

基于人工代谢算法的 TSP 问题求解分析

胡 扬 桂卫华 蔡自兴

(中南大学信息科学与工程学院 长沙 410083)

摘 要 通过分析生物体新陈代谢的生理机能,建立人工代谢算法模型。通过分析底物和生成物之间的浓度差建立多步催化反应动力学模型。通过对城市网络和代谢网络进行类比,建立基于浓度差的 TSP 问题寻优模型。实例推导表明,人工代谢算法能有效地实现 TSP 问题的寻优规划。

关键词 人工代谢算法, TSP 问题, 代谢算子, 酶

中图分类号 TP182 文献标识码 A

Analysis for TSP Problem Based on Artificial Metabolic Algorithm

HU Yang GUI Wei-hua CAI Zi-xing

(School of Information Science and Engineering, Central South University, Changsha 410083, China)

Abstract Artificial metabolic algorithm(AMA) model was built based on physiological metabolic functions in organism. Multi-step catalytic reaction kinetics model was obtained by analysis concentration difference between substrate and product. Searching optimization model for traveling salesman problem(TSP) based on concentration difference was established by using analogous analysis between city network and metabolic network. One real example states that optimization programming for traveling salesman problem(TSP) can be implemented efficiently through artificial metabolic algorithm.

Keywords Artificial metabolic algorithm(AMA), Traveling salesman problem(TSP), Metabolic operator, Enzyme

1 引言

TSP 问题(旅行商问题)一直以来都作为验证人工智能算法正确性的重要问题。其一般表述是:假设有 n 个城市,分别设为 $C_1, C_2, \dots, C_n$ 。现有一旅客希望从城市 C_1 出发,依次拜访这 n 个城市,最后回到城市 C_1 。希望所经过的路程为最短。其中这 n 个城市之间的相互距离为已知^[1]。

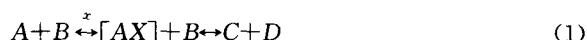
针对 TSP 问题,人工智能领域不断涌现出新的寻优解法。遗传算法对所经过的城市序号进行编码来组成 DNA 序列。利用遗传算子对 DNA 序列进行相关操作。以最短路径为适应度实现 TSP 问题的寻优^[2]。免疫算法通过将城市序列看作抗体,抗原则对应目标函数,通过不断调整抗体使之与抗原的亲密度达到最大来解决 TSP 问题^[3,4]。蚁群算法通过对多只蚂蚁气味浓度的辨识来实现并行搜索,识别最短路径,从而解决 TSP 问题^[5]。

本文提出的人工代谢算法是一种基于生化反应浓度驱动的仿生算法。它利用生化反应中底物和生成物浓度平衡调节原则,将两两相连的路径分别视作底物和生成物,通过寻找代谢平衡路径来实现代谢方程的浓度平衡^[6]。当各个代谢反应均实现平衡时,中间代谢物量达到最大值。此时对应的代谢通路即为所求 TSP 问题的优化路径。

2 人工代谢算法概述

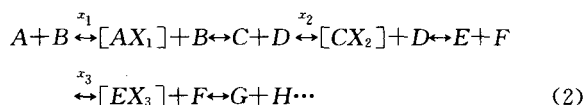
2.1 生化级联反应

生物体内部时时刻刻都进行着大量的生物化学反应。生物体的新陈代谢功能保证着生物体的正常运转。而这些反应都是通过酶来催化进行的。其中最基本也是最核心的生化反应方程式如式(1)所示。



式中, A, B 是反应过程的底物, C, D 是反应过程的生成物, X 是催化酶, AX 称为反应的中间代谢物。上述反应过程为可逆反应,用双向箭头表示。当反应达到平衡时,所生成的中间代谢物的量达到最大。此时酶的催化能力达到最高。

一般而言,生物体内还进行着另一类也是较为常见的生化反应:级联生化反应。在这类反应中,前一个反应过程生成的生成物将作为下一个生化反应过程的底物。以此类推,各生化反应构成连锁的反应链。当实现反应平衡时,最终反应过程的中间代谢物的量达到最大值。最优代谢路径形成。



如式(2)所示,在第一个反应中,在酶 x_1 的催化下,生成

收稿日期:2009-07-27 返修日期:2009-10-14 本文受国家杰出青年科学基金项目(60425310)资助。

胡 扬(1978—),男,讲师,主要研究方向为人工智能、离散事件动态系统等,E-mail:foxgoat@yahoo.com.cn;桂卫华(1950—),男,博士生导师,主要研究方向为工业过程集成建模与优化控制等;蔡自兴(1938—),男,博士生导师,主要研究方向为人工智能与机器人控制等。

中间代谢物 AX_1 , 同时第一个反应的生成产物为 C, D ; 同时 C, D 作为底物在酶 x_2 的催化下又参与第二个反应, 生成中间代谢物 CX_2 , 第二个反应的生成产物为 E, F ; 同时 E, F 作为底物在酶 x_3 的催化下又参与第三个反应, 此时生成的产物为 G, H 。以此类推, 直至反应完成^[7]。

