On the Quantification of Relevant Information

Daniel Polani, Thomas Martinetz and Jan Kim
Institute for Neuro- and Bioinformatics, University of Lübeck, Germany

Much effort has been spent on trying to quantify the amount of semantic information processed by a system. The quantification of semantic information has been the topic of many discussions on the philosophical, conceptual or predicate logic level (e.g. Bar-Hillel 1964). However, what is frequently sought is just a simple operational quantification of information, conceptually similar to Shannon's information, but differing in its (non-)treatment of semantics and being able to quantify *relevance*.

A new incentive to develop a quantification of relevant information has been brought forward due to the efforts to study the evolution of sensors, which have been increasing recently (Pask 1959; Cariani 1993; Lee et al. 1996; Lund et al. 1997; Mark et al. 1998; Liese et al. 2000). In that context one is interested in agents that evolve to solve a certain task (e.g. maximizing survival time or the number of offspring) and that, for this purpose, evolve appropriate sensors to solve that task. Such sensor optimization procedures can be driven by explicit sensor quality measures (Nilsson and Pelger 1994), but often the task is implicitly formulated and there is no a priori knowledge about which sensor properties are going to be relevant for the agents to be able to solve it. Now the question arises, when will sensors evolve that attempt to capture as much information about the agent environment as possible and when will they limit themselves to capture only a very restricted bandwidth of information channels.

Among other factors, this intuitively depends on what amount of information in the environment is *relevant* for the agent. What we are interested in is to quantify this amount of information. A couple of properties would be desirable for this quantity. What one would like to have is a measure that quantifies the relevant, the "valuable" information in the environment. One would like to be able to interpret it in a sense similar to the classical Shannon information; loosely speaking, its units should be expressable in "bits".

A theory of information value has been developed in the context of decision theory (Howard 1966; Poh and Horvitz 1996). Here, the value of a piece of information, (modeled by some random variable possibly providing evidence about a given system) is measured in terms of the utility function defined for that problem. However, the information value approach introduces a considerable degree of arbitrariness in form of the utility function (Bertsekas 1976). The utility scale differs from scenario to scenario and is not transferable in general. Also, information value does not measure the *amount* of relevant information carried by a variable, but just its *value*.

Therefore, here, to define a notion of relevant information, we adopt a stance between decision and information theory. Tishby et al. (1999); Kaski and Sinkkonen (2000) define a measure of relevance in a random variable X, given some relevance indicator variable Y by the mutual information $I(X;Y)$ between X and Y. In their work, the relevance indicator variable has to be obtained from some additional source, e.g. by appropriate tagging of samples for X by a priori knowledge or human intervention. Making use of that knowledge, they then extract that information in X *relevant* for Y.

In our approach, we wish to give a specific, decision-theoretic instantiation for Y. We assume a decision system with (for simplicity of exposition) finite state and action spaces. We assume that for a given state x, there exists a set $\mathcal{Y}^*$ of equivalently optimal actions $\mathcal{Y}^* := \{y_{i_1^*} \dots y_{i_l^*}\} \subseteq \mathcal{Y} = \{y_1 \dots y_k\}$, where $\mathcal{Y}$ is the set of possible actions. At this point we make no assumptions about the model that selects the optimal actions; one could, for instance, imagine the optimal actions to be maximizing the expectation value of some suitable utility function. Given that state x, we define a conditional probability $p(y|x)$ of selecting an action y by $p(y|x) := 1/|\mathcal{Y}^*|$ if $y \in \mathcal{Y}^*$ and setting $p(y|x) := 0$ else. In other words, given a state x, let $p(y|x)$ denote the probability that y will be selected as optimal, assuming the principle of insufficient reason (maximum entropy) if several optimal actions exist. Let $p(x)$ be the probability of the system being in the state x; denote the respective random variable by X. Let the random variable Y^* be jointly distributed with X via $p(x, y) = p(y|x)p(x)$, where the $p(y|x)$ is given as above. Y^* denotes the random variable signifying the selection of an optimal action. Then define the information in X *relevant* for the decision system by the mutual information $I(X; Y^*)$ between the state variable X and the action selection variable Y^*.

