

The main idea of the Las Vegas algorithm used here [1] is to perform several trials (10 by default, maximum 30 can be implemented in one cycle) of optimization for different splits into an active training set, a passive training set, a calibration set, and a validation set. In this case, the best split for the calibration set is remembered. Below is an approximate scheme for using the Las Vegas algorithm for compounds with some biological activity. *We thank Professor Veselinovic for the idea to use the Las Vegas algorithm for the CORAL software.*

Run the program (CORALSEA.exe) and click where the pointer is

The screenshot displays the CORAL software interface, titled "CORAL: Please load method or make a random split". The interface is divided into several sections:

- Left Panel:** Contains three vertically stacked plots, each with a Y-axis labeled "MODEL" and an X-axis labeled "EXPERIMENT". The plots are labeled "Active Training set (A)", "Passive Training set (P)", and "Calibration set (C)".
- Top Center:** A button labeled "Load method" next to a text field "Method.txt".
- Center:** A large blue arrow points upwards from a button labeled "Las Vegas algorithm" (with a value of 10) to a button labeled "Split #TotalSet.txt into training and validation sets".
- Right Panel:** Contains various settings and options:
 - SMILES/quasi-SMILES on A,P,C sets:** Includes checkboxes for GRAPH, HSG, HFG, GA0, R3, R4, R5, R6, R7.
 - SMILES:** Includes checkboxes for Chaos, s, ss, sss, BOND, NOSP, HALO, HARD.
 - Atoms pairs proportions:** A table for Cl, Br, N, O, S, P, =, #.
 - Contributions:** Includes checkboxes for N, O, S, F, Cl, Br, =, #, xyx, xyxx, xyzyx.
 - Classification model:** Includes checkboxes for CLASSIC SCHEME and CORRELATION BALANCE.
 - Weights:** Includes sliders for dR weight, IIC weight, CII weight, CCCP weight.
 - Thresholds:** Includes sliders for Threshold start, Threshold maximal, and Number of probes.
 - Model Details.txt:** A text field containing "So Las Vegas June3,2024".
 - Buttons:** Includes "STOP and SAVE current CWs", "DemoDCW", "EvolutionCorr", and "EXIT".
- Bottom:** Includes checkboxes for "Search for duplicates in SMILES" and "Search for duplicates in CAS (ID)", and a table for "Split Info" with columns W%, N111, N110, N101, N100, Nall, Defect, and Density.

click where the pointer is

The screenshot displays the CORAL software interface. A dialog box titled "Preparation of split April 13, 2022" is open, showing instructions and a workflow for data splitting. The main application window is visible in the background, showing various settings and a data table.

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

1. **Input**: #TotalSet.txt **Load** **Output**: #TotalSet'.txt

2. **Do distribution**

3. **Save files**

Dispersion Limit = 0.01

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.25	0	0
Invisible training set	-	0.25	0	0
Calibration set	#	0.25	0	0
#ValidationSet.txt	*	0.25	0	0

Loading of vector of Distribution Save vector of Distribution EXIT

Model Details.txt So Las Vegas June 3, 2024

Number of promoters increase/decrease

STOP and SAVE current CW's DemoDCW EvolutionCorr EXIT

click where the pointer is

CORAL: Please load method or make a random split

Don't show optimization

MODE

Load method

Method.txt

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

Input Output

+1. #TotalSet.txt Load #TotalSet.txt N = 295

2. Do distribution

3. Save files

Dispersion Limit = 0.005

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	0	0
Invisible training set	-	0.250	0	0
Calibration set	#	0.250	0	0
#ValidationSet.txt	*	0.250	0	0

Loading of vector of Distribution Save vector of Distribution EXIT

to training and validation sets

s algorithm 10

SMILES/quasi-SMILES on A,P,C sets

CORALSEA-2024v

GRAPH HSG HFG GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

R3 R4 R5 R6 R7

SMILES Chaos s ss sss BOND NOSP HALO HARD

Atoms pairs proportions Cl Br N O S P = #

F. Cl. Br. N. O. S. P. =

Contributions

N O S F Cl Br = # xyx xyxx xyzyx

Classification model

CLASSIC SCHEME

CORRELATION BALANCE

Index of Ideality of Correlation

Correlation Intensity Index

Conformism coefficient

dR weight IIC weight CII weight CCCP weight

d start d limit Nepoch

Threshold start Threshold maximal Number of probes

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease

STOP and SAVE current CWs DemoDCW EvolutionCorr EXIT

Search for duplicates in SMILES Search for duplicates in CAS (ID)

