

求解多维 0-1 背包问题的一种改进的遗传算法^{*})

曾 智 杨小帆 陈 静 陈文斌 唐荣旺

(重庆大学计算机学院 重庆 400044)

摘 要 针对多维 0-1 背包问题,通过应用贪心法和二分搜索法的思想,本文提出了一种新的杂交算子——中值杂交,并且基于此算子提出了求解多维 0-1 背包问题的一种改进的遗传算法。最后本文通过一系列数值实验,把改进算法与传统的遗传算法以及其他最新的遗传算法进行比较,经过对求得近似解的精度及计算所需时间两方面的对比,验证了其有效性。

关键词 多维 0-1 背包问题,遗传算法,中值杂交算子

An Improved Genetic Algorithm for the Multidimensional 0-1 Knapsack Problem

ZENG Zhi YANG Xiao-Fan CHEN Jing CHEN Wen-Bin TANG Rong-Wang

(Computer College of Chongqing University, Chongqing 400044)

Abstract This paper proposes a new crossover operator which is called median crossover. It is designed by using the ideas of the greedy algorithm and the 2-division search algorithm, and aims at the multidimensional 0-1 knapsack problem. Based on this median crossover operator, the paper also presents an improved genetic algorithm for solving the multidimensional 0-1 knapsack problem. Furthermore, the efficiency of the presented method is investigated by comparing it with the traditional simple genetic algorithm and other latest methods, in respect of the accuracy of an approximate solution and the processing time through series of numerical experiments.

Keywords Multidimensional 0-1 knapsack problem, Genetic algorithms, Median crossover

1 引言

多维 0-1 背包问题(the Multidimensional 0-1 Knapsack Problem, MKP)又名多约束 0-1 背包问题,是世界上最知名的整数规划问题之一,在最近这几十年来世界上许多研究机构和个人对它做了大量的研究。但是,由于 MKP 问题是一个 NP 难问题,很可能不存在多项式时间算法,因此现在获得的一些进步虽然可以解决一些中等规模的 MKP 问题,但是要完全解决它仍然是一个很大的挑战。

MKP 问题最早由 Lorie 和 Savage^[1]、Manne 和 Markowitz^[2]分别于 1955 年和 1957 年作为资金预算模型提出来。基本上, MKP 问题是一个资源分配的模型,它一般如下定义:

$$\begin{aligned} \max \quad & F = \sum_{j=1}^n c_j x_j \\ \text{s. t.} \quad & \sum_{j=1}^n a_{ij} x_j \leq b_i, i \in M = \{1, 2, \dots, m\}, \\ & x_j \in \{0, 1\}, j \in N = \{1, 2, \dots, n\}, \end{aligned} \quad (1)$$

其中 n 是可选物体的个数, m 是背包约束条件的个数。注意,对于多维 0-1 背包问题, a_{ij} 、 b_i 、 c_j 都是非负数。

由于 MKP 问题是一个 NP 难问题,当 m 和 n 都很大的时候,现有的大多数精确的和启发式的算法就失效了。一些精确算法只能工作于 m 和 n 很小的一类 MKP 问题,如 Gavish 和 Pirkul 提出的精确算法^[3]。Balas 和 Martin 提出了一种启发式的算法^[4],用于求解 MKP 问题,它把 MKP 问题的整数约束 $x_j \in \{0, 1\}$ 放松为 $0 \leq x_j \leq 1$,使之变为线形规划

问题,求解后再利用一些启发式的方法把解变回 0、1,并保持可行性。更多的例子见文[5]。

20 世纪 70 年代初期, Holland 基于自然界的遗传和自然选择,提出了遗传算法(Genetic Algorithms, GAs)^[6]。遗传算法作为搜索、优化、学习的一种方法,吸引了大量研究者的关注。对于求解 0-1 背包问题,遗传算法非常的适合。最近十几年,许多研究者将遗传算法用于求解多维 0-1 背包问题,如文[7~9]等。其中 Masatoshi Sakawa 和 Kosuke Kato 把染色体编码为双串(double strings),提出了一种使用双串的遗传算法(genetic algorithms with double strings, 即 GADS)^[8,9]。本文将基于贪心法和二分法的思想,提出一种改进的遗传算法(IGA)。

