TOPICS IN INTELLIGENT COMPUTING/TOPICS IN SOFT COMPUTING:
FOUNDATIONS OF EXPLAINABLE FUZZY AI
Syllabus for the course CS 5354/4365, Summer 2021

CLASS TIME: MTWRF 11:40-1:50 pm

INSTRUCTOR: Vladik Kreinovich, email vladik@utep.edu, office phone (915) 747-6951.

- The instructor's office hours are MTWR 10:40-11:40 am or by appointment.
- Preferable way of contact is email to vladik@utep.edu
- If you want to contact during the scheduled office hours, there is no need to schedule an appointment.
- If you are not available during the instructor's scheduled office hours, please schedule an appointment in the following way:
    - use the instructor's appointments page http://www.cs.utep.edu/vladik/appointments.html to find the time when the instructor is not busy (i.e., when he has no other appointments), and
    - send him an email, to vladik@utep.edu, indicating the day and time that you would like to meet.

  He will then send a reply email, usually confirming that he is available at this time, and he will place the meeting with you on his schedule.

CONTENTS

Need for Explainable AI. Modern AI techniques -- especially deep learning -- provide, in many cases, very good recommendations where a self-driving car should go, whether to give a company a loan, etc. The problem is that these techniques are not (yet) perfect.

In some cases, the recommendations generate by an AI system are not good. Of course, as the famous Marilyn Monroe movie says, "Nobody's perfect". Human experts are not perfect either. However, when a human expert -- be it a banking official or a medical doctor -- makes a recommendation, he or she can, if asked, provide an explanation. If you find the explanation not sufficiently convincing, you can ask for someone else's advice.

Unfortunately, recommendations provided by an AI system (such as a deep neural network) usually come without an explanation. So we cannot so easily separate good and bad advice. It is therefore desirable to make AI more explainable.

Why Fuzzy Techniques. Providing an explanation means finding natural-language rules and ideas which are, in some reasonable sense, equivalent to the numerical results provided by the AI tools. The problem of connecting natural language rules and numerical decisions is known since 1960s. Then, the need was recognized to incorporate expert knowledge into control and decision making.

Experts use imprecise words like "small". For this incorporation, a special technique was invented -- known as fuzzy techniques. This technique led to many successful applications. It is therefore reasonable to use this techniques in designing explainable AI.

What We Will Study in This Class. If we knew how to make AI explainable, teaching this class would be easier. We would just teach the corresponding algorithms and methods.

At present, explainable AI remains largely an ultimate goal. We do not yet know which tools will work better. So, instead of studying specific tools, it makes sense to study the *foundations* for these tools, so that we will know why we need to use these tools, and we will know which tools are better in which situations. This will help us select appropriate tools for making current AI applications more explainable.

First Topic: Introduction to Fuzzy Techniques. We want to better understand how fuzzy techniques can help with explainable AI. For this, we need to have a good understanding of these techniques. We will learn the corresponding techniques and how they are used in control and in other applications.

We will also try to make these techniques themselves more explainable. Namely, we will explain first-principle motivations for these techniques.

We will study all three main stages of fuzzy techniques:

1. describing the original imprecise words like "small" in numerical terms,
2. combining the corresponding numbers; to describe boolean (and- and or-) combinations of the corresponding properties, special "and" and "or" operations are used for this;
3. "defuzzification" -- transforming imprecise recommendations into a precise control value.

Second Topic: Which Version of Fuzzy Technique to Select. In all three stages of fuzzy techniques, there are several different options. Empirically, in different situations, different options work best. This makes sense, since in different situations, we have different objectives. For example, if we launch a single drone to inspect an area, the main objective is to maximize the probability that its mission succeeds. On the other hand, if we launch a swarm of drones to inspect the same area, it is probably Ok if one of them does not do much -- as long as, on average, the overall mission is successful.

How do we select the best techniques? In some cases, we have finitely many parameters. So, we need to find the best values of these parameters. To find the largest and the smallest values of a function of several such variables, we can use calculus. (Do not worry if you have forgotten some of it, we will refresh.)

In many other cases, however, we need to select a function -- e.g., the best "and"- and "or"-operations. There is a natural generalization of calculus that deals with such optimization problems. It is known as *variational calculus*, and it is actively used in control. We will learn the basics of this techniques. As an example, we will use this technique to come up with optimal ``and"- and ``or"-operations for the two above-described drone situations.

Third Topic: Towards Explainable Machine Learning. The ultimate goal is to make the *results* of machine learning (and other AI techniques) explainable. We are still working on this.

Meanwhile, an important help would be to make the machine learning techniques themselves explainable. At present, in many cases, the only reason we select some techniques and some parameters of these technique is that these techniques empirically work well on several problems. This is not as convincing as when we prove that these techniques are, in some reasonable sense, optimal. We will analyze deep learning from this viewpoint.

We will show that many empirically successful features of deep learning can indeed be proven to be optimal. The corresponding proofs require the use of other techniques widely used in applications to the physical world, namely, the technique of symmetries. So, we will learn symmetry-related optimization techniques as well.

Projects. This course is somewhat on the theoretical side. However, it is important to have applications in mind. Students will be therefore encouraged to work on projects.

Ideally, projects should be related to real-life applications. Purely theoretical projects are also OK. A project may consist of reviewing some paper(s) on explainable AI, on its applications, and on its foundations.

