Researchers unmaks lobbyists' model legislation

August 29 2016, by Heather Newman

Since 2005, about two dozen states have created "Stand Your Ground" laws authorizing deadly force in self-defense. And nearly all of those laws have similar language.

So, how did that happen? They all shared an author: The American Legislative Exchange Council drafted "model" legislation and lobbied states to pass it.

This tactic of using model laws to influence legislation is common in the U.S. ALEC, for its part, doesn't hide the fact that it provides this service in a "model policy library." But this kind of influence can be hard to detect.

Eytan Adar, associate professor at the University of Michigan School of Information, calls this the "Dark Corpora" problem: spotting laws where clusters of similar language exist, and reconstructing what the source material might have looked like.

The Dark Corpora problem just got a lot easier to solve—thanks to new research led by Adar; Matthew Burgess, who recently received his Ph.D. in computer science from the U-M College of Engineering; and Eugenia Giraudy of internet market research firm YouGov.

"As bipartisan politics leads to more gridlock at the federal level, lobbyists are focusing more and more on influencing state legislation," Burgess said. "We hope that social scientists, watchdog groups and

journalists use the tools and data we built."

The researchers developed LobbyBack: a new system of [natural language processing](#) analysis that takes clusters of bills and identifies common language, attempting to reconstruct those sentences into something that might resemble the original model legislation.

"This is about trying to figure out what the law looked like in the original document and how laws propagate from these organizations to state legislatures," Adar said. "There are two reasons why it's exciting: First, this is an area where that kind of influence has a huge effect. Because it's not transparent, it's problematic. It's also a hard, interesting problem that exists in areas ranging from meme diffusion to plagiarism detection."

The system could be used by journalists or researchers to identify the sources behind new bills, the study suggests.

"We look at all the laws and find groups of laws that seem more similar than we would expect due to random chance. Those we examine sentence by sentence," Adar said. "We use an algorithm used in DNA analysis to put things in order and find common sentences. For the ones that are obvious, we can say, we're certain this is where it came from. But we can't always confirm."

Still, common language strongly suggests a common source, even if that source is not immediately clear. Identifying those occurrences can give interested parties pointers on where to look.

The team used the entire database of state legislation from openstates.org to test the system, including some 550,000 bills and 200,000 resolutions for all 50 states. The algorithm discards boilerplate and other text frequently reused, concentrating on long passages that can

differ in details and are related to the topic of the bill.

Once they analyzed the complete collection of bills and created clusters of documents that shared similar language, they then compared the cluster to a database of more than 1,800 pieces of model legislation collected from a variety of lobbying groups. Of those, 360 matched clusters in the state bill collection.

Adar said the researchers plan to make code and data available to others who might be interested in using the tool.

The researchers recently presented two papers on the work. One, titled "Prototype Synthesis for Model Laws," was released at the Association for Computational Linguistics in Berlin earlier this month. Another authored by Burgess and others, titled "The Legislative Influence Detector: Finding Text Reuse in State Legislation," was presented earlier this month at the ACM SIGKDD Conference on Knowledge Discovery and Data Mining. This paper more fully presents the algorithms used to match text that researchers manually provided to a database of bills.

  More information: The Legislative Influence Detector: Finding Text Reuse in State Legislation: www.kdd.org/kdd2016/subtopic/v … in-state-legislation

Provided by University of Michigan

provided for information purposes only.