

文章编号:1001-9081(2006)07-1682-04

基于 QPSO 算法的多阶段投资组合优化

须文波,江家宝,孙俊

(江南大学 信息工程学院,江苏 无锡 214122)

(jiabaojiang2004@sina.com)

摘要:研究了基于量子行为的微粒群优化(QPSO)算法在多阶段投资组合优化中制定投资决策的方法,目标函数是最大化个人经济效益或最大化周期结束时个人财富。通过比较用 QPSO 算法和遗传算法优化美国标准普尔指数 100 的不同股票和现金分配所得到的期望收益率均值与方差,证实了该方法的优越性。

关键词:粒子群优化;量子;多阶段投资组合;资产分配

中图分类号: TP18 **文献标识码:** A

Multi-stage portfolio optimization using quantum-behaved partical swarm optimization

XU Wen-bo, JIANG Jia-bao, SUN Jun

(School of Information Technology, Southern Yangtze University, Wuxi Jiangsu 214122, China)

Abstract: The method of decision-making in the field of multi-stage portfolio optimization using Quantum-behaved Partical Swarm Optimization(QPSO) was studied. Its objective function was to maximize one's economic utility or end-of-period wealth. How to use QPSO to find best portfolio according to objective function was introduced. By comparing the expect return and their variances that come from optimizing the allocation of cash and various stocks in the market of USA, QPSO algorithm with genetic algorithms was demonstrated superior to genetic algorithms.

Key words: partical-swarm-optimization; quantum; multi-stage-portfolio; asset allocation

0 引言

投资计划的制定包括资产的分配和风险管理。资产配置问题就是决定总投资价值分配到每一个投资资产的比例。风险管理就是根据历史数据计算出不同投资工具的投资风险并由此产生具有一定风险回报特征的投资组合。多阶段随机优化是一个定量的模型,它结合了资产配置策略和补救策略。该模型在不断变化的金融市场中通过周期性地再平衡资产组合使期望回报最大化和风险最小化方式管理投资组合。多阶段优化技术注重问题的动态性从而产生最佳投资组合。

资产分配的最优化问题是一个复杂的 NP 难解问题,它是非线性的,具有许多局部最优点。为了解决资产配置问题,可以使用线性归化解决方法,即在目标函数中非线性期间被分段地线性化,但时间复杂度大,且易于陷入局部最优点。我们提出一种新的能保证全局收敛的粒子群算法,即量子粒子群算法(Quantum-behaved Particle Swarm Optimization, QPSO)^[1,2],将它运用于解决多阶段投资组合优化问题。实验证明,QPSO 的性能优于遗传算法。

1 多阶段投资组合优化模型

由于风险随着时间的推移是不固定的,所以多阶段随机模型^[3]考虑了资产配置问题的动态性,在不断变化的金融市场中通过周期性地再平衡资产投资组合方式管理投资组合以

达到回报最大化和(或)风险最小化,从而产生最佳投资组合。模型结合了各种特定情况的随机性分析,每一特定情况 s 描述了整个多阶段计划周期从开始到结束的某一阶段特定情况,即图 1 中从开始到结束的一条路径的某一段。在本文中,某一特定情况是由市场指数的变化定义的,为了简洁我们设置两个特定情况为市场指数下降和市场指数上升。

假设我们必须优化 A 组资产,用 0 号资产表示现金,用别的符号代表其他任何投资工具,比如:债券、基金、期货和股票。把整个计划的时间范围分为 T 阶段;每个阶段制定投资决策,每个阶段可能有不同的特定情况。建立一个描绘成图的特定情节树,以便形象化地表示动态优化资产配置投资决策。

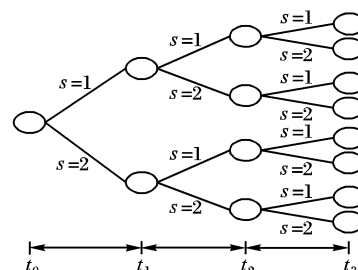


图 1 拥有 2 种特定情况和 3 个阶段的特定情节树

投资组合优化的数学模型描述如下:

1) 参数

$r_{i,t}^s = 1 + p_{i,t}^s$,这里的 $p_{i,t}^s$ 是第 t 阶段在特定情况 s 下资产

收稿日期:2006-01-15;修订日期:2006-04-09 基金项目:国家自然科学基金资助项目(60474030)

