Research Summary

Thu Trang Nguyen
thu.nguyen@ucdconnect.ie

Understanding how a predictive model makes a decision is an increasingly important area of machine learning research. This task is crucial in time series classification applications in which users need to know which parts of the time series are relevant for the classification prediction. For example, when doing a physio exercise, a patient receives feedback on whether the execution is correct or not (classification), and if not, which parts of the motion are incorrect (explanation), so the patient can take remedial action. While there is a plethora of techniques to generate explanations for time series classification algorithms, their outcomes are often inconsistent. This inconsistency reduces trust of application users with regard to the explanation techniques. Automated, computer-aided methods of evaluating and comparing explanations, can help to filter out problematic explanation techniques and to identify trustworthy ones. My research aims to study and develop approaches to automatically and computationally assess saliency-based explanation methods for time series classification. We extract explanation weights for each point in the time series, use these weights to perturb specific parts of the time series, and measure the impact on classification accuracy. By this perturbation process, we show that explanations that highlight discriminative sections of the time series lead to significant changes in classification accuracy, enabling the objective quantification and ranking of different explanations.

My current research proposes a method to efficiently evaluate any saliency-based explanation methods for time series classification, objectively quantify their usefulness with appropriate metrics, and assist application users in focusing on the most useful explanations for their specific use cases. Future works include designing an optimized explanation method that maximize informativeness metric and demonstrating an use-case in which explanation can assist human in learning new knowledge from data, for example, the subsequences DNA that is responsible for a specific medical condition.