Complexity and Scale: Understanding the Creative

Oliver Richardson

Abstract. This paper proposes a method of examining large informational bodies through the eyes of an information consumer. There is a well-studied cryptographic dichotomy between the random and the predictable, and between the simple and the complex, but autonomous information consumers (people, for instance), tend not to prefer data too far towards either end of the spectrum. They cognitively detect some over property of data and call it “interesting”.

By defining a quantity called *novelty,* as the integral of Kolmogorov complexity with respect to scale, we can capture many of the properties of the common-sense notion of the word. The quantity so defined is somewhat subject-relative, though it has a well-defined objective component. The paper will apply this analysis to the appreciation of narrative, shedding light on the meaning of the word “interesting”. All of this demonstrates how mathematics can reveal aspects of the nature of creativity and intellectual motivation.

Introduction

You are a pilot. Every day you go to work at the same time. Every day you check the weather and flight schedule, board the plane, press the same switches, and say the same thing to your passengers.You like your job, but the piece that bothers you is its completely static nature.So you decide to do something bold.To make life interesting and reach your maximum potential, you quit your job and resolve to work in a new place every day. You become a barber, then a political organizer, then a janitor, an accountant, a chemical engineer… and by the end of the week, you are happy with the amount of progress you have made towards having a stimulating life. A year later, however, you begin to feel bored with your life, just as you were before, although you are not immediately sure why this might be the case. At this point, some people mightcrawl back to their old jobs, muttering that there is no escape from life’s crushing monotony. Others might continue to look for new jobs, insisting that the jadedness was an artifact of their imaginations. But you do neither - instead, you make a crucial connection between these two pieces of your life: while every day is new and different, every week feels the same. These lifestyles differ only in time scale. You begin to see that interestingness lies neither in absolute pattern, nor in complete randomness.

adfa, p. 1, 2011.

The purpose of this paper is to give a mathematical account of this insight, in terms of computational complexity – an account that will be very widely applicable to many areas where the terms "interesting" and "novelty" are routinely applied. Though it is necessary to lay out some formal structural framework in order to clearly define the quantities that we wish to discuss, the majority of the interesting ideas in this piece are approximations and observations that stem from the core definitions.

The Basic Technical Apparatus: String Complexity

The conception to be introduced here relies heavily on the notion of the *complexity* of anobject ***o***as represented in terms of a string of characters ***s***. For this, I will use Kolmogorov's definition:

Let P be the set of all programs in any universal programming language that deterministically output***s***. P is a non-empty set, as the line "print ***s***" in python, for example, is in P. Choose $\boldsymbol{d} \in \{ x : |x| \leq |y| \ \forall \ y \in P \}$.We can now define the K-complexity by $K(\mathbf{o}) = |\mathbf{d}|$, where $|\ \mathbf{m}\ |$ is the number of bytes in the program **m.**

Though it may be obvious what this operation does, the following examples serve to illustrate that although this definition is extremely useful, it is not a great measure of the "interestingness" of a string. Consider the following three alphanumeric strings:

(1)`AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA`
(2)`0CrMQXAhpkzGllzYcW3N wcy6TONwPn79IyxzJdW0F BZot3XPTpWrKi0Q78`
(3)`a mathematician is a device for turning coffee into theorems`[1]

String (1) has low relatively complexity, as the following coarse python snippet already does a decent job of compressing the string into a description.

```
for x in range (1, 60)
    print "A"
```

By contrast, string (2) was randomly generated, and hence has extremely high complexity. It is not apparent that it is possible to provide any description shorter than the length of the string itself.String (3) is by far the most interesting of the three strings, but it is not maximally complex. All of the words are in an English dictionary, delimited by spaces, and utilize only lower-case letters –all patterns that a program could easily exploit to generate the same data in a fewer number of bytes.

Observe: complexity and interestingness are not the same thing. A random sequence is maximal in K-complexity but can be quite low in interestingness, but a

[1] This is a famous Paul Erdős joke. Using this example is cheating somewhat because it draws outside source of information for context (knowledge of the English language, mathematics, coffee) – but the point is obvious anyway.

considerably more interesting string (Moby Dick) will be less complex than an equally long string of random characters selected from the same pool.

