Knowledge Geometry Is Similar to General Relativity: Both Mass and Knowledge Curve the Corresponding Spaces

Francisco Zapata[1] and Vladik Kreinovich[2]
[1]Research Institute for Manufacturing and Engineering Systems
[2]Department of Computer Science
University of Texas at El Paso
El Paso, TX 79968, USA
fazg74@gmail.com, vladik@utep.edu

Abstract

In this paper, we explain and explore the idea that knowledge is similar to mass in physics: similarly to how mass curves space-time, knowledge curves the corresponding knowledge space.

1 Knowledge Geometry

How to detect that two objects are different? Let us start with the following sample situation. To observe wild parrots, an ornithologist sets up a feeder. Every morning, a parrot appears, and every evening, a parrot appears. These two parrots look similar. A natural question is: is the same parrot coming in the mornings and in the evenings, or these are two different parrots?

Natural idea: observe properties. A natural way to answer the above question is to observe various properties of these two parrots. For example, if the morning parrot has a red spot, and the evening parrot does not, this means that they are different birds. If the wing span of evening parrot is smaller than the wing span of the morning parrot, they are different birds.

Without losing generality, we can consider binary properties. In general, properties can be of binary (yes-no) type, e.g., "has a red spot". We can also consider numerical properties like the wing span. In the computer, whatever information we have can be represented in terms of binary digits (bits), i.e., in terms 0s and 1s:

- on the one hand, the property of having or not having a red spot is represented by a single bit;

- on the other hand, the numerical value of the wingspan is represented by several bits.

Instead of considering all these different types of properties, let us simply consider all the information as the sequence of bits. From this viewpoint, measuring the wing span means determining the values of several binary properties:

- the first of these binary properties is the value of the 1st bit in the binary expansion of the measured value,
- the second of these binary properties is the value of the 2nd bit in the binary expansion of the measured value,
- etc.

Representing knowledge about each object. Let N be the total number of binary properties. Let us denote these properties by $P_1, \ldots, P_N$. For each object a and for each property P_i, we have the following three possibilities:

- the first possibility is that we know that the property P_i holds for the object a, i.e., that the value $P_i(a)$ is "true"; in the computers, the value "true" is usually represented by 1;
- the second possibility is that we know that the property P_i does not hold for the object a, i.e., that the value $P_i(a)$ is "false"; in the computers, the value "false" is usually represented by 0;
- the third possibility is that we do not know whether the object a satisfies the property P_i; let us denote this case by $P_i(a) = *$.

Gauging difference between the two objects. In the first approximation, it is reasonable to gauge the difference between the two objects by the number of properties in which these two objects differ. In other words, if the object a is characterized by the values $P_1(a), \ldots, P_N(a)$, and the object b is characterized by the values $P_1(b), \ldots, P_N(b)$, then we take $D(a, b) \stackrel{\text{def}}{=} \sum\limits_{i=1}^{N} d(P_i(a), P_i(b))$, where we define:

- $d(v, v') = 1$ if we know that the values v and v' are different, i.e., when either $v = 0$ and $v' = 1$, or $v = 1$ and $v' = 0$; and
- $d(v, v') = 0$ for all other pairs v and v'.

Some properties may be more important, some less important. To take the difference in importance into account, we can assign weights $w_i > 0$ to different properties P_i, so that differences in the more important properties will be added with more weight. In other words, it makes sense to consider the following formula for the distance between the two objects:

$$D(a, b) = \sum_{i=1}^{N} w_i \cdot d(P_i(a), P_i(b)). \tag{1}$$

Comment. The idea of a reasonable knowledge-based distance between objects is not new; it has been described, e.g., in [8].

2 Both Mass and Knowledge Curve the Corresponding Spaces: An Idea

Observation: additional knowledge increases distances. Let us analyze what happens to thus defined knowledge-based distance between objects if we gain additional knowledge. Gaining additional knowledge means that for some properties and for some objects:

- where we previously did not know whether this property is satisfied or not (i.e., we had the unknown value $*$),
- now we know that this property is satisfied or that it is not satisfied.

How does this change in knowledge affect the distance $D(a, b)$, i.e., a weighted sum of the distances $d(P_i(a), P_i(b))$?

If for some property P_i, we had $d(P_i(a), P_i(b)) = 1$, this means that one of the values $P_i(a)$ and $P_i(b)$ was equal to 0 ("false") and another to 1 ("true"). In other words, if $d(P_i(a), P_i(b)) = 1$, this means that we already know both truth values $P_i(a)$ and $P_i(b)$. For this property, the additional knowledge will not change these truth values and thus, the distance $d(P_i(a), P_i(b)) = 1$ will remain unchanged. So:

- values $d(P_i(a), P_i(b)) = 1$ remain unchanged, while
- the values $d(P_i(a), P_i(b)) = 0$ may increase to 1: e.g., if some truth values were unknown $P_i(a) = P_i(b) = *$, and we found out that the property P_i is false for a and true for b.