W% N111 N110 N101 N100 Nall Defect Density

Split Info 0 0 0 0 0 0 0 0

click where the pointer is

CORAL: Please load method or make a random split

Don't show optimization

MODE ↑

Load method

Method.txt

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

Input Output

+1. #TotalSet.txt Load #TotalSet.txt N = 295

+2. Do distribution

3. Save files

Dispersion Limit = 0.005

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	74	0.251
Invisible training set	-	0.250	75	0.254
Calibration set	#	0.250	73	0.247
#ValidationSet.txt	*	0.250	73	0.247

Loading of vector of Distribution Save vector of Distribution EXIT

to training and validation sets

s algorithm 10

SMILES/quasi-SMILES on A,P,C sets

CORALSEA-2024v

GRAPH HSG HFG GAO SMILES

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn R3 R4 R5 R6 R7

Chaos Chaos s ss sss BOND NOSP HALO HARD

Atoms pairs proportions Cl Br N O S P = #

Contributions

N O S F Cl Br = # xyx xyxx xyzyx

Classification model CLASSIC SCHEME CORRELATION BALANCE

Index of Ideality of Correlation IICweight

Correlation Intensity Index CIIweight

Conformism coefficient CCCPweight

d_start d_limit Nepoch Threshold start Threshold maximal Number of probes

Model Details.txt So Las Vegas June 3, 2024

Number of promoters increase/decrease

STOP and SAVE current CWs DemoDCW EvolutionCorr EXIT

Search for duplicates in SMILES Search for duplicates in CAS (ID)

click where the pointer is

The screenshot displays the CORAL software interface. A dialog box titled "Preparation of split April 13, 2022" is open, showing the steps for preparing the data split. The main application window is visible in the background, showing various settings and a table of results.

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

Input: #TotalSet.txt Load Output: #TotalSet'.txt N = 295

+2. Do distribution

+3. Save files

Dispersion Limit = 0.005

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	74	0.251
Invisible training set	-	0.250	75	0.254
Calibration set	#	0.250	73	0.247
#ValidationSet.txt	*	0.250	73	0.247

Loading of vector of Distribution Save vector of Distribution EXIT

SMILES/quasi-SMILES on A,P,C sets

GRAPH HSG HFG GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nm

R3 R4 R5 R6 R7

Atoms pairs proportions Cl Br N O S P = #

Contributions

N O S F Cl Br = # xyx xyyx xyzyx

Classification model CLASSIC SCHEME CORRELATION BALANCE

Index of Ideality of Correlation Correlation Intensity Index Conformism coefficient

dR weight IIC weight CII weight CCCP weight

D_start d_limit Nepoch Threshold start Threshold maximal Number of probes


Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease

STOP and SAVE current CW/s DemoDCW EvolutionCort EXIT

Search for duplicates in SMILES Search for duplicates in CAS (ID)

click where the pointer is

 CORAL: select Phase (1 or 2); or change and save method; or import current model

Show optimization

DotSize-1

MODEL

Y↑

X→

Active Training set (A)

EXPERIMENT

...

MODEL

Y↑

X→

Passive Training set (P)

EXPERIMENT

...

MODEL

Y↑

X→

Calibration set (C)

EXPERIMENT

...

Load method

Save method

Method.txt

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Las Vegas algorithm

10

Import of current model

W%	N111	N110	N101	N100	Nall	Defect	Density
Split Info	0	0	0	0	0	0	0

☒ Search for duplicates in SMILES
 ☐ Search for duplicates in CAS (ID)

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH
 ☐ HSG
 ☐ HFG
 ☐ GAO

☒ SMILES
 ☐ Chaos
 ☒ s
 ☐ Chaos
 ☒ ss
 ☐ BOND
 ☐ NOSP
 ☐ HALO
 ☐ HARD

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

e0									
e1									
e2									
e3									
p2									
p3									
p4									
s2									
s3									
nn									

R3 R4 R5 R6 R7

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F...	0	0	0	0	0	0	0	0
Cl...	0	0	0	0	0	0	0	0
Br...	0	0	0	0	0	0	0	0
N...	0	0	0	0	0	0	0	0
O...	0	0	0	0	0	0	0	0
S...	0	0	0	0	0	0	0	0
P...	0	0	0	0	0	0	0	0
=...	0	0	0	0	0	0	0	0

