

2015/04/10

MOE 操作マニュアル

# *QuaSAR-Evolution*

Version 2011.04.15

株式会社 モルシス

ライフサイエンス部

〒104-0033 東京都中央区新川1-28-38 東京ダイヤビル

TEL : 03-3553-8030

FAX : 03-3553-8031

E-mail : support@molsis.co.jp

URL : <https://www.molsis.co.jp/>

- 本編の内容は、改善のため予告なく変更されることがあります。
- 本編の内容を無断で転載することを禁止します。



QSAR における記述子選択の問題について、現バージョンの MOE では、以下の 4 種類の方法をとることができます。

- *Correlation Matrix* で共線性のある記述子をチェックする
- *QuaSAR-Contingency* で、従属変数と相関性がなさそうなものを除外する
- *QuaSAR-Model* 実行後、寄与の少ない記述子を順次削除する
- *QuaSAR-Classify* で、判別に大きく寄与する記述子をチェックする

しかしながら、重回帰分析において、記述子の自動選択が可能な *Stepwise regression* などの一般的な手法は、MOE2005.06 版では搭載されていません。

*QuaSAR-Evolution* は、ご要望の多い記述子の自動選択が可能な QSAR の手法として、遺伝的アルゴリズム(GA)を弊社でコード化したものです。GA は、*Stepwise regression* や PLS に比べてモデル構築に時間はかかりますが、一般的に、よりシンプルで予測精度の高いモデルを得ることができます。また、複数のモデルを同時に進化させるという特徴から、モデル構築に必須な記述子を容易に見つけ出すことも可能です。

このアプリケーションに関するご意見・ご要望が御座いましたら下記アドレス、Fax までご連絡ください。皆様からお寄せいただいたご意見は、弊社から CCG 社にレポートし、MOE のバージョンアップの参考にさせていただきます。

*E-mail* : [support@molsis.co.jp](mailto:support@molsis.co.jp)

## 1. 機能概要

### ➤ 遺伝的アルゴリズム

#### ☐ 手法：

ランダムに選択した記述子を組み合わせ、多数の回帰式を作成し母集団とする。この中から 2 つの式をランダムに（ただし、優れた固体がより多く選択されるようなバイアスをかけて）選択し、かけ合わせ、突然変異を経て、母集団の中の最も劣った式よりも優れた式が得られた場合に、世代交代させる。

#### ☐ 設定可能なパラメータ：

初期集団の式の長さ、個体数、突然変異確率、優れた固体が次世代に子孫を残す確率、世代数、式を固定長とするか可変長とするかなどを設定可能。

#### ☐ 評価関数：式が進化しているかどうかの判定基準

AIC, LOF(Lack of Fit), R2, 自由度修正済みから選択可能。

テストセットの指定により、進化の判断基準としてテストセットに対する RMSE も同時に考慮することが可能。この場合は、指定評価関数からみて良い式が生まれた場合でも、テストセットに対する RMSE\_TestSet が悪化する場合は、世代交代をさせないという判断を加えている。なお、テストセットを指定しない場合は、LOF 以外の評価関数で Variable Length GA を実行すると、オーバーフィッティングとなるモデルになりやすい。

#### ☐ 進化の継続：

いったん QuaSAR-Evolution でモデル化を行ったデータセットについては、終了時の母集団をもとに、進化を継続させることができる。この際、記述子を追加するなどの条件変更を行ったり、MOE を再起動した後から継続することも可能。

### ➤ 結果出力

#### ☐ 以下の 2 つの .mdb ファイルに出力

##### ● モデル式集団            ga\_\*\*\*\_eq.mdb

下記の項を出力し、100 世代ごとに更新

Equation (モデル式)、RMSE、RMSE\_TestSet、RR (決定係数)、adjRR (自由度調整済決定係数)、AIC (赤池情報量基準)、LOF (Lack of Fit)、F 値

##### ● トラジェクトリ            ga\_\*\*\*\_trj.mdb

下記の項を出力し、100 世代ごとに更新

Generation (世代数)、式集団中の各記述子のポピュレーション

➤ **モデルの評価**

☐ モデル式による推算

指定したモデル式による推算値と Z スコアをデータベースに出力

☐ Cross Validation

Leave One Out, Leave 20% Out 実行可能

ga\_\*\*\*\_eq.mdb に、この結果に基づく x\_RR と x\_adjRR が出力される

➤ **モデルの保存と利用**

☐ モデルの保存

ga\_\*\*\*\_eq.mdb に出力された中で、後に物性推算に使用したいモデルが得られた場合、別のファイル (デフォルト : mymodels.mdb) に出力可能。

