

基于有序划分编码的图着色算法

韩丽霞^{1,2},王宇平²,兰绍江¹

(1. 中国矿业大学计算机科学与技术学院,江苏徐州 221116; 2. 西安电子科技大学计算机学院,陕西西安 710071)

摘要: 针对整数编码的冗余性,提出了求解图着色问题的一种新的编码方式.采用有序划分编码问题的解,编码后的个体具有与问题的潜在解一一对应的特点.与整数编码相比,新的编码避免了冗余性,将搜索空间缩小了 $k!$ 倍.对5个标准图着色问题的仿真结果表明,基于有序划分编码的新算法是求解图着色问题的一种有效的算法.

关键词: 图着色问题; 进化算法; 编码

中图分类号: TB18 **文献标识码:** A **文章编号:** 0372-2112 (2010) 01-0146-05

Graph Coloring Algorithm Based on Ordered Partition Encoding

HAN Li-xia^{1,2}, WANG Yu-ping², LAN Shao-jiang¹

(1. School of Computer Science and Technology, China University of Mining and Technology, Xuzhou, Jiangsu 221116, China;

2. School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi 710071, China)

Abstract: A novel encoding scheme was presented to solve the redundancy of the integer representation for the graph coloring problem. The ordered partition representation has the characteristic of one-to-one correspondence to the valid solution of the related problem. Compared with the integer representation, the new encoding avoided the redundancy completely and shrunk the search space $k!$ times of that of the integer representation. The simulation results on five standard benchmark problems demonstrated the proposed algorithm was effective for the graph coloring problem.

Key words: graph coloring problem; evolutionary algorithm; encoding

1 引言

图着色问题是一个具有良好应用背景的组合优化问题,在排课表、电路布线、频率安排、调度安排以及计算机存储等方面都有着广泛的应用.由于图着色问题是NP-难的问题^[1],不存在多项式时间的精确算法,因而,应用启发式算法寻求其近似解成为现实的解决方案.

近几年来,受自然界进化理论的启发,仿生算法如DNA算法^[2]、蚂蚁算法^[3]、进化算法^[4~10]成为研究的热点,并应用于求解图着色问题.高琳和许进^[2]通过对图的顶点和颜色进行恰当的编码,依据分子生物学的实验方法,提出了图着色问题的DNA算法;廖飞雄和马良^[3]根据蚂蚁找寻食物的行为,提出了求解图着色问题的蚂蚁算法,有效的避免了启发式搜索易陷入局部极小的缺陷;Galiner和Hao^[4]提出了求解图着色问题的混合进化算法,用禁忌算法提高杂交算子产生的后代个体,对大规模的图着色问题取得了不俗的成果;在此基础上,Celia和Adam^[5]提出了求解图着色问题的改进遗传算法,在求解性能和时间上都有所提高;Mumford^[6]提出了

图着色问题的序编码和序杂交算子,结合模拟退火算法,新的混合遗传算法缩短了找到问题近似解的时间.

进化算法是抽象于自然界生物体进化过程,通过全面模拟自然选择和进化机制,形成的一种具有‘生成+检验’特征的仿生优化算法.它以编码空间代替问题的参数空间,以适应度函数为评价依据,以种群为进化基础,以对种群中染色体的进化操作实现选择和遗传机制,建立起一个迭代过程.在进化过程中,染色体中重要的基因被有效的利用,从而使新一代的种群优于上一代的种群,群体中的染色体不断进化,逐渐逼近问题的最优解,达到优化的目的.目前,进化算法已被成功的应用于工程、计算机科学、图像处理等领域众多问题的求解,并且取得了很好的效果.而编码方式作为进化算法流程中的第一步,决定了进化算子的操作方式,对算法的效果具有极为重要的影响.根据图着色问题对顶点划分的本质特征,提出了图着色问题的有序划分编码方式,可以有效的避免整数编码的冗余性,将 k 种颜色的图着色问题的搜索空间缩小了 $k!$ 倍.同时,设计了简单的划分杂交算子和减少违反量的贪婪局部搜索算子,新的

混合进化算法有效的保留父代个体的特征,大大提高了求解大规模图着色问题的速度.对 5 个大规模标准算例的仿真结果表明新算法是求解图着色问题的一种有效的算法.

