Practical Uses For Machine Learning In Health Care Cases

By Mihran Yenikomshian, Lisa B. Pinheiro, Jimmy Royer and Paul E. Greenberg; Analysis Group, Inc.

| Mihran Yenikomshian | Lisa B. Pinheiro | Jimmy Royer | Paul E. Greenberg |

As a society we have come a long way from the days of manual pencil and paper calculations. We are now accustomed to computations performed with the push of a button, whether by way of a high-end calculator or a sophisticated computer running electronic spreadsheets and statistical software. However, the growing volume and complexity of available data are testing the capacity of these familiar tools.

Traditional spreadsheets and statistical software are limited in terms of the magnitude of data they can handle. And, it would be a laborious process to turn an exabyte — or even many gigabytes or terabytes — of textual material into something that can be meaningfully interpreted with automated computer processes. Even then, some narrative details are likely to get lost.

Machine learning — which draws on insights from a variety of disciplines including computer science, mathematics and engineering — represents a breakthrough in this context. Machine learning can efficiently turn enormous volumes of both structured (coded) and unstructured (narrative) data into something meaningful, without losing underlying details.

We recently described the use of machine-learning algorithms in health care litigation.[1] In this follow-up article, we elaborate on some practical applications of machine learning in the courtroom in terms of informing legal strategy, identifying relevant materials for experts, and enhancing expert testimony.

Informing Legal Strategy

Health care, in particular, has experienced an explosion in terms of the variety and richness of newly available data. This has been driven in part by the advent of electronic medical records (EMR) and introduction of industry reporting requirements, such as the Sunshine Act. In addition, it has been further fueled by technological innovations that allow for greater data storage and rely on ever-increasing computing power. For an attorney trying to get a handle on the most relevant facts of the case and develop winning legal theories, this proliferation of data can be daunting.

Machine learning algorithms offer the opportunity to derive insights from complex, voluminous data that might otherwise be elusive at the earliest stages of discovery. Consulting experts can use these tools to generate hypotheses even before knowing all the factual underpinnings of the case. This dynamic can provide the extra confidence for counsel to pursue a novel theory of the case.

One potential application in life sciences litigation involves disputes over alleged off-label promotion of prescription drugs. While conventional analyses might group together all patients with a particular condition (e.g., lung cancer), machine learning methods can be used to identify other similarities among patients that lead to finer groupings. Such clustering could reveal clinical differences (e.g., advanced age, failure on other cancer therapies, genetic markers) among groups of patients that might explain use of the drug independent of any promotion. Uncovering these types of patterns at an early stage in the litigation can be beneficial to attorneys as they contemplate crafting the case narrative.

Machine learning can also be used to pinpoint helpful facts and witnesses for further examination. For example, attorneys defending a pharmaceutical manufacturer against allegations of kickbacks paid to physicians might benefit from machine learning's ability to identify doctors who did not receive any payments but had similar prescribing patterns to those who did. Deposing such physicians could shed light on factors that drive prescribing patterns in the absence of any possible inducements. With conventional methods, this could be a cumbersome process requiring the analyst to specify selected parameters of interest and might not generate ideal candidates. But, with machine learning, there is no limit to the number of — or interrelationship among — parameters the computer can account for, which increases the potential for a more fertile as well as efficient exercise.

Identifying Relevant Materials for Experts

Machine leaning can also be a valuable tool when it comes to identifying relevant materials to share with the expert. With the proliferation of enormously detailed data, it may be tempting to invoke rough distinctions in response to an expert's request for materials. But, this approach risks producing too much or too little information. Machine learning can help determine precisely what information is central to the query and alleviate the excess costs associated with production of extraneous materials. Since a seemingly innocuous data omission could prove to be the missing statistical link on which the quality of the prediction depends, it is particularly important to get this right.

