Information and Computation

Carlos Gershenson

Introduction

Before delving into the role of information theory as a descriptive tool for human computation (von Ahn 2009), we have to agree on at least two things: what is human, and what is computation, as human computation is at its most general level computation performed by humans. It might be difficult to define what makes us human, but for practical purposes we can take an "I-know-it-when-I-see-it" stance. For computation, on the other hand, there are formal definitions, tools and methods that have been useful in the development of digital computers and can also be useful in the study of human computation.

Information

Information has had a long and interesting history (Gleick 2011). It was Claude Shannon (1948) who developed mathematically the basis of what we now know as *information theory* (Ash 1990). Shannon was interested in particular on how a message could be transmitted reliably across a noisy channel. This is very relevant for telecommunications. Still, information theory has proven to be useful beyond engineering (von Baeyer 2005), as anything can be described in terms of information (Gershenson 2012).

C. Gershenson (✉)

Departamento de Ciencias de la Computación, Centro de Ciencias de la Complejidad,
Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional
Autónoma de México, A.P. 20-726, 01000 México, D.F. México
e-mail: cgg@unam.mx

A brief technical introduction to Shannon information H is given in Appendix A. The main idea behind this measure is that messages will carry more information if they reduce uncertainty. Thus, if some data is very regular, i.e. already certain, more data will bring few or no new information, so H will be low, i.e. few or no new information. If data is irregular or close to random, then more data will be informative and H will be high, since this new data could not have been expected from previous data.

Shannon information assumes that the meaning or decoding is fixed, and this is generally so for information theory. The study of meaning has been made by semiotics (Peirce 1991; Eco 1979). The study of the evolution of language (Christiansen and Kirby 2003) has also dealt with how meaning is acquired by natural or artificial systems (Steels 1997).

Information theory can be useful for different aspects of human computation. It can be used to measure, among other properties: the information transmitted between people, novelty, dependence, and complexity (Prokopenko et al. 2009; Gershenson and Fernández 2012). For a deeper treatment of information theory, the reader is referred to the textbook by Cover and Thomas (2006).

Computation

Having a most general view, computation can be seen simply as the transformation of information (Gershenson 2012). If anything can be described in terms of information, then anything humans do could be said to be human computation. However, this notion is too broad to be useful.

A formal definition of computation was proposed by Alan Turing (1936). He defined an abstract "machine" (a Turing machine) and defined "computable functions" as those which the machine could calculate in finite time. This notion is perhaps too narrow to be useful, as Turing machines are cumbersome to program and it is actually debated whether Turing machines can model all human behavior (Edmonds and Gershenson 2012).

An intermediate and more practical notion of computation is *the transformation of information by means of an algorithm or program.* This notion on the one hand tractable, and on the other hand is not limited to abstract machines.

In this view of computation, the algorithm or program (which can be run by a machine or animal) defines rules by which information will change. By studying at a general level what happens when the information introduced to a program (input) is changed, or how the computation (output) changes when the program is modified (for the same input), different types of dynamics of information can be identified:

- Static. Information is not transformed. For example, a crystal has a pattern which does not change in observable time.
- Periodic. Information is transformed following a regular pattern. For example, planets have regular cycles which in which information measured is repeated every period.

- Chaotic. Information is very sensitive to changes to itself or the program, it is difficult to find patterns. For example, small changes in temperature or pressure can lead to very different meteorological futures, a fact which limits the precision of weather prediction.
- Complex. Also called critical, it is regular enough to preserve information but allows enough flexibility to make changes. It balances robustness and adaptability (Langton 1990). Living systems would fall in this category.

Wolfram (2002) conjectured that there are only two types of computation: universal or regular. In other words, programs are either able to perform any possible computation (universal), or they are simple and limited (regular). This is still an open question and the theory of computation is an active research area.