このファイルには、決定係数など ga\_\*\*\*\_eq.mdb に出力された全情報が、モデル名、トレーニングセットとテストセットのファイル名とともに保存され、複数の QuaSAR-Evolution の結果を整理して保存、利用が可能。

☐ モデル式による物性推算

mymodels.mdb などのファイルに保存された式は、QuaSAR-ModelCalculation から利用可能なので、物性推算の際には、QuaSAR-Evolution を起動する必要はない。

## 2. インストール方法

インストール方法は二種類ある。インストールした場所によって起動方法が若干異なるが、2-1の方法でインストールを行うと、MOE 起動と同時に svl ファイルが Load される。

### ➤ 2-1 \$MOE/custom/フォルダ中にインストール

- ① ディレクトリ \$MOE/custom/svl/rsi.svl/ を作成する。(MOE2016)
- ② ga.svl を \$MOE/custom/svl/rsi.svl/ に保存する。  
(※ MOE2015 は\$MOE/custom/lib/svl/内に rsi.svl フォルダを作ります。)
- ③ MOE を起動する。

### ➤ 2-2 一時的に使用

- ① ga.svl を作業ディレクトリ中 (例 C:/ccg/work など) におく。作業ディレクトリ中に保存した場合は、QuaSAR-Evolution を起動する前に MOE | File | Open から ga.svl を選択し、Load SVL File で svl ファイルを Load する。

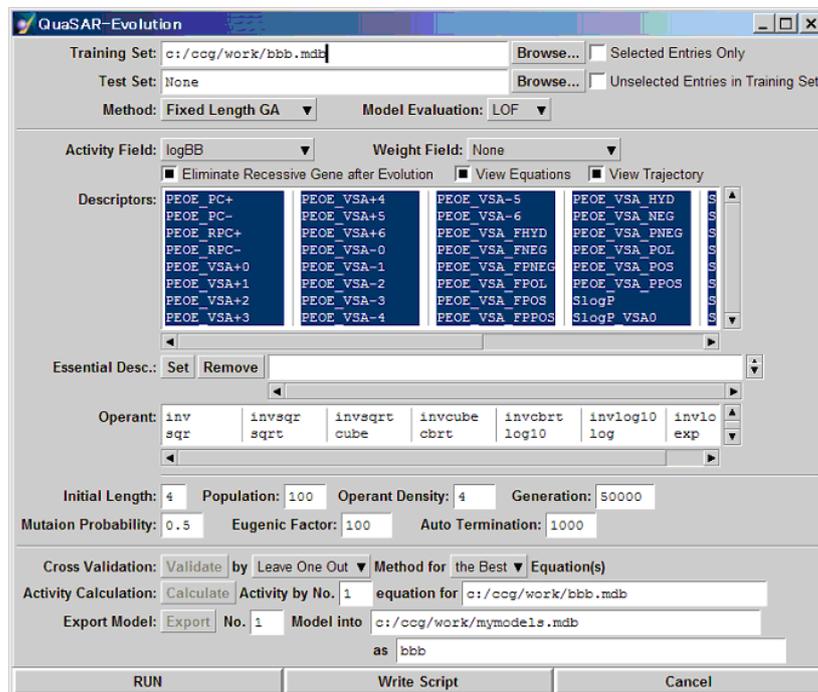
## 3. 操作法

### ➤ 3-1 QuaSAR-Evolution の起動

- 解析するデータベースは、あらかじめ記述子を算出しておく。
- コマンドラインに GA[ ] と入力し Enter キーを押して、QuaSAR-Evolution パネルを起動する。



### ➤ 3-2 QuaSAR-Evolution パネルの説明



#### ☐ Training Set

解析対象となるデータベースファイル名

QuaSAR-Evolution 起動時に開いているファイルがあれば自動認識される。何もファイルが開いていない場合は、Open Database Viewer パネルが開き、ファイル選択が促される。また Browse ボタンを押すことで対象データベースを選択、変更することもできる。

#### ☐ Selected Entries Only

このオプションを選択すれば、Training Set 中の選択されたエントリのみがトレーニングセットとして扱われる。

#### ☐ Test Set

Test Set のファイル名が指定されていれば、モデル構築と同時にこのファイルのデータを使用して RMSE\_TestSet が算出され、評価関数とともに進化の判定基準となる。Training Set と同様、記述子は算出済みであること。