在第 2 节中,将对多维 0-1 背包问题进行分析,并对其初始化排序;在第 3 节中,将提出改进的遗传算法,包括新提出的中值杂交算子等组成部分;在第 4 节中,将通过数值实验对新的改进遗传算法进行评估;最后对本文进行总结,并给出今后工作的方向。

2 问题分析

对于 KMP 问题(1),假设其具有如下特性:

任给两个问题(1)的解 $x^1 = (x_0^1, x_1^1, \dots, x_{n-1}^1)^T$, $x^2 = (x_0^2, x_1^2, \dots, x_{n-1}^2)^T$, 设 $G(x^1)$ 、 $G(x^2)$ 分别为把 x^1 、 x^2 看成 n 位二进制数时所表示的数值。不妨设 $G(x^1) > G(x^2)$, 则如果 x^1 为可行解, x^2 一定为可行解,且 $F(x^1) > F(x^2)$; x^2 为不可行解, x^1 一定为不可行解。

^{*}) 重庆市自然科学基金资助课题(编号: CSTC, 2005BB2191)。曾 智 硕士研究生, 研究方向: 计算智能和神经网络; 杨小帆 教授、博导, 研究方向: 并行与容错计算、动力系统稳定性。

这时,我们称 KMP 问题(1)具有单调性。

当 KMP 问题(1)具有单调性时,它就可以使用二分搜索法求解,从 $(0,0,\dots,0)^T$ 到 $(1,1,\dots,1)^T$ 进行 n 次二分之后就能找到最优可行解。

但是一般的 KMP 问题(1)是不具有单调性的,二分法是不能应用到一般的 KMP 问题(1)的。所以,只能采用近似计算,让问题近似地具有此特性。这里,我们可以使用贪心法的思想,通过对每一位计算出其有效度,然后按照每位的有效度从大到小把问题的系数重新排序,这样让 KMP 问题(1)近似地具有单调性。

为了使问题具有单调性,解的高位对应的 c_j 应该更大,这样才能使解所对应的二进制数越大,所求函数值 F 越大。所以,每位的有效度应该随着其对应的 c_j 增大而增大。同时,解的高位对应的约束度也应该更大,这样才能使当 $G(x^1) > G(x^2)$, x^2 为不可行解时, x^1 成为不可行解的概率更大。所以,每位的有效度可以按照这样计算:

$$E(j) = c_j \cdot \epsilon + 500 \cdot CD(j) \cdot (1 - \epsilon) \quad (2)$$

其中约束度 $CD(j) = \frac{1}{m} \sum_{i=1}^m a_{ij} / b_i$, ϵ 为给定的系数。

计算出每位的有效度后,按照它从大到小把问题的系数重新排序。这样,按照有效度从大到小排序后的 KMP 问题就可以使用二分法近似地求得最优解了。

不过,这样求得的最优解精度不高,这主要是因为使用贪心法的思想一般只能达到局部最优,并不能达到全局最优。那么如何利用上面的思想,跳出局部最优,得到高精度的近似最优解呢? 启发式方法,特别是遗传算法为我们提供了这样的一种方法。下面,本文将提出一种使用以上思想的杂交算子——中值杂交算子,并以此为基础,提出求解多维 0-1 背包问题的一种改进的遗传算法(IGA)。

3 改进的遗传算法

3.1 染色体编码与解码

使用遗传算法求解多维 0-1 背包问题,染色体通常被编码为长度为 n 的 0-1 串^[7],但是这样编码很少产生原象是可行解的染色体,降低了遗传算法的能力。于是,Masatoshi Sakawa 和 Kosuke Kato 提出了一种使用双串的遗传算法(genetic algorithms with double strings,即 GADS)^[8,9],它把染色体编码为一种叫做双串(double strings)的结构,它通过特定的解码后产生的原象总是可行解。

