Fuzzy Ideas Explain a Complex Heuristic Algorithm for Gauging Pavement Conditions

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

Carlos M. Chang Albitres
Department of Civil Engineering
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
cchangalbitres2@utep.edu

Olga Kosheleva[1]
Vladik Kreinovch[2]
[1]Department of Teacher Education
[2]Department of Computer Science
University of Texas at El Paso
500 W. University
El Paso, TX 79968, USA
olgak@utep.edu, vladik@utep.edu

***Abstract*—To gauge pavement conditions, researchers have come up with a complex heuristic algorithm that combines several expert estimates of pavement characteristics into a single index – which is well correlated with the pavement's durability and other physical characteristics. While empirically, this algorithm works well, it lacks physical or mathematical justification beyond being a good fit for the available data. This lack of justification decreases our confidence in the algorithm's results – since it is known that often, empirically successful heuristic algorithms need change when the conditions change. To increase the practitioners' confidence in the resulting pavement condition estimates, it is therefore desirable to come up with a theoretical justification for this algorithm. In this paper, we show that by using fuzzy techniques, it is possible to come up with the desired justification.**

I. Formulation of the Problem

It is important to gauge pavement conditions. Most roads are heavily used. Heavy traffic stresses the pavement. As a result, after several years, it is necessary to maintain – or sometimes even repair – the roads.

Roads repairs are expensive. It is therefore important to adequately gauge pavement conditions – so that we will be able to correctly decide which road segments need maintenance or repair, and which can wait a few more years.

This is especially important since it is known that a proper maintenance can make the road last much longer and thus, drastically decrease the need for expensive road repairs.

How pavement conditions are gauged now. One of the most frequently used technique for gauging pavement conditions is based on visual inspection of the pavement.

Visual inspection enables the inspectors to detect different types of problems – known as *distresses*. We can have buckling, we can have potholes, we can have cracks, etc. For each type of distress, inspectors:

- measure the area affected by this type of distress (or the length, for linear distresses like linear cracks), and
- use the results of these measurement to evaluate the severity of the corresponding distress.

The resulting data is then combined into a single *pavement condition index* (PCI).

The combination rules used in the computation of the PCI are selected so as to provide the most accurate prediction of the pavement durability. To improve the predictive quality, more and more complex algorithms are used; see, e.g., the latest international standard [1].

Problem. The problem is that the existing algorithm for gauging the pavement condition is heuristic. This algorithm has been selected purely empirically, it does not have any physical or mathematical justification – beyond being a good fit for the available data.

In general, heuristic methods often work well, but they are usually less reliable than theoretically justified algorithms – since they rely solely on the past experiences and when the situations change, we may need to change the algorithms as well. To increase the user's confidence in the PCI algorithm, it is thus desirable to come up with a theoretical justification for this algorithm.

What we do in this paper. In this paper, we provide the desired theoretical justification for the current state-of-the-art complex heuristic algorithm for gauging pavement conditions.

In this justification, we take into account the fact that this algorithm combines – somewhat subjective – inspector observations, observations which include information described not in numerical terms, but rather in terms of imprecise ("fuzzy') words from natural language, such as "high", "low", and "medium". Thus, to analyze this problem, it is reasonable to use techniques specifically designed for translating such knowledge into precise numbers – namely, *fuzzy* techniques; see, e.g., [2], [3], [4], [6], [7], [9].

These techniques are what we will use in our justification.

II. The Current State-of-the-Art Algorithm for Gauging Pavement Conditions: A Brief Reminder

What we start with. For each road segment, this algorithm starts with the numbers

$$x_1, \ldots, x_n$$

that describe the relative areas (or relative lengths) of the distresses within this segment.

First step: a non-linear transformation. First, an appropriate non-linear transformation $f_i(x_i)$ is applied to each value x_i, resulting in so-called *deduct values*

$$s_i = f_i(x_i)$$

ranging from 0 to 100 (or, equivalently, from 0 to 1). These non-linear transformations

$$f_i(x_i)$$

are selected so that the resulting PCI have the largest correlation with the pavement's durability.

The deduct values are selected in such as way that larger values of the scores correspond to more severe distresses:

- the value 100 (or 1) corresponds to the most sever distress, while
- the value 0 corresponds to the absence of distress.

