



(12) 发明专利

(10) 授权公告号 CN 117648552 B

(45) 授权公告日 2025. 02. 25

(21) 申请号 202311750776.0

G06F 18/241 (2023.01)

(22) 申请日 2023.12.18

G06N 3/006 (2023.01)

(65) 同一申请的已公布的文献号

申请公布号 CN 117648552 A

(56) 对比文件

CN 107734035 A, 2018.02.23

CN 109614445 A, 2019.04.12

(43) 申请公布日 2024.03.05

审查员 唐进岭

(73) 专利权人 深圳世纪前沿量化科技有限公司

地址 518000 广东省深圳市福田区福田街

道皇岗社区益田路3008号皇都广场3

号楼1114D7

(72) 发明人 陈家馨

(74) 专利代理机构 广东皓行知识产权代理事务

所(普通合伙) 441037

专利代理师 李俊杰

(51) Int. Cl.

G06F 18/20 (2023.01)

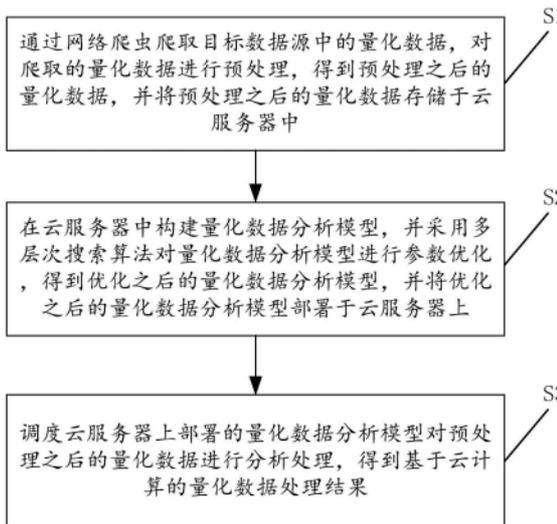
权利要求书4页 说明书10页 附图1页

(54) 发明名称

一种基于云计算的量化数据处理方法

(57) 摘要

本发明公开了一种基于云计算的量化数据处理方法,属于云计算以及量化数据处理技术领域,通过采用深度学习算法以及采用多层次搜索算法对量化数据分析模型进行参数优化,从而实现对历史数据的学习,形成可以处理量化数据的规则,再通过该规则对量化数据进行处理,从而实现量化数据的自动化处理,不仅能够有效地提高量化数据的分析效率,还能够消除人为分析带来的误差。



1. 一种基于云计算的量化数据处理方法,其特征在于,包括:

通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行预处理,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中;

在云服务器中构建量化数据分析模型,并采用多层次搜索算法对量化数据分析模型进行参数优化,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上;

调度云服务器上部署的量化数据分析模型对预处理之后的量化数据进行分析处理,得到基于云计算的量化数据处理结果;

通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行预处理,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中,包括:

通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行缺失值处理、异常值处理以及归一化处理,得到初始处理之后的量化数据;

将初始化之后的量化数据统一存储为预设数据结构,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中;

在云服务器中构建量化数据分析模型,包括:构建深度学习模型,并将构建的深度学习模型作为量化数据分析模型;

采用多层次搜索算法对量化数据分析模型进行参数优化,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上,包括:

初始化量化数据分析模型的模型参数,得到模型参数向量,并重复获取多个模型参数向量,得到种群;

获取历史数据以及历史数据对应的标签数据,并采用历史数据以及历史数据对应的标签数据,获取每个模型参数向量对应的适应度值;

针对每个模型参数向量,采用固定视野以及自适应步长的局部搜索方法对模型参数向量进行局部搜索,并根据模型参数向量对应的适应度值对局部搜索过程进行控制,获取局部搜索之后的模型参数向量;

针对局部搜索之后的模型参数向量,采用变动视野的全局平衡搜索方法对模型参数向量进行平衡搜索,并根据模型参数向量对应的适应度值对平衡搜索过程进行控制,获取平衡搜索之后的模型参数向量;

针对平衡搜索之后的模型参数向量,采用种群信息融合的全局搜索方法对模型参数向量进行全局搜索,并根据模型参数向量对应的适应度值对全局搜索过程进行控制,获取全局搜索之后的模型参数向量;

针对全局搜索之后的模型参数向量,重新获取模型参数向量的适应度值,并根据适应度值获取个体扩展数量;

根据个体扩展数量,采用正态分布方法产生多个新模型参数向量,并将新模型参数向量与全局搜索之后的模型参数向量进行融合,再采用优胜劣汰的方法抉择出新种群;

针对新种群,获取每个模型参数向量对应的适应度值,并根据每个模型参数向量对应的适应度值获取最优模型参数向量;

判断是否满足算法终止条件,若是,则输出最优模型参数向量,否则返回局部搜索的步骤;

将输出的最优模型参数向量作为量化数据分析模型的最终模型参数,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上。

2. 根据权利要求1所述的基于云计算的量化数据处理方法,其特征在于,针对每个模型参数向量,采用固定视野以及自适应步长的局部搜索方法对模型参数向量进行局部搜索,并根据模型参数向量对应的适应度值对局部搜索过程进行控制,获取局部搜索之后的模型参数向量,包括:

针对每个模型参数向量,生成一个随机向量为: $(\Delta x_{i,1}, \Delta x_{i,2}, \dots, \Delta x_{i,d}, \dots, \Delta x_{i,D})$, 其中, $\Delta x_{i,d}$ 表示第i个模型参数向量中第d维参数对应的随机参数, $i=1, 2, \dots, L$, L表示模型参数向量的总数, $d=1, 2, \dots, D$, D表示模型参数向量中参数总维度,且 $\Delta x_{i,d}$ 按各0.5的概率随机为 α^t 或者 $-\alpha^t$, α^t 表示第t次优化过程中的自适应步长;

