Proceedings of the Fifth Annual Conference on Human Interaction with Complex Systems,
UIUC, Illinois, April 30 - May 2, 2000

AN APPLICATION OF RISSANEN'S COMPLEXITY MEASURE TO OPTIMIZING ABSTRACTION LEVELS

Ricardo Gutierrez-Osuna
Oscar N. Garcia

Department of Computer Science and Engineering
Wright State University
Dayton, OH 45435

ABSTRACT

In previous work we have considered the user-centered complexity of a hierarchy of abstractions as observed from a user's perspective. We hypothesized that, given a perspective, there should be a variety of complexities depending on the granularity of the abstraction and the measure of complexity chosen. In this paper we consider an optimization based on the Minimum Description Length complexity measure and its implications for choosing the proper level of abstraction.

BACKGROUND

We have hypothesized ([GARC99], [GARC98]) that, given a user's perspective, there is a level of complexity, related to the level of abstraction of an object or system description, which facilitates operations with the object or system in the sense of that perspective. Here we will advance this hypothesis by utilizing a measure of complexity that illustrates the problem. A user or intelligent agent interacts with a given world through observations and forms a model. That model may be considered at different levels of abstraction (see cone on the right of Figure 1 below, analogous to Figure 4 in [GARC99]), which the user or agent uses at a specifically chosen level.

MINIMUM DESCRIPTION LENGTH

Corresponding to each of the hierarchical levels of abstraction in that figure there is a corresponding (bar) level of complexity for instantiations of the model at different granularities. We expect an instantiation where interactions with the model by the user or agent are easier. An intuitive approach to measuring complexity is the Minimum Description Length (MDL) principle [RISS89]. Imagine a communications game where a sender is to transmit a sequence of observations to a receiver. MDL states that the (stochastic) complexity of the problem can be measured by the number of bits used to encode a theory, plus the number of bits used to encode the observations with the aid of the theory. Thus, the optimal hypothesis *h* (out of a set H) for a given observation sequence D is:

$$h_{MDL} = \arg\min_{h \in H} \left\{ \underbrace{L(h)}_{regularity} + \underbrace{L(D|h)}_{error} \right\}$$

where *L(h)* is the length of the hypothesis and *L(D|h)* is the length of the data given the theory. *L(h)* captures regularity (the inverse of complexity) in the observations, whereas *L(D|h)* measures those aspects of the observations not predicted by the hypothesis. When the number of observations is small, the term *L(h)* dominates, biasing the selection criterion towards more regular, less complex hypotheses. As the number of observations increases, the term *L(D|h)* becomes increasingly important, allowing more complex theories to be selected. The MDL principle is, thus, a data-driven instantiation of Occam's Razor: choose the simplest theory that explains a phenomenon for a given amount of data. MDL controls over-fitting by limiting the size (number of parameters) of the acceptable hypotheses based on the amount of data available for reliably estimating their parameters. In this regard, MDL is related to PAC learning, which provides lower bounds for the minimum amount of data needed for a given learning problem.

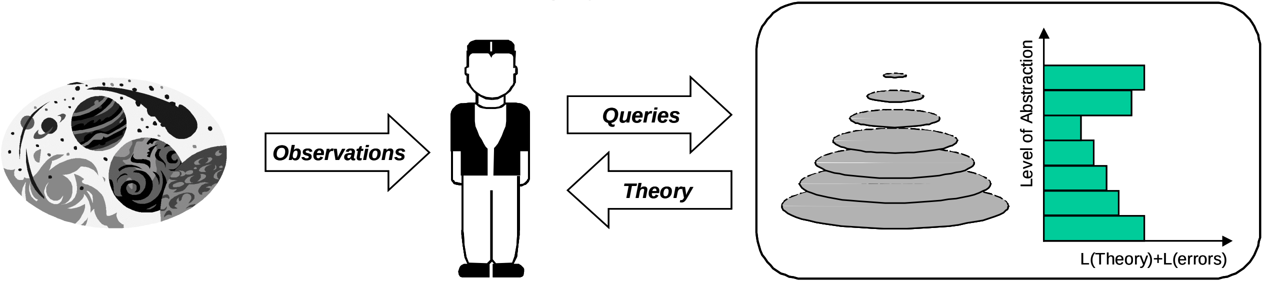

Figure 1. A user-centered perspective of the hierarchical abstract model with different MDL complexities

