Explainability: New Application and New Promise of Fuzzy Techniques

Julia Taylor Rayz, Victor Raskin, Scott Dick, and Vladik Kreinovich

World has always been run by experts. For millennia, the world has been run by experts: medical doctors used their skills and their intuition to cure patients, engineers used their skills and their intuition to design building, bridges, etc., entrepreneurs used their intuition to run companies, generals use their skills and their intuition to win battles, Sherlock Holmes's used their skills and their intuition to catch criminals, etc.

Why is this a problem? An obvious problem with this arrangement is that there are very few very good top experts, not enough to solve all the problems. As a result, the rest of us has to rely on the services of less experienced experts. Kings could use the best medical doctors, but when a simple peasant got sick, the person who took care of his or her illness was clearly not so skilled.

Another problem is that even top experts are sometimes wrong. John Le Carre, a popular author of spy novels, describes – in his novel A Little Town in Germany – a top expert as a person who only asks a question when he/she knows the answer. Largely, this is true: a top medical doctor usually knows the diagnosis before the tests confirm it, a top physicist intuits the results of the experiment, a top mathematician knows whether the statement is true before a proof is found, etc. But even top experts make mistakes. Most bridges designed by top engineers cause our awe but some of them spectacularly collapsed. Genius 20 century airplane designers had many brilliant successes – and several catastrophic designs. David Hilbert, the top mathematician of the late 1890s, when asked to present 23 important challenges to the 20 century mathematicians, guessed most answers correctly, but not all: e.g., his 13th problem was to find a function of three variables that cannot be represented as compositions of functions of two variables – and it turned out that this is impossible. One may say that $1/23 \approx 5\%$ error rate is very low – but do we really want 5% of the population to be sitting in jail without any crime just because of Sherlock Holmes's mistakes?

Leinbiz's dream. As science and engineering developed, many things that were previously based on intuition became the subject of exact equations. Engineers used formulas to design buildings and bridges, even medical doctors started using some formulas to decide on the dosage of medicine. The famous 16-17 century philosopher Leibniz – a co-author of calculus and the author of the binary system that all computers use – had a dream that some day, mathematization will reach a level at which we would not need experts, we would not need informal arguments – we will just calculate and see who is right and what to do.

Leibniz's dream starts coming true. This is exactly what happened in the 19 century and in the first half of the 20 century: more and more equations have been discovered, better and better computational devices enables us to solve these equations. Control was no longer the domain only of experienced operators – automatic controllers successfully operated factories and even airplanes. Companies were no longer run based by intuition – technocrats provided mathematical models that led to new successes. Even in military applications, game theory – heaving financed by defense all over the world – promised to largely replace the generals' intuition.

By the 1960s, results were not always perfect, but the feeling was that with new faster computers – and computers did become faster year after year – Leibniz's dream will finally come true. Medical expert systems will replace human doctors, robots will replace skilled workers.

But it turned out that experts are still needed. However, by the mid-1960s, it became clear to several researchers that the original hopes were too optimistic. One of these researchers was Lotfi Zadeh, one of the authors of the then most popular book on automatic control. He realized that one of the reasons why even the best automatic controllers were sometimes not as efficient as human experts is that human experts possess additional knowledge – which is difficult to incorporate into an automatic controller because it uses imprecise ("fuzzy") words from natural language like "small".

To describe such knowledge in precise (and thus, computer-understandable) terms, Zadeh came up with the ideas of fuzzy logic and fuzzy techniques. This technique establishes a correspondence between natural-language knowledge and precise numerical formulas and dependencies.

Fuzzy boom and why it slowed down. After a few years, Zadeh's ideas led to numerous successful applications, from fuzzy rice cookers and washing machines to fuzzy controllers for elevators, cars, and trains.

Every time we did not have exact equations, exact data, using expert intuition – translated, by fuzzy techniques, into precise control strategy – helped a lot. But of course, more and more equations became known, so the need for expert knowledge decreased – in many cases when previously we had to rely on expert knowledge, now exact equations and known and an optimal control is possible.

What is happening now. The more adequately we describe a system, the more complex the corresponding equations. At some point, it becomes not realistic to solve these equations by exact guaranteed methods.

Instead, practitioners started using efficient – although not guaranteed – methods of machine learning, where, based on several known cases with known solutions, the system tries to find a similar solution for new cases. Many current machine learning methods like deep learning are very good – they control self-driving cars, they help companies decide who to hire, they help banks decide to whom to give loans – they are everywhere.

And again, there is a problem. Many systems based on machine learning are very good – but they are not perfect. If a system for solving crimes is 95% accurate, this is a great achievement – but we can now repeat the same question: do we really want to see 5% of the population in jail just because of the program's errors? More generally, do we want 5% of the population – millions – to be treated unfairly just because of the program's imperfection?

So what is a solution? Modern machine learning techniques are like top experts in the old days – they are usually good, but sometimes, they fail. So what can we do?

For a human expert – e.g., a medical doctor – the solution was to ask for his/her arguments, to have him/her discuss it with other medical doctors. Unfortunately, we cannot do it for a computer program, they are more like idiot savants rather than top experts – they tell us the recommendations but they cannot explain how they came up with these recommendations.

Since the current techniques do not provide such explanations, we need to learn how to produce them. How can we do that? What we need is to translate numerical results into a natural-language description. In other words, what we need is to invoke a correspondence between natural-language knowledge and precise numerical formulas and dependencies – and this correspondence is exactly what fuzzy techniques provide! So, what we need is to learn how to use fuzzy techniques to make results of machine learning explainable.

What is in this book. Some of the papers presented in this book do exactly this: they show how fuzzy techniques can lead to explainable. At present, there are not too many such cases, this is still work in progress.

For this ultimate goal to succeed, we need to solve many challenging problems in fuzzy techniques itself – and be able to better tune the current methods on new applications. This is what most other papers are about – a steady progress in many aspects of fuzzy.

We hope that this book – and similar books on explainable fuzzy AI – will boost this important research area.

Our thanks. This book is based on papers from the 2021 Annual Conference of the North American Fuzzy Information Processing Society. We want to thank everyone who helped organize this conference; we also want to thanks the authors for selecting this venue for their interesting results, we want to thank the reviewers for their hard work, we want to thank the conference participants for their interest, and – last but not the least – we want to thank Professor Janusz Kacprzyk and the Springer staff for agreeing to (and helping to) produce this volume. Our sincere thanks for all of you!