Contributions

☐ N
 ☐ O
 ☐ S
 ☐ F
 ☐ Cl
 ☐ Br
 ☐ =
 ☐ #
 ☐ xyx
 ☐ xyxx
 ☐ xyzyx

☐ Classification model
 ☐ CLASSIC SCHEME
 ☒ CORRELATION BALANCE

dR weight 0.1

☐ Index of Ideality of Correlation
 ☐ Correlation Intensity Index
 ☒ Conformism coefficient

CCCP_C 0.25 0

D_{start} 0.5

d_{limit} 0.1

N_{epoch} 15

Threshold start 3

Threshold maximal 3

Number of probes 1

Model Details.txt


So Las Vegas June3,2024

Number of promoters increase/decrease 15

STOP and SAVE current CWs

☒ DemoDCW
 ☒ EvolutionCorr

EXIT

 CORAL: select Phase (1 or 2); or change and save method; or import current model

Show optimization DotSize-1

MODEL \uparrow Y

Active Training set (A) \rightarrow X EXPERIMENT

...

MODEL \uparrow Y

Passive Training set (P) \rightarrow X EXPERIMENT

...

MODEL \uparrow Y

Calibration set (C) \rightarrow X EXPERIMENT

...

Load method Save method Method.txt

Split #TotalSet.txt into training and validation sets

Start Las Vegas Calculations 10

APCV style

Import of current model

W%	N111	N110	N101	N100	Nall	Defect	Density
Split Info	0	0	0	0	0	0	0

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

SMILES/quasi-SMILES on A,P,C sets

#TotalSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

R3 R4 R5 R6 R7

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE dR weight 0.1

☐ Index of Ideality of Correlation

☐ Correlation Intensity Index

☒ Conformism coefficient CCCP_C 0.25 0

D_{start} 0.5 d_{limit} 0.1 N_{epoch} 15

Threshold start 3

Threshold maximal 3

Number of probes 10

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease 15

☒ DemoDCW ☒ EvolutionCorr

EXIT

click where the pointer is

Please wait until the calculation is finished

click where the pointer is

CORAL: Wait please...

Show optimization | DotSize=1

Active Training set (A) | EXPERIMENT

Passive Training set (P) | EXPERIMENT

Calibration set (C) | EXPERIMENT

Load method | Save method | Method.txt

Status of the calculations:

Threshold.....3 to 3
Epoch.....16 to 15
Probe.....2 to 10

Las Vegas Calculations in process | 10

APCV style

SMILES/quasi-SMILES on A,P,C sets

#TotalSet'.txt

☐ GRAPH ☐ HSG ☐ HFG ☒ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

R3 R4 R5 R6 R7

☒ SMILES

☐ Chaos ☒ s

☐ Chaos ☒ ss

☐ BOND

☐ NOSP

☐ HALO

☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F.....	0	0	0	0	0	0	0	0
Cl.....	0	0	0	0	0	0	0	0
Br.....	0	0	0	0	0	0	0	0
N.....	0	0	0	0	0	0	0	0
O.....	0	0	0	0	0	0	0	0
S.....	0	0	0	0	0	0	0	0
P.....	0	0	0	0	0	0	0	0
=.....	0	0	0	0	0	0	0	0

All

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyyx ☐ xyzyx

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE dR weight 0.1

☐ Index of Ideality of Correlation

☐ Correlation Intensity Index

☒ Conformism coefficient CCCPC 0.25 0

D_start 0.5 d_limit 0.1 N_epoch 15

Threshold start 3

Threshold maximal 3

Number of probes 10

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease 15

☒ DemoDCW

☒ EvolutionCorr

EXIT

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
98	54	1	0	0	55	45.22	0.556

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

click where the pointer is
click where the pointer is

click where the pointer is

CORAL: Wait please...

