پژوهشگاه دانش‌های بنیادی
پژوهشکده علوم کامپیوتر

سخنرانی علمی

Concept-based Methods for Explainability of Deep Neural Networks

فاطمه آقایی پور

پژوهشکده علوم کامپیوتر، پژوهشگاه دانش‌های بنیادی (IPM)

Abstract

Explainability of Deep Neural Networks (DNNs) has been garnering increasing attention in recent years. Of the various explainability approaches, concept-based techniques stand out for their ability to utilize human-meaningful concepts instead of focusing solely on individual pixels. However, there is a scarcity of methods that consistently provide both local and global explanations. Moreover, most of the methods have no offer to explain misclassification cases. Considering these challenges, we present a unified concept-based system for unsupervised learning of both local and global concepts. Our primary objective is to uncover the intrinsic concepts underlying each data category by training surrogate explainer networks to estimate the importance of the concepts. Our experimental results substantiated the efficacy of the discovered concepts through diverse quantitative and qualitative assessments, encompassing faithfulness, completeness, and generality. In this talk, we will delve into this method as well as other related concept-based approaches.

Biography

Fatemeh Aghaeipoor is currently a postdoctoral researcher at IPM, school of Computer Science. She received her B.Sc. degree in Software Engineering from Iran University of Science & Technology (IUST), and her M.Sc. and Ph.D. in Artificial Intelligence from Shahid Bahonar University. She was also a visiting researcher in the CITIC research center at the University of Granada, Spain, from 2019 to 2020 and in the Aristotle University (AUTH), Thessaloniki, Greece, in 2018. Her main areas of interest are eXplainable AI (XAI), ML/DL, Computer Vision, Fuzzy Logic.

زمان : چهارشنبه ۱۴۰۲/۰۹/۲۹– ساعت ۱۵:۰۰

ارائه به صورت مجازی انجام خواهد شد.

https://vmeeting2.ipm.ir/b/com-hh1-n07-vil

*****شرکت برای عموم علاقه‌مندان آزاد است*****