

Everything Is a Matter of Degree: The Main Idea Behind Fuzzy Logic Is Useful in Geosciences and in Authorship

Christian Servin¹, Aaron Velasco², Edgar Daniel Rodriguez Velasquez³,
and Vladik Kreinovich⁴

¹Information Technology Systems Department, El Paso Community College (EPCC)
919 Hunter Dr., El Paso, TX 79915-1908, USA, cservin1@epcc.edu

²⁻⁴Departments of ²Earth, Environmental, and Resource Sciences,

³Civil Engineering, and ⁴Computer Science

University of Texas at El Paso, El Paso, Texas 79968, USA

aavelasco@utep.edu, edrodriguezvelasquez@miners.utep.edu, vladik@utep.edu

1. Formulation of the problem

- One of the main ideas behind fuzzy logic is Zadeh's idea that everything is a matter of degree.
- This idea has been very fruitful in many application areas.
- In this paper, we show that there are still many new areas where this idea can be successfully applied.
- Specifically, we show that it can help to solve puzzling questions in such diverse applications areas as geosciences and authorship.

2. Importance of earthquake studies

- Earthquakes are among the most devastating events.
- Their effect depends on our preparedness:
 - if we know that a certain area is prone to have earthquakes of certain magnitude,
 - then we can strengthen all the buildings and structures,
 - so as to minimize the earthquake's damaging effect.
- This strengthening is reasonably expensive.
- So it is only used when we are reasonably confident that earthquakes of such magnitude are possible.
- Because of this, predicting the magnitudes and location of possible future earthquakes is one of the main objectives of geosciences.

3. Seismogenic zones: traditional approach to earthquake study

- Up to the 19th century, scientists believed that continents remain the same.
- Then it turned out that continents drift with time.
- To be more precise, plates containing continents (or parts of the continents) drift with time.
- This knowledge formed what is now called *plate tectonics*.
- It was noticed that most strong earthquakes – as well as most volcanos – occur in the borders between these plates.
- Later on, it was found that plates themselves are not unchangeable.
- They consist of smaller pieces called *terranes* that can also drift with respect to each other.
- The vast majority of earthquakes occurs close to the *faults* – boundaries of terranes.

4. Seismogenic zones: traditional approach to earthquake study (cont-d)

- This is still the main approach to predicting earthquakes.
- Researchers usually assume that the earthquakes occur only at the faults; so:
 - at the faults, they recommend engineering measures to mitigate the effect of possible earthquakes, while
 - in the areas inside the terranes no such measures are recommended.

5. Recent discovery

- A recent statistical analysis of earthquake records has shown that:
 - contrary to the above-described traditional beliefs,
 - earthquakes are not limited to the faults.
- Most earthquakes do occur at the faults.
- However, there have been earthquakes in other areas as well:
 - as we move away from the fault,
 - the probability of an earthquake decreases but it never goes to 0.

6. What does this mean in terms of seismogenic zones?

- In the traditional approach, there was a clear (crisp) distinction between:
 - seismogenic zones where earthquakes are possible and
 - other areas where earthquakes are not possible.
- The recent discovery shows that earthquakes are possible literally everywhere.
- At first glance, this implies that the whole Earth is a seismogenic zone.
- However, this would be a useless conclusion.
- There should be a difference between:
 - zones where earthquakes are frequent – e.g., near the major faults
 - and
 - zones where earthquakes are so rare that it took several decades to notice them.

7. What does this mean in terms of seismogenic zones (cont-d)

- In other words, some areas are clearly seismogenic zones, while other are barely seismogenic.
- In other words, being a seismogenic zone is not a crisp property, it is a matter of degree:
 - some areas are more seismogenic,
 - some are less seismogenic.
- This is perfectly in line with the main idea that Zadeh placed in the foundation of fuzzy logic – that everything is a matter of degree.

8. What is the physical meaning of this phenomenon

- The traditional approach implicitly assumes that a fault is a line.
- In such a description, we can easily separate regions close to the line from regions which are far away from the line.
- A similar description was thought to hold when we describe visible cracks: e.g., cracks in rocks, cracks in pavement, etc.
- A more detailed analysis has shown that visible cracks actually have a fractal structure:
 - there is a main crack line, along which the stress is high,
 - at several points in the main line, it branches into then smaller-size crack lines, along which the stress is somewhat smaller;
 - at several points in each of these “second-order” lines, the line itself branches into even smaller-size crack lines, etc.