本文提出的改进遗传算法(IGA)还是采用 0-1 串的编码方法,但是吸取 GADS 的解码方法。原象为不可行解的 0-1 串,将通过特定的解码算法,把它修补为近似最好的可行解。如何使其修补为近似最好的可行解呢? 这里,就需要对问题的每位按照下面的优先度再一次排序:

$$P(j) = c_j / CD(j) \quad (3)$$

其中约束度 $CD(j) = \frac{1}{m} \sum_{i=1}^m a_{ij} / b_i$

这次排序不像上次那样改变问题,而是得到序列 $\{k_0, k_1, k_2, \dots, k_n\}$, 其中 k_i 表示排序后的第 i 位为排序前的第 k_i 位。排序后,越在高位的,它的 c_j 越可能大,而约束越可能小,所以越应该优先选取。于是,我们可得解码算法如下,其中 s 为解码前的个体, x 为解码后的解, sum 是 m 维列向量, c_j 是约束系数矩阵的第 j 列列向量, b 为约束向量。

解码算法

Step 1: 令 $j := 0, sum := 0$ 。

Step 2: 如果 $s_{k_j} = 0$, 则 $x_{k_j} := 0, j := j + 1$, 然后跳转到 step 4; 否则, 如果 $s_{k_j} = 1$, 则执行 step 3;

Step 3: 如果 $sum + a_{k_j} \leq b$, 则 $x_{k_j} := 1, sum := sum + a_{k_j}, j := j + 1$, 然后执行 step 4; 否则, $x_{k_j} := 0, j := j + 1$, 然后执行 step 4;

Step 4: 如果 $j \geq n$, 则 stop; 否则, 跳转到 step 2。

3.2 适应度函数和选择算子

通常,对个体 s , 可以简单地用 $f(s) = \sum_{j=1}^n c_j \cdot s_j$ 求得其实适应度。但是, s 不一定是可行解, 所以应该先解码, 将其修补为可行解 x , 然后再用上式计算其适应度。于是得到本文改进遗传算法(IGA)的适应度函数:

$$f(s) = \sum_{j=1}^n c_j \cdot x_j \quad (4)$$

对种群中的每个个体计算出其适应度后, 就可以评估每个个体, 计算出其选择概率。选择概率应该按照如下公式计算:

$$p(s) = f(s) / \sum_{s \in P} f(s') \quad (5)$$

其中 $s \in P$ 表示 s 为种群中的一个个体。

遗传算法的选择算子有许多种, 如期望值选择(expected value selection)、赌轮选择(roulette wheel selection)等。本文改进遗传算法(IGA)采用精英赌轮选择(elitist roulette wheel selection), 它包括两方面: 精英主义和赌轮选择。

精英主义(Elitism): 如果上一代种群中的个体的适应度大于这一代种群中的每一个个体, 则将它保留到这一代种群中。

赌轮选择(Roulette wheel selection): 赌轮上面有许多格, 小球在上面滚动后会随机落在其中一个格子里。而小球落在哪个格子里的概率, 取决于格子的大小。同理, 我们按照种群中每个个体的选择概率划分它所代表的格子的大小, 然后滚动小球, 小球最后落在哪个格子里就代表选择哪个个体。进行 N 次后就选择出了 N 个个体。

3.3 杂交算子

可应用于遗传算法的杂交算子有许多种, 传统的有单点杂交、双点杂交等。Masatoshi Sakawa 和 Kosuke Kato 提出的使用双串的遗传算法(GADS)^[8,9]中, 使用了部分匹配杂交算子(Partially matched crossover, PMX)。本文在将问题(1)重新按位有效度排序后, 结合二分搜索算法的思想, 设计出了一种新的杂交算子——中值杂交算子。

杂交算法

Step 1: 令 $l := 0$;

Step 2: 从选择算子选出的 N 个个体中, 取出两个 s^1, s^2 , 其中 $f(s^1) > f(s^2)$;

令 $x := s^1, y := s^2$;

Step 3: 随机产生一个实数 r , 使 $0 \leq r \leq 1$, 如果 $r \leq p_c$, 则执行 step 4; 否则, 跳转到 step 5;

Step 4: 令 $y := \text{mean}(s^1, s^2)$;

Step 5: 保存 x 和 y 作为 s^1 和 s^2 杂交产生的后代, 然后令 $l := l + 1$;

Step 6: 如果 $l \geq N/2$, stop; 否则, 跳转到 step 2。

其中, p_c 为给定的杂交概率, 函数 $\text{mean}(x^1, x^2)$ 表示把 x^1 和 x^2 看作 n 位二进制数, 先相加, 然后除 2 (即右移一位), 求得它们的平均值。

