Euclidean-Distance 1.0
(ED)

NANOBRIDGES
-ACollaborative Project

UNIWERSYTET GDAŃSKI

Marie Curie Actions
EUROPEAN UNION

THE AUTHORS ARE GRATEFUL FOR THE FINANCIAL SUPPORT FROM THE EUROPEAN COMMISSION THROUGH THE MARIE CURIE **IRSES** PROGRAM, NANOBRIDGES PROJECT (FP7-PEOPLE-2011-IRSES, GRANT AGREEMENT NUMBER 295128).

Last updated: 22.05.2015

Euclidean-Distance 1.0

Background: The multivariate chemometrical analysis involves the measurement of distances between objects and variables. The two widely and most commonly used distance measures are Euclidean distance (ED) [1] and the Mahalanobis distance (MD) [2]. The ED is easy to compute and interpret but the calculation of MD takes into account the correlation in the data, since it is calculated by using the inverse of the variance-covariance matrix of the dataset.

About the algorithm: The Euclidean distance or Euclidean metric is the ordinary distance between two points in the Euclidean space. The equations involved in the calculation of Euclidean distance are as follows:

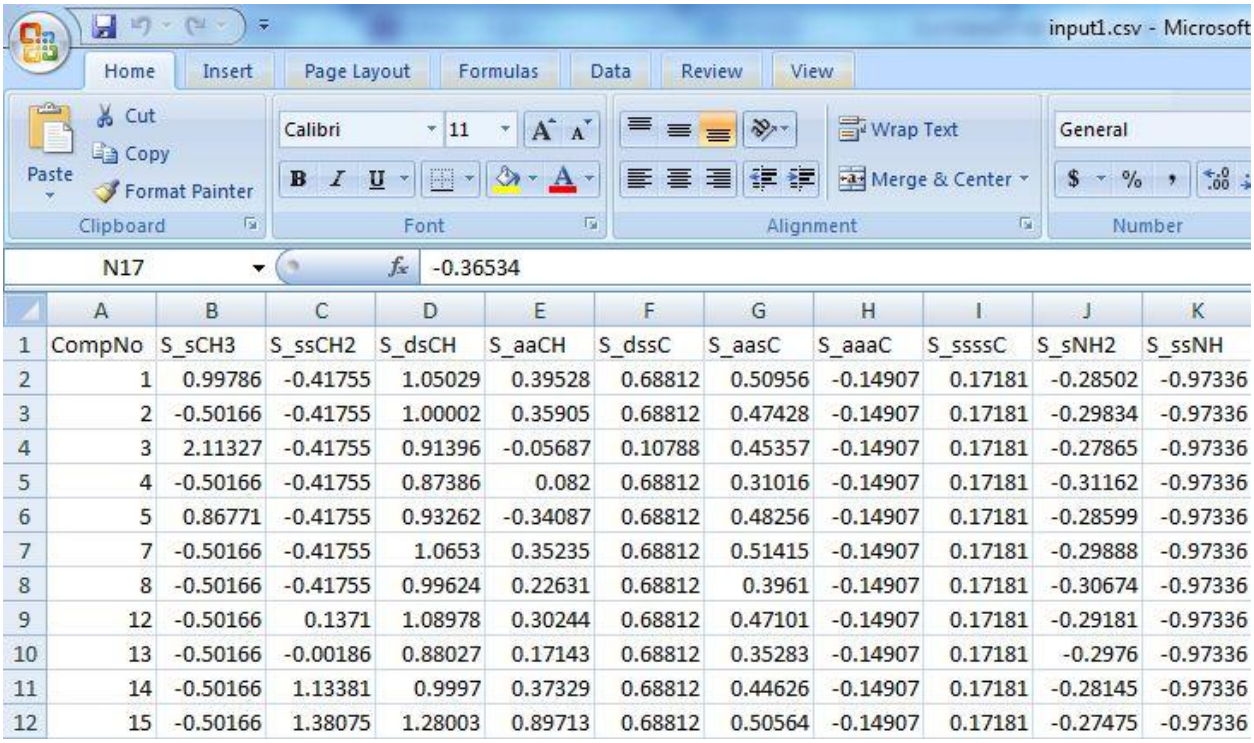
$$d_{ij} = \left\| X_i - X_j \right\| = \sqrt{\sum_{k=1}^{m} (X_{ik} - X_{jk})^2} \qquad \text{(i)}$$

$$\overline{d_i} = \frac{\sum_{j=1}^{n} d_{ij}}{n-1} \qquad \text{(ii)}$$

$$MeanDistance_{(Normalized)} = \frac{(MeanDistance - Min\_MeanDistance)}{(Max\_MeanDistance - Min\_MeanDistance)} \qquad \text{(iii)}$$