#### ☐ Unselected Entries in Training Set

このオプションを選択すれば、Training Set 中の選択されていないエントリがテストセットとして扱われる。

なお、このオプションを選択すれば自動的に、Test Set には Training Set と同じファイル名が設定され、書き換えできなくなる。また、Training Set の Selected Entries Only オプションも選択状態になる。

- ⇒ **Method**  
進化の過程で式の長さを一定に保つかどうかを指定。式を可変長とした場合は、Variable Length GA に変更する。
- ⇒ **Model Evaluation**  
世代交代の基準となる評価関数  
Variable Length 法で Test Set がない場合は、LOF を推奨。 AIC、adjRR とも、式が長くなることへのペナルティが LOF よりも少ないため、Test Set の指定がない場合は、オーバーフィッティングしたモデルになりやすい。一方、LOF では十分な R2 の式が得られなかった場合、Test Set を指定した上で、他の評価関数を使用すれば、より多くの記述子が含まれた式が得られる可能性が高い。結果は、Cross Validation で比較する必要がある。
- ⇒ **Activity Field**  
デフォルトでは、Database ファイルで数値が保存されたカラムのもっとも左のものが自動的に認識されるので、必要に応じて変更する。
- ⇒ **Weight Field**  
QuaSAR-Model と同様、エントリごとの重み付けが必要な場合があれば、あらかじめ 0~1 の範囲を入力したカラムを用意しておき、ここで設定する。デフォルトは None。
- ⇒ **Eliminate Recessive Gene after Evolution**  
このパラメータが選択されていると、指定した世代の進化が終了した時点で、モデル式集団に一つも含まれていない記述子は、劣性遺伝子とみなし、Descriptors の選択状態を解除する。
- ⇒ **View Equations, View Trajectory**  
これらのオプションを選択すると、該当する出力ファイルが開く。
- ⇒ **Descriptors**  
Database ファイルの中の数値カラムで、Activity 以外の全てフィールド名 (QuaSAR-Evolution のモデルから推算された値のカラムは除く) のリスト。モデル構築に使用する記述子は、このリストの中で選択。なお、新しい Training Set について解析を開始する際は、全記述子が選択された状態でパネルが開くが、既に QuaSAR-Evolution でモデルが構築されたファイルを指定した場合は、保存された式集団で使われている記述子のみが選択された状態となる。
- ⇒ **Essential Desc.**  
式の中に必ず加えたい記述子を指定。上記 Descriptors のリストの中から記述子を選択し、Set ボタンを押すことで Essential Descriptor に登録す

ることができる。削除は、登録記述子を選択し **Remove** を押す。

⇒ **Operant**

モデル式はデフォルトでは1次式で構築されるが、**Operant** 内のファンクションを指定することで二乗、ルート、逆数、**log**などを考慮してモデルを構築することができる。選択状態のファンクションが計算に考慮される。各ファンクションについては、オンラインヘルプ [\\$MOE¥html¥fcnindex.html](#) を参照。

⇒ **Initial Length**

式の母集団を構築する際に、一つの式に含まれる記述子の数。

⇒ **Population**

母集団に含まれるモデル式の数。

⇒ **Operant Density**

オペラントを使用する記述子の比率。この数に1個の割合でオペラントを使用する記述子を選択される。

⇒ **Generation**

世代交代を試みるステップ数。

⇒ **Mutation Probability**

突然変異確率。0.5では、2ステップに1回の割合で、記述子の一つが、両親に含まれていないものと置き換わる可能性がある。

⇒ **Eugenic Factor**

母集団の中の最も優れた式と、最も劣る式が、世代交代に関与する可能性の比率。デフォルトの100では、最も優れた親は、100倍の確率で子供を産む。

⇒ **Auto Termination**

ここで指定したステップ数の間、評価関数の値が母集団の全ての式について変化しない場合は、進化を終了する。

⇒ **Cross Validation**

**Validate** ボタンをクリックすると、指定したモデル式について、**Leave One Out** ないしは **Leave 20% Out** でクロスバリデーションを実行。

なおこの機能は、モデルを構築する前は、使用できない。

**Leave 20% Out** は、テストセットとしてエントリの20%をランダムに選び、他をトレーニングセットとする。全ての組み合わせについて実行すると膨大な計算量になるため、ここでは、500回の平均を出力している。従って、計算するたびに若干結果は異なる。

