

Implicit-descriptor ligand-based virtual screening by means of collaborative filtering**Raghuram Srinivas**

Southern Methodist University, USA

Current ligand-based machine learning methods in virtual screening rely heavily on molecular fingerprinting preprocessing, i.e. explicit description of ligands structural and physicochemical properties in a vectorized form. Of particular importance to these current methods are the extent to which molecular fingerprints describe a particular ligand and what metric sufficiently captures similarity among ligands. In this work, we propose and evaluate methods that do not require explicit feature vectorization through fingerprinting, but, instead, provide implicit descriptors based only on other known assays. Our methods are based upon collaborative filtering algorithms. Our implicit descriptor method does not require the fingerprint similarity search, which makes the method free of the bias arising from the empirical nature of the fingerprint models and similarity search assumptions. The main strengths of this method are its resilience to target-ligand sparsity and high potential for spotting promiscuous ligands.

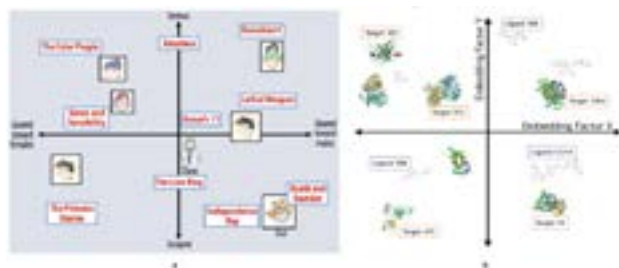


Figure 1: The figure a, illustrates the concept of latent factor; the latent factor recommendation for a movie recommendation engine. The latent factor method relies on learning hidden factors from the user-movie ratings alone. In this simplified example, the system learns two dimensions from the ratings and places the movies and users in this 2D space. The users predicted rating of the movie would be a dot product of the user's and movie's location in the 2D space. Figure b illustrates the same concept for a target-ligand embedding. Here the latent factors correspond to properties that the ligands and protein targets can be jointly modeled with. The properties may correspond to a distinct chemical property, but might also pertain to a factor not well described by traditional cheminformatics

Biography

Raghuram Srinivas is a Research Assistant with the Master of Science in Data Science program at Southern Methodist University, Dallas, Texas. He has extensive experience in Cognitive Computing and Artificial Intelligence Space with his work at the Watson Group at International Business Machines and contributed seven patents in the field.

rsrinivas@smu.edu