On the Foundations and Philosophy of Info-Metrics

Amos Golan

Info-Metrics Institute and Department of Economics, American University, 4400 Massachusetts Avenue, Washington, DC 20016, United States of America
agolan@american.edu

1 Background, Objective and Motivation

Among the set of open questions in philosophy of information posed by Floridi [5,6] is the question of "What Is the Dynamics of Information?" For recent discussion see Crnkovic and Hofkirchner [2]) and a complimentary summary of open questions in the interconnection of philosophy of information and computation (Adriaans [1]). The broad definition of "dynamics of information" includes the concept of "information processing." This article concentrates on that concept, redefines it as "info-metrics," and discusses some open questions within info-metrics.

Inference and processing of limited information is one of the most fascinating universal problems. We live in the information age. Information is all around us. But be it much or little information, perfect or blurry, complementary or contradicting, the main task is always how to process this information such that the inference – derivation of conclusions from given information or premises – is optimal.

The emerging field of info-metrics is the science and art of inference and quantitatively processing information. It crosses the boundaries of all sciences and provides a mathematical and philosophical foundation for inference with finite, noisy or incomplete information. Info-metrics is at the intersection of information theory, statistical methods of inference, applied mathematics, statistics and econometrics, complexity theory, decision analysis, modeling and the philosophy of science. From mystery solving to the formulation of all theories – we must infer with limited and blurry observable information. The study of info-metrics helps in resolving a major challenge for all scientists and all decision makers of how to reason under conditions of incomplete information. Though optimal inference and efficient information processing are at the heart of info-metrics, these issues cannot be developed and studied without understanding information, entropy, statistical inference, probability theory, information and complexity theory as well as the meaning and value of information, data analysis and other related concepts from across the sciences. Info-metrics is based on the notions of information, probabilities and relative entropy. It provides a unified framework for reasoning under conditions of incomplete information.

Though much progress has been made, there are still many deep philosophical and conceptual open questions in info-metrics: What is a correct inference

S.B. Cooper, A. Dawar, and B. Löwe (Eds.): CiE 2012, LNCS 7318, pp. 238–245, 2012.

method? How should a new theory be developed? Is a unified approach to inference, learning and modeling necessary? If so, does info-metrics provide that unified framework? Or even simpler questions related to real data analyses and correct processing of different types and sizes of blurry data fall at the heart of info-metrics. Simply stated, modeling flaws and inference with imperfect information is a major challenge. Inconsistencies between theories and empirical predictions are observed across all scientific disciplines. Info-metrics deals with the study of that challenge. These issues – the fundamental problem, current state, thoughts on a framework for potential solutions and open questions – are the focus of that article.

The answer to the above questions demands a better understanding of both information and information processing. That includes understanding the types of information observed, connecting it to the fundamental – often unobserved – entities of interest and then meshing it all together and processing it in a consistent way that yields the best theories and the best predictions. Info-metrics provides that mathematical and philosophical framework. It generalizes the Maximum Entropy (ME) principle (Jaynes [10,11]) which builds on the principle of insufficient reason. The ME principle states that in any inference problem, the probabilities should be assigned by maximizing Shannon's information measure called entropy (Shannon [13]) subject to all available information. Under this principle only the relevant information is used. All information enters as constraints in an optimization process: constraints on the probability distribution representing our state of uncertainty. Maximizing the entropy subject to no constraints yields the uniform distribution representing a state of complete uncertainty. Introducing meaningful information as constraints in the optimization takes the distribution away from uniformity. The more information there is, the further away the resulting distribution is from uniformity or from a state of complete uncertainty. (For detailed discussion on the philosophy of information, as well as dynamic information, across the sciences see van Benthem, and Adriaans [4]; the resent text of van Benthem [3]; the proceedings of the Info-Metrics Workshop on the Philosophy of Information [12].)

In this article, I will provide a summary of the state of info-metrics, the universality of the problem and a solution framework. I will discuss some of the open questions and provide a number of examples from across the scientific spectrum throughout this article.

2 Examples and Framework

Much has been written on information. Much has been written on inference and prediction. But the interdisciplinary combined study of information and efficient information processing is just starting. It is based on very sound foundations and parts of it have long history within specific disciplines. It provides the needed link to connect all sciences and decision processes. To demonstrate the basic issues in info-metrics and the generality of the problem, consider the following very simple examples.

