Why Storytelling Is Good for Education? Why Practitioners Stopped Eliciting Membership Functions and “And”- and “Or”-Operations from the Experts?

Christian Servin[1], Olga Kosheleva[2], and Vladik Kreinovich[3]

Information Technology Systems Department
El Paso Community College (EPCC)
919 Hunter Dr., El Paso, TX 79915-1908, USA, cservin1@epcc.edu
Departments of [2]Teacher Education and [3]Computer Science,
University of Texas at El Paso,
500 W. University, El Paso, TX 79968, USA,
olgak@utep.edu, vladik@utep.edu

1. The first problem

- In many disciplines – such as mathematics – we want to teach students rigorous thinking.
- However, empirical evidence shows that:
  - many successful teaching techniques – including techniques for teaching mathematics
  - use informal storytelling.
- A natural question is: why?

2. The second problem

- There are many successful applications of fuzzy techniques.
- In particular, there are many successful applications to control.
- These applications usually have three stages:
  - first, we elicit imprecise ("fuzzy") rules form the experts,
  - then, we use fuzzy techniques to transform these rules into a precise control strategy, and
  - finally, if needed, we tune the resulting controller to make it most effective.
- The first and the third stages are, in essence, the same now as they were decades ago.
- However, in implementing the second stage, there is a drastic different between what was done earlier, and what is done now:

3. The second problem (cont-d)

- When fuzzy applications started, designers of the corresponding fuzzy systems spent a lot of times eliciting,
  - from the experts,
  - the exact shapes of the membership functions
  - and, often, the exact "and"- and "or"-operations that best describe the expert's thinking.
- In contrast, at present, this is practically never done.
- The designers use simple – e.g., triangular – membership functions and simple – min or product – "and"-operations (t-norms).
- This change saves the designers a lot of effort.

4. The second problem (cont-d)

- For example, the designers of the world's first effective medical expert system MYCIN:
  - spent several years – and several million dollars
  - finding out which "and"-operation best describes the reasoning of medical doctors.
- This does not mean that the designers use only simple membership functions and "and"- and "or"-operations in their designs.
- The resulting systems often use complex membership functions and complex "and"- and "or"-operations.
- However, this adjustment of membership functions and "and"- and "or"-operations happens only on the third – tuning – stage.
- This is no longer done by first eliciting this information from the users.

5. The second problem (cont-d)

- So, now we elicit much less information from the experts than in the past.
- However, still the results are as good.
- Intuitively, the more information we get form the experts, the better our system should be.
- However, it looks like in this case, the corresponding additional information does not improve the quality of the resulting system.
- Indeed, the absence of this additional information does not make the systems perform any worse.
- But why? How can we explain this?
- In this talk, we provide answers to both questions – and we explain the relation between these two answers.

6. Why practitioners stopped eliciting membership functions and "and"- and "or"-operations

- Both membership functions and "and"- and "or"-operations deal with fuzzy degrees.
- Fuzzy degrees that describe the expert's degree of confidence in different statements.
- According to the well-known seven-plus-minus-two law, a person can meaningful distinguish between 5 and 9 different degrees.
- For most people, it is 7 degrees.
- From the purely mathematical viewpoint, fuzzy degrees can take infinitely many values.
- Namely, they can take all possible real values from the interval $[0, 1]$.
- However, in reality, most experts can only meaningfully distinguish 7 different degrees.

7. Why practitioners stopped eliciting membership functions and “and”- and “or”-operations (cont-d)

- One of these degrees is “absolutely true” – that corresponds to 1, another is “absolutely false”, which corresponds to 0.
- This leave us with 5 degrees corresponding to uncertainty – and thus, located inside the open interval $(0, 1)$.
- To describe the most appropriate numerical values corresponding to these 5 degrees, we need to select 5 points in this open interval.
- We have no reason to believe that the distance between any pair of neighboring degrees is larger than for any other pair.
- So, it makes sense to assume that all these distances are equal.
- So, each of the 7 degrees (including 0 and 1) is located at the same distance d from the neighboring degrees.
- Since the first value is 0, the next is thus d, the next after that is $2d$, etc.

8. Why practitioners stopped eliciting membership functions and "and"- and "or"-operations (cont-d)

- The final seventh degree is $6d$.
- Since we know that the final seventh degree is 1, this means that $6d = 1$, so $d = 1/6$.
- Thus, the corresponding degrees are 0, 1/6, 1/3, 1/2, 2/3, 5/6, and 1.
- To make computations more convenient, let us use the scale $[0, 6]$ instead of the usual $[0, 1]$.
- In this scale, the actual degrees are integers: 0, 1, 2, 3, 4, 5, and 6.
- In this scale, the two simplest "and"-operations – min and product – get the following form:
  - minimum and maximum remains minimum and maximum, while
  - the product becomes $r/6$ for $r = a \cdot b$.
- For integers a and b, the value $v \stackrel{\text{def}}{=} (a \cdot b)/6$ is sometimes not an integer.

