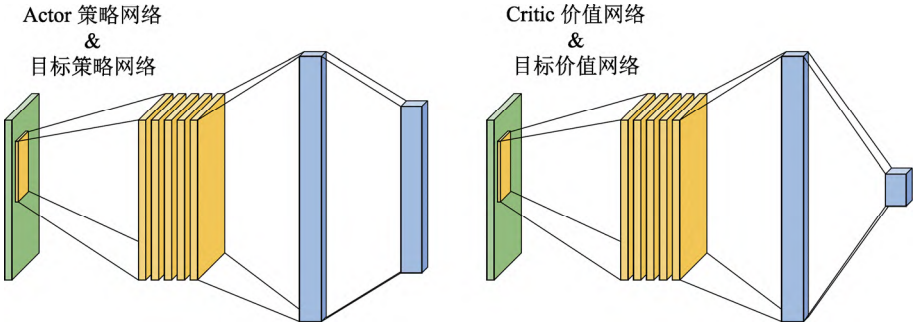

图 5　DDPG_KNN 的 CNN 结构

Fig. 5　Structure of CNN in DDPG_KNN

程序运行的服务器环境为 Tesla P40. 经过多次调参，最终 DDPG_KNN 算法的参数配置如表 4 所示。

表 4 CNN 参数

Tab. 4 CNN parameters

变量名	值	说明
task_num	3	航班状态由 b 个航班资源图拼接构成，$b=3$
resize_radio	(2,2)	环境状态图在输入 CNN 网络前缩小为原来的一半
action_coder	2	动作空间由 2 位编码构成
Reward	[10,0,−10]	奖励函数，停留机位为近机位获得奖励 10，停留机位为 I 型远机位获得奖励 0，停留机位为 II 型远机位获得奖励−10
max_ep	300	最大迭代次数
learning_rate	0.000 1,0.00 1	Actor 网络优化器 Adam 的学习率为 0.000 1，Critic 网络优化器 Adam 的学习率为 0.000 1
gamma	0.99	gamma 值
tau	0.005	τ
capacity	3 000	经验存储容量
target_update_interval	8	每 8 个 step 更新一次参数
batch_size	128	一次批处理抽取样本数
exploration_noise	0.5	设置噪声高斯分布的方差
update_iteration	2	更新网络参数采样次数
conv2d_in_channels	1	CNN 卷积层输入通道数
conv2d_out_channels	5	CNN 卷积层输出通道数
conv2d_kernel_size	5	CNN 卷积层 kernel 值
conv2d_stride	(1,1)	CNN 卷积层 stride 值
pi_hidden2_output	1 000	第 1 层全连接的 Critic 神经节点数
v_hidden2_output	500	第 1 层全连接的 Actor 神经节点数

5.2 结果分析

以 2019 年 1 月 12 日作为案例 1 进行实验，图 6 所示为靠桥率变化情况，其中深紫色线为 10 次实验的均值 μ，浅紫色范围表示由 10 个种子控制的算法的离散程度，该范围为$[\mu-\sigma,\mu+\sigma]$。图 6 的靠桥率变化情况和图 7 的反馈变化情况相同，不同种子的结果的离散程度也相同，主要原因是将靠桥率作为奖励函数设计依据，靠桥率的高低表现为反馈的升降。

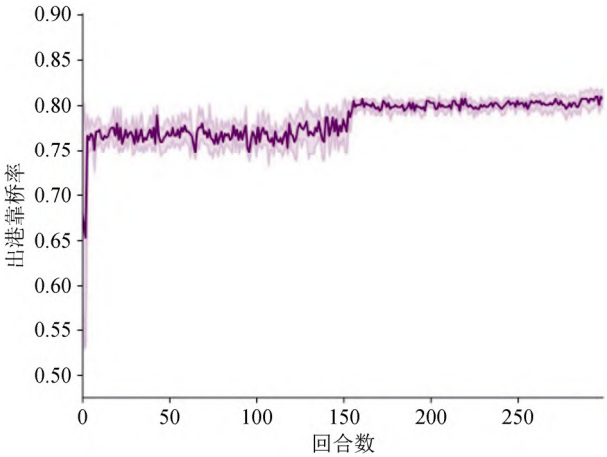

图 6 靠桥率变化情况

Fig. 6 Change of airport bridge rate

在图 6 靠桥率的变化情况中，从 10 次不同种子的平均结果来看，初始结果平均靠桥率为 0.67，初始最大靠桥率为 0.80，之后靠桥率平均值一直在增大，在 170 代时收敛速度降低，最终收敛至 0.80，即有 70 架飞机靠桥。从 10 次不同种子的结果范围来看，初始的离散程度范围较大，在 0.50～0.80 之间，稳定后的离散范围在 0.77～0.81 之间。

