Medical Xpress

Computing toxic chemicals

July 18 2013

A new computational method for working out in advance whether a chemical will be toxic will be reporting in a forthcoming issue of the *International Journal of Data Mining and Bioinformatics*.

There is increasing pressure on the chemical and related industries to ensure that their products comply with increasing numbers of safety regulations. Providing regulators, intermediary users and consumers with all the necessary information to allow them to make informed choices with respect to use, disposal, recycling, environmental issues and human health issues is critical. Now, Meenakshi Mishra, Hongliang Fei and Jun Huan of the University of Kansas, in Lawrence, have developed a computational technique that could allow the industry to predict whether a given compound will be toxic even at a low dose and thus allow alternatives to be found when necessary.

Toxicity is almost always an issue of availability and dosage. Whether or not a compound is natural or synthetic it can be toxic from snake venom and jellyfish stings to petrochemicals and pesticides. However, some chemicals are more toxic than others, exposure to a lower dose will cause health problems or potentially be lethal. It is very important to find a way to determine whether a newly discovered synthetic or natural chemical might cause toxicity problems.

The team also points out that the US Environmental Protection Agency (EPA) and the Office of Toxic Substances (OTS) in the USA had listed 70,000 industrial chemicals in the 1990s, with 1000 chemicals added each year for which even simple toxicological experiments had not been

carried out. This is largely a problem of logistics and costs as well as the ethical question of whether so many tests, which would have to be carried out on laboratory animals, should be done at all.

Now, Huan and colleagues in the Department of Electrical Engineering and Computer Science at Kansas, have successfully tested a statistical algorithm against more than 300 chemicals for which the toxicity profile is already known. Their technique offers a computational method of screening a large number of compounds for obvious toxicity very quickly and might preclude the need for animal testing of the compounds, provided regulators don't insist on such "in vivo" data from the latter.

The research builds on well-established principles from the pharmaceutical industry known as Quantitative structure-activity relationships (QSARs) in which the type of atoms and how they are connected together can be correlated with the activity of a drug molecule. Certain molecular shapes and types are soluble in water, for instance, or interact in a certain way with different enzymes and other proteins in the body, leading to their overall activity. Different molecular features will make a similar molecule behave in a different way - more or less soluble, stronger or weaker acting. The team has now turned the QSAR around so that instead of searching for the features in a molecule that make it of benefit in medicine they look for the atomic groups and the type of bonds that hold them together to find associations with toxicity.

The team points out that few earlier attempts at predicting toxicity of chemicals have proved successful, most approaches are no better than random guessing. The team's new statistical approach combines "Random Forest" selection with "Naïve Bayes" statistical analysis to boost the predictions well beyond random. They team saw prediction accuracy in 2 out of 3 chemicals tested. Given that there are around

100,000 industrial chemicals that need toxicity profiling, this result should allow the industry and regulators to focus on a large number of the most pressing of those, the ones predicted to have greatest toxicity and leave the less likely until additional resources are available.

The researchers are now tuning the algorithm to work faster and with greater precision so that it ignores common molecular features now known not to contribute to toxicity characteristics in the chemicals they have studied so far.

 More information: *Int. J. Data Mining and Bioinformatics*, 2013, 8, 338-348

Provided by Inderscience Publishers