2.2 人工代谢算法的编码

对于如式(1)所示的单步反应而言, 人工代谢算法的编码规则如下: 底物 A 和 B 的浓度在 $[0, 1]$ 之间, 底物 A 和 B 的量各占整个反应平衡方程中全部物理量的比例可计算得出, 即为其浓度, 用二进制编码表示。假设采用 n 位二进制代码的编码方式, 即将实际浓度映射成 $[0, 2^n - 1]$ 上的 $0, 1$ 组成的编码串。酶的浓度也用 n 位二进制代码的编码方式表达, 也即为 $[0, 2^n - 1]$ 上的 $0, 1$ 组成的编码串。酶的浓度的初始值为随机给出。根据酶与底物的匹配情况, 通过代谢因子对酶的编码进行逻辑关系的调节, 使酶与当前底物更好地匹配。匹配的原则是 A 的编码和 X 的编码进行逻辑同或的结果所对应的数位上的二进制数应尽可能逼近 1。其中平衡算子 ke 的作用是随机选择酶 X 中的二进制数位, 使对应的数位上的数值与底物对应数位上的数值相等, 即: 若底物某一位上的数值是 0, 平衡算子将酶 X 中与底物对应数位上的数值也转化为 0; 若底物某一位上的数值是 1, 平衡算子将酶 X 中与底物对应数位上的数值也转化为 1。抑制算子 ki 的作用是随机选择酶 X 中的二进制数位, 使对应的数位上的数值与底物对应数位上的数值不相等。即: 若底物某一位上的数值是 0 时, 抑制算子则将酶 X 中与底物对应数位上的数值转化为 1; 若底物某一位上的数值是 1 时, 抑制算子则将酶 X 中与底物对应数位上的数值转化为 0。这种编码的思想是使酶 X 与底物之间的作用关系更好地与真实的“共价催化”原理相一致, 使得通过调整酶 X 的变化达到调整化学平衡成为可能。平衡算子 ke 和抑制算子 ki 的值在 $[0, 1]$ 之间。平衡算子 ke 在一个计算周期中的值为当前全部底物的浓度; 抑制算子 ki 在一个计算周期中的值为全部生成物的浓度。在得到平衡算子 ke 和抑制算子 ki 的值后, 决定所要变换的酶的二进制编码的位数。当代谢反应达到稳态, 中间代谢物浓度变化率趋于不变。此时中间代谢物的量已达到稳态最大值, 代谢寻优终止。

3 人工代谢算法的收敛性分析

3.1 模式定理

(1) 基础定义

从以上的分析可以看出: 人工代谢算法的操作过程较为简单, 它通过对酶的编码不断执行操作, 再使酶和底物的编码进行同或运算。实际上, 在算法执行的过程中蕴含了大量模式的隐含操作。这里给出相关定义。

定义 1(模式, schema) 模式就是指一些相同的构型, 它描述了在某些相同位置上具有相似结构特征的个体编码串的一个子集。以二进制编码串为例, 个体由二值字符集 $V = \{0, 1\}$ 组成编码串。其模式由三值字符集 $V_+ = \{0, 1, *\}$ 组成。其中 $*$ 为通配符, 既可以表示 0, 也可以表示 1。例如模式 $*01*$ 描述了 4 个元的子集 $\{0010, 0011, 1010, 1011\}$ 。定义在串长为 L 的二进制串上的模式共有 3^L 个。一般而言, 定义在含 k 个基本字符上长度为 L 的串上的模式有 $(k+1)^L$ 个。在

M 个二进制串的群体中至多有 $2^L M$ 个模式。

定义 2(模式阶, schema order) 对代谢算法而言, 在模式 H 中具有确定浓度值的位置数目称为该模式的模式阶, 记为 $o(H)$ 。

定义 3(模式定义长度, schema defining length) 模式 H 中从左到右第一个确定浓度值位置和最后一个确定浓度值的位置间的距离称为该模式的模式定义长度, 记为 $\sigma(H)$ 。

在引入了模式的概念后, 代谢算法的运行过程就是对模式的运算过程。首先假设在代谢过程的第 t 轮中, 当前浓度集合 $P(t)$ 中能匹配模式 H 的个体数为 $m(H, t)$, 下一轮代谢浓度 $P(t+1)$ 中能匹配模式 H 的个体数为 $m(H, t+1)$ 。

(2) 各算子的作用

① 平衡算子的作用

平衡算子 ke 的作用是随机选择酶 E 中的二进制数位, 使对应的数位上的数值与底物对应数位上的数值相等。平衡算子 ke 在一个计算周期中的值为当前全部底物的浓度。设酶 E 由 n 个二进制位组成, 当前全部底物的浓度为 c_s , 则在酶 E 中有 $n \times c_s$ 个二进制位与底物发生同或作用, 由 0 转变成 1 或由 1 转变成 0。设酶 E 二进制位为 0 对应底物的二进制位为 0 的情形有 n_1 个; 酶 E 二进制位为 0 对应底物的二进制位为 1 的情形有 n_2 个; 酶 E 二进制位为 1 对应底物的二进制位为 0 的情形有 n_3 个; 酶 E 二进制位为 1 对应底物的二进制位为 1 的情形有 n_4 个; 满足 $n_1 + n_2 + n_3 + n_4 = n$ 。在平衡算子作用下, 模式 H 所匹配的中间复合物浓度能够增加的概率为: $(n_1 + n_4) \times c_s$ 。由于运用平衡算子的条件是底物浓度大于或远大于生成物浓度, 因此中间复合物浓度变化率(浓度差)减少的概率也为: $(n_1 + n_4) \times c_s$ 。此时, 模式 H 所匹配的复合物浓度能进入下一次代谢平衡选择的个数为:

$$m(H, t+1) = m(H, t) * (n_1 + n_4) * c_s \quad (3)$$

由于底物浓度大于或远大于生成物浓度, 故 $(n_1 + n_4) > 1$, 有: $m(H, t+1) = m(H, t) * (1 + k_1)$, 其中 k_1 大于 1。因此, 从 $t=0$ 开始, 有: $m(H, t+1) = m(H, 0) * (1 + k_1)^t$ 。由此可以认为: 在平衡算子作用下, 平均适应度高于群体平均适应度的模式的个数将呈指数级增长; 同时, 平均适应度低于群体平均适应度的模式的个数将呈指数级减少。

