动态规划(Dynamic Programming, DP)

作者：凯鲁嘎吉 - 博客园 http://www.cnblogs.com/kailugaji/

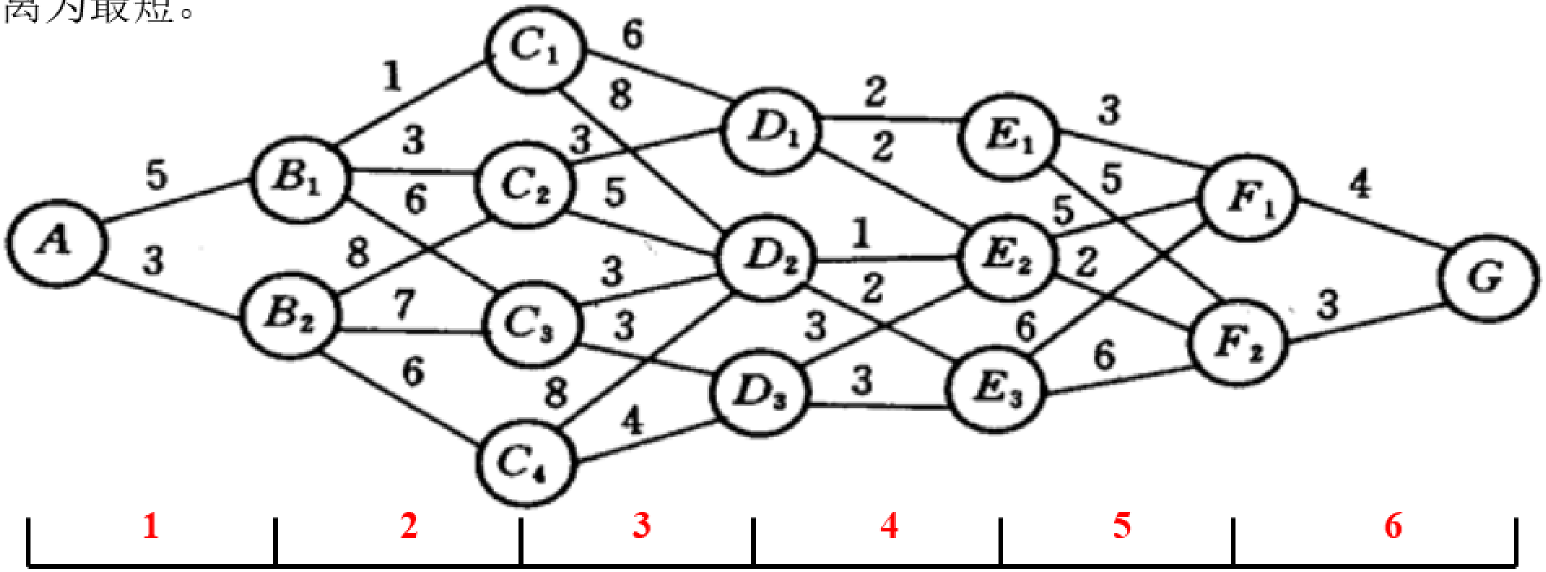
动态规则是运筹学的一个分支, 它是解决多阶段决策过程最优化的一种数学方法。大约产生于20世纪50年代。1951年美国数学家贝尔曼(R .Bellman)等人, 根据一类多阶段决策问题的特点, 把多阶段决策问题变换为一系列互相联系的单阶段问题, 然后逐个加以解决。与此同时, 他提出了解决这类问题的"最优性原理", 研究了许多实际问题, 从而创建了解决最优化问题的一种新的方法——动态规划(Dynamic Programming, DP)。他的名著"动态规划"于1957年出版, 该书是动态规划的第一本著作。

动态规划模型的分类, 根据多阶段决策过程的时间参量是离散的还是连续的变量, 过程分为离散决策过程和连续决策过程。根据决策过程的演变是确定性的还是随机性的, 过程又可分为确定性决策过程和随机性决策过程。组合起来就有离散确定性、离散随机性、连续确定性、连续随机性四种决策过程模型。本博客主要参考[1], 研究离散决策过程, 介绍动态规划的基本概念、理论和方法, 并通过最短路线问题来说明它的应用。

