

基于正交多主体遗传算法的业务规则引擎推理方法^{*}

张磊¹ 张瑞生² 李廉²

(兰州大学数学与统计学院 兰州 730000)¹ (兰州大学信息科学与工程学院 兰州 730000)²

摘要 现阶段业务规则引擎多采用前向链接算法实现模式匹配的推理过程。本文提出了一种基于正交多主体遗传算法的业务规则引擎推理方法,以解决规则引擎无法求解类似组合服务竞标的竞胜标确定问题。该方法以多主体系统为基础,利用正交试验设计方法产生较好的初始种群,设计正交交叉操作数以获得更好的后代;通过每个主体的逐步演化和自学习功能来提高算法的全局优化能力和收敛速度。

关键词 规则引擎,竞胜标确定,正交试验设计,多主体系统,正交多主体遗传算法

The Business Rules Engine Inference Approach Based on Orthogonal Multi-agent Genetic Algorithm

ZHANG Lei¹ ZHANG Rui-Sheng² LI Lian²

(School of Mathematics and Statistics, Lanzhou University, Lanzhou 730000)¹

(School of Information Science and Engineering, Lanzhou University, Lanzhou 730000)²

Abstract It uses Forward-Chaining arithmetic to realize inference in business rules engine. It proposes a inference approach of business rules engine based on orthogonal multi-agent genetic algorithm in this paper to solve the limitation that business rules engine can not solve The Winner Determination Approach of Combinatorial Service Compete. Orthogonal design is introduced to generate an initial population of points that are scattered uniformly over the feasible solution space and to generate a crossover operator. It realizes the global optimal computation via the local interacting agents with abilities of local perceptivity, competition and evolvement, self-learning etc.

Keywords Rules engine, Winner determination, Orthogonal experimental design, Multi-agents system, Orthogonal multi-agent genetic algorithm

业务规则管理技术(BRM, Business Rules Management)为企业级软件开发提供了快速便捷的变化管理,它将业务逻辑和技术逻辑进行分离并统一存储,利用业务规则引擎在运行时动态管理和修改业务规则,从而提高软件系统柔性和适应性。

现阶段主流规则引擎多基于前向链接 Rete 算法及其变体实现模式匹配的推理过程,而该算法无法解决类似组合服务竞标的竞胜标确定问题。本文提出了一种基于正交多主体遗传算法的业务规则引擎推理方法,以解决规则引擎无法实现上述业务规则求解的问题。

1 业务规则引擎分析

业务规则管理技术将规则的实现从具体程序代码中抽取出来,将业务逻辑当作结构化对象进行处理,生成对应业务规则,将其复杂逻辑关系交给规则引擎处理。

规则引擎由推理引擎发展而来,是一种嵌入在应用程序中的组件,使用预定义语义模块接受数据输入,并与加载在引擎中的业务规则进行测试匹配,启动那些符合当前数据状态下的业务规则,根据业务规则中声明的执行逻辑触发对应操作。

现阶段规则引擎多基于前向链接算法实现模式匹配的推理过程,即从已知事实出发推出解答或建议,结构多为一组条件模板和操作,当条件模板和工作空间中元素匹配时,就运行

其对应操作。Rete 算法及其变体是在模式匹配中利用推理机的时间冗余性和规则结构的相似性,通过保存中间去处来提高推理效率的一种模式匹配算法,其核心思想是将分离的匹配项根据内容动态构造匹配树,以达到显著降低计算量的效果^[1]。

在企业运行环境尤其是电信运营环境中,类似于电信组合服务竞标的竞胜标确定问题是在进行服务竞标时需要解决的一类关键业务决策问题。在运营支撑系统中需要针对服务提供商所提供的服务开通时间、服务价格及其它相关 QoS 指标来确定服务竞标归属,类似于组合拍卖问题。电信组合服务竞标可以使电信服务竞标方把多个服务看作一个标的来竞标。这种方式不仅能够提高服务竞标的效率,而且能降低服务竞标方的竞标风险,因而这种服务竞标方式在实际应用中具有广阔的应用前景。