获取第i个模型参数向量的伪梯度为:

$$F(X_i^t)_d = \frac{f(X_i^t + \Delta X_i^t) - f(X_i^t - \Delta X_i^t)}{2\Delta x_{i,d}}$$

其中, $F(X_i^t)_d$ 表示第i个模型参数向量中第d维参数对应的为梯度, $f(*)$ 表示适应度函数, X_i^t 表示第t次优化过程中的第i个模型参数向量, ΔX_i^t 表示随机向量;

根据预设固定视野 β_0 获取第t+1次优化过程中的自适应步长 α^{t+1} , 并根据自适应步长 α^{t+1} 以及第i个模型参数向量的伪梯度 $F(X_i^t)_d$, 对第i个模型参数向量进行局部搜索为:

$$x_{i,d}^{t+1} = x_{i,d}^t + \alpha^{t+1} * \text{sign}(F(X_i^t)_d)$$

其中, $x_{i,d}^t$ 表示第t次优化过程中的第i个模型参数向量中第d维参数, $x_{i,d}^{t+1}$ 表示 $x_{i,d}^t$ 对应的局部搜索值;

判断局部搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该局部搜索,得到局部搜索之后的模型参数向量,否则直接将原有的模型参数向量作为局部搜索之后的模型参数向量。

3. 根据权利要求2所述的基于云计算的量化数据处理方法,其特征在于,根据预设固定视野 β_0 获取第t+1次优化过程中的自适应步长 α^{t+1} 为:

$$\alpha^{t+1} = \frac{\alpha_0}{20} + \beta_0 * \sin\left(\frac{\pi}{2} * \left(1 - \frac{t-1}{T}\right)\right)$$

其中, α_0 表示初始步长, π 表示圆周率, T 表示最大优化次数。

4. 根据权利要求3所述的基于云计算的量化数据处理方法,其特征在于,针对局部搜索之后的模型参数向量,采用变动视野的全局平衡搜索方法对模型参数向量进行平衡搜索,并根据模型参数向量对应的适应度值对平衡搜索过程进行控制,获取平衡搜索之后的模型参数向量,包括:

以当前优化次数t为基础,获取变动视野为:

$$\beta^{t+1} = \frac{1}{1 + e^{2\ln 100(t/T) * \ln 1000}} * \frac{T-t+1}{T} * \beta^t$$

其中, β^t 表示第t次优化过程中的变动视野, β^{t+1} 表示第t+1次优化过程中的变动视野,且变动视野初始值为 β_0 , e 表示自然常数;

根据变动视野 β^{t+1} ,获取平衡搜索范围为:

$$lb_{i,d} = x_{i,d}^{t+1} - \beta^{t+1}$$

$$ub_{i,d} = x_{i,d}^{t+1} + \beta^{t+1}$$

其中, $lb_{i,d}$ 表示第i个模型参数向量中第d维参数的平衡搜索下界, $ub_{i,d}$ 表示第i个模型参数向量中第d维参数的平衡搜索上界, $x_{i,d}^{t+1}$ 表示局部搜索之后的第i个模型参数向量中第d维参数;

根据平衡搜索范围,对局部搜索之后的模型参数向量进行平衡搜索为:

$$x_{i,d}^{t+1'} = lb_{i,d} + \text{rand}(0,1) * (ub_{i,d} - lb_{i,d})$$

其中, $\text{rand}(0,1)$ 表示(0,1)之间的随机数, $x_{i,d}^{t+1'}$ 表示平衡搜索之后的 $x_{i,d}^{t+1}$;

判断平衡搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该平衡搜索,得到平衡搜索之后的模型参数向量,否则直接将局部搜索之后的模型参数向量作为平衡搜索之后的模型参数向量。

5.根据权利要求4所述的基于云计算的量化数据处理方法,其特征在于,针对平衡搜索之后的模型参数向量,采用种群信息融合的全局搜索方法对模型参数向量进行全局搜索,并根据模型参数向量对应的适应度值对全局搜索过程进行控制,获取全局搜索之后的模型参数向量,包括:

根据所有平衡搜索之后的模型参数向量,获取种群的平均位置为 \bar{X} ; \bar{X} 中每一维参数为其他所有模型参数向量在同维参数的平均值;

根据种群的平均位置 \bar{X} ,对平衡搜索之后的模型参数向量进行全局搜索为:

$$x_{i,d}^{t+1''} = x_{i,d}^{t+1'} + \text{rand}(x_{\min}, x_{\max}) * (\bar{x}_d - x_{i,d}^{t+1'})$$

其中, $x_{i,d}^{t+1''}$ 表示全局搜索之后的模型参数向量, $\text{rand}(x_{\min}, x_{\max})$ 表示 (x_{\min}, x_{\max}) 之间的随机数, x_{\min} 表示全局搜索区间下限, x_{\max} 表示全局搜索区间上限, \bar{x}_d 表示平均位置 \bar{X} 中第d维参数;

判断全局搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该全局搜索,得到全局搜索之后的模型参数向量,否则直接将平衡搜索之后的模型参数向量作为全局搜索之后的模型参数向量。

6.根据权利要求5所述的基于云计算的量化数据处理方法,其特征在于,针对全局搜索之后的模型参数向量,重新获取模型参数向量的适应度值,并根据适应度值获取个体扩展数量为:

$$\gamma_i = \xi \left(\frac{f_i - f_{\min}}{f_{\max} - f_{\min}} * (S_{\max} - S_{\min}) + S_{\min} \right)$$

其中, γ_i 表示第i个全局搜索之后的模型参数向量对应的个体扩展数量, $\xi(*)$ 表示取整函数, f_i 表示第i个全局搜索之后的模型参数向量对应的适应度值, f_{\max} 表示全局搜索之后的模型参数向量对应的最大适应度值, f_{\min} 表示全局搜索之后的模型参数向量对应的最小适应度值, S_{\max} 表示预设的最大扩展整数, S_{\min} 表示预设的最小扩展整数。