2 图着色问题

给定无环图 $G=(V, E)$ (其中 $V=\{v_1, v_2, \dots, v_n\}$ 为顶点集, $E=\{e_1, e_2, \dots, e_m\}$ 为边集), 图着色问题就是要用最少的颜色对图 G 的顶点进行着色, 满足相邻的顶点(同一条边的两个端点称为相邻顶点)着不同颜色的约束条件.

定义 1 设 $\pi=\{V_1, V_2, \dots, V_k\}$ 是集合 V 中的子集组成的集合, 若满足

- (1)(非空性) $V_i \neq \emptyset, \forall i$;
- (2)(完整性) $\bigcup_{i=1}^k V_i = V$;
- (3)(不交性) $V_i \cap V_j = \emptyset (i \neq j, \forall i, j)$;

则称集合 π 是集合 V 的一个 k -划分 (V_i 称为划分块). 若划分块 V_i 中的任意两顶点均不相邻, 则划分块 V_i 又称为独立集; 若相邻的两个顶点 u 和 v 同处于划分块 V_i 中, 则连接这两个顶点的边 $e=(u, v) \in E$ 称为违反边, 记为 $q(e)=1$; 否则 $q(e)=0$. 设划分块 $V_i (i=1, 2, \dots, k)$ 中违反边的数量为 $Q(V_i)$, 因此, k -划分 $\pi=\{V_1, V_2, \dots, V_k\}$ 的违反量为

$$w(\pi) = \sum_{i=1}^k Q(V_i) = \sum_{j=1}^m q(e_j)$$

不难发现, k -划分 $\pi=\{V_1, V_2, \dots, V_k\}$ 的违反量为 0 当且仅当各划分块 V_i 的违反量均为 0, 即所有划分块 $V_i (i=1, 2, \dots, k)$ 均为独立集.

定义 2 图 G 的一个 k -正常着色, 就是把图 G 的顶点集分成 k 个独立集的一个划分 $\pi=\{V_1, V_2, \dots, V_k\}$.

定义 3 使图 G 为 k -正常着色的最小颜色数 k 称为图 G 的色数, 记为 $\chi(G)$.

给定颜色集合 $C=\{c_1, c_2, \dots, c_k\}$, 图着色问题就是将图的顶点集 V 分为 k 个独立集的过程, 即寻找违反量为 0 的 k -划分 $\pi=\{V_1, V_2, \dots, V_k\}$ 的过程. 若找到, 令 $k=k-1$, 直到找到图的 $\chi(G)$ -正常着色为止.

3 有序划分编码

以 7 个顶点的图着色问题为例, 考虑整数编码的如下着色方案,

$$a_1 = (1213333), a_2 = (1312222), a_3 = (2123333), \\ a_4 = (2321111), a_5 = (3132222), a_6 = (3231111)$$

对于图着色问题来说, 本质就是图的顶点集的划分, 而顶点所着的颜色本身没有任何意义, 上述的 6 个

3-着色方案均是将顶点集划分为 3 个划分块 $\{1, 3\}$, $\{2\}$, $\{4, 5, 6, 7\}$, 只是对各划分块所着的颜色不同而已(同一划分块中的顶点着相同的颜色). 因此, 上述的 6 个 3-着色方案对应于图着色问题的同一个有效解. 一般的, 对图着色问题的任意一个 k -划分, 整数表示法的编码空间中都有 $k!$ 个着色方案与之相对应, 即整数编码的冗余性. 为了避免整数编码的这种冗余性, 提出了图着色问题的有序划分编码, 编码后的个体与问题的有效解是一一对应的关系.

为了编码的规范性, 我们约定按照划分块的基数(顶点数量)的降序排列, 对基数相同的划分块, 最小顶点较小的划分块排列在前. 如个体 $\{2, 5, 6\}$, $\{1, 3\}$, $\{4, 8\}$, $\{7\}$ 就是有序划分编码后的个体. 划分块 $\{2, 5, 6\}$ 基数为 3, 排列在前, 划分块 $\{1, 3\}$ 和 $\{4, 8\}$ 基数相同, 划分块 $\{1, 3\}$ 的最小顶点为 1, 排列在划分块 $\{4, 8\}$ 之前, 划分块 $\{7\}$ 基数为 1, 是最小的划分块, 排在最后.