1．问题陈述——最短路线问题及穷举法求解

➢ 问题陈述：最短路线问题

- 给定一个线路网络，两点之间连线上的数字表示两点间的距离，试求一条由A到G的铺管线路，使总距离为最短。



➢ 求解一：穷举法

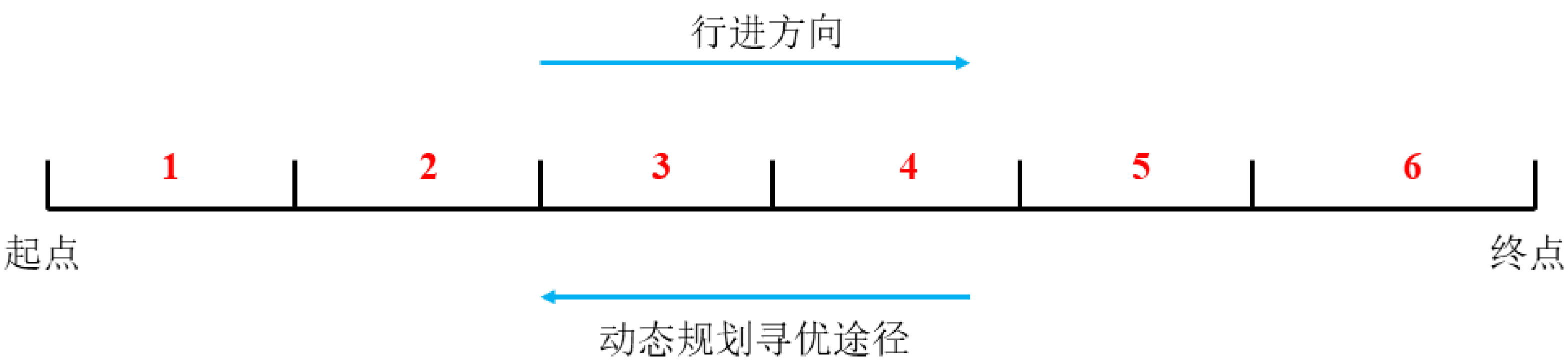
- 把由A到G所有可能的每一条路线的距离都算出来，然后互相比较找出最短者，相应地得出了最短路线。这样，由A到G的6个阶段中，一共有2×3×2×2×2×1 = 48条不同的路线，比较48条不同的路线的距离值，才找出最短路线为$A\to B_1\to C_2\to D_1\to E_2\to F_2\to G$，相应最短距离为18。

参考：《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.

2．动态规划求解

➢ 求解二：动态规划

- 穷举法计算是相当繁杂的。如果当段数很多，各段的不同选择也很多时，这种解法的计算将变得极其繁杂，甚至在电子计算机上计算都是不现实的。因此，为了减少计算工作量，需要寻求更好的算法，这就是下面要介绍的动态规划的方法。
- 动态规划方法的基本思想：若找到了$A \to B_1 \to C_2 \to D_1 \to E_2 \to F_2 \to G$是由A到G的最短路线，则$D_1 \to E_2 \to F_2 \to G$应该是由$D_1$出发到G点的所有可能选择的不同路线中的最短路线。
- 根据最短路线这一特性，寻找最短路线的方法，就是从最后一段开始，用由后向前逐步递推的方法，求出各点到G点的最短路线，最后求得由A点到G点的最短路线。所以，动态规划的方法是从终点逐段向始点方向寻找最短路线的一种方法。



参考：《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.

➢ 求解二：动态规划

- 先介绍动态规划的基本概念：

阶段

- 把所给问题的过程，恰当地分为若干个相互联系的阶段，以便能按一定的次序去求解。描述阶段的变量称为阶段变量，常用k表示。阶段的划分，一般是根据时间和空间的自然特征来划分，但要便于把问题的过程能转化为多阶段决策的过程。最短路线问题可分为6个阶段来求解，k分别等于1、2、3、4、5、6。

状态

- 状态表示每个阶段开始所处的自然状况或客观条件，它描述了研究问题过程的状况，又称不可控因素。最短路线问题的状态就是某阶段的出发位置。它既是该阶段某支路的起点，又是前一阶段某支路的终点。通常一个阶段有若干个状态，第一阶段有一个状态就是点A，第二阶段有两个状态，即点集合$\{B_1, B_2\}$，一般第k阶段的状态就是第k阶段所有始点的集合。状态应具有下面的性质：如果某阶段状态给定后，则在这阶段以后过程的发展不受这阶段以前各段状态的影响。换句话说，过程的过去历史只能通过当前的状态去影响它未来的发展，当前的状态是以往历史的一个总结。这个性质称为无后效性(即马尔科夫性)。

决策

- 决策表示当过程处于某一阶段的某个状态时，可以作出不同的决定(或选择)，从而确定下一阶段的状态，这种决定称为决策。在最优控制中也称为控制。描述决策的变量，称为决策变量。它可用一个数、一组数或一向量来描述。常用$u_k(s_k)$表示第k阶段当状态处于s_k时的决策变量。

参考：《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.

