IPM
پژوهشگاه دانش‌های بنیادی
پژوهشکده علوم کامپیوتر

سخنرانی علمی

Explainability of Deep Neural Networks using Fuzzy Systems

فاطمه آقایی پور
پژوهشگاه دانش های بنیادی

Abstract

Explainability of deep neural networks has been receiving increasing attention with regard to auditability and trustworthiness purposes. Of the various post-hoc explainability approaches, rule extraction methods assist to understand the logic that underpins their functioning. Whereas the rule-based solutions are directly managed and understood by practitioners, the use of intervals or crisp values in the antecedents that rely on numerical values might not be intuitive enough. In this case, the benefits of a linguistic representation based on fuzzy sets/rules are straightforward, as these semantically meaningful components ease the model understanding. In this regard, we propose fuzzy rule-based explainer systems for deep neural networks. The algorithm learns a compact yet accurate set of fuzzy rules based on features' importance (i.e., attribution values) distilled from the trained networks. These systems can be used for both local and global explainability purposes. The evaluation results of different applications revealed that the fuzzy explainers maintained the fidelity and accuracy of the original deep neural networks while implying lower complexity and better comprehensibility.

Biography

Fatemeh Aghaeipoor is currently postdoctoral researcher at IPM, school of Computer Science. She received her B.Sc. degree in Computer Engineering from Iran University of Science & Technology (IUST), and her M.Sc. and Ph.D. in Artificial Intelligence from Shahid Bahonar University of Kerman. She was working as a researcher at the University of Granada, Spain, in 2020. Her main areas of interest are Machine Learning, XAI, Big Data, Fuzzy Logic.

زمان : چهارشنبه ۱۴۰۱/۰۶/۳۰– ساعت ۱۵:۰۰

ارائه به صورت مجازی انجام خواهد شد.

https://vmeeting.ipm.ir/b/com-uwd-qpp

شرکت برای عموم علاقه‌مندان آزاد است