Venue: ASCE International Conference on Computing in Civil Engineering, 2024

Testing of Generative AI for Disaster Scene Computing and Human-in-the-loop Post-disaster Decision-Making

Kailynn Barnt and ZhiQiang Chen

Generative AI, specifically large multimode models (LMM), is transforming the practice of machine learning or conventional AI for applying predictive analytics in engineering and many other fields. In Civil Engineering, one traditionally hard problem, image-based built-environment understanding, has become surprisingly easy. In our testing, using an LMM engine, one can readily obtain semantics-rich outputs, including pixel-level parsing and descriptive captions. The underlying learning is realized through a hybrid set of generative supervised deep learning and zero-shot learning (ZSL) models; for the latter, an inference is created based on the parsed patterns and spatial contexts, and auxiliary information is generated for a general description. However, it is recognized that, due to the underlying complexity of mixed learning models in an LMM architecture, no confidence or posterior-probability scores are generated, which hinders the design of a human-in-the-loop (HITL) decision-making process. In our case, we focus on applying LMM for disaster-scene understanding, including image-based parsing and damage recognition toward rapid decision-making for post-disaster recovery. We argue that in such a process, humans should not wait for the output from a GAI-driven workflow as 'end-users'; rather, they should collaboratively participate in different phases, including crowd-based data collection, cleaning, report generation, and visual analytics creation. To realize this process, generating confidence scores from GAI models is crucial as it provides the mechanism for deciding when a human intervention is necessary when a low GAI score is encountered (for example, a human user needs to provide a new class label for a miss in detecting an object or a wrong caption over a disaster scene image. Using a data-centric approach and an open-source LMM framework, we curate a disaster scene database and test the LMM outputs against human generation, based on which we manually grade GAI outputs as the ground-truth confidence scores. By differentiating the instances of largely supervised inference and ZSL inference, we further develop a predictive model to generate the GAI confidence score in terms of a set of intermediate measurements from the LMM model. We will demonstrate that such GAI confidence scores are instrumental in developing a HITL decision-making workflow for achieving rapid and objective disaster-scene-enabled post-disaster response.

Short version:
Generative AI, specifically large multimode models (LMM), is transforming the practice of machine learning or conventional AI for applying predictive analytics in engineering and many other fields. In Civil Engineering, one traditionally hard problem, image-based built-environment understanding, has become surprisingly easy. In our testing, using an LMM engine, one can readily obtain semantics-rich outputs, including pixel-level parsing and descriptive captions. The

underlying learning is realized through a hybrid set of generative supervised deep learning and zero-shot learning (ZSL) models. However, it is recognized that, due to the underlying complexity of mixed learning models in an LMM architecture, no confidence or posterior-probability scores are generated, which hinders the design of a human-in-the-loop (HITL) decision-making process. In our case, we focus on applying LMM for disaster-scene understanding to achieve rapid post-disaster recovery. We argue that in such a process, humans should not wait for the output from a GAI-driven workflow as 'end-users'; instead, they should collaboratively participate in different phases, including crowd-based data collection, cleaning, report generation, and mapping product creation. To realize this process, generating confidence scores from GAI models is crucial as it provides the mechanism for deciding whether a human intervention is necessary when a low GAI score is encountered. Using a data-centric approach and an open-source LMM framework, we manually grade GAI outputs as the ground-truth confidence scores given a curated disaster-scene database. We further develop a predictive model to generate the GAI confidence score relating to a set of intermediate measurements from the LMM model. We will demonstrate that such GAI confidence scores are instrumental in developing a HITL decision-making workflow for rapid and objective disaster-scene-enabled post-disaster response.