作者简介:须文波(1946-),男,江苏无锡人,教授,博士生导师,主要研究方向:人工智能、计算机控制技术、嵌入式操作系统;江家宝(1972-),男,安徽无为,人,硕士研究生,主要研究方向:金融计算、嵌入式操作系统;孙俊(1974-),男,江苏无锡人,讲师,博士研究生,主要研究方向:人工智能、生物信息。

i 的收益率; π_s^t 是第 t 阶段特定情况 s 发生的概率,且 $\sum_s \pi_s^t = 1$; W_0 是投资开始时的财富总量; W_t^s 是第 t 阶段结束时在 s 特定情况下所有财富总量; $X_{i,t}^s$ 是第 t 阶段开始时在特定情况 s 下重新平衡资产分配之前资产 i 的财富量; T 是阶段总数。

2) 决策变量

$Y_{i,t}^s$ 是第 t 阶段开始时在特定情况 s 下重新平衡资产分配之后资产 i 的财富量; $B_{i,t}^s$ 是第 t 阶段开始时在特定情况 s 下为了重新平衡资产而需要买入资产 i 的总量; $D_{i,t}^s$ 是第 t 阶段开始时在特定情况 s 下为了重新平衡资产而需要卖出资产 i 的总量。

各个变量必须满足以下关系式:

$$\sum_i X_{i,t}^s = \sum_i Y_{i,t}^s = W_0, s = 1, 2, \dots, S \quad (1)$$

$$\sum_i Y_{i,t}^s = W_{t-1}^s, s = 1, 2, \dots, S, t = 1, 2, \dots, T, i \in A \quad (2)$$

$$X_{i,t}^s = r_{i,t-1}^s \times Y_{i,t-1}^s, s = 1, 2, \dots, S, t = 1, 2, \dots, T, i \in A \quad (3)$$

$$Y_{1,t}^s = X_{1,t}^s + \sum_{i \neq 1} D_{i,t}^s - \sum_{i \neq 1} B_{i,t}^s, s = 1, 2, \dots, S, t = 1, 2, \dots, T, i \in A \quad (4)$$

$$Y_{i,t}^s = X_{i,t}^s + B_{i,t}^s - D_{i,t}^s, s = 1, 2, \dots, S, t = 1, 2, \dots, T, i \neq 1 \text{ 且 } i \in A \quad (5)$$

$$X_{i,t}^s = X_{i,t-1}^s, s_1 \neq s_2, t = 1, 2, \dots, T, i \in A \quad (6)$$

关系式(1)和(2)描述了初始化时的总财富量和每个时间阶段 t 开始时在特定情节 s 下的总财富量。关系式(3)描述了在特定情况 s 下资产 i 平衡前的财富量和前一阶段积聚的财富量。关系式(4)和(5)描述了现金和其他种类资产在特定情况 s 下平衡后的财富量。关系式(6)描述了任何阶段平衡前每一个资产在所有情况下的价值相等。

3) 数学模型

$$\text{Max } Z = \sum_s \pi_s \text{Measure}_T^s \quad (7)$$

式中 Measure_T^s 是在特定情节 s 下 T 阶段所用方法的性能度量函数,我们采用效益函数作为性能度量函数,其好处是能够在一定风险忍受程度较宽的范围内制定投资决策,其表达式如下:

$$\text{Measure} = \begin{cases} \ln(w), & \text{if } \gamma = 0 \\ \frac{(w)^\gamma}{\gamma}, & \text{otherwise} \end{cases} \quad (8)$$

式中 γ 是投资者根据他对投资风险的容忍程度设定的参数: γ 值越小,表示投资者对投资风险越敏感, $\gamma = 1$ 时效益函数值是阶段结束时的财富量并且没有被考虑投资风险。因此目标函数描述如下:

$$Z = \begin{cases} \sum_s \pi_s \ln(w_T^s), & \text{if } \gamma = 0 \\ \sum_s \pi_s \frac{(w_T^s)^\gamma}{\gamma}, & \text{otherwise} \end{cases} \quad (9)$$

目标函数是以特定情节的发生概率为权重的性能度量函数。概率权重优化意味着优化更多地集中在发生概率比较高的那些特定情况上。每个特定情况的发生概率可以由历史统计数据或任何预测系统产生。本文采用历史统计数据计算每个特定情况的发生概率。