Consider "betting" on the state of the economy as is conveyed by the data. This is expressed nicely in a recent Washington Post editorial (August 9, 2011): "Top officials of the Federal Reserve, and their staff, assembled around a gigantic conference table to decide what, if anything, they should do to help the flagging U.S. economy. Just about every member of this group, headed by Chairman Ben S. Bernanke, was an expert in economics, banking, finance or law. Each had access to the same data. And yet, after hours of discussion, they could not agree...." What would you do? How would you interpret the data? Would you follow the Chairman or would you construct your own analysis? There is not enough information to ensure one solution. There are many stories that are fully consistent with the available information and with the data in front of each one of the Fed members. The Chairman's solution is only one such story. The Fed members face blurry and imperfect data. Simply phrased, the question is what is the solution to X+Y = more-or-less 10. Unlike the previous case, even if you have more information (say, X = more-or-less 3), there are still infinitely many solutions to this two blurry equations problem. Using information theory and the tools of info-metrics yields the most conservative solution.

Now, rather than "betting" on the story the "data" represent, consider a judge "betting" on the District Attorney's story based on the observable evidence, or a Justice "betting" on the correct interpretation of the constitution as is nicely expressed by Justice Souter's remarks at the 2011 Harvard Commencement: "...The reasons that constitutional judging is not a mere combination of fair reading and simple facts extend way beyond the recognition that constitutions have to have a lot of general... . Another reason is that the Constitution contains values that may well exist in tension with each other, not in harmony. Yet another reason is that the facts that determine whether a constitutional provision applies may be very different from facts like a person's age or the amount of the grocery bill; constitutional facts may require judges to understand the meaning that the facts may bear before the judges can figure out what to make of them. And this can be tricky..." The justices are faced with the same evidence (information). But in addition to the hard evidence, the justices incorporate in their decision process other information such as their own values, their own subjective information, and their own interpretation of the written word. They end up disagreeing even though they face the same hard evidence. Is there a trivial way of solving it? No. But, there is a way to consistently incorporate the hard evidence and the softer information such as prior beliefs, value judgment, and imprecise meanings of words within the information-theoretic and info-metrics framework mentioned earlier. Building on these examples, I discuss the universal problem of information processing and a framework for inference with the available information.

3 Universality

Generally speaking, what is common to decisions made by Supreme Court justices, communication among individuals, scientific theories, literary reviews, art critiques, image reconstruction, data analyses, and decision making? In each case, inference is made and information is processed. This information may be complete or incomplete, perfect or blurry, objective or subjective at the decision time or at the moment of analysis. The justices are faced with the same evidence (information). But in addition to the evidence, each justice decides the case using her/his own values and her/his own subjective information and interpretation of the written word. They end up disagreeing even though they face the same hard evidence. When communicating with one another, individuals use their own interpretation and understanding of each word, phrase, gesture and facial expression. Again, the subjective information is incorporated with the hard evidence – the words. Scientific theories are based on observed information (and observable data) together with sets of axioms and assumptions. Some of these axioms and assumptions are unobservable and cannot be verified. They reflect the very basic beliefs together with the minimally needed information the scientist needs to deduce and infer a new theory. They are the fundamentals necessary for building a theory but they are not always observed. Again, the subjective information (beliefs, assumptions, axioms) is incorporated with the hard evidence – the observed data. Literary and art reviewers are also faced with the same evidence: the painting, the text, the dance, the poem. But, in addition to evidence, they evaluate it based on their own subjective information. Image and data processing involve observable, unobservable and subjective information. Since all data processing involve noisy information, it is impossible to do inference without incorporating subjective and other structural information. To complicate things further, some of the observed or unobserved information may incorporate contradicting pieces of information making it practically impossible to reconcile it within a single, unified theory. Even worse: if the different pieces of information are complementary, how much value should each one receive? The common thread across these examples is the need for coherent and efficient information processing and inference. The problem is universal across all disciplines and across all decision makers.

4 The Starting Premise – The Observer

If one views the scientist as an observer (not a common view within the social sciences) then the universal problem of inferring with limited information can be resolved. That solution framework serves as a common thread across all sciences. The idea is very simple. Regardless of the system or question studied, the researcher observes only a certain amount of information or evidence. The observer needs to process that information in a scientifically honest way. That means that she has to process the observed information in conjunction with the unobserved information coming from theoretical considerations, intuition or subjective beliefs. That processing should be done within an optimal inference framework while taking into account the relationship between the observable and the unobservable; the relationship of the fundamental (often unobserved) entity of interest and its relationship to the observable quantities; the connection between the micro state (usually unobserved) and the global macro state of the system (usually observed); the value of the observed information; specification of the theoretical and other subjective information; and finally even if one has all the answers and inference is made, it needs to be verified.