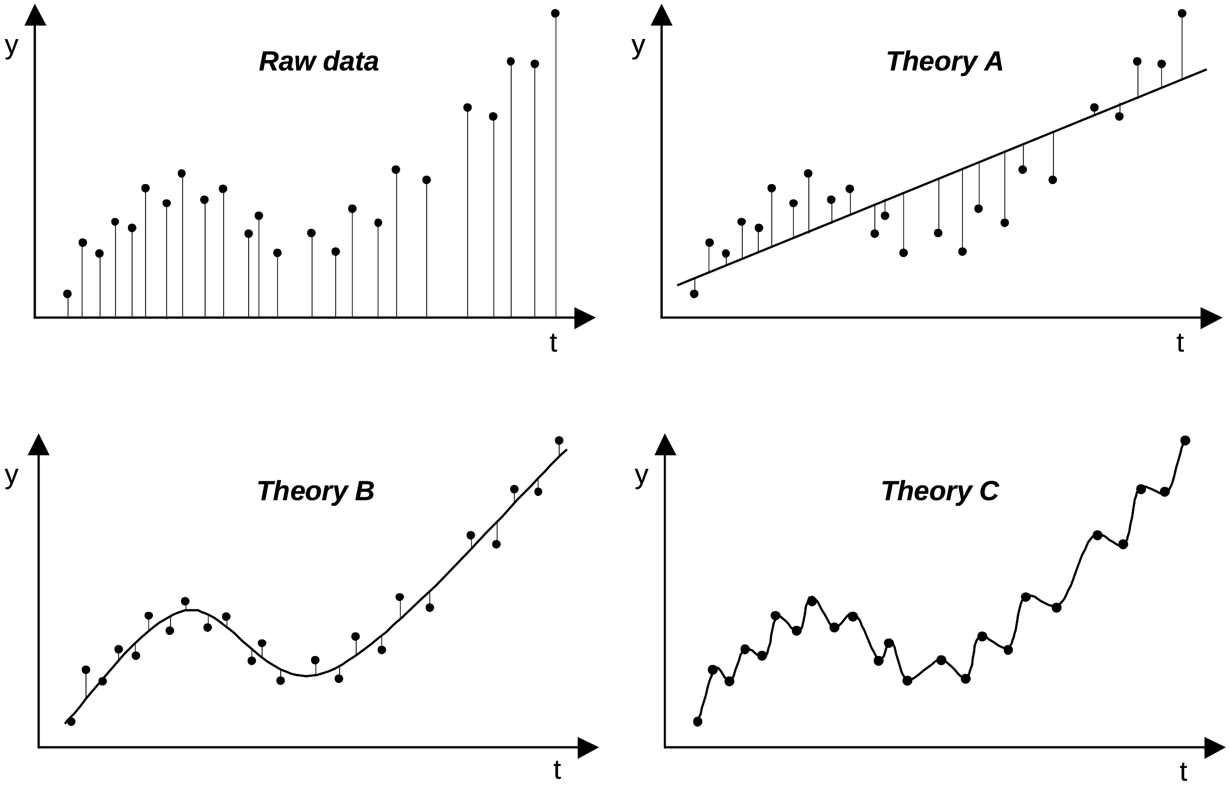

Figure 2. Different theories for transmitting a set of data across a fixed quantization error channel

EXAMPLE

To illustrate the MDL principle let us assume the problem shown in Figure 2, where the goal is to transmit the observation sequence $\{y_1, y_2, \ldots, y_N\}$, shown as raw data, across a channel with a fixed quantization error. The naïve alternative would be to transmit the binary codes of the raw data after it has been digitized with the required quantization level. Three other alternatives are shown in this figure. Theory A (least complex theory) encodes the series with a linear model $y = a_1 t + a_0$, Theory B performs a cubic fit $y = a_3t^3 + a_2t^2 + a_1t + a_0$, and Theory C (