显然, 有序划分编码后的个体与图着色问题的有效解之间是一一对应的关系. 与整数编码的搜索空间相比, 有序划分编码的搜索空间缩小了 $k!$ 倍, 有利于减少算法的复杂度, 节省时间, 对大规模图着色问题求解, 具有重要的现实意义.

4 进化算法

与传统的优化算法相比, 进化算法的主要本质特征在于群体的搜索策略和简单的进化算子. 它随机的产生一组初始解作为初始种群, 在杂交算子和变异算子的作用下, 产生新的个体; 以适应度函数作为评价依据来衡量个体的好坏, 通过选择算子产生下一代种群. 这样一代代的不断进化, 最后收敛到最好的个体, 即问题的最优解或次优解.

4.1 初始种群的产生

一般的, 初始种群通常是随机产生的. 为了保证种群的多样性和种群中个体的质量, 采用随机贪婪算法产生初始解, 具体过程如下:

- (1) 令划分块 $V_i = \emptyset (i=1, 2, \dots, k)$;
- (2) 随机的产生 n 个顶点的一个排列 $p=(p_1, p_2, \dots, p_n)$, 依次安排顶点 p_j 至与其相邻的顶点最少的划分块中;

(3) 将划分 $\{V_1, V_2, \dots, V_k\}$ 有序化, 得到有序划分编码的个体 π .

重复上述步骤, 得到的 N 个个体 $\pi_1, \pi_2, \dots, \pi_N$ 作为初始种群(其中 N 为种群规模).

4.2 评价函数

为了执行适者生存的原则, 进化算法需要对个体的适应值进行评价, 据此来决定个体在下一代的生存能力. 对极小化问题来说, 个体的适应值越小, 该个体

就具有越强的生存能力. 针对图着色问题就是要将图的顶点集 V 分为 k 个独立集, k 动态减小的这一特征, 定义了图着色问题的动态适应值函数:

$$f(\pi) = w(\pi) + |k - l| \cdot m$$

其中, 个体 π 为顶点集的 l -划分, $w(\pi)$ 为个体 π 的违反量, m 为图的边数, $(k+1)$ 为已找到的正常着色所用颜色数的最小值. 由于算法致力于寻找图的 k -正常着色, 因此, 所有 k -划分的个体构成了当前的可行解集, 参数 $|k - l| \cdot m$ 保证了可行解永远优于不可行解, 即 k -划分个体的适应值总是小于非 k -划分个体的适应值, 在进化过程中, 具有更强的生存能力.

每当找到违反量为 0 的 k -划分, 令 $k = k - 1$, 该适应度函数均会随之变化, 算法就致力于寻找图的新的 k -正常着色. 对可行的个体 (k -划分), 适应值函数等于违反量, 因此, 对可行的个体, 违反量较小的个体, 适应值更小; 对种群中的不可行 (非 k -划分) 的个体, 其适应值永远大于 k -划分个体的适应值, 在进化过程中, 逐渐被淘汰, 以达到寻找图图的 k -正常着色的目的. 由此可见, 该动态适应度函数可以贴切的描述图着色问题, 是一个非常有效的适应度函数.

4.3 划分杂交算子

有效的杂交算子应根据问题的特征保留父辈个体中好的基因, 同时, 可以引入新的信息, 以维持种群的多样性, 避免陷入随机搜索. 图着色问题的本质是顶点集的划分, 提出了划分杂交算子, 既保留父母个体的部分划分块信息, 同时, 又改变某些顶点隶属的划分块.

对参加杂交的两个父母个体 $\pi_1 = (V_1^1, V_2^1, \dots, V_k^1)$ 和 $\pi_2 = (V_1^2, V_2^2, \dots, V_k^2)$, 划分杂交算子具体的过程如下:

(1) 令 $V_i = \emptyset$ ($i = 1, 2, \dots, k$), $j = 1$;

(2) 随机产生 $(0, 1)$ 之间的随机数 r , 若 $r < 0.5$, 令 $V_j = V_j^1$; 否则, $V_j = V_j^2$;

(3) 将 V_j 中的顶点从父母个体中删除, 令 $j = j + 1$. 若 $j \leq k$, 转(2); 否则, 将 $V - \{V_1, V_2, \dots, V_k\}$ 的顶点随机插入划分块 V_i ($i = 1, 2, \dots, k$) 中.