然而 Rete 算法及其变体无法解决类似电信组合服务竞标的竞胜标确定问题。在电信组合服务竞标的机理设计中,竞胜标确定问题与组合拍卖竞胜标确定问题类似,是一个 NP 难题^[2,3]。现代启发式算法给组合拍卖竞胜标确定问题的求解带来十分光明的科学前景。文[4~6]对该问题的精确算法和近似算法做了分析研究。文[7]在深入研究该问题模型的特性基础上,给出了求解该问题的启发式遗传算法。文[8]采用改进遗传算法求解组合拍卖竞胜标问题。

本文在现有遗传算法基础上,将多主体系统^[9~11]、正交

^{*}国家自然科学基金重大研究计划重点项目(90612016)。张磊 博士生,研究方向为支撑系统与软件、人工智能和多主体系统;张瑞生 博士,博导,研究方向为支撑系统与软件、人工智能和化学信息学。

试验设计^[12]、遗传算法^[13,14]相结合,提出了一种基于正交多主体遗传算法的电信组合服务竞标的竞标标确定方法,以解决业务规则引擎无法实现这类复杂业务规则求解的问题。

2 正交多主体遗传算法

为了有效求解电信组合服务竞标的竞标标确定问题,笔者特设计了一种正交多主体遗传算法。该方法以多主体系统为基础,通过每个主体的逐步演化和自学习功能来提高算法的全局优化能力和收敛速度;利用正交试验设计方法产生较好的初始种群,并设计正交交叉算子以获得更好的后代。在该算法中,单个染色体对应的是待优化问题的一组输入;染色体个体的适应度评价是指,对应于该染色体所刻画的一组输入,待优化问题目标函数的输出值。该方法的计算流程图如图1所示。

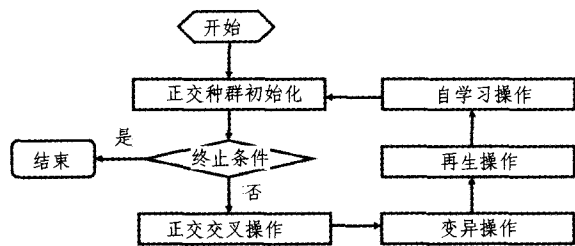


图1 正交多主体遗传算法的计算流程

2.1 问题描述

假设 M 为待竞标电信服务的集合;对任意服务竞标方 i ,可对 M 中的单项服务或服务的组合 $S \subseteq M$ 进行竞标,其标价为 $b_i(S)$ 。为了便于处理,这里对待研究问题进行一定的简化。

对“服务标的”相同的服务标,只保留其最高价格的服务标(设这样的服务标有 n 个)。电信组合服务竞标的竞标标确定问题的一般模型可描述如下:

$$\max_{x \in A} \sum_{S \in X} b(S) \quad (1)$$

电信组合服务竞标的竞标标确定问题一般模型的三点说明:①模型的优化目标为最大化服务运营商的收益;②模型中的 X 为一个可行解,每个项目至多可分配给一个服务标;③模型中的 A 为可行解空间,因问题特性,有如下定义:

$$A = \{W \subseteq \Gamma \mid \forall S, S' \in W, S \cap S' = \emptyset\} \quad (2)$$

这里, W 是 M 的族集(即 $\bigcup_{S \in W} S = M$), Γ 是 W 的集合。

从上面的模型中可以看出,电信组合服务竞标的竞标标确定问题是一个组合优化问题。该问题的求解是一个十分复杂的计算问题,是一个 NP 难题。

2.2 编码方式

编码是设计遗传算法的一个关键步骤,需要针对问题设计染色体,包括基因字串的长度以及基因代表的含义,即对要搜索空间的可行解以编码的形式来呈现,笔者采用直接编码方法来实现本问题可行解的编码。每个染色体对应一个可行解,每个基因对应于不同的“服务标的”。在本问题中,由于染色体中的基因数是不固定的,因此染色体的长度是可变的;而且在染色体中基因的顺序没有意义。假设待竞标服务数为 m 个,其集合为 M ,简化后所剩的“有竞争力的服务标”数为 n ,对应每个服务标 $b_i (i=1, 2, \dots, n)$ 的“服务标的”为 $S_i (i=1, 2, \dots, n)$,则可以用下面式子来表示一个染色体:

$$\{b_1(S_1), b_6(S_6), b_9(S_9), \dots, b_k(S_k)\} \quad (3)$$

这里, $S_i \in \{S_k \mid S_k \subseteq M, k=1, 2, \dots, n\}$,对于集合 $\{1, 6, 9, \dots, k\}$ 中任意的 $i \neq j$,有 $S_i \cap S_j = \emptyset$ 和 $\bigcup_{i \in \{1, 6, 9, \dots, k\}} S_i = M$ 。

2.3 正交种群初始化

在完成染色体编码以后,需要产生初始种群作为起始解,所以首先需要决定初始化种群的数目,染色体表示个体的代码,初始群体的染色体一般采用随机初始化的方式产生。笔者将每个染色体看作是一个主体(Agent)。由于染色体的长度是可变的,因而每个主体的大小也是不同的。本文正交种群初始化的操作思路如下所示。

(1)将 m 个服务标的随机分成 t (t 为设计参数)个子集:

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	...	S_{m-2}	S_{m-1}	S_m
P_1			P_2			P_3		P_t			

(2)根据“优先适合启发式”规则,将子集 P_i 转化为 B_i 个“服务标”的集合,在每个集合中,子集 P_i 内的每个服务标的不重复、不遗漏;

(3)将每个子集内的各个“服务标”的集合组合起来,共生成 $\prod_{i=1}^t B_i$ 个可行解;

(4)从这些可行解中选取最好的 G 个主体(可行解)作为初始种群。

其“优先适合启发式”规则的操作步骤如下所示。

STEP1:保留服务标的排列中第一个服务标 $b_1(S_1)$, 把它的“服务标的”作为染色体中第一个基因,并令 $U = \{S_1\}$;

STEP2:在服务标的排列中顺序选取余下的服务标 $b_j(S_j)$, 若 $U \cap S_j \neq \emptyset$, 则保留 $b_j(S_j)$, 把它的“服务标的”作为染色体中基因,顺序地补充到染色体中去,并令 $U = U \cup S_j$; 否则,进行下一步;

STEP3:重复 STEP2,直到 $U = M$ 。

2.4 正交交叉操作

交叉操作是遗传算法中起核心作用的遗传操作,是指两个父代个体的部分结构加以替换重组而生成新个体的操作。本文借助正交试验的实际思想,定义了正交交叉操作,可使本算法在参数寻优过程中,总向好的方向进行,以获得最优参数。本文正交交叉操作的操作思路如下:

(1)将当前染色体中的 m 个服务标的随机分成 t (t 为设计参数)个子集:

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	...	a_{m-2}	a_{m-1}	a_m
P_1			P_2			P_3		P_t			

(2)生成正交表 $L_F(2^t) = [\beta_j]_{F \times 2^t}$, 这里的 Q 取奇数, $F = Q^J$, J 是满足关系式 $(Q^J - 1)/(Q - 1) = 2$ 的正整数;

(3)按照 $L_F(2^t)$ 中每一行的两个数值,选取相应的两个子集(例如:如果正交表中当前行中的元素为 1 和 3,则选取子集 P_1 和 P_3);

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	...	a_{m-2}	a_{m-1}	a_m
P_1			P_2			P_3		P_t			

(4)交换这两个子集上的服务标,这样就形成一个新的服务标的排列;

(5)将新产生服务标的排列按照优先启发式构造一个新的可行解(即主体);

(6)从这 F 个新主体中选取一个最好的主体作为最终的交叉结果。

(下转第 173 页)