出力としては、**x\_RR** と **x\_adjRR** が **GA\_\*\*\*\_eq.mdb** に書き出される

⇒ **Activity Calculation**

Calculate ボタンをクリックすると、指定したモデル式による推算値と Z スコアを、指定した Database へ出力する。 カラム名は、\*\*\_eq\_1、Z\_\*\*\_eq\_1 のようになる。 なお、推算の前に、必要な記述子が算出済みでない場合は自動算出される。 Z スコア  $\geq 2.5$  は、外れ値とみなされる。

☐ **Export Model**

Export ボタンをクリックすると、指定したモデル式が、R2 などのデータとともに別ファイルへ出力される。 ‘as’欄で指定したモデル名や Training Set, Test Set で使用したファイル名もあわせて書き出されるので、多数のモデルを整理して保存し、下記の QuaSAR-ModelCalculation パネルから推算に使用するために利用する。

➤ **3-1 計算の実行**

☐ **RUN**

RUN ボタンを押すと計算が開始する。



モデル式集団のファイル ga\_\*\*\*\_eq.mdb とトラジェクトリファイル ga\_\*\*\*\_trj.mdb が作成される。

☐ **CONTINUE**

一旦計算が終了すると、ボタンが下図のようになる。

CONTINUE ボタンを押すと継続した計算を行う。その際に条件の変更も可能。



☐ **Write Script**

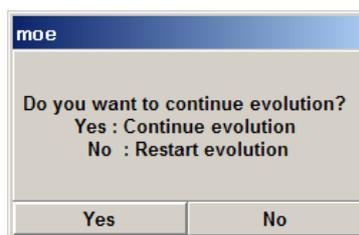
Write Script ボタンを押すと、バッチジョブのスクリプトを出力。

新規の場合、OK ボタンでスクリプトを出力。

一旦計算が終了した状態では、下図のようになり、

Yes で継続計算のスクリプトを出力し、

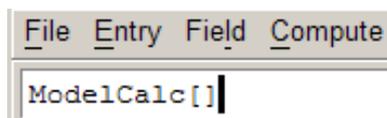
No で新規の計算のスクリプトを出力。



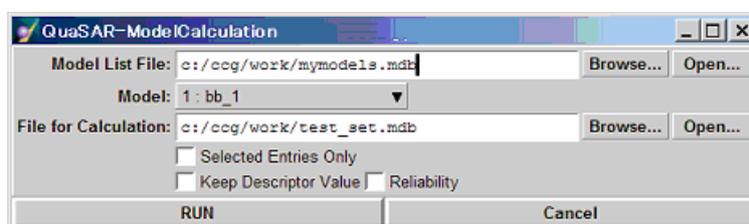
➤ **3-3 QuaSAR-ModelCalculation の起動**

QuaSAR-Evolution から Export された式は、QuaSAR-ModelCalculation を使用することで、分子構造のみを含む.mdb ファイルから直接物性推算を行うことができる。 起動コマンドは、

```
svl> ModelCalc []
```



➤ **3-4 QuaSAR-ModelCalculation パネルの説明**



☐ **Model List File**

QSAR モデルが保存された.mdb ファイル。

QuaSAR-Evolution から出力されたものか、同じフォーマットで式が記述されているもの。 ファイル名カラムの入力の方法は、QuaSAR-Evolution と同様。

☐ **Model**

Model List File で指定したファイルに保存されている QSAR モデル名がリストされているので、使用するモデルを選択。

☐ **File for Calculation**

物性推算する分子構造を含む.mdb ファイル。 記述子は、算出済みでなくとも良い。

- ⇒ **Selected Entries Only**  
このオプションを選択すれば、選択されたエントリのみが処理される。
- ⇒ **Keep Descriptor Value**  
推算時に自動的に算出された記述子を、そのまま保持するか、物性推算後にデータベースから削除するかを指定。
- ⇒ **Reliability**  
トレーニングセットの活性値の範囲に入っているかをチェック。モデルに使用された記述子のうち、範囲に入っている記述子の比率を算出。

➤ **3-5 mymodels.mdb の .fit ファイルへの変換**

.fit ファイルに書き出すためのユーティリティが `mdb2fit`。  
`mymodels.mdb` を開き、変換するエントリを選択、コマンドウインドウにおいて `mdb2fit[]` とタイプする。