在文[10]中,作者指出,进行杂交的两父个体,如果它们太接近,则属于近亲繁殖,会导致杂交失败,从而降低遗传算法的效率。在中值杂交算子中,如果两个个体相等,或者所对应的二进制数相差1时,将不会杂交产生新个体。所以,需要在杂交算法的 step 2 中对取出的两个个体进行比较,如果出现上述情况,则把适应度小的 s^2 进行修正,使它与 s^1 高位有不相同的位。

这样,使用中值杂交算子后,就产生了新一代种群的 N 个个体。

3.4 变异及加强算子

经过杂交产生的新一代种群,应该在一定概率下随机发生变异,这样有助于提出局部最优,达到全局最优。本文提出的改进遗传算法(IGA)同样需要变异算子,算法如下。

变异算法

Step 1: 令 $l := 0$;

Step 2: 令 $j := 0$;

Step 3: 随机产生一个实数 r , 使 $0 \leq r \leq 1$, 如果 $r \leq p_m$, 则执行 step 4; 否则跳转到 step 5;

Step 4: 如果 $s_j^l = 0$, 则 $s_j^l = 1$, 然后执行 step 5; 否则如果 $s_j^l = 1$, 则 $s_j^l = 0$, 然后执行 step 5;

Step 5: $j := j + 1$, 如果 $j \geq n$, 则执行 step 6; 否则, 跳转到 step 3;

Step 6: $l := l + 1$, 如果 $l \geq N$, 则 stop; 否则, 跳转到 step 2。

其中 N 是种群大小, s_j^l 表示种群中第 l 个个体的第 j 位, p_m 表示变异概率。

除了变异外,本文提出的改进遗传算法(IGA)还将使用一个加强算子来对种群中的个体进行一定概率下的随机加强,算法如下。

加强算法

Step 1: 令 $l := 0$;

Step 2: 随机产生一个实数 r , 使 $0 \leq r \leq 1$, 如果 $r \leq p_i$, 则执行 step 3; 否则跳转到 step 7;

Step 3: 如果 s^l 为不可行解, 则跳转到 step 4; 否则, 执行 step 4;

Step 4: 令 $j := 0$;

Step 5: 如果 $s_j^l = 0$, 则 $s_j^l = 1$; 如果这时 s^l 变为不可行解, 则 $s_j^l = 0$;

Step 6: $j := j + 1$, 如果 $j \geq n$, 则执行 step 7; 否则, 跳转到 step 5;

Step 7: $l := l + 1$, 如果 $l \geq N$, 则 stop; 否则, 跳转到 step 2。

其中 s^l 表示种群中第 l 个个体, p_i 表示变异概率。

3.5 算法运行及其终止

在算法开始的时候,先要随机地产生 N 个互不相同的个体,作为初试种群,然后依次对它们进行评估、选择、杂交、变异及加强后,产生新一代个体。如此循环 G (迭代次数)次后,在最后一代种群中选出适应度最大的个体,经过加强、解码后,即得到所求近似解。

4 数值实验及其结果

下面,为了检验本文提出的改进遗传算法(IGA)的可行性和有效性,将通过求解变量数分别为 50、100、150、200,约束条件数分别为 20、30、40、50 的多维 0-1 背包问题(1),来与

经典简单遗传算法(SGA)^[6] 和使用双串的遗传算法(GADS)^[8,9] 比较。

其中,问题(1)的参数 c_j 、 a_{ij} 分别为区间 $[0, 999]$ 内的随机数, b_i 由下式确定:

$$b_i = \gamma \cdot \sum_{j=1}^n a_{ij} \quad (6)$$

这里 γ 为一个给定的正实数,表示约束条件的严厉度。

本文的数值实验在一台 PC 机上进行,环境如下: CPU: P4 2.4G(C); Memory: 512MB; OS: Windows XP; 使用 Visual C++ 6.0 编译器。

IGA 算法参数设置如下: 种群大小 $N = 100$, 迭代次数 $G = 1600$, 杂交概率 $p_c = 0.9$, 变异概率 $p_m = 0.01$, 加强概率 $p_i = 0.3$, $\epsilon = 0.15$ 。