➢ 求解二：动态规划

策略

- 策略是一个按顺序排列的决策组成的集合。由过程的第k阶段开始到终止状态为止的过程，称为问题的后部子过程(或称为k子过程)。当$k = 1$时，此决策函数序列称为全过程的一个策略，简称策略，记为$p_{1,n}(s_1)$。即$p_{1,n}(S_1) = \{u_1(s_1), u_2(s_2), \ldots, u_n(s_n)\}$

状态转移方程

- 状态转移方程是确定过程由一个状态到另一个状态的演变过程。若给定第k阶段状态变量s_k的值，如果该段的决策变量u_k一经确定，第$k + 1$阶段的状态变量s_{k+1}的值也就完全确定。即s_{k+1}的值随s_k和u_k的值变化而变化。这种确定的对应关系，记为$s_{k+1}= T_k(s_k, u_k)$，该式描述了由k阶段到$k + 1$阶段的状态转移规律，称为状态转移方程。T_k称为状态转移函数。最短路径问题的状态转移方程为$s_{k+1}=u_k(s_k)$。

指标函数和最优值函数

- 用来衡量所实现过程优劣的一种数量指标，称为指标函数。它是定义在全过程和所有后部子过程上确定的数量函数。常用$V_{k,n}$表示。即$V_{k,n} = V_{k,n}(s_k, u_k, s_{k+1}, \ldots, s_{n+1}), k = 1, 2, \ldots, n$，指标函数的最优值，称为最优值函数，记为$f_k(s_k)$。它表示从第$k$阶段的状态$s_k$开始到第$n$阶段的终止状态的过程，采取最优策略所得到的指标函数值。即$f_k(s_k) = \underset{\{u_k, \ldots, u_n\}}{opt} V_{k,n}(s_k, u_k, s_{k+1}, \ldots, s_{n+1})$. 其中“opt” 可根据题意而取min或max。

参考：《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.

➢ 求解二：动态规划

- 下面按照动态规划的方法，从最后一段开始计算，由后向前逐步推移至A 点
- $V_{k,n}$：在第k阶段由点s_k至终点G的距离。$d_k(s_k, u_k) = v_k(s_k, u_k)$：在第$k$阶段由点$s_k$到点$s_{k+1} = u_k(s_k)$的距离，如$d_5(E_1, F_1)$=3表示在第5阶段由点$E_1$到点$F_1$的距离为3。$f_k(s_k)$：从第$k$阶段点$s_k$到终点G的最短距离，如$f_4(D_1)$表示从第4阶段中的点$D_1$到点G的最短距离
- 当$k = 6$ 时，由F_1到终点G只有一条路线，故$f_6(F_1) = 4$。同理，$f_6(F_2) = 3$， $u_6(F_2) = G$
- 当$k = 5$ 时，出发点有E_1、E_2、E_3三个。若从E_1出发，则有两个选择：①至F_1；②至F_2，则

$$f_5(E_1) = \min\begin{Bmatrix} d_5(E_1, F_1) + f_6(F_1) \\ d_5(E_1, F_2) + f_6(F_2) \end{Bmatrix} = \min\begin{Bmatrix} 3+4 \\ 5+3 \end{Bmatrix} = 7$$

- 其相应的决策为$u_5(E_1) = F_1$，这说明，由E_1至终点G的最短距离为7，其最短路线是$E_1 \to F_1 \to G$
- 同理，从E_2和E_3出发，则有

$$f_5(E_2) = \min\begin{Bmatrix} d_5(E_2, F_1) + f_6(F_1) \\ d_5(E_2, F_2) + f_6(F_2) \end{Bmatrix} = \min\begin{Bmatrix} 5+4 \\ 2+3 \end{Bmatrix} = 5$$

- 其相应的决策为$u_5(E_2) = F_2$

$$f_5(E_3) = \min\begin{Bmatrix} d_5(E_3, F_1) + f_6(F_1) \\ d_5(E_3, F_2) + f_6(F_2) \end{Bmatrix} = \min\begin{Bmatrix} 6+4 \\ 6+3 \end{Bmatrix} = 9$$

- 且$u_5(E_3) = F_2$

参考：《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.

