

# 基于 K-均值聚类的多雷达数据融合算法研究

舒红平, 王运锋, 徐振明, 邹书蓉

SHU Hong-ping, WANG Yun-feng, XU Zhen-ming, ZOU Shu-rong

1.四川大学 计算机学院, 成都 610064

2.成都信息工程学院 计算机系, 成都 610041

1.College of Computer Science, Sichuan University, Chengdu 610064, China

2.Department of Computers, Chengdu University of Information Technology, Chengdu 610041, China

E-mail: cqshp@cuit.edu.cn

SHU Hong-ping, WANG Yun-feng, XU Zhen-ming, et al. Research on multiradar data fusion algorithm based on K-mean clustering. Computer Engineering and Applications, 2007, 43(16): 183-185.

**Abstract:** This paper studies the discrimination of the different kind goal observation data using the k-mean clustering method, and realizes to the multi-objectives real-time track through the kind of data fusion. The basic thought, cluster treating processes and the algorithm realization of the observation data k-mean clustering are studied. Filter Equations for maneuvering target tracking are described; Parameter matrix theory for the Simplified calculation and corresponding initial value are given in air traffic control system. The discovery which the K-mean clustering be able to discriminate different goal well, and after clustering the kind to carry on the track fusion to be more accurate are found. The simulation result indicates that, the filter track trace is good after the K-mean clustering.

**Key words:** air traffic control; K-mean clustering; data fusion

**摘 要:** 应用 K-均值聚类的方法区分源于不同目标的观测数据, 通过类间数据融合, 实现对多目标的实时跟踪。研究了观测数据 K-均值聚类的基本思想、聚类处理过程及算法实现, 讨论了对机动目标跟踪的 Kalman 滤波方程及空管系统中易于计算的各参数矩阵理论依据及相应的初值。发现通过 K-均值聚类能很好区分不同目标, 聚类后再进行跟踪融合更加准确。仿真结果表明, 经 K-均值聚类处理后的滤波跟踪航迹效果较好。

**关键词:** 空中交通管理; K-均值聚类; 数据融合

文章编号: 1002-8331(2007)16-0183-03 文献标识码: A 中图分类号: TP301

## 1 引言

多雷达数据融合处理是整个雷达信息处理的中心, 它把多通道雷达接口处理单元送来的多路雷达信息进行实时融合处理成综合雷达数据, 并将处理结果送往冲突、低高、区域预警系统及有关管制员席位作进一步处理。

多部雷达同时监控某一区域的多批目标时, 由于通信延迟, 雷达天线扫描周期不同及各雷达系统误差不同等多种原因, 在某一时刻情报融合中心接收到的多部雷达测量数据呈现团状, 且分布在目标真实值的周围。如何做到既要区分同一传感器观测的不同目标, 又要关联不同传感器观测的同一目标就成为数据融合中目标观测数据关联的难点。当目标数量少, 且相互位置较远时容易实现; 而当目标数量多且相互位置较近时, 由于信息处理中心接收到的同一传感器对不同目标的观测数据可能比不同传感器对同一目标的观测数据相似度更近, 此

时关联就容易产生错误。为避免产生错误的关联结果, 本文提出通过 K-均值聚类分析方法对目标观测数据进行聚类, 通过对每一时刻传感器数据进行聚类, 以区分源于不同目标的数据, 通过类间数据融合, 实现对多目标的实时跟踪。

## 2 测量数据的 K-均值聚类

聚类需要解决的问题是将已给定的若干无标记的模式聚集起来使之成为有价值的类<sup>[1]</sup>。利用动态聚类方法, 对各雷达站录取的观测数据点聚类, 类的个数就是发现目标的批数。此时, 每个时间周期的类的个数可能不同, 目标观测数据到达时, 会出现下述情况<sup>[2]</sup>:

(1) 初始阶段, 没有形成任何目标航迹, 所以接收到的观测数据都是新目标, 建立新航迹;

(2) 已经建立了目标航迹, 有些观测数据是这些目标航迹

基金项目: 国家自然科学基金(the National Natural Science Foundation of China under Grant No.60572175); 四川省青年科技基金(No.06ZQ026-054); 成都信息工程学院自然科学与技术发展基金项目(No.CSRF200504)。