The definition is natural in many senses. It has an immediate interpretation as an information quantity of Shannon type. Via the decision mechanism it addresses and incorporates the subjectivity of relevance (i.e. it takes into account the question "*relevant for what purpose?*"). When there is no available choice in actions or no correlation between state and possible actions, the relevant information vanishes. This fits intuition that, if by knowing X the choice of Y^* is not affected, there is no relevant information in X with respect to Y^*.

The notion is readily generalizable to the situation when the exact state X is not known to the decision system, like in typical POMDP tasks (Kaelbling et al. 1996) or sensor evolution contexts. The notion of relevant information can be strengthened to quantify the *minimum* required information about the system state X to choose a right action Y^*. The case when the precise optimal action is not known or when there are non-optimal actions whose utility is only insignificantly smaller than those of the optimal ones can also be fitted into this framework. These aspects will be treated in a forthcoming paper.

References

Bar-Hillel, Y., (1964). *Language and Information*. Addison-Wesley.

Bertsekas, D. P., (1976). *Dynamic Progamming and Stochastic Control*, vol. 125 of *Mathematics in Science and Engineering*. New York: Academic Press, Inc. First edition edition.

Cariani, P., (1993). To evolve an ear: epistemological implications of Gordon Pask's electrochemical devices. *Systems Research*, 10(3):19–33.

Howard, R. A., (1966). Information value theory. *IEEE Transactions on Systems Science and Cybernetics*, SSC-2:22–26.

Kaelbling, L. P., Littman, M. L., and Moore, A. W., (1996). Reinforcement learning: a survey. *Journal of Artificial Intelligence Research*, 4:237–285.

Kaski, S., and Sinkkonen, J., (2000). Metrics that learn relevance. In *IJCNN*.

Lee, W.-P., Hallam, J., and Lund, H. H., (1996). A Hybrid GP/GA Approach for Co-evolving Controllers and Robot Bodies to Achieve Fitness-Specified Tasks. In *Proc. IEEE 3rd Int. Conference on Evolutionary Computation*. NJ: IEEE Pr.

Liese, A., Polani, D., and Uthmann, T., (2000). On the development of spectral properties of visual agent receptors through evolution. In Whitley, D., Goldberg, D., Cantú-Paz, E., Spector, L., Parmee, I., and Beyer, H.-G., editors, *Proc. Genetic and Evolutionary Computation Conference (GECCO), Las Vegas, Nevada*, 857–864. Morgan Kaufmann.

Lund, H. H., Hallam, J., and Lee, W.-P., (1997). Evolving Robot Morphology. In *Proc. IEEE 4th Int. Conference on Evolutionary Computation*. NJ: IEEE Pr.

Mark, A., Polani, D., and Uthmann, T., (1998). A Framework for Sensor Evolution in a Population of Braitenberg Vehicle-like Agents. 428–432.

Nilsson, D.-E., and Pelger, S., (1994). A pessimistic estimate of the time required for an eye to evolve. *Proc. Roy. Soc. Lond.*, 256:53–58.

Pask, G., (1959). Physical Analogues to the Growth of a Concept. In *Mechanisation of Thought Processes: Proceedings of a Symposium held at the National Physical Laboratory (No. 10)*, vol. II, 877–928. London: Her Majesty's Stationary Office.

Poh, K. L., and Horvitz, E., (1996). A Graph-Theoretic Analysis of Information Value. In Horvitz, E., and Jensen, F., editors, *Proc. Twelfth Conf. on Uncertainty in Artificial Intelligence*, 427–435. San Francisco, CA: Morgan Kaufman.

Tishby, N., Pereira, F. C., and Bialek, W., (1999). The Information Bottleneck Method. In *Proc. 37th Annual Allerton Conference on Communication, Control and Computing, Illinois*.