7. 根据权利要求6所述的基于云计算的量化数据处理方法,其特征在于,根据个体扩展数量,采用正态分布方法产生多个新模型参数向量,并将新模型参数向量与全局搜索之后的模型参数向量进行融合,再采用优胜劣汰的方法抉择出新种群,包括:

针对每个全局搜索之后的模型参数向量,以正态分布 $N(\mathbf{0}, \sigma_t^2)$ 在模型参数向量周围产生与个体扩展数量相等的多个新模型参数向量;

其中, σ_t 表示标准差,且 $\sigma_t = ((T-t)^h / (T^h)) * (\sigma_c - \sigma_z) + \sigma_z$, h 表示非线性调和指数, σ_c 表示标准差起始值, σ_z 表示标准差终止值;

将新模型参数向量与全局搜索之后的模型参数向量进行融合,得到融合种群,并保留融合种群中适应度值最大的 L 个模型参数向量,得到新种群。

一种基于云计算的量化数据处理方法

技术领域

[0001] 本发明属于云计算以及量化数据处理技术领域,具体涉及一种基于云计算的量化数据处理方法。

背景技术

[0002] 量化数据一般是指用数值表示的数据,在各行各业中都存在量化数据。通常,为了提升企业生产的效率,常常会由专门的职能部门对量化数据进行分析(例如,对生产过程中设备运行参数进行采集,以分析生产是否异常;又例如,对海量交易数据进行分类,从而可以使工作人员根据分类类别能够快速进行决策以及策略优化),从而实现数据分类或者产生决策建议,以辅助领导进行生产调度决策,或者实现生产参数优化。

[0003] 但是,采用职能部门对量化数据进行分析,不仅仅存在效率低下的问题,还会因为人员的差异,导致最终的数据分类、分析结果不同。

发明内容

[0004] 本发明提供一种基于云计算的量化数据处理方法,用以解决现有技术中存在的问题。

[0005] 一种基于云计算的量化数据处理方法,包括:

[0006] 通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行预处理,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中;

[0007] 在云服务器中构建量化数据分析模型,并采用多层次搜索算法对量化数据分析模型进行参数优化,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上;

[0008] 调度云服务器上部署的量化数据分析模型对预处理之后的量化数据进行分析处理,得到基于云计算的量化数据处理结果。

[0009] 进一步地,通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行预处理,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中,包括:

[0010] 通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行缺失值处理、异常值处理以及归一化处理,得到初始处理之后的量化数据;

[0011] 将初始处理之后的量化数据统一存储为预设数据结构,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中。

[0012] 进一步地,在云服务器中构建量化数据分析模型,包括:构建深度学习模型,并将构建的深度学习模型作为量化数据分析模型。

[0013] 进一步地,采用多层次搜索算法对量化数据分析模型进行参数优化,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上,包括:

[0014] 初始化量化数据分析模型的模型参数,得到模型参数向量,并重复获取多个模型

参数向量,得到种群;

[0015] 获取历史数据以及历史数据对应的标签数据,并采用历史数据以及历史数据对应的标签数据,获取每个模型参数向量对应的适应度值;

[0016] 针对每个模型参数向量,采用固定视野以及自适应步长的局部搜索方法对模型参数向量进行局部搜索,并根据模型参数向量对应的适应度值对局部搜索过程进行控制,获取局部搜索之后的模型参数向量;

[0017] 针对局部搜索之后的模型参数向量,采用变动视野的全局平衡搜索方法对模型参数向量进行平衡搜索,并根据模型参数向量对应的适应度值对平衡搜索过程进行控制,获取平衡搜索之后的模型参数向量;

[0018] 针对平衡搜索之后的模型参数向量,采用种群信息融合的全局搜索方法对模型参数向量进行全局搜索,并根据模型参数向量对应的适应度值对全局搜索过程进行控制,获取全局搜索之后的模型参数向量;

[0019] 针对全局搜索之后的模型参数向量,重新获取模型参数向量的适应度值,并根据适应度值获取个体扩展数量;

[0020] 根据个体扩展数量,采用正态分布方法产生多个新模型参数向量,并将新模型参数向量与全局搜索之后的模型参数向量进行融合,再采用优胜劣汰的方法抉择出新种群;

[0021] 针对新种群,获取每个模型参数向量对应的适应度值,并根据每个模型参数向量对应的适应度值获取最优模型参数向量;

[0022] 判断是否满足算法终止条件,若是,则输出最优模型参数向量,否则返回局部搜索的步骤;

[0023] 将输出的最优模型参数向量作为量化数据分析模型的最终模型参数,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上。

[0024] 进一步地,针对每个模型参数向量,采用固定视野以及自适应步长的局部搜索方法对模型参数向量进行局部搜索,并根据模型参数向量对应的适应度值对局部搜索过程进行控制,获取局部搜索之后的模型参数向量,包括:

[0025] 针对每个模型参数向量,生成一个随机向量为: $(\Delta x_{i,1}, \Delta x_{i,2}, \dots, \Delta x_{i,d}, \dots, \Delta x_{i,D})$, 其中, $\Delta x_{i,d}$ 表示第*i*个模型参数向量中第*d*维参数对应的随机参数, $i=1, 2, \dots, L$, L 表示模型参数向量的总数, $d=1, 2, \dots, D$, D 表示模型参数向量中参数总维度,且 $\Delta x_{i,d}$ 按各 0.5 的概率随机为 α^t 或者 $-\alpha^t$, α^t 表示第*t*次优化过程中的自适应步长;

[0026] 获取第*i*个模型参数向量的伪梯度为:

$$[0027] \quad F(X_i^t)_d = \frac{f(X_i^t + \Delta X_i^t) - f(X_i^t - \Delta X_i^t)}{2\Delta x_{i,d}}$$