Show optimization **DotSize-1**

MODEL Y↑

Active Training set (A) EXPERIMENT X

MODEL Y↑

Passive Training set (P) EXPERIMENT X

MODEL Y↑

Calibration set (C) EXPERIMENT X

Load method Save method Method.txt

Status of the calculations:

Threshold 3 to 3
THE CALCULATION IS COMPLETED
Probe 10 to 10

Las Vegas Calculations in process 10

APCV style

Coralsea

#TotalSetLV.txt contains the split recommended by Las Vegas algorithm

OK

SMILES/quasi-SMILES on A,P,C sets

#TotalSetLV.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO ☒ SMILES

☐ Chaos ☒ s

☐ Chaos ☒ ss

☐ sss

☐ BOND

☐ NOSP

☐ HALO

☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE

dR weight 0.1

Correlation Index

cient CCCPC 0.25 0

mit 0.1 Nepoch 15

Threshold start 3

Threshold maximal 3

Number of probes 10

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease 15

☒ DemoDCW

☒ EvolutionCorr

EXIT

W% N111 N110 N101 N100 Nall Defect Density

Split Info 100 50 0 0 0 50 49.14 0.0000

☒ Search for duplicates in SMILES

☐ Search for duplicates in CAS (ID)

click where the pointer is

run the program and click where the pointer is

click where the pointer is

CORAL: Please load method or make a random split

☐ Don't show optimization

MODEL \uparrow Y

Active Training set (A) \rightarrow X EXPERIMENT

...

Load method

Method.txt

MODEL \uparrow Y

Passive Training set (P) \rightarrow X EXPERIMENT

...

Split #TotalSet.txt into training and validation sets

MODEL \uparrow Y

Calibration set (C) \rightarrow X EXPERIMENT

...

Las Vegas algorithm

10

SMILES/quasi-SMILES on A,P,C sets

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

e0									
e1									
e2									
e3									
p2									
p3									
p4									
s2									
s3									
nn									

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyxx ☐ xyzyx

☐ Chaos ☐ s

☐ Chaos ☐ ss

☐ BOND

☐ NOSP

☐ HALO

☐ HARD

Atoms pairs proportions

Cl Br N O S P = #

F...					
Cl...					
Br...					
N...					
O...					
S...					
P...					
=					

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE

☐ Index of Ideality of Correlation

☐ Correlation Intensity Index

☐ Conformism coefficient

d_start d_limit Nepoch

Threshold start

Threshold maximal

Number of probes

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease

☐ Search for duplicates in SMILES

☐ Search for duplicates in CAS (ID)

☐ STOP and SAVE current CWs

☐ DemoDCW

☐ EvolutionCort

EXIT

click where the pointer is

add 'LV' to the file name and click where the pointer is

click where the pointer is

CORAL: Please load method or make a random split

Don't show optimization

MODE ↑

Load method Method.txt

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.

The second action should be "Do distribution".

The third action should be "Save files".

Input Output

1. #TotalSetLV.txt Load #TotalSet.txt

2. Do distribution

3. Save files

Dispersion Limit = 0.01

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.25	0	0
Invisible training set	-	0.25	0	0
Calibration set	#	0.25	0	0
#ValidationSet.txt	*	0.25	0	0

Loading of vector of Distribution Save vector of Distribution EXIT

to training and validation sets

s algorithm 10

SMILES/quasi-SMILES on A,P,C sets

CORALSEA-2024lv

GRAPH HSG HFG GAO SMILES

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn R3 R4 R5 R6 R7

Chaos Chaos s ss sss BOND NOSP HALO HARD

Atoms pairs proportions Cl Br N O S P = #

F... Cl... Br... N... O... S... P... =...

Contributions

N O S F Cl Br = # xyx xyxx xyzyx

Classification model

CLASSIC SCHEME

CORRELATION BALANCE dR weight IIC weight CII weight CCCP weight

Index of Ideality of Correlation

Correlation Intensity Index

Conformism coefficient

D_start d_limit Nepoch Threshold start Threshold maximal Number of probes

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease

STOP and SAVE current CW's DemoDCW EvolutionCorr EXIT

W% N111 N110 N101 N100 Nall Defect Density

Split Info 0 0 0 0 0 0 0 0

Search for duplicates in SMILES Search for duplicates in CAS (ID)

click where the pointer is

The screenshot displays the CORAL software interface. A dialog box titled "Preparation of split April 13, 2022" is open, showing instructions and a table for the planned distribution of training and validation sets. The main application window in the background shows various settings for SMILES/quasi-SMILES on A,P,C sets, including checkboxes for different features and a table for atom pairs proportions.

Preparation of split April 13, 2022

The first action is "Load". File "#TotalSet.txt" must exist in your folder.
 The second action should be "Do distribution".
 The third action should be "Save files".