显然, 该划分杂交算子可以有效的保留父辈个体中的部分划分块, 而基于有效划分的编码方式, 使得较大的划分块优先得到保留, 同时, $V - \{V_1, V_2, \dots, V_k\}$ 中顶点的随机插入, 又可以引入新的个体, 有利于保持种群的多样性.

4.4 贪婪局部搜索

为了加快算法的收敛速度, 结合了一个新的贪婪局部搜索算子来改进杂交产生的后代个体. 图着色问题就是寻找违反量为 0 的划分. 因此, 贪婪局部搜索的思想是: 随机选取违反边的一个顶点 v , 将其安排到与

v 相邻的顶点最少的划分块中. 设 $\pi = \{V_1, V_2, \dots, V_k\}$ 是杂交产生的后代个体, 贪婪局部搜索的伪码过程如下:

$Ls = 1$

While not met the stop criteria ($Ls < L$)

Select a node $v \in V_i$ of conflict edge randomly

Choose the partition block V_j ($j \neq i$) with the least number of nodes adjacent to v

Remove v to the partition block V_j , and

obtain a new individual π'

If $f(\pi') \leq f(\pi)$, Let $\pi = \pi'$.

End

该局部搜索算子将违反边的端点 v 移动到与其相邻的顶点数最小的划分块中, 贪婪的减少个体的违反量. 其中 L 为局部搜索的步长, 与求解的问题密切相关. 对于大规模的图着色问题, L 的取值较大.

4.5 变异算子

变异算子一般是对个体的一个小小的扰动, 旨在增加种群的多样性, 帮助算法跳出局部最优解. 对参加变异的个体, 随机的改变一个顶点隶属的划分块, 得到的新的个体作为变异算子产生的后代个体. 该变异算子简单, 易执行.

4.6 基于有序划分编码的混合进化算法 (PEA, Evolutionary Algorithm based on ordered partition encoding)

(1) (初始化): 确定参数. 种群规模 N , 杂交概率、变异概率分别为 p_c 、 p_m , 随机贪婪算法产生 N 个个体作为初始种群, 记为 $P(0)$, 令进化代数 $t = 0$;

(2) (杂交): 以杂交概率 p_c 在 $P(t)$ 中选择参加杂交的个体, 并两两配对, 用杂交算子产生后代. 杂交产生的后代集合记为 O_1 ;

(3) (局部搜索): 对杂交产生的每个后代个体进行局部搜索, 得到新的后代的集合记为 O_2 ;

(4) (变异): 以变异概率 p_m 在 O_2 中选择参加变异的个体, 用变异算子产生变异后代. 变异后代的集合记为 O_3 ;

(5) (选择): 在 $P(t) \cup O_1 \cup O_2 \cup O_3$ 中选择最好的 N 个个体作为下一代种群 $P(t+1)$, 令 $t = t + 1$;

(6) (终止): 若终止条件成立, 则停; 否则, 令 $t = t + 1$, 转(2).

5 数值结果

为了验证算法的有效性, 从图着色国际标准数据库 (<http://mat.gsia.cmu.edu/COLOR//instances.html>) 中选取了 5 个经典大规模图着色问题进行仿真实验.

5.1 标准算例

(1) DSJC250.5 (.5 表示边的稠密度), 目前求得的

最好解为 28(28-正常着色);

(2)DSJC500.5,目前求得的最好解为 48;

(3)DSJC1000.5,目前求得的最好解为 83;

(4)Le450_15c,图的色数为 15(最优解);

(5)Le450_25c,图的色数为 25(最优解).

上述 5 个算例是验证图着色算法中通用的问题,便于与其他算法进行比较.

5.2 参数对照

为了对比的公平性,各算法的参数如下:

(1)种群规模:PEA 算法,种群规模 $N = 10$;

GPB 算法,种群规模 $N = 100 (n \leq 500)$,

$N = 500 (n = 1000)$;

GSA1 算法和 GSA2 算法,种群规模 $N = 300$;

(2)终止条件:采用最大进化代数为 T_{\max} .

PEA 算法, $T_{\max} = 50$;

GSA1 算法和 GSA2 算法,最大进化代数 $T_{\max} = 500$;

GPB 算法的终止条件未知.