作者简介: 舒红平(1974-), 男, 副教授, 博士研究生, 主要研究领域为数据融合、数据挖掘及智能决策等; 王运锋(1975-), 男, 博士, 从事雷达信号(C)1994处理、多源数据融合等方面的研究; 徐振明(1956-), 副教授, 从事可视化项目管理方面的研究; 邹书蓉(1964-), 副教授, 从事基于网络的计算机应用研究。

的后续点,有些是新目标的观测点;

(3)所有的观测数据都是已建立的目标航迹后续点。

上述3种情况可以根据类的个数来确定。具体确定方法如下:

(1)当类的个数大于目标航迹数时,此时有新目标出现;

(2)当类的个数小于等于目标航迹数时,这些点迹分别分配给目标航迹;

(3)当目标航迹数为零时,根据类的个数建立新航迹。

聚类之前,先假定所有雷达测量数据已经被校准过,所有传感器的扫描周期相等,如果不等,需进行时空对准处理,且要统一转换到一个统一的坐标系下,故需进行坐标变换处理。

## 2.1 算法基本思想

K-均值聚类法首先由 MacQueen 提出,它以 K 为参数,将 n 个对象分成 K 个类,以使类内具有较高的相似度,而类间的相似度较低。其相似度的计算根据一个类中对象的平均值来进行。此方法能有效地处理类内密集,但类间区别明显的数据的聚类,其时间复杂度为  $O(nkt)$  (其中 t 是迭代次数),因此有相对较高的可伸缩性和高效率。

假定  $N_i$  为第  $i_p$  聚类  $C_i$  中的样本数目,  $m_i$  是这些样本的均值。若分成 k 类,则各聚类的均值表示为:

$$m_i = \frac{1}{N_i} \sum_{p \in C_i} p \quad (1)$$

误差平方和准则定义如下:

$$E = \sum_{i=1}^k \sum_{p \in C_i} |p - m_i|^2 \quad (2)$$

上式中的 E 是雷达观测数据中所有对象的平方误差的总和, p 是空间中的点,表示给定的雷达数据对象,  $m_i$  是聚类  $C_i$  的平均值(p 和  $m_i$  都是多维的)。上式准则试图使生成的结果取聚类尽可能地紧凑和独立。对于不同的分类, k 值不同,则 E 的值就不相同,使 E 达到最小的分类数 k 认为是当前的目标数。

在第一个时间周期后,首先要进行初始分类,根据聚集情况选择代表点。在以后的每一个时间周期内,只需以前面的分类数为初始分类数,综合航迹当前点为代表点。代表点确定后,则自成一类,将其它原始情报点依顺序归入与其距离最近的代表点一类,并重新计算该类的重心以代替原来的代表点,依此类推,直到原始情报点都归到相应的类中为止。分类完成后,则要进行类代表点与上一时间周期内形成的综合航迹归并,从第一个代表点开始,用站号和批号去匹配已建立的航迹,如果有一批匹配,则是该航迹的延续点,否则建立新批和主辅批档案。依此类推,将所有的代表点匹配完。

## 2.2 K-均值聚类处理过程

针对雷达观测数据,利用 K-均值聚类处理过程如图 1 所示。

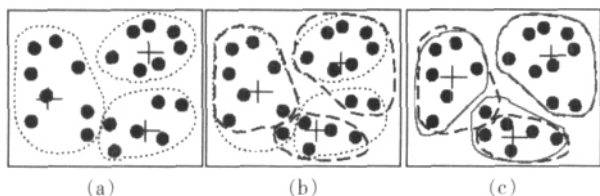


图1 雷达观测数据 K-均值聚类处理过程图

(1)从观测数据集中任意选择 3 个观测值作为初始聚类中心(图 1 中这些观测值被标上了“+”);其余观测值则根据与这 3 个聚类中心的距离,根据最近距离原则,逐个分别聚类到这 3 个聚类中心所代表的(3 个)聚类中;由此获得了如图 1(a)所示的 3 个聚类(以虚线圈出)。

(2)在完成第一轮聚类之后,各聚类中心发生了变化;继而更新 3 个聚类的聚类中心(图 1(b)中这些观测点被标上了“+”);也就是分别根据各聚类中的观测值计算相应聚类的均值。根据所获得的 3 个新聚类中心,以及各对象与这 3 个聚类中心的距离,根据最近距离原则对所有观测值进行重新归类。有关变化情况如图 1(b)所示(用粗虚线圈出)。