For convenience, we will define a quantity called "information density" as the ratio of a string's complexity to its length.

$$\boldsymbol{\sigma}(\boldsymbol{s}) = \frac{K(\boldsymbol{s})}{|\boldsymbol{s}|}$$

Complexity-Scale Curves

Let us now consider an arbitrarily long random string $\boldsymbol{s}$ – but this time instead of immediately looking at its complexity, we will first divide it into uniform mutually exclusive regions of size $\boldsymbol{n}$(Assume that for any given $\boldsymbol{n}$, $|\boldsymbol{s}| \gg \boldsymbol{n}$). Furthermore, we will introduce a new alphabet in which to represent our new meta-strings, composed of sorted[2]$\boldsymbol{n}$-tuples of the letters in the original language[3]. We also wish to exploit the fact that (AAABBBC) is "kind of close" to (AAAABBC), so we optimize our language such that we represent sequences of similar meta-letters with lists of replacements, hence decreasing the amount of information in sequences of close letters, and, by extension, the size of the minimal description. To denote the i^{th} such partition of $\boldsymbol{s}$, we write $[s]_i$. Although this may seem like an unnecessarily complicated formulation, this optimization gives us a significant asymptotic informational improvement as $\boldsymbol{n}$ goes to infinity. Now: what happens as we examine the information density of the new string comprised of these meta-letters?

For $\boldsymbol{n}$ = 1, we know that this abstraction is no different from examining the string with itself. Since it is a giant random string, we know that we know that we have maximal K-complexity, and so $\boldsymbol{\sigma}([\boldsymbol{s}]_1) = 1$.

On the other end of the spectrum, if we were to chunk our arbitrarily large string into arbitrarily large pieces, we would find that every chunk (or meta-letter) looks roughly the same. This isdue to the fact that each meta-letter is now also an arbitrarily large random string, and hence should have equal distributions of letters. The K complexity of the object constructed as a sequence of these pieces is minimal ($\lim_{n\to\infty} \boldsymbol{\sigma}([\boldsymbol{s}]_{\boldsymbol{n}}) = 0$); this case is analogous to the case of the sequence of a repeated single character.

---

[2] This is necessary to eliminate order as a discriminating factor within chunks. For example, (A, B, C), and (C, B, A), become the same triplet when sorted.

[3] For n = 3, for example, an element in our new alphabet becomes something of the form (x,y,z), where x, y, and z are members of the old one, such that for our sorting metric, x ≤ y ≤ z. For brevity's sake, we will omit commas.

Now that we know what our endpoints look like, we can make some decent predictions about the intermediate behavior. For this random string ***s*****,** partitioned into chunks of size ***m*** or ***n***,

$$m > n \Rightarrow \sigma([s]_m) < \sigma([s]_n),$$

since a larger partition size indicates a larger stability and grouping of the characters relative to their possible configurations.

It is much faster and more intuitive to visualize this process, which is what we will refer to as a “Complexity-scale curve”. The culmination of our previous discussion of a random string is shown below, in a rather cartoonish fashion.

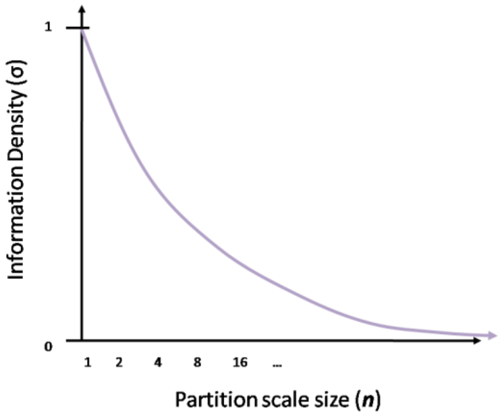


Fig. 1. Above is the complexity-scale curve for an infinitely long string of random numbers.

But the fun doesn’t end there! We can represent *any* string of text or information in this manner. Bellow we have a diagram of our other, more trivial extreme case, and also a slightly more interesting one that illustrates the fact that these curves are not, in general, strictly decreasing. By trippling every letter, the string is random when looking at sets of three letters at once, and hence is maximized at ***n*** = 3.