Computing Networks

Computing networks (CNs) are a formalism proposed to compare different types of computing structures (Gershenson 2010). CNs will be used to compare neural computation (information transformed by neurons), machine distributed computation (information transformed by networked computers), and human computation.

In computing networks, nodes can process information (compute) and exchange information through their edges, each of which connects the output of node with the input of another node. A computing network is defined as **a set of nodes N linked by a set of edges K used by an algorithm a to compute a function f** (Gershenson 2010). Nodes and edges can have internal variables that determine their state, and functions that determine how their state changes. CNs can be stochastic or deterministic, synchronous or asynchronous, discrete or continuous.

In a CN description of a **neural network** (NN) model, *nodes* represent neurons. Each neuron i has a continuous state (output) determined by a function y_i which is composed by two other functions: the weighted sum S_i of its inputs $\bar{x}_i$ and an activation function A_i, usually a sigmoid. Directed *edges ij* represent synapses, relating outputs y_i of neurons i to inputs x_j of neurons j, as well as external inputs and outputs with the network. Edges have a continuous state w_{ij} (weight) that relates the states of neurons. The *function f* may be given by the states of a subset of N (outputs $\bar{y}$), or by the complete set N. NNs usually have two dynamical scales: a "fast" scale where the network function f is calculated by the functional composition of the function y_i of each neuron i, and a "slow" scale where a learning *algorithm a* adjusts the weights w_{ij} (states) of edges. There is a broad diversity of algorithms a used to update weights in different types of NN. Figure 1 illustrates NNs as CNs.

Digital machines carrying out **distributed computation** (DC) can also be represented as CNs. *Nodes* represent computers while *edges* represent network connections between them. Each computer i has information H_i which is modified by a program $P_i(H_i)$. Physically, both H_i and P_i are stored in the computer memory, while the information transformation is carried out by a processor. Computers can share
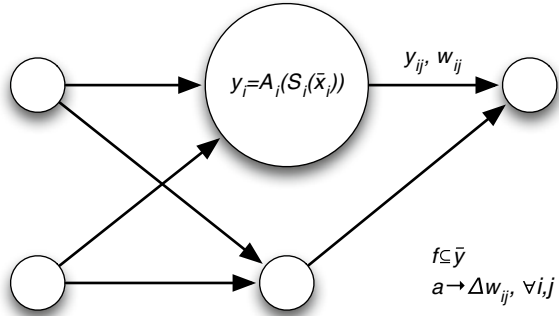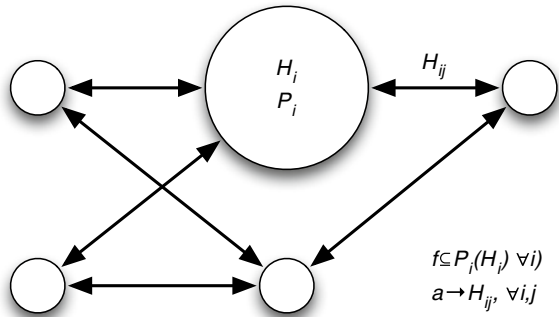
Fig. 1 A NN represented
as a CN

Fig. 2 A DC system or a HC
system represented as a CN

information H_{ij} across edges using a communication protocol. The *function f* of the
DC will be determined by the output of $P_i(H_i)$ of some or all of the nodes, which can
be seen as a "fast" scale. Usually there is an algorithm *a* working at a "slower" scale,
determining and modifying the interactions between computers, i.e. the network
topology. Figure 2 shows a diagram of DC as a CN.

 Human computation (HC) can be described as a CN in a very similar way than
DC. People are represented as *nodes* and their interactions as *edges*. People within
a HC system transform information H_i following a program $P_i(H_i)$. In many cases,
the information shared between people H_{ij} is transmitted using digital computers,
e.g. in social networks, wikis, forums, etc. In other cases, e.g. crowd dynamics,
information H_{ij} is shared through the environment: acustically, visually (Moussaïd
et al. 2011), stigmergically (Doyle and Marsh 2013), etc. The function *f* of a HC
system can be difficult to define, since in many cases the outcome is observed and
described only a posteriori. Still, we can say that *f* is a combination of the computa-
tion carried out by people. An algorithm *a* would determine how the social links
change in time. Depending on the system, *a* can be slower than *f* or vice versa.