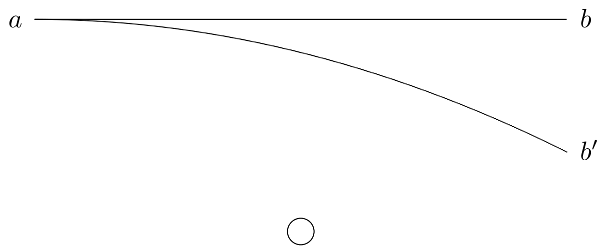
In both cases, the value $d(P_i(a), P_i(b))$ either remains the same or increases – and, as a result, the distance $D(a, b)$ between the two objects either remains the same (if we did not learn any new information about their difference) or increases. In short, in general, *additional knowledge increases distance.*

Conclusion: shortest paths change. In the vicinity of the object c for which we gained the new knowledge, distance increases. As a result, if originally, the shortest path between some objects a and b passed through c (or near c), its length increases – and an alternative path which does not pass near c becomes now the shortest.

Example. The changing of the shortest path can be illustrated on the example of traffic. Let us measure the distance $d(a, b)$ between the two points by the time that it takes to travel between the locations a and b. In the absence of heavy traffic (e.g., at night), the shortest path, e.g., between a location to the South of downtown and a location to the North of it goes through downtown.

However, during the rush hours, the traffic in downtown is usually congested. As a result, the path through downtown becomes much longer. In this case, a different path will be the shortest: the one which goes around downtown.

This is similar to curving of space-time. This phenomenon is similar to the geometric interpretation of gravity in General Relativity; see, e.g., [4, 6]. In the absence of masses, a body follows the straight line – which happens to be the shortest path between the initial position a and the final position b. In the presence of a heavy mass (e.g., the Earth or the Sun), the bodies start falling on this mass.



Shortest paths are a specific example of *geodesics*, i.e., paths on which the length is stationary (e.g., the smallest or the largest). In General Relativity, particles follow the geodesics in space-time. Because of the curving of space-time (largely time), spatial projections of geodesics are not straight lines – i.e., the actual path of a body in a curved space is different from the shortest path as measured by the spatial distance.

3 Both Mass and Knowledge Curve the Corresponding Spaces: Towards a Quantitative Description

Towards a quantitative description. In the ideal case, we know the exact values $v_i(a)$ of all the quantities describing an object a. In such an ideal situation, it is easy to check whether two objects a and b are identical or not:

- if for all quantities, we get $v_i(a) = v_i(b)$, then the objects a and b are indistinguishable – i.e., in effect, a and b are the same object;
- if for some quantity i, we have $v_i(a) \neq v_i(b)$, this means that the objects a and b are different from each other.

In practice, the values come from measurements, and measurements are never 100% exact; there is always a measurement error due to which the measured value $\widetilde{v}_i(a)$ is, in general, somewhat different from the actual (unknown) value

$v_i(a)$ of this quantity; see, e.g., [7]. In many cases, we know the probability distribution of the measurement error $\Delta v_i(a) \stackrel{\text{def}}{=} \widetilde{v}_i(a) - v_i(a)$. Often, this distribution is Gaussian; this is in line with the fact that usually, many different phenomena contribute to the measurement error, and, according to the Central Limit Theorem, the distribution of the sum of a large number of small independent random variables is close to Gaussian; see, e.g., [10]. It is usually assumed that the bias (mean error) has been compensated, so the mean value of the measurement error is 0.

Because of the measurement errors, even if a and b are the same object, i.e., even if the actual values of the corresponding quantities coincide $v_i(a) = v_i(b)$, the measured values $\widetilde{v}_i(a) = v_i(a) + \Delta v_i(a)$ and $\widetilde{v}_i(b) = v_i(b) + \Delta_i(b)$ will be, in general, slightly different. Vice versa, when the objects differ and $v_i(a) \neq v_i(b)$ for some i, it is possible that we will have $\widetilde{v}_i(a) = v_i(a)+\Delta v_i(a) = v_i(b)+\Delta_i(b) = \widetilde{v}_i(b)$, i.e., the observed values will be the same.

So, based on the measurement results, we can never know for sure whether the two objects a and b are identical or not, we can only make this conclusion with a certain probability. Specifically, based on the known probability distribution of the measurement error, we can estimate what is the probability that the observed values $\widetilde{v}_i(a)$ and $\widetilde{v}_i(b)$ come from the same object.

- When this probability is high, we conclude that the objects are most probably the same.
- When this probability is low, we conclude that the objects are different.