As an observer the scientist and the decision maker become detectives that must combine all the types of information, process it all and make their inference. When possible they validate their theory or decision and update their inference accordingly. With that view we are back at the universal problem of inference with limited information. But now we already acknowledge the advantage of tackling that challenge as external observers. Since we deal with information, it seems natural that the study of this problem takes us back to information theory and the earlier work of Jaynes on the Maximum Entropy formalism.

5 Open Questions

Keeping in mind that all information is finite and that the observed information is often very limited and blurry, all inferential problems are inherently underdetermined (as discussed above). Further, the information observed is of different types: "hard" data, "soft" data and prior information. The "hard" data are just the traditional observed quantities. The "soft" data represent our own subjective beliefs and interpretation as well as axioms, assumptions and possible unverifiable theories. The prior information represents our prior understanding of the process being studied (often it is based on some verifiable data or on information coming from previous studies or experiments).

I will start with a few examples of small and large data and then I will provide a short list of open questions.

A number of simple representative examples capture some of the basic problems raised so far. Consider for example analyzing a game between two – or more – individuals. Traditionally, one starts by assuming the players are rational or maximize a certain objective function. Other structure (on the information set each one of the players has) is also assumed. But in reality the researcher does not observe the individual's preferences (or objectives) but rather observes the actions taken by these players. But similar to previous examples, this means that the problem is inherently underdetermined: there are many stories/games that are consistent with the observed actions. The info-metrics framework described here allows one to analyze such problems. Bono and Golan (2011) formulate that problem. They formulate a solution concept without making assumptions about expected utility maximization, common knowledge or beliefs. Beliefs, strategies and the degree to which players maximize expected utility are endogenously determined as part of the solution. To achieve this, rather than solving the game from the players' point of view, they analyze the game as an "observer" who is not engaged in the process of the game but observes the actions taken by the players. They use an information-theoretic approach which is based on the Maximum Entropy principle. They also compare their solution concept with Bayesian Nash equilibrium and provide a way to test and compare different models and different modeling assumptions. They show that alternative uses of the observer's information lead to alternative interpretations of rationality. These alternative interpretations of rationality may prove useful in the context of ex post arbitration or the interpretation of experimental data because they indicate who is motivating whom.

This example brings out some of the fundamental questions a researcher has to deal with when constructing a theory and processing information in the behavioral or social sciences. For example, how can one connect the unobserved preferences (and rationality) with the observed actions? Or, how much observed information one needs in order to solve such a problem? Or what inference method is the "correct" method to use? Or how can the theory be validated with the observed information? All of these questions arise naturally when one deals with small and noisy data within the social sciences.

Consider now a different example dealing with decomposing mass spectra of gas mixtures from noisy measurements (Toussaint, Golan and Dose [14]). Given noisy observations the objective here is estimate the unknown cracking patterns and the concentrations of the contributing molecules. Again, unless more structure is imposed, the problem is underdetermined. The authors present an information-theoretic inversion method called generalized maximum entropy (GME) for decomposing mass spectra of gas mixtures from noisy measurements. In this GME approach to the noisy, underdetermined inverse problem, the joint entropies of concentration, cracking, and noise probabilities are maximized subject to the measured data. This provides a robust estimation for the unknown cracking patterns and the concentrations of the contributing molecules. The method is applied to mass spectroscopic data of hydrocarbons, and the estimates are compared with those received from a Bayesian approach. The authors also show that the GME method is efficient and is computationally fast. Like the previous example, an information-theoretic constrained optimization framework is developed. Here as well some fundamental questions arise. For example, how should one handle the noise if the exact underlying distribution is unknown? How can one connect the observed noisy data with the basic entities of interest (concentration and cracking in this case)?

Similar inferential problems exist also with big data. One trivial example is image reconstruction or a balancing of a very large matrix. The first problem is how to reduce the data (or the dimensionality of the problem) so the reconstruction will be computationally efficient. Within the information-theoretic constrained optimization (or inverse) framework discussed here one can solve that problem as well. Generally speaking, the inverse problem is transformed into a generalized moment problem, which is then solved by an information theoretic method. This estimation approach is robust for a whole class of distributions and allows the use of prior information. The resulting method builds on the foundations of information-theoretic methods, uses minimal distributional assumptions, performs well and uses efficiently all the available information (hard and soft data). This method is computationally efficient. For more image reconstruction examples see Golan, Bhati and Buyuksahin [8]. For other examples within the natural sciences see Golan and Volker [9]. For more derivations and examples (small and large data, mostly within the social sciences) see Golan [7].