- 求解二：动态规划

- 类似地, 当$k=4$时, 有

$$f_4(D_1)=7, u_4(D_1)=E_2$$
$$f_4(D_2)=6, u_4(D_2)=E_2$$
$$f_4(D_3)=8, u_4(D_2)=E_2$$

当$k=3$时, 有

$$f_3(C_1)=13, u_3(C_1)=D_1$$
$$f_3(C_2)=10, u_3(C_2)=D_1$$
$$f_3(C_3)=9, u_3(C_3)=D_2$$
$$f_3(C_4)=12, u_3(C_4)=D_3$$

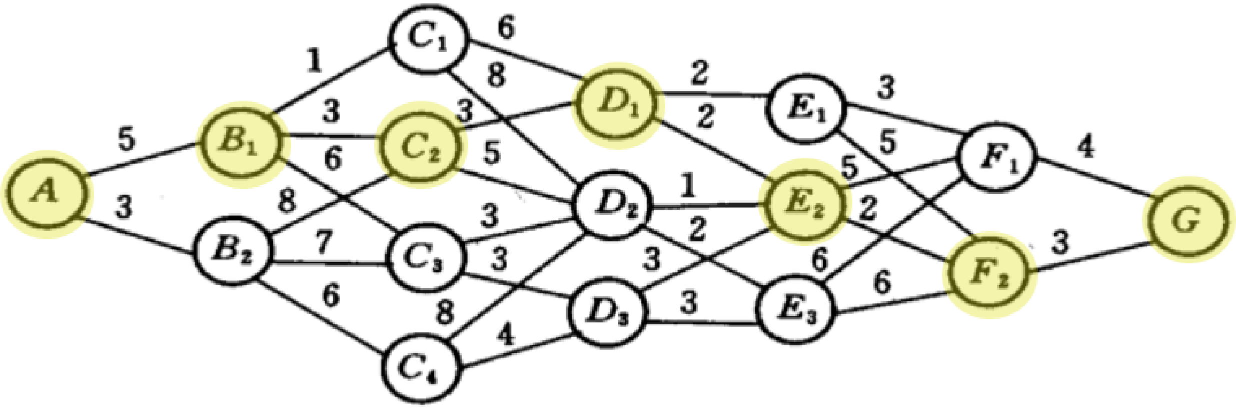
- 当$k=2$时, 有

$$f_2(B_1)=13, u_2(B_1)=C_2$$
$$f_2(B_2)=16, u_2(B_2)=C_3$$

当$k=1$时, 出发点只有一个A点, 则

$$f_1(A)=\min\begin{Bmatrix} d_1(A,B_1)+f_2(B_1) \\ d_1(A,B_2)+f_2(B_2) \end{Bmatrix}=\min\begin{Bmatrix} 5+13 \\ 3+16 \end{Bmatrix}=18$$

- 且$u_1(A) = B_1$。于是得到从起点A到终点G的最短距离为18。为了找出最短路线，再按计算的顺序反推之，可求出最优决策函数序列$\{u_k\}$，即由$u_1(A) = B_1$, $u_2(B_1) = C_2$, $u_3(C_2) = D_1$, $u_4(D_1) = E_2$, $u_5(E_2) = F_2$, $u_6(F_2) = G$组成一个最优策略。因而，找出相应的最短路线为$A \to B_1 \to C_2 \to D_1 \to E_2 \to F_2 \to G$



➢ 求解二：动态规划

- 从上面的计算过程中可以看出，在求解的各个阶段，我们利用了k阶段与$k+1$阶段之间的递推关系：

$$\begin{cases} f_k(s_k) = \min\limits_{u_k \in D_k(s_k)} \{d_k(s_k, u_k(s_k)) + f_{k+1}(u_k(s_k))\}, k = 6,5,4,3,2,1 \\ f_7(s_7) = 0 \end{cases}$$

- 一般情况,k阶段与$k+1$阶段的递推关系式可写为

$$\boxed{f_k(s_k) = \underset{u_k \in D_k(s_k)}{opt} \{v_k(s_k, u_k(s_k)) + f_{k+1}(u_k(s_k))\}, k = n, n-1, \cdots, 1}$$

- 其中边界条件为$\boxed{f_{n+1}(s_{n+1}) = 0}$
- 这种递推关系式称为动态规划的基本方程(也叫贝尔曼方程)。已知初始状态A，则按下面箭头所指的方向逐次变换有

$$\begin{array}{cccccccccccc} & & u_1(A) & & u_2(B_1) & & \cdots & & u_6(F_2) & \text{策略} \\ & \nearrow & \downarrow & \nearrow & \downarrow & \nearrow & & \nearrow & \downarrow & \\ A & & B_1 & & C_2 & & \cdots & & G & \text{状态(前提：无后效性)} \\ \text{(已知)} & & & & & & & & & \end{array}$$

- 从而可得最优策略为$\{u_1(A), u_2(B_1), \cdots, u_6(F_2)\}$，相应的最短路线为$A \to B_1 \to C_2 \to D_1 \to E_2 \to F_2 \to G$

参考：《运筹学》教材编写组.运筹学(第三版)[M].清华大学出版社,2005.