It is also possible to select a more creative project. In such a project, students -- individually or in groups -- will come up with something new. Explainable AI is a new developing topic, there are many more open problems than results. Any progress in any of these open problems will bring us closer to the goal of making AI explainable.

MATERIALS USED IN THE CLASS: there is no textbook; instead, we will use several papers, see below

HOMEWORKS. Each topic means home assignments. I will post correct solutions. Since I will be posting correct solutions to homeworks, it does not make any sense to accept late assignments: once an assignment is posted, it make no sense for you to copy it in your own handwriting, this does not indicate any understanding. So, please try to submit your assignments on time.

Things happen. If there is an emergency situation and you cannot submit it on time, let me know, you will then not be penalized -- and I will come up with a similar but different assignment that you can submit when you become available again.

TESTS. There will be three tests. If you are unable to attend the test, let me know, I will organize a different version of the text at a time convenient for you.

GRADES: Maximum number of points:

- first test: 10
- second test: 10
- third test: 15
- home assignments: 10
- final exam: 35
- project: 20

(smart projects with ideas that can turn into a serious scientific publication get up to 40 points).

A good project can help but it cannot completely cover possible deficiencies of knowledge as shown on the test and on the homeworks. In general, up to 80 points come from tests and home assignments. So:

- to get an A, you must gain, on all the tests and home assignments, at least 90% of the possible amount of points (i.e., at least 72), and also at least 90 points overall;
- to get a B, you must gain, on all the tests and home assignments, at least 80% of the possible amount of points (i.e., at least 64), and also at least 80 points overall;
- to get a C, you must gain, on all the tests and home assignments, at least 70% of the possible amount of points (i.e., at least 56), and also at least 70 points overall.

WE WILL OVERCOME. Topics that we study in this class are not easy topics, but hopefully, you will all do well, it is not as difficult as many things you have successfully mastered in the CS classes so far.

SPECIAL ACCOMMODATIONS: If you have a disability and need special accommodations -- e.g., extra time on the exams -- please contact the Center for Accommodations and Support Services (CASS) at 747-5148 or by email to cass@utep.edu. For additional information, please visit the CASS website at http://www.sa.utep.edu/cass. CASS's staff are the only individuals who can validate and if need be, authorize accommodations for students.

SCHOLASTIC DISHONESTY: Any student who commits an act of scholastic dishonesty is subject to discipline. Scholastic dishonesty includes, but not limited to cheating, plagiarism, collusion, submission for credit of any work or materials that are attributable to another person.

Cheating is:

- copying from the test paper of another student;
- communicating with another student during a test to be taken individually;
- giving or seeking aid from another student during a test to be taken individually;
- possession and/or use of unauthorized materials during tests (i.e. crib notes, class notes, books, etc.);
- substituting for another person to take a test;
- falsifying research data, reports, academic work offered for credit.

Plagiarism is:

- using someone's work in your assignments without the proper citations;

- submitting the same paper or assignment from a different course, without direct permission of instructors.

To avoid plagiarism see: https://www.utep.edu/student-affairs/osccr/_Files/docs/Avoiding-Plagiarism.pdf

Collusion is unauthorized collaboration with another person in preparing academic assignments.

Instructors are required to -- and will -- report academic dishonesty and any other violation of the Standards of Conduct to the Dean of Students.

Notes: When in doubt on any of the above, please contact your instructor to check whether you are following an authorized procedure.

MATERIALS.

Kelly Cohen, Laxman Bokati, Martine Ceberio, Olga Kosheleva, and Vladik Kreinovich, "Why Fuzzy Techniques in Explainable AI? Which Fuzzy Techniques in Explainable AI?", Proceedings of the Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2021, West Lafayette, Indiana, June 7-9, 2021, to appear.
pdf file

Lectures on fuzzy logic
pdf file

Hung T. Nguyen, Vladik Kreinovich, Bob Lea, Dana Tolbert. "How to control if even experts are not sure: robust fuzzy control". Proceedings of the Second International Workshop on Industrial Applications of Fuzzy Control and Intelligent Systems, College Station, December 2-4, 1992, pp. 153-162.
pdf file

Hung T. Nguyen, Vladik Kreinovich, Dana Tolbert. "On robustness of fuzzy logics". Proceedings of the 1993 IEEE International Conference on Fuzzy Systems FUZZ-IEEE'93, San Francisco, California, March 1993, Vol. 1, pp. 543-547.
pdf file

Hung T. Nguyen, Vladik Kreinovich, and Dana Tolbert, "A measure of average sensitivity for fuzzy logics", International Journal on Uncertainty, Fuzziness, and Knowledge-Based Systems, 1994, Vol. 2, No. 4, pp. 361-375.
pdf file

Olga Kosheleva and Vladik Kreinovich, "Finding the Best Function: A Way to Explain Calculus of Variations to Engineering and Science Students", Applied Mathematical Sciences, 2013, Vol. 7, No. 144, pp. 7187-7192.
pdf file

Vladik Kreinovich and Olga Kosheleva, "Optimization under uncertainty explains empirical success of deep learning heuristics", In: Panos Pardalos, Varvara Rasskazova, and Michael N. Vrahatis (eds.), Black Box Optimization, Machine Learning and No-Free Lunch Theorems, Springer, Cham, Switzerland, 2021, pp. 195-220.
pdf file