图 8 为遗传算法（genetic algorithm，GA）、DQN 与 DDPG_KNN 的结果比较图。观察主要的优化目标效果（最大出港靠桥航班率），可以看到 DDPG_KNN 从一开始的效果就优于另外 2 种算法。

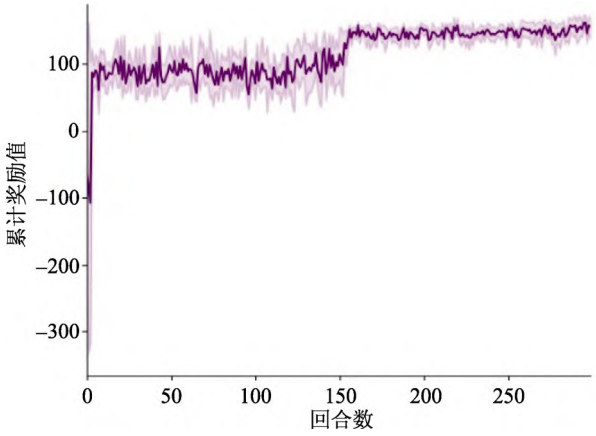

图7　反馈变化情况

Fig. 7　Change of reward

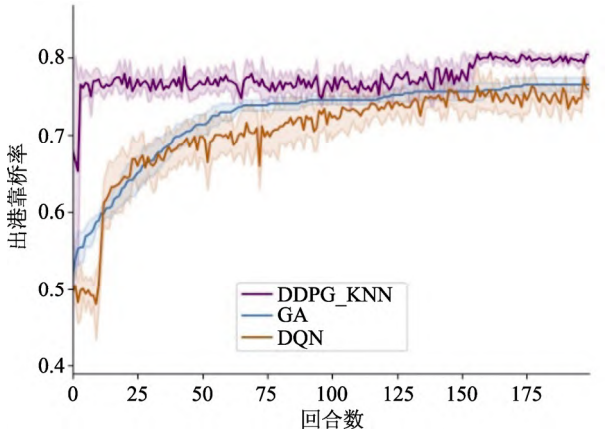

图8　GA、DQN、DDPG_KNN的结果对比

Fig. 8　Results comparison of GA, DQN, DDPG_KNN

6　结论

机位分配是影响机场运行安全和效率的重要因素之一，本文对机场高峰期航班的停留机位的分配进行研究。从航班降落到起飞将可能需要经过多个机位，同时也需要考虑机场实际的业务约束，机位分配是一个目标、多约束的组合优化问题。本文在此基础上进行研究。

对机场机位分配流程进行分析，考虑到同时刻航班唯一性、机型机位大小匹配性、机型机位业务匹配性等业务约束，将出港靠桥的航班数最大作为优化目标，建立机位分配的基本模型。

考虑到智能体做决策时会出现动作空间过大的问题，借鉴KNN的思路，对2位编码的动作空间进行特征构造，将原来离散的动作空间用连续向量表示，让网络中输出层的神经节点数减少，使得参数更少。由于动作空间由连续向量表示，采用同样基于Actor-Critic的DDPG算法进行实验。实验验证了DDPG_KNN算法的有效性，其求解的出港靠桥率结果也高于遗传算法和DQN算法求解的结果。

[参考文献] (References)

[1] HAGHANI A, CHEN M. Optimizing gate assignments at airport terminals[J]. Transportation Research Part A: Policy and Practice, 1998, 32(6): 437-454.

[2] AOUN O, EL AFIA A. Using Markov decision processes to solve stochastic gate assignment problem[C]//2014 International Conference on Logistics Operations Management. New York: IEEE, 2014: 42-47.

[3] AOUN O, EL AFIA A. Application of multi-agent Markov decision processes to gate assignment problem[C]//2014 Third IEEE International Colloquium in Information Science and Technology (CIST). New York: IEEE, 2014: 196-201.

[4] AOUN O, EL AFIA A. Time-dependence in multi-agent MDP applied to gate assignment problem[J]. International Journal of Advanced Computer Science & Applications, 2018, 9(2): 331-340.

[5] 许永磊. 基于遗传算法与强化学习的机位分配研究[D]. 武汉：华中科技大学，2019.
XU Y L. Research on airport gate assignment based on genetic algorithm and reinforcement learning[D]. Wuhan: Huazhong University of Science and Technology, 2019. (in Chinese)

[6] 赵家明. 机场停机位智能分配方法研究及实现[D]. 北京：北京工业大学，2019.
ZHAO J M. Research and implementation of air gate intelligent assignment method[D]. Beijing: Beijing University of Technology, 2019. (in Chinese)

[7] 程博. 基于深度强化学习的停机位分配算法研究[D]. 北京：中国科学院大学，2020.