SGA 算法参数设置如下: 种群大小 $N = 100$, 迭代次数 $G = 5000$, 杂交概率 $p_c = 0.9$, 变异概率 $p_m = 0.01$ 。

GADS 算法参数设置如下: 种群大小 $N = 100$, 代沟 $G = 0.9$, 杂交概率 $p_c = 0.9$, 变异概率 $p_m = 0.01$, 反向变异概率 $p_r = 0.03$, $\epsilon = 0.02$, 最大迭代次数 $I_{\max} = 1000$, 最小迭代次数 $I_{\min} = 100$, $c_{\text{mult}} = 1.8$ 。

实验结果见表 1 和表 2。

表 1 求解 50 个变量、20 个约束条件的 KMP 问题(1)的计算结果(计算 10 次)

γ	方法	最好解	最差解	平均解	计算时间(s)
0.25	IGA	10125	10125	10125	1.391
	SGA	9385	8465	8838.6	2.463
	GADS	10125	10125	10125	1.875
0.5	IGA	18405	18349	18388.2	1.563
	SGA	17676	16999	17383.4	2.453
	GADS	18405	18405	18405	1.952
0.75	IGA	23922	23922	23922	1.890
	SGA	23588	22682	23137.5	2.953
	GADS	23922	23883	23908	2.028

表 2 求解 200 个变量、50 个约束条件的 KMP 问题(1)的计算结果(计算 10 次)

γ	方法	最好解	最差解	平均解	计算时间(s)
0.25	IGA	40196	39837	39959.3	7.395
	SGA	31297	27569	29059.3	11.578
	GADS	40112	37891	38774.5	8.763
0.5	IGA	70787	70095	70409.7	9.083
	SGA	51435	48413	50039.5	15.141
	GADS	70276	66774	68437.0	11.015
0.75	IGA	100734	100408	100577.4	12.531
	SGA	91257	83335	86921.8	21.782
	GADS	100245	98467	99512.7	15.439

从数值实验中可以清楚地看到,本文提出的改进遗传算法(IGA)明显优于传统的简单遗传算法,解的精确度更高,计算所花时间更短。同 GADS 算法相比也很有优势,解的精确度和计算时间都有提高,从而表明了本文提出的改进遗传算法(IGA)对于求解多维 0-1 背包问题不但可行,而且非常有效。

结论 本文针对多维 0-1 背包问题的特点,通过结合贪

一种新的核线性鉴别分析算法及其在人脸识别上的应用^{*}

郑宇杰¹ 杨静宇¹ 吴小俊² 王卫东¹ 张丽丽¹

(南京理工大学计算机系 南京 210094)¹ (江苏科技大学电子信息学院 镇江 212003)²

摘要 基于核策略的核 Fisher 鉴别分析(KFD)算法已成为非线性特征抽取的最有效方法之一。但是先前的基于核 Fisher 鉴别分析算法的特征抽取过程都是基于 2 值分类问题而言的。如何从重叠(离群)样本中抽取有效的分类特征没有得到有效的解决。本文在结合模糊集理论的基础上,利用模糊隶属度函数的概念,在特征提取过程中融入了样本的分布信息,提出了一种新的核 Fisher 鉴别分析方法——模糊核鉴别分析算法。在 ORL 人脸数据库上的实验结果验证了该算法的有效性。

关键词 核策略,核 Fisher 鉴别分析,模糊核 Fisher 鉴别分析,特征提取,人脸识别

A New Kernel Discriminant Analysis Algorithm and its Application to Face Recognition

ZHENG Yu-Jie¹ YANG Jing-Yu¹ WU Xiao-Jun² WANG Wei-Dong¹ ZHANG Li-Li¹

(Department of Computer Science, Nanjing University of Science & Technology, Nanjing 210094)¹

(School of Electronics and Information, Jiangsu University of Science & Technology, Zhenjiang 212003)²

Abstract Kernel Fisher discriminative analysis(KFD)algorithm based on kernel trick has been one of the effective nonlinear feature extraction methods. But all previous nonlinear feature extraction methods based on KFD algorithm which procedures are based on solving binary classification problem. How to extract effective discriminative information from overlapping(outlier)samples is still open. In this paper, a new KFD algorithm named fuzzy kernel discriminative analysis(FKFD)is proposed. In the proposed algorithm, fuzzy K-nearest neighbor(FKNN)algorithm is incorporated into the process of KFD algorithm and the corresponding fuzzy membership degrees are gained. Therefore, distribution information of samples is embedded in the proposed algorithm through fuzzy membership degrees. Experimental results on ORL face database demonstrate the effectiveness of the proposed algorithm.