Second step: sorting the deduct values. The deduct values corresponding to distresses of different types are then sorted in the decreasing order, from the most severe to the least severe:

$$s_{(1)} \ge s_{(2)} \ge \ldots$$

Third step: deciding how many deduct values to use. Based on the largest deduct value, we then decide how many deduct values to use. This number m of used deduct values is found from a formula

$$m = 1 + \frac{9}{98} \cdot (100 - s_{(1)}). \tag{1}$$

We then use only the values

$$s_{(1)} \ge s_{(2)} \ge \ldots \ge s_{(m)}.$$

Final step: combining deduct values. To combine the values $s_{(1)}$, ..., $s_{(m)}$, we do the following:

- first, we compute the sum of the largest deduce value $s_{(1)}$ and of $m-1$ small values (equal to 2); we apply an appropriate non-linear transformation to transform this sum into the interval $[0, 100]$; thus, we get the first combined deduct value c_1;
- then, we compute the sum of the two largest deduct values and of $m-1$ 2s – and apply a different non-linear transformation to the resulting sum; thus, we get the second combined deduct value c_2;
- after that, we compute the sum of 3 largest deduct values and $m-2$ 2s, and apply a yet different non-linear transformation to the resulting sum; thus, we get the third combined deduct value c_3;
- then we repeat the same procedure for 4 largest deduct values, for 5 largest deduct values, etc., until we are combine all m deduct values.

As a result, we get m combined deduct values

$$c_1, c_2, , \ldots, c_m.$$

After that, we take the largest of the resulting combined distress values

$$c \stackrel{\text{def}}{=} \max_i c_i.$$

The PCI is simply 100 minus this largest value:

$$\text{PCI} \stackrel{\text{def}}{=} 100 - c.$$

The resulting combination of somewhat subjective estimates is indeed well-correlated with physical properties. The algorithm has been selected so as to provide the largest correlation with the pavement durability and other physical characteristic. For example, it has been shown that PCI is strongly correlated with the International Roughness Index that measures the passing vehicle's vibrations caused by the pavement's imperfection; see, e.g., [8].

Towards reformulating the final step. Our ultimate goal is to decide when a road segment needs maintenance or repair. This decision is made by comparing the PCI estimated for this segment with a certain threshold t_0. The condition that

$$100 - c \ge t_0$$

is equivalent to

$$c \le 100 - t_0.$$

In its turn, the condition that

$$c = \max_i c_i \le 100 - t_0$$

is equivalent to requiring that

$$c_i \le 100 - t_0$$

for each i.

Each value c_i is obtained from the sum

$$s_{(1)} + \ldots + s_{(i)}:$$

- by adding $(m-i)$ values of 2 and
- by applying an appropriate non-linear transformation to the resulting sum.

Thus, the condition

$$c_i \le 100 - t_0$$

is equivalent to requiring that the sum

$$s_{(1)} + \ldots + s_{(i)}$$

is greater than or equal to some threshold t_i. Thus, we can reformulate the final step as follows.

Reformulation of the final step. To decide whether the given road segment needs repairs or maintenance, we check, for each i from 1 to m, whether

$$s_{(1)} + s_{(2)} + \ldots + s_{(i)} \geq t_i$$

for the corresponding threshold t_i.

What needs explanation. Natural questions are:

- Why should we use sum and not any other combination function?
- Why should we consider the sum of a few largest distress values and not of all these values?
- Why should we consider several sums instead of just one?
- Where does the formula for the number m of considered deduct values come from?

There can be many other questions, since the above procedure, with its emphasis on sorting and maxima, does not look like any physical formula – physics formulas very rarely use maxima.

III. Why Should We Use Sum and Not Any Other Combination Function: An Explanation

Let us start analyzing the problem. The road segment is good if there are not too many distresses of each type, i.e., if there are:

- few distresses of the first type *and*
- few distresses of the second type, etc.

In other words, the pavement is good if:

- the first value x_1 is small *and*
- the second value x_2 is small, etc.

This looks like a typical phrase to be analyzed by fuzzy techniques. Namely, phrase is an "and"-combination of simpler phrases like "the value x_1 is small", "the value x_2 is small", etc. To assign a numerical value to the validity of this phrase, it makes sense:

- first, to estimate the degree to which each simple statement "x_i is small" is true, and then
- combine these degree of confidence into a single degree.

This is exactly what we will do.

We need different membership functions for different i. In accordance with the usual fuzzy techniques, for each i and for each x_i, we need to come with a number d_i describing to what extend the given value x_i is small. Let us denote this number by $\mu_i(x_i)$. In fuzzy techniques, the corresponding function $\mu_i(x_i)$ is known as the *membership function* corresponding to the notion "small".

In the traditional application of fuzzy techniques, when we have several occurrences of the same word like "small", we use the same membership function. However, most fuzzy textbooks emphasize that this is not necessarily the case: for example, then transforming the size in meters into a number, "small" means two different things when referring to cats or to people – a cat the size of a small human being is, by cats' standards, a giant.

