

改进的量子遗传算法及其在测试数据生成中的应用

周 绮*, 姜淑娟, 赵雪峰

(中国矿业大学 计算机科学与技术学院, 江苏 徐州 221116)

(* 通信作者电子邮箱 computer_zq@163.com)

摘 要:针对测试数据自动生成中收敛速度不够快的缺点,提出一种改进的量子遗传算法(IQGA),其对量子遗传算法的主要改进是:1)在个体更新时,对个体的某一位取反,将取反后的个体用于指导下一代个体的进化;2)对测量后的二进制个体进行变异,而不是传统的互换量子比特的概率幅。将 IQGA 用于测试数据生成,通过对三个基础程序进行实验,结果表明 IQGA 在覆盖率和迭代次数两个方面都优于传统量子遗传算法。IQGA 不仅能保证种群朝着正确的方向进化,同时有效地避免了早熟现象,能以更快的速度搜索到目标解。

关键词:量子遗传算法;测试数据生成;取反指导更新;二进制变异;快速收敛

中图分类号: TP311.52 **文献标志码:** A

Improved quantum genetic algorithm and its application in test data generation

ZHOU Qi*, JIANG Shu-juan, ZHAO Xue-feng

(School of Computer Science and Technology, China University of Mining and Technology, Xuzhou Jiangsu 221116, China)

Abstract: This paper proposed an Improved Quantum Genetic Algorithm (IQGA) for the problem of slow convergence in test data generation. There are two main improvements. First, every bit of every individual was reversed to conduct the evolution; second, the binary individuals were mutated after measurement, instead of the traditional exchange of the probability amplitude of quantum bits. IQGA was applied into test data generation. The experiments on three basic programs prove that IQGA is better than QGA in terms of coverage rate and the number of iterations. IQGA can not only ensure the right direction of the evolution of populations, but also avoid premature phenomenon, and it can get the solution at a faster convergence speed.

Key words: Quantum Genetic Algorithm (QGA); test data generation; updating conducted by reversing; binary mutation; rapid convergence

0 引言

软件测试是软件质量的重要保障。在软件测试中,测试用例的设计与生成是一个非常重要的环节。人工生成测试数据是一项繁琐复杂的任务,测试数据自动生成技术可以极大地提高测试效率,因此一直都是研究热点。随着一系列算法的引入,如:禁忌搜索、模拟退火、遗传算法、粒子群算法等^[1-2],测试数据生成技术得到了很大的发展。

量子遗传算法(Quantum Genetic Algorithm, QGA)将量子计算与遗传算法相结合,采用量子比特的概率幅对染色体进行编码,单个染色体可以表示多个染色体的叠加,使得算法具有很好的种群多样性。同时,坍缩操作和量子门更新操作使得算法兼具强大的全局搜索能力和局部搜索能力。基于此,大量学者对其进行了深入的研究^[3-11]。

文献[4]将带有变异的量子遗传算法用于解决组合优化问题,其实验结果表明 QGA 优于遗传算法和遗传量子算法 Genetic Quantum Algorithm, GQA)。文献[6]提出了一种解决组合优化问题的改进型量子遗传算法(Novel Improved Quantum Genetic Algorithm, NIQGA),算法引入了动态调整量子门旋转角步长机制、量子交叉操作和量子变异操作;文献

[7]采用量子比特相位比较法更新量子门和自适应调整搜索网格的策略对量子遗传算法进行改进,并将其用于无线传感器网络感知节点的分布优化中,得到了良好的性能;文献[8]提出了一种基于量子遗传算法混合核聚算法的 XML 文档聚类新方法;文献[9]将量子遗传算法应用到测试数据生成当中,其实验结果表明量子遗传算法的性能优于传统遗传算法。

上述工作表明了量子遗传算法的广泛应用,同时也证实了量子遗传算法的优良性能。但是在测试数据生成中,对于量子遗传算法的研究还不够深入。基于此,本文提出一种改进的量子遗传算法 IQGA,算法不仅能保证向着目标解的方向搜索,而且种群多样性好,收敛速度快。将改进后的算法 IQGA 应用于测试数据生成,良好的收敛速度和全局搜索能力保证了算法总能搜索到目标解,实验证明算法综合性能优于传统量子遗传算法。