(3)重复上述过程就可获得如图 1(c)所示的聚类结果(用实线圈出)。这时由于各聚类中观测值归属已不再变化,整个聚类操作结束。

## 2.3 算法实现

基于 K-均值的雷达数据聚类算法分为 7 步,具体算法如下:

Step1 确定初始分类,假设为 k 类,根据聚集情况任意选择 k 个观测值作为初始的聚类中心代表点,计算每个聚类的均值  $m_1, m_2, \dots, m_k$  和  $E$ 。

Step2 Repeat

Step3 选择一个备选区观测值 p, 设 p 现在在  $C_i$  中。

Step4 计算

$$\rho = \begin{cases} \frac{N_i}{N_i+1} |p - m_i|^2 & j=i \\ \frac{N_j}{N_j-1} |p - m_j|^2 & j=i \end{cases} \quad (3)$$

Step5 对于所有的 j, 若  $\rho_i < \rho_j$ , 表示 p 应归入  $C_i$  类,需将 p 从  $C_j$  移到  $C_i$  中。

Step6 重新计算  $m_i$  和  $m_j$  的值,并修改 E。

Step7 Until 连续迭代 N 次 E 不再发生变化。

## 3 数据融合和目标跟踪

### 3.1 雷达观测数据融合

飞机在空中飞行时可以设定它是等速运动的,由机动和大气湍流引起的加速可以看作对等速运动的一种摄动。机动的快速程度用  $\alpha = \frac{1}{T_c}$  来反映,  $T_c$  是相关函数的时间常数。对于快速程度不同的机动,  $\alpha$  取值不同。在一个  $T_c$  时间内,可得到 N 个观测值构成的一个类  $C = \{Z^i(t_k) | i=1, 2, \dots, N\}$ 。为了充分利用各个雷达目标发现的信息,以便得到更加精确的目标航迹,应对这 N 个测量进行融合。采用状态方程描述目标的运动特性为:

$$\begin{cases} X^i(t_{k+1}) = \Phi(T, \alpha) X^i(t_k) + U^i(t_k) \\ Z^i(t_k) = HX^i(t_k) + W^i(t_k) \end{cases} \quad (4)$$

式中:  $U^i(t_k)$  为非齐次驱动输入,是一种离散的白噪声矢量序列。

$W^i(t_k)$  为  $t_k$  时刻的观测噪声。

$$E[U^i(t_k)] = 0, E[W^i(t_k)] = 0, E[W^i(t_k) W^{iT}(t_k)] = R_i = \sigma_w^2(i) I$$

$$E[U^i(t_k) U^{iT}(t_k)] = Q$$

在式(4)中,由于每部雷达的测量误差不尽相同,因此  $\sigma_w^2$

(i) ( $i=1, 2, \dots, N$ ) 可能两两不等。作线性组合:

$$Z(t_k) = \sum_{i=1}^N \alpha_i Z^i(t_k) \quad (5)$$

且令  $\sum_{i=1}^N \alpha_i = 1$ , 可以得到<sup>[9]</sup>:

$$\alpha_i = \frac{1}{\sigma_w^2(i) \sum_{j=1}^N \frac{1}{\sigma_w^2(j)}} \quad (6)$$

### 3.2 目标跟踪

基于式(4)对目标进行跟踪, 进行目标状态估计和预测。卡尔曼滤波是一种线性、无偏, 以误差方差最小为准则的最优估计算法, 利用卡尔曼滤波器可得到基于融合测量的目标状态滤波值。

滤波估值方程为:

$$\hat{X}(k+1|k+1) = \hat{X}(k+1|k) + K(k+1)[Z(k+1) - H(k+1)\Phi(T, \alpha)\hat{X}(k|k)] \quad (7)$$