② 抑制算子的作用

抑制算子 ki 的作用是随机选择酶 E 中的二进制数位, 使对应的数位上的数值与底物对应数位上的数值不相等。抑制算子 ki 在一个计算周期中的值为当前全部生成物的浓度。设酶 E 由 n 个二进制位组成, 当前全部生成物的浓度为 c_p , 则在酶 E 中有 $n \times c_p$ 个二进制位与生成物发生同或作用, 满足 $c_s + c_p = 1$ 。设酶 E 二进制位为 0 对应生成物的二进制位为 0 的情形有 n_5 个; 酶 E 二进制位为 0 对应生成物的二进制位为 1 的情形有 n_6 个; 酶 E 二进制位为 1 对应生成物的二进制位为 0 的情形有 n_7 个; 酶 E 二进制位为 1 对应生成物的二进制位为 1 的情形有 n_8 个; 满足 $n_5 + n_6 + n_7 + n_8 = n$ 。在抑制算子作用下, 模式 H 所匹配的各个个体浓度能够减少的概率为: $(n_6 + n_7) \times c_p$ 。由于运用抑制算子的条件是底物浓度小于或远小于生成物浓度, 因此中间复合物浓度变化率(浓度差)减少的概率也为: $(n_6 + n_7) \times c_p$ 。此时, 模式 H 所匹配的复合物浓度能进入下一次代谢平衡选择的个数为:

$$m(H, t+1) = m(H, t) * (n_6 + n_7) * c_p \quad (4)$$

由于底物浓度小于或远小于生成物浓度,故 $(n_0 + n_7) \gg 1$,有: $m(H, t+1) = m(H, t) * (1 + k_2)$,其中 k_2 大于1。因此,从 $t=0$ 开始,有: $m(H, t+1) = m(H, 0) * (1 + k_2)^t$ 。由此可以认为:在抑制算子作用下,平均适应度高于群体平均适应度的模式的个数将呈指数级增长;同时,平均适应度低于群体平均适应度的模式的个数将呈指数级减少。

③ 凋亡算子的作用

由于酶催化的专一性,当某一代谢物不存在或其浓度小于某一极小值时,对应的酶分子应该开始退化。这种机理保证了正常细胞的发育。在代谢平衡实现以前,凋亡算子 k_a 为一不为0的随机数, $k_a \in (0, 1)$ 。当出现某一代谢物浓度小于某一极小值,即 $\sigma(H) \rightarrow 0$ 或 $\sigma(H) \rightarrow 0$ 时,凋亡算子 k_a 的作用是将自身编码各数位全变为1,其意义在于当某一代谢物浓度在代谢反应中可以忽略不计时,对应的催化酶可以实时地退出代谢网络,降低代谢运算的复杂性。

(3) 模式定理

定义4 “在人工代谢算法中,在平衡算子、抑制算子和凋亡算子的作用下,最终代谢物浓度差低于起始代谢物浓度差的模式将呈指数级增长”,该结论称为模式定理。

模式定理的意义在于阐述了代谢算法的理论基础,保证了代谢过程向浓度平衡方向移动。其核心是将浓度差的倒数作为适应度函数。当浓度差越大时,适应度函数值越小,当浓度差达到零或小于某一个指定极小值时,实现浓度平衡,适应度函数值取最优值。代谢寻优调节过程结束。

3.2 算法收敛性分析

由于代谢算法的平衡、抑制、凋亡操作都是独立随机进行的,新代谢物浓度只与上一次代谢物浓度和本次代谢操作算子相关,而与更早之前的代谢浓度与代谢操作无关。同时,各轮代谢过程与时间起点无关。这样就可以将其描述成一个齐次的马尔可夫链 $P_t = \{P_t, t \geq 0\}$,通过利用对齐次马尔可夫链的性质进行理论分析,就可以得到代谢算法的收敛性质^[8]。

定理1 代谢算法能以概率1收敛到最优解。

证明: 设浓度差的各种可能状态的集合为 I 。在代谢算法中,浓度差从一个状态 $i \in I$,经过代谢算子的作用而转变到 $j \in I$ 。其中,平衡、抑制和凋亡算子的转移概率分别为: e_{ij} , r_{ij} 和 a_{ij} ,相应的随机矩阵为: $E = \{e_{ij}\}$, $R = \{r_{ij}\}$, $A = \{a_{ij}\}$,则算法的浓度差状态转移矩阵为: $C = ERA = \{c_{ij}\}$ 。易知, C 是正定的。

在 t 时刻,浓度差是状态 j 的概率 $P_j(t)$ 为: $P_j(t) = \sum_{i \in I} P_i(0) c_{ij}^t$, ($t=0, 1, 2, \dots$)。由齐次马尔可夫链的性质可知, $P_j(t)$ 的稳定概率分布与其初始概率分布无关,即有: $P_j(\infty) = \sum_{i \in I} P_i(\infty) c_{ij} > 0$ 。

对于集合 $P^*(t) = [U(t), P(t)]$,其中 $U(t)$ 为当前集合中适应度最高的个体,它的转移规则如下:

① 依据浓度差状态转移矩阵, $P(t)$ 由 C 产生 $P(t+1)$ 。

② $U(t)$ 是从上一代浓度差集合和本代浓度差集合中选出的一个具有最大适应度的个体,即 $U(t+1) = \max\{U(t), U_0\}$, U_0 是 $P(t+1)$ 中适应度最高的个体。

随机过程 $\{P^*(t), t \geq 0\}$ 仍然是一个齐次的马尔可夫链,即有: $P^*(t) = P^*(0)(C^*)^t$ 。假设集合状态中包含有最优解 I_0 ,则该随机过程的状态转移概率为:

$$\begin{cases} c_{ij}^* > 0, \forall i \in I, \forall j \in I_0 \\ c_{ij}^* = 0, \forall i \in I_0, \forall j \notin I_0 \end{cases} \quad (5)$$

即从任意状态向含有最优解的状态转移概率大于0,而从含有最优解的状态向不含最优解的状态转移的概率为0。此时,对于任意 $i \in I, j \notin I_0$,都有:

$$\begin{cases} (c^*)_{ij}^t \rightarrow 0 (t \rightarrow \infty) \\ P_j^*(\infty) = 0 (j \notin I_0) \end{cases} \quad (6)$$

即算法总能以概率1收敛到最优解,证毕。

4 基于酶操作的 TSP 问题求解

4.1 问题分析

假设存在一最短路径,能遍历这 n 个城市,且路径无重复。若将每个城市之间的距离看成代谢物,则这一最短距离应是整个代谢网络实现最优平衡时的代谢路径。也即当代谢物在整个网络中流通完毕时,所对应的代谢中间产物应为最大。根据这种思路,假设有路径从城市 A_1 到城市 A_2 再到城市 A_3 然后到城市 A_4 ,经过 n 个城市后,最后回到城市 A_1 ,则可在一个代谢反应中将路径 $A_1 A_2$ 的长度看成底物,路径 $A_2 A_3$ 的长度看成生成物;在下一个代谢反应中将路径 $A_2 A_3$ 的长度看成底物,路径 $A_3 A_4$ 的长度看成生成物;在接着的代谢反应中又将路径 $A_4 A_{i+1}$ 的长度看成底物,路径 $A_{i+1} A_{i+2}$ 的长度看成生成物;通过不断调节酶的浓度使各代谢路径上的物质流量实现最大。这实际上是代谢反应中的级联多步反应。其中通过对比底物和生成物的浓度,决定酶的催化作用方向,并根据浓度差决定酶催化作用的强弱^[9,10]。

4.2 人工代谢算法在 TSP 问题的推导过程分析

为简单起见,假设一个网络由A,B,C,D 4个城市组成。已知AB间距离为2;AD间距离为5;AC间距离为3;BC间距离为4;BD间距离为4;CD间距离为3。易得遍历该网络有6种方式,分别为:AB—BC—CD—DA(2—4—3—5),总距离为14; AB—BD—DC—CA(2—4—3—3),总距离为12; AC—CB—BD—DA(3—4—4—5),总距离为16; AD—DC—CB—BA(5—3—4—2),总距离为14; AD—DB—BC—CA(5—4—4—3),总距离为16; AC—CD—DB—BA(3—3—4—2),总距离为12。现分别对这6种模式运用代谢算法进行分析,采用浓度差驱动的方式实现代谢过程。

路径1 AB—BC—CD—DA(2—4—3—5),共14个单位长度

① $t=0$ 到 $t=1$ 过程

由 $y(0) = 0010B = 2D$, $y(1) = 0100B = 4D$, $y(0) < y(1)$ 可得,在该过程中酶的作用表现为抑制催化作用。设酶 E 在 $t=0$ 时刻的二进制编码为 $E(0) = 0001B$,则酶 $E(0)$ 能发挥抑制反应的位数 n 为:

$$n = \frac{y(1) - y(0)}{y(0) + y(1)} * 4 = \frac{4 - 2}{4 + 2} * 4 \approx 1$$

式中,4表示二进制编码的位数为4位。 n 的值通过取整得到。对比 $y(1) = 0100B = 4D$,则能催化 $t=1$ 时刻代谢反应的酶 $E(1)$ 的二进制编码为 $E(1) = 0011B$ 。

② $t=1$ 到 $t=2$ 过程

令 $E(1)$ 与 $y(1)$ 进行“同或”,得到 $t=1$ 到 $t=2$ 过程的代谢中间物二进制编码为:1000B。

由 $y(1) = 0100B = 4D$, $y(2) = 0011B = 3D$, $y(1) > y(2)$ 可

得,在该过程中酶的作用表现为平衡催化作用。由①推出酶 E 在 $t=1$ 时刻的二进制编码为 $E(1)=0011B$,则酶 $E(1)$ 能发挥平衡反应的位数 n 为:

$$n = \frac{y(1)-y(2)}{y(1)+y(2)} * 4 = \frac{4-3}{4+3} * 4 \approx 1$$

对比 $y(2)=0011B=3D$,则能催化 $t=2$ 时刻代谢反应的酶 $E(2)$ 的二进制编码为 $E(2)=0011B$ 。

③ $t=2$ 到 $t=3$ 过程

令 $E(2)$ 与 $y(2)$ 进行“同或”,得到 $t=2$ 到 $t=3$ 过程的代谢中间物二进制编码为:1111B。

由 $y(2)=0011B=3D, y(3)=0101B=5D, y(2) < y(3)$ 可得,在该过程中酶的作用表现为抑制催化作用。由②推出酶 E 在 $t=2$ 时刻的二进制编码为 $E(2)=0011B$,则酶 $E(2)$ 能发挥抑制反应的位数 n 为:

$$n = \frac{y(3)-y(2)}{y(3)+y(2)} * 4 = \frac{5-3}{5+3} * 4 \approx 1$$

对比 $y(3)=0101B=5D$,则能催化 $t=3$ 时刻代谢反应的酶 $E(3)$ 的二进制编码为 $E(3)=0010B$ 。

④ $t=3$ 到 $t=4$ 过程

令 $E(3)$ 与 $y(3)$ 进行“同或”,得到 $t=3$ 到 $t=4$ 过程的代谢中间物二进制编码为:1000B。

路径 2 AB—BD—DC—CA(2—4—3—3),共 12 个单位长度

① $t=0$ 到 $t=1$ 过程

由 $y(0)=0010B=2D, y(1)=0100B=4D, y(0) < y(1)$ 可得,在该过程中酶的作用表现为抑制催化作用。设酶 E 在 $t=0$ 时刻的二进制编码为 $E(0)=0001B$,则酶 $E(0)$ 能发挥抑制反应的位数 n 为:

$$n = \frac{y(1)-y(0)}{y(1)+y(0)} * 4 = \frac{4-2}{4+2} * 4 \approx 1$$

式中,4 表示二进制编码的位数为 4 位。 n 的值通过取整得到。对比 $y(1)=0100B=4D$,则能催化 $t=1$ 时刻代谢反应的酶 $E(1)$ 的二进制编码为 $E(1)=0011B$ 。

② $t=1$ 到 $t=2$ 过程

令 $E(1)$ 与 $y(1)$ 进行“同或”,得到 $t=1$ 到 $t=2$ 过程的代谢中间物二进制编码为:1000B。

由 $y(1)=0100B=4D, y(2)=0011B=3D, y(1) > y(2)$ 可得,在该过程中酶的作用表现为平衡催化作用。由①推出酶 E 在 $t=1$ 时刻的二进制编码为 $E(1)=0011B$,则酶 $E(1)$ 能发挥平衡反应的位数 n 为:

$$n = \frac{y(1)-y(2)}{y(1)+y(2)} * 4 = \frac{4-3}{4+3} * 4 \approx 1$$

对比 $y(2)=0011B=3D$,则能催化 $t=2$ 时刻代谢反应的酶 $E(2)$ 的二进制编码为 $E(2)=0011B$ 。

③ $t=2$ 到 $t=3$ 过程

令 $E(2)$ 与 $y(2)$ 进行“同或”,得到 $t=2$ 到 $t=3$ 过程的代谢中间物二进制编码为:1111B。

由 $y(2)=0011B=3D, y(3)=0011B=3D, y(2)=y(3)$ 可得,在该过程中酶不发挥作用。由②推出酶 E 在 $t=2$ 时刻的二进制编码为 $E(2)=0011B$,则酶 $E(2)$ 能发挥催化反应的位数 n 为:

$$n = \frac{y(3)-y(2)}{y(3)+y(2)} * 4 = \frac{3-3}{3+3} * 4 \approx 0$$

对比 $y(3)=0011B=3D$,则能催化 $t=3$ 时刻代谢反应的

酶 $E(3)$ 的二进制编码为 $E(3)=0011B$ 。

④ $t=3$ 到 $t=4$ 过程

令 $E(3)$ 与 $y(3)$ 进行“同或”,得到 $t=3$ 到 $t=4$ 过程的代谢中间物二进制编码为:1111B。

路径 3 AC—CB—BD—DA(3—4—4—5),共 16 个单位长度

① $t=0$ 到 $t=1$ 过程

由 $y(0)=0011B=3D, y(1)=0100B=4D, y(0) < y(1)$ 可得,在该过程中酶的作用表现为抑制催化作用。设酶 E 在 $t=0$ 时刻的二进制编码为 $E(0)=0001B$,则酶 $E(0)$ 能发挥抑制反应的位数 n 为:

$$n = \frac{y(1)-y(0)}{y(1)+y(0)} * 4 = \frac{4-3}{4+3} * 4 \approx 1$$

式中,4 表示二进制编码的位数为 4 位。 n 的值通过取整得到。对比 $y(1)=0100B=4D$,则能催化 $t=1$ 时刻代谢反应的酶 $E(1)$ 的二进制编码为 $E(1)=0011B$ 。

② $t=1$ 到 $t=2$ 过程

令 $E(1)$ 与 $y(1)$ 进行“同或”,得到 $t=1$ 到 $t=2$ 过程的代谢中间物二进制编码为:1000B。

由 $y(1)=0100B=4D, y(2)=0100B=4D, y(1)=y(2)$ 可得,在该过程中酶不发挥催化作用。由①推出酶 E 在 $t=1$ 时刻的二进制编码为 $E(1)=0011B$,则酶 $E(1)$ 能发挥催化反应的位数 n 为:

$$n = \frac{y(1)-y(2)}{y(1)+y(2)} * 4 = \frac{4-4}{4+4} * 4 \approx 0$$

对比 $y(2)=0011B=3D$,则能催化 $t=2$ 时刻代谢反应的酶 $E(2)$ 的二进制编码为 $E(2)=0011B$ 。

③ $t=2$ 到 $t=3$ 过程

令 $E(2)$ 与 $y(2)$ 进行“同或”,得到 $t=2$ 到 $t=3$ 过程的代谢中间物二进制编码为:1111B。

由 $y(2)=0100B=4D, y(3)=1001B=5D, y(2) < y(3)$ 可得,在该过程中酶发挥抑制催化作用。由②推出酶 E 在 $t=2$ 时刻的二进制编码为 $E(2)=0011B$,则酶 $E(2)$ 能发挥抑制催化反应的位数 n 为:

$$n = \frac{y(3)-y(2)}{y(3)+y(2)} * 4 = \frac{5-4}{5+4} * 4 \approx 0$$

对比 $y(3)=1001B=5D$,则能催化 $t=3$ 时刻代谢反应的酶 $E(3)$ 的二进制编码为 $E(3)=0011B$ 。

④ $t=3$ 到 $t=4$ 过程

令 $E(3)$ 与 $y(3)$ 进行“同或”,得到 $t=3$ 到 $t=4$ 过程的代谢中间物二进制编码为:0101B。

路径 4 AD—DC—CB—BA(5—3—4—2),共 14 个单位长度

① $t=0$ 到 $t=1$ 过程

由 $y(0)=0101B=5D, y(1)=0011B=3D, y(0) > y(1)$ 可得,在该过程中酶的作用表现为平衡催化作用。设酶 E 在 $t=0$ 时刻的二进制编码为 $E(0)=0001B$,则酶 $E(0)$ 能发挥平衡反应的位数 n 为:

$$n = \frac{y(0)-y(1)}{y(0)+y(1)} * 4 = \frac{5-3}{5+3} * 4 \approx 1$$

式中,4 表示二进制编码的位数为 4 位。 n 的值通过取整得到。对比 $y(1)=0011B=3D$,则能催化 $t=1$ 时刻代谢反应的酶 $E(1)$ 的二进制编码为 $E(1)=0011B$ 。

② $t=1$ 到 $t=2$ 过程

令 $E(1)$ 与 $y(1)$ 进行“同或”，得到 $t=1$ 到 $t=2$ 过程的代谢中间物二进制编码为：1111B。

由 $y(1)=0011B=3D, y(2)=0100B=4D, y(1)<y(2)$ 可得，在该过程中酶的作用表现为抑制催化作用。由①推出酶 E 在 $t=1$ 时刻的二进制编码为 $E(1)=0011B$ ，则酶 $E(1)$ 能发挥抑制反应的位数 n 为：

$$n = \frac{y(2)-y(1)}{y(2)+y(1)} * 4 = \frac{4-3}{4+3} * 4 \approx 1$$

对比 $y(2)=0100B=4D$ ，则能催化 $t=2$ 时刻代谢反应的酶 $E(2)$ 的二进制编码为 $E(2)=1011B$ 。

③ $t=2$ 到 $t=3$ 过程

令 $E(2)$ 与 $y(2)$ 进行“同或”，得到 $t=2$ 到 $t=3$ 过程的代谢中间物二进制编码为：0000B。

由 $y(2)=0100B=4D, y(3)=0010B=2D, y(2)>y(3)$ 可得，在该过程中酶的作用表现为平衡催化作用。由②推出酶 E 在 $t=2$ 时刻的二进制编码为 $E(2)=1011B$ ，则酶 $E(2)$ 能发挥平衡反应的位数 n 为：

$$n = \frac{y(2)-y(3)}{y(2)+y(3)} * 4 = \frac{4-2}{4+2} * 4 \approx 1$$

对比 $y(3)=0010B=2D$ ，则能催化 $t=3$ 时刻代谢反应的酶 $E(3)$ 的二进制编码为 $E(3)=1010B$ 。

④ $t=3$ 到 $t=4$ 过程

令 $E(3)$ 与 $y(3)$ 进行“同或”，得到 $t=3$ 到 $t=4$ 过程的代谢中间物二进制编码为：0111B。

路径 5 AD—DB—BC—CA(5—4—4—3)，共 16 个单位长度

① $t=0$ 到 $t=1$ 过程

由 $y(0)=0101B=5D, y(1)=0100B=4D, y(0)>y(1)$ 可得，在该过程中酶的作用表现为平衡催化作用。设酶 E 在 $t=0$ 时刻的二进制编码为 $E(0)=0001B$ ，则酶 $E(0)$ 能发挥平衡反应的位数 n 为：

$$n = \frac{y(0)-y(1)}{y(0)+y(1)} * 4 = \frac{5-4}{5+4} * 4 \approx 0$$

式中，4 表示二进制编码的位数为 4 位。 n 的值通过取整得到。对比 $y(1)=0100B=4D$ ，则能催化 $t=1$ 时刻代谢反应的酶 $E(1)$ 的二进制编码为 $E(1)=0001B$ 。

