Departmental Technical Reports (CS)   Department of Computer Science

5-1-2000

Granularity as an Optimal Approach to Uncertainty - A General Mathematical Idea with Applications to Sleep, Consumption, Traffic Control, Learning, etc.

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

Hung T. Nguyen

Recommended Citation

Granularity as an Optimal Approach to Uncertainty – A General Mathematical Idea With Applications to Sleep, Consumption, Traffic Control, Learning, etc.

Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
email vladik@cs.utep.edu

Hung T. Nguyen
Department of Mathematical Sciences
New Mexico State University
Las Cruces, NM 88003, USA
email hunguyen@nmsu.edu

Abstract

Traditional statistical and fuzzy approaches to describing uncertainty are continuous *in the sense that we use a (potentially* infinite*) set of values from the interval* $[0, 1]$ *to characterize possible degrees of uncertainty. In reality, experts describe their degree of belief by using one of the* finitely *many words from natural language; in this sense, the actual description of expert uncertainty is* granular.

In this paper, we show that in some reasonable sense, granularity is the optimal way of describing uncertainty. A similar mathematical idea explains similar "granularity" in such diverse areas as sleep, consumption, traffic control, and learning.

1. Introduction

In statistical approach to uncertainty, all values from [0,1] are needed. In the traditional *statistical* approach to uncertainty, the uncertainty of an event E is described by its probability, which is a number from the interval $[0, 1]$. In many physical situations, this probability gradually (and continuously) changes from 0 to 1.

As a result, due to the known property of continuous functions, for each number from the interval $[0, 1]$, there is a moment of time when this particular number represents a probability of a real-life event. Thus, all numbers from the interval $[0, 1]$ are needed to describe probabilistic uncertainty.

At first glance, it may seem that in fuzzy approach to uncertainty, also all values from [0,1] are needed. In *fuzzy* approach, uncertainty is also described by a number from the interval $[0, 1]$ which describes the degree to which the expert believes that a certain property holds.

For many reasonable properties (like "small"), as the value of the corresponding physical quantity increase, the corresponding degree continuously changes from 1 to 0 (or from 0 to 1). Therefore, similarly to the probabilistic case, it looks like we need all (infinitely many) numbers from the interval $[0, 1]$ to describe fuzzy uncertainty.

In reality, experts' description of their uncertainty is granular. There is a problem with this conclusion. One of the main goals of fuzzy logic is to formalize *expert knowledge* (and its uncertainty). Real-life experts, however, do not use infinitely many different values to describe their degree of certainty. They normally use a small finite number of alternatives: namely, one of the words describing uncertainty.

Instead of infinitely many possible values from the interval $[0, 1]$, we only get finitely many words; each word therefore corresponds to a whole set (granule) of possible values. In other words, the actual description of uncertainty is granular.

From the traditional viewpoint, granularity is a crude approximation. From the viewpoint of traditional probabilistic or fuzzy approach, this granular description is a crude approximate description of the continuous uncertainty. One can expect such an approximate behavior from a simple crude system which does not have enough memory or computing power to process too many possible degrees of certainty.

But is it? However, it is unclear why a human brain, the result of billions of years of evolution from simple one-cell organisms to sophisticated thinking abilities, would use such a low quality crude approximation scheme.

So maybe granularity is not a crude approximation scheme? Maybe granularity is, vice versa, a high quality (or even optimal) scheme for describing uncertainty?

What we are planning to do. In this paper, we show that granularity is, indeed, a high quality (and even optimal) scheme for describing uncertainty.

In our derivation, we will use the standard techniques of continuous mathematics, in line with our general results from [3] showing that continuous mathematics is a very helpful tool in justifying different techniques for handling uncertainty.

2. Towards the Formalization of the Problem

Preliminary idea: experts strife to increase the certainty. How can we formalize this problem? We want to describe the uncertainty of human knowledge. This knowledge consists of several different statements $S_1, \ldots, S_n$. We are considering the situation in which the uncertainty of each statement is characterized by a number from the interval $[0, 1]$; in other words, the uncertainty of human knowledge is described by assigning n numbers $d_1, \ldots, d_n$ to the statements which form a knowledge base.

