Docket #: S09-278

SCISSORS - Rapid Estimation of Chemical Similarities

Researchers in Prof. Vijay Pande's laboratory developed a novel computational technique (“SCISSORS”) that affords several orders of magnitude acceleration in chemical library screening. SCISSORS uses a linear algebra algorithm to approximate a given molecular similarity measure to achieve accurate estimations of shape. The generic technology can enable a large-scale cheminformatics analyses applicable to a variety of similarity measures. SCISSORS also allows compact and opaque representation which reduces storage requirements and allows relevant information to be shared without disclosing proprietary structures.

Ongoing Research
The inventors are improving the accuracy of the similarities obtained by SCISSORS. - see Kearnes et al. 2013 for the latest.

Applications

- **Cheminformatics:**
  - fast similarity searches of large chemical libraries
  - scaling of all-pairs comparisons within molecular libraries
  - opaque data sharing with outside entities
  - machine learning using vectors calculated with SCISSORS as molecular features for linear computational learning algorithms

Advantages

- **Fast:**
  - similarity comparisons over large molecular libraries are orders of magnitude faster than existing techniques

    - reduced number of conventional slow similarity evaluations required from quadratic to linear scaling
- **Low computational cost**
- **Small data storage:**
    - compact vector representations
    - faster to recalculate similarity search results than to store them
- **Confidential sharing of chemical data** - opaque representation allows safe disclosure of relevant information without revealing proprietary structures
- **Few basis molecules** needed to achieve accurate estimation of shape (~500 - 1000)

Publications

- Kearnes SM, Haque IS and Pande VS. "SCISSORS: Practical Considerations", *Journal of Chemical Information and Modeling.* 2013 Dec 1; Web release.
- Haque IS and Pande VS. "Error Bounds on the SCISSORS Approximation Method", *Journal of Chemical Information and Modeling.* 2011; 51 (9): pp 2248-53.
- Haque IS and Pande VS. "SCISSORS: a linear-algebraical technique to rapidly approximate chemical similarities", *Journal of Chemical Information and Modeling.* 2010; 50 (6): pp1075-88.

Patents

- Published Application: 20110213567

Innovators

- Imran Haque
- Vijay Pande

Licensing Contact

Imelda Oropeza

Senior Licensing Manager, Physical Sciences

Email