1 量子遗传算法简介

QGA 是建立在量子的态矢量表达基础上,将量子比特的概率幅表示应用于染色体的编码,并利用量子旋转门来实现染色体的更新,从而求得目标解。实践证明 QGA 具有种群多样性好、种群规模小、寻优能力强的特点。

收稿日期:2011-07-29;修回日期:2011-09-23。

基金项目:国家自然科学基金资助项目(60970032);江苏省自然科学基金资助项目(BK2008124);江苏省“青蓝工程”资助项目。

作者简介:周绮(1987-),女,湖南望城人,硕士研究生,主要研究方向:软件测试;姜淑娟(1966-),女,山东莱阳人,教授,博士生导师,博士,主要研究方向:软件分析与测试、程序设计语言、编译技术;赵雪峰(1986-),男,江西九江人,硕士研究生,主要研究方向:软件工程、软件测试。

QGA 算法的基本流程如下:

```

步骤1 初始化种群  $Q(t)$ ;
步骤2 While 非结束状态 do
Begin
    1) 对种群  $Q(t)$  实施一次测量, 得到一组二进制串个体  $P(t)$ ;
    2) 对各二进制串个体进行适应度评估;
    3) 记录最佳个体及其适应度值;
    4) 利用量子门对种群进行更新, 得到子代种群  $Q(t+1)$ ;
    5)  $t = t + 1$ 
End

```

2 改进的量子遗传算法

对 QGA 的改进主要体现在: 添加新的变异操作和个体更新。

2.1 变异操作

传统的量子变异是通过操作量子染色体, 即互换量子比特概率幅 (α, β) 的值来实现, 通过这种方式, 使得测量时得到状态“0”和状态“1”的概率发生改变。由于变异的随机性, 改变的结果可能使个体朝着错误的方向进化, 最终可能导致收敛速度慢, 甚至找不到目标解。

IQGA 中, 先对量子染色体进行测量, 获得个体的二进制串表示, 再对该二进制串个体进行变异, 即在给定的变异概率下, 改变二进制串个体某一位或者多位的值, 若变异位的值为 0, 变异后则为 1; 反之, 变异后则为 0, 此变异与传统遗传算法的变异类似。

2.2 个体更新

量子遗传算法中, 个体的更新是通过旋转门来实现的。通常需要存储最优个体, 比较最优个体与当前个体某一位的值及其适应度值的大小, 然后通过查表来实现个体某一位的更新。依次对每一位进行更新后, 即可实现整个个体的更新。

IQGA 中, 个体的更新同样通过旋转门来实现, 但是与当前个体进行比较的并不是最优个体, 而是对当前个体的某一位进行取反之后的个体 (记作: CompareIndividual), 计算 CompareIndividual 的适应度值后, 再将其与当前个体比较, 通过查询表 1 进而实现该位的更新。 x_i 和 c_i 分别表示当前个体与 CompareIndividual 的第 i 位, $f(x)$ 和 $f(c)$ 表示当前个体与 CompareIndividual 的适应度值, $\Delta\theta_i$ 表示旋转角度的大小, α_i 和 β_i 分别表示当前个体第 i 位的概率幅, $S(\alpha_i, \beta_i)$ 表示旋转角度的方向。

表 1 旋转角度

x_i	c_i	$f(x) \geq f(c)$	$\Delta\theta_i$	$S(\alpha_i, \beta_i)$			
				$\alpha_i\beta_i > 0$	$\alpha_i\beta_i < 0$	$\alpha_i = 0$	$\beta_i = 0$
0	1	假	0.007	+1	-1	0	± 1
0	1	真	0.007	-1	+1	± 1	0
1	0	假	0.007	-1	+1	± 1	0
1	0	真	0.007	+1	-1	0	± 1

2.3 算法

改进的量子遗传算法 (IQGA) 步骤如下所示:

```

步骤1 初始化种群;
步骤2 While 非结束状态 do
Begin
    1) 对种群  $Q(t)$  实施一次测量, 得到一组二进制串个体  $P(t)$ ;
    2) 对  $P(t)$  进行变异;
    3) 对各二进制串个体进行适应度评估;
    4) 对个体进行更新, 具体如下:

```

For 各个体的每一位

Begin

- ① 计算出 CompareIndividual;
- ② 对 CompareIndividual 进行解码, 求适应度值等操作;
- ③ 根据 CompareIndividual 及当前个体, 查询表 1 对个体进行更新操作。

End

- 5) $t = t + 1$;

