“Grider”, a visual data mining tool for inspection of chemical spaces. Extension to diversity exploration

Alban Arrault, Véronique Hamon, Philippe Vayer, Marc Bertrand

Technologie Servier, rue E. Vignat 25-27, 45000 Orléans, France

Pharmaceutical industries face up to millions of records, or even far more, generated by experimental or *in silico* models. This emerging situation has encouraged multiple disciplines to develop new data mining methods, designed for the extraction into relevant knowledge. In numerous applications, a data cloud entity is driven by its geo-spatial coordinates and its density. Most of time, human visual perception is too limited to quantify both parameters especially while the set is tightly packed. Thereby, it remains crucial for the analyst to gain insight into these data and consequently to manage but also compare different spaces.

In the present study, we describe a visual data mining tool, called “Grider”, able to represent and analyze geo-spatial localization and density of a data cloud. In brief, the algorithm throws up a chemical space data into an automatically profiled grid. Each cell of the grid is regarded as an originally switched-off pixel (no element present in this pixel) which can be switched-on when a minimum of one element is localized in this pixel. The merging of the whole switched-on pixels provides a general shape which characterizes the geo-spatial occupancy of the chemical space cloud. In addition, switched-on pixels can be coloured according to the “k” times they were activated by “k” molecule(s). A key feature of this functionality is to inform about the density of the cloud.

Working with such a pixel-based grid clearly offers flexibility when handling large ensembles. This motivated us to work on “Grider” extensions. Here, some examples of the use of this approach are presented.