Consider a dispute over best efforts in the context of a co-promotion agreement for a prescription pharmaceutical, for example. To estimate sales but for one party's alleged failure to perform, an expert might want to develop a model of consumer demand. This historically would have involved combining different types of quantitative data, such as shipments of drugs, marketing expenditures, and prices and typically would not make use of the vast amounts of qualitative data in a company's possession. Machine learning, however, has the potential to highlight relevant features of otherwise difficult-to-use data. Thus, information that might have once been discarded as impractical or irrelevant for expert modeling purposes, such as documentation of patient/physician perceptions, can be identified in the discovery record with the benefit of machine learning and can serve to increase the predictive power of counterfactual scenarios.

Enhancing Expert Testimony

Expert testimony can also be enhanced by machine learning techniques. Whereas historically, liability and damages experts have formed their opinions based on analyses they have prespecified and validated with traditional data sets, machine learning algorithms provide the opportunity to learn from data and experiences without the imposition of restrictions and assumptions.

Suppose, for example, that an expert wants to undertake a causation analysis focusing on a specific patient population with a particular disease. EMR data paint a rich portrait of a patient's medical history, but they contain an intricate mix of structured and unstructured elements that present challenges for conventional analytics.[2] Such methods cannot harness all of this rich qualitative information as they require an analyst to specify certain criteria in advance and/or perform character string searches to identify appropriate patients. As such, insights from potentially meaningful textual notes can easily be missed. With machine learning's ability to perform natural language processing, on the other hand, all of this information can be brought to bear on the analysis.

To be sure, machine learning is not a replacement for expert judgment nor should it be considered a wholesale substitute for traditional methods. Data need to be understood, cleaned, coded and analyzed before even thinking about employing any computer

algorithm. Moreover, after implementing a methodology, the expert will need to rigorously validate the chosen model and evaluate whether results are meaningful and sufficiently accurate (e.g., a model that accurately predicts an outcome 90 percent of the time but has a high false positive rate might not be appropriate). Such considerations need to be factored into any methodological decisions.

Testifying experts using machine learning methods will also need to educate and convince the court of the validity of these less familiar models. Like the introduction of other new technologies (e.g., fingerprints, DNA evidence), testifying experts' reliance on machine learning might invite initial skepticism in the courtroom. Machine learning algorithms can be very technical and complicated with somewhat ominous sounding names (e.g., support vector machine). This feature, combined with fear eliciting pop-culture portrayals of artificial intelligence (e.g., think Skynet from the Terminator movies),[3] may initially evoke an even more profound adverse reaction.

Accordingly, experts relying on these models will need to be excellent communicators and decompose structures into easily understandable components to dispel the feeling of a black box that cannot be trusted.[4] Furthermore, since there are many different machine learning algorithms and implementations, experts will need to become familiar with their inner workings and articulate the rationale for choosing a specific method.

Conclusion

In the ever-increasingly complex and technical world of litigation, the widespread adoption of machine learning will no doubt prove to be a significant advance. These new techniques can be harnessed by nontestifying experts to help attorneys develop better legal strategies, conduct informed fact discovery, and provide testifying experts with the most complete set of relevant information.

Testifying experts opining on liability and damages can also take advantage of these new techniques that, taken together with traditional methods, can bolster their opinions. The role of testifying experts will be enhanced in this context despite the additional automated processes involved. Their expertise will continue to form the basis for selecting data and analytical methods that will need to be communicated persuasively to relevant parties.

Mihran Yenikomshian is a vice president in Analysis Group's Boston office. Lisa Pinheiro and Jimmy Royer are vice presidents with Groupe d'Analyse, Analysis Group's Montreal office. Paul Greenberg is a managing principal in Analysis Group's Boston office.

Endnotes

1   Pinheiro et al., "Machine-Learning Algorithms Can Help Health Care Litigation," Law360, June 8, 2016.

2   Walsh, Stephen H. "The clinician's perspective on electronic health records and how they can affect patient care." Bmj 328.7449 (2004): 1184-1187.

3   The sentient artificial intelligence computer system that was responsible for the destruction of mankind.

4   For example, an unhelpful decomposed explanation of what a support vector machine does would be "a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression or other tasks" per Wikipedia.