End

2.4 算法分析

在新变异中, 与传统变异一样, 变异过程的随机性导致变异结果同样具有一定随机性。因此, 新变异也能维持种群多样性, 并能在一定程度上避免早熟。随机性的另一个结果是变异后个体的适应度值可能增大, 也可能减小。传统算法中, 无论变异结果如何, 其都被保存到下一代个体中, 若变异后个体适应度值增大, 则有利于进化; 但变异后个体适应度值也可能减小, 则该变异被保存到下一代个体中, 可能会导致收敛速度减慢甚至找不到目标解。IQGA 通过个体更新有效地避免了该问题。因为量子染色体是以量子比特概率幅的形式存在, 不是二进制形式, 故无论变异结果如何, 变异改变的只是二进制形式个体, 变异并未保存到量子染色体中。但在个体更新过程中, 变异能够起到指导作用。

个体更新时, IQGA 首先计算出 CompareIndividual, 然后对 CompareIndividual 进行解码、求适应度值等操作, 根据 CompareIndividual 及变异后个体, 查询表 1 对个体进行更新操作。假定某位从 0 变异到 1, 更新时, 经过适应度值的计算与比较, 若对比的结果是该位为 1 时的适应度值比该位为 0 时的适应度值大, 则说明变异是朝着正确的方向, 通过旋转门的操作, 增大该位坍缩为 1 的概率; 若对比的结果是该位为 0 时的适应度值比该位为 1 时的适应度值大, 则说明变异是朝着错误的方向, 那么量子染色体会朝着相反的方向进化, 即增大该位坍缩为 0 的概率。如此, 算法保证了量子染色体总是朝着正确的方向进化。由于 CompareIndividual 是对变异后个体的某位取反之后得到的, 故 CompareIndividual 与变异后个体中正在更新的位不可能出现同为 0 或者同为 1 的情况, 所以表 1 中没有 x_i 和 c_i 同为 1 或者同为 0 的情况。

以下举例说明新变异及个体更新。

例如, 现有一个长度为 6 的量子染色体, 假定目标解的第一位为 0。进化到某代时, 其染色体如下:

$$\begin{bmatrix} 0.8925 & 0.2459 & 0.1005 & 0.7761 & 0.8587 & 0.9923 \\ 0.4509 & 0.9693 & 0.9949 & 0.6307 & 0.5125 & 0.1238 \end{bmatrix} \quad (1)$$

若用传统的方法对其进行变异, 即在给定的变异概率下, 对调某位量子比特概率幅 (α, β) 的值。变异的结果可能是:

$$\begin{bmatrix} 0.4509 & 0.2459 & 0.1005 & 0.7761 & 0.8587 & 0.9923 \\ 0.8925 & 0.9693 & 0.9949 & 0.6307 & 0.5125 & 0.1238 \end{bmatrix} \quad (2)$$

该变异导致量子染色体第一位坍缩为 0 的概率变小, 而目标解第一位为 0, 故该变异对个体来说是不利的; 并且此结果会被保存到下一代个体中, 最终可能导致算法找不到目标解。

而在 IQGA 中, 首先对该量子染色体进行测量, 得到一个二进制串的测量结果, 假定测量得到的结果是: 000111。接下来对该二进制串进行变异, 即对以上二进制串, 在给定的变异概率下, 改变某位或者多位的值, 若某位为 0, 变异后则为 1, 反之则为 0。假定变异结果为: 100111。第一位由 0 变异为

1,而目标解第一位为0,说明变异是朝着错误的方向。随后对其进行解码、评估等操作,并开始更新。首先更新第一位:计算此时的 CompareIndividual,即 000111,然后对其进行解码,计算适应度值等操作,此时, $x_i = 0, c_i = 1$,由目标解第一位为0, $f(x) > f(c)$,根据表1应增大该位取0的概率,即 $|\alpha|^2$ 变大,这说明即使朝着错误方向变异时,量子染色体仍朝着正确的方向进化。

综上,该算法有以下优势:

- 1) 变异操作维持了种群的多样性,有效避免了早熟现象;
- 2) 新的变异方法虽然也有随机性,可能导致变异后的二进制串的适应度值减小,但随后的个体更新保证了无论变异结果如何,量子染色体总是朝着正确的方向进化;
- 3) 种群多样性以及进化方向的正确提高了算法性能,加快了收敛速度。