Input: #TotalSetLV.t Load Output: #TotalSet'.txt N = 295

2. Do distribution

3. Save files

Dispersion Limit = 0.005

		Planned Distribution	Frequency	Actual Distribution
#TrainingSet.txt	+	0.250	74	0.251
Invisible training set	-	0.250	75	0.254
Calibration set	#	0.250	73	0.247
#ValidationSet.txt	*	0.250	73	0.247

Loading of vector of Distribution Save vector of Distribution EXIT

SMILES/quasi-SMILES on A,P,C sets

GRAPH HSG HFG GAO SMILES
 e0 e1 e2 e3 p2 p3 p4 s2 s3 mn R3
 e0 R4
 e1 R5
 e2 R6
 e3 R7
 p2
 p3
 p4
 s2
 s3
 mn

Atoms pairs proportions
 Cl Br N O S P = #

Contributions
 N O S
 F Cl Br
 = #
 xyx xyx xyzx

Classification model
 CLASSIC SCHEME
 CORRELATION BALANCE dR weight
 Index of Ideality of Correlation IIC weight
 Correlation Intensity Index CII weight
 Conformism coefficient CCCP weight

D_start d_limit Nepoch
 Threshold start
 Threshold maximal
 Number of probes

Model Details.txt So Las Vegas June 3, 2024

Number of promoters increase/decrease

STOP and SAVE current CWs DemoDCW
 EvolutionCorr EXIT

Search for duplicates in SMILES Search for duplicates in CAS (ID)

click where the pointer is

 CORAL: select Phase (1 or 2); or change and save method; or import current model

MODEL

Y↑

X→

EXPERIMENT

Show optimization

DotSize-1

Active Training set (A)

...

MODEL

Y↑

X→

EXPERIMENT

Passive Training set (P)

...

MODEL

Y↑

X→

EXPERIMENT

Calibration set (C)

...

Load method

Save method

Method.txt

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Las Vegas algorithm

10

Import of current model

W%	N111	N110	N101	N100	Nall	Defect	Density
Split Info	0	0	0	0	0	0	0

☒ Search for duplicates in SMILES
 ☐ Search for duplicates in CAS (ID)

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH
 ☐ HSG
 ☐ HFG
 ☐ GAO
 ☒ SMILES

☐ Chaos
 ☒ s
 ☐ Chaos
 ☒ ss
 ☐ BOND
 ☐ NOSP
 ☐ HALO
 ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F..	0	0	0	0	0	0	0	0
Cl...	0	0	0	0	0	0	0	0
Br.....	0	0	0	0	0	0	0	0
N.....	0	0	0	0	0	0	0	0
O.....	0	0	0	0	0	0	0	0
S.....	0	0	0	0	0	0	0	0
P.....	0	0	0	0	0	0	0	0
=.....	0	0	0	0	0	0	0	0

Contributions

☐ N
 ☐ O
 ☐ S
 ☐ F
 ☐ Cl
 ☐ Br
 ☐ =
 ☐ #
 ☐ xyx
 ☐ xyxx
 ☐ xyzyx

☐ Classification model
 ☒ CLASSIC SCHEME
 ☒ CORRELATION BALANCE
 ☐ Index of Ideality of Correlation
 ☐ Correlation Intensity Index
 ☒ Conformism coefficient
 ☐ CCCP_C

dR weight

0.1

D_{start}

0.5

d_{limit}

0.1

Nepoch

15

Threshold start

3

Threshold maximal

3

Number of probes

10

Model Details.txt

So Las Vegas June3,2024

Number of promoters increase/decrease

15

STOP and SAVE current CW's

☒ DemoDCW
 ☒ EvolutionCorr

EXIT

click where the pointer is

CORAL: Building up preferable model

Show optimization **DotSize-1**

Active Training set (A) **EXPERIMENT**

Passive Training set (P) **EXPERIMENT**

Calibration set (C) **EXPERIMENT**

Phase 2: Building up preferable model (T*,N*)

Define threshold and press Continue **3** **Continue**

Import of current model

	W%	N111	N110	N101	N100	Nall	Defect	Density
Split Info	0	0	0	0	0	0	0	0

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ **GRAPH** ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nm

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☐ Chaos ☒ **SMILES** ☐ Chaos ☒ **ss** ☐ Chaos ☒ **sss** ☐ **BOND** ☐ **NOSP** ☐ **HALO** ☐ **HARD**