[0028] 其中, $F(X_i^t)_d$ 表示第*i*个模型参数向量中第*d*维参数对应的为梯度, $f(*)$ 表示适应度函数, X_i^t 表示第*t*次优化过程中的第*i*个模型参数向量, ΔX_i^t 表示随机向量;

[0029] 根据预设固定视野 β_0 获取第*t*+1次优化过程中的自适应步长 α^{t+1} , 并根据自适应步长 α^{t+1} 以及第*i*个模型参数向量的伪梯度 $F(X_i^t)_d$, 对第*i*个模型参数向量进行局部搜索为:

$$[0030] \quad x_{i,d}^{t+1} = x_{i,d}^t + \alpha^{t+1} * \text{sign}(F(X_i^t)_d)$$

[0031] 其中, $x_{i,d}^t$ 表示第t次优化过程中的第i个模型参数向量中第d维参数, $x_{i,d}^{t+1}$ 表示 $x_{i,d}^t$ 对应的局部搜索值;

[0032] 判断局部搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该局部搜索,得到局部搜索之后的模型参数向量,否则直接将原有的模型参数向量作为局部搜索之后的模型参数向量。

[0033] 进一步地,根据预设固定视野 β_0 获取第t+1次优化过程中的自适应步长 α^{t+1} 为:

$$[0034] \quad \alpha^{t+1} = \frac{\alpha_0}{20} + \beta_0 * \sin\left(\frac{\pi}{2} * \left(1 - \frac{t-1}{T}\right)\right)$$

[0035] 其中, α_0 表示初始步长, π 表示圆周率, T 表示最大优化次数。

[0036] 进一步地,针对局部搜索之后的模型参数向量,采用变动视野的全局平衡搜索方法对模型参数向量进行平衡搜索,并根据模型参数向量对应的适应度值对平衡搜索过程进行控制,获取平衡搜索之后的模型参数向量,包括:

[0037] 以当前优化次数t为基础,获取变动视野为:

$$[0038] \quad \beta^{t+1} = \frac{1}{1 + e^{2 \ln 100(t/T) * \ln 1000}} * \frac{T-t+1}{T} * \beta^t$$

[0039] 其中, β^t 表示第t次优化过程中的变动视野, β^{t+1} 表示第t+1次优化过程中的变动视野,且变动视野初始值为 β_0 , e 表示自然常数;

[0040] 根据变动视野 β^{t+1} , 获取平衡搜索范围为:

$$[0041] \quad lb_{i,d} = x_{i,d}^{t+1} - \beta^{t+1}$$

$$[0042] \quad ub_{i,d} = x_{i,d}^{t+1} + \beta^{t+1}$$

[0043] 其中, $lb_{i,d}$ 表示第i个模型参数向量中第d维参数的平衡搜索下界, $ub_{i,d}$ 表示第i个模型参数向量中第d维参数的平衡搜索上界, $x_{i,d}^{t+1}$ 表示局部搜索之后的第i个模型参数向量中第d维参数;

[0044] 根据平衡搜索范围,对局部搜索之后的模型参数向量进行平衡搜索为:

$$[0045] \quad x_{i,d}^{t+1}' = lb_{i,d} + \text{rand}(0,1) * (ub_{i,d} - lb_{i,d})$$

[0046] 其中, $\text{rand}(0,1)$ 表示 (0,1) 之间的随机数, $x_{i,d}^{t+1}'$ 表示平衡搜索之后的 $x_{i,d}^{t+1}$;

[0047] 判断平衡搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该平衡搜索,得到平衡搜索之后的模型参数向量,否则直接将局部搜索之后的模型参数向量作为平衡搜索之后的模型参数向量。

[0048] 进一步地,针对平衡搜索之后的模型参数向量,采用种群信息融合的全局搜索方法对模型参数向量进行全局搜索,并根据模型参数向量对应的适应度值对全局搜索过程进行控制,获取全局搜索之后的模型参数向量,包括:

[0049] 根据所有平衡搜索之后的模型参数向量,获取种群的平均位置为 \bar{X} ; \bar{X} 中每一维参数为其他所有模型参数向量在同维参数的平均值;

[0050] 根据种群的平均位置 \bar{X} , 对平衡搜索之后的模型参数向量进行全局搜索为:

$$[0051] \quad x_{i,d}^{t+1} = x_{i,d}^{t+1} + \text{rand}(x_{\min}, x_{\max}) * (\overline{x_d} - x_{i,d}^{t+1})$$

[0052] 其中, $x_{i,d}^{t+1}$ 表示全局搜索之后的模型参数向量, $\text{rand}(x_{\min}, x_{\max})$ 表示 (x_{\min}, x_{\max}) 之间的随机数, x_{\min} 表示全局搜索区间下限, x_{\max} 表示全局搜索区间上限, $\overline{x_d}$ 表示平均位置 \overline{X} 中第d维参数;

[0053] 判断全局搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该全局搜索,得到全局搜索之后的模型参数向量,否则直接将平衡搜索之后的模型参数向量作为全局搜索之后的模型参数向量。

[0054] 进一步地,针对全局搜索之后的模型参数向量,重新获取模型参数向量的适应度值,并根据适应度值获取个体扩展数量为:

$$[0055] \quad \gamma_i = \xi \left(\frac{f_i - f_{\min}}{f_{\max} - f_{\min}} * (S_{\max} - S_{\min}) + S_{\min} \right)$$

[0056] 其中, γ_i 表示第i个全局搜索之后的模型参数向量对应的个体扩展数量, $\xi(*)$ 表示取整函数, f_i 表示第i个全局搜索之后的模型参数向量对应的适应度值, f_{\max} 表示全局搜索之后的模型参数向量对应的最大适应度值, f_{\min} 表示全局搜索之后的模型参数向量对应的最小适应度值, S_{\max} 表示预设的最大扩展整数, S_{\min} 表示预设的最小扩展整数。