3 实验结果及分析

为了验证算法的有效性,将改进的量子遗传算法用于测试数据生成。本文在三个程序上进行实验,测试的准则是分支覆盖,实验结果是运行20次的平均值。

本文的实验程序是文献[2]中的3个程序:三角形分类程序(Triangle Classifier Program, TCP)、判断直线与矩形关系(Line Rectangle Classifier, LRC)程序和计算两日期相隔天数程序(Number of Days Between Two Dates, Date Problem),具体程序控制流图参考文献[2]。

3.1 三角形分类程序

三角形分类程序是一个非常典型的基准程序,很多研究都用到了这个程序。但是,不同的研究,程序的具体实现也不尽相同^[2],故可比性不大。

本文中的三角形分类程序与文献[2]的相同,该程序接收三个参数:A,B,C,程序首先判断输入数据能否组成一个三角形,若能,则继续判断是一般三角形、等腰三角形或者等边三角形。

该程序共有12个分支,最深嵌套5层,且节点5,6,9,11都包含相等测试,增大了搜索难度。其中分支6的嵌套层次为4,且包含两个相等测试,即要求三个输入数据都相等,故最难覆盖。

实验结果如图1所示,可以看出,随机数据生成算法和QGA很难达到100%的覆盖率,而IQGA平均只要迭代500多次即能达到100%的覆盖率。

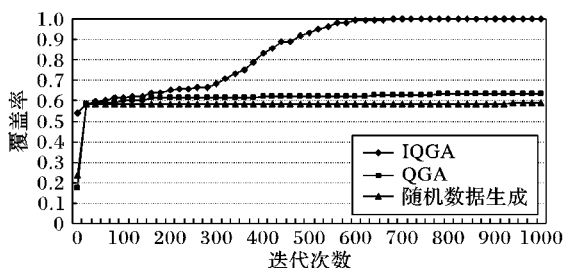


图1 覆盖率随迭代次数变化图(TCP)

为了验证算法是否朝着目标解的方向搜索,本文对迭代过程中的量子比特概率幅进行了记录。此实验中,最难覆盖的节点是判断等边三角形的节点,而等边三角形的条件是三条边相等。实验中用一个长度为16的量子染色体来表示每条边,要得到等边三角形,需满足三条边各对应位进化时同趋

于0或者同趋于1。

实验中记录的某个体三条边的迭代情况如图2所示。横坐标表示迭代次数,纵坐标表示各位坍塌为0的概率。从图2可以看出,各条边的第一位坍塌为0的概率逐渐迭代为1,而第16位坍塌为0的概率逐渐迭代为0。其余各对应位的迭代趋势也大致类似,迭代的最终结果是产生三个相等的数值,即覆盖等边三角形的分支,由此可以看出算法是朝着目标解的方向搜索的。