9. What is the physical meaning of this phenomenon (cont-d)

- Because of this structure, there is, in effect, no area completely without cracks and without stress:
 - there are areas around the main fault line, in which the crack is the most visible and the stress is the highest;
 - there are areas around the second-order fault lines, where the crack is less visible and the stress is somewhat lower;
 - there are areas around the “third-order” fault lines, where the crack is even less visible and the stress is even lower, etc.,
 - all the way to areas where cracks are microscopic and the stress is barely measurable.
- This is what we directly observe in rock cracks, in the pavement cracks, and this is what we indirectly observe for earthquakes.

10. What is the physical meaning of this phenomenon (cont-d)

- Earthquakes can appear everywhere, just:
 - in some areas they are more frequent and stronger, while
 - in other areas they are less frequent and weaker.
- This means, in effect, that faults are everywhere, just:
 - in some areas they are larger and correspond to larger stress, while
 - in other areas, they are weaker and the corresponding stress is smaller.
- In other words, for this phenomenon, physics is in perfect agreement with Zadeh's principle – at least on the qualitative level.

11. Possible application to authorship: formulation of the problem

- In the past, most researchers worked on their own, and most papers had just one author.
- Nowadays, research is often performed by big groups of researchers:
 - some of them make significant contributions to the research results, while
 - contributions of others are not as significant.
- So, a question naturally appears:
 - when the results of this joint research are formulated in a paper,
 - who should be included in the list of the paper's authors?
- This is a subject of many serious discussions.

12. Why is this a difficult problem?

- In our opinion, the problem is difficult because there is no crisp, discrete separation between:
 - authors on the one hand and
 - contributors who end up being thanked (but who are not listed as authors) on the other hand.
- In each group, there is an implicit threshold, so that:
 - participants whose contribution level is above this threshold are listed as authors, while
 - those whose level of contribution is below this threshold are not.
- This threshold level varies between different research communities, between different research groups.
- For example, many experimental papers have dozens of authors, while most theoretical papers usually have much fewer ones.
- And within each group, there is a certain level of subjectivity.

13. Why is this a difficult problem (cont-d)

- Being an author in several papers is critically important:
 - for students to defend their dissertations,
 - important for job search,
 - for promotion.
- As a result, the degree of subjectivity in deciding who is listed as an author often causes conflicts within the research groups.
- These conflicts hinder possible collaboration and thus, in the long run, slow down the progress of science.
- How can we avoid this subjectivity?

14. What we propose

- We propose to take into account that being an author is, in effect, not a crisp notion.
- In many cases, it is a matter of degree, so:
 - instead of listing some collaborators as authors and others as non-authors,
 - why not list everyone who contributed something intellectual to the result as authors
 - but list them with the corresponding degree of authorship?
- To some extent, this is already done in some journals.
- There, for each submitted paper, the authors have to agree on percentages of their contributions.
- But at present, this is only done with respect to participants who have already been declared authors.

15. What we propose (cont-d)

- We propose to extend this idea to all the participants, including those who are usually not included in the authors' list.
- Of course, for this idea to work, we need to take into account this degree of authorship when evaluating:
 - the quality of a student's dissertation work, or
 - the quality of the researcher.
- We believe that this – yet another – example of using the above-mentioned Zadeh's principle will help resolve this issue.

16. How to assign the degree of authorship: main idea

- For this idea to work, we need to have an acceptable way to assign degrees of authorship.
- In some cases, the authoring group includes a leader whose opinion everyone respects.
- In such cases, we can simply ask this trusted leader to provide the degrees of authorship.
- However, the very fact that often conflicts appear around this issue means that in many cases, people's opinions differ.
- In such cases, a natural idea is:
 - to ask different participants of the research group to provide such estimates, and
 - then we need to come up with combined estimates that take into account all the opinions.

17. How to assign the degree of authorship: first approximation

- For every two participants i and j , let us denote the degree assigned to the participant i by the participant j by d_{ij} .
- In the beginning, we do not know a priori who contributed more – and thus, whose opinion is more informed and more valuable.
- In the first approximation, we can therefore simply take the average of all the assigned degrees.