- ing Incremental Wavelet Approximations. In: Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003
- 13 王永利,周景华,徐宏炳,等. 时间序列数据流的自适应预测. 自动化学报(已录用,待发表)
 - 14 Trudnowski J D, McReynolds W L, Johnson M J. Real-time very short-term load prediction for power-system automatic generation control [J]. IEEE Transactions on Control Systems Technology, 2001, 9(2): 254~260
 - 15 徐科,徐金梧,班晓娟. 基于小波分解的某些非平稳时间序列预测方法[J]. 电子学报, 2001, 29(4): 266~268
 - 16 贺国光,马寿峰,李宇. 基于小波分解与重构的时间序列预测法[J]. 自动化学报, 2002(6): 1012~1014
 - 17 Liu K, Subbarayan S, Shoultz R R, et al. Comparison of Very Short-term Load Forecasting Techniques [J]. IEEE Transactions on Power System, 1996, 11(2): 232~239
 - 18 Fisher D H. Knowledge Acquisition via Incremental Conceptual Clustering. Machine Learning, 1987, 139~172
 - 19 Utgoff P E. An Improved Algorithm for Incremental Induction of Decision Trees. In: Proc. of the Eleventh International Conference on Machine Learning, 1994, 318~325
 - 20 Wang H, Fan W, Yu P, et al. Mining Concept-Drifting Data Streams Using Ensemble Classifiers. In: 9th ACM International Conference on Knowledge Discovery and Data Mining (SIGKDD), Washington DC, USA, Aug. 2003
 - 21 Gehrke J, Rastogi R. Querying and mining data stream: you only get one look. A tutorial. In: Franklin MJ, Moon B, Ailamaki A,

- eds. Proc. of the 2002 ACM SIGMOD Int'l Conf. on Management of Data. Madison: ACM Press, 2002. 635
- 22 Domingos P, Hulten G. Mining High-speed Data Streams. In: Proceedings of the Association for Computing Machinery Sixth International Conference on Knowledge Discovery and Data Mining, 2000
- 23 Papadimitriou S, Faloutsos C, Brockwell A. Adaptive, Hands-off Stream Mining. In: 29th International Conference on Very Large Data Bases VLDB, 2003
- 24 Aggarwal C, Han J, Wang J, et al. On Demand Classification of Data Streams. In: Proc. 2004 Int Conf. on Knowledge Discovery and Data Mining, Seattle, WA, Aug. 2004
- 25 Last M. Online Classification of Nonstationary Data Streams. Intelligent Data Analysis, 2002, 6(2): 129~147
- 26 Ding Q, Ding Q, Perrizo W. Decision Tree Classification of Spatial Data Streams Using Peano Count Trees. In: Proceedings of the ACM Symposium on Applied Computing, Madrid, Spain, March 2002
- 27 Gaber M M, Krishnaswamy S, Zaslavsky A. On-board Mining of Data Streams in Sensor Networks. S Badhyopadhyays, Maulik U, Haolder L, et al. eds. Advanced Methods of Knowledge Discovery from Complex Data, Springer Verlag
- 28 Last M, Klein Y, Kandel A. Knowledge Discovery in Time Series Databases. IEEE Transactions on Systems, Man, and Cybernetics, 31(Part B, 1): 160~169, 20

(上接第 155 页)

2.5 变异操作

变异操作的基本内容是对群体中个体串的某些基因座的基因值作变动。本文变异操作的操作思路如下:

STEP1:在父代染色体对应的服务标的排列上随机选择某一段子排列;

STEP2:使得这个服务标的排列上的子排列顺序产生倒转,这样就形成一个新的服务标的排列;

STEP3:将新产生服务标的排列按优先适合启发式规则构造一个新的主体(即染色体)。

2.6 再生操作

为了提高本文算法的收敛速度,笔者在主体完成演化操作后,对当代主体中适应度最低的主体执行一次再生操作。再生操作建立在群体中个体的适应度评估的基础上,目的是把优化的个体直接遗传到下一代,或者通过配对交叉产生新的个体再遗传到下一代。本算法再生操作就是对当前主体再次进行正交交叉操作和变异操作,希冀能通过再生操作来提高该主体的适应度。

