9-2015

Occam's Razor Explains Matthew Effect

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

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

Comments:
Technical Report: UTEP-CS-15-67
Published in *Journal of Innovative Technology and Education*, 2015, Vol. 2, No. 1, pp. 47-50.

Occam's Razor Explains Matthew Effect

Olga Kosheleva and Vladik Kreinovich
University of Texas at El Paso
500 W. University
El Paso, Texas 79968, USA
olgak@utep.edu, vladik@utep.edu

Abstract

Sociologists of science noticed that the results of many collaborative projects and discoveries are often attributed only to their most famous collaborators, even when the contributions of these famous collaborators were minimal. This phenomenon is known as the *Matthew effect*, after a famous citation from the Gospel of Matthew. In this article, we show that Occam's razor provides a possible explanation for the Matthew effect.

1 Matthew Effect: A Brief Description

In the 1960s, Robert K. Merton, a sociologist of science, observed that there is a tendency to remember only the most famous contributors to a collaborative project, even when the contribution of these famous contributors was minimal. As a result, the famous researchers become even more famous, while others are largely forgotten [3]. He called this phenomenon *Matthew effect*, after a citation from the Gospel of Matthew 25:29: "For whoever has will be given more, and they will have an abundance. Whoever does not have, even what they have will be taken from them."

This phenomenon has been observed in many areas of science. For example, it has recently been emphasized in an article about the contribution of female scientists to computing which was published in Communications of the ACM, the main Computer Science magazine [1].

2 How Can We Explain the Matthew Effect?

Our knowledge decays with time. In the beginning, we may be well aware of the contributions of different scientists to a project. However, as time goes by, we tend to forget the details – and this explains why with time, our perception of contributions of different scientists changes.

To understand why these changes follow the Matthew-effect pattern, let us describe the knowledge decay in precise terms.

How to describe knowledge decay in precise terms. Let us describe this knowledge decay in precise terms. In the beginning, for each project, we know:

- the names of all its contributors, and
- the percentages $p_1, p_2, \ldots$ that describe the relative contribution of each of the contributors.

These percentages add up to 1: $\sum_i p_i = 1$.

As the knowledge decays, for those projects that we still remember, instead of the *exact* values $p_i \in [0, 1]$, we have only *partial* information about these values. The corresponding uncertainty increases with time, until we reach the maximal uncertainty, when about each of the percentages p_i, we only know that this value is between 0 and 1.

Occam's razor: a way to select a possible description in case of uncertainty. As our knowledge decays, uncertainty follows. Uncertainty means that there are are several different descriptions which are consistent with our uncertain knowledge. In many cases, we select a description which is – in some reasonable sense – the simplest.

The idea of selecting the simplest explanation was first explicitly described William of Ockham (c. 1287–1347) and is thus known as *Occam's razor*. In the 20th century, it was shown that when we interpret complexity (and simplicity) as algorithmic complexity, then the correspondingly formalized Occam's razor is indeed an (asymptotically) optimal way to gaining knowledge about the world; see, e.g., [2]. Let us therefore apply Occam's razor to our problem.

Possible descriptions of contributions to different projects. We would like to apply Occam's razor and select the simplest description of contributions which is consistent with our uncertain knowledge. To make this selection, we need to find out what are these possible descriptions – and what is their algorithmic complexity.

Let us assume that the set of projects is fixed. Let P be the number of the projects. If we use b-bit strings to identify a project, then we can have at most 2^b different strings, and thus, we can identify at most 2^b different projects. Thus, to identify a project, we need to have $2^b \geq P$, i.e., we need to use at least $b_0 \stackrel{\text{def}}{=} \lceil \log_2(P) \rceil$ bits. The simplest possible representation is when we use exactly this many bits, e.g., if we identify each project simply by its ordinal number 0, 1, 2, $\ldots$ (in binary form).

We also need to have a list C of contributors. Contributors can then also be identified by their ordinal numbers.

In these terms, each possible description of contributions means that for each of P projects, we have:

- a list of contributors, and
- for each of these contributors, the corresponding percentage $p_i > 0$.

We want descriptions to be *correct*, i.e., we want this list to only contain the actual contributors to the project – although we allow the possibility that not all original contributors are remembered.

Which of these possible descriptions is the simplest? For each project, the fewer contributors we mention, the fewer bits we need to store this information. Thus, the simplest possible description of each project is when we attribute it to a single contributor – for which, then, the corresponding percentage is then equal to 1 and thus, does not need to be explicitly listed.

The listing of only one contributor is consistent with our uncertainty, since the only remaining information about each percentage p_i is that this percentage can be any number between 0 and 1 – in particular, it can be equal to practically 0, meaning that the actual contribution of this contributor was negligible.

Thus, in the simplest possible description, we have:

- a list of C contributors, and
- for each of P projects, an ordinal number describing this project's (single) contributor.

Let us count how many bits we need for this description; then, we will then be able to select the simplest of such descriptions as the one that requires the smallest number of bits to store.

Let w be an average number of bits needed to store a contributor's name. Then, the listing of all C contributors requires $C \cdot w$ bits.

For each of the P projects, we need to store the ordinal number of the corresponding contributor. For C contributors, we need $c = \lceil \log_2(C) \rceil$ bits to store each such ordinal number, so overall, we need $P \cdot c = P \cdot \lceil \log_2(C) \rceil$ bits.

Thus, the total number of bits that we need to store a possible description is equal to the sum $C \cdot w + P \cdot \lceil \log_2(C) \rceil$. This number of bits is increasing with C, so the smallest number of bits corresponds to the case when the number C of listed contributors is the smallest possible.

Some projects have originally had a single contributor. The corresponding contributors have to be included in this list anyway. Let us call these contributors *famous*. So, if for some other project p:

- one of the original contributors is famous in this sense (i.e., is also a sole contributor to some other project $p' \neq p$), while
- other original contributors have only contributed to this particular project p,

the simplest possible description is when the project p is attributed only to the famous collaborator. Indeed, otherwise, if we attributed this project to one of the other original contributors, we would need to add the name of that collaborator to the list C and thus, take extra bits to store this additional information.

Conclusion: Occam's razor indeed explains Matthew effect. We have shown that Occam's razor – in this case, the idea of selecting the simplest possible model consistent with our imprecise memory – indeed explains why many collaborative scientific projects are often contributed solely to their most famous collaborators. In other words, Occam's razor indeed explains the Matthew effect.

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.

References

[1] T. Haigh and M. Priestley, "Innovators assemble: Ada Lovelace, Walter Isaacson, and the superheroines of computing", *Communications of the Association for Computing Machinery (ACM)*, 2015, Vol. 58, No. 9, pp. 20–27.

[2] M. Li and P. M. B. Vitányi, *An Introduction to Kolmogorov Complexity*, Springer-Verlag, Berlin, Heidelberg, New York, 2008.

[3] R. K. Merton, "The Matthew effect in science", *Science*, 1968, Vol. 159, No. 3810, pp. 56–63.