**A new quantum chaotic cuckoo search algorithm for data clustering**

**文献概述：**

本文提出了一种新的用于数据聚类的量子混合布谷鸟搜索算法（QCCS）。布谷鸟搜索（CS）在聚类问题上优于传统的元启发式算法。但所有的布谷鸟都有相同的搜索行为，这可能导致算法收敛到局部最优，此外，收敛速度对随机生成的初始质心种子很敏感。

因此，本文使用受量子理论启发的非均匀更新来扩展CS能力，以解决布谷鸟搜索聚类问题的全局搜索能力。此外，还定制了有效的策略来管理边界。

**相关介绍：**

1. means是一种非常简单，让人容易理解的分区聚类算法，但该算法存在随机初始化问题，有时会陷入局部解。为了克服这一缺点，基于元启发式的算法已经证明了它们对NP-hard问题。这些算法能够在合理的时间内达到此类问题的最佳或接近最佳的解决方案。为此，我们可以将聚类视为一类优化问题。对此我们称适应度函数的目标函数最大化或最小化。许多研究者可以证明，元启发式算法能够找到聚类问题的全局解决方案，而不是局部解决方案。 元启发式算法之一是布谷鸟搜索优化算法。它基于布谷鸟的繁殖行为，如雏鸟的寄生和Levy飞行的典型特征。

通过CS与三种广泛使用的优化算法的比较和证明分析来看，CS在统计上比微分进化（DE）、粒子群优化（PSO）和人工蜂群（ABC）更成功。但CS的主要缺点就是，所有的布谷鸟有相同的搜索行为，易陷入局部最优解。另外，为了找到局部最优解，迭代次数也会增加。

过去的十年里，有两种理论在不同的研究领域取得了巨大的成功应用：量子理论和混沌理论。量子理论为进化计算提供了新的概念，为提高典型优化算法的效率和速度，它们被归为两个不同的类别：

第一类：结合遗传量子计算算子的原理，如量子比特表示法、状态的叠加、干扰、测量和突变等提出的不同的元启发式算法来解决组合优化问题。

第二类：采用描述量子系统行为的薛定谔方程，以增强原始元启发式算法的搜索程序，包括量子布谷鸟搜索（QCS）和量子粒子群优化（QPSO）。

**相关工作：**

Sun, Xu, and Ye探索了量子化粒子群优化对数据聚类的适用性。他们的实验结果证明了两个结论：

一是量子化粒子群优化（OPSO）的整体性能优于K-means和PSO聚类算法。

二是OPSO和K-means方法的混合可以极大地提高OPSO聚类算法。

同年，Paterlini和Krink报告了GA、PSO和DE对一个聚类问题地性能比较结果。结果显示，DE一直处于优势地位。

Ishak Boushaki等人，提出了一种基于布谷鸟搜索优化算法地数据聚类地新算法。在四个不同地数据集上评估了所提出地算法的性能，并于众所周知的K-means、PSO、引力搜索算法（GSA）、大爆炸-大紧缩算法（BB-BC）和黑洞算法（BH）。实验证明了新方法在几乎所有数据集上都能达到最佳值。

对于所有经典的元启发式优化算法，它们都会为了找到最佳解决方案使得迭代次数增加。此外，几乎所有的元启发式方法都有两个主要的缺点：

一：它们设计许多难以调整的参数（个人学习系数，突变概率等）。

二：由于搜索空间大，它们通常需要进行大量迭代，这使得它们在计算量上令人忘而却步。

因此这些方法大多数通过与K-means算法或其它替代算法混合使用来扩展原始算法的能力。然而，混合算法复杂度高，需要更多的计算量。与推广某种元启发式方法相比，其它方法采用混合和量子理论等最新理论启发的新概念，以简单的方式增强其自身的性能。

本文选择第二种解决方案来扩展标准的CS的能力，主要是量子理论和混合映射。

**混合映射和量子布谷鸟搜索算法：**

1. 混合映射（chaotic map）

大多数元启发式算法几乎不依赖于初始种子，因此，如果初始解决方案没有很好地进行选择，会造成局部最小值和较差的结果。

混沌映射是一种数学函数，随着时间的推移表现出某种混沌行为。混沌函数的不同取值构成混沌序列。在混沌序列中，n+1时刻的混沌值(Cn+1)只取决于时间n的混沌值（Cn）。著名的Logistic图由以下公式给出：