18. How to assign the degree of authorship: first approximation (cont-d)

- In other words, to compute the first approximation to the authorship degree assigned to each participant i , we can do the following:

– first, we compute the average values $a_i^{(1)} = \frac{1}{n-1} \cdot \sum_{j \neq i} d_{ij}$, where

n is the number of contributors;

– then, we normalize these values, to make sure the resulting degrees $d_i^{(1)}$ add up to 1:

$$d_i^{(1)} = \frac{a_i^{(1)}}{\sum_{j=1}^n a_j^{(1)}}.$$

19. A natural example shows that we need to go beyond the first approximation

- Let us consider a simple case when three people worked on a project:
- Professor Einstein ($i = 1$) came up with the great design idea.
- Engineer Edison ($i = 2$) transformed this idea into the actual step-by-step design.
- A skilled worker Mr. Dexterous ($i = 3$) actually design this device – which worked exactly as Professor Einstein expected.
- How should we allocate authorship of the resulting paper?
- Professor Einstein understands that:
 - while his was the main idea,
 - this idea would not have been implemented without the ingenuity of the engineer and the skills of the worker.
- In his opinion, the engineer's task was clearly more creative.

20. A natural example shows that we need to go beyond the first approximation (cont-d)

- So he assigns, to the engineer, the weight $d_{12} = 0.2$ and to the worker the weight $d_{13} = 0.1$.
- This implicitly assumes that his own contribution was 70%.
- Engineer Edison largely agrees with this assessment, so he assigns $d_{21} = 0.7$ and $d_{23} = 0.1$.
- On the other hand, Mr. Dexterous did not communicate with Professor Einstein at all.
- All he saw was a great design given to him by the engineer.
- While the engineer have probably praised Professor Einstein's contribution, Mr. Dexterous attributes this praise to engineers' modesty.

21. A natural example shows that we need to go beyond the first approximation (cont-d)

- As many other people, Mr. Dexterous believes that:
 - academicians are talking a lot and getting too much praise for their mostly impractical (thus, largely useless) ideas, while
 - engineers (and, to some extent, workers) are the ones who contribute to the society's progress.
- So:
 - he assigns, to Professor Einstein, the same small degree as to himself $d_{31} = 0.1$, while
 - he assigns the rest of the degree to the engineer: $d_{32} = 0.8$.
- As a result of taking the average, we get
$$a_1^{(1)} = \frac{0.7 + 0.1}{2} = 0.4, \quad a_2^{(1)} = \frac{0.2 + 0.8}{2} = 0.5, \quad a_3^{(1)} = \frac{0.1 + 0.1}{2} = 0.1.$$
- These averages happen to add up to 1.

22. A natural example shows that we need to go beyond the first approximation (cont-d)

- So after normalization, we get the exact same degrees: $d_1^{(1)} = 0.4$, $d_2^{(1)} = 0.5$, and $d_3^{(1)} = 0.1$.
- So now it looks like the engineer was the major contributor to the project.
- This is not right.

23. How can we get more adequate estimates: idea

- The problem with the above first-approximation estimate is that in this estimate,
 - the opinion of someone whose contribution to the paper was very small was given the same weight as
 - the opinion of the major contributors.
- We should give more weight to the opinions of major contributors and less weight to the opinions of minor contributors.
- A natural idea is to use the degree of authorship as this weight.
- Of course, we do not yet know this degree – the whole purpose of this procedure is to come up with such a degree.
- However, we do know approximate values of these degrees, so let us use them as weight.
- This way, we can get adjusted – hopefully more adequate – degrees.
- Thus, we arrive at the following procedure.

24. How can we get more adequate estimates: algorithm

- The above formulas show how to compute the degrees $d_i^{(1)}$ corresponding to the first approximation.
- Based on these degrees, we can compute the next approximation values $d_i^{(2)}$ as follows.
- First, we compute the weighted averages

$$a_i^{(2)} = \frac{\sum_{j \neq i} d_i^{(1)} \cdot d_{ij}}{\sum_{j \neq i} d_i^{(1)}}.$$

- Then, we normalize these values, to make sure the resulting degrees $d_i^{(2)}$ add up to 1:

$$d_i^{(2)} = \frac{a_i^{(2)}}{\sum_{j=1}^n a_j^{(2)}}.$$

25. Example

- In the above example, we get:

$$a_1^{(2)} = \frac{0.5 \cdot 0.7 + 0.1 \cdot 0.1}{0.5 + 0.1} = \frac{0.36}{0.6} = 0.6,$$

$$a_2^{(2)} = \frac{0.4 \cdot 0.2 + 0.1 \cdot 0.8}{0.4 + 0.1} = \frac{0.16}{0.5} = 0.32, \text{ and}$$

$$a_3^{(2)} = \frac{0.4 \cdot 0.1 + 0.5 \cdot 0.1}{0.4 + 0.5} = \frac{0.09}{0.9} = 0.1.$$

- These degrees add up to 1.02, so normalization leads to

$$d_1^{(2)} \approx 0.59, \quad d_2^{(2)} \approx 0.31, \quad d_3^{(2)} \approx 0.10.$$

- Good news is that:

- we now recognize Professor Einstein as the main author, and
- the corresponding degrees are very close to the values $d_1 = 0.7$, $d_2 = 0.2$, and $d_3 = 0.1$ agreed upon by the two major contributors.