 In DC, the algorithm *a* is centrally determined by a designer, while in most HC
systems, the *a* is determined and executed by people (nodes) themselves.

Using information theory, we can measure how much information H_{ij} is transmitted between people, how much each person receives and produces, and how much the entire system receives and produces. In many cases, machines enable this transmission and thus also facilitate its measurement. Comparing the history of information transfers and current information flows can be used to measure the novelty in current information.

Examples

Social Networks

A straightforward example of human computation can be given with online social networks. There are key differences, e.g. links are bidirectional in Facebook (my friends also have me as their friend) and unidirectional in Twitter (the people I follow do not necessarily follow me, I do not necessarily follow my followers). People and organizations are represented with their accounts in the system as nodes, and they receive information through their incoming links, They can share this information with their outgoing links and also produce novel information that their links may receive. People can decide how to create or eliminate social links, i.e. *a* is decided by individuals.

These simple rules of the information dynamics on social networks are able to produce very interesting features of human computation (Lerman and Ghosh 2010), which can be described as functions *f*. For example, non-official news can spread very quickly through social networks, challenging mass media dominated by some governments. On the other hand, false rumors can also spread very quickly, potentially leading to collective misbelief. Nevertheless, it has been found that the dynamics of false rumors spreading is different from that of verifiable information (Castillo et al. 2011).

Describing social networks as CNs is useful because interactions are stated explicitly. Moreover, one can relate different scales with the same model: local scale (nodes), global scale (networks), and meso scales (modules); and also temporal scales: fast (*f*) and slow (*a*). Information theory can be used to detect novelty in social interactions (high *H* values in edges), imitation (low *H* values in edges), unusual patterns ("fake" information), correlations (with mutual information), and communities (modules (Newman 2010)).

Wikipedia

Wikipedia gives a clear example of the power of human computation. Millions of people (nodes) from all over the world have collaboratively built the most extensive encyclopedia ever. The sharing of information is made through editable webpages on a specific topic. Since these pages can potentially link more than two people

(editing the webpage), the links can be represented as those of a hypernetwork (Johnson 2009), where edges can link more than two nodes (as in usual networks). The information in pages (hyperedges) can be measured, as it changes over time with the editing made by people linked to them. The information content delivered by different authors can be measured with H. When this is increased, it implies novelty. The complexity of the webpages, edits, and user interactions can also be measured, seen as a balance between maximum information (noise) and minimum information (stasis) (Fernández et al. 2013).

The function f of Wikipedia is its own creation, growth, and refinement: the pages themselves are the output of the system. Again, people decide which pages to edit, so the algorithm a is also decided by individuals.

Traditionally, Wikipedia—like any set of webpages—is described as a network of pages with directional edges from pages that link to other pages. This is a useful description to study the structure of Wikipedia itself, but it might not be the most appropriate in the context of human computation, as no humans are represented. Describing Wikipedia as a CN, the relationships between humans and the information they produce collaboratively is explicit, providing a better understanding of this collective phenomenon.

Conclusions

Concepts related to information and computation can be applied to any system, as anything can be described in terms of information (Gershenson 2012). Thus, HC can also benefit from the formalisms and descriptions related to information and computation.

CNs are general, so they can be used to describe and compare any HC system. For example, it is straightforward to represent online social networks such as Facebook, Twitter, LinkedIn, Google+, Instagram, etc. as CNs. As such, their structure, functions, and algorithms can be contrasted, and their local and global information dynamics can be measured. The properties of each of these online social networks could be compared with other HC systems, such as Wikipedia.