5.3 数值结果

对 5 个标准算例,各运行 10 次,记录了 PEA 算法求得的正常着色所用颜色的最小值和运行时间,并将结果与文献[5]中的 GPB 算法,文献[6]中的 Mutation GSA(简记为 GSA1)算法,Order Based GSA(简记为 GSA2)算法,Iterated Greedy(简记为 IG)算法和 DSat 算法结果进行比较.表 1 记录了各算法求得的正常着色所用颜色的最小值.

表 1 各算法对 5 个算例的结果对照

算例	DSat	GSA1	IG	GSA2	GPB	PEA
DSJC250.5	37	31	29	29	28	28
DSJC500.5	65	55	52	50	48	48
DSJC1000.5	115	100	95	86	83	83
Le450_15c	23	25	23	15	15	15
Le450_25c	29	30	29	29	26	26

由表 1 可以看出,除 le450_25c 外,PEA 算法和 GPB 算法均能找到 5 个算例已知的最好解(对问题 le450_15c 找到了最优解).从总体来看,PEA 算法和 GPB 算法的求解性能最好,其次是 GSA2 算法、IG 算法、GSA1 算法,求解性能最差的是 DSat 算法.值得注意的是,除问题 le450_25c 外,PEA 算法和 GPB 算法均能找到问题已知的最好解或最优解;对问题 le450_25c,上述的两种算法求得的最好结果与最优解也只是相差 1,是求解性能比较稳定的算法.

表 2 对算法 GSA2,GPB 和 PEA 的运行时间进行了对照(其余算法时间未知).由表 2 可以看出,PEA 算法对各问题的运行时间最短,其次是 GSA2 算法,GPB 算法所用的时间最长.与 GSA2 算法相比,PEA 算法的求解时间至少缩短了 10 倍;除 le450_15c 外,PEA 算法所

用的时间比 GPB 算法的所需时间至少缩短了 158 倍.由此可见,PEA 算法具有快速收敛的优点.

表 2 各算法的运行时间表

算例	n	m	GSA2(se)	GPB	PEA(se)
DSJC250.5	250	15668	1618	9min	8
DSJC500.5	500	62624	6473	11hr	239
DSJC1000.5	1000	154940	25777	49.9hr	684
Le450_15c	450	16680	4007	55se	265
Le450_25c	450	17343	4563	19hr	431

注: n 为图的顶点数, m 为边数

从表 1 和表 2 可以看出,PEA 算法和 GPB 算法的求解性能优于其余 4 种算法,但是,在求解性能相当的情况下,GPB 算法所用的时间远远超过 PEA 算法所用的时间.图 1 为 PEA 算法种群中最好个体的违反量随进化代数的

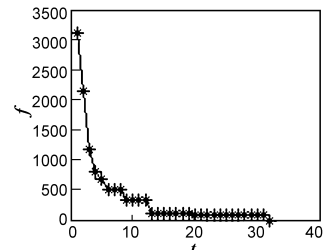


图 1 PEA 算法求解 DSJC1000.5 适应度变化曲线

变化曲线图,说明了 PEA 算法快速收敛的特点.

5.4 PEA 算法在新旧框架下的结果对照

进化算法求解图着色问题的旧框架:以往的遗传算法在求解图着色问题时,对给定的 k_0 种颜色(令 $k = k_0$),产生初始种群,在遗传算子作用下,求解正常 k -点着色方案;若找到正常 k -着色方案,令 $k = k - 1$,重新产生初始种群,求解正常 k -着色,直至找到问题的正常 $\chi(G)$ -着色方案为止.显然,多次产生初始种群,损失了当前种群中已有的有效信息,浪费了计算时间. PEA 算法得到问题的正常 k -着色方案后,划分杂交算子可以保留种群中的有效信息,快速找到正常 $(k - 1)$ -着色方案,大大节省了求解的时间,与动态的适应度函数协同作用,避免了重复产生初始种群,不失为进化算法求解图着色问题提供了一种新的框架.图 2 给出了 PEA 算法在求解图着色问题标准算例 Myciel16 在旧框架下(记为 PEA-old)运行过程中适应度函数的变化曲线.

