WHAT IS INFORMATION?

Summary

The notion of information has so far been quantified mostly in statistical terms, giving rise to Shannon's information theory and the principles of digital data transmission. Modern communication systems involving complex, intelligent, and autonomous agents call for a new look at the measures of information, where context, semantics, and agent rationality are not entirely abstracted from. In this essay we propose a framework for measuring information inspired by the event-driven paradigm.

1. INTRODUCTION

In this brief note, deliberately nontechnical and vague at times, we muse on the notion of information in its generality, hoping to capture some of its essential aspects and provoke a discussion leading to a new definition of information that might be applied in contemporary science and engineering. We shall proceed inductively, giving examples from which hopefully a formal framework will arise.

Advances in information technology, the abundance of information systems and services, the much-trumpeted advent of information society, or even the Information Age (recently embodied in the realm of Web 2.0), almost obscure the fact that the common buzzword — information — remains undefined in precise terms. This stands in the way of a systematic study. Shannon's statistical information quantifies to what extent a recipient of data can reduce statistical uncertainty associated with its source by observing the output of a source-recipient channel. Hence, the semantics of data are irrelevant — as Shannon puts it in his 1948 paper: "These semantic aspects of communication are irrelevant to the engineering problem." The channel error rate, on the other hand, does matter: for example, with a 50% binary error rate, the amount of statistical information sent through the channel is zero. But it seems that the intuitive understanding of information cannot be formalized without bringing into the picture the timing of data (consider a notice stating the train departure is at 6 o'clock served a recipient after that time), the activity its recipients are to undertake (consider the same notice served a recipient just intending to take a nap), and the knowledge of the recipient's internal rules of conduct (consider a recipient of binary data at the output of a channel with a 50% error rate who, because he wrongly estimates the channel to be perfect, uses the received data bona fide).
The context of data cannot be abstracted from, either. Even at a 50% error rate some information may be recovered from the context (e.g., a math textbook transmitted over such a channel might still be recognized as such). This point becomes particularly valid in the realm of biosystems – most biological information depends on where it is retrieved e.g., its location within a cell, a piece of DNA or protein. Unfortunately, this important aspect is sometimes left out of information theory, where a purely random sequence is known to contain the most statistical information. Biology is above all about context, and so a periodic pattern, while containing less statistical information than a random sequence, may contain a lot more biological information.

2. EVENT-DRIVEN PARADIGM

An intuitive relationship between data (any sequence of interpretable symbols) and information is that data may or may not carry information. One may observe that a piece of data carries information if it can impact a recipient's ability to achieve the objective of some rational activity within a given context. In fact, this observation, stated more or less explicitly, was the point of departure of some early textbooks on information technology. There seems to be little formal apparatus, however, to quantitatively account for all its constituent facets.

Thus information has a flavor of relativity (depends on the recipient's knowledge and capability) and a flavor of rationality (depends on the activity undertaken). Underlying the latter is also a flavor of timeliness, for the value of data may depend on its temporal structure and the timing of its generation and reception.

We offer more examples to illustrate a further subtlety. Clearly, a speaker of Chinese (a more knowledgeable recipient) can make out a lot more of a textbook on VLSI circuit design written in that language than a non-speaker (a less knowledgeable recipient). However, the latter can by default regard any string of symbols that does not look like an ethnic language as a blueprint of a complex VLSI circuit. Furthermore, a duplicate notice of a train departure time does not contribute to the objective of catching that train and therefore is of no informational value (the recipient already knows it), unless the recipient's internal rules of conduct require that at least one confirmation of the train departure time be received. Finally, decryption keys separated in time and space seem to carry zero information until they are brought together in one location at the same time. Obviously, information carried by data is related to its context, but also to a recipient's internal rules of conduct, as well as to where and how these rules are applied.

Having said this, we still need a quantitative definition, an analogue of Shannon's statistical information, retaining the flavors of relativity, rationality, and timeliness, and accounting for context and recipient's behavior. Can we attempt formal definitions of the amount of information and maximum transferable information – capacity – without a lengthy specification of the semantics of data? One possibility is to adopt an event-driven paradigm which we sketch below.