Moreover, CNs and Information Theory can be used to design and self-monitor HC systems (Gershenson 2007). For example, information overload should be avoided in HC systems. The formalisms presented in this chapter and in the cited material can be used to measure information inputs, transfers, and outputs to avoid not only information overload, but also information poverty (Bateson 1972).

In our age where data is overflowing, we require appropriate measures and tools to be able to make sense out of "big data". Information and computation provide some of these measures and tools. There are still several challenges and opportunities ahead, but what has been achieved so far is very promising and invites us to continue exploring appropriate descriptions of HC systems.

Appendix

Shannon Information

Given a string *X*, composed by a sequence of values *x* which follow a probability distribution *P*(*x*), information (according to Shannon) is defined as:

$$H = -\sum P(x) \log P(x). \tag{1}$$

For binary strings, the most commonly used in ICT systems, the logarithm is usually taken with base two. For example, if the probability of receiving ones is maximal ($P(1) = 1$) and the probability of receiving zeros is minimal ($P(0) = 0$), the information is minimal, i.e. $H = 0$, since we know beforehand that the future value of *x* will be 1. Information is zero because future values of *x* do not add anything new, i.e. the values are known beforehand. If we have no knowledge about the future value of *x*, as with a fair coin toss, then $P(0) = P(1) = 0.5$. In this case, information will be maximal, i.e. $H = 1$, because a future observation will give us all the relevant information, which is also independent of previous values. Equation 1 is plotted in Fig. 3. Shannon information can be seen also as a measure of uncertainty. If there is absolute certainty about the future of *x*, be it zero ($P(0) = 1$) or one ($P(1) = 1$), then the information received will be zero. If there is no certainty due to the probability distribution ($P(0) = P(1) = 0.5$), then the information received will be maximal. Shannon used the letter *H* because equation 1 is equivalent to Boltzmann's entropy in thermodynamics, which is also defined as *H*. The unit of information is the bit. One bit represents the information gained when a binary random variable becomes known.
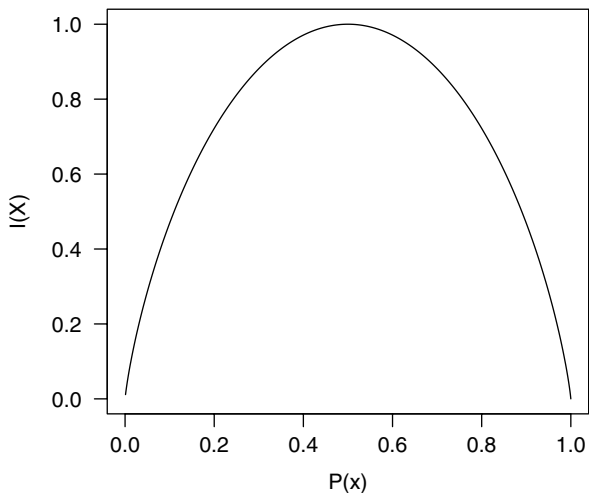

Fig. 3 Shannon's information *H*(*X*) of a binary string *X* for different probabilities *P*(*x*). Note that $P(0) = 1 - P(1)$

A more detailed explanation of information theory, as well as measures of complexity, emergence, self-organization, homeostasis, and autopoiesis based on information theory can be found in Fernández et al. (2013).

