Transactions on Engineering and Computing Sciences - Vol. 11, No. 2
Publication Date: April 25, 2023
DOI:10.14738/tecs.112.14499.
Diamant, E. (2023). Data-Driven Bioinformatics is a Popular but Wrong and Misleading Attempt to Assess the Information-Handling Abilities of Natural Biological Systems. Transactions on Engineering and Computing Sciences, 11(2). 90-92.

SERVICES FOR SCIENCE AND EDUCATION

Data-Driven Bioinformatics is a Popular but Wrong and Misleading Attempt to Assess the Information-Handling Abilities of Natural Biological Systems

Emanuel Diamant
Independent Research Engineer
Kiryat Ono, Israel

It is generally agreed and accepted that Information is an indispensable part of our modern everyday life. But that does not mean that a generally agreed and accepted understanding of what information is or may be does exist in the general public or in the specific scientific communities. Norbert Wiener's dictum "Information is information, not matter or energy" is not a definition, but expresses pretty well the existing ambiguity and embarrassment about the discoursed matters.

The grounding fathers of the Theory of Information (Shannon, Kolmogorov, Fisher, Chaitin, Renyi) also keep silent about the issue, devoting themselves only to the "information measure" problem. The newly invented notion of "information measure" has served very well the purposes of a reliable message exchange over a communication channel. But in all other cases, it led to a long-lasting improper mixing and merging between the notion of "information" and the notion of "information measure", which, in turn, made the relations between the notion of "information" and notions of "data", "knowledge", and "semantics", blurred, intuitive and undefined.

In such circumstances, I have decided to develop my own definition of Information and the consequent corollaries that come out from this definition. The adventure has taken some time, but finally, I have published on several occasions (almost over a span of 15 years) my results, which today sound as such:

"Information is a linguistic description of structures observable in a given data set".

I would spend the rest of my time discussing some of the consequences of my definition that deserve (as it seems to me) more attention and concern.

So, in a data set, the data elements are not distributed randomly, but due to the similarity of their physical parameters, are naturally grouped into some kind of clusters or bundles. (In an electronic image these could be groups of pixels of the same color or brightness, for example). I propose to call these clusters **primary or physical data structures**.

In the eyes of an external observer, these primary data structures are organized into larger and more complex agglomerations, which I propose to call **secondary data structures**.

These secondary structures reflect the observer's view on the grouping of the primary data structures, and so they can be called **meaningful or semantic data structures**.

While the formation of primary (physical) data structures is determined by the objective (natural, physical) properties of the data, the subsequent formation of secondary (semantic) data structures is a subjective process governed by the conventions and habits of the observer (or a mutual agreement in an observers' group).

As it was said, **the description of the structures observed in the data set should be called "Information".** In this regard, it is necessary to distinguish between two types of information - **physical information and semantic information**.

Both are language descriptions; however, physical information can be described using a variety of languages (recall that mathematics is also a language), and semantic information can be described only using the observer's natural language.

It must be said that, as it follows from the above-mentioned definition of information, physical and semantic pieces of information form a single composite notion of information. Not a dual notion, where both forms coexist in a superposition of quantum states, but a composition of two sequential states with different levels of structural organization (where physical information is at the lowest level of this organization).

Information processing is carried out in a hierarchical fashion, where the semantic information of a lower level is transferred to the next higher level, where it becomes part of a structure of a higher complexity. The process of such a new arrangement is carried out according to subjective rules retained in a prototypical (referential) structure stored in the observer's (system's) memory.

Thus, in an information processing system, memory is the conserved information retained from a previous attempt of semantic information processing.

An important consequence of the statement that "**Information is a linguistic description**" immediately implies that information descriptions always materialize as a set of words, a fragment of text, or a narrative. In this regard, an important note should be made - these text sequences are written with nucleotide letters and amino acid signs. That turns the biological information into a physical entity, into a "thing", with its weight, length, and other physical properties. For the purposes of our discussion, this is an extremely important remark, because when we talk about information processing in biological systems we must take into account the physical attributes of the processed information packages, which means a persistent demand for a physical space for information flow arrangement, cell local spaces for information (memory) deposition and handling, and so on space demanding requirements.

What follows from this is: **Information processing is text processing!** Not data processing, to which we all are accustomed, but an unknown yet text processing paradigm. Lotfi Zadeh has proposed to call it **Computing with Words**. But in the past 50 years, it has not taken off.

By clarifying and disentangling the physical and semantic information processing paths we immediately place in the right light the true abilities of new sciences and technologies developed to extract information and knowledge from huge data volumes produced by contemporary biological sciences and their experiments. All of them are being declared as data-driven undertakings. In light of our findings, that means that only physical information can be derived as the result of their efforts. Not the desired semantic information, which (as you now know) is the real goal of contemporary information handling and processing.

Only being equipped with the new knowledge about "What is information" we can now take part in deciphering and explaining real biological experiments and trials – for example, experiments of transferring memory from one group of songbirds or smiles to another group. Such mechanical memory transfer from one group of living beings to another group can be accomplished only if the memory is a physical thing, a material entity, and not an imagined engram of spiking neurons, as it is usually believed today.

Only being equipped with the new knowledge about "What is information" we can now understand the attempts to create artificial memory storage units similar to biological storage units with an unprecedented storage capacity density – information volume vs involved nucleotides' physical volume.

Brain information processing, brain cognitive disorders, and brain inflammations can be understood and successfully treated also only if information processing modeling rules inspired by our latest findings would be brought to the possession and knowledge of the neurologists' communities. But that is not in the scope of this conference.

References

[1] Paulien Hogeweg, The Roots of Bioinformatics in Theoretical Biology, PLoS Computational Biology, https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1002021

[2] Emanuel Diamant, Advances in Bioinformatics and Computational Biology: Don't take them too seriously anyway, https://arxiv.org/abs/1505.04785

[3] Emanuel Diamant, Does Wu's Philosophy Define What Is Information? *Proceedings* 2017, *1*(3), 88; http://www.mdpi.com/2504-3900/1/3/88

[4] Emanuel Diamant, Shannon's Definition of Information is Obsolete and Inadequate. it is Time to Embrace Kolmogorov's Insights on the Matter, https://vixra.org/abs/1712.0494

[5] Emanuel Diamant, The brain is processing information, not data: Does anybody knows about that? https://vixra.org/abs/1712.0493

[6] Emanuel Diamant, Unveiling the mystery of visual information processing in human brain, Brain Research 1225(10):171-178, August 2008, https://www.researchgate.net/publication/222567891

[7] Diamant, E. Designing Artificial Cognitive Architectures: Brain Inspired or Biologically Inspired? https://www.researchgate.net/publication/329582475