## 4. 例題

ここでは、\$MOE/sample/mol/blood\_brain.mdb（出典：J. Chem. Inf. Comput. Sci. 39(2),396-402, (1999)）を例にして一連の操作法の流れを習得し、あわせて GA 法による記述子選択の有効性を検証する。

### データの準備

- ✓ \$MOE/sample/mol/blood\_brain.mdb を開き、作業ディレクトリに保存する。  
以下では、このファイルを bbb.mdb とする。
- ✓ 以下の 51 種の 2D 記述子を算出する。  
SlogP 関連 11 種、SMR 関連 9 種、PEOE 関連 30 種及び TPSA

### Lesson1: QuaSAR-Evolution の基本操作

- ☐ GA[] コマンドで、QuaSAR-Evolution パネルを開く。
- ☐ ここでは、まず TPSA 以外の 50 記述子を使ってモデル化を試みる。  
デフォルトでは全ての記述子が選択された状態になっているので、TPSA のみ選択を解除し、RUN ボタンをクリックする
- ☐ ga\_bbb\_eq.mdb と ga\_bbb\_trj.mdb が重なって開くので、ga\_bbb\_trj.mdb を適当な位置にマウスで移動する
- ☐ ga\_bbb\_trj.mdb のデータ領域をスクロールし、母集団中に存在量が増えつつある記述子を調べる。Plot Fields には、モデル式の集団の中で使用頻度の高い上位 10 個の記述子のプロットを自動的に行われます。
- ☐ ga\_bbb\_eq.mdb の中で、進化の過程を反映して、次第に式が書き換えられ、AIC、LOF などが改善されていくことを確認する。
- ☐ 1000 ステップの間、母集団が改善されなくなった時点で、計算は終了する (End of Evolution と表示)。
- ☐ 計算終了後、ga\_bbb\_eq.mdb をスクロールして観察し、同程度の R<sup>2</sup> でさまざまなモデル式が得られていることを確認する。この例では、最もフィッティングの良いエントリ番号 1 の式では、R<sup>2</sup>=0.71~0.73 程度が得られる。

### Lesson2: モデルの評価と保存

- ☐ 計算が終了した時点で、QuaSAR-Evolution のパネルの Calculate、Validate ボタンがアクティブになっていることを確認。
- ☐ QSAR モデルによる物性推算  
Calculate ボタンをクリックすることで、1 番目のモデル式による推算結果と Z スコアが、bbb.mdb の bbb\_eq\_1、Z\_bbb\_eq\_1 カラムにそれぞれ書き出さ

れる。 ボタン右の欄のモデル式番号を変えることで、他のモデルによる推算値も同様に計算可能。

☐ 結果解析

bbb.mdb で DatabaseViewer | Compute | Analysis | Correlation Plot... コマンドを実行し、logBB と bbb\_eq\_1 の相関をプロットして確認する。 また、bbb.mdb のプロット領域に Z\_ bbb\_eq\_1 をプロットし(DatabaseViewer | Display | Plot を ON にして、Z\_bbb\_eq\_1 をプロット)、外れ値を検出する。(通常は、異常値と考えられるデータがあれば、それを除いて再度モデル化を行う。 この例題では、この過程は省略する。)

☐ Cross Validation

QuaSAR-Evolution パネルの Validate ボタンをクリックすると、ga\_bbb\_eq.mdb 中の 1 番目の式に対して、x\_RR と x\_adjRR が計算され、表に書き出される。

他の式に対しても同様の Cross Validation を行う場合は、the Best のボタンをクリックして対象となるモデルを切り替える。

また、必要に応じて Leave 20% Out も可能。

☐ QSAR モデルの保存

これまでの操作で、集団進化した 100 個の式が ga\_bbb\_eq.mdb に保存されているが、後日、新規化合物の物性予測に利用するのは、通常、最もフィッティングの良いモデルのみであろう。 ここでは、エントリ番号 1 の式を別ファイルに保存しておく。

QuaSAR-Evolution のパネルの as 欄に、bbb\_1 と入力し、Export ボタンをクリックする。

この操作で、mymodels.mdb に該当するモデルの情報が書き出されるので、ファイルを開いて確認する。 Lesson5 では、このファイルに保存された式による新規化合物の物性予測を行う。