[0057] 进一步地,根据个体扩展数量,采用正态分布方法产生多个新模型参数向量,并将新模型参数向量与全局搜索之后的模型参数向量进行融合,再采用优胜劣汰的方法抉择出新种群,包括:

[0058] 针对每个全局搜索之后的模型参数向量,以正态分布 $N(0, \sigma_t^2)$ 在模型参数向量周围产生与个体扩展数量相等的多个新模型参数向量;

[0059] 其中, σ_t 表示标准差,且 $\sigma_t = ((T-t)^h / (T^h)) * (\sigma_c - \sigma_z) + \sigma_z$, h 表示非线性调和指数, σ_c 表示标准差起始值, σ_z 表示标准差终止值;

[0060] 将新模型参数向量与全局搜索之后的模型参数向量进行融合,得到融合种群,并保留融合种群中适应度值最大的L个模型参数向量,得到新种群。

[0061] 本发明提供了一种基于云计算的量化数据处理方法,通过采用深度学习算法以及采用多层次搜索算法对量化数据分析模型进行参数优化,从而实现对历史数据的学习,形成可以处理量化数据的规则,再通过该规则对量化数据进行处理,从而实现量化数据的自动化处理,不仅能够有效地提高量化数据的分析效率,还能够消除人为分析带来的误差。

附图说明

[0062] 此处的附图被并入说明书中并构成本说明书的一部分,示出了符合本发明的实施例,并与说明书一起用于解释本发明的原理。

[0063] 图1为本发明实施例提供一种基于云计算的量化数据处理方法的流程图。

[0064] 通过上述附图,已示出本发明明确的实施例,后文中将有更详细的描述。这些附图和文字描述并不是为了通过任何方式限制本发明构思的范围,而是通过参考特定实施例为本领域技术人员说明本发明的概念。

具体实施方式

[0065] 这里将详细地对示例性实施例进行说明,其示例表示在附图中。下面的描述涉及附图时,除非另有表示,不同附图中的相同数字表示相同或相似的要素。以下示例性实施例中所描述的实施方式并不代表与本发明相一致的所有实施方式。相反,它们仅是与如所附权利要求书中所详述的、本发明的一些方面相一致的装置和方法的例子。

[0066] 下面结合附图详细说明本发明的实施例。

[0067] 如图1所示,一种基于云计算的量化数据处理方法,包括:

[0068] S1、通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行预处理,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中。

[0069] 量化数据可以为数字数据(如交易数据、设备运行数据等),也可以为图像数据等等。目标数据源用于表征待分析的数据源,通过固定的采样频率,可以对某个时间段内的数据进行采集,并自动化的分析出对应的分类结果。例如:对图像数据进行分析,可以实现零件分类、产品表面缺陷识别(如玻璃、车漆面的缺陷识别)等等分类结果;或者,对某条生产线的设备运行参数进行分析,从而可以实现设备是否正常运行的分类;又或者,对交易数据进行分析,从而获取不同的预设决策分析,基于云计算的量化交易数据处理能大幅提高数据处理速度和分析效率,使投资者能够在关键时刻更快地做出交易决策。

[0070] S2、在云服务器中构建量化数据分析模型,并采用多层次搜索算法对量化数据分析模型进行参数优化,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上。

[0071] 云计算具有强大的计算性能,通过在云服务器中构建量化数据分析模型,能够快速地对量化数据进行分析,从而提高数据分析处理效率。虽然云计算具有强大的计算性能,但是在部署量化数据分析模型的过程中,需要耗费大量的资源进行模型优化,且优化效果不佳的话,会导致后续计算机资源以及算力的浪费,因此,本发明实施例提出了多层次搜索算法,能够有效地提升优化速度以及优化效果,从而有效地节约了计算机资源。

[0072] S3、调度云服务器上部署的量化数据分析模型对预处理之后的量化数据进行分析处理,得到基于云计算的量化数据处理结果。

[0073] 优化之后的量化数据分析模型具备对量化数据处理分析的能力,因此可以调度优化之后的量化数据分析模型进行数据自动处理,从而提高数据处理效率。

[0074] 可选的,为了提高识别准确率(如图像分类准确率),还可以根据数据对量化数据分析模型的参数进行更新,从而进一步地优化模型。

[0075] 本发明提供了一种基于云计算的量化数据处理方法,通过采用深度学习算法以及采用多层次搜索算法对量化数据分析模型进行参数优化,从而实现对历史数据的学习,形成可以处理量化数据的规则,再通过该规则对量化数据进行处理,从而实现量化数据的自动化处理,不仅能够有效地提高量化数据的分析效率,还能够消除人为分析带来的误差。

[0076] 在本实施例中,通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行预处理,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中,包括:

[0077] 通过网络爬虫爬取目标数据源中的量化数据,对爬取的量化数据进行缺失值处理、异常值处理以及归一化处理,得到初始处理之后的量化数据;

[0078] 将初始化之后的量化数据统一存储为预设数据结构,得到预处理之后的量化数据,并将预处理之后的量化数据存储于云服务器中。

[0079] 对量化数据进行归一化处理,保证了数据的识别速率,从而节约了计算机处理资源。值得说明的是,不论是用于优化量化数据分析模型的量化数据,都应当处理为统一数据长度以及统一数据结构,从而保证量化数据分析模型能够正常识别量化数据。例如:对图像数据进行分类时,需要处理为统一尺寸。

