

## 一种基于模糊神经网络的化学分子生物毒性预测模型的方法

### 技术领域

本发明属于化工技术领域，具体是一种基于模糊神经网络的化学分子生物毒性预测模型的方法。

### 背景技术

近些年来，化学物质合成技术飞速发展，合成物质已经应用于社会的各行各业，随着工业的高度发展，全球化学品的种类和年产量在急剧增长，化学污染物在品种和数量上日益增多，化学物质的污染处理早已受到了世界重点关注，其中研究如何减小化学物质的毒性已经成为了当今社会急需解决的问题。因此，在研究化学物质合成过程中运行控制技术，在保证精确控制的前提下，实现化学物质排放时毒性达标，减小环境污染，是未来环保型化学物质研究的必然趋势。

目前，针对化学物质生物毒性控制通常前段检测控制和后端排放处理。前端检测控制常采用的方法就是通过利用各种仪器对已合成的化学物质毒性检测，超过指标则重新合成或者禁止使用，这种方法容易造成大量经济损失，并且合成过程比较复杂麻烦；而后端排放处理通常是物质排放后对其造成的污染进行处理，这种手段容易造成环境污染和加大处理经济成本。姚碧云等人在论文《QSAR 方法在毒性预测中的应用和进展》中给出了一种基于 QSAR 的毒性预测方法，但其给出的是一种思路，未给出任何实质性的步骤和研究数据，无法得到其方案可行的结论，专利号为 CN201510347479.0、CN201310705921.3 等文献也给出了类似的方案，但其计算预测精度较小，误差较大。针对这一种情况，本文中给出了一种基于模糊神经网络预测模型进行预测控制算法，利用模糊神经网络预测模型对建立的疏水性与生物毒性 QSAR 模型关系加以精确计算，比原来的预测手段提高了精确度，减小了预测误差，实现了对化学物质生物毒性的提前预测，降低了化学合成的成本和减小了环境污染。

### 发明内容

针对上述问题，本发明设计了一种基于模糊神经网络的化学分子生物毒性预测模型的方法，通过构建模糊神经网络模型，利用模型预测控制的方法实现生物毒性的提前预测。

本发明的技术方案如下：

本发明获得了一种基于模糊神经网络的化学分子生物毒性预测模型的方法，该方法是基于模糊神经网络，构建了化学分子生物毒性预测模型，通过改变神经网络参数，达到预测化学分子生物毒性的目的；提高了化学分子的预测精度，减小了预测误差。

## 说明书

本发明采用了如下的技术方案和步骤：

1、建立生物毒性与辛醇/水分配系数的 QSAR 模型关系。具体是指，建立用于化学分子合成过程中生物毒性（用  $LC_{50}$  半数致死浓度）与辛醇/水分配系数的模型关系；其步骤为：利用回归分析法推导  $\log Kow$  和毒性  $EC_{50}$  之间的数学表达式，建立方程： $y = ax + b$ ，其中  $y$  表示毒性效应浓度，即为  $\log LC_{50}$ ，单位为  $mmol/L$ ， $x$  是  $\log Kow$  值；最终得到公式：

$$\log(LC_{50}) = a\log(Kow) + b \quad (1)。$$

式（1）中的  $a$ ， $b$  为线性回归系数，应用这些导出的线性方程后，用于计算未测试化学品的毒性值， $mmol/L$ 。

2、根据步骤 1 的生物毒性与辛醇/水分配系数的 QSAR 模型关系，建立模糊神经网络化学分子毒性预测模型，目的是使得计算的生物毒性值更加精确。这里我们选取了包含 5 层的 NFN 结构，包括输入层，可信度层，归一化层，解模糊层，输出层。下面是各层的计算函数：

网络第一层为输入层，主要目的是将输入信号模糊化，得到信号隶属度。我们可以采用某种函数来实现，通常包括高斯函数，钟形函数、Sigmoid 函数等，这里我们采用高斯函数来实现：

$$\mu(x_i) = e^{-((x_i - c)/a)^2} \quad (2)$$

式中， $i$  取值 1、2， $c$  为函数的中心值， $a$  为函数的中心宽度值。

网络第二层是可信度层，目的是把模糊化的信号隶属度结果相乘，实际得到的是每条模糊推理规则的可信度。

$$\omega_1 = \mu_1\mu_3, \quad \omega_2 = \mu_1\mu_4, \quad \omega_3 = \mu_2\mu_3, \quad \omega_4 = \mu_2\mu_4 \quad (3)$$