### Lesson3: 記述子の追加と進化の継続

☐ TPSA を選択に加え、CONTINUE をクリックする。

先ほどからプロットしている記述子に加えて TPSA をプロットさせると、突然変異によって母集団にこの記述子が加わった後は、次第に式集団全体に広まっていく様子が観察される。 同時に、TPSA に引きずられるように、次第に SMR\_VSA4 も優勢になっていく。

☐ Lesson2 と同様に、クロスバリデーションなどを実行し、TPSA を使用しないモデルに比べて改善されていることを観察する。

この例の SMR\_VSA4 のように、ある記述子が、単独ではさほどモデル化に重

要な要素というわけではなく、他の記述子と組み合わせられることにより優れたモデルが得られるような場合、他のアルゴリズムでは、このようなモデルを見出すことは困難。

- ☐ Lesson2 と同様に、最もフィッティングの良いモデルを **bbb\_2** として、**mymodels.mdb** に書き出す。
- ☐ さらに **Method** を **Variable Length** に変更して **CONTINUE** をクリックする。**TPSA** と **SMR\_VSA4** の組み合わせが重要であることは変わらないが、より多数の記述子を含む式が発生し、**R2** が改善されていることを確認する。
- ☐ 最も良いモデルを、**bbb\_3** として **mymodels.mdb** に書き出す。
- ☐ いったん全てのファイルを閉じ、**QuaSAR-Evolution** を終了する。

#### Lesson4: テストセットを併用したモデル化

オプションとしてテストセットを併用することにより、評価関数とともにテストセットに対する **RMSE** を考慮して進化させることが可能。（ただし、この手法でも、全てのデータをトレーニングセットとした場合よりも、必ずしもさらに良いモデルが得られるという保証はないように思われる。）

- ☐ **bbb.mdb** を開き、エントリ番号 1~61 までを選択する。
- ☐ **QuaSAR-Evolution** を起動する。  
**ga\_bbb\_trj.mdb** と **ga\_bbb\_eq.mdb** が自動的に開き、Lesson3 終了時に **QuaSAR-Evolution** で使用されていた（式に取り込まれていた）記述子のみが選択されていることを確認。
- ☐ 再度、全記述子を選択する。
- ☐ **Unselected Entries in Training Set** オプションを選択。
- ☐ **Test Set** 欄が書き込み禁止状態になるとともに、**Training Set** と同じファイル名が表示されていること、また、**Training Set** の **Selected Entries Only** オプションが選択状態に変わったことに注意。
- ☐ **RUN** ボタンをクリックする。
- ☐ これまでと同様に、式の検証を行い、最もフィッティングの良いモデルを **mymodels.mdb** に書き出す。  
(なお、ここで使用するトレーニングセットの出典となる論文では、61 化合物のうちから、異常値と思われる 3 点を除いて 58 点でモデルを構築している。)
- ☐ **QuaSAR-Evolution** を終了し、ファイルを閉じる。

#### Lesson5: モデル式による物性予測

- ☐ 分子構造のみを含む任意の **.mdb** ファイルを用意する。

- ⇒ ModelCalc [ ]で QuaSAR-ModelCalculation パネルを開く。
- ⇒ Model List File に mymodels.mdb を指定し、Enter キーを押す。
- ⇒ 推算に使用するモデルを、Model オプションから選ぶ。
- ⇒ 計算対象とする.mdb ファイルを File for Calculation 欄で指定する。
- ⇒ RUN ボタンをクリック
- ⇒ 推算結果を確認する。