Atoms pairs proportions

	Cl	Br	N	O	S	P	#
F..	0	0	0	0	0	0	0
Cl...	0	0	0	0	0	0	0
Br.....	0	0	0	0	0	0	0
N.....	0	0	0	0	0	0	0
O.....	0	0	0	0	0	0	0
S.....	0	0	0	0	0	0	0
P.....	0	0	0	0	0	0	0
=.....	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ = ☐ # ☐ xyx ☐ xyxx ☐ xyzxx

☐ **Classification model** ☐ **CLASSIC SCHEME** ☒ **CORRELATION BALANCE**

☐ Index of Ideality of Correlation ☐ Correlation Intensity Index ☒ **Conformism coefficient** **CCCC_C** 0,25 0

dR weight 0,1

D_{start} 0,5 **d_{limit}** 0,1 **Nepoch** 15

Threshold start 3 **Threshold maximal** 3 **Number of probes** 10

Model Details.txt **So Las Vegas June3,2024**

Number of promoters increase/decrease 15

STOP and SAVE current CWs ☒ **DemoDCW** ☒ **EvolutionCorr** **EXIT**

Please wait until the calculation is finished and click where the pointer is

CORAL: you should save model now

Show optimization DotSize-1

MODEL Y↑

Active Training set (A) EXPERIMENT X

Phase 2: Building up preferable model (T*,N*)

Define threshold and press Continue 3 Continue

Selected threshold is 3
THE CALCULATION IS COMPLETED

MODEL Y↑

Passive Training set (P) EXPERIMENT X

MODEL Y↑

Calibration set (C) EXPERIMENT X

C0 = -2.8843809 C1 = 0.3947727

Insert a SMILES for calculation of DCW and EndPoint

Demo of calculation of DCW and endpoint will be in file

Start of DCW and Endpoint calculation for SMILES from file

Save Model

Split Info

W%	N111	N110	N101	N100	Nall	Defect	Density
100	51	0	0	0	51	43.21	0.543

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO ☒ SMILES

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

☐ R3 ☐ Chaos ☒ s

☐ R4 ☐ Chaos ☒ ss

☐ R5 ☐ sss

☐ R6 ☐ BOND

☐ R7 ☐ NOSP

☐ HALO

☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0
#	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE dR weight 0.1

☐ Index of Ideality of Correlation

☐ Correlation Intensity Index

☒ Conformism coefficient CCCP_C 0.25 0

D start 0.5 d limit 0.1 Nepoch 15

DemoDCW.txt Threshold start 3

Threshold maximal 3

Number of probes 1

#ValidationSet.txt #ModelForValidationSet.txt

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease: 15

☒ DemoDCW

☒ EvolutionCorr

STOP and SAVE current CWs

EXIT

click where the pointer is twice 1.Load method; and 2 Import of current model

CORAL: select Phase (1 or 2); or change and save method; or import current model

Show optimization **DotSize-1**

MODEL **Y** **X** **EXPERIMENT**

Active Training set (A)

MODEL **Y** **X** **EXPERIMENT**

Passive Training set (P)

MODEL **Y** **X** **EXPERIMENT**

Calibration set (C)

Load method **Save method** **Method.txt**

Phase 1: Search for preferable model (T^*, N^*)

Phase 2: Building up preferable model (T^*, N^*)

Las Vegas algorithm **10**

Import of current model

Split Info	W%	N111	N110	N101	N100	Nall	Defect	Density
	0	0	0	0	0	0	0	0

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

CORALSEA-2024lv

☐ **GRAPH**
☐ HSG
 ☐ HFG
 ☐ GAO

	e0	e1	e2	e3	p2	p3	p4	s2	s3	nn
e0	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e1	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
e3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
p2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
p3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
p4	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
s2	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
s3	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
nn	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Contributions

☐ N
 ☐ O
 ☐ S
☐ F
 ☐ Cl
 ☐ Br
☐ =
 ☐ #
☐ xyx
 ☐ xyxx
 ☐ xyzyx

☐ **Classification model**
☐ CLASSIC SCHEME
☒ **CORRELATION BALANCE**

☐ Index of Ideality of Correlation
☐ Correlation Intensity Index
☒ Conformism coefficient

CCCP_C

D_{start}
 d_{limit}
 N_{epoch}

Threshold start
 Threshold maximal
 Number of probes

☒ **SMILES**
☐ Chaos
 ☒ \$
☐ Chaos
 ☒ ss
☐ sss
☐ **BOND**
☐ **NOSP**
☐ **HALO**
☐ **HARD**

