Mining the Internet for Speedier Safety Alerts on Drugs

Scientists are sifting through massive quantities of freely available data scattered across the Internet, aiming to catch potentially deadly problems with prescription drugs more quickly—even ahead of federal regulators.

Researchers from the University of Virginia and West Virginia University have developed mathematical recipes that computers use to filter billions of pieces of data from patient comments in online chats, websites and news stories to detect serious adverse drug reactions.

They then catalog the complaints and determine when they rise to a level that deserves medical or regulatory attention.

Similar techniques have been used to detect bioterrorism threats, study crowd and consumer behavior and map out how infections spread.

The goal is to be able to alert the medical community about possible negative drug reactions much earlier so they can use it their clinical practices, as well as to warn the U.S. Food and Drug Administration about potential new adverse reactions.

It can sometimes take years for enough verified evidence to build up for the FDA to act and issue a warning. For instance, the antidepressant Prozac was approved in 1987, but the FDA's most serious "black box" warning, stating that some children, teens and young adults who take it could be at higher risk of suicidal thoughts or behavior, didn't come until 2005.

The challenge for researchers, they say, is to sort through tons of "noise," the reams of information that aren't relevant, accurate or important, and to recognize useful signals. Researchers hope to use continually improving computer algorithms—programs that detect key
patterns or relationships between words—to make recognition of important signals more accurate, says Ahmed Abbasi, a professor of information technology at the University of Virginia’s McIntire School of Commerce in Charlottesville.

"How can we make the needle bigger and the haystack smaller?" asks Dr. Abbasi, who has published papers using similar methods to detect fraudulent medical websites and financial reporting by companies.

Dr. Abbasi, West Virginia’s Donald Adjeroh and colleagues have been testing their strategy by analyzing Web data on 20 drugs, such as Prozac, that are already on the market for which the FDA has issued its strongest "black box" warnings.

They culled tens of millions of pieces of data on each of the drugs from websites over the last decade and determined, using their algorithms, at what point they would have been alerted to possible links between drugs and adverse reactions.

In preliminary data, they found that 80% of the time, their formulas detected potential adverse event patterns three months earlier than the FDA issued warnings, said Dr. Adjeroh, a computer science professor at West Virginia University in Morgantown. In some cases, they were years ahead of the FDA’s warnings. The researchers were recently awarded a $130,000 grant from the National Science Foundation Smart Health and Wellbeing program to launch a bigger project in this area.

The length of time that the FDA takes to investigate reports and issue a warning hinges on numerous factors, including the frequency and severity of an adverse event, internal and external discussions and legal reviews, says spokeswoman Sandy Walsh. "There is not a one-size-fits-all approach."

The agency’s detection process primarily relies on doctors, companies and consumers to report problems directly to the FDA. If enough adverse events are recorded, the agency launches an investigation into the potential connection and can warn the public about the possibility of an adverse event or pull the drug from the market.

"With many safety issues, we try to issue early notices of ‘ongoing safety reviews’…which we will update once our analysis is complete," says Ms. Walsh.

Still, the FDA’s process for reporting adverse events has been criticized by doctors and consumer groups for being slow. "The issue with the FDA process is that it relies on spontaneous reporting, which can take a long time," says David
Bates, chief quality officer at Brigham and Women's Hospital in Boston, who studies improvements in adverse-event detection.

In recent years, the FDA has pushed to improve post-marketing surveillance. Its Sentinel Initiative, a public/private partnership that searches medical-claims data for potential adverse events, has identified several potential links between drugs and negative effects. To date, Sentinel investigators have created computer programs to examine some 40 medical products or conditions, Ms. Walsh says. They recently completed a review of certain blood-pressure medications and will discuss their findings in a webinar next week.

Dr. Bates, who is also a professor of medicine at Harvard Medical School, and his team are working on other approaches to improve detection of adverse drug events. He says the work of Drs. Abbasi and Adjeroh and their big-data approach is "exciting" and "will likely be very helpful for detecting signals." But challenges remain, says Dr. Bates, who isn't involved in their work. With drugs and adverse events, "sometimes it's hard to put two and two together," he says.

Drs. Abbasi and Adjeroh say they are honing their computer algorithms to search more accurately for genuine signals in the data and tune out extraneous information. This is a major challenge since there is a "potentially infinite number of matches between drugs and side effects," says Dr. Abbasi.

They are developing a more sophisticated keyword search to help eliminate words appearing close together on a website that have no real connection. Instead of just looking for the words "Prozac" and "suicide" that appear anywhere on a page, adding "sentiment" information, such as "I've been feeling" may be helpful, says Dr. Abbasi.

They also have to separate out legitimate reports of concerns from fraudulent or intentionally misleading sources—using cues they have identified for fake websites and known fake Web addresses.

Such big-data approaches move away from a reliance on voluntary reporting and clinical intuition. "There's always been this struggle between intuition and data," says Dr. Abbasi. "There's a paradigm shift where data-driven decision-making is being increasingly adopted."

Write to Shirley S. Wang at shirley.wang@wsj.com