

New scientific approach reduces bias in training data for improved machine learning

September 1 2021



Gautam Thakur leads a team of ORNL researchers who have developed a new scientific method for identifying bias in human data annotators to ensure high-quality data inputs for machine learning applications. Credit: Carlos Jones/ORNL, U.S. Dept. of Energy

As companies and decision-makers increasingly look to machine

learning to make sense of large amounts of data, ensuring the quality of training data used in machine learning problems is becoming critical. That data is coded and labeled by human data annotators—often hired from online crowdsourcing platforms—which raises concerns that data annotators inadvertently introduce bias into the process, ultimately reducing the credibility of the machine learning application's output.

A team of scientists led by Oak Ridge National Laboratory's Gautam Thakur has developed a new scientific method to screen human data annotators for bias, ensuring high-quality data inputs for machine learning tasks. The researchers have also designed an [online platform](#) called ThirdEye that allows for scaling up the screening process.

The team's results were published in the *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*.

"We have created a very systematic, very scientific method for finding good data annotators," Thakur said. "This much-needed approach will improve the outcomes and realism of machine learning decisions around public opinion, online narratives and perception of messages."

The Brexit vote in fall 2016 provided an opportunity for Thakur and his colleagues Dasha Herrmannova, Bryan Eaton and Jordan Burdette and collaborators Janna Caspersen and Rodney "RJ" Mosquito to test their method. They investigated how five common attitude and knowledge measures could be combined to create an anonymized profile of data annotators who are likely to label data used for machine learning applications in the most accurate, bias-free way. They tested 100 prospective data annotators from 26 countries using several thousand social media posts from 2019.

"Say you want to use machine learning to detect what people are talking about. In the case of our study, are they talking about Brexit in a positive

or negative way? Are data annotators likely to label data as only reflecting their beliefs about leaving or staying in the EU because their bias clouds their performance?" Thakur said. "Data annotators who can put aside their own beliefs will provide more accurate data labels, and our research helps find them."

The researchers' mixed-method design screens data annotators with qualitative measures—the Symbolic Racism 2000 Scale, Moral Foundations Questionnaire, social media background test, Brexit knowledge test and demographic measures—to develop an understanding of their attitudes and beliefs. They then performed statistical analyses on the labels annotators assigned to [social media posts](#) against a subject matter expert with extensive knowledge of Brexit and Britain's geopolitical climate and a social scientist with expertise in inflammatory language and online propaganda.

Thakur stresses that the team's method is scalable in two ways. First, it cuts across domains, impacting data quality for machine learning problems related to transportation, climate and robotics decisions in addition to health care and geopolitical narratives relevant to national security. Second, ThirdEye, the team's open-source interactive web-based platform, scales up the measurement of attitudes and beliefs, allowing for profiling of larger groups of prospective data annotators and faster identification of the best hires.

"This research strongly indicates that data annotators' morals, prejudices and prior knowledge of the narrative in question significantly impact the quality of labeled data and, consequently, the performance of machine learning models," Thakur said. "Machine learning projects that rely on labeled data to understand narratives must qualitatively assess their data annotators' worldviews if they are to make definitive statements about their results."

More information: Gautam Thakur et al, A Mixed-Method Design Approach for Empirically Based Selection of Unbiased Data Annotators, *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021* (2021). [DOI: 10.18653/v1/2021.findings-acl.169](https://doi.org/10.18653/v1/2021.findings-acl.169)

Provided by Oak Ridge National Laboratory

Citation: New scientific approach reduces bias in training data for improved machine learning (2021, September 1) retrieved 18 April 2024 from <https://techxplore.com/news/2021-09-scientific-approach-bias-machine.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.