Computer Science
Defense

Interpretable and Interactive Representation Learning on Geometric Data

Yuyang Guo
Emory University

Abstract: Abstract:
In recent years, representation learning on geometrics data, such as image and graph-structured data, are experiencing rapid developments and achieving significant progress thanks to the rapid development of Deep Neural Networks (DNNs), including Convolutional Neural Networks (CNNs) and Graph Neural Networks (GNNs). However, DNNs typically offer very limited transparency, imposing significant challenges in observing and understanding when and why the models make successful/unsuccessful predictions. While we are witnessing the fast growth of research in local explanation techniques in recent years, the majority of the focus is rather handling how to generate the explanations, rather than understanding whether the explanations are accurate/reasonable, what if the explanations are inaccurate/unreasonable, and how to adjust the model to generate more accurate/reasonable explanations. ¡br¿ To explore and answer the above questions, this dissertation aims to explore a new line of research called Explanation-Guided Learning (EGL) that intervenes the deep learning models' behavior through XAI techniques to jointly improve DNNs in terms of both their explainability and generalizability. Particularly, we propose to explore the EGL on geometric data, including image and graph-structured data, which are currently under-explored in the research community due to the complexity and inherent challenges in geometric data explanation. ¡br¿
To achieve the above goals, we start by exploring the interpretability methods for geometric data on understanding the concepts learned by the deep neural networks (DNNs) with bio-inspired approaches and propose methods to explain the predictions of Graph Neural Networks (GNNs) on healthcare applications. Next, we design an interactive and general explanation supervision framework GNES for graph neural networks to enable the learning to explain pipeline, such that more reasonable and steerable explanations could be provided. Finally, we propose two generic frameworks, namely GRADIA and RES, for robust visual explanation-guided learning by developing novel explanation model objectives that can handle the noisy human annotation labels as the supervision signal with a theoretical justification of the benefit to model generalizability. ¡br¿
This research spans multiple disciplines and promises to make general contributions in various domains such as deep learning, explainable AI, healthcare, computational neuroscience, and human-computer interaction by putting forth novel frameworks that can be applied to various real-world problems where both interpretability and task performance are crucial.

Thursday, December 1, 2022, 11:00 am
https://emory.zoom.us/j/5693008550

Computer Science
Emory University