在本文中，Cn∈（0，1），a = 4，这几个是大多数研究工作中所建议的取值。

1. 布谷鸟搜索算法（cuckoo search）和莱维飞行（levy flight）

布谷鸟搜索算法是最新的自然启发式搜索算法之一，其灵感来自鸟类的两种特殊行为：布谷鸟的寄生行为和莱维飞行。它集中在三个理想化的规则上。

1. 每只布谷鸟一次只产一个蛋，并将蛋随机扔在某个宿主鸟巢中。
2. 具有最佳蛋的宿主鸟巢将传给下一代。
3. 宿主鸟巢的数量是固定的，布谷鸟下的蛋被发现的概率pa∈[0,1]。在被发现的情况下，宿主鸟要么将蛋扔掉，要么放弃自己的巢穴，建立一个全新的巢穴。为简单起见，后一个假设可以用被新的随机鸟巢续贷的较差鸟巢的种群规模的一部分来模拟。（For simplicity, this last assumption can be simulated by the fraction (pa) of the population size of worse nests that are replaced by new random nests.）

布谷鸟使用从莱维飞行分布中得出的随机步长从当前巢穴移动到新巢穴。随机游走的是一条马尔科夫链，其下一个位置仅取决于当前位置和转移概率。对于标准布谷鸟搜索来说，新巢（Xi+1）根据当前巢（Xi）和全局最佳巢（Xb）确定，确定的公式如下：

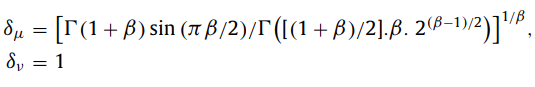


此处α>0，表示的是步长的比例因子。S是一个基于莱维飞行的随机行走距离。S使用Mantegna算法进行计算，公式如下：



其中β∈[1,2]中的固定参数，一般取为3/2，μ和v从正态分布中得出。





CS伪代码如下：

开始：

目标函数f(x), x = (x1,x2,...,xd)’

生成n个宿主巢穴。Xi(i=1,2,...,n);

While(t < MaxGeneration) or (stop criterion){

1. 通过莱维飞行获取一直布谷鸟；
2. 评估其自适应F (i)
3. 在n个鸟巢中随机选择一个巢穴(X j)进行寄生；

If( F (i) > F (j)){

替换鸟巢j，作为一个新的解决方案。

}else{

1. Pa的概率让该巢穴被遗弃并建造新的巢穴
2. 保留最优质的巢穴
3. 对巢穴进行优质度的排序，找到全局最优

}

}

结束：

数据结果的处理。

3. 量子布谷鸟搜索算法

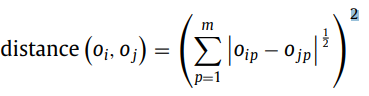
针对标准的CS算法的实现可以看出，所有的布谷鸟都有相同的搜索行为。因此种群中所有个体都有相同的机制来进行更新最优解。算法的更新是基于全局最优解和当前局部最优解。因此该算法任意陷入局部最优，导致过早的收敛。

在量子布谷鸟搜索（QCS）中，非同性质的更新是通过三个具有相同概率的更新中随机选择一个进行的。

粒子位置与时间的关系由薛定谔波动方程决定。

1. The global overview of QCCS(QCCS算法概要)

聚类算法为了确定相似簇的对象，通常使用它们之间的相似性度量。通常使用欧几里得距离。



m是属性的数量，Ojp是对象i的属性p的值。确定大量可能分区之间的最佳分区可以让我们将聚类转化为优化问题。本文使用欧氏距离测量的簇内总和（SSE）。SSE越小，说明簇内之间越紧密。

与绝大多数优化算法一样，首先初始化鸟巢，重复执行一组指令，直到满足某些条件，最后显示全局最佳鸟巢。在整个过程中，最先初始化的种群是通过混的序列生成的，而不是随机序列生成的。而新的最优解的更新是由QCS中提出的非齐次量子更新执行。