预测估值方程为:

$$\hat{X}(k+1|k) = \Phi(T, \alpha)\hat{X}(k|k) \quad (8)$$

增益方程为:

$$K(k+1) = P(k+1|k)H^T(k+1)[H(k+1)P(k+1|k)H^T(k+1) + R(k+1)]^{-1} \quad (9)$$

滤波估值协方差矩阵方程为:

$$P(k+1|k+1) = [I - K(k+1)H(k+1)]P(k+1|k) \quad (10)$$

预测估值协方差矩阵方程为:

$$P(k+1|k) = \Phi(T, \alpha)P(k|k)\Phi^T(T, \alpha) + Q(k) \quad (11)$$

## 4 仿真结果

应用式(7)~式(11)的目标跟踪算法, 取观测转移矩阵为:

$$\Phi(T, \alpha) = \begin{bmatrix} 1 & T & \frac{T^2}{2} & 0 & 0 & 0 \\ 0 & 1 & T & 0 & 0 & 0 \\ 0 & 0 & e^{-\alpha T} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & T & \frac{T^2}{2} \\ 0 & 0 & 0 & 1 & T & 0 \\ 0 & 0 & 0 & 0 & 0 & e^{-\alpha T} \end{bmatrix} \quad (12)$$

取观测噪声协方差矩阵为:

$$Q(k) = \sigma_m^2 \begin{bmatrix} q_{11} & q_{12} & q_{13} & 0 & 0 & 0 \\ q_{21} & q_{22} & q_{23} & 0 & 0 & 0 \\ q_{31} & q_{32} & q_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & q_{11} & q_{12} & q_{13} \\ 0 & 0 & 0 & q_{21} & q_{22} & q_{23} \\ 0 & 0 & 0 & q_{31} & q_{32} & q_{33} \end{bmatrix}$$

式中

$$q_{11} = (1 - e^{-2\alpha T} + 2\alpha T + 2\alpha^3 T^3 / 3 - 2\alpha^2 T^2 - 4\alpha T e^{-\alpha T}) / 2\alpha^5$$

$$q_{12} = (1 + e^{-2\alpha T} - 2e^{-\alpha T} + 2\alpha T e^{-\alpha T} - 2\alpha T + \alpha^2 T^2) / 2\alpha^4$$

$$q_{13} = (1 - e^{-\alpha T} - 2\alpha T e^{-\alpha T}) / 2\alpha^3 = q_{31}$$

$$q_{22} = (4e^{-\alpha T} - 3e^{-2\alpha T} + 2\alpha T) / 2\alpha^3$$

$$q_{23} = (1 + e^{-2\alpha T} - 2e^{-\alpha T}) / 2\alpha^2 = q_{32}$$

$$q_{33} = (1 - e^{-2\alpha T}) / 2\alpha$$

取估值协方差矩阵初值为:

$$P_0 = \begin{bmatrix} \sigma_{x2}^2 & \frac{\sigma_{x2}^2}{T} & 0 & \sigma_{xy2}^2 & \frac{\sigma_{xy2}^2}{T} & 0 \\ \frac{\sigma_{x2}^2}{T} & \frac{\sigma_{x1}^2 + \sigma_{x2}^2}{T^2} & 0 & \frac{\sigma_{xy2}^2}{T} & \frac{\sigma_{xy1}^2 + \sigma_{xy2}^2}{T^2} & 0 \\ 0 & 0 & \sigma_m^2 & 0 & 0 & 0 \\ \sigma_{xy2}^2 & \frac{\sigma_{xy2}^2}{T} & 0 & \sigma_{y2}^2 & \frac{\sigma_{y2}^2}{T} & 0 \\ \frac{\sigma_{xy2}^2}{T} & \frac{\sigma_{xy1}^2 + \sigma_{xy2}^2}{T^2} & 0 & \frac{\sigma_{y2}^2}{T} & \frac{\sigma_{y1}^2 + \sigma_{y2}^2}{T^2} & 0 \\ 0 & 0 & 0 & 0 & 0 & \sigma_m^2 \end{bmatrix}_{6 \times 6}$$