Event-driven paradigm offers a few advantages. First, it is well-established among the computer science community thanks to the work of Hoare and others in the field of operating systems and distributed algorithms. Second, it is discrete and timeless in nature, yet allows for dynamic characterization of systems evolving in continuous time. Finally, it is able to formalize such intuitions as causality and consistency of local views without specifying the semantics of involved events. At the same time, it generalizes the data-information relationship: now it is events that may or may not carry information; in
What is Information?

particular, an event may correspond to reception of a piece of data. The event-driven paradigm-inspired formalization goes along the following lines:

- Consider a universe populated by systems performing activities in order to achieve objectives (examples are living organisms, institutions, communities, software agents, Internet domains etc.).
- A system is characterized by its current state, expressible through a number of system variables (e.g., current memory content, life-critical characteristics of the body, traffic conditions or operational status of network links).
- Activities include observations of changes in system state; any observable change will be referred to as the occurrence of an event (e.g., clock tick, execution of a specific operation, reception of a message – a piece of data – from another system).
- Partial order on the set of events within a system may be defined as the order in which the events occur (simultaneous events not precluded); the set of events preceding event E is called the context of event E (this, too, is a widely accepted notion among the computer science community).
- Events may have attributes e.g., time of occurrence, as perceived by a system in question, and semantics (type of task being executed, type of a received message, prescription how the message is to be reacted to etc.), as defined by the system's rules of conduct i.e., the specification of the activities the system performs.

We would like to regard information as another (measurable) attribute of an event in a way reflecting our previous discussion. To this end, define an objective functional that maps a system's rules of conduct and the present context (set of events that have occurred so far) into any space with a well-defined order of points. The idea is that the rules of conduct R along with the present context C determine to what extent a given objective has been achieved, as expressed by objective(R, C). Before defining a possible measure of information we look at more examples from which we derive our approach.

Example 1. [Decimal Representation] Assume that one's objective is to learn the number pi and the rules of conduct include executing an algorithm to compute successive decimal digits using a finite-speed processor. Each computed digit is then regarded as an event and objective(R, C) is a real-valued function monotonically increasing and asymptotically stabilizing in C.

Example 2. [Train Departure] In our previous discussion, the event corresponding to the reception of a 6 o'clock departure notice may have different impact upon objective(R, C): for example, it increases it if and only if the following conjunction holds: (1) there is no such event in C or R requires confirmation of a previous notice, and (2) the clock ticks in C imply it is still possible to get to the station in time. The other examples given previously can easily be re-thought along similar lines.

Example 3. [Synergy of Data] In a secret sharing scheme, parts of a decryption key are dispersed among subsystems of a distributed system. The event corresponding to the reception of another part from a fellow subsystem does not improve the ability to decrypt a given cipher text unless all the other parts of the key are already in C. Likewise, an observed pixel of a digital image may increase a viewer's ability to understand the image depending on how many neighboring pixels have already been observed (this example
illustrates that the event-driven paradigm also covers apparently spatial, rather than
temporal, contexts — in general, there is no difficulty evaluating the objective functional as
long as events are processed locally or received from a distributed total order protocol).

Example 4. [Life Science, Multimedia over Packet Networks] The impulses exchanged
along nerves or processed within neural cells of a living organism critically depend on
timing e.g., a stimulus generated by a pain receptor is useless if it arrives too late to
administer a defensive gesture. Clearly, there must be a timing mechanism in neurons,
meaning that clock ticks should be part of the context sets arising in the modeling of
neurons. Similarly, clock ticks are relevant when judging the usefulness of successive
speech or video frames sent over a packet network. Since they share network resources
with unpredictable data traffic, the frames arrive at the destination irregularly, as quantified
by the delay jitter. Premature and overdue arrivals (events with too few or too many clock
ticks in the context) are equally unwelcome, but contribute to the perceived quality of
network service in a different way: the former have to be buffered before delivery and the
latter are typically discarded.