Below I provide a partial list of the fundamental open questions in info-metrics. Naturally, some of these questions are not independent of one another, but it is helpful to include each one of these questions separately.

1. What is information?
2. What information do we observe?
3. How can we connect the observed information to the basic unit (or entity) of interest?
4. How should we process the information we observe while connecting it to the basic unit of interest?
5. Can we quantify all types of information?
6. How can we handle contradicting evidence (or information)?
7. How can we handle complimentary evidence (or information)?
8. Can we define a concept of "useful" information?
9. Is there a way to assign value to information? If so is it an absolute or a relative value?
10. How is the macro level information connected to the basic micro level information?
11. How to do inference with finite information?
12. Is there a way for modelling (and developing theories) with finite/limited information?
13. How can we validate our theories?
14. Is there a unique relationship between information and complexity?
15. What is a correct inference method? Is it universal to all inferential problems? (What are the mathematical foundations of that method?)
16. Is the same information processing and inference framework applies across all sciences?
17. Is a unified approach to inference and learning necessary and useful?

The above list is not complete but it provides a window toward some pressing issues within info-metrics that need more research. A more detailed discussion and potential answers to some of these questions is part of a current and future research agenda.

6 Summary

In this paper, I summarized some of the fundamental and philosophical principles of info-metrics. In addition to presenting the basic ideas of info-metrics (including its precise definition), I demonstrated some of the basic problems via examples taken from across the disciplines. A short discussion of the different types of information is provided as well. These include the "hard" information (data), "soft" information (possible theories, intuition, beliefs, conjectures, etc), and priors. A generic framework for an information-theoretic inference is discussed. The basic premise is that since all information is finite/limited (or since we need to process information in finite time) we can construct all information processing as a general constrained optimization problem (information-theoretic inversion procedure). This framework builds on Shannon's entropy (Shannon [13]), on Bernoulli's principle of insufficient reason (published eight years after his death in 1713) and on Jaynes principle of maximum Entropy (Jaynes [10,11]) and further generalizations (e.g., Golan [7]).

The paper concludes with a partial list of open questions in info-metrics. These questions are related directly to quantitative information processing and inference and to the meaning of information. In future research these questions will be tackled one at a time.

References

1. Adriaans, P.: Some open problems in the study of information and computation (2011), `http://staff.science.uva.nl/~pietera/open_problems.html`
2. Crnkovic, G.D., Hofkirchner, W.: Floridi's: Open Problems in Philosophy of Information. Ten Years Later. Information 2, 327–359 (2011)
3. van Benthem, J.: Logical Dynamics of Information and Interaction, pp. 1–384. Cambridge University Press, Cambridge (2011)
4. van Benthem, J., Adriaans, P.: Philosophy of Information. North Holland, Amsterdam (2008)
5. Floridi, L.: Open problems in the Philosophy of Information. Metaphilosophy 35, 554–582 (2004)
6. Floridi, L.: The Philosophy of Information, pp. 1–432. Oxford University Press, Oxford (2011)
7. Golan, A.: Information and Entropy Econometrics – A Review and Synthesis. Foundations and Trends in Econometric 2(1-2), 1–145 (2008); Also appeared as a book - paperback
8. Golan, Bhati, A., Buyuksahin, B.: An Information-Theoretic Approach for Image Reconstruction: The Black and White Case. In: Knuth, K. (ed.) 25th International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering, MAXENT (2005)
9. Golan, A., Volker, D.: A Generalized Information Theoretical Approach to Tomographic Reconstruction. J. of Physics A: Mathematical and General, 1271-1283 (2001)
10. Jaynes, E.T.: Information Theory and Statistical Mechanics. Physics Review 106, 620–630 (1957a)
11. Jaynes, E.T.: Information Theory and Statistical Mechanics II. Physics Review 108, 171–190 (1957b)
12. Proceedings of Info-Metrics Institute Workshop on the Philosophy of Information. American University, Washington, DC (2011)
13. Shannon, C.E.: A Mathematical Theory of Communication. Bell System Technical Journal 27, 379–423 (1948)
14. Toussaint, U.V., Golan, A., Dose, V.: Maximum Entropy Decomposition of Quadruple Mass Spectra. Journal of Vacuum Science and Technology A 22(2), 401–406 (2004)