

A new method of modeling drug-target interactions fixes a detrimental bias of past techniques

June 28 2017

"Drug discovery is a very long process. At each stage, you may find your drug is not good enough and you need to seek another candidate," explains A*STAR's Xiao-Li Li. His team won 'best paper' at the 2016 International Conference on Bioinformatics for a novel approach to correcting an intrinsic problem with machine learning methods.

Computer simulation, or 'in silico' [drug discovery](#) techniques, can improve accuracy and reduce the drawn out, hugely expensive road to bringing a [drug](#) to market—averaging more than 12 years and \$US1.8 billion.

Many computer simulations however first require 'training' on datasets of known drugs and their targets. This data can include additional information on 3-D structure, chemical composition, and other molecular properties. Drawing on trends from this database of known data, the simulation can then predict the interactions of unknown molecules—leading to [new drugs](#) and new [target](#) proteins.

However, of all the drugs and targets in the database, only certain combinations will interact. Potential pairings are far outweighed by non-interacting pairs referred to as 'between-class imbalance'. Further imbalance is present in the form of different and unequal subtypes of interaction, dubbed 'within-class imbalance'.

"Any computational models that are designed to optimize accuracy will be biased and will tend to classify unknown pairs into majority or non-interaction class," says Li. "Majority classes are better represented in data than minority interaction classes—this skews these models and produces errors. Data imbalance is a challenging issue."

Li's team at the A*STAR Institute for Infocomm Research, sought to overcome this by developing an 'imbalance-aware' algorithm that more accurately predicted drug-target interactions based on a database of 12,600 known interactions and around 18 million known non-interacting pairs. The algorithm was designed to better recognize underrepresented interaction groups and enhance the data within them.

By improving the ability of the computer model to focus on the most useful data (the interactions), the team created a system that outperformed existing modeling techniques, predicting new, unknown drug-target interactions with high accuracy.

The future of machine learning depends on artificial intelligence and advanced learning such as 'deep learning.' Nevertheless, as Li adds: "data is key. In order to further enhance our predictive capability, the first thing we can do is collect more relevant data about drugs and targets."

More information: Ali Ezzat et al. Drug-target interaction prediction via class imbalance-aware ensemble learning, *BMC Bioinformatics* (2016). [DOI: 10.1186/s12859-016-1377-y](https://doi.org/10.1186/s12859-016-1377-y)

Provided by Agency for Science, Technology and Research (A*STAR), Singapore

Citation: A new method of modeling drug-target interactions fixes a detrimental bias of past

techniques (2017, June 28) retrieved 26 April 2024 from <https://phys.org/news/2017-06-method-drug-target-interactions-detrimental-bias.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.