➢ 求解二：动态规划

□ 动态规划基本思想

(1) 动态规划方法的关键在于正确地写出基本的递推关系式和恰当的边界条件(简言之为基本方程)。要做到这一点，必须先将问题的过程分成几个相互联系的阶段，恰当地选取状态变量和决策变量及定义最优值函数，从而把一个大问题化成一族同类型的子问题，然后逐个求解。即从边界条件开始，逐段递推寻优，在每一个子问题的求解中，均利用了它前面的子问题的最优化结果，依次进行，最后一个子问题所得的最优解，就是整个问题的最优解。

(2) 在多阶段决策过程中，动态规划方法是既把当前一段和未来各段分开，又把当前效益和未来效益结合起来考虑的一种最优化方法。因此，每段决策的选取是从全局来考虑的，与该段的最优选择答案一般是不同的。

(3) 在求整个问题的最优策略时，由于初始状态是已知的，而每段的决策都是该段状态的函数，故最优策略所经过的各段状态便可逐次变换得到，从而确定了最优路线。

□ 在明确了动态规划的基本概念和思想之后，给一个实际问题建立动态规划模型时，需做到下面五点

(1) 将问题的过程划分成恰当的阶段；

(2) 正确选择状态变量s_k，使它既能描述过程的演变，又要满足无后效性；

(3) 确定决策变量u_k及每阶段的允许决策集合$D_k(s_k)$；

(4) 正确写出状态转移方程；

(5) 正确写出指标函数$V_{k,n}$的关系，它应满足下面三个性质：

① 是定义在全过程和所有后部子过程上的数量函数；

② 要具有可分离性, 并满足递推关系。即 $V_{k,n}(s_k,u_k,\ldots,s_{n+1})=\psi_k\left[s_k,u_k,V_{k+1,n}(s_{k+1},u_{k+1},\ldots,s_{n+1})\right]$

③ 函数 $\psi_k(s_k,u_k,V_{k+1,n})$ 对于变量$V_{k+1,n}$要严格单调。

参考：《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.

3. 动态规划 vs 穷举法

➢ 求解二：动态规划

- 以A为始端，G为终端，从G到A的解法称为逆序解法。以G为始端、A为终端的从A到G的解法称为顺序解法。
- 顺序解法和逆序解法只表示行进方向的不同或对始端终端看法的颠倒。但用动态规划方法求最优解时，都是在行进方向规定后，均要逆着这个规定的行进方向，从最后一段向前逆推计算，逐段找出最优途径。

➢ 动态规划的方法比穷举法有以下优点

- 减少了计算量。若用穷举法，就要对48条路线进行比较，运算在计算机上进行时，比较运算要进行47次；求各条路线的距离，即使用逐段累加方法，也要进行6 + 12 + 24 + 48 + 48 = 138次加法运算。用动态规划方法来计算，比较运算(从k = 5段开始向前算)共进行3 + 3 + 4 + 4 + 1 =15次。每次比较运算相应有两次加法运算，再去掉中间重复两次(即$B_1 \to C_1$，$B_2 \to C_4$各多算了一次)，实际只有28次加法运算。可见，动态规划方法比穷举法减少了计算量。而且随着段数的增加，计算量将大大地减少。
- 丰富了计算结果。在逆序(或顺序)解法中，我们得到的不仅仅是由A点(或G点)出发到G点(或A点)的最短路线及相应的最短距离，而且得到了从所有各中间点出发到G点(或A点)的最短路线及相应的距离。这就是说，求出的不是一个最优策略，而是一族的最优策略。这对许多实际问题来讲是很有用的，有利于帮助分析所得结果。

参考：《运筹学》教材编写组.运筹学(第三版)[M].清华大学出版社, 2005.

4. 参考文献

[1]《运筹学》教材编写组. 运筹学(第三版)[M]. 清华大学出版社, 2005.