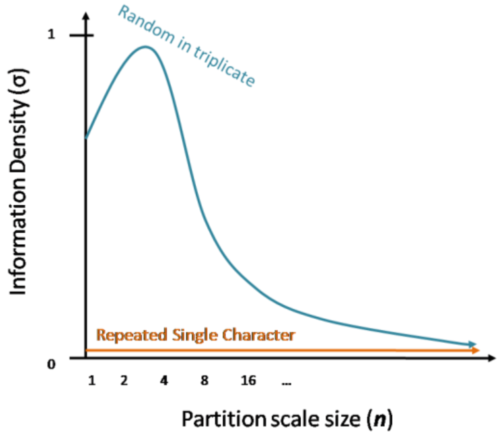
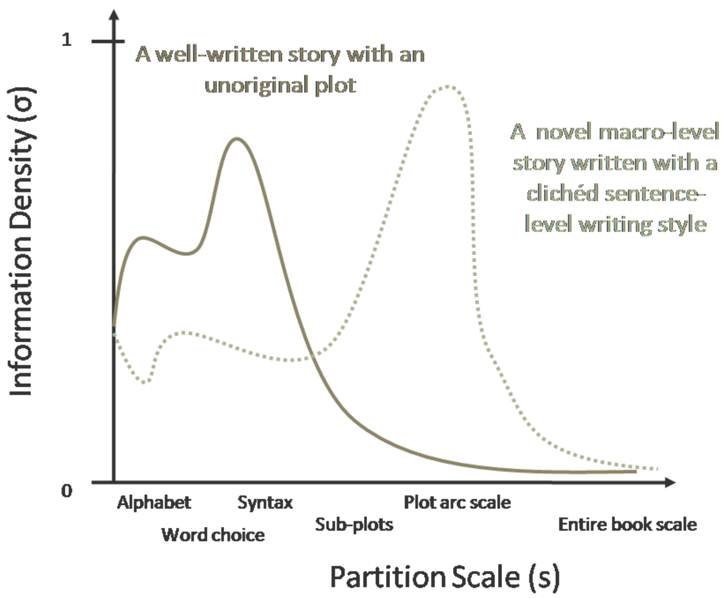


Fig. 2. A few more illustrative complexity-scale diagrams

But these analyses of data streams apply (in theory, at least) to data streams that look far more human. Literature, for example, is one field in which the strings of information are long enough to merit a scale-based analysis, and in which being interesting is of primary importance. With this in mind, we begin to examine some complexity-scale diagrams. In order to do this, we will need to mentally extend our definition of partition size to apply to any scale $n \in \mathbb{R}$. Let us start by labeling a few commonly cited narrative traits on our scale-complexity diagram[4].



In the above diagram, there are several things to take away. Firstly, as one would expect of an analysis of literature, different traits are independent degrees of freedom – but not completely independent from one another, due to the continuity of the function[5]. This is one sign that we're on the right track.

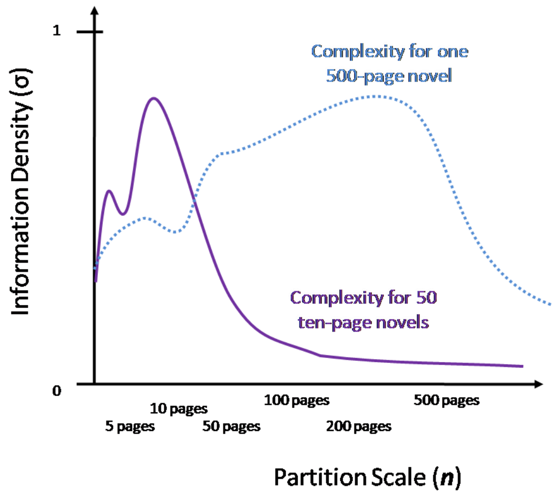
Furthermore, by being slightly less pedantic about the formal framework definitions provided earlier, you can use this technique to quickly understand some very important intrinsic qualities about a large quantity of information. This model also

[4] Note that this diagram extrapolates a little bit. In particular, "Clichéd writing style" reacts against other people's work, not one's own, and to correct for this, you provide the relevant context in the beginning of the string ***s***

[5] This comes from our informal extension of the partition scale domain to $\mathbb{R}$.

does a good job of explaining the relationship between the length of a piece and the time that a person puts into it.