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F.....	0	0	0	0	0	0	0	0
Cl.....	0	0	0	0	0	0	0	0
Br.....	0	0	0	0	0	0	0	0
N.....	0	0	0	0	0	0	0	0
O.....	0	0	0	0	0	0	0	0
S.....	0	0	0	0	0	0	0	0
P.....	0	0	0	0	0	0	0	0
=.....	0	0	0	0	0	0	0	0

All

dR weight

Model Details.txt

So Las Vegas June3,2024

Number of promoters increase/decrease

STOP and SAVE current CWs

☒ DemoDCW
☒ EvolutionCorr

EXIT

click where the pointer is to check model on validation set

CORAL: Calculation of model for external substances

Hide DotSize-1

Active Training set (A) EXPERIMENT
n=74: R2=0.5860: s=1.37: MAE=1.19: F=102

Passive Training set (P) EXPERIMENT
n=75: R2=0.6976: s=1.15: MAE=0.983: F=168

Calibration set (C) EXPERIMENT
n=73: R2=0.8655: s=0.708: MAE=0.551: F=457

XY-dots analysis **Marker-0**

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets.

Phase 2: Building up preferable model (T*,N*)

Calculation model for a list of SMILES

list.txt ListModel.txt

C0 = -2.8843809 C1 = 0.3947727

Insert a SMILES for calculation of DCW and EndPoint

Demo of calculation of DCW and endpoint will be in file

DCW(3,15)= EndPoint =

Start of DCW and Endpoint calculation for SMILES from file

Import of current model

W% N111 N110 N101 N100 Nall Defect Density

Split Info 100 51 0 0 0 51 43.21 0.543

Continue optimization

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 mn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES

☐ Chaos ☒ s

☐ Chaos ☒ ss

☐ BOND

☐ NOSP

☐ HALO

☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F	0	0	0	0	0	0	0	0
Cl	0	0	0	0	0	0	0	0
Br	0	0	0	0	0	0	0	0
N	0	0	0	0	0	0	0	0
O	0	0	0	0	0	0	0	0
S	0	0	0	0	0	0	0	0
P	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0
#	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE dR weight 0.1

☐ Index of Ideality of Correlation

☐ Correlation Intensity Index

☒ Conformism coefficient CCCPC 0.25 0

D_start 0.5 d_limit 0.1 Nepoch 15

DemoDCW.txt Threshold start 3

Threshold maximal 3

Number of probes 1

#ValidationSet.txt #ModelForValidationSet.txt

Model Details.txt So Las Vegas June 3, 2024

Number of promoters increase/decrease 15

☒ DemoDCW

☒ EvolutionCorr

EXIT

STOP and SAVE current CWs

Please see results for validation set in file .../model/#ModelForValidationSet.txt; Now you can study plots "expr vs calc"

Hide

DotSize-1

Active Training set (A) EXPERIMENT

MODEL Y

n=74: R2=0.5860; s=1.37; MAE=1.19; F=102

Passive Training set (P) EXPERIMENT

MODEL Y

n=75: R2=0.6976; s=1.15; MAE=0.983; F=168

Calibration set (C) EXPERIMENT

MODEL Y

n=73: R2=0.8655; s=0.708; MAE=0.551; F=457

Validation set (V) EXPERIMENT

MODEL Y

n=73: R2=0.8377; s=0.776; MAE=0.589; F=367

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Calculation model for a list of SMILES

list.txt ListModel.txt

C0 = -2.8843809 C1 = 0.3947727

Insert a SMILES for calculation of DCW and EndPoint

c1ccc([N])(=O)O)c(c1)Cl x

Demo of calculation of DCW and endpoint will be in file

DCW(3.15)= 14.2090 EndPoint =

Start of DCW and Endpoint calculation for SMILES from file

Import of current model

W%	N111	N110	N101	N100	Nall	Defect	Density
100	51	0	0	0	51	43.21	0.543

Split Info

Search for duplicates in SMILES ☒ Search for duplicates in CAS (ID) ☐

Continue optimization

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☐ GRAPH ☐ HSG ☐ HFG ☐ GAO

e0 e1 e2 e3 p2 p3 p4 s2 s3 nn

☐ R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☒ SMILES ☐ Chaos ☐ Chaos ☐ s ☐ ss ☐ sss ☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F...	0	0	0	0	0	0	0	0
Cl...	0	0	0	0	0	0	0	0
Br...	0	0	0	0	0	0	0	0
N...	0	0	0	0	0	0	0	0
O...	0	0	0	0	0	0	0	0
S...	0	0	0	0	0	0	0	0
P...	0	0	0	0	0	0	0	0
=	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S ☐ F ☐ Cl ☐ Br ☐ = ☐ # ☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model ☒ CLASSIC SCHEME ☒ CORRELATION BALANCE dR weight 0.1

