ISA Explainability workshop, Mar 2020 - abstracts

Title: Opening the blackbox - explaining supervised advanced machine learning models

Speaker: Ephraim Goldoin

Abstract:

בשנים האחרונות הסתבר יותר ויותר שמודלים חדשניים מתחום לימוד המכונה מאפשרים ניבוי טוב יותר מהמודלים הסטטיסטיים הקלאסיים כמו אלו ממשפחת GLM, הכוללת את הרגרסייה הלינארית, הלוגיסטית, הפואסונית ועוד. התופעה נצפית בעוצמה רבה יותר במקרים שבהם כמות הנתונים לתהליך הנתוח גדולה מבחינת כמות התצפיות ומספר המשתנים המסבירים. מצד שני למודלים אלו ישנו חיסרון מהותי. היתרון בניבוי מושג באמצעות אלגוריתמים מורכבים שתוצאתם הסופית איננה מאפשרת להבין ולפרש את משמעות המודל. המודל הינו ברוב המקרים black box. הדבר מביא לכך שבעולם המחקר המדעי ובתחומים מפוקחים, כמו ברפואה, היכולת להשתמש במודלים אלו מוגבלת שכן הם לא מספקים לעולם המחקר ולמפקחים כלים לבדיקת קשרים ונסיבתיות בין מסבירים למוסבר, וכימות קשרים אלו.

קהילת לומדי המכונה ראתה בעבר את תפקידה העיקרי ביצירת מודלים בעלי כושר ניבוי משופר, אולם ככל שהשימוש במודלים אלו גובר, עולה יותר ויותר הצורך ביכולת להסביר ולפרש את המודלים הללו. בשנים האחרונות החלו אכן להתפתח כלים חדשניים המתחילים לתת מענה לבעייה ומאפשרים את פתיחת ה- black boxes של המודלים המתקדמים.

בהרצאה נסקור את המודלים המורכבים מעולם לימוד המכונה ונציג ונדגים את הכלים החדשניים שמתפתחים בימים אלו ממש לפרושם.

Title: Meta Decision Trees for Explainable Recommendation Systems

Speaker: Eyal Shulman

Abstract:

We tackle the problem of building explainable recommendation systems that are based on a per-user decision tree, with decision rules that are based on single attribute values. We build the trees by applying learned regression functions to obtain the decision rules as well as the values at the leaf nodes. The regression functions receive as input the embedding of the user's training set, as well as the embedding of the samples that arrive at the current node. The embedding and the regressors are learned end-to-end with a loss that encourages the decision rules to be sparse. By applying our method, we obtain a collaborative filtering solution that provides a direct explanation to every rating it provides. With regards to accuracy, it is competitive with other algorithms. However, as expected, explainability comes at a cost and the accuracy is typically slightly lower than the state of the art result reported in the literature.

Title: Explaining visual understanding: Learn to reason about the perceived world.

Speaker: Prof. Gal Chechik

Abstract:

AI aims to build systems that can interact with their environment, with people and with other agents in the real world. This vision poses hard algorithmic challenges for learning. Such systems are often required to learn to generalize effectively from few samples, to communicate their understanding in ways that are natural to people, and "use common sense" to take into account the unseen context of an image. I will discuss several research thrusts for facing these challenges. First, an unsupervised model of the expectations of a human listener that yields informative communication about images. Second, a cooperative learning algorithm to teach networks to communicate about images using natural language. Finally, I will discuss leveraging compositional structures in attribute space to learn from descriptions without any visual samples.

The work is a summary of a series of papers, done in collaboration with
Y. Atzmon, S. Bengio, L. Bracha, J. Berant, A. Globerson, R. Hertzig, K. Murphi, D. Parikh, M. Raboh, R. Vednatam, G. Vered,

Bio:

Gal Chechik is an Assoc. Prof at the Gonda Brain Institute at Bar-Ilan University and a director of AI research at NVIDIA. His current research spans learning in brains and machines, including large-scale learning algorithms for machine perception, and analysis of changes of mammalian brains.
In 2018, Gal joined NVIDIA as the founder and head of nvidia's research in Israel. Prior to that, Gal was a staff research scientist at Google Brain and Google research developing large-scale algorithms for machine perception, used by millions daily. Gal earned his PhD in 2004 from the Hebrew University, and completed his postdoctoral training at Stanford CS department. In 2009, he started the learning systems lab at the Gonda center of Bar Ilan university, and was appointed an associate professor in 2013 Gal authored ~85 refereed publications, ~35 patents, including publications in Nature Biotechnology, Cell and PNAS.
http://chechiklab.biu.ac.il

Title: One Explanation Does Not Fit All: A Toolkit and Taxonomy of AI Explainability Techniques

Speaker: Dr. Rony Luss

Abstract:

As artificial intelligence and machine learning algorithms make further inroads into society, calls are increasing for these algorithms to explain their decisions. Whether you are the loan officer that needs to understand why an algorithm accepted a particular application or the applicant that wants to know what they could've done different to avoid their rejected application, the need for explanations is clear.

This talk gives an overview to AI Explainability 360, IBM's open-source software toolkit featuring eight diverse and state-of-the-art explainability methods and two evaluation metrics. We further provide a taxonomy to help those requiring explanations figure out which explanation method will best serve their purposes.

Data scientists and other users will learn about a new toolkit that offers hands-on experience with some of the latest explainability methods through various demos and tutorials in Jupyter notebooks. Taken together, our toolkit and taxonomy can help identify gaps where more explainability methods are needed and provide a platform to incorporate them as they are developed.