2 算法简介

2.1 粒子群优化算法

粒子群优化算法(Particle Swarm Optimization, PSO)是在

1995年由美国社会心理学家 James Kennedy 和电气工程师 Russell Eberhart 共同提出的,基本思想是模拟鸟类群体行为,并利用了生物群体模型,与其他进化类算法相类似,也是用“群体”与“进化”的概念,同样也是依据个体(微粒)的目标函数值大小进行操作。区别在于,微粒群算法将每个个体看作是在 N 维搜索空间中的一个无重量无体积的微粒,并在搜索空间中以一定的速度飞行,飞行速度由个体的飞行经验和群体的飞行经验进行动态调整,假设: $X_i = (X_{i1}, X_{i2}, \dots, X_{in})$ 为微粒 i 的当前位置; $V_i = (V_{i1}, V_{i2}, \dots, V_{in})$ 为微粒 i 的当前飞行速度; $P_i = (P_{i1}, P_{i2}, \dots, P_{in})$ 为微粒 i 所经历的当前最佳位置(即经历过的具有最好适应值的位置)。

设 $f(X)$ 为最优化的目标函数,则微粒 i 所经历的当前最佳位置由下式确定:

$$P_i(t+1) = \begin{cases} P_i(t), & f(X_i(t+1)) \leq f(P_i(t)) \\ X_i(t+1), & f(X_i(t+1)) > f(P_i(t)) \end{cases} \quad (10)$$

设群体中的微粒数为 N ,群体中所有微粒所经历过的最佳位置 $P_g(t)$,称为全局最佳位置。则:

$$P_g(t) \in \{P_0(t), P_1(t), \dots, P_N(t)\} \mid f(P_g(t)) = \max\{f(P_0(t)), f(P_1(t)), \dots, f(P_N(t))\} \quad (11)$$

算法的进化方程可描述如下:

$$V_{i,j}(t+1) = w \times V_{i,j}(t) + c_1 \times r_{1j}(t) \times (P_{i,j}(t) - X_{i,j}(t)) + c_2 \times r_{2j}(t) \times (P_{g,j}(t) - X_{i,j}(t)) \quad (12)$$

$$X_{i,j}(t+1) = X_{i,j}(t) + V_{i,j}(t+1) \quad (13)$$

其中,下标“ i ”表示微粒 i ,下标“ j ”表示微粒的第 j 维, t 表示第 t 代, c_1, c_2 为加速常数, $r_1 \sim U(0,1), r_2 \sim U(0,1)$ 为两个相互独立的随机数。速度的计算以下面三个基值为基础:第一部分 $V_{i,j}(t)$ 为微粒先前的速度,第二部分 $P_{i,j}(t)$ 为“认知”部分,因为它仅考虑了微粒自身的经验,是局部最优位置,第三部分 $P_{g,j}(t)$ 是“社会”部分,表示微粒间的社会信息共享,是全局最优位置。实验用到的参数 w 可以是固定值,一般是线性地从 0.9 递减到 0.4,具体取值视情况而定, c_1 和 c_2 为待定常数,一般取 $c_1 = 1.6, c_2 = 2.0$ 。

2.2 QPSO 算法

笔者提出的 QPSO 算法^[1,2]不仅参数个数少,并且在搜索能力上优于所有已开发的 PSO 算法。在 QPSO 中,粒子群按照下面的三个公式移动:

$$mbest = \frac{1}{M} \sum_{i=1}^M P_i = \left(\frac{1}{M} \sum_{i=1}^M P_{i1}, \frac{1}{M} \sum_{i=1}^M P_{i2}, \dots, \frac{1}{M} \sum_{i=1}^M P_{in} \right) \quad (14)$$

$$PP_{ij} = f \times P_{ij} + (1-f) \times P_{gj}; f = rand \quad (15)$$

$$x_{ij} = PP_{ij} \pm a \times |mbest_j - x_{ij}| \times \ln(1/u), u = rand \quad (16)$$

这里 $mbest$ 是粒子群 $pbest$ 的中间位置。 PP_{ij} 为 P_{ij} 和 P_{gj} 之间的随机点。 a 为 QPSO 的收缩扩张系数,它是 QPSO 收敛的一个重要的参数,第 T 次迭代时一般可取 $a = 0.5 + 0.5 \times (MAXTIME - T) / MAXTIME$ (具体 a 取值视情况而定),其中 $MAXTIME$ 是迭代的最大次数。QPSO 的算法流程如下:

输入 粒子的维数(等于选择的资产数),粒子的个数,选择的资产号,初始总投资价值,投资者对风险容忍的水平(即 γ 值)。

输出 最佳目标函数值,各阶段在各种情况下需要买入和卖出的资产号与资产量,各阶段平衡后各资产的投资比例以及各资产量,最终结束时各资产的期望财富。

初始化

For $T = 1$: MAXTIME

根据公式(9) 计算粒子的目标函数值;
更新每个粒子的新局部最优位置 $P_{i,j}$ 及其他相关变量;
更新全局最优位置 $P_{g,j}$ 及其他相关变量;
根据公式(14) 计算 $mbest$;
根据公式(15) 计算每个粒子随机点 $PP_{i,j}$;
根据公式(16) (以一定的概率取加和减) 更新每个粒子的新位置;

End

PSO 和 QPSO 的不同的地方在于“进化”方法不同,即更新粒子位置的方法不同。

3 多阶段投资组合优化

3.1 基于 QPSO 算法的多阶段投资组合优化

根据投资者的选择从 M 种总的可选资产中选择 N 种资产并随机指定各资产各阶段在各种情况下的权重(即投资比例),这样构成一个粒子(即一种投资组合),按此方法构成 K 个粒子。根据自然选择进化原理,种群中适应性最强者被优先选择,而适应性最差者被首先淘汰掉,本文中的适应性大小是由目标函数: $\text{Max } Z = \sum_s \pi_s \text{Measure}_t^s$ 确定的。这样如此下去直到满足一定的条件(如迭代进化一定的次数、目标函数值达到一定的期望值) 算法结束,本文的结束条件是:迭代进化达到一定的次数。

算法开始于各个粒子的初始化过程,包括以下三步:1) 根据投资者的选择从 M 种有价证券中选取 N 种有价证券构成粒子,然后随机初始化粒子的坐标(即资产的投资比例) $0 < p[i].tsw[j][s][t] < 1$,并对粒子的坐标进行归一化(即: $\sum_{j=1}^N p[i].tsw[j][s][t] = 1$),计算出以下各个变量:各阶段各种情况下平衡后各资产产量 $p[i].ts_after[j][s][t]$ 、各阶段各种情况下买入卖出的资产产量 $p[i].ts_buy[j][t]$ 和 $p[i].ts_sale[j][t]$ 、各阶段结束时各种情况下各资产的期望价值量 $p[i].ts_expect[j][t]$ 、目标函数值($p(i).ad$),依此方法形成 K 个粒子。2) 初始化各个粒子的当前最佳位置坐标 $p[i].tsw_pbest[j][s][t]$ 和当前最佳目标函数值 $p[i].p_ad$ 为粒子的初始位置及目标函数值。3) 初始化全局的最佳目标函数值 Gb_ad 以及全局最佳位置 $Gbest.tsw[j][s][t]$ 为第一个粒子的目标函数值及初始位置;以后每次迭代都依次进行以下三步:“评估”、“取代”和“修正”。

评估 对各粒子进行如下处理:首先对各维坐标进行归一化,然后计算出各阶段各种情况下平衡后各资产产量 $p[i].ts_after[j][t]$ 以及各阶段各种情况下买入卖出的资产产量 $p[i].ts_buy[j][t]$ 和 $p[i].ts_sale[j][t]$,计算出各阶段结束时各种情况下各资产的期望价值量 $p[i].ts_expect[j][t]$,最后根据公式(9) 求出各个粒子的当前目标函数值。

取代 首先比较各个粒子当前目标函数值($p[i].ad$) 与其自身当前最佳目标函数值($p[i].p_ad$),若粒子当前目标函数值大,则用该粒子的当前坐标 $p[i].tsw[j][s][t]$ 、各种情况下平衡后各资产产量 $p[i].ts_after[j][t]$ 、各阶段各种情况下买入卖出的资产产量 $p[i].ts_buy[j][t]$ 和 $p[i].ts_sale[j][t]$ 、各阶段结束时各种情况下各资产的期望价值量 $p[i].ts_expect[j][t]$ 、目标函数值 $p[i].ad$ 取代其当前最