Experts do not simply keep the knowledge about their area of expertise, they also strife to increase the amount of knowledge, either by extracting new pieces of knowledge from the experimental data, or by using logical arguments to extract new knowledge and new statements from the already known ones.

Since the memorization abilities of each individual expert are limited, an expert invariable forgets some of the knowledge that he previously knew. This "forgetting" is important because it allows us to clear the memory for new ideas and new results. However, while individual experts may forget part of the knowledge, from the viewpoint of the whole community of experts, the total amount of knowledge (normally) increases.

The final goal of experts as a group is to attain full knowledge of a certain domain, when we would be able, given each statement, to decide whether this statement is true or not. From this viewpoint, the ideal degree of certainty for each statement is either 0 or 1.

If the degree of certainty is 0, this means that for the negation of this statement, the degree of certainty is equal to 0. Thus, without losing generality, we can say that an expert strives to make his degrees of certainty d_i as close to 1 as possible, i.e., as large as possible.

Ramification of the preliminary idea: experts strife to increase the certainty within limited resources. Each increase in the degrees of certainty (i.e., each decrease in uncertainty) requires a certain effort, often, a very substantial effort. The expert's resources are usually limited. So, an optimal behavior for an expert would be: *Within the given total effort, to maximize the resulting certainty.* Let us formalize this requirement.

First step towards formalization: describing effort. We are trying to formalize the fact that experts have a limited number of resources and that, therefore, they can only use a limited amount of effort.

The effort $e(E, d)$ which is necessary to achieve a given level of certainty d for a given statement E depends both on the level d and on the statement E:

- the larger required degree of certainty, the more efforts are needed, so $e(E, d)$ is an increasing function of the degree d;

- also, for some statements, their checking requires much more time and effort than for the others, so the necessary amount of effort $e(E, d)$ depends not only on the degree d, but also on the statement E.

This function describes what effort is necessary for a *single* statement. If *two* statements E_1 and E_2 are completely unrelated, then clarifying one of them does not in any way help us to clarify the second one. Thus, for unrelated statements, if we want to achieve the degree of certainty d_1 for the first statement and the degree of certainty d_2 for the second statement, then the required effort can be simply computed as a sum of the efforts corresponding to these individual statement, i.e., as $e(E_1, d_1) + e(E_2, d_2)$.

In real life, statements are rarely unrelated.

- Sometimes, the statements are logically related; e.g., E_1 implies E_2. In this case, if we increase the degree of certainty d_1 in the statement E_1, we automatically increase the degree of certainty in the statement E_2. Therefore, the effort used to increase the degree of certainty in E_1 helps in increasing the degree of certainty in E_2. As a result, the amount of efforts required to achieve *both* the degrees d_1 and d_2 is (when $d_1 \approx d_2$) practically equal to the effort $e(E_1, d_1)$ necessary to achieve the degree d_1 and thus, much smaller than the sum $e(E_1, d_1) + e(E_2, d_2)$.

- In some other cases, there is a different ("competition") relation between the statements E_1 and E_2: attempts to confirm the statement E_1 compete for the same resources as attempts to confirm the statement E_2. For example, if we are interested in the effect of a certain medicine on a rare disease, then we can test this medicine on patients; we can use the same (small) population of patients to test the second medicine. However, if we want to test both medicines, we cannot simply combine these two testings because we do not have enough patients for that. Therefore, if we want to test both medicines, we need to use a lot of additional efforts, e.g., test the medicines on animals, computer models, etc. In

this case, the total effort required to achieve both degrees of certainty is much larger than simply the sum of the efforts $e(E_1, d_1) + e(E_2, d_2)$.