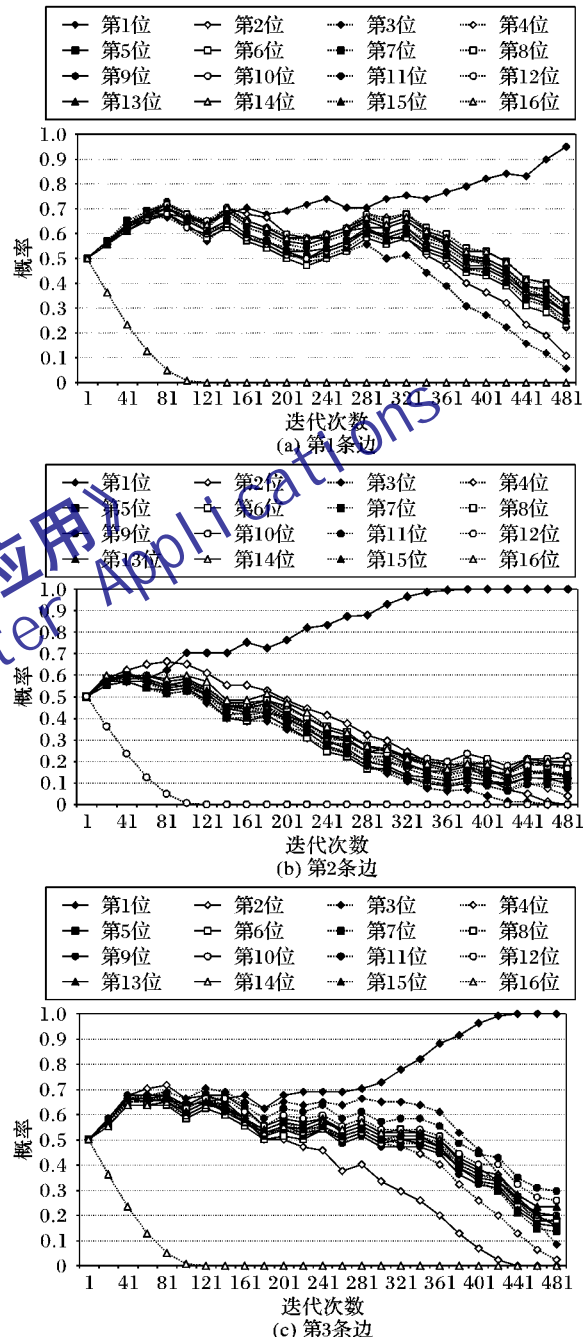


图2 各边的迭代情况

3.2 判断直线与矩形关系程序

该程序的作用是判断直线与矩形的相对关系,对文献[2]中的程序进行了修改,修改部分在文献[2]的控制流图中体现为:

- 1) 节点33和节点35左分支的判断条件由 $(x > xr1 \ \&\& \ x < xr2)$ 修改为 $(x > xr1 \ \&\& \ x < xr2 \ \&\& \ x > xl1 \ \&\& \ x < xl2)$,右分支的判断条件则相应地修改为 $!(x > xr1$

$\&\& x < = xr2 \&\& x > = xl1 \&\& x < = xl2$);

2) 节点 37 和节点 39 左分支的判断条件由 $(y > = yr1 \&\& y < = yr2)$ 修改为 $((y > = yr1 \&\& y < = yr2) \&\& ((yl1 > yl2 \&\& y > = yl2 \&\& y < = yl1) \parallel (yl1 < yl2 \&\& y > = yl1 \&\& y < = yl2)))$, 右分支的判断条件相应地修改为 $!((y > = yr1 \&\& y < = yr2) \&\& ((yl1 > yl2 \&\& y > = yl2 \&\& y < = yl1) \parallel (yl1 < yl2 \&\& y > = yl1 \&\& y < = yl2)))$ 。

程序的输入包括 8 个变量: $(xr1, xr2, yr1, yr2)$, 这 4 个变量确定矩形的位置; $(xl1, xl2, yl1, yl2)$, 这 4 个变量确定直线的位置。输出为直线与矩形的关系, 一共包括四种情况: 1) 直线被矩形包围; 2) 直线完全在矩形外; 3) 直线部分在矩形中, 部分在矩形外; 4) 输入的变量不能构成直线或矩形。

程序共有 36 个分支, 嵌套层次非常深, 最深为 12, 而嵌套层次中也包含相等测试, 故该程序比 3.1 节中三角形分类程序复杂。

实验结果如图 3 所示, 可以看出, 随机数据生成算法和 QGA 很难达到 100% 的覆盖率, 而 IQGA 只需要平均 100 次即可完成 100% 的覆盖率。

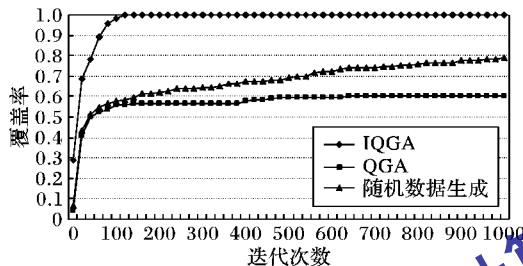


图3 覆盖率随迭代次数变化图(LRG)

3.3 计算两日期相隔天数程序

该程序作用是计算两个日期相隔的天数。程序的输入包括 6 个参数: $(day0, mon0, year0), (day1, mon1, year1)$, 两组参数分别确定一个日期。程序首先判断日期是否有效; 之后, 对两个日期进行排序, 确保第二个日期比第一个大。在程序中, 需要判断年份是否为同年来进行不同的加减操作。这些操作使得程序很复杂, 嵌套层次更深。该实验需要大量的数据才能达到 100% 的覆盖率, 但 IQGA 平均只要 200 多次迭代即可完成 100% 覆盖, 说明 IQGA 不仅种群多样性好, 而且收敛速度快。

