4-1-2023

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

Christian Servin
*El Paso Community College*, cservin1@epcc.edu

Aaron Velasco
*The University of Texas at El Paso*, aavelasco@utep.edu

Edgar Daniel Rodriguez Velasquez
*The University of Texas at El Paso*, edrodriguezvelasquez@miners.utep.edu

Vladik Kreinovich
*The University of Texas at El Paso*, vladik@utep.edu



Comments:

Technical Report: UTEP-CS-23-06a

To appear in *Proceedings of the 2023 Annual Conference of the North American Fuzzy Information Processing Society NAFIPS'2023*, Cincinnati, Ohio, May 31 - June 2, 2023.

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

Christian Servin, Aaron Velasco, Edgar Daniel Rodriguez Velasquez, and Vladik Kreinovich

**Abstract** This paper presents two applications of the general principle – the everything is a matter of degree – the principle that underlies fuzzy techniques. The first – qualitative – application helps explain the fact that while most earthquakes occur close to faults (borders between tectonic plates or terranes), earthquakes have also been observed in areas which are far away from the known faults. The second – more quantitative – application is to the problem of which of the collaborators should be listed as authors and which should be simply thanked in the paper. We argue that the best answer to this question is to explicitly state the degree of authorship – in contrast to the usual yes-no approach. We also show how to take into account that this degree can be estimated only with some uncertainty – i.e., that we need to deal with interval-valued degrees.

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

Aaaron Velasco
Department of Earth, Environmental, and Research Sciences, University of Texas at El Paso
500 W. University, El Paso, TX 79968, USA, e-mail: aavelasco@utep.edu

Edgar Daniel Rodriguez Velasquez
Department of Civil Engineering, Universidad de Piura in Peru (UDEP), Av. Ramón Mugica 131
Piura, Peru, e-mail: edgar.rodriguez@udep.pe
and
Department of Civil Engineering, University of Texas at El Paso, 500 W. University
El Paso, TX 79968, USA, e-mail: edrodriguezvelasquez@miners.utep.edu

Vladik Kreinovich
Department of Computer Science, University of Texas at El Paso, 500 W. University
El Paso, Texas 79968, USA, e-mail: 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; see, e.g., [1, 4, 9, 12, 13, 16].

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 Possible Application to Geosciences

**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 the 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.

**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 – or, 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.

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.

**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; see, e.g., [2]. Most earthquakes do occur at the faults, but 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.

**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, but 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. 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 (see, e.g., [1, 4, 9, 12, 13, 16]) – that everything is a matter or degree.

**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 (see, e.g., [7, 15]):

- 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, with even smaller stress, etc.

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: 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.

## 3 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, see, e.g., [14].

**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: e.g., 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.

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 (and who is not) often causes conflicts within the research groups – and these conflicts hinder possible collaboration and thus, in the long run, slow down the progress of science. How can we avoid this subjectivity?

**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 with the corresponding degree of authorship?

To some extent, this is already done in some journals – where 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. 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.

**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.

**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. 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

$$a_i^{(1)} = \frac{1}{n-1} \cdot \sum_{j \neq i} d_{ij}, \tag{1}$$

  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:

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

**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, and
- 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, so he assigns, to the engineer, the weight $d_{12} = 0.2$ and to the worker the weight $d_{13} = 0.1$ – thus implicitly assuming 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. 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

$$a_1^{(1)} = \frac{0.7+0.1}{2} = 0.4,\ \ a_2^{(1)} = \frac{0.2+0.8}{2} = 0.5,\ \ a_3^{(1)} = \frac{0.1+0.1}{2} = 0.1.$$

These averages happen to add up to 1, 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.

**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.

**How can we get more adequate estimates: algorithm.** Formulas (1) and (2) 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

$$a_i^{(2)} = \frac{\sum\limits_{j\neq i} d_i^{(1)} \cdot d_{ij}}{\sum\limits_{j\neq i} d_i^{(1)}}, \tag{3}$$

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

$$d_i^{(2)} = \frac{a_i^{(2)}}{\sum\limits_{j=1}^{n} a_j^{(2)}}. \tag{4}$$

**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

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

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.