In some cases, the total effort is smaller than the sum; in some other cases, the total effort is much larger than the sum. It is reasonable to assume that for a large number of statements which may be related in different ways, on average, these positive and negative deviations from the sum more or less compensate each other. In the resulting approximation, the total effort of achieving certainty degree d_1 for statement E_1, certainty degree d_2 on statement E_2, ..., and certainty degree d_n on a statement E_n can be described as the sum of the corresponding individual efforts, i.e., as $e(E_1, d_1) + \ldots + e(E_n, d_n)$.

We are interested in large bodies of knowledge, which contain many different statements E_i.

- Some of these statements are easier to analyze and require fewer effort to attain the given degree of certainty d.

- Some of these statements are more difficult to analyze and require more effort to attain the given degree of certainty d.

For a large body of knowledge, which contains both easy-to-analyze and difficult-to-analyze statement, it is reasonable, as a first approximation to the sum $e(E_1, d_1) + \ldots + e(E_n, d_n)$, to replace individual dependencies $d \rightarrow e(E, d)$ with a single "average" dependence $e(d)$ – i.e., by an effort $e(d)$ required to achieve a level of certainty d for an "average" statement. In other words, as a first approximation to the desired sum, we take the sum $e(d_1) + \ldots + e(d_n)$.

When we replace $e(E_i, d_i)$ by $e(d_i)$, then:

- for easy-to-analyze statements, for which the actual effort $e(E_i, d)$ is smaller than average, this replacement overestimates the effort;

- for difficult-to-analyze statements, for which the actual effort $e(E_i, d)$ is larger than average, this replacement overestimates the effort.

It is reasonable to expect that on average, these positive and negative corrections to the sum formula more or less compensate each other. Therefore, as a reasonable first approximation to the total effect needed to achieve the levels of certainty $d_1, \ldots, d_n$, we can take the sum $e(d_1) + \ldots + e(d_n) = E$.

Second step towards formalization: describing satisfaction. Similarly, the amount of satisfaction $s(E, d)$ resulting from achieving the degree of certainty d for a statement E, depends both on the statement E and on the degree d.

This function describes what satisfaction we get from a *single* statement. If *two* statements E_1 and E_2 are completely unrelated, then the resulting satisfaction can be simply computed as a sum of the satisfaction levels corresponding to these individual statement, i.e., as $s(E_1, d_1) + s(E_2, d_2)$.

In real life, as we have mentioned, statements are rarely unrelated.

- Sometimes, the statements are logically related; e.g., E_1 implies E_2. In this case, if we increase the degree of certainty d_1 in the statement E_1, we automatically increase the degree of certainty in the statement E_2. Therefore, the satisfaction that we get from achieving *both* the degrees d_1 and d_2 is (when $d_1 \approx d_2$) practically equal to the satisfaction $s(E_1, d_1)$ of achieving the degree d_1 for the first statement, and thus, smaller than the sum $s(E_1, d_1) + s(E_2, d_2)$.

- In some other cases, there is a synergy between the statements E_1 and E_2. For example, these statement may cover two possible cases of some general interesting statement, and thus, the satisfaction of confirming both statement E_i with certain degrees of certainty means that we have covered the general statement as well. As a result, in this case, the total amount of satisfaction obtained from achieving both degrees of certainty is much larger than simply the sum of the satisfaction values $s(E_1, d_1) + s(E_2, d_2)$.

In some cases, the total satisfaction is smaller than the sum; in some other cases, the total satisfaction is much larger than the sum. It is reasonable to assume that for a large number of statements which may be related in different ways, on average, these positive and negative deviations from the sum more or less compensate each other. In the resulting approximation, the total satisfaction of achieving certainty degree d_1 for statement E_1, certainty degree d_2 on statement E_2, ..., and certainty degree d_n on a statement E_n can be described as the sum of the corresponding individual satisfaction levels, i.e., as $s(E_1, d_1) + \ldots + s(E_n, d_n)$.

We are interested in large bodies of knowledge, which contain many different statements E_i.

- Some of these statements are more relevant and interesting, and therefore their confirmation brings in more satisfaction.

- Some of these statements are more technical, less relevant and less interesting, and therefore their confirmation brings in less satisfaction.