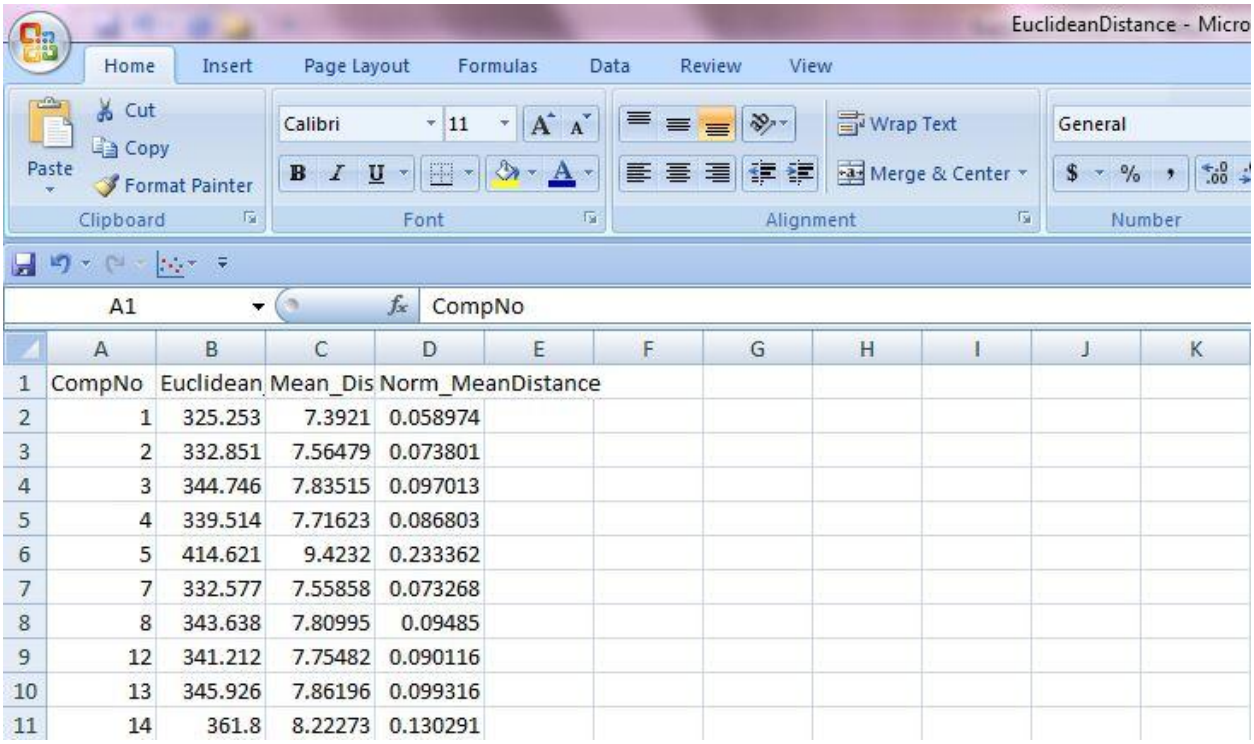
Where,
d_{ij}: the distance score between two compounds
$\overline{d_i}$: Mean distance

Input file format: This program takes a input file in CSV format.