26. Algorithmic comment

- If needed, we can iterate further.
- Once we know the degrees $d_i^{(k)}$ corresponding to the k -th approximation, we can compute the next approximation values $d_i^{(k+1)}$ as follows.
- First, we compute the weighted averages

$$a_i^{(k+1)} = \frac{\sum_{j \neq i} d_i^{(k)} \cdot d_{ij}}{\sum_{j \neq i} d_i^{(k)}}.$$

- Then, we normalize these values, to make sure the resulting degrees $d_i^{(k+1)}$ add up to 1:

$$d_i^{(k+1)} = \frac{a_i^{(k+1)}}{\sum_{j=1}^n a_j^{(k+1)}}, \text{ etc.}$$

27. General comment

- What we propose is similar to the iterative process that leads to PageRank.
- PageRank is a numerical criterion that Google search uses to rank possible answers to queries.
- Crudely speaking, the PageRank algorithm boils down to the following.
- In the first approximation, we view the importance $a_i^{(i)}$ of a webpage by the number of other pages j that link to it ($j \rightarrow i$):

$$a_i^{(1)} = \sum_{j:j \rightarrow i} 1.$$

- In the next approximation:
 - we take into account that the linking pages have, in general, different importance,
 - so we use this importance as a weight.

28. General comment (cont-d)

- We do not yet know the actual importance, so we use approximate importance values obtained on the previous step:

$$a_i^{(2)} = \sum_{j:j \rightarrow i} a_j^{(1)}.$$

- If needed, we can continue this procedure.
- Once we know the k -th approximation, we can compute the next approximation values as

$$a_i^{(k+1)} = \sum_{j:j \rightarrow i} a_j^{(k)}.$$

29. What if we take into account the uncertainty of the degrees?

- It is difficult for people to come up with exact numbers d_{ij} describing contributions of others.
- This is very subjective, and we do not think that it is possible to distinguish between, e.g., 50% and 51%.
- People are much more comfortable providing a range $[\underline{d}_{ij}, \bar{d}_{ij}]$ of possible values, such as $[0.6, 0.7]$.
- In this case, in the first approximation, we come up with interval of possible values of $a_i^{(1)}$:

$$[\underline{a}_i^{(1)}, \bar{a}_i^{(1)}] = \frac{1}{n-1} \cdot \sum_{j \neq i} [\underline{d}_{ij}, \bar{d}_{ij}]$$

30. What if we take into account the uncertainty of the degrees (cont-d)

- Here, by an operation \oplus between intervals, we mean the range of the values $a \oplus b$ when a and b lie in the corresponding intervals:

$$[\underline{a}, \bar{a}] \oplus [\underline{b}, \bar{b}] \stackrel{\text{def}}{=} \{a \oplus b : a \in [\underline{a}, \bar{a}], b \in [\underline{b}, \bar{b}]\}.$$

- Once we have an approximation $[\underline{a}_i^{(1)}, \bar{a}_i^{(1)}]$, we need to compute the intervals of possible values of the normalized degrees.
- Each normalized degree is a fraction of two expressions which are linear in $a_i^{(1)}$.
- There exists an efficient algorithm for computing this range.
- This algorithm is, in effect, what is used when we extend centroid defuzzification to the interval-valued fuzzy case.

31. What if we take into account the uncertainty of the degrees (cont-d)

- Once we have the interval-valued degrees $d_i^{(1)}$, we can take into account that the weighted average is monotonic in d_{ij} ; thus:
 - to find the largest possible value of $a_i^{(2)}$, it is sufficient to consider the upper bound \bar{d}_{ij} , while
 - to find the smallest possible value of $a_i^{(2)}$, it is sufficient to consider the lower bound \underline{d}_{ij} .
- Once we fix the values d_{ij} this way, the formula for the normalized degrees also becomes fractionally linear.
- So we can use the same algorithm to compute the interval of possible values of $a_i^{(2)}$, etc.

32. Acknowledgments

This work was supported in part by:

- National Science Foundation grants 1623190, HRD-1834620, HRD-2034030, and EAR-2225395;
- AT&T Fellowship in Information Technology;
- program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and
- a grant from the Hungarian National Research, Development and Innovation Office (NRDI).