CHENG B. Gate assignment problem research via deep reinforcement learning[D]. Beijing: University of Chinese Academy of Sciences, 2020. (in Chinese)

[8] WIKIPEDIA F, PROGRAMMING D, PROCESSES M. Markov decision process[J]. European Journal of Operational Research, 1989, 39(1): 1-16.

[9] KINGMA D P, BA J L. Adam: a method for stochastic optimization[M]. Ithaca: arXiv, 2015.

[附录] 2019 年 1 月 12 日的数据

航站楼	航班号	任务	机号	机型	起飞机场	预起时间	降落机场	预降时间
T1	GS7563	正班	B3263	E190	乌鲁木齐	7:25	喀什	0:05
T1	GS7551	正班	B1619	A320	乌鲁木齐	7:45	和田	0:40
T1	GS7513	正班	B3171	E190	乌鲁木齐	7:50	阿克苏	0:10
T1	GS7497	正班	B1620	A320	乌鲁木齐	7:50	西安	22:20
T1	GS7571	正班	B3257	E195	乌鲁木齐	8:05	塔城	1:05
T1	GS7521	正班	B8062	A320	乌鲁木齐	8:25	郑州	23:20
T1	GS7535	正班	B3213	E190	乌鲁木齐	8:45	伊宁	23:00
T1	GS7493	正班	B3108	E190	乌鲁木齐	9:50	西安	23:35
T1	GS7577	正班	B8389	A321	乌鲁木齐	9:55	天津	2:00
T1	GS7505	正班	B3267	E190	乌鲁木齐	10:15	博乐	21:55
T1	GS7555	正班	B3160	E190	乌鲁木齐	10:40	阿勒泰	23:55
T1	GS7557	正班	B3168	E190	乌鲁木齐	11:00	阿勒泰	0:45
T2	DZ6248	正班	B1462	B738	乌鲁木齐	6:55	郑州	22:40
T2	SC2281	正班	B1230	B738	乌鲁木齐	6:55	若羌	23:55
T2	SC8825	正班	B1360	B738	乌鲁木齐	7:00	青岛	22:30
T2	UQ3563	正班	B205V	B738	乌鲁木齐	7:00	兰州	21:15
T2	UQ2657	正班	B7198	B738	乌鲁木齐	7:05	武汉	1:45
T2	UQ2603	正班	B205T	B738	乌鲁木齐	7:10	和田	1:35
T2	HU7889	正班	B1995	B738	乌鲁木齐	7:20	喀什	23:55
T2	SC8812	正班	B1441	B738	乌鲁木齐	7:25	南京	23:55
T2	MU2397	正班	B8119	A320	乌鲁木齐	7:30	西安	23:00
T2	SC8755	正班	B1359	B738	乌鲁木齐	7:35	济南	22:00
T2	SC4705	正班	B1437	B738	乌鲁木齐	7:45	合肥	23:40
T2	FM9138	正班	B1265	B738	乌鲁木齐	7:45	银川	0:40
T2	SC8839	正班	B1443	B738	乌鲁木齐	7:45	伊宁	0:25
T2	CA4369	正班	B5325	B738	乌鲁木齐	7:50	广州	20:00
T2	SC4699	正班	B1442	B738	乌鲁木齐	7:50	兰州	0:25
T2	HU6091	正班	B5853	B738	乌鲁木齐	7:55	喀什	23:15
T2	UQ2519	正班	B205U	B738	乌鲁木齐	8:00	兰州	1:05
T2	HU7895	正班	B5636	B738	乌鲁木齐	8:05	西安	1:10
T2	UQ2539	正班	B2159	B738	乌鲁木齐	8:20	万州	23:45
T2	UQ2507	正班	B1569	B738	乌鲁木齐	8:35	西安	0:55
T2	MU2784	正班	B2220	A320	乌鲁木齐	8:35	太原	22:45
T2	SC4903	正班	B7806	B738	乌鲁木齐	8:45	石家庄	1:40
T2	HU7346	正班	B5852	B738	乌鲁木齐	9:00	北京	1:40
T2	HU7821	正班	B5581	B738	乌鲁木齐	9:00	济南	23:50