References

Ash RB (1990) Information theory. Dover, New York

Bateson G (1972) Steps to an ecology of mind. Ballantine, New York

Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on world wide web. WWW '11. ACM, New York, pp 675–684. http://doi.acm.org/10.1145/1963405.1963500

Christiansen MH, Kirby S (2003) Language evolution, vol 3. Oxford University Press, Oxford/New York

Cover TM, Thomas JA (2006) Elements of information theory. Wiley-Interscience. http://www.elementsofinformationtheory.com/

Doyle MJ, Marsh L (2013) Stigmergy 3.0: from ants to economies. Cogn Syst Res 21:1–6. http://dx.doi.org/10.1016/j.cogsys.2012.06.001

Eco U (1979) A theory of semiotics. Indiana University Press, Bloomington

Edmonds B, Gershenson C (2012) Learning, social intelligence and the Turing test – why an "out-of-the-box" Turing machine will not pass the Turing test. In: How the world computes: Turing centenary conference and 8th conference on computability in Europe, CiE 2012, Cambridge, 18–23 June 2012. In: Cooper SB, Dawar A, Löwe B (Eds) Proceedings. Lecture notes in computer science, vol 7318/2012. Springer, Berlin/Heidelberg, pp 182–192. http://arxiv.org/abs/1203.3376

Fernández N, Maldonado C, Gershenson C (2013) Information measures of complexity, emergence, self-organization, homeostasis, and autopoiesis. In: Prokopenko M (Ed) Guided self-organization: inception. Springer (in press). http://arxiv.org/abs/1304.1842

Gershenson C (2007) Design and control of self-organizing systems. CopIt Arxives, Mexico. http://tinyurl.com/DCSOS2007, http://tinyurl.com/DCSOS2007

Gershenson C (2010) Computing networks: a general framework to contrast neural and swarm cognitions. Paladyn. J Behav Robot 1(2):147–153. http://dx.doi.org/10.2478/s13230-010-0015-z

Gershenson C (2012) The world as evolving information. In: Minai A, Braha D, Bar-Yam Y (Eds) Unifying themes in complex systems, vol VII. Springer, Berlin/Heidelberg, pp 100–115. http://arxiv.org/abs/0704.0304

Gershenson C, Fernández N (2012) Complexity and information: measuring emergence, self-organization, and homeostasis at multiple scales. Complexity 18(2):29–44. http://dx.doi.org/10.1002/cplx.21424

Gleick J (2011) The information: a history, a theory, a flood. Pantheon, New York

Johnson J (2009) Hypernetworks in the science of complex systems. Imperial College Press, London

Langton C (1990) Computation at the edge of chaos: phase transitions and emergent computation. Physica D 42:12–37

Lerman K, Ghosh R (2010) Information contagion: an empirical study of the spread of news on digg and Twitter social networks. In: Proceedings of 4th international conference on weblogs and social media (ICWSM), Washington, DC

Moussaïd M, Helbing D, Theraulaz G (2011) How simple rules determine pedestrian behavior and crowd disasters. PNAS 108(17):6884–6888. http://dx.doi.org/10.1073/pnas.1016507108

Newman M (2010) Networks: an introduction. Oxford University Press, Oxford

Peirce CS (1991) Peirce on signs: writings on semiotic by Charles Sanders Peirce. University of North Carolina Press, Chapel Hill

Prokopenko M, Boschetti F, Ryan AJ (2009) An information-theoretic primer on complexity, self-organisation and emergence. Complexity 15(1):11–28. http://dx.doi.org/10.1002/cplx.20249

Shannon CE (1948) A mathematical theory of communication. Bell Syst Tech J 27:379–423, 623–656. http://tinyurl.com/6qrcc

Steels L (1997) The synthetic modeling of language origins. Evol Commun 1(1):1–34

Turing AM (1936) On computable numbers, with an application to the entscheidungsproblem. Proc Lond Math Soc Ser 2 42:230–265. http://www.abelard.org/turpap2/tp2-ie.asp

von Ahn L (2009) Human computation. In: 46th ACM/IEEE design automation conference, 2009, DAC '09, San Francisco, pp 418–419

von Baeyer HC (2005) Information: the new language of science. Harvard University Press, Cambridge. http://www.hup.harvard.edu/catalog.php?isbn=9780674018570

Wolfram S (2002) A new kind of sciene. Wolfram Media. http://www.wolframscience.com/