9. Why practitioners stopped eliciting membership functions and “and”- and “or”-operations (cont-d)

- Since in this scale, only integer degree values make sense, we will round the value into an integer V by using the usual rounding rules:

v	$[0, 0.5)$	$[0.5, 1.5)$	$[1.5, 2.5)$	$[2.5, 3.5)$	$[3.5, 4.5)$	$[4.5, 5.5)$	[5.
V	0	1	2	3	4	5	

- One can easily show that in terms of the integer-valued r, this rounding leads to the following rules:

r	$[0, 2]$	$[3, 8]$	$[9, 14]$	$[15, 20]$	$[21, 26]$	$[27, 33]$	$[33, 36]$
V	0	1	2	3	4	5	6

10. So how will the product look like in this framework?

- Let us use this framework to see how the product will look like.
- For example, for $a = 3$ and $b = 5$, the actual product is $r = 15$.
- So, based on the above rules, we take the degree 3 as the resulting value.
- By repeating these calculations for all possible pairs (a, b), we get the following table:

	0	1	2	3	4	5	6
0	0	0	0	0	0	0	0
1	0	0	1	1	1	1	1
2	0	1	1	1	2	2	2
3	0	1	1	2	2	3	3
4	0	1	1	2	2	3	4
5	0	1	1	3	3	4	5
6	0	1	2	3	4	5	6

11. So how will the product look like in this framework (cont-d)

- One can see that all these values are either equal to $\min(a, b)$ or differ from $\min(a, b)$ by one – i.e., by the smallest distinguishable difference.

12. What about other "and"-operations? what about "or"-operations?

- One of the possible ways to elicit a fuzzy degree is to ask n experts.
- If m of them say "yes", take the value m/n.
- In this case, from the purely mathematical viewpoint, the degree can be viewed as a probability.
- Namely, it is the probability that a randomly selected expect says "yes".
- In general, for random events, the probability that both occur is between $\max(a+b-1,0)$ and $\min(a,b)$.
- When the correlation is non-negative, this probability is between $a \cdot b$ and $\min(a,b)$.
- In fuzzy application, "and" is usually applied to conditions from the same rule provided by the same expert.

13. What about other "and"-operations? what about "or"-operations (cont-d)

- So it is reasonable to expect these conditions to be non-negatively correlated.
- Thus, the corresponding "and"-operations are between $a \cdot b$ and min.
- We already know that the results of $a \cdot b$ and $\min(a, b)$ differ by no more than one level.
- Thus, the same is true for any two operations between these two.
- So, no matter what "and"-operation the expert actually uses, the result is minimally different from min.
- This is why current practitioners simply use min instead of the time-consuming elicitation.
- For any "or"-operation $f_\vee(a, b)$, we can use the same result.

14. What about other "and"-operations? what about "or"-operations (cont-d)

- Indeed, as is well known, its dual $1 - f_\vee(1 - a, 1 - b)$ is an "and"-operation – and thus, close to min.
- Therefore, the original "or"-operation is close to the dual of min, i.e., to max.

15. What about membership functions?

- Why do practitioners use piece-wise linear membership functions?
- Any smooth function can be expanded into Taylor series, in which the next term after linear is quadratic, proportional to x^2.
- For values $a \in [0, 6]$, the original term x^2 corresponds to $a^2/6$ – the diagonal terms in the above table.
- We can see that these terms are minimally different from a.
- So, quadratic terms lead to an almost the same result as linear.
- This explains why practitioners do not consider such terms – and thus, do not consider higher order terms.
- As a result, they only use piecewise linear functions.
- All this explains why practitioners nowadays do not elicit membership functions and "and"- and "or"-operations.

16. But why storytelling?

- In precise sciences, learning means moving from a wrong precise statement to a correct precise statement.
- For example, it means moving from the wrong $3 \cdot 3 = 6$ to the correct $3 \cdot 3 = 9$.
- This means changing, for each of these statements, our degree of confidence in them from 0 to 1 or from 1 to 0.
- To many students, such an abrupt change is difficult.
- To make learning easier, it is desirable to replace a single abrupt change with two or more changes which are smaller than by 1.
- For each intermediate statement, the degree should thus be inside $(0, 1)$.
- What does that mean?
- Degrees inside $(0, 1)$ were introduced by Zadeh to describe imprecise ("fuzzy") natural-language statements.

17. But why storytelling (cont-d)

- So, intermediate statements should be imprecise natural-language statements.
- This is exactly what storytelling is about – so this explains why storytelling is effective in education.

18. How is this related to the second problem?

- In the traditional approach to fuzzy, it is not enough to say that we have degrees between 0 and 1 to be able to apply fuzzy techniques.
- We also need to select the shapes of the membership functions, to select "and"- and "or"-operations, etc.
- What we showed is that for the simple purpose of describing expert knowledge, this additional knowledge is, in effect, not needed.
- Different selections lead to almost same (minimally distinguishable) results.
- From this viewpoint, once we see that it is reasonable to use degrees between 0 and 1, we can already use fuzzy techniques.
- For storytelling, this means that we do not need to worry about membership functions or about "and"- and "or"-operations.
- We can, in effect, use whatever we want, and the results will be good – as indeed they are when we use storytelling in education.

19. Acknowledgments

This work was supported by:

- the AT&T Fellowship in Information Technology,
- the Institute for Risk and Reliability, Leibniz Universitaet Hannover, Germany,
- the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) Focus Program SPP 100+ 2388, Grant Nr. 501624329,
- the European Union under the project ROBOPROX (No. CZ.02.01.01/00/22 008/0004590),
- the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand,
- the Ho Chi Minh City University of Banking, Vietnam, and
- Thang Long University, Hanoi, Vietnam.