[0080] 在本实施例中,在云服务器中构建量化数据分析模型,包括:构建深度学习模型,并将构建的深度学习模型作为量化数据分析模型。

[0081] 可选的,当量化数据为数值数据时,可以将量化数据组成向量,并采用BP (Back Propagation,反向传播)神经网络、BRF (Radial Basis Function,径向基函数)神经网络等等深度学习模型作为量化数据分析模型,从而实现向向量数据的识别。也可以将量化数据按照固定格式组装为矩阵数据,采用卷积神经网络作为量化数据分析模型,从而实现向向量数据的识别。值得说明的是,当量化数据为图像时,相当于限定了数据输入形式,需要采用能够识别图像数据的深度学习模型构建量化数据分析模型。

[0082] 在本实施例中,采用多层次搜索算法对量化数据分析模型进行参数优化,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上,包括:

[0083] 初始化量化数据分析模型的模型参数,得到模型参数向量,并重复获取多个模型参数向量,得到种群。

[0084] 例如,在量化数据分析模型的模型参数上限与模型参数下限之间,随机生成模型参数,从而可以得到模型参数向量。

[0085] 获取历史数据以及历史数据对应的标签数据,并采用历史数据以及历史数据对应的标签数据,获取每个模型参数向量对应的适应度值。

[0086] 将历史数据作为输入,将历史数据对应的标签数据作为期望输出,可以获取量化数据分析模型的误差函数值,从而可以将误差函数值的负数作为适应度值。按照此方法,逐一或者并行获取每个模型参数向量对应的适应度值即可。

[0087] 针对每个模型参数向量,采用固定视野以及自适应步长的局部搜索方法对模型参数向量进行局部搜索,并根据模型参数向量对应的适应度值对局部搜索过程进行控制,获取局部搜索之后的模型参数向量。

[0088] 针对局部搜索之后的模型参数向量,采用变动视野的全局平衡搜索方法对模型参数向量进行平衡搜索,并根据模型参数向量对应的适应度值对平衡搜索过程进行控制,获取平衡搜索之后的模型参数向量。

[0089] 针对平衡搜索之后的模型参数向量,采用种群信息融合的全局搜索方法对模型参数向量进行全局搜索,并根据模型参数向量对应的适应度值对全局搜索过程进行控制,获取全局搜索之后的模型参数向量。

[0090] 针对全局搜索之后的模型参数向量,重新获取模型参数向量的适应度值,并根据适应度值获取个体扩展数量。

[0091] 根据个体扩展数量,采用正态分布方法产生多个新模型参数向量,并将新模型参数向量与全局搜索之后的模型参数向量进行融合,再采用优胜劣汰的方法抉择出新种群。

[0092] 针对新种群,获取每个模型参数向量对应的适应度值,并根据每个模型参数向量

对应的适应度值获取最优模型参数向量。

[0093] 判断是否满足算法终止条件,若是,则输出最优模型参数向量,否则返回局部搜索的步骤。

[0094] 可选的,判断是否满足算法终止条件可以包括:判断当前优化次数是否大于最大优化次数,若是,则满足算法终止条件,否则不满足算法终止条件。

[0095] 将输出的最优模型参数向量作为量化数据分析模型的最终模型参数,得到优化之后的量化数据分析模型,并将优化之后的量化数据分析模型部署于云服务器上。

[0096] 在现有技术中,常常采用梯度下降法、粒子群算法等算法对深度学习模型进行训练,不仅仅训练速度慢,还存在训练效果差的问题,从而导致计算机资源大量浪费。因此,本发明实施例提出了多层次搜索算法对量化数据分析模型进行参数优化,以解决现有技术中存在的问题,提升量化数据的分类能力。

[0097] 在本实施例中,针对每个模型参数向量,采用固定视野以及自适应步长的局部搜索方法对模型参数向量进行局部搜索,并根据模型参数向量对应的适应度值对局部搜索过程进行控制,获取局部搜索之后的模型参数向量,包括:

[0098] 针对每个模型参数向量,生成一个随机向量为: $(\Delta x_{i,1}, \Delta x_{i,2}, \dots, \Delta x_{i,d}, \dots, \Delta x_{i,D})$,其中, $\Delta x_{i,d}$ 表示第*i*个模型参数向量中第*d*维参数对应的随机参数, $i=1,2,\dots,L$,*L*表示模型参数向量的总数, $d=1,2,\dots,D$,*D*表示模型参数向量中参数总维度,且 $\Delta x_{i,d}$ 按各0.5的概率随机为 α^t 或者 $-\alpha^t$, α^t 表示第*t*次优化过程中的自适应步长;

[0099] 获取第*i*个模型参数向量的伪梯度为:

$$[0100] \quad F(X_i^t)_d = \frac{f(X_i^t + \Delta X_i^t) - f(X_i^t - \Delta X_i^t)}{2\Delta x_{i,d}}$$

[0101] 其中, $F(X_i^t)_d$ 表示第*i*个模型参数向量中第*d*维参数对应的为梯度, $f(*)$ 表示适应度函数, X_i^t 表示第*t*次优化过程中的第*i*个模型参数向量, ΔX_i^t 表示随机向量;

[0102] 根据预设固定视野 β_0 获取第*t*+1次优化过程中的自适应步长 α^{t+1} , 并根据自适应步长 α^{t+1} 以及第*i*个模型参数向量的伪梯度 $F(X_i^t)_d$, 对第*i*个模型参数向量进行局部搜索为:

$$[0103] \quad x_{i,d}^{t+1} = x_{i,d}^t + \alpha^{t+1} * \text{sign}(F(X_i^t)_d)$$

[0104] 其中, $x_{i,d}^t$ 表示第*t*次优化过程中的第*i*个模型参数向量中第*d*维参数, $x_{i,d}^{t+1}$ 表示 $x_{i,d}^t$ 对应的局部搜索值;

[0105] 通过自适应步长进行搜索,不仅能够不断改变搜索方向,还能够不断改变搜索步长,从而实现局部范围内的精细搜索,更有助于寻找到最优值。

[0106] 判断局部搜索之后的第*i*个模型参数向量对应的适应度是否增大,若是,则接受该局部搜索,得到局部搜索之后的模型参数向量,否则直接将原有的模型参数向量作为局部搜索之后的模型参数向量。