It is therefore reasonable to define the distance between the objects a and b in such a way that the larger the distance, the smaller the corresponding probability. Namely:

- we assume that we observe the exact values of the corresponding quantities $v_i(a)$ and $v_i(b)$, and
- we compute the probability that the difference between these values can be explained by the measurement errors.

Gaussian distributions correspond to Riemannian geometry, more general distributions to more general (Finsler) geometry. Let us assume that we observe the values $v_i(a) \neq v_i(b)$. Under the hypothesis that the difference between these values can be explained by the measurement error, i.e., that $v_i(a) = v_i + \Delta v_i(a)$ and $v_i(b) = v_i + \Delta v_i(b)$ for some values v_i, we conclude that $\Delta v_i(a) - \Delta v_i(b) = d_i \stackrel{\text{def}}{=} v_i(a) - v_i(b)$.

Measurement errors $\Delta v_i(a)$ and $\Delta v_i(b)$ corresponding to different measurements are usually independent. So, when the measurement errors Δh_j are normally distributed (with 0 mean), the difference $d_i = \Delta v_i(a) - \Delta v_i(b)$ is also normally distributed (and also with 0 mean). The corresponding probability density function has the form $\text{const} \cdot \exp(-c_i \cdot (d_i)^2)$ for an appropriate value c_i.

Since we assumed that measurement errors corresponding to different measurements are independent, we conclude that the overall probability that the objects a and b are different is equal to the product of the corresponding probabilities, i.e., to the value $\prod\limits_i \text{const} \cdot \exp(-c_i \cdot (d_i)^2) = \text{const} \cdot (-s)$, where $s \stackrel{\text{def}}{=} \sum\limits_i c_i \cdot (d_i)^2$.

So, the probability is uniquely determined by the weighted sum s. In general, each object a can be characterized by the values of the corresponding parameters $a_1, \ldots, a_n$, and the quantities $v_i(a)$ smoothly depend on these parameters. As a result, for close objects $a = (a_1, \ldots, a_n)$ and

$$b = a + \Delta a = (a_1 + \Delta a_1, \ldots, a_n + \Delta a_n),$$

we get

$$d_i = v_i(a + \Delta a) - v_i(a) = \sum_{j=1}^{n} D_{ij} \cdot \Delta a_j + o(\Delta a),$$

where $D_{ij} \stackrel{\text{def}}{=} \dfrac{\partial v_i}{\partial a_j}$, and thus,

$$s = \sum_i c_i \cdot (d_i)^2 = \sum_i c_i \cdot \left(\sum_{j=1}^{n} D_{ij} \cdot \Delta a_j \right)^2 + o(\Delta a).$$

Therefore, s is a quadratic function of Δa_j, $s = \sum\limits_{jk} g_{jk} \cdot \Delta a_j \cdot \Delta a_k$ for some g_{jk}. Thus, the value s is naturally related to the Riemannian distance $d(a, a + \Delta a) = \sqrt{\sum\limits_{jk} g_{jk} \cdot \Delta a_j \cdot \Delta a_k}$.

For more general probability distributions, we get a more general formula for the corresponding metric – i.e., in the smooth case, a general Finsler space [1, 2, 3, 5, 9] instead of a specific Riemannian space.

Acknowledgments. This work was supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

The authors are thankful to Luis Rocha and Ron Yager for stimulating discussions.

References

[1] P. L. Anotonelli et al. (eds.), *Handbook of Finsler geometry*, Vols. 1, 2, Boston: Kluwer Academic Publishers, 2003.

[2] D. Bao, S. S. Chern and Z. Shen, *An Introduction to RiemannFinsler Geometry*, Springer-Verlag, 2000.

[3] S. Chern, "Finsler geometry is just Riemannian geometry without the quadratic restriction", *Notices AMS*, 1996, Vol. 43, pp. 959–963.

[4] Feynman R., Leighton R., and Sands M. The Feynman lectures on physics, Boston, Massachisetts:Addison Wesley, 2005.

[5] P. Finsler, *Über Kurven und Flächen in allgemeinen Räumen*, PhD Dissertation, Göttingen University, 1918; reprinted by Birkhäuser, 1951.

[6] Misner C.W., Thorne K.S., and Wheeler J.A. Gravitation, W. H. Freeman:New York, 1973.

[7] S. Rabinovich, *Measurement Errors and Uncertainties: Theory and Practice*, American Institute of Physics, New York, 2005.

[8] L. M. Rocha, "Combination of evidence in recommendatuon systems characterized by distance functions", *Proceedings of the 2002 International IEEE Conference on Fuzzy Sets and Systems FUZZ-IEEE'2002*, Honolulu, Hawaii, May 2002, pp. 203–208.

[9] Z. Shen, *Lectures on Finsler Geometry*, World Scientific Publishers, Singapore, 2001.

[10] D. J. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, Chapman and Hall/CRC Press, Boca Raton, Florida, 2011.