2.7 自学习操作

为了提高本文算法的全局优化能力和收敛速度,笔者在主体完成演化操作后,对每个主体都执行一次自学习操作。

该操作的具体流程如下所示。

STEP1:在父代染色体对应的服务标的排列上随机选择某一段子排列;

STEP2:将这个服务标的排列上的子排列中的最后一个服务标插入到子排列的最前面位置,子排列中的其他服务标依次向后移一位,这样就形成一个新的服务标的排列;

STEP3:将新产生服务标的排列按优先适合启发式规则构造一个新的主体(即染色体)。

2.8 终止条件

严格地讲,遗传算法的迭代终止条件目前尚无定论,一般通过多次进化逐渐逼近最优解而不是恰好等于最优解,因此需要确定其终止条件,最常用的终止方法是规定遗传的代数。在本算法求解过程中,笔者始终保存历代最优解,同时指定最大演化代数作为停止准则,即选一个大的正整数 Max-Gen 为最大迭代次数,若当前迭代次数 k 大于 Max-Gen,则停止迭代并输出历史最优解作为最终的结果。

结论 本文主要创新点:提出了一种基于正交多主体遗传算法的电信组合服务竞标的竞胜标确定方法,以解决业务规则引擎无法实现这类复杂业务规则求解的问题。该方法以多主体系统为基础,通过每个主体的逐步演化和自学习功能来提高算法的全局优化能力和收敛速度;利用正交试验设计方法产生较好的初始种群和设计正交交叉算子以获得更好的后代。

参考文献

- 1 张渊,夏清国. 基于 Rete 算法的 JAVA 规则引擎. 科学技术与工程, 2006, 6(11): 2
- 2 Mito M, Fujita S. On heuristics for solving winner determination problem in combinatorial auctions [J]. Journal of Heuristics, 2002, 10(5): 507~523
- 3 Sandholm T, Suri S, Gilpin A, et al. CABOB: A fast optimal algorithm for winner determination in combinatorial auctions. Management Science (S0025-1909), 2005, 51(3): 374~390
- 4 Sandholm T. Approaches to winner determination in combinatorial auctions [J]. Decision Support Systems (S0167-9236), 2000, 28(1): 165~176
- 5 Kelly F, Steinberg R. A combinatorial auction with multiple winners for universal service [J]. Management Science, 2000, 46(4): 586~596
- 6 Kuyanoglu E, Wu S D. On combinatorial auction and Lagrangian relaxation for distributed resource scheduling [J]. IIE Transactions, 1999, 31(9): 813~826
- 7 陈培友,汪定伟. 用遗传算法求解组合拍卖竞胜标[J]. 东北大学学报(自然科学版), 2003, 24(1): 7~10
- 8 陈培友,汪定伟. 用改进遗传算法求解组合拍卖竞胜标[J]. 东北大学学报(自然科学版), 2004, 25(4): 349~352
- 9 Roias E M, Mukherjee A. Multi-agent framework for general-purpose situational simulations in the construction management domain [J]. Journal of Computing in Civil Engineering, 2006, 20(3): 165~176
- 10 Ota J. Multi-agent robot systems as distributed autonomous systems [J]. Advanced Engineering Informatics, 2006, 20(1): 59~70
- 11 Camacho D, Aler R, Borrajo D, et al. Multi-agent plan based information gathering [J]. Applied Intelligence, 2006, 25(1): 59~71
- 12 Kan H, Shen H. Lower bounds on the minimal delay of complex orthogonal designs with maximal rates [J]. IEEE Transactions on Communications, 2006, 54(3): 383~388
- 13 何大阔,王福利,毛志忠. 遗传算法在离散变量优化问题中的应用研究[J]. 系统仿真学报, 2006, 18(5): 1154~1156
- 14 姚钦,史仪凯,夏锐. 多目标交互式遗传算法在测试点确定问题中的应用[J]. 系统仿真学报, 2006, 18(6): 1469~1472