佳位置的坐标 $p[i].tsw_pbest[j][s][t]$ 、各阶段各种情况下平衡后各资产产量 $p[i].ts_pbest_after[j][t]$ 、各阶段各种情况下买入卖出的资产产量 $p[i].ts_pbest_buy[j][t]$ 和 $p[i].ts_pbest_sale[j][t]$ 、各阶段结束时各种情况下各资产的期望价值量 $p[i].ts_pbest_expect[j][t]$ 及目标函数值 $p[i].p_ad$;然后比较所有粒子当前最佳位置的目标函数值 $p[i].p_ad$ 与全局最佳目标函数值 Gb_ad ,若粒子当前最佳位置的目标函数值大,则用该粒子的当前最佳位置的坐标、各阶段各种情况下平衡后各资产产量、各阶段各种情况下买入卖出的资产产量、各阶段结束时各种情况下各资产的期望价值量及目标函数值替代全局最佳位置的坐标 $Gbest.tsw[j][s][t]$ 、各阶段平衡后各资产产量 $Gbest.ts_after[j][t]$ 、各阶段各种情况下买入卖出的资产产量 $Gbest.ts_buy[j][t]$ 和 $Gbest.ts_sale[j][t]$ 、各阶段结束时各种情况下各资产的期望价值量 $Gbest.ts_expect[j][t]$ 。

修正 这一步是算法的核心,对每个粒子 i 进行如下操作:首先计算粒子群局部最佳位置的中间位置 $mbest[j][s][t] = (1/K) \times \sum_{i=1}^K p[i].tsw_pbest[j][s][t]$,然后计算 P_{ij} 和 P_{gj} 之间的随机点,令: $pp_w[j][s][t] = f \times p[i].tsw_pbest[j][s][t] + (1-f) \times Gbest.tsw[j][s][t]$ (其中 f 是 $0 \sim 1$ 的随机数),最后计算粒子的新位置为: $p[i].tsw[j][s][t] = pp_w[j][s][t] + a \times |p[i].tsw[j][s][t] - mbest[j][s][t]| \times \ln(1/u)$,公式右边以 0.5 的概率取加,以 0.5 的概率取减, $a = 0.9 - 0.8 \times \text{times}/\text{TIMES}$, u 为 $(0,1)$ 之间的随机数。每个粒子的新位置计算完后处理各维坐标值出现大于 1 和小于等于 0 的情况,方法是若某维大于 1 则随机赋给它一个 $0 \sim 1$ 的随机数,但保持粒子中原先各维大小的顺序不变;若某维小于等于 0 时作类似的处理。

3.2 基于遗传算法的多阶段投资组合优化

遗传算法实现的流程如下:

- 1) 随机初始化一个种群;
- 2) 评价每个染色体;
- 3) 应用适应性较强者优先选择的标准,产生最好的个体到新一代中去;
- 4) 为繁殖而选择染色体;
- 5) 运用交叉和变异操作产生新一代;
- 6) 评价新的染色体;
- 7) 如果终止条件满足,返回最佳结果;如果不满足,把精华带到新一代中去;
- 8) 如果达到种群的改组条件,随机产生剩余个体并返回 6),否则,返回 4)。

4 实证结果

为了验证算法,我们以美国标准普尔指数 100 的 100 只成分股票自 1999 年 12 月 5 日至 2005 年 12 月 5 日 6 年间的每周开盘价和收盘价以及市场指数为原始数据进行测试,把整个周期分别分成 3 个和 4 个阶段(每个阶段时间长度都是一周),分别计算出各个阶段市场指数上升和下降的概率以及市场指数上升和下降时 100 只成分股票的历史同期收益率的均值,然后从中选取不同行业的 9 种股票并把银行存款作为一项投资初始化 30 个粒子(粒子数越多效果越好,但运行时间越长)。很明显,每种投资组合中的投资资产的权值有无数种可能,所以可能的粒子数即投资组合数是无穷大。笔者在实

验过程中,参数 a 随着迭代线性地从 0.9 递减到 0.1 (多次试验证明这样效果最佳)。

遗传算法参数设定如下:终止条件:进化代数为 10000;种群大小:4;交叉概率:0.2;变异概率:0.01;用于决定改组的非改进代数最大值:20;用于决定终止的非进化代数最大值:200。

图 1 表明:若阶段数设置为 3,各阶段发生的情况有两种,则从整个周期开始到周期结束有 8 条路径;若阶段数设置为 4,各阶段发生的情况有两种,则从整个周期开始到周期结