[0107] 局部搜索往往不影响算法的全局性,因此本发明实施例引入贪心算法对局部搜索进行贪心选择,从而提升算法的收敛速度。

[0108] 在本实施例中,根据预设固定视野 β_0 获取第*t*+1次优化过程中的自适应步长 α^{t+1} 为:

$$[0109] \quad \alpha^{t+1} = \frac{\alpha_0}{20} + \beta_0 * \sin\left(\frac{\pi}{2} * \left(1 - \frac{t-1}{T}\right)\right)$$

[0110] 其中, α_0 表示初始步长, π 表示圆周率, T 表示最大优化次数。

[0111] 在本实施例中, 针对局部搜索之后的模型参数向量, 采用变动视野的全局平衡搜索方法对模型参数向量进行平衡搜索, 并根据模型参数向量对应的适应度值对平衡搜索过程进行控制, 获取平衡搜索之后的模型参数向量, 包括:

[0112] 以当前优化次数 t 为基础, 获取变动视野为:

$$[0113] \quad \beta^{t+1} = \frac{1}{1 + e^{2 \ln 100(t/T) * \ln 1000}} * \frac{T-t+1}{T} * \beta^t$$

[0114] 其中, β^t 表示第 t 次优化过程中的变动视野, β^{t+1} 表示第 $t+1$ 次优化过程中的变动视野, 且变动视野初始值为 β_0 , e 表示自然常数;

[0115] 根据变动视野 β^{t+1} , 获取平衡搜索范围为:

$$[0116] \quad lb_{i,d} = x_{i,d}^{t+1} - \beta^{t+1}$$

$$[0117] \quad ub_{i,d} = x_{i,d}^{t+1} + \beta^{t+1}$$

[0118] 其中, $lb_{i,d}$ 表示第 i 个模型参数向量中第 d 维参数的平衡搜索下界, $ub_{i,d}$ 表示第 i 个模型参数向量中第 d 维参数的平衡搜索上界, $x_{i,d}^{t+1}$ 表示局部搜索之后的第 i 个模型参数向量中第 d 维参数;

[0119] 根据平衡搜索范围, 对局部搜索之后的模型参数向量进行平衡搜索为:

$$[0120] \quad x_{i,d}^{t+1}' = lb_{i,d} + \text{rand}(0,1) * (ub_{i,d} - lb_{i,d})$$

[0121] 其中, $\text{rand}(0,1)$ 表示 $(0,1)$ 之间的随机数, $x_{i,d}^{t+1}'$ 表示平衡搜索之后的 $x_{i,d}^{t+1}$;

[0122] 判断平衡搜索之后的第 i 个模型参数向量对应的适应度是否增大, 若是, 则接受该平衡搜索, 得到平衡搜索之后的模型参数向量, 否则直接将局部搜索之后的模型参数向量作为平衡搜索之后的模型参数向量。

[0123] 通过设置变动视野 β^{t+1} 进行平衡搜索, 可以在算法前期拥有一个比较大的搜索范围, 从而提升算法的全局搜索能力, 随着算法的进行, 变动视野 β^{t+1} 逐渐缩小, 从而逐渐缩小搜索范围, 从而使算法拥有局部精细搜索的能力。并且平衡搜索的随机性更强, 能够实现局部区域的全面探索。

[0124] 在本实施例中, 针对平衡搜索之后的模型参数向量, 采用种群信息融合的全局搜索方法对模型参数向量进行全局搜索, 并根据模型参数向量对应的适应度值对全局搜索过程进行控制, 获取全局搜索之后的模型参数向量, 包括:

[0125] 根据所有平衡搜索之后的模型参数向量, 获取种群的平均位置为 \bar{X} ; \bar{X} 中每一维参数为其他所有模型参数向量在同维参数的平均值;

[0126] 根据种群的平均位置 \bar{X} , 对平衡搜索之后的模型参数向量进行全局搜索为:

$$[0127] \quad x_{i,d}^{t+1}'' = x_{i,d}^{t+1}' + \text{rand}(x_{\min}, x_{\max}) * (\bar{x}_d - x_{i,d}^{t+1}')$$

[0128] 其中, $x_{i,d}^{t+1}''$ 表示全局搜索之后的模型参数向量, $\text{rand}(x_{\min}, x_{\max})$ 表示 (x_{\min}, x_{\max}) 之

间的随机数, x_{\min} 表示全局搜索区间下限, x_{\max} 表示全局搜索区间上限, \bar{x}_d 表示平均位置 \bar{X} 中第d维参数;

[0129] 判断全局搜索之后的第i个模型参数向量对应的适应度是否增大,若是,则接受该全局搜索,得到全局搜索之后的模型参数向量,否则直接将平衡搜索之后的模型参数向量作为全局搜索之后的模型参数向量。

[0130] 虽然平衡搜索具备一定的全局搜索作用,但是随着算法地进行衰减,因此本发明实施例引入全局搜索,能够使算法始终具备全局搜索能力,从而提升收敛效果。

[0131] 可选的,在全局搜索时,当全局搜索之后的第i个模型参数向量对应的适应度未增大时,也可以以预设的概率接收全局搜索,从而避免算法停滞。

[0132] 在本实施例中,针对全局搜索之后的模型参数向量,重新获取模型参数向量的适应度值,并根据适应度值获取个体扩展数量为:

$$[0133] \quad \gamma_i = \xi \left(\frac{f_i - f_{\min}}{f_{\max} - f_{\min}} * (S_{\max} - S_{\min}) + S_{\min} \right)$$