For a large body of knowledge, which contains both more interesting and less interesting statements, it is reasonable, as a first approximation to the sum $s(E_1, d_1) +$

$\ldots + s(E_n, d_n)$, to replace individual dependencies $d \to s(E, d)$ with a single "average" dependence $s(d)$ – i.e., by a satisfaction $s(d)$ coming from achieving a level of certainty d for an "average" statement. In other words, as a first approximation to the desired sum, we take the sum $s(d_1) + \ldots + s(d_n)$.

When we replace $s(E_i, d_i)$ by $s(d_i)$, then:

- for more interesting statements, for which the actual satisfaction level $s(E_i, d)$ is larger than average, this replacement underestimates the satisfaction level;

- for less interesting statements, for which the actual satisfaction level $s(E_i, d)$ is smaller than average, this replacement overestimates the satisfaction level.

It is reasonable to expect that on average, these positive and negative corrections to the sum formula more or less compensate each other. Therefore, as a reasonable first approximation to the total satisfaction corresponding to the levels of certainty $d_1, \ldots, d_n$, we can take the sum $s(d_1) + \ldots + s(d_n)$.

Final formalization. In this formalization, in order to find the optimal choice of degrees of certainty $d_1, \ldots, d_n$, we must solve the following conditional (constrained) optimization problem:

$$s(d_1) + \ldots + s(d_n) \to \max \qquad (1)$$

under the condition (constraint)

$$e(d_1) + \ldots + e(d_n) = E. \qquad (2)$$

3. Solution of the Formalized Problem Explains Granularity

Analytical solution to the above problem. The above conditional optimization problem can be easily solved by using the standard calculus technique of Lagrange multipliers, according to which the above conditional optimization problem can be reduced to an unconditional one

$$s(d_1) + \ldots + s(d_n) + C \cdot [e(d_1) + \ldots + e(d_n) - E] \to \max,$$

where C is a constant (Lagrange multiplier).

For this unconditional optimization problem, the maximum can be computed by simply differentiating the objective function with respect to d_i and equating the resulting partial derivative to 0. As a result, we get the following equation: $s'(d_i) + C \cdot e'(d_i) = 0$. So, all the degrees d_i corresponding to the optimal degrees selection must be the solutions to the equation $F(d) = 0$, where we denoted $F(d) = C - s'(d)/e'(d)$. In other words, all these degrees must be the roots of a function $F(d)$.

This analytical solution explains granularity. Intuitively, small changes in the certainty degree d should

not drastically affect neither the average effort $e(d)$ required to achieve this degree, nor the average satisfaction $s(d)$ resulting from achieving this degree. Therefore, it is reasonable to assume that the functions $e(d)$ and $s(d)$ are smooth and probably even analytical (i.e., can be expanded in Taylor series). In this case, the function $F(d)$ is also an analytical function defined on the interval $[0, 1]$. It is known that an anlytical function which is not identically 0 can only have finitely many roots on an interval. Thus, all the optimal degrees of certainty d_i must belong to the finite set of these solutions.

For usual analytical functions, this set of solutions is small. Indeed, an arbitrary analytical function, by definition, is equal to its Taylor series and therefore, can be approximated, with an arbitrary accuracy, by a polynomial. A polynomial of degree n can have no more than n roots; so, e.g., if a cubic polynomial is a reasonable approximation for the function $F(d)$, then, in this approximation, the function $F(d)$ has no more than 3 roots, so we use no more than three different levels of certainty. A more accurate approximation, e.g., by a 7-th order polynomial (which is usually enough to visually coincide for most known analytical functions on $[0, 1]$ such as $\sin$, $\cos$, etc.), would reveal no more than 7 different degrees of certainty, etc.

In other words, no matter how many statement we consider, for each of these statements, the optimal degree of certainty d_i should belong to the same (small) set. Thus, even if we start with the degrees which can, in principle, take arbitrary values from the interval $[0, 1]$, we end up showing that in the optimal assignment, only a few of these values will be actually used.