续表

航站楼	航班号	任务	机号	机型	起飞机场	预起时间	降落机场	预降时间
T2	FM9220	正班	B1451	B738	乌鲁木齐	9:05	上海虹桥	22:25
T2	CA1292	正班	B5398	B738	乌鲁木齐	9:15	北京	18:25
T2	UQ2513	正班	B1568	B738	乌鲁木齐	9:15	绵阳	0:05
T2	HU7819	正班	B6103	B738	乌鲁木齐	9:20	西安	0:35
T2	CA4192	正班	B6036	A319	乌鲁木齐	9:45	成都	21:45
T2	ZH9242	正班	B5440	B738	乌鲁木齐	10:10	郑州	22:55
T2	CA1296	正班	B1761	B738	乌鲁木齐	10:15	北京	0:10
T2	GJ8668	正班	B8898	A320	乌鲁木齐	10:55	银川	23:55
T2	UQ2503	正班	B6268	B738	乌鲁木齐	10:55	郑州	4:00
T2	HU7833	正班	B1102	B738	乌鲁木齐	11:00	汉中	0:45
T2	FM9222	正班	B5370	B738	乌鲁木齐	11:10	兰州	0:50
T3	3U8859	正班	B8330	A320	乌鲁木齐	6:55	泸州	20:05
T3	CZ6861	正班	B1520	B738	乌鲁木齐	7:15	阿克苏	1:15
T3	3U8061	正班	B8681	A20N	乌鲁木齐	7:30	兰州	20:40
T3	CZ6671	正班	B1285	B738	乌鲁木齐	7:30	库尔勒	1:25
T3	CZ6811	正班	B5285	B737	乌鲁木齐	8:00	和田	22:10
T3	CZ6821	正班	B7995	B738	乌鲁木齐	8:00	伊宁	0:55
T3	CZ6877	正班	B1367	B738	乌鲁木齐	8:00	莎车	0:50
T3	CZ6967	正班	B5762	B738	乌鲁木齐	8:00	兰州	1:50
T3	CZ6681	正班	B5291	B737	乌鲁木齐	8:05	克拉玛依	1:25
T3	CZ8601	正班	B5250	B737	乌鲁木齐	8:05	喀什	0:10
T3	CZ6007	正班	B1362	B738	乌鲁木齐	8:10	伊斯兰堡	0:35
T3	CZ6871	正班	B1747	B738	乌鲁木齐	8:10	库车	0:55
T3	CZ6827	正班	B7970	B738	乌鲁木齐	8:15	图木舒克	21:00
T3	CZ6889	正班	B6281	A320	乌鲁木齐	8:15	长沙	0:40
T3	CZ6841	正班	B1952	B738	乌鲁木齐	8:20	阿勒泰	0:50
T3	CZ6939	正班	B5761	B738	乌鲁木齐	8:20	武汉	0:05
T3	3U8901	正班	B6433	A319	乌鲁木齐	8:30	西宁	21:00
T3	CZ6941	正班	B1407	B738	乌鲁木齐	8:55	成都	23:35
T3	CZ6024	正班	B1283	B738	乌鲁木齐	9:00	广州	21:15
T3	CZ6857	正班	B5241	B737	乌鲁木齐	9:00	塔城	21:30
T3	CZ6803	正班	B206C	B738	乌鲁木齐	9:05	喀什	23:35
T3	CZ6817	正班	B5283	B737	乌鲁木齐	9:05	和田	0:55
T3	CZ6863	正班	B7967	B738	乌鲁木齐	9:05	阿克苏	0:50
T3	CZ6949	正班	B1363	B738	乌鲁木齐	9:05	重庆	21:15
T3	CZ3436	正班	B1653	A320	乌鲁木齐	9:15	贵阳	21:40
T3	CZ6975	正班	B1237	B738	乌鲁木齐	9:15	南阳	0:55
T3	CZ6026	正班	B7997	B738	乌鲁木齐	9:30	北京	22:00
T3	CZ6981	正班	B2733	B787-8	乌鲁木齐	9:30	上海虹桥	0:20
T3	CZ6959	正班	B1410	B738	乌鲁木齐	9:55	西安	23:35
T3	CZ6679	正班	B5290	B737	乌鲁木齐	10:00	库尔勒	0:30
T3	CZ6805	正班	B5240	B737	乌鲁木齐	10:00	喀什	23:55
T3	CZ6859	正班	B5767	B738	乌鲁木齐	10:00	西宁	7:20

续表

航站楼	航班号	任务	机号	机型	起飞机场	预起时间	降落机场	预降时间
T3	CZ6240	正班	B301K	A20N	乌鲁木齐	10:05	沈阳	22:20
T3	CZ6925	正班	B1365	B738	乌鲁木齐	10:10	成都	7:00
T3	CZ6961	正班	B1736	B738	乌鲁木齐	10:10	银川	20:45
T3	CZ6979	正班	B7971	B738	乌鲁木齐	10:15	广州	0:40
T3	CZ6995	正班	B1748	B738	乌鲁木齐	10:20	上海虹桥	7:30
T3	CZ5691	正班	B7972	B738	乌鲁木齐	10:30	南京	23:40
T3	CZ6901	正班	B2727	B787	乌鲁木齐	10:30	北京	23:50
T3	CZ6005	正班	B7968	B738	乌鲁木齐	11:35	比什凯克	22:45
T3	CZ5355	正班	B1782	B738	乌鲁木齐	11:40	三亚	7:30

（责任编辑：段桃）