Bowen Zhang

Topic: ER Explanations

Why should I trust you? Explaining the predictions of any classifier

Q1: How can be data leakage be distinguished? Would it be done manually or can it be identified by some learning methods?

Q2: Would keep the global fidelity with local fidelity together for the fidelity measurement represent better how good the model is?

Q3: How will the result of *LIME* guide us to change the training process?

On the robustness of interpretability methods

Q1: How will all the locally Lipschitz be summarized as the robustness?

Q2: How will the robustness method guide the learned model to be more robust?

Q3: Would it be more informative if the robustness is calculated as the information gain of the model/explanation from the different input?

Interpreting deep learning models for entity resolution: an experience report using LIME

Q1: Recall the paper “Deep Learning for Entity Matching: A Design Space Exploration” we read in last lecture, could breaking the class matching into textual matching a good idea for weight the matching and get the explaination?

Q2: How could *Mojito* deal with the problem of data leakage as mentioned in the first paper?

Q3: How can we use *Mojito* to make our learning more interpretable?