*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

$$a_i^{(k+1)} = \frac{\sum\limits_{j \neq i} d_i^{(k)} \cdot d_{ij}}{\sum\limits_{j \neq i} d_i^{(k)}}, \tag{5}$$

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

$$d_i^{(k+1)} = \frac{a_i^{(k+1)}}{\sum\limits_{j=1}^{n} a_j^{(k+1)}}, \text{ etc.} \tag{6}$$

*General comment.* What we propose is similar to the iterative process that leads to PageRank – a numerical criterion that Google search uses to rank possible answers to queries; see, e.g., [6]. 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 ($j \to i$):

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

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

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

If needed, we can continue this procedure: once we know the $k$-th approximation, we can compute the next approximation values as

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

**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 $[\underline{d}_{ij},\overline{d}_{ij}]$ 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 $a_i^{(1)}$:

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

where by sum – or any other operation $\oplus$ – between intervals we mean the range of the values $a\oplus b$ when $a$ and $b$ lie in the corresponding intervals $[\underline{a},\overline{a}]$ and $[\underline{b},\overline{b}]$ (see, e.g., [3, 5, 8, 10]):

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

Once we have an approximation $[\underline{a}_i^{(1)},\overline{a}_i^{(1)}]$, we need to compute the intervals of possible values of the normalized degrees. Each normalized degree (2) is a fraction of two expressions which are linear in $a_i^{(1)}$. There exists an efficient algorithm for computing this range – see, e.g., [11]. This algorithm is, in effect, what is used when we extend centroid defuzzification to the interval-valued fuzzy case; see, e.g., [9].

Once we have the interval-valued degrees $d_i^{(1)}$, we can take into account that the expression (3) is monotonic in $d_{ij}$. Thus:

- to find the largest possible value of $a_i^{(2)}$, it is sufficient to consider the upper bound $\overline{d}_{ij}$, 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 (3) also becomes fractionally linear, so we can use the same algorithm to compute the interval of possible values of $a_i^{(2)}$, etc.

## Acknowledgments

This work was supported in part by the National Science Foundation grants 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in

Computer Science), HRD-1834620 and HRD-2034030 (CAHSI Includes), and by the AT&T Fellowship in Information Technology.

It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478, and by a grant from the Hungarian National Research, Development and Innovation Office (NRDI).

The authors are greatly thankful to the anonymous referees for valuable suggestions.

## References

1. R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
2. W. Fan, A. J. Barbour, J. J. McGuire, Y. Huang, G. Lin, E. S. Cochran, and R. Okuwaki, "Very low frequency earthquakes in between the seismogenic and tremor zones in Cascadia?", *AGU Advances*, 2022, Vol. 3, e2021AV000607.
3. L. Jaulin, M. Kiefer, O. Didrit, and E. Walter, *Applied Interval Analysis, with Examples in Parameter and State Estimation, Robust Control, and Robotics*, Springer, London, 2001.
4. G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
5. B. J. Kubica, *Interval Methods for Solving Nonlinear Contraint Satisfaction, Optimization, and Similar Problems: from Inequalities Systems to Game Solutions*, Springer, Cham, Switzerland, 2019.
6. A. N. Langville and C. D. Meyer, *Google's PageRank and Beyond: The Science of Search Engine Rankings*, Princeton University Press, Princeton, New Jersey, and Oxford, UK, 2012.
7. B. B. Mandelbrot, *Fractals: Form, Chance and Dimension*, Freeman Co., New York, 2020.
8. G. Mayer, *Interval Analysis and Automatic Result Verification*, de Gruyter, Berlin, 2017.
9. J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
10. R. E. Moore, R. B. Kearfott, and M. J. Cloud, *Introduction to Interval Analysis*, SIAM, Philadelphia, 2009.
11. H. T. Nguyen, V. Kreinovich, B. Wu, and G. Xiang, *Computing Statistics under Interval and Fuzzy Uncertainty*, Springer Verlag, Berlin, Heidelberg, 2012.
12. H. T. Nguyen, C. L. Walker, and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2019.
13. V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
14. M. A. Parsons, D. S. Katz, M. Langseth, H. Ramapriyan, and S. Ramdeen, "Credit where credit is due", *EOS: Science News from the American Geophysical Union*, 2022, Vol. 103, No. 11, pp. 20–23.
15. L. E. Vallejo, "Fractal analysis of the cracking and failure of asphalt pavements", In: *Proceedings of the 2016 Geotechnical and Structural Engineering Congress*, Phoenix, Arizona, February 14–17, 2016.
16. L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.