EXPLAINABLE AI: MAKING VISUAL QUESTION ANSWERING SYSTEMS MORE TRANSPARENT

Raymond J. Mooney is a Professor in the Department of Computer Science at the University of Texas at Austin. He received his Ph.D. in 1988 from the University of Illinois at Urbana/Champaign. He is an author of over 170 published research papers, primarily in the areas of machine learning and natural language processing.


Artificial Intelligence systems’ ability to explain their conclusions is crucial to their utility and trustworthiness. Deep neural networks have enabled significant progress on many challenging problems such as visual question answering (VQA), the task of answering natural language questions about images. However, most of them are opaque black boxes with limited explanatory capability. The goal of Explainable AI is to increase the transparency of complex AI systems such as deep networks.

We have developed a novel approach to XAI and used it to build a high-performing VQA system that can elucidate its answers with multi-modal natural-language and visual explanations that faithfully reflect important aspects of its underlying reasoning while capturing the style of comprehensible human explanations. Crowd-sourced human evaluation of these explanations demonstrate the advantages of our approach.