A NOVEL METHOD FOR ARTIFICIAL INTELLIGENCE LITERATURE SURVEY

TITLE: Joint Optimization of Artificial Intelligence Fairness and Utility

AUTHOR: Yunfeng Zhang, Rachel K. E. Bellamy, Kush R. Varshney

DESCRIPTION:

AI is increasingly being used in many high stakes decision-making applications in which fairness is an important concern. Already, there are many examples of AI being biased and making questionable and unfair decisions. The AI research community has proposed many methods to measure and mitigate unwanted biases, but few of them involve inputs from human policy makers. We argue that because different fairness criteria sometimes cannot be simultaneously satisfied, and because achieving fairness often requires sacrificing other objectives such as model accuracy, it is key to acquire and adhere to human policy makers' preferences on how to make the tradeoff among these objectives. In this paper, we propose a framework and some exemplar methods for eliciting such preferences and for optimizing an AI model according to these preferences. Artificial intelligence (AI) systems are already working alongside humans as trusted advisors in high-stakes decision-making applications such as mortgage lending, hiring, and prison sentencing. In these domains, fairness is an important concern and there are many examples of AI being biased and making arguably unfair decisions. Decisions in these domains are social constructions, and as such need to incorporate multiple viewpoints. Current AI technology does not readily allow for inclusion of viewpoints from those other than the developer of the AI model. The need for deep technical expertise is a barrier to participation for people outside of technical domains. Research is needed to bridge this gap and allow for voices from multiple domains to influence the creation and adoption of fair AI. This is critical for AI that will advise on decisions that have the potential to discriminate against certain populations, such as racial minorities or people with disabilities.

PUBLISHED IN: 2020

TITLE **:** Accountability and Evaluation of AI Algorithms Minitrack

AUTHOR **:** Radmila Juric, Robert Steel

DESCRIPTION:

The minitrack on the Accountability and Evaluation of AI Algorithms is a new within the track on Decision Analytics and Service Science. This is a rather novel but important topic which concerns many of us, but it did not attract too many papers, for many reasons. One of them is that the problem of AI accountability still does not appear as the mainstream research topic across our academic institutions, and consequently, we could not find academically viable research results which could crack this quite difficult problem. There are only two papers accepted in our session. They are different, they do not directly resolve the problem o AI evaluation and accountability, but they address issues which can have an impact on the quality and reliability of algorithms we run behind the spectrum of AI computational models The paper entitled "A comparison between a two-feedback control loop and a reinforcement learning algorithm for compliant low-cost series elastic actuators" proposes a novel type of a controller for an elastic actuator which uses an artificial neural network, trained with reinforcement learning. The authors claim that there is significance in using knowledge generated through machine learning procedures and instantly create properties of the elastic actuator to be controlled. This in turn reduces the need of manually defining fuzzy rues for controlling actuators. Trustworthy and reliable results from such AI algorithms, in safety critical applications, could be obtain by building more robust ML systems, particularly in the presence of hardware level faults. A non-biased comparison between AI and classical two-loop control systems does give indications that with reinforcement learning, ANN algorithm might excel, particularly if the model of the system is difficult or impossible to obtain. The calculation of adhesion and friction, which appear between hard surfaces, should be sufficient for building an ML classifier to predict slippery conditions in rail transport.

PUBLISHED IN: 2020

TITLE **:** Artificial Intelligence for enhancing clinical medicine

AUTHOR **:** Roxana Daneshjou, Lukasz Kidzinski,, Olga Afanasiev, Jonathan H. Chen

DESCRIPTION:

Machine learning and deep learning have revolutionized our ability to analyze and find patterns in multi-dimensional and intricate datasets. As such, these methods have the ability to help us decipher the large volume of data generated within healthcare. These tools hold the promise of enhancing patient care through several modalities, including clinical decision support, monitoring tools, image interpretation, and triaging capabilities. For the 2020 Pacific Symposium on Biocomputing's session on Artificial Intelligence for Enhancing Clinical Medicine, we highlight novel research on the application of artificial intelligence to solve problems within the field of medicine. Artificial intelligence aims to create computer programs that can abstractly reason and make generalizations about the world.1 While general artificial intelligence remains elusive, we observe continued advances and applications in machine learning, the science of creating algorithms that improve in performance with increasing data and experience.1 Deep learning, a subset of machine learning that uses multi-layer networks, has particularly gained popularity in the past several years for solving complex tasks with large amounts of data. As the emerging potential of such technology advances, researchers have been fascinated with the idea of harnessing artificial intelligence for accomplishing tasks in biomedical research and clinical medicine, an idea first proposed in the 1970s.2 Medicine is now ripe for the use of such technology with the rise in large volumes of multi-dimensional data which has grown beyond the limitations of humans to interpret.3 A wide range of studies applying artificial intelligence to medicine have been published, including algorithms to predict radiology imaging findings, dermatologic diagnoses, and patient outcomes from electronic medical record data.3 Here we highlight cutting edge research in the field selected for presentation at the 2020 Pacific Symposium on Biocomputing session on "Artificial Intelligence for Enhancing Clinical Medicine."

PUBLISHED IN: 2022

TITLE **:** Artificial Intelligence Pipeline to Bridge the Gap between Bench Researchers and Clinical Researchers in Precision Medicine

AUTHORS **:** Lewis J. Frey, Douglas A. Talbert

DESCRIPTION:

Precision medicine informatics is a field of research that incorporates learning systems that generate new knowledge to improve individualized treatments using integrated data sets and models. Given the ever-increasing volumes of data that are relevant to patient care, artificial intelligence (AI) pipelines need to be a central component of such research to speed discovery. Applying AI methodology to complex multidisciplinary information retrieval can support efforts to discover bridging concepts within collaborating communities. This dovetails with precision medicine research, given the information rich multi-omics data that are used in precision medicine analysis pipelines. In this perspective article we define a prototype AI pipeline to facilitate discovering research connections between bioinformatics and clinical researchers. We propose building knowledge representations that are iteratively improved through AI and human-informed learning feedback loops supported through crowdsourcing. To illustrate this, we will explore the specific use case of nonalcoholic fatty liver disease, a growing health care problem. We will examine AI pipeline construction and utilization in relation to bench-to bedside bridging concepts with interconnecting knowledge representations applicable to bioinformatics researchers and clinicians.

PUBLISHED IN: 2019

TITLE: Efficiently Sampling Functions from Gaussian Process Posteriors

AUTHORS: Daniel Adiwardana, Minh-Thang Luong, David R. So, Jamie Hall, Noah Fiedel Romal Thoppilan, Zi Yang, Apoorv Kulshreshtha, Gaurav Nemade, Yifeng Lu, Quoc V. Le

DESCRIPTION:

A multi-turn open domain chatbot trained end-to-end on data mined and filtered from public domain social media conversations. This 2.6B parameter neural network is simply trained to minimize perplexity of the next token. We also propose a human evaluation metric called Sensibleness and Specificity Average (SSA), which captures key elements of a human-like multi-turn conversation. Our experiments show strong correlation between perplexity and SSA. The fact that the best perplexity end-to-end trained Meena scores high on SSA (72% on multi-turn evaluation) suggests that a human-level SSA of 86% is potentially within reach if we can better optimize perplexity. Additionally, the full version of Meena (with a filtering mechanism and tuned decoding) scores 79% SSA, 23% higher in absolute SSA than the existing chatbots we evaluated.

In contrast to most modern conversational agents, which are highly specialized, the Google research team introduces a chatbot **Meena** that can chat about virtually anything. It's built on a large neural network with 2.6B parameters trained on 341 GB of text. The researchers also propose a new human evaluation metric for open domain chatbots, called Sensibleness and Specificity Average (SSA), which can capture important attributes for human conversation. They demonstrate that this metric correlates highly with perplexity, an automatic metric that is readily available. Thus, the Meena chatbot, which is trained to minimize perplexity, can conduct conversations that are more sensible and specific compared to other chatbots. Particularly, the experiments demonstrate that Meena outperforms existing state-of-the-art chatbots by a large margin in terms of the SSA score (79% vs. 56%) and is closing the gap with human performance (86%).

PUBLISHED IN: 2022