proceedings

MDPI

Proceedings

Web Server and R Library for the Calculation of Markov Chains Molecular Descriptors †

Paula Carracedo-Reboredo [1], Cristian R. Munteanu [1,2], Humbert González-Díaz [3,4] and Carlos Fernandez-Lozano [1,2,*]

[1] Departament of Computer Science and Information Technologies, Faculty of Computer Science, Universidade da Coruña, Campus Elviña s/n, 15071 A Coruña, Spain; paula.carracedo@udc.es (P.C.-R.); c.munteanu@udc.es (C.R.M.)

[2] CITIC-Research Center of Information and Communication Technologies, Universidade da Coruña, 15071 A Coruña, Spain

[3] Basque Center for Biophysics, University of the Basque Country UPV/EHU, 48940 Leioa, Bilbao, Spain; humberto.gonzalezdiaz@ehu.es

[4] IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain

* Correspondence: carlos.fernandez@udc.es

† Presented at the 3rd XoveTIC Conference, A Coruña, Spain, 8–9 October 2020.

Published: 20 August 2020

Abstract: Markov Chain Molecular Descriptors (MCDs) have been largely used to solve Cheminformatics problems. The software to perform the calculation is not always available for general users. In this work, we developed the first library in R for the calculation of MCDs and we also report the first public web server for the calculation of MCDs online that include the calculation of a new class of MCDs called Markov Singular values. We also report the first Cheminformatics study of the biological activity of 5644 compounds against colorectal cancer.

Keywords: Markov Chains; online tool; R; colorectal cancer

1. Introduction

Cheminformatics models are able to predict different outputs in complex molecular systems. On the other hand, colorectal cancer (CRC) is the third most commonly occurring cancer in men and the second in women, having a mortality of approximately 56% of the patients [1]. Although a number of compounds for anti-CRC activity have been synthetized and tested, the possibility of coming across an effective drug is still too low [2]. Markov Chain Molecular Descriptors (MCDs) have been largely used to solve Cheminformatics problems, and the calculation is done very often using specific software not ever available for general users. In this work, we developed the first library in R [3] for the calculation of MCDs and the first public web server for the calculation online that includes the calculation of a new class of MCDs called Markov Singular indices. We report a case study; we illustrated the use of those molecular descriptors in the study of active compounds against colorectal cancer (CRC).

2. Materials and Methods

We proposed an implementation in R of the algorithm for calculation of MCDs that can calculate two drug topological indices (TIs) families: Markov Mean Properties (MMPs) and Markov Singular Values of Transition Probabilities (MMSVs). The combination of the RMarkov.mol with the RRegrs package generates a powerful and fast R tool for designing QSAR (Quantitative

structure-activity relationship) regression models. We obtained 5644 preclinical assays of CRC active compounds from ChEMBL and calculated the MCDs using our web server to simplify the process.

3. Results

Figure 1 shows the user interface of Markov Chemical Descriptors Calculator (MCDCalc) web server. This allows the calculation of molecular descriptors for each atomic property and type of atom. Smiles formulas can be read from the text file or can be individually pasted on screen textbox.

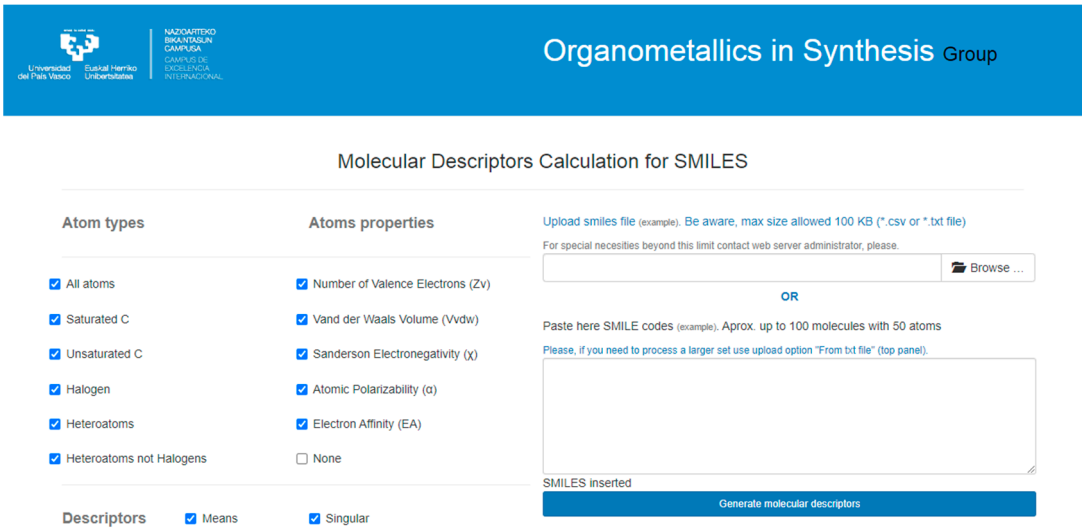

Figure 1. Online web server.

We also used the molecular descriptors as input for the RRegrs [4] package in order to find better regression models. Random Forest (RF), Support Vector Machines (SVM), Neural Networks (NN), and Partial Least Squares (PLS) regression methods have been tested and the results are presented in Table 1.

Table 1. Results for RF, SVM, NN and PLS.

Method	Training		Test	
	R^2	RMSE	R^2	RMSE
RF	0.907	0.101	0.926	0.093
SVM	0.868	0.122	0.866	0.128
NN	0.849	0.132	0.829	0.143
PLS	0.801	0.155	0.775	0.167

4. Conclusions

We have developed the first library in R for the calculation of MCDs, and the first public web server for the calculation of MCDs online that includes the calculation of Markov Singular values which are useful to predict the activity prediction of anti-colorectal cancer compounds. The RF regression model showed the best results.

References

1. Riihimaki, M.; Hemminki, A.; Sundquist, J.; Hemminki, K. Pat-terns of Metastasis in Colon and Rectal Cancer. *Sci. Rep.* **2016**, *6*, 1–9, doi:10.1038/srep29765.
2. Jhanwar, B.; Sharma, V.; Singla, R.K.; Shrivastava, B. QSAR-Hansch Analysis and Related Approaches in Drug Design. *Phar.-Macol. Online Newsl.* **2011**, *1*, 306–344.

3. Santiago, C.B.; Guo, J.Y.; Sigman, M.S. Predictive and Mechanistic Multivariate Linear Regression Models for Reaction Development. *Chem. Sci.* **2018**, *9*, 2398–2412, doi:10.1039/c7sc04679k.
4. Tsiliki, G.; Munteanu, C.R.; Seoane, J.A.; Fernandez-Lozano, C.; Sarimveis, H.; Willighagen, E.L. RRegrs: An R package for computer-aided model selection with multiple regression models. *J. Cheminform.* **2015**, *7*, 46, doi:10.1186/s13321-015-0094-2.