从图 2(a)中可以看出,相较于 PEA-old 中(图 2(b)),PEA 对 MYCIE16 的搜索曲线平缓了许多,而 PEA

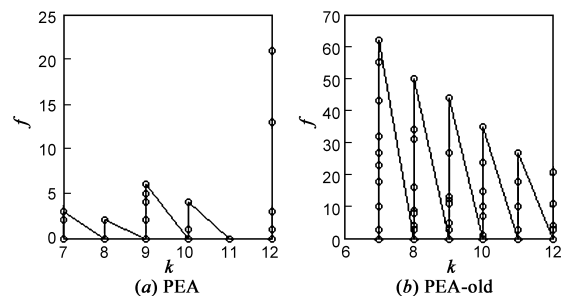


图 2 PEA 和 PEA-old 求解 Myciel16 的过程

重复产生初始种群,导致了违反约束的边的数目达到了 60 之多.由此可见,每得到颜色数更小的着色方案,重新产生初始种群,造成了巨大的时间浪费,损失了有效的遗传信息,新算法的提出有效的提高了进化算法求解图着色问题的潜力.

6 结论

针对图着色问题对顶点集划分的本质特征,设计了符合图着色问题的编码方式和进化算子,提出了求解图着色问题的一种有效的进化算法.无论在算法求解性能或是时间上,与以往的算法相比,都有所改进.

参考文献:

- [1] R M Karp. Reducibility among Combinatorial Problems[M]. Raymond E Miller and James W Thather; eds. Complexity of Computer Computations, Plenum Press, 1972. 85 – 103.
- [2] 高琳,许进.图的顶点着色问题的 DNA 算法[J].电子学报,2003,31(4):494 – 497.
Gao Lin, Xu Jin. A DNA algorithm for graph vertex coloring problem[J]. Acta Electronica Sinica, 2003, 31(4): 494 – 497. (in Chinese)
- [3] P Galinier, J K Hao. Hybrid evolutionary algorithms for graph coloring[J]. Journal of Combinatorial Optimization, 1999, 3: 379 – 397.
- [4] A G Celia, P B Adam. Genetic algorithm for graph coloring: exploration of Galinier and Hao's algorithm[J]. Journal of Combinatorial Optimization, 2003, 7: 229 – 236.
- [5] L Christine. Mumford. New order-based crossover for the graph coloring problem[A]. In Proc. PPSN' 06[C]. Springer-Verlag Berlin Heidelberg, 2006, LNCS 4193: 880 – 889.
- [6] Anna M, Adam P B, Celia A G. Improve graph colouring with linear programming and genetic algorithms[A]. Proceeding of the 2000 Genetic and Evolutionary Computation Conference [C]. Las Vegas, Nevada, USA, 2000. 240 – 245.

- [7] Celia A G, Adam P B. Genetic algorithm for graph coloring: exploration of Galinier and Hao's algorithm[J]. Journal of Combinatorial Optimization, 2003, 7: 229 – 236.
- [8] Zbigniew K, Marcin K, Krzysztof K. Parallel genetic algorithm for graph coloring problem[A]. M. Bubak et al. (Eds.): ICCS [C], Springer-Verlag Berlin Heidelberg, 2004, LNCS 3036: 215 – 222.
- [9] Yu J Q, Yu P N. A novel parallel genetic algorithm for the graph coloring problem in VLSI channel routing[A]. Proceeding of the Third International Conference on National Computation (ICCN)[C]. 2007. 101 – 105.
- [10] Christine L Mumford. New Order-based crossover for the graph coloring problem[A]. Springer-Verlag Berlin Heidelberg, T. P. Runarsson et al. (Eds): The ninth international conference on parallel problem solving from nature PPSN IX [C], Reykjavik, Iceland, September 9 – 13, 2006, LNCS 4193: 880 – 889.

作者简介:



韩丽霞 女,1980年2月出生于威海荣成,2009年在西安电子科技大学获理学博士学位,现为中国矿业大学计算机学院讲师,从事组合优化、智能计算方面的研究.

E-mail: lxhan2006@yahoo.com.cn

王宇平 男,1961年7月出生于陕西西安.1993年6月在西安交通大学获博士学位,现为西安电子科技大学计算机学院教授,博士生导师,全国经济数学与管理数学学会副理事长.主要从事进化计算和优化新方法的研究.在《IEEE Transactions on Evolutionary Computation》,《IEEE Transactions on System, Man and Cybernetics》,《Engineering Optimization》,《计算机学报》,《电子学报》等国内外刊物发表论文 100 多篇.