It is perhaps interesting to note that there is another way of diagramming novels, commonly used in the study of narrative: an intensity-time diagram. This construction allows analysts to see similarities in plot intensity over the course of the novel, but it does a poor job of showing the smaller details, which are often just as important. The transformation from intensity-time to complexity-scale functions as an informal analogue of a Fourier or LaPlace transform, in that it decomposes a large number of data into their frequencies. We will continue with one more example in literature.



This illustrates something that many people have found in their experience of both genres of writing. Across smaller time scales, reading smaller books is more rewarding. If this is the order of time that you like to focus on, be it due to scheduling or intrinsic mental processes, short stories and poems are more rewarding on the smaller scale. Longer books, on the other hand, have more total complexity over a long period of time, if you would prefer to focus on five hundred pages of narrative at once and see the over-arching connections between smaller pieces.

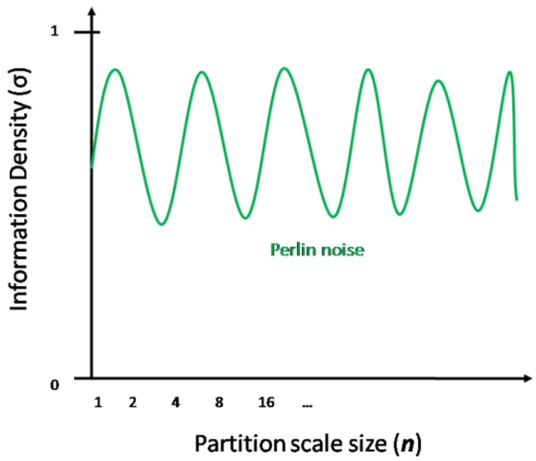
It is superficially useful to think of information density as a scoring mechanism for how interesting something is. However, one needs to recognize that it's a rather poor scoring mechanism, since changes based on the perspective and scale that a person cares about. What we need is something that takes all the scales into account and gives us a single score for "interestingness".

Novelty

Finally, we have a formal description of the "interestingness" property. We will call this the *novelty* of an object $\boldsymbol{o}$.

$$\mathrm{N}(\boldsymbol{o}) = \int_{-\infty}^{\infty} \boldsymbol{\sigma}([\boldsymbol{s}]_{\mathbf{x}}) e^{x} dx$$

This is the total area underneath our complexity-novelty curve, finally giving us the desired scoring algorithm that takes into account complexities for all breakdowns of an object into pieces on a given scale. From its definition alone, after a little bit of fiddling, it is fairly easy to find a way to kindly persuade this metric of *novelty* to give you infinite values over the absolute range, and overly large ones over a given scale range. This exploit is to design our noise with fractals, so that we hit every scale with a value that is close to our maximum information density. Fractally generated noise, also known as "Perlin Noise" does this, more or less.



Well, first of all: If we're going to fail at classifying human information, then this is a pretty good way to do it. Perlin noise is some of the prettiest, most interesting noise that a person could come up with. It is often used to generate aesthetically pleasing constructs such as terrain, lightening, and clouds in video games, and when done right, it looks absolutely brilliant.