实验结果如图 4 所示, 可以看出, 从迭代次数和覆盖率两方面考虑, IQGA 都明显优于另外两种算法。

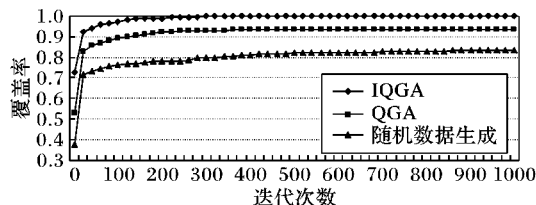


图4 覆盖率随迭代次数变化图(Date Problem)

由以上实验结果可知, 三个实验中, IQGA 都能以较少的迭代次数得到目标解, 表现出了良好的性能。

4 结语

本文提出了一种改进的量子遗传算法 IQGA, 并将其用于测试数据生成。其特点主要体现在个体更新和变异操作上。变异操作不仅能维持种群的多样性, 同时也能在一定程度上

避免早熟现象。尽管变异的随机性可能导致不利的变异, 但是变异结果并未被保存下来, 而是在个体更新中指导进化方向, 使得在出现不利变异时量子染色体仍能朝着正确的方向进化。文中的三个实验证实了算法的良好性能。

参考文献:

- [1] 伦立军, 丁雪梅, 李英梅. 基于遗传算法的测试数据生成研究[J]. 计算机工程, 2005, 31(25): 82-84.
- [2] DÍAZ E, TUYA J, BLANCO R, *et al.* A tabu search algorithm for structural software testing[J]. Computers and Operations Research, 2008, 35(10): 3052-3072.
- [3] 邹谊, 庄镇泉, 李斌. 基于量子遗传算法的嵌入式系统软硬件划分算法[J]. 电路与系统学报, 2004, 9(5): 1-7.
- [4] 熊焰, 陈欢欢, 苗付友, 等. 一种解决组合优化问题的量子遗传算法 QGA[J]. 电子学报, 2004, 32(11): 1855-1858.
- [5] 王宇平, 李英华. 求解 TSP 的量子遗传算法[J]. 计算机学报, 2007, 30(5): 748-755.
- [6] 邢焕来, 潘炜, 邹喜华. 一种解决组合优化问题的改进型量子遗传算法[J]. 电子学报, 2007, 35(10): 1999-2002.
- [7] 付华, 韩爽. 基于新量子遗传算法的无线传感器网络感知节点的分布优化[J]. 传感技术学报, 2008, 21(7): 1259-1263.
- [8] 蒋勇, 谭怀亮, 李光文. 基于量子遗传算法的 XML 聚类方法[J]. 计算机应用, 2011, 31(2): 446-449.
- [9] 王梅萍, 刘新贵. 基于量子遗传算法的软件测试数据自动生成[J]. 计算机测量与控制, 2011, 18(1): 34-37.
- [10] 李盼池, 李士勇. 基于量子遗传算法的正规模糊神经网络控制器设计[J]. 系统仿真学报, 2007, 19(16): 3710-3714.
- [11] 杨俊安, 庄镇泉. 量子遗传算法现状[J]. 计算机科学, 2003, 30(11): 13-15.

第五届中国计算机网络与信息安全学术会议

(CCNIS'2012) 征文通知

第五届中国计算机网络与信息安全学术会议 (CCNIS2012) 将于 2012 年 8 月上旬在中国福建省武夷山举行, 承办单位为福建师范大学(软件学院)。CCNIS2012 现公开征集会议论文, 征文范围涵盖计算机网络与信息安全及相关领域。

征文范围(不限于这些领域)

- 信息安全算法
- 信息安全协议
- 信息系统安全
- 网络系统安全
- 云计算与物联网安全
- 安全专用芯片
- 数据与应用安全
- 隐私、风险与信任
- 新型网络与计算技术

重要日期

征文截止时间: 2012 年 5 月 8 日

录用通知时间: 2012 年 6 月 8 日

会议召开时间: 2012 年 8 月 3—5 日

投稿方式

请作者选择以下两种方式之一:

1) 发送稿件至 ccnis2012@vip.163.com, 主题请注明 CCNIS2012 投稿。

2) 发送稿件至在线投稿系统的在线投稿栏目:

<https://www.easychair.org/conferences/?conf=ccnis2012>

详情请见: <http://ccnis2012.fjnu.edu.cn>