② $t=1$ 到 $t=2$ 过程

令 $E(1)$ 与 $y(1)$ 进行“同或”，得到 $t=1$ 到 $t=2$ 过程的代谢中间物二进制编码为：1010B。

由 $y(1)=0100B=4D, y(2)=0100B=4D, y(1)=y(2)$ 可得，在该过程中酶不发挥催化作用。由①推出酶 E 在 $t=1$ 时刻的二进制编码为 $E(1)=0001B$ ，则酶 $E(1)$ 能发挥催化反应的位数 n 为：

$$n = \frac{y(2)-y(1)}{y(2)+y(1)} * 4 = \frac{4-4}{4+3} * 4 \approx 0$$

对比 $y(2)=0100B=4D$ ，则能催化 $t=2$ 时刻代谢反应的酶 $E(2)$ 的二进制编码为 $E(2)=0001B$ 。

③ $t=2$ 到 $t=3$ 过程

令 $E(2)$ 与 $y(2)$ 进行“同或”，得到 $t=2$ 到 $t=3$ 过程的代谢中间物二进制编码为：1010B。

由 $y(2)=0100B=4D, y(3)=0011B=3D, y(2)>y(3)$ 可得，在该过程中酶的作用表现为平衡催化作用。由②推出酶

E 在 $t=2$ 时刻的二进制编码为 $E(2)=0001B$ ，则酶 $E(2)$ 能发挥平衡反应的位数 n 为：

$$n = \frac{y(2)-y(3)}{y(2)+y(3)} * 4 = \frac{4-3}{4+2} * 4 \approx 1$$

对比 $y(3)=0011B=3D$ ，则能催化 $t=3$ 时刻代谢反应的酶 $E(3)$ 的二进制编码为 $E(3)=1011B$ 。

④ $t=3$ 到 $t=4$ 过程

令 $E(3)$ 与 $y(3)$ 进行“同或”，得到 $t=3$ 到 $t=4$ 过程的代谢中间物二进制编码为：0111B。

路径 6 AC—CD—DB—BA(3—3—4—2)，共 12 个单位长度

① $t=0$ 到 $t=1$ 过程

由 $y(0)=0011B=3D, y(1)=0011B=3D, y(0)=y(1)$ 可得，在该过程中酶不发挥催化作用。设酶 E 在 $t=0$ 时刻的二进制编码为 $E(0)=0001B$ ，则酶 $E(0)$ 能发挥催化反应的位数 n 为：

$$n = \frac{y(0)-y(1)}{y(0)+y(1)} * 4 = \frac{3-3}{3+3} * 4 \approx 0$$

式中，4 表示二进制编码的位数为 4 位。 n 的值通过取整得到。对比 $y(1)=0011B=3D$ ，则能催化 $t=1$ 时刻代谢反应的酶 $E(1)$ 的二进制编码为 $E(1)=0001B$ 。

② $t=1$ 到 $t=2$ 过程

令 $E(1)$ 与 $y(1)$ 进行“同或”，得到 $t=1$ 到 $t=2$ 过程的代谢中间物二进制编码为：1101B。

由 $y(1)=0011B=3D, y(2)=0100B=4D, y(1)<y(2)$ 可得，在该过程中酶发挥抑制催化作用。由①推出酶 E 在 $t=1$ 时刻的二进制编码为 $E(1)=0001B$ ，则酶 $E(1)$ 能发挥抑制催化反应的位数 n 为：

$$n = \frac{y(2)-y(1)}{y(2)+y(1)} * 4 = \frac{4-3}{4+3} * 4 \approx 1$$

对比 $y(2)=0100B=4D$ ，则能催化 $t=2$ 时刻代谢反应的酶 $E(2)$ 的二进制编码为 $E(2)=0011B$ 。

③ $t=2$ 到 $t=3$ 过程

令 $E(2)$ 与 $y(2)$ 进行“同或”，得到 $t=2$ 到 $t=3$ 过程的代谢中间物二进制编码为：1000B。

由 $y(2)=0100B=4D, y(3)=0010B=2D, y(2)>y(3)$ 可得，在该过程中酶的作用表现为平衡催化作用。由②推出酶 E 在 $t=2$ 时刻的二进制编码为 $E(2)=0011B$ ，则酶 $E(2)$ 能发挥平衡反应的位数 n 为：

$$n = \frac{y(2)-y(3)}{y(2)+y(3)} * 4 = \frac{4-2}{4+2} * 4 \approx 1$$

对比 $y(3)=0010B=2D$ ，则能催化 $t=3$ 时刻代谢反应的酶 $E(3)$ 的二进制编码为 $E(3)=0010B$ 。

④ $t=3$ 到 $t=4$ 过程

令 $E(3)$ 与 $y(3)$ 进行“同或”，得到 $t=3$ 到 $t=4$ 过程的代谢中间物二进制编码为：1111B。

由对上述 6 种路径的分析表明：当中间代谢物取到最大值 1111 时，所对应的路径为 AB—BD—DC—CA(2—4—3—3)，共 12 个单位长度；或 AC—CD—DB—BA(3—3—4—2)，共 12 个单位长度。此时 TSP 问题取得最优解，与实际情形相吻合。这也从另外一个方面说明了人工代谢算法的可靠性和有效性。

(下转第 242 页)

均 F_1 值的比较。表 3 给出了算法的训练时间和分类时间的比较。

表 2 宏平均准确率、宏平均召回率和宏平均 F_1 值比较

算法	宏平均准确率(%)	宏平均召回率(%)	宏平均 F_1 值(%)
1-a-l	68.27	62.91	64.71
1-a-r	68.51	64.11	65.61
DAGSVM	68.22	62.75	64.23
HBTSVM	69.07	64.15	65.83

表 3 训练时间和分类时间比较

算法	训练时间(ms)	分类时间(ms)
1-a-l	1038	469
1-a-r	2109	422
DAGSVM	1038	357
HBTSVM	986	315