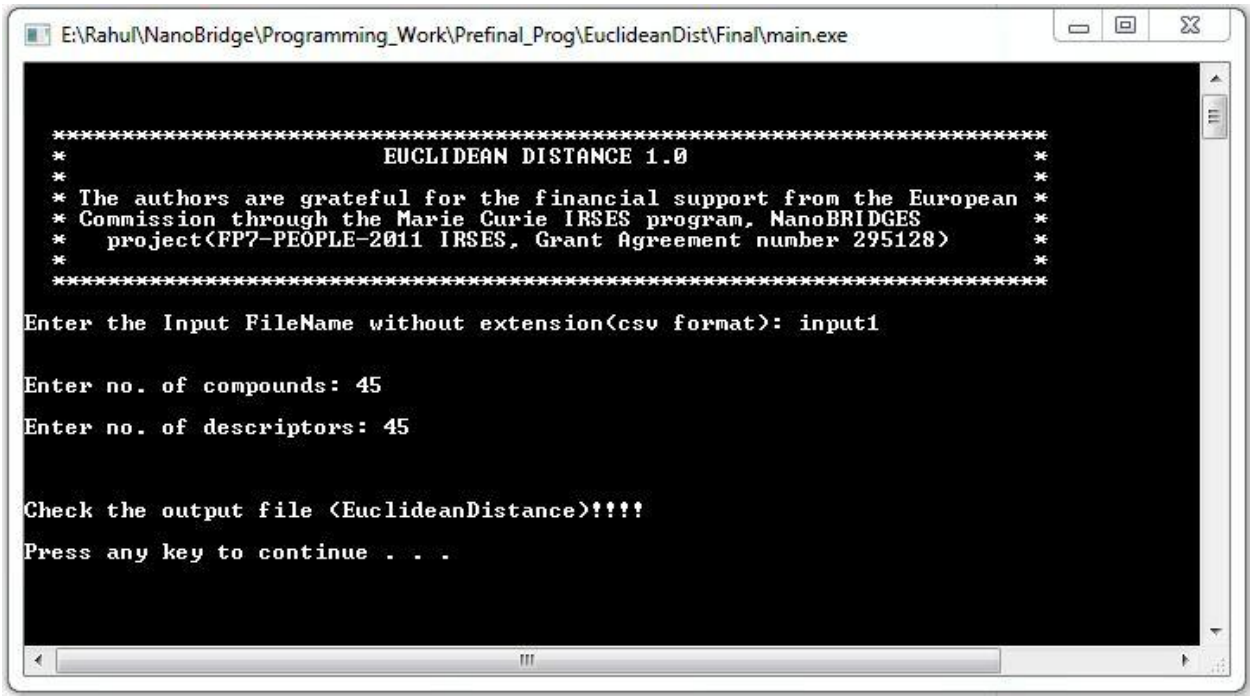
	A	B	C	D	E	F	G	H	I	J	K
1	CompNo	S_sCH3	S_ssCH2	S_dsCH	S_aaCH	S_dssC	S_aasC	S_aaaC	S_ssssC	S_sNH2	S_ssNH
2	1	0.99786	-0.41755	1.05029	0.39528	0.68812	0.50956	-0.14907	0.17181	-0.28502	-0.97336
3	2	-0.50166	-0.41755	1.00002	0.35905	0.68812	0.47428	-0.14907	0.17181	-0.29834	-0.97336
4	3	2.11327	-0.41755	0.91396	-0.05687	0.10788	0.45357	-0.14907	0.17181	-0.27865	-0.97336
5	4	-0.50166	-0.41755	0.87386	0.082	0.68812	0.31016	-0.14907	0.17181	-0.31162	-0.97336
6	5	0.86771	-0.41755	0.93262	-0.34087	0.68812	0.48256	-0.14907	0.17181	-0.28599	-0.97336
7	7	-0.50166	-0.41755	1.0653	0.35235	0.68812	0.51415	-0.14907	0.17181	-0.29888	-0.97336
8	8	-0.50166	-0.41755	0.99624	0.22631	0.68812	0.3961	-0.14907	0.17181	-0.30674	-0.97336
9	12	-0.50166	0.1371	1.08978	0.30244	0.68812	0.47101	-0.14907	0.17181	-0.29181	-0.97336
10	13	-0.50166	-0.00186	0.88027	0.17143	0.68812	0.35283	-0.14907	0.17181	-0.2976	-0.97336
11	14	-0.50166	1.13381	0.9997	0.37329	0.68812	0.44626	-0.14907	0.17181	-0.28145	-0.97336
12	15	-0.50166	1.38075	1.28003	0.89713	0.68812	0.50564	-0.14907	0.17181	-0.27475	-0.97336

Output file format:



	A	B	C	D	E	F	G	H	I	J	K
1	CompNo	Euclidean	Mean_Dis	Norm_MeanDistance							
2	1	325.253	7.3921	0.058974							
3	2	332.851	7.56479	0.073801							
4	3	344.746	7.83515	0.097013							
5	4	339.514	7.71623	0.086803							
6	5	414.621	9.4232	0.233362							
7	7	332.577	7.55858	0.073268							
8	8	343.638	7.80995	0.09485							
9	12	341.212	7.75482	0.090116							
10	13	345.926	7.86196	0.099316							
11	14	361.8	8.22273	0.130291							

How to run a program:



Disclaimer

For academic purpose only.

The program Euclidean-Distance1.0 has been developed in C++ language and is validated on the known data sets. This program is compatible with both 32- and 64-bit Windows operating system. Please report for discrepancy of result for any other dataset. Contact us at any of the following addresses:

Dr. Tomasz Puzyn,
NanaBRIDGES Project
Coordinator,
Faculty of Chemistry,
University of Gdansk,
Gdansk,
Poland 80-952
Email Id: puzi@qsar.eu.org

Dr. Kunal Roy,
Drug Theoretics and Cheminformatics Lab.,
Dept. of Pharmaceutical Technology,
Jadavpur University,
Kolkata, West Bengal,
INDIA-700032
Email Id: kunalroy_in@yahoo.com

Rahul Balasaheb Aher,
Phd Research Scholar,
Drug Theoretics and Cheminformatics Lab.,
Dept. of Pharmaceutical Technology,
Jadavpur University,
Kolkata, West Bengal,
INDIA-700032
E-mail Id: rahulba26@gmail.com (*for any queries regarding the program)

References:

1. Golmohammadi, Hassan, Zahra Dashtbozorgi, and William E. Acree Jr. "Quantitative structure–activity relationship prediction of blood-to-brain partitioning behavior using support vector machine." *European Journal of Pharmaceutical Sciences* 47, no. 2 (2012): 421-429.
2. De Maesschalck, Roy, Delphine Jouan-Rimbaud, and Désiré L. Massart. "The mahalanobis distance." *Chemometrics and intelligent laboratory systems* 50, no. 1 (2000): 1-18.