取量测矩阵为:

$$H(k) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}$$

应用 2.3 小节的 K- 均值聚类算法, 利用上述的目标跟踪算法及雷达观测数据融合方法进行处理, 得到的融合航迹如图 2 所示。

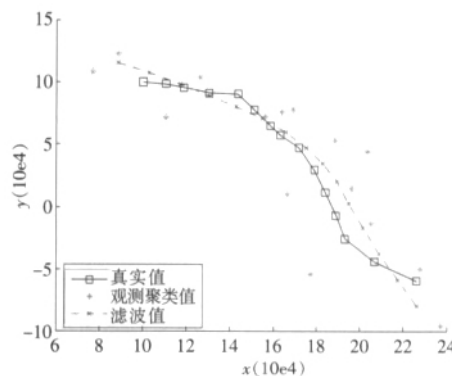


图 2 K-均值聚类滤波跟踪航迹

## 5 结束语

本文讨论了 K- 均值聚类的基本思想、K- 均值聚类处理过程及具体的算法实现步骤, 在对观测数据聚类处理基础上, 讨论了数据融合和目标跟踪的算法, 确定了滤波方程中的观测转移矩阵、观测噪声协方差矩阵、估值协方差矩阵初值及量测矩阵。通过 MATLAB 仿真, 得到了 K- 均值聚类处理后的滤波跟踪航迹图, 仿真结果表明, 对观测值进行 K- 均值聚类处理后的滤波跟踪航迹效果较好。(收稿日期: 2006 年 11 月)

## 参考文献:

- [1] MacQueen J. Some methods For classification and analysis of multivariate observations[J]. Proc 5th Berkeley Symp Math Statist Prox. 1967: 281-297.

- [2] Kaufman L, Rousseeuw P J. Finding groups in data: an introduction

(下转 195 页)

$$P_{S_{m+2}} = \frac{S_{m+1} + H_{m+1} - D_{m+1}}{C} = \frac{S_{m+2}}{C} \quad (3)$$

$$P_{g_m} = \frac{E_m}{N_m} \quad (4)$$

式中  $y$  表示年份;  $m$  表示月份;  $S_m$  表示月份  $m$  的起始蓄水量;  $I_m$  表示月份  $m$  的水库入水量;  $R_m$  表示月份  $m$  实际的供水量;  $D_m$  表示月份  $m$  的需水量;  $H_m$  表示月份  $m$  的预测流量;  $E_m$  表示月份  $m$  的总发电量;  $N_m$  表示月份  $m$  的需求电量;  $C$  表示水库容量。

采用成本函数作为目标规划模型, 系统决策者可以通过调整水价、电价或小时发电量来分析各个目标对于系统成本产生的影响, 从而得出系统的最佳成本收益。

## 4.2 效用函数

以成本函数作为目标规划模型解决多目标的水资源优化调度策略时, 在出现诸如水价补贴、电价补贴等情况的地区, 单纯以目标成本的优先级为决策依据是不切实际的。因此, 基于决策者偏好的效用函数可以用来解决多目标优化决策问题。在 WSDSS 系统中, 采用了需求驱动的随机动态决策模型, 实现实时水库调度优化策略。该模型不仅解决了系统需求预测的不确定性, 而且考虑到各种参数(如入水量、用水需求等)的季节性变化以及决策者偏好, 采用贝叶斯决策理论(BDT)按月份更新水库流的转换概率。该模型中的决策变量决定于系统状态  $D_m$ 、 $S_m$ 、 $I_m$  和  $R_m$ 。通过对各个目标根据重要性分配不同的权重值, 得出优化的调度策略。效用函数  $f_{y,m}(\cdot)$  反映了在多目标的相互作用下, 在某年度  $y$  的某月份  $m$  期望的最小损耗值。

$$f_{y,m}(D_m, S_m, I_m, H_{m+1}) = \min(E(D_{m+1}) \cdot [C_{y,m}(Pr_m, D_{S_{m+2}}, P_{g_m}) + (E(I_{m+1}|H_{m+1}, I_m) \cdot \{E(H_{m+2}|I_{m+1}) \cdot [f_{y,m}(D_{m+1}, S_{m+1}, I_{m+1}, H_{m+2})]\})]) \quad (5)$$

式中,  $y$  表示年份;  $m$  表示月份; 成本函数  $C_{y,m}(\cdot)$  表示月份  $m$  的供水成本、月份  $m+1$  的期望蓄水成本以及月份  $m$  的发电成本之和。  $D_m$  表示月份  $m$  的需水量;  $S_m$  表示月份  $m$  的起始蓄水量;  $I_m$  表示月份  $m$  的水库入水量;  $H_m$  表示月份  $m$  的预测流量;  $E$  表示数学期望。