Example 5. [Path Discovery in MANETs] Each of a collection of mobile terminals
interconnected by a wireless transmission medium can physically communicate with the
(usually small) fraction of terminals currently within its transmission range. To accomplish
full connectivity, source-to-destination paths are to be discovered and set up if need be, on
the assumption that the terminals currently between the source and destination will act as
relays. The more time (clock ticks) is allowed for path discovery at a given terminal, the
more path discovery events will occur as the terminal moves around and encounters more
terminals within transmission range. This reflects upon the objective of sending a
maximum data flow to the destination, a thought at the core of the so-called time capacity
paradox. Our example is somewhat similar to the decimal representation of \( \pi \) except that,
somewhat counterintuitively, objective\((R, C)\) need not increase in \( C \): new paths are
discovered, but previously discovered ones may get disconnected and drop the data. To
make objective\((R, C)\) increase in \( C \), the rules of conduct \( R \) may be designed so as to store
the data at each relay pending discovery of an alternative path; in this way no part of the
data is ever dropped.

Example 6. [Herding Effects, Web 2.0, DNA] The conclusion of the previous example
suggests that objective\((R, C)\) increases in \( C \) provided that the rules of conduct \( R \) are
somehow "rational." Unfortunately, studies of so-called herding effects disprove that
intuition too: an individual contemplating an action behaves rationally by observing
successive decisions of other rational individuals (as shown by a Bayesian analysis), only to
find, and very quickly at that, that further observations bring no useful information as to the
benefits of the action. Perhaps, then, one can only assert that objective\((R, C)\) is
nondecreasing in \( C \) given rational \( R \)? There are examples that run counter even that
intuition. Imagine a user session with a Web search engine in which too much data, or the
presence of conflicting data, paralyze the user's ability to perform any useful activity; from
another perspective, a growing number of users contributing their ideas to a digital Web 2.0
community may at some point prevent a broad consensus on the issue being resolved.
Equally daunting is the well-known fact that the sheer amount of data contained in a
biological database (e.g., human genome) may blur patterns leading to the identification of
relevant human traits.
What is Information?

Example 7. [Cooperative and Noncooperative Settings] Consider now a system where the objective functionals defined at different subsystems are in conflict (e.g., the problem of Byzantine generals, DoS or selfish attacks on communication protocols). The simplest example are two data sources contending for a multiple access channel. The various forms of the rules of conduct R may then calibrate the sources' behavior from cooperative (where \( \text{objective}(R, C) \) increases in the total amount of data successfully sent over the channel i.e., in the overall channel utilization) to noncooperative (where \( \text{objective}(R, C) \) increases in the amount of own data sent, leading to a congestion game) to malicious (where \( \text{objective}(R, C) \) decreases in the amount of data sent by the other source).

Example 8. [Stochastic Complexity] Included in \( \text{objective}(R, C) \) may be the cost of the very recognition and interpretation of C. Imagine a recipient knowing that the source uses an optimal code for its stream of data, but having to learn on-the-fly the stochastic mechanism according to which the source generates data. As time passes, the model reveals itself to the recipient who can then recover the length of the optimal code, hence correctly interpret C.

Example 9. [Quantum Information and Communication] In some communication environments an attempt to interpret data leads to the distortion of the data itself. This is the case when the data is encoded in the polarization of photons, since the principles of quantum mechanics leave only a 50% chance of correctly copying a photon's polarization (vertical or diagonal). Even though complex protocols enable exchange of correct data between two parties, each event (arrival of another photon) may affect \( \text{objective}(R, C) \) either favorably or adversely, depending on pure chance. As Pauli notes, "... any attempt to measure [that] property destroys (at least partially) the influence of earlier knowledge of the system." Furthermore, the order in which experiments are performed may change information (as in the spin experiment).

3. INFORMATION AND CAPACITY

We are now in a position to give a framework for defining the amount of information in a way consistent with intuitions based on the previous examples and discussion.

**Definition:** The amount of information carried by the event E in the context C as perceived at a system with the rules of conduct R is

\[
\text{info}_{R,C}(E) = \text{distance}([\text{objective}(R, C), \text{objective}(R, C + E)],
\]

where the function \( \text{distance} \) measures the difference between two points according to the order defined on the space of objectives.