束有 16 条路径。为了方便,每条路径用 0 表示市场指数下降,用 1 表示市场指数上升,(如 011 表示:第 1 阶段市场指数下降,第 2 阶段市场指数上升,第 3 阶段市场指数上升),表 4 仅显示两条路径的结果。

当 γ 值分别为 -1,0,1 时,等权重算法程序、遗传算法程序和 QPSO 程序都是运行 500 次,每次 QPSO 程序主循环迭代次数为 5000 次,试验得到的投资期望收益率的均值和方差如表 4 所示。

表 1 选择的美国标准普尔指数 100 的 9 只成分股票

Stock No.	Corporation	Sector	Industry
1	Anheuser-Busch Companies Inc. (BUD)	Consumer Goods	Beverages-Brewers
2	Burlington Northern Santa Fe Corp. (BNI)	Services	Railroads
3	Eastman Kodak Co. (EK)	Consumer Goods	Photographic Equipment & Supplies
4	Ford Motor Co. (F)	Consumer Goods	Auto Manufacturers-Major
5	Honeywell International Inc. (HON)	Conglomerates	Conglomerates
6	Microsoft Corp. (MSFT)	Technology	Application Software
7	Merrill Lynch & Co. Inc. (MER)	Financial	Investment Brokerage-National
8	Texas Instruments Inc. (TXN)	Technology	Semiconductor-Broad Line
9	United Technologies Corp. (UTX)	Conglomerates	Conglomerates

表 2 3 个阶段各种情节(即路径)发生概率

路径	π_s	路径	π_s
(0,0,0)	0.116505	(1,0,0)	0.184466
(0,0,1)	0.155340	(1,0,1)	0.135922
(0,1,0)	0.067961	(1,1,0)	0.067961
(0,1,1)	0.15534	(1,1,1)	0.116505

表 3 4 个阶段各种情节(即路径)发生概率

路径	π_s	路径	π_s
(0,0,0,0)	0.0389610	(1,0,0,0)	0.0909091
(0,0,0,1)	0.0909091	(1,0,0,1)	0.0519481
(0,0,1,0)	0.0779221	(1,0,1,0)	0.0649351
(0,0,1,1)	0.0519481	(1,0,1,1)	0.0259740
(0,1,0,0)	0.0519481	(1,1,0,0)	0.1038960
(0,1,0,1)	0.0519481	(1,1,0,1)	0.0649351
(0,1,1,0)	0.0519481	(1,1,1,0)	0.0389610
(0,1,1,1)	0.0649351	(1,1,1,1)	0.0779221

表 4 显示了在 γ 值不同情况下 QPSO 算法和遗传算法运行结果,通过对比和分析可知:对于每一个 γ 值,无论是分成 4 个阶段还是 3 个阶段, QPSO 算法得到的期望收益率均值都比遗传算法和等权重算法得到的均值大,且 QPSO 算法得到的期望收益率方差比遗传算法得到的期望收益率方差小得多,均值大说明算法的收敛性能好,方差小说明算法的鲁棒性好,比较结果可知在多阶段投资组合优化领域 QPSO 比遗传算法更加优越。

5 结语

这篇论文阐述了基于量子行为的微粒群优化算法(QPSO),以及如何运用 QPSO 算法制定多阶段投资决策,同时比较了 QPSO 算法与经典的遗传算法和等权重算法的运行结果,通过以上运行结果的对比和分析可以得出如下结论:与经典的遗传算法和等权重算法相比, QPSO 算法能够较好地收敛于全局最优点,而且鲁棒性好,在解决多阶段投资组合优

