

Solution to Homework 1

First question. Why do we need explainable AI in the first place?

Answer. We all make mistakes:

- humans make mistakes,
- computer systems give wrong answers.

We can often deal with human mistakes:

- If you do not agree with the instructor's grade, you can ask why the grade was lowered and, based on this explanation, argue that you deserve more credit.
- If you do not get the bank loan and the folks explain the reason, you have a chance to argue – or at least to know how to get a better chance next time.

The problem is that many AI programs – e.g., the ones that use deep learning – just provide a recommendation, they do not come with any explanation. We know that these programs are not perfect, that sometimes their recommendations are wrong – but since there are no explanations, we do not know which recommendations are wrong.

It is therefore desirable to have such explanations.

Second question. Why is it a reasonable idea to try to use fuzzy techniques when designing explainable AI?

Answer. Desire for explanations means that we need to be able to transform numerical recommendations into natural-language explanations. In other words, we need to connect numerical recommendations with natural-language rules.

Such a connection has been explored before: this is exactly what fuzzy techniques are about.

Third question. Why were fuzzy technique invented in the first place – and who was their inventor?

Answer. Fuzzy techniques were designed by Lotfi Zadeh who realized that a large part of expert experience – namely, the rules the experts formulate in terms of imprecise words from natural language – are not used in automatic control. So, he designed fuzzy techniques to translate experts' natural-language rules into precise control strategies.