Keywords Kernel trick, Kernel Fisher discriminative analysis, Fuzzy kernel Fisher discriminative analysis, Feature extraction, Face recognition

1 引言

基于核函数的学习算法已经成为目前模式识别和机器学习研究领域的研究热点,特别是随着统计学习理论研究的成熟,涌现了大量的基于核函数的学习算法,如支持向量机^[1](SVMs)、核主成分分析^[2]、核独立成分分析^[3]以及核线性鉴别分析^[4~6]等等^[7]。其中核线性鉴别分析由于可以抽取得到原始样本的有效非线性分类信息,被广泛应用于模式识别的各个领域。如 Mika 等提出了针对两类鉴别分析的方法;Baudat 等进一步提出了多类的核线性鉴别分析方法。最近, Yang 提出了一个有效的核线性鉴别分析的框架^[8,9]: KP-

CA+LDA,该框架揭示了核线性鉴别分析算法的本质,并且给出了一个简单的特征提取步骤。但是,先前的关于核线性鉴别分析算法的特征提取过程都是基于 2 值分类问题而言的,也就是说它是用来解决是否属于还是不属于某一类的特征提取问题。当数据样本存在重叠点或者离群点的情况下,这种简单的处理 2 值问题的算法显然会遇到困难。如何克服这个缺点以及从这些重叠(离群)样本中抽取得到有效的分类特征是本文要关注的问题。

基于对此类问题的关注,本文引入了模糊集^[10]的思想,提出了一种新的基于模糊集理论的模糊核线性鉴别分析算法。该算法充分利用了样本的分布信息,有助于更有效地得

^{*}国家自然科学基金资助(编号:60472060)、南京理工大学科研发展基金。郑宇杰 博士研究生。

心法和二分法思想,设计出了求解多维 0-1 背包问题的一种改进的遗传算法。它通过计算 KMP 问题(1)的位有效度,然后按位有效度重新排序,使其近似具有单调性。这样,使用含中值杂交算子等组成部分的改进遗传算法求解,能更快地收敛到近似最优解。实验结果表明,其对于求解多维 0-1 背包问题不但可行,而且非常有效。

参考文献

- Lorie J, Savage L. Three problem in capital rationing. Journal of Business, 1955, 28: 229~239
- Manne A, Markowitz H. On the solution of discrete programming problem. Econometrica, 1957, 25: 84~110
- Gavish B, Pirkul H. Efficient algorithms for solving multiconstraint 0-1 knapsack problems to optimality. Mathematical Programming, 1985, 31: 78~105
- Balas E, Martin C. Pivot and Complement - a Heuristic for 0-1

- Programming. Management Science, 1980, 26(1): 86~96
- Freville A. The multidimensional 0-1 knapsack problem: An overview. European Journal of Operational Research, 2004, 155: 1~21
- Holland J. Adaptation in Natural and Artificial Systems. MI: Ann Arbor, MI: University of Michigan Press, 1975; Cambridge, MA: MIT Press, 1992
- Goldberg D. Genetic Algorithms in Search, Optimization, and Machine Learning. Reading, MA: Addison-Wesley, 1989
- Sakawa M, Kato K, Shibano T. An interactive fuzzy satisficing method for multiobjective multidimensional 0-1 knapsack problems through genetic algorithms. In: Proceedings of 1996 IEEE International Conference on Evolutionary Computation, Piscataway, NJ, USA, 1996
- Sakawa M, Kato K. Genetic algorithms with double strings for 0-1 programming problems. European Journal of Operational Research, 2003, 144: 581~597
- Bian R, Chen Z, Yuan Z. Improved Crossover Strategy of Genetic Algorithms and Analysis of Its Performance. In: Proceedings of the 3th world congress on Intelligent Control and Automation, Hefei, P. R. China, 2000