化问题中, QPSO 算法的投资组合优化性能比经典的遗传算法更优越。

表 4 选择银行存款作为一项投资的运行结果对比表

比较内容	阶段数设置为 3			
	等权重	遗传算法	QPSO 算法	
路径均值	$\gamma = -1$	-0.0160686	-0.00634997	0.00781309
	$\gamma = 0$	-0.0160686	-0.00657826	0.00781309
	$\gamma = 1$	-0.0160686	-0.00739237	0.00781309
0 方差	$\gamma = -1$	2.1079e-033	9.1494e-006	2.52707e-029
	$\gamma = 0$	4.8148e-035	7.3952e-006	3.10514e-028
	$\gamma = -1$	4.8148e-035	1.0644e-005	6.98818e-029
路径均值	$\gamma = -1$	-0.00537535	0.0120198	0.0234148
	$\gamma = 0$	-0.00537535	0.0124407	0.0234148
	$\gamma = 1$	-0.00537535	0.0115636	0.0234148
1 0 方差	$\gamma = -1$	3.0927e-034	2.2629e-005	4.8759e-027
	$\gamma = 0$	1.0833e-034	1.6571e-005	4.21215e-029
	$\gamma = 1$	1.0833e-034	3.1704e-005	6.30861e-029
比较内容	阶段数设置为 4			
	等权重	遗传算法	QPSO 算法	
路径均值	$\gamma = -1$	-0.0178517	-0.00897002	0.0142964
	$\gamma = 0$	-0.0178517	-0.00893489	0.0142964
	$\gamma = 1$	-0.0178517	-0.00935024	0.0142602
0 方差	$\gamma = -1$	2.0268e-033	1.5319e-005	3.38669e-016
	$\gamma = 0$	5.8981e-034	9.8758e-006	2.05478e-016
	$\gamma = 1$	2.1625e-033	1.0154e-005	4.9113e-008
路径均值	$\gamma = -1$	0.0146243	0.0157732	0.0558503
	$\gamma = 0$	0.0146243	0.0157362	0.0558503
	$\gamma = 1$	0.0146243	0.0159392	0.0558503
1 1 0 方差	$\gamma = -1$	1.1948e-033	1.0526e-005	8.34164e-016
	$\gamma = 0$	1.9259e-034	1.4535e-005	9.59182e-016
	$\gamma = 1$	2.8563e-033	1.4818e-005	1.03957e-014

参考文献:

[1] SUN J, XU WB. A Global Search Strategy of Quantum-behaved Particle Swarm Optimization[A]. Proceedings of IEEE Conference on Cybernetics and Intelligent Systems[C], 2004. 111 - 116.

的似然函数。因此在本算法中采用一个近似方式表示似然函数,即:

$$L(y | \theta^{(t)}) \approx \bar{L}(y | \theta^{(t)}) = \bar{L}(y | \theta^{(t-1)}) - L(y_i | \theta^{(t-1)}) + L(y_i | \theta^{(t)}) \quad (6)$$

Neal^[6]证明对于所有 $N > k, L^{(t)} < L^{(t-k)}$ 成立。因此当 $C_k^{(t)} = \frac{L^{(t)} - L^{(t-k)}}{L^{(t)}}$ 小于某极限值 ε 时,EM 算法收敛,在本文中取:

$$C_k^{(t)} = \frac{\bar{L}^{(t)} - \bar{L}^{(t-k)}}{\bar{L}^{(t)}} \quad (7)$$

这种划分数据块的方法非常适合于样本数据量很大的情况,通过将数据集划分为较小的块,分别对每个块进行块内优化处理,并且在块间进行合成,划分数据块的方法在很大程度上减少了 EM 算法的计算量。

3 实验结果及分析

3.1 算法计算精度比较

使用本文提出的改进的 EM 算法与标准 EM 算法进行比较,分析本文算法的参数学习性能,用于测试的贝叶斯网络如图 1 所示。

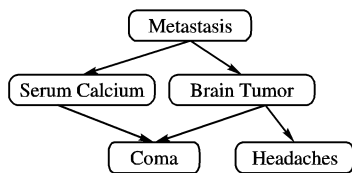


图 1 Cancer 的网络拓扑结构

利用概率逻辑抽样算法分别产生 40 个训练样本集,分别包括 25, 50, 75, ..., 400, 425, 450, ..., 950, 975, 1 000 个样本。使用 KL 距离作为参数学习结果与实际节点条件概率的接近程度,采用改进的 EM 算法与 EM 算法对节点 SerumCalcium 参数学习的 KL 距离如图 2 所示。

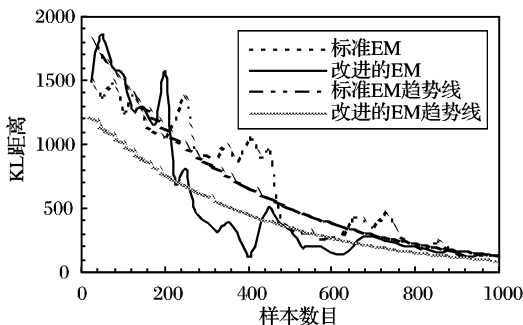


图 2 SerumCalcium 节点两种算法参数学习的 KL 距离

图 2 中,横坐标表示训练样本集的大小(所含样本的数目),纵坐标表示参数学习结果 $p'(V | Pa(v))$ 与真实值 $p(V | Pa(v))$ 之间的 K-L 距离。从图可以清楚地看出,对于相同的