网络第三层为归一化层，目的是将可信度层的可信度进行归一化计算。

$$\bar{\omega}_1 = \omega_1 \sum \omega_i, \quad \bar{\omega}_2 = \omega_2 \sum \omega_i, \quad \bar{\omega}_3 = \omega_3 \sum \omega_i, \quad \bar{\omega}_4 = \omega_4 \sum \omega_i \quad (4)$$

网络第四层为解模糊层，目的是计算每条规则的输出，把输入信号  $x_1$ ， $x_2$  引入解模糊网络模型，通常包括一阶 Sugeno 模型或者二阶 Sugeno 模型，这里我们采用一阶 Sugeno 模型，其函数如下：

$$O_n = \bar{\omega}_n (P_n x_1 + Q_n x_2 + R_n) \quad (5)$$

式中， $n$  取值 1、2、3、4， $P$ ， $Q$ ， $R$  为线性回归系数。

网络第五层为输出层，是各条规则的累加， $y = \sum O_n$ 。

3、根据步骤 2 的模糊神经网络化学分子毒性预测模型，建立优化 NFN 参数的遗传算法模型，具体是修正了模糊神经网络参数，使得计算值更加精确；建立符合度函数：

$$F=1/(1+E) \quad (6)$$

$$E = \sum_{i=1}^{\infty} [(F(i) - F_p)^2] \quad (7)$$

E 为与控制目标的误差, F(t) 是模糊神经网络模型的输出;  $F_p$  是输出的控制目标。

4、根据步骤 2 所述的一种基于模糊神经网络的化学分子生物毒性预测模型的方法, 完成上述步骤后, 进行:

步骤 S4、利用优化后的模糊神经网络模型对新分子的生物毒性值进行计算预测, 根据所计算的生物毒性值判断该分子是否符合要求。如《中国环境保护行业标准》的要求。

本发明的有益效果是:

- 1、本发明中研究了化学分子的疏水性(用辛醇/水分配系数来表示)与生物毒性的关系;
- 2、本发明建立了化学分子的疏水性(用辛醇/水分配系数来表示)与生物毒性的 QSAR 模型关系, 并利用线性回归法解出了具体的方程值;
- 3、本发明中建立了模糊神经网络的生物毒性预测模型算法, 并通过优化参数后的新模型对生物毒性进行了计算和预测, 比原来的生物毒性 QSAR 模型相比, 具有精度高、误差小、高效稳定等特点。

## 附图说明

图 1 为本发明的流程结构图;

图 2 为本发明的模糊神经网络拓扑结构;

图 3 为本发明的不同结构化合物的生物毒性测量值与预测值;

图 4 为本发明的不同结构化合物的生物毒性测量值与预测值的误差。

## 具体实施方式

下面结合附图和实施例对本发明进一步说明。

在本次实例中, 我们选取含氯乙烷的不同结构的化学分子, 通过实验我们可以得出不同分子的疏水性, 即  $\log Kow$  值, 以及部分物质的毒性检测值, 同时, 采用以上方法我们可以对不同分子进行毒性预测, 最终通过对比可以看出预测值与检测值的误差较小, 如表 1 所示。这里以计算预测氯取代基的乙烷分子毒性为例。

根据如图 1 所示, 本发明的基于模糊神经网络的化学分子生物毒性预测模型的方法, 主要步骤如下内容如下:

- 1、建立用于化学分子合成过程中生物毒性(用  $LC_{50}$  半数致死浓度)与辛醇/水分配系数( $Kow$ )的 QSAR 模型关系, 利用回归分析法推导  $\log Kow$  和毒性  $EC_{50}$  之间的数学表达式, 建立方程:  $y = ax + b$ , 其中“y”表示毒性效应浓度(即,  $\log LC_{50}$ , 单位为  $mmol/L$ ), “x”

# 说明书

是 Log Kow 值。将现有的部分常用的结构化合物的辛醇/水分配系数进行统计，并制作下表 1，根据表 1 中的辛醇/水分配系数的数据，采用线性回归法中的最小二乘法对一元方程“ $y = ax + b$ ”进行推导，可以计算得出回归系数的值 a 为-0.746，b 为 0.89。

则最终可写为：

$$\text{Log}(LC_{50}) = -0.746\text{Log}(Kow) + 0.89 \quad (8)$$

计算含氯乙烷的不同结构的化学分子的 Log LC<sub>50</sub> 值，如表 1 所示。

表 1

不同结构化合物	辛醇-水分配系数	LC <sub>50</sub> 测量值	QSAR 预测值
氯乙烷	1.54407	0.27251	-0.26187
1,1 -二氯乙烷	1.79029	0.06416	-0.44555
1,2 -二氯乙烷	1.48001	0.32672	-0.21409
1,1,1-三氯乙烷	2.49969	-0.53614	-0.97477
1,1,2-三氯乙烷	2.46982	-0.51086	-0.95249
1,1,1,2-四氯乙烷	2.38917	-0.44261	-0.89232
五氯乙烷	2.88986	-0.86630	-1.26584
六氯乙烷	4.59988	-2.31332	-2.54151