Thus an event only carries nonzero information if it changes \( \text{objective}(R, C) \), a property consistent with the intuitive flavors of relativity, rationality, and timeliness. Note that conditioning on C is necessary given that there may be multiple occurrences of the event E, each in a different context. Also note that in view of Examples 5 and 6, negative information is not unthinkable. In fact, this might lead to an interesting distinction of systems: nonconfoundable systems, contrasted with confoundable ones, are those whose rules of conduct R preclude negative information regardless of C. One can imagine a smart Web user always able to remove conflicting data from the context and proceed monotonically towards her objective. Whether and for what types of data sources and objective functionals such R exist is an open problem.
Finally, it is natural to surmise that both R and C are subject to various constraints implied, respectively, by the systems' architecture and the nature of the event sources. In the spirit of Shannon, one may define the channel capacity between the event source and the recipient as a maximum-type measure on a collection of amounts of information carried by successive events, within the regions of feasible R and C (i.e., subject to the said constraints). Thus the capacity of the channel between the event source and recipient is

\[
\text{capacity} = \max_{\text{feasible } R} \max_{\text{feasible } C} \max_{E \in C} \text{info}_{R,C}(E). \tag{2}
\]

With so structured a definition it is possible to confine interest to one or both of the inner maxima if for some reasons the systems' architecture and/or the nature of the event source are regarded as fixed. We should point out that calculating the capacity in the above framework seems to be particularly difficult in a distributed environment featuring multiple autonomous systems (such as a multi-agent system).

4. FINAL REMARKS

Our definition (1) is somewhat similar in spirit to that of the value of information, independently proposed by Luenberger. In the presence of a single source of uncertainty (in the form of different possible states of the world), he considers a decision-maker maximizing her average payoff and calculates the net benefit of receiving an imperfect signal about the true state of the world. Clearly, the net benefit is zero if the signal does not reduce the uncertainty, just as Shannon's statistical information is zero if the binary error rate is 50%. In such a Bayesian setting, negative values of information are impossible. Although we propose a broader framework, with context and rules of conduct explicitly accounted for, we still need a generalization of imperfect signals (or imperfect events in our wording); ours is a faultless communication system, where events do not get corrupted or misinterpreted. While partly justified given contemporary high-quality transmission and processing infrastructure, this is a serious restriction.

There are many other challenging questions. What is the capacity of the channel between a source and recipient with conflicting objectives (cf. Example 7)? In a simple entry deterrence game the capacity may compute to zero even though there is no physical obstacle to communication! A no less fundamental question is, How are structural information and communicable information related? The foregoing discussion is biased toward the latter, which is why the notion of event plays so central a role: a system remaining in one and the same state cannot change its perception of the achieved objective. However, another strong intuition of information holds it to be embedded in the very structure of an object and thus independent of any rational activity – this we may refer to as structural information. In this sense one attributes structural information to a book volume or graph connectivity (as expressed e.g., by Koerner entropy). Structural information can alternatively be regarded as a potential to achieve objectives assuming the existence of a rational recipient (observer). A measure of structural information contained in an object could be taken to be the upper bound on the communicable information given by (1) over all possible sequences of events resulting from observation of the object. For example, reception of summaries of successive chapters of a book may in some cases yield more communicable information than reception of the whole book content word by word (by instantly creating a complete picture of the topic or provoking free associations), thus better approximate the amount of structural information.
REFERENCES

We list below some relevant books and articles on the general notion of information that the reader might find useful. They are quoted with the understanding that many other references exist.


CZYM JEST INFORMACJA?

Streszczenie

Pojęcie informacji było dotąd kwantyfikowane głównie w terminach statystycznych, dając początek shannonowski teorii informacji i zasad cyfrowej transmisji danych. Współcześnie systemy komunikacyjne łączące skomplikowane, inteligentne i autonomiczne podmioty wymagają nowego spojrzenia na miary informacji, uwzględniającego kontekst i semantykę przekazu oraz racjonalne działania podmiotów. W niniejszym eseju zaproponowano podejście do obliczania miar informacji w paradygnacie sterowania zdarzeniowego.