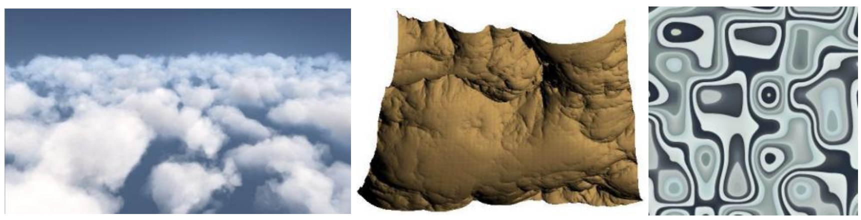


Fig. 3. Examples of why Perlin noise isn't so far from maximally interesting

But we haven't failed yet. Perhaps this is the best our novelty function can do with only the data stream we care about as input – but people have a huge backlog of informational context from their culture and experiences. Obviously, we can't take into account cultural facets and clichés without access to some source of contextual information. Obtaining warped cultural information is clearly extremely difficult, and perhaps worth an entire paper unto itself, so we will proceed without it[6]. But there is also a second reason that Perlin Noise hasn't broken our novelty system yet, stemming from the subjectivity that was hinted at earlier.

The Eye of the Beholder

Some people enjoy 1000-page novels; others prefer shorter stories. We can now explain this in mathematical terms.

Information consumers often have different affinities for different scales. For example, extremely detail-oriented people and specialists focus on smaller scales, whereas other groups of people are better described as generalists, focusing on larger scales. In general let's define the *interest function*, of a scale $\boldsymbol{s}$, as the amount of interest, affinity, or attention that an information consumer has for a given scale. This now allows us to define a more precise version of novelty tailored towards an individual, by masking off the pieces that are

$$\mathrm{N}(\boldsymbol{o}) = \int_{-\infty}^{\infty} \boldsymbol{\sigma}([\boldsymbol{s}]_{\mathbf{x}}) * I(x) * e^{x} dx\,,$$
$$where\ 0 \leq I(x) \leq 1\ \forall\ x \in \mathbb{R}$$

One benefit of this construction is that we can ensure that with a finitely large zone of interest, the derived novelty will also be finite. It is also nice that this definition allows us to apply the concept to quite different forms of information consumers, such as individuals with smaller memory, children, and perhaps even larger god-scale entities.

Creativity

Until this point in the essay, we have dealt only with information consumers, and the way in which they interact with the world. Now it is time to look at the production of information in a way such that it is valuable to the consumers. Creativity is a coveted aspect of the information producer. Not only do we want our artists to be creative, but also our scientists, our journalists, and even people we talk to. Though there

[6] However, if we had it in the right format, we could place it at the beginning of our data stream, so as to simulate having encountered those data first.

are many good definitions in other places for what this actually means, we will provide our own, in terms of the quantities that we have spent so long formalizing.

Creativity is increasing the novelty of an existing string of information, by appending something new to the end. It is often easy to append something to the end of a string which does not change its novelty, particularly if it's already uniformly low: you need merely to follow the pattern that has already been laid out by the previous data in the string.

It is also significant that we have not restricted ourselves to a length of the "something new". This is not only immediately necessary in order to be able to make any impact on larger scales of complexity, but also a key property that allows creativity to flourish. Because we have a finite alphabet, novelty-focused producers soon exhaust their supply of novel one and two-letter ideas, and quickly move on to bigger things, recycling these smaller ideas as pieces of a bigger idea. This shows us something about a creative process: even though all of the good ideas and possibilities have been exhausted, you can always move to a new scale and positively impact novelty in a bigger or smaller place than previous producers have tried to explore – and while the permutations on a given scale are numbered, the infinitude of scales makes up for it.

Conclusion

Not only do live in a world riddled with information, but we live in a world where information is immensely popular. Producing good information gets you food, and sometimes mates. Consuming information is what people do for pleasure, for work, and practically for everything else. It is clear that an understanding of how people deal with information could be extremely useful.

This construction has applications in advertising and art distribution, where companies can learn their subscribers' interest curves and thereby filter through data to find the ones that customers are most likely to enjoy. The concept of novelty has applications in personalized lossy compression, where object could be compressed in such a way that information that consumers don't care about and that is masked away by their preference functions is the information that is dropped.

It is a very effective way of describing different scales of musical structure, and matching similarly constructed pieces. It effectively describes and breaks symmetries between political ideologies, by showing where people want to place their governmental complexity on a complexity-scale curve. It is a tool for analyzing the effectiveness of organizational command hierarchy, and is extremely applicable in describing fashion and artistic movements.

In the end, having an understanding of what people look for in their information is both extremely useful and explanatorily powerful.