采用基于决策者偏好的效用函数作为目标规划模型, 可以充分利用各类预测方法和工具提供更加灵活和适应性的决策, 在真实的水库约束条件的基础上提出优化的资源调度策略。图 4 显示了利用成本函数和效用函数得到的年发电量曲线, 为了简化问题分析, 对民用供水量、工业供水量、农业供水量以及洪水调控等赋予相同的权重值, 使得多项目标具有相同的优先级。图 3 反映出应用效用函数得到的发电量高于成本函数的结果, 即由效用函数得到的水资源优化策略要优于成本函数。

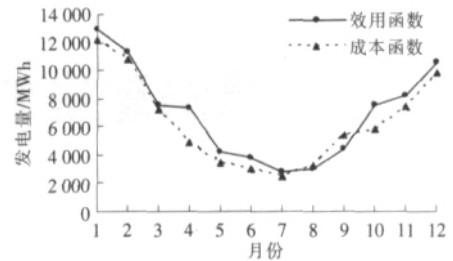


图 3 成本函数和效用函数决策结果比较

成本函数基于各关键决策目标的成本以及系统资源调度及维护成本, 主要通过对水价和电价的合理有效的调整, 产生水资源调度的优化策略。而效用函数基于多目标决策理论, 使决策者可以把明确或不明确的各目标优先级与决策支持策略相关联, 根据对不同目标优先级的偏好程度做出决策结果。

## 5 结论

本文介绍了以供水调度为中心的山西张峰水库资源调度决策支持系统的体系结构, 并采用多目标决策方法为水资源优化调度提供决策支持。由于系统的复杂性, 本文侧重介绍了成本函数和效用函数, 分别从经济成本收益的角度和决策者对各目标的偏好做出优化的目标决策结果。WSDSS 系统的研究开发提高了张峰水库资源调度决策的灵活性, 并能够从决策者的角度对水资源的供给调配做出决策。然而由于水库资源调度的决策问题时一个复杂的系统问题, 受多种环境的、人为的因素的制约, 需要根据水库的运行状态以及环境实际需要进一步改进和充实模型库和方法库, 以便更好地应用于水资源调度辅助决策中。(收稿日期: 2007 年 3 月)

## 参考文献:

- [1] Palomo J, Rios Insua D, Salewicz K A. Reservoir management decision support[C]//iEMSs 2002 Integrated Assessment and Decision Support, Lugano, Switzerland, 2002: 229-234.
- [2] Westphal K S, Vogel R M, Kirshen P, et al. Decision support system for adaptive water supply management[J]. Journal of Water Resources Planning and Management, 2003: 165-177.
- [3] Chung Tae Hwan, Abraham D M, Gokhale S B. Decision support system for microtunneling applications[J]. Journal of Construction Engineering and Management, 2004: 835-843.
- [4] Karamouz M, Zahraie B, Araghinejad S. Decision support system for monthly operation of hydropower reservoirs: a case study[J]. Journal of computing in civil engineering, 2005: 194-207.
- [5] 北京交通大学电子信息工程学院, 水利部山西水利水电勘测设计研究院. 山西张峰水库综合信息管理与决策支持系统研究报告[R], 2004-07.
- [6] 舒红平. 空中交通管理系统中航迹发散判据及抑制处理[J]. 控制与决策, 2005, 20(3): 345-348.
- [7] Han Jiawei, Kamber M. Data mining: concept and techniques[M]. [SI.]: Morgan Kaufmann Publishers, 2001.
- [8] Shu Hong-ping. Research on Multirada Data Fusion Algorithm based on K-central Clustering[C]//IECT 2006, Cheng du, Southwest Jiaotong University, September 28, 2006: 47-50.
- [9] to cluster analysis[M]. New York: John Wiley & Sons, 1990.
- [3] 王海, 王忠民. 一种基于密度和网格的聚类算法在 KDD 中的应用[J]. 计算机工程与应用, 2004, 40(24): 179-183.
- [4] 田启明, 王丽珍, 尹群. 基于网格距离的聚类算法的设计、实现和应用[J]. 计算机应用, 2005, 25(2): 295-297.
- [5] 赵国庆. 雷达对抗原理[M]. 西安: 西安电子科技大学出版社, 1999.

(上接 185 页)