This is exactly the case here. For example, if x_1 describes the relative area of severe distress, then x_1 should really be small for this distress to be acceptable and not requiring any maintenance. However, for low severity distress x_2, even if this distress takes a significant part of the road segment, by itself, this may not necessarily trigger any need for maintenance. Thus, in our case, we need different membership functions $\mu_i(x_i)$ for different i.

How to combine the degrees. In general, the problem of combining the degrees is as follows:

- we know the degrees a and b to which statements A and B are true, and
- we want to use these values a and b to estimate the degree to which a composite statement $A \,\&\, B$ is true.

In fuzzy logic, the corresponding estimate is called an *"and"-operation* (or, for historical reasons, a *t-norm*); let us denote it by $f_{\&}(a, b)$.

In these terns, the desired degree of confidence that the road segment is good is equal to

$$f_{\&}(\mu_1(x_1), \mu_2(x_2), \ldots). \tag{2}$$

Natural conditions on an "and"-operation. The "and"-operation should satisfy several conditions. First, since $A \,\&\, B$ and $B \,\&\, A$ mean the same, it is reasonable to expect that the corresponding estimates for their degrees should be the same, i.e., that we should have

$$f_{\&}(a, b) = f_{\&}(b, a)$$

for all a and nb. In other words, the "nd"-operation should be *commutative*.

Similarly, since $A \,\&\, (A \,\&\, C)$ and $(A \,\&\, B) \,\&\, C$ means the same, we expect that the estimates of the degree of these two statement should be the same, i.e., that for all a, b, and c, we should have

$$f_{\&}(a, f_{\&}(b, c)) = f_{\&}(f_{\&}(a, b), c).$$

In other words, an "and"-operation should be *associative*.

There are several other reasonable properties; see, e.g., [2], [3], [4], [6], [7], [9]. An "and"-operation that satisfies all these properties is usually what is called a t-norm.

Structure of a generic t-norm. Some t-norms have the form

$$f_{\&}(a, b) = g^{-1}(g(a) + g(b)) \tag{3}$$

for some increasing function $g(z)$, where $g^{-1}(z)$ indicates an inverse function, for which $g^{-1}(g(z)) = z$. Such t-norms are know as *Archimedean*.

For example, for the probability-inspired operation $f_{\&}(a, b) = a \cdot b$, we get this form with $g(z) = \ln(z)$.

It is known (see, e.g., [5]) that for every t-norm $f_{\&}(a, b)$ and for every $\varepsilon > 0$, there exists an Archimedean t-norm $f'_{\&}(a, b)$ which is ε-close to $f_{\&}(a, b)$, i.e., for which

$$|f'_{\&}(a, b) - f_{\&}(a, b)| \leq \varepsilon$$

for all a and b. Since the expert's degrees of confidence are always approximate, and ε can be arbitrary small, in practice, we can safely replace the original t-norm with an ε-close Archimedean one – as long as ε is small enough. Thus, without losing generality, we can safely assume that the t-norm $f_{\&}(a,b)$ is Archimedean.

This explains why in gauging pavement conditions, we use sum. Indeed, the degree of confidence that the road segment is good is determined by the formula (2). As we have discussed, we can safely assume that the corresponding t-norm is Archimedean, i.e., that it is described by the formula (3).

Substituting the expression (3) into the formula (2), we conclude that the desired degree d has the form

$$d = g^{-1}(g(\mu_1(x)) + g(\mu_2(x_2)) + \ldots),$$

i.e., equivalently, the form

$$d = g^{-1}(s), \text{ where } s = s_1 + s_2 + \ldots,$$

$$s_i = f_i(x_i), \text{ and } f_i(z) \stackrel{\text{def}}{=} g(f_i(z)).$$

In particular, since the function $g(z)$ is increasing, the condition that road is good enough, i.e., that

$$d \geq d_0$$

for some threshold d_0, can be equivalently reformulated as

$$s \geq t_0 \stackrel{\text{def}}{=} g(d_0).$$

In other words, we get

$$s_1 + s_2 + \ldots \geq t_0.$$

This is exactly the sum-based formula used to estimate the desired degree – which is thus explained by fuzzy ideas.

IV. Why Should We Consider the Sum of a Few Largest Distress Values And Not of All These Values?

Analysis of the problem: analyzed road segments are reasonably good. The whole procedure makes sense when roads are reasonably well maintained and are in reasonable condition. If the road is in a clearly bad condition, there is no need to accurately gauge its quality, we just need to repair it.

The need for an accurate estimate of the road's quality occurs when we have several segments of reasonably good quality, and we need to find the way to maintain them and making them even better.