## 5. その他の機能およびコメント

- ⇒ Initial Length の欄でエンターキーを押すと、Population のデフォルト値が変わる。これは、選択された記述子の数と Initial Length から、母集団中に、各遺伝子（記述子）をもつ個体（式）が、平均 10 程度は存在するような値が選ばれる。（この数値に関して、理論的な裏づけがあるわけではない）
- ⇒ 選択された記述子の数が Initial Length の値と同じであれば、Population を無視して 1 つのモデルだけを出力する。すなわち、単純な MLR を実行する。
- ⇒ モデル化の前に、全てのエントリに同じ値（例えば、全ての分子についてゼロなど）をとる記述子は自動的に除外され、さらに、複数の記述子間で、全エントリにわたって同じ値をとる場合があれば、一つだけが使用される。
- ⇒ Eliminate Recessive Gene....オプションを利用することにより、母集団から完全に消え去った劣性遺伝子は、Descriptors リストの選択状態を自動的に解除することできるので、記述子が非常に多い場合は、例えば 3000 回程度の進化を何回か繰り返すことで、効率的な記述子選択が可能と考えられる。RUN ボタンをクリックすると、母集団の形成から再スタートするが、CONTINUE ボタンを使えば、前回の集団について進化を継続させることができる。
- ⇒ 異常値を除いたモデル化の手順  
Database Viewer の Correlation Plot コマンドを使用して実験値と推算値を比較し、異常値と考えられるデータをグラフ上で選択。DBV | Entry | Invert Entry Selection コマンドで異常値以外のエントリを選択。Selected Entries Only オプションを選択して QuaSAR\_Evolution を実行。なお、Test Set を併用する系でこの操作を行うには、Test Set を別ファイルにしておく必要がある。
- ⇒ 記述子が多い場合は、Fixed Length GA である程度記述子を絞り込み、Variable Length GA で、さらにフィッティングの良いモデルを探すというような流れも考えられる。
- ⇒ これまでの流れが、Global Minimum 近傍を探すものと考えると、最後に、突然変異確率 = 0 で CONTINUE することにより、その範囲の Local Minimum に達するような最終的な収束が可能か。

- ≡ LOF は、モデル内で使用される記述子の数が増えるとペナルティがかかるような評価関数となっているが、ここで使用しているペナルティは、本来の定義とは若干異なる。

$$LOF = \frac{LSE}{\left(1 - \frac{c + dp}{N}\right)^2}$$

ここで、

$C$  … 基底関数の数（定数項を除く）

$p$  … 全ての基底関数に含まれる特性の数の合計

$d$  … *smoothing parameter*

であるが、QuaSAR-Evolution では、常に  $c = p$  であるという前提で

$c + dp = 2p$  とした。なお、 $d$  は任意のパラメータであり、現在の設定では、3～4 項の式が最適となることが多い。例えば、SMR\_VSA0 ～ SMR\_VSA 7 の 8 記述子は、同じ基底関数に含まれると考えられるので、この中の複数が組み込まれた式の LOF は、本来の定義より高めに出力される。したがって、モデル式は、より短いものが優先される可能性が高い。

この定義を厳密に当てはめるためには、記述子の名称だけから、どの記述子が同じ基底関数に含まれているかを判断するルーチンを組む必要があり、かなり煩雑なばかりでなく、一定のルールを組み込んでも、MOE 以外で作成した記述子を併用したモデルを構築する場合などは、正確な判断ができない場合があると予想される。 $d$  の設定で、ある程度任意性が残る評価関数である以上、シンプルな定義にすることを優先した。

LOF は、AIC や自由度修正済み R2 ほど一般的ではなく、LOF による判定が絶対という根拠もないようなので、他の評価関数との併用が望ましいと判断し、このアプリケーションではそれらをあわせて出力することとした。

なお、クロスバリデーションなどで評価した結果、より長い式の方が望ましいと判断された場合は、テストセットを併用して AIC などを評価関数とするか、LOF の係数 2 を小さくすることで、より長い式に収束させることは可能。

## 6. リリースノート

2001/12/17

- リリース

2003/9/24

- 必須の記述子を指定する **Essential Descriptor** オプション追加

2003/12/7

- 予測時にトレーニングセットの活性値の範囲に入っているかを示す **reliability factor** を出力。

2005/10/2

- **.fit** ファイルに書き出すためのユーティリティ **mdb2fit** を追加しました。  
**mymodels.mdb** を開き、変換するエントリを選択、コマンドウインドウにおいて **mdb2fit[]** とタイプします。

2006/1/9

- バッチモードで計算ができるように、GUI からスクリプトが書き出せるようにしました。**Write Script** ボタンで **Yes: Continue No: Run** のバッチを書き出すことができます。

2006/6/12

- トラジェクトリファイルに、**LOF** などの評価関数のその時点における最良モデルの値と全モデルの平均値を書き出すようにしました。
- 重要な記述子については、トラジェクトリファイルのプロット領域に、全モデル中の存在量を、自動的にプロットするようにしました。

2006/3/5

- **Training Set** と異なる **MDB** を **Test Set** と使用した際に、**Expression is not a finite scalar number** のエラーが出力される不具合を解消しました。



<CCG 社日本総代理店>  
株式会社 モルシス  
ライフサイエンス部

Phone: 03-3553-8030

URL: <http://www.molsis.co.jp/>

FAX: 03-3553-8031

E-mail: [support@molsis.co.jp](mailto:support@molsis.co.jp)

---