Journal of Chemical Information and Modeling 2015 vol.55 N11, pages 2403-2410

Stargate GTM: Bridging Descriptor and Activity Spaces

Gaspar H., Baskin I., Marcou G., Horvath D., Varnek A.
Kazan Federal University, 420008, Kremlevskaya 18, Kazan, Russia

Abstract

© 2015 American Chemical Society. Predicting the activity profile of a molecule or discovering structures possessing a specific activity profile are two important goals in chemoinformatics, which could be achieved by bridging activity and molecular descriptor spaces. In this paper, we introduce the "Stargate"version of the Generative Topographic Mapping approach (S-GTM) in which two different multidimensional spaces (e.g., structural descriptor space and activity space) are linked through a common 2D latent space. In the S-GTM algorithm, the manifolds are trained simultaneously in two initial spaces using the probabilities in the 2D latent space calculated as a weighted geometric mean of probability distributions in both spaces. S-GTM has the following interesting features: (1) activities are involved during the training procedure; therefore, the method is supervised, unlike conventional GTM; (2) using molecular descriptors of a given compound as input, the model predicts a whole activity profile, and (3) using an activity profile as input, areas populated by relevant chemical structures can be detected. To assess the performance of S-GTM prediction models, a descriptor space (ISIDA descriptors) of a set of 1325 GPCR ligands was related to a B-dimensional (B = 1 or 8) activity space corresponding to pKi values for eight different targets. S-GTM outperforms conventional GTM for individual activities and performs similarly to the Lasso multitask learning algorithm, although it is still slightly less accurate than the Random Forest method.

http://dx.doi.org/10.1021/acs.jcim.5b00398