In such situations, most distress values x_i are small. When a distress value is very small, it does not affect the overall quality of a road segment.

Computational consequences of this analysis. Since small distress values do not affect the quality of a road segment, taking them into account would be a waste of computational resources.

To avoid this waste, it makes sense to ignore these very small values, and consider only a few largest distress values. This is exactly whet is usually done: instead of taking the sum of all the values $s_1 + s_2 + \ldots$, we only consider the sum of the m largest values

$$s_{(1)} + s_{(2)} + \ldots + s_{(m)}.$$

This is exactly what practitioners do.

V. Why Should We Consider Several Sums Instead of Just One?

General idea. If, based on the largest distress, we know that the road segment need repair or maintenance, there is no need to consider all other distresses. In this case, taking other distresses into account would be a waste of computational resources.

If, based on the first distress, we cannot make a definite conclusion, it is reasonable to also consider the second distress, etc.

Thus, instead of always taking all m distresses into account, it makes sense to first check just the largest distress, then two largest, then three largest. etc.

This is exactly what is done in practice.

This is a fuzzy analog of lazy logical operations. In classical 2-valued logic, if we want to find the truth value of a statement $A \,\&\, B$ and we know that A is false, there is no need to find the truth value of B – we can already conclude that the composite statement $A \,\&\, B$ is also false.

This simple observation saves us computation time. The corresponding operation is known as a *lazy* "and". This is the most commonly used "and"-operation in programming languages such as C or Java.

What we are describing here is the fuzzy analogue of such lazy "and"-operations. Indeed, when the first values $s_{(1)}$, $s_{(2)}$, ... are already large – corresponding to close-to-false(0) values of the corresponding degrees $\mu_i(x_i)$ – then there is no need to compute any further terms, we know that the road segment needs repair or maintenance.

VI. Where Does the Formula for the Number m of Considered Deduct Values Come From?

Analysis of the problem. Suppose that we know the largest distress $s_{(1)}$. Let us denote, by S_0, the overall distress level after which the road segment needs repairs or maintenance.

Let us denote, by s_0, the smallest value of an individual distress that is still worth taking into account, so that values smaller than s_0 can be safely set to 0. Then, if, in addition to the largest distress, we take into account $m-1$ other non-zero distresses, we get the overall value

$$s_{(1)} + (m-1) \cdot s_0.$$

If this value is already larger than or equal to the threshold S_0, this means that there is no need to consider any additional distresses – we already know that the road segment needs repairs or maintenance.

On the other hand, if among the m largest distresses, the smallest is already below s_0 – and can hence be safely ignored – this means that all smaller distresses can also be ignored.

So, considering more than m distresses also does not make sense.

Thus, in all possible cases, the largest number of distresses to be continued is the smallest m for which

$$s_{(1)} + s_0 \cdot (m-1) \geq S_0.$$

In terms of m, this inequality can be reformulated in the equivalent form

$$m \geq 1 + \frac{1}{s_0} \cdot (S_0 - s_{(1)}).$$

So, the smallest possible value m that satisfies this property has the form

$$m = 1 + \frac{1}{s_0} \cdot (S_0 - s_{(1)}). \tag{4}$$

This analysis explains the formula for the number m of considered deduct values. Indeed, (4) is exactly the formula used to estimate how many deduct values we need to take into account.

ACKNOWLEDGMENTS

This work was supported in part by the National Science Foundation grant HRD-1242122 (Cyber-ShARE Center of Excellence).

REFERENCES

[1] ASTM International, *Standard Practice for Roads and Parking Lots Pavement Condition Index Surveys*, International Standard D6433-18.
[2] R. Belohlavek, J. W. Dauben, and G. J. Klir, *Fuzzy Logic and Mathematics: A Historical Perspective*, Oxford University Press, New York, 2017.
[3] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.
[4] J. M. Mendel, *Uncertain Rule-Based Fuzzy Systems: Introduction and New Directions*, Springer, Cham, Switzerland, 2017.
[5] H. T. Nguyen, V. Kreinovich, and P. Wojciechowski, "Strict Archimedean t-norms and t-conorms as universal approximators", *International Journal of Approximate Reasoning*, 1998, Vol. 18, Nos. 3–4, pp. 239–249.
[6] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.
[7] V. Novák, I. Perfilieva, and J. Močkoř, *Mathematical Principles of Fuzzy Logic*, Kluwer, Boston, Dordrecht, 1999.
[8] K. Park, N. E. Thomas, and K. W. Lee, "Applicability of the International Roughness Index as a predictor of asphalt pavement condition", *Journal of Transportation Engineering*, 2007, Vol. 133, No. 12, pp. 706–709.
[9] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.