训练样本集,采用改进的 EM 算法所得结果要比采用标准 EM 算法所得结果与真实概率间的 KL 距离要小,这表明其结果也更为精确。而从趋势线也可以看出,当训练样本集较小时,改进的 EM 算法所得参数学习结果明显优于最大似然估计算法所得结果,而当训练样本集大小增加时,两种算法所得参数结果趋于相同。

3.2 算法时间性能比较

对节点 SerumCalcium,分别采用改进的 EM 算法和标准 EM 算法对计算似然函数所用时间进行比较,数据集取样本数目为 1000 的集合,似然函数与所用时间变化情况如图 3 所示。

从图 3 中可以发现在节点 SerumCalcium,改进的 EM 算法使用约 300s 就估计出最大似然函数,而 EM 算法使用了约 600s,说明改进的 EM 算法速度上明显优于标准的 EM 算法。

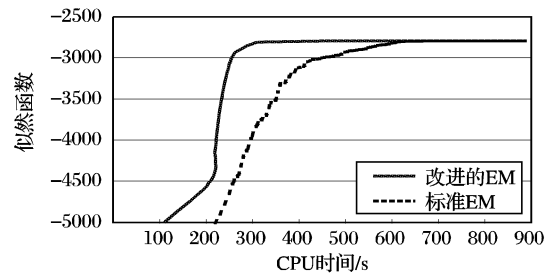


图 3 SerumCalcium 节点两种算法的似然函数与时间关系

4 结语

EM 算法是处理不完整数据集贝叶斯网络参数学习的常用方法,通过将大规模数据集划分为较小的数据块,并通过块内优化,块外合成的方法对 EM 算法进行了改进,实验表明改进的 EM 算法在速度和精度方面都具有较好的性能。

参考文献:

- [1] COOPER GF, HERSKOVITS E. A Bayesian Method for the induction of probabilistic networks from data[J]. Machine Learning, 1992, 9(4): 309 - 347.
- [2] SPIEGELHALTER DJ, LAURITZEN SL. Sequential Updating of Conditional Probabilities on Directed Graphical Structures[J]. Networks, 1990, 20(5): 579 - 605.
- [3] GOLMARD JL, MALLET A. Learning probabilities in causal trees from incomplete databases[J]. Artificial Intelligence, 1991, 5(1): 93 - 106.
- [4] LAURITZEN SL. The EM algorithm for graphical association models with missing data[J]. Computational Statistics and Data Analysis, 1995, 19(2): 191 - 201.
- [5] THIESSON B, MEEK C, HECKERMAN D. MSR-TR-99-31, Accelerating EM for large databases[R]. Microsoft Research, 2001.
- [6] NEAL R, HINTON G. A view of the EM algorithm that justifies incremental, sparse, and other variants[A]. Learning in Graphical Models[C]. Kluwer Academic Publishers, 1998. 355 - 371.

(上接第 1685 页)

- [2] SUN J, FENG B, XU WB. Particle Swarm Optimization with Particles Having Quantum Behavior[A]. Proceedings of 2004 Congress on Evolutionary Computation[C], 2004. 325 - 331.
- [3] MULVEY JM, ROSENBAUM DP, SHETTY B. Strategic Financial risk management and operations research[J]. European Journal of Operational Research, 1997, 97(1): 1 - 16.
- [4] 曾建潮,介婧,崔志华. 微粒群算法[M]. 第 1 版. 北京:科学出版社, 2004.
- [5] CHAN MC, WONG CC, CHEUNG BK-S, et al. Genetic Algorithms in

- Multi-Stage Portfolio Optimization System[A]. Computing in Economics and Finance[C]. Society for Computational Economics, 2002.
- [6] BERGER A, MULVEY J. Integrative Risk Management for Individual[A]. ZIEMBA W, MULVEY J. Worldwide Asset and Liability Management[C]. Cambridge University Press, 1996.
- [7] CARINO DR, ZIEMBA WT. Formulation of the Russell-Yasuda Kasai Financial Planning Model[Z]. Frank Russell Company, 1995.