Hence, granularity is indeed optimal.

4. Similar Ideas Can Be Applied to Sleep, Consumption, Traffic Control, Learning

Application to sleep. We can apply similar ideas to the description of other biological processes. For example, every biological creature has a certain level of activity. It cannot maintain the highest possible level of activity all the time, because its resources are limited. Therefore, it must distribute these resources in such a way that the overall efficiency is the largest. Let $d_1, \ldots, d_n$ denote levels of activity at different consequent moments of time. Let $s(d)$ denote the productivity of the activity with level d, and let $e(d)$ denote the effort needed to maintain this activity level.

Then, the optimal levels of activity can be determined by solving the optimization problem (1), (2). We already know, from the solution to this problem, that in the optimal solution, the levels d_i cannot take arbitrary values, they should all belong to a small set of values. Therefore, the optimal activity schedule consists not of slowly

changing activity from one level to another, but of switching between several discrete levels of activity.

This conclusion explains why, instead of a slow transition between high and low activity, most living creature have an abrupt transition between activity and sleep.

If we take subtler details into consideration, then we can say that the above conclusion explains why living creatures have an abrupt transition between activity and several levels of sleep such as a REM phase and a normal sleep.

Similarity with "bang-bang" control in control theory. The above conclusion is consistent with the fact that in control theory, the optimal control often involves abrupt changes from one regime to another. For example, when driving a car, stability means, in particular, that once the car swerved, it should return to the original trajectory. The faster it returns, the more stable is the system. Therefore, from the viewpoint of stability only, the ideal (optimal) control would be the one that brings the car back on track in the shortest possible time (i.e., with the largest possible acceleration).

The non-smoothness of the optimal control is not a peculiar feature of the car example: in control theory, there are general theorems that show that under certain (reasonably general) conditions, the optimal control is indeed of the above-described "bang-bang" type (see, e.g., [2]; not incidentally, the word "bang-bang" is an "official", well-defined and widely used term in control theory).

Application to consumption. Similarly, for a person with limited resources, the consumption schedule which leads to the largest satisfaction is not the schedule in which these resources are equally distributed, but rather the schedule in which periods of higher consumption ("feasts") are abruptly changing to periods of lower consumptions ("fasts"). This conclusion is in good accordance with the results obtained by economists who analyzed more complicated economic models (see, e.g., [1, 4]).

Application to traffic control. In traffic, similar idea explains why the optimal traffic arrangement means that we fix a small number of speed levels, and assign (maybe dynamically) each road to one of these levels. In real life, such levels are freeway, city limits, school zone, etc.

Application to learning. In learning, the optimal distribution in learning activity is not a steadfast study, but rather periods of intense study separated by periods of relative rest.

Similarly, if we analyze the distribution of learning activity in the population, so that d_i described the amount of effort used to educate i-th person, we come to the conclusion that the optimal arrangement is not when the teaching efforts are uniformly distributed among students or when there is a continuous change from one student to another, but rather when there are a few levels and each student is assigned to a certain level of studying.

This result is in good accordance with the discrete system of university education, where the possible levels of education in a given domain are described in terms of a small list of degrees (e.g., BSc, MSc, Ph.D.).

5. Possible Future Work

So far, all we got was a justification of granularity as a high-quality approach to the description of uncertainty. However, since we know where this granularity comes from, for new problems, we can not only *justify* granularity, but we can also find the *optimal* level of granularity.

The authors are thankful to the anonymous referees for valuable comments.

References

[1] J. Elster, "Rationality and the emotions", *The Economic Journal*, 1996, Vol. 438, pp. 1386–1397.

[2] R. R. Mohler, *Nonlinear systems. Vol. 1. Dynamics and control*, Prentice Hall, Englewood Cliff, NJ, 1991.

[3] H. T. Nguyen and V. Kreinovich, *Applications of continuous mathematics to computer science*, Kluwer, Dordrecht, 1997.

[4] T. Scitkovsky, *The Joyless Economy*, Oxford University Press, Oxford, 1992.