☐ Index of Ideality of Correlation ☐ Correlation Intensity Index

☒ Conformism coefficient CCCPc 0.25 0

D_start 0.5 d_limit 0.1 Nepoch 15

DemoDCW.txt Threshold start 3

2.7249 Threshold maximal 3

#ValidationSet.txt Number of probes 1

#ModelForValidationSet.txt

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease 15

☒ DemoDCW ☒ EvolutionCorr

STOP and SAVE current CWs

EXIT

Reference

1. Toropova A.P., Toropov A.A. The coefficient of conformism of a correlative prediction (CCCP): Building up reliable nano-QSPRs/QSARs for endpoints of nanoparticles in different experimental conditions encoded via quasi-SMILES (2024) Science of the Total Environment, 927, art. no. 172119, DOI: 10.1016/j.scitotenv.2024.172119

December 13,2024

There are two changes in this version of the program.

First, the Las Vegas process is controlled by two parameters: the number of samples (tests) and the percentage of object movements between the active training, passive training, calibration and validation sets.

CORAL: select Phase (1 or 2); or change and save method; or import current model

Show optimization DotSize-1

MODEL Y↑

Active Training set (A) EXPERIMENT

MODEL Y↑

Passive Training set (P) EXPERIMENT

MODEL Y↑

Calibration set (C) EXPERIMENT

Load method Save method Method.txt

Phase 1: Search for preferable model (T*,N*)

Split #TotalSet.txt into training and validation sets

Phase 2: Building up preferable model (T*,N*)

Las Vegas algorithm

Ten tests

50%

SMILES/quasi-SMILES on A,P,C sets

#TrainingSet.txt

☒ GRAPH ☒ HSG ☐ HFG ☐ GAO

e0 ☒ e1 ☐ e2 ☐ e3 ☐ p2 ☐ p3 ☐ p4 ☐ s2 ☐ s3 ☐ nn

R3 ☐ R4 ☐ R5 ☐ R6 ☐ R7

☐ Chaos ☒ s ☐ Chaos ☒ ss ☐ sss

☐ BOND ☐ NOSP ☐ HALO ☐ HARD

Atoms pairs proportions

	Cl	Br	N	O	S	P	=	#
F...	0	0	0	0	0	0	0	0
Cl...	0	0	0	0	0	0	0	0
Br...	0	0	0	0	0	0	0	0
N...	0	0	0	0	0	0	0	0
O...	0	0	0	0	0	0	0	0
S...	0	0	0	0	0	0	0	0
P...	0	0	0	0	0	0	0	0
=...	0	0	0	0	0	0	0	0

Contributions

☐ N ☐ O ☐ S

☐ F ☐ Cl ☐ Br

☐ = ☐ #

☐ xyx ☐ xyxx ☐ xyzyx

☐ Classification model

☐ CLASSIC SCHEME

☒ CORRELATION BALANCE

☒ Index of Ideality of Correlation

☒ Correlation Intensity Index

☒ Conformism coefficient

dR weight 0.1

IIC_C 0.3 0

CIIC_C 0.3 0

CCCP_C 0.3 0

D_start 0.5 d_limit 0.1 Nepoch 15

Threshold start 3

Threshold maximal 3

Number of probes 1

Model Details.txt So Las Vegas June3,2024

Number of promoters increase/decrease 15

STOP and SAVE current CW's

☒ DemoDCW

☒ EvolutionCorr

EXIT

W% N111 N110 N101 N100 Nall Defect Density

Split Info 0 0 0 0 0 0 0 0

☒ Search for duplicates in SMILES ☐ Search for duplicates in CAS (ID)

Second, the data on correlation weights are now presented in two files, for example, Ws3-1.txt and Wv3-1.txt instead of the previously used w3-1.txt.

Ws3-1.txt this is statistics, that is, the distribution by sets and the corresponding development into active and passive attributes;

Wv3-1.txt these are directly the numerical values of the correlation weights for calculating the optimal descriptor