[0134] 其中, γ_i 表示第i个全局搜索之后的模型参数向量对应的个体扩展数量, $\xi(*)$ 表示取整函数, f_i 表示第i个全局搜索之后的模型参数向量对应的适应度值, f_{\max} 表示全局搜索之后的模型参数向量对应的最大适应度值, f_{\min} 表示全局搜索之后的模型参数向量对应的最小适应度值, S_{\max} 表示预设的最大扩展整数, S_{\min} 表示预设的最小扩展整数。

[0135] 由于高维数学问题难以求解,因此本发明实施例提出了一种种群扩展策略,通过在模型参数向量附近的固定区域内生成多个均匀分布的新个体,并采用优胜劣汰的方法进行抉择,能够有效地提升收敛速度以及局部搜索精度。

[0136] 在本实施例中,根据个体扩展数量,采用正态分布方法产生多个新模型参数向量,并将新模型参数向量与全局搜索之后的模型参数向量进行融合,再采用优胜劣汰的方法抉择出新种群,包括:

[0137] 针对每个全局搜索之后的模型参数向量,以正态分布 $N(0, \sigma_t^2)$ 在模型参数向量周围产生与个体扩展数量相等的多个新模型参数向量;

[0138] 其中, σ_t 表示标准差,且 $\sigma_t = ((T-t)^h / (T^h)) * (\sigma_c - \sigma_z) + \sigma_z$, h 表示非线性调和指数, σ_c 表示标准差起始值, σ_z 表示标准差终止值;

[0139] 将新模型参数向量与全局搜索之后的模型参数向量进行融合,得到融合种群,并保留融合种群中适应度值最大的L个模型参数向量,得到新种群。

[0140] 本领域内的技术人员应明白,本发明的实施例可提供为方法、系统、或计算机程序产品。因此,本发明可采用完全硬件实施例、完全软件实施例、或结合软件和硬件方面的实施例的形式。而且,本发明可采用在一个或多个其中包含有计算机可用程序代码的计算机可用存储介质(包括但不限于磁盘存储器、CD-ROM、光学存储器等)上实施的计算机程序产品的形式。

[0141] 本发明是参照根据本发明实施例的方法、设备(系统)、和计算机程序产品的流程图和/或方框图来描述的。应理解可由计算机程序指令实现流程图和/或方框图中的每一流程和/或方框、以及流程图和/或方框图中的流程和/或方框的结合。可提供这些计算机程序指令到通用计算机、专用计算机、嵌入式处理机或其他可编程数据处理设备的处理器以产

生一个机器,使得通过计算机或其他可编程数据处理设备的处理器执行的指令产生用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的装置。

[0142] 这些计算机程序指令也可存储在能引导计算机或其他可编程数据处理设备以特定方式工作的计算机可读存储器中,使得存储在该计算机可读存储器中的指令产生包括指令装置的制造品,该指令装置实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能。

[0143] 这些计算机程序指令也可装载到计算机或其他可编程数据处理设备上,使得在计算机或其他可编程设备上执行一系列操作步骤以产生计算机实现的处理,从而在计算机或其他可编程设备上执行的指令提供用于实现在流程图一个流程或多个流程和/或方框图一个方框或多个方框中指定的功能的步骤。

[0144] 本领域普通技术人员可以理解实现上述事实和方法中的全部或部分步骤是可以通程序来指令相关的硬件来完成,涉及的程序或者所述的程序可以存储于一计算机所可读存储介质中,该程序在执行时,包括如下步骤:此时引出相应的方法步骤,所述的存储介质可以是ROM/RAM、磁碟、光盘等等。

[0145] 以上所述的具体实施方式,对本发明的目的、技术方案和有益效果进行了进一步详细说明,所应理解的是,以上所述仅为本发明的具体实施方式而已,并不用于限定本发明的保护范围,凡在本发明的精神和原则之内,所做的任何修改、等同替换、改进等,均应包含在本发明的保护范围之内。

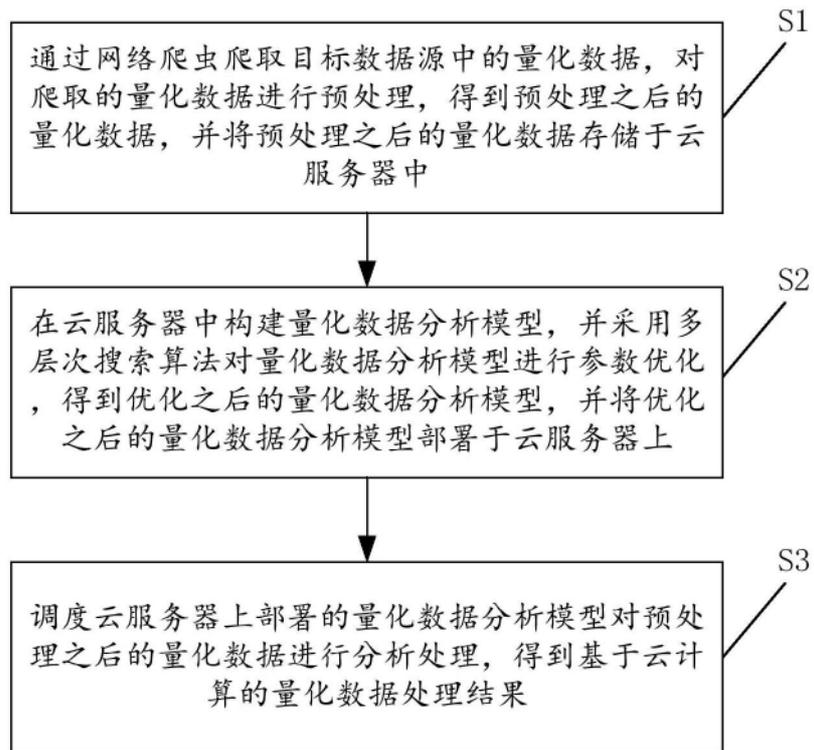


图1