2、根据步骤 1 建立模糊神经网络化学分子毒性预测模型，建立了基于 Sugeno 模型，这里我们选取了包含 5 层的 NFN 结构，包括输入层，可信度层，归一化层，解模糊层，输出层。如图 2 所示，下面是各层的计算函数：

网络第一层为输入层，主要目的是将输入信号模糊化，得到信号隶属度。我们可以采用某种函数来实现，通常包括高斯函数，钟形函数、Sigmoid 函数等，这里我们采用高斯函数来实现：

$$\mu(x_i) = e^{-((x_i - c)/a)^2} \quad (9)$$

式中，i 取值 1、2，c 为函数的中心值，a 为函数的中心宽度值。

对上式函数两边求取 Log 函数，通过积分求得 c=1.6，a=0.14。得到函数

$$\mu(x_i) = e^{-((x_i - 1.6)/0.14)^2}$$

网络第二层是可信度层，目的是把模糊化的信号隶属度结果相乘，实际得到的是每条模糊推理规则的可信度。

## 说 明 书

$$\omega_1 = \mu_1\mu_3, \quad \omega_2 = \mu_1\mu_4, \quad \omega_3 = \mu_2\mu_3, \quad \omega_4 = \mu_2\mu_4 \quad (10)$$

网络第三层为归一化层，目的是将可信度层的可信度进行归一化计算。

$$\bar{\omega}_1 = \omega_1 / \sum \omega_i, \quad \bar{\omega}_2 = \omega_2 / \sum \omega_i, \quad \bar{\omega}_3 = \omega_3 / \sum \omega_i, \quad \bar{\omega}_4 = \omega_4 / \sum \omega_i \quad (11)$$

网络第四层为解模糊层，目的是计算每条规则的输出，把输入信号  $x_1, x_2$  引入解模糊网络模型，通常包括一阶 Sugeno 模型或者二阶 Sugeno 模型，这里我们采用一阶 Sugeno 模型，其函数如下：

$$O_n = \bar{\omega}_n (P_n x_1 + Q_n x_2 + R_n) \quad (12)$$

式中，n 取值 1、2、3、4，P，Q，R 为线性回归系数。

网络第五层为输出层，是各条规则的累加， $y = \sum O_n$ 。

3、根据步骤 2 建立优化 NFN 参数的遗传算法模型，其特征在于修正了模糊神经网络参数，使得计算值更加精确。建立符合度函数：

$$F = 1 / (1 + E) \quad (13)$$

$$E = \sum_{i=1}^{\infty} [(F(i) - F_p)^2] \quad (14)$$

E 为与控制目标的误差，F(t) 是模糊神经网络模型的输出； $F_p$  是输出的控制目标。

4、根据步骤 2 所述的一种基于模糊神经网络的化学分子生物毒性预测模型的方法，完成上述步骤后，进行：

步骤 S4、利用优化后的模糊神经网络模型对含氯取代基的乙烷的生物毒性值进行计算预测，代入步骤 3 中，最终计算出  $LC_{50}$ 。如表 2 所示。根据图 3 和图 4 所示可以发现，经过模糊神经网络模型计算测得的值比原有的函数值更加精确，误差值更小。

表 2

不同结构化合物	辛醇-水分配系数	$LC_{50}$ 测量值	神经模糊网络优化的 $LC_{50}$ 预测值
氯乙烷	1.54407	0.27251	0.26125
1,1-二氯乙烷	1.79029	0.06416	0.07815
1,2-二氯乙烷	1.48001	0.32672	0.30672
1,1,1-三氯乙烷	2.49969	-0.53614	-0.6375

## 说明书

1,1,2-三氯乙烷	2.46982	-0.51086	-0.5367
1,1,1,2-四氯乙烷	2.38917	-0.44261	-0.0313
五氯乙烷	2.88986	-0.86630	-0.8863
六氯乙烷	4.59988	-2.31332	-2.4130

以上所述，仅是本发明的较佳实施例而已，并非对本发明作任何形式上的限制，虽然本发明已以较佳实施例揭露如上，然而并非用以限定本发明，任何熟悉本专业的技术人员，在不脱离本发明技术方案范围内，当可利用上述揭示的技术内容作出些许更动或修饰为等同变化的等效实施例，但凡是未脱离本发明技术方案的内容，依据本发明的技术实质对以上实施例所作的任何简单修改、等同变化与修饰，均仍属于本发明技术方案的范围内。