由实验结果可知,在分类精度上,HBTSVM 与 1-a-r 相当,略高于 1-a-l 和 DAGSVM;在训练速度上,HBTSVM 高于 1-a-l, 1-a-r 和 DAGSVM,其原因是 1-a-l 和 DAGSVM 需要训练的子分类器数目较多,虽然 HBTSVM 和 1-a-r 需要训练的子分类器的数目相同,但 1-a-r 的每个子分类器的训练都要求全部训练样本,而 HBTSVM 的子分类器中只有一个要求全部训练样本,其余只要求部分训练样本;在分类速度上,HBTSVM 明显高于 1-a-l,1-a-r 和 DAGSVM,其原因是在对新样本进行分类时,其它 3 种方法要么需要较多的子分类器,要么需要进行较复杂的计算。由此可知,HBTSVM 在保证分类精度的前提下,分类速度有了较大的提高,在应用方面具有一定的优越性。若类别较多、各类训练样本的规模相同,则生成的二叉树为完全二叉树,分类速度的优势更明显。

结束语 多类分类在许多应用领域具有重要的价值。本文对支持向量机多类分类方法进行深入研究,提出了一种新的基于二叉树的支持向量机多类分类算法。在保证分类精度的同时,分类速度有较大提高。采用该算法在 Rutgers 21578 标准数据集上进行了分类实验,结果证明了该算法的有效性。进一步的研究工作是在类别数较多、规模较大的数据集上进行试验。

参 考 文 献

[1] Vapnik V. The Nature of Statistical Learning Theory[M]. New

(上接第 199 页)

结束语 TSP 问题一直是作为检验人工智能算法优劣的标志性问題。本文通过对人工代谢算法的思想来源、编码规则、运行模型、收敛性能地阐述和证明,说明了算法的运作过程,并对 TSP 问题的求解过程进行了详细分析,给出了相关的结论。

参 考 文 献

[1] 王凌. 智能优化算法及其应用[M]. 北京:清华大学出版社,2001
 [2] 陈国良,王煦法,庄镇泉,等. 遗传算法及其应用[M]. 北京:人民邮电出版社,1996
 [3] 焦李成,杜海峰,刘芳,等. 免疫优化计算、学习与识别[M]. 北京:科学出版社,2006
 [4] 丁永生,任立红. 人工免疫系统:理论与应用[J]. 模式识别与人

York, Springer, 1995

[2] Bennett K P. Combining Support Vector and Mathematical Programming Methods for Classification[A]// Advances in Kernel Methods; Support Vector Learning[C]. Cambridge, MA; MIT press, 1999; 307-326
 [3] Krebel U G. Pairwise Classification and Support Vector Machines [A]// Advances in Kernel Methods; Support Vector Learning[C]. Cambridge, MA; MIT press, 1999; 255-268
 [4] Platt J C, Cristianini N, Shawe-Taylor J. Large Margin DAGs for multiclass classification[A]// Advances in Neural Information Processing Systems[C]. Cambridge, MA; MIT Press, 2000; 547-553
 [5] Ghani R. Using error-correcting codes for text classification[C]// Proc of the 17th International Conference on Machine Learning, 2002
 [6] Cheong S. Support vector machines with Binary tree architecture for multi-class classification[J]. Neural Information Processing Letters and Reviews, 2004
 [7] 孟媛媛,刘希玉. 一种新的基于二叉树的 SVM 多类分类方法[J]. 计算机应用, 2005, 25(11): 2653-2657
 [8] 马笑潇,黄席樾,柴毅. 基于 SVM 的二叉树多类分类算法及其在故障诊断中的应用[J]. 控制与决策, 2003, 18(3): 272-276
 [9] 孙德山,吴今培. 基于线性规划的多支持向量机算法[J]. 计算机科学, 2005, 32(10): 160-163
 [10] 朱远平,戴汝为. 基于 SVM 决策树的文本分类器[J]. 模式识别与人工智能, 2005, 18(4): 412-416
 [11] 应伟,王正欧,安金龙. 一种基于改进的支持向量机的多类文本分类方法[J]. 计算机工程, 2006, 32(16): 74-76
 [12] 张永,迟忠先,闫德勤. 一种新的模糊补偿多类支持向量机[J]. 计算机科学, 2006, 33(12): 152-155
 [13] 唐发明,王仲东,陈锦云. 一种新的二叉树多类支持向量机算法[J]. 计算机工程与应用, 2005, 41(7): 24-26
 [14] 唐发明,王仲东,陈锦云. 支持向量机多类分类算法研究[J]. 控制与决策, 2005, 20(7): 746-749
 [15] Chang Chinchang, Lin Chihjen. LIBSVM: a Library for Support Vector Machines[EB/OL]. <http://www.csie.ntu.tw/~cjlin/libsvm>, 2005

工智能, 2000, 13(1): 52-59

[5] 李士勇. 蚁群算法及其应用[M]. 哈尔滨:哈尔滨工业大学出版社, 2004
 [6] 埃伯哈德. O. 沃伊特. 生物化学系统的计算分析[M]. 储炬,李友荣,译. 北京:化学工业出版社, 2006
 [7] 内斯托尔 V. 托雷斯,埃伯哈德. O. 沃伊特. 代谢工程的途径分析与优化[M]. 修志龙,滕虎,译. 北京:化学工业出版社, 2005
 [8] 王翼飞,史定华. 生物信息学——智能化算法及其应用[M]. 北京:化学工业出版社, 2006
 [9] 梅乐和,岑沛霖. 现代酶工程[M]. 北京:化学工业出版社, 2006
 [10] Chekhova E, Barton P I, Gorak A. Optimal operation processes of discrete-continuous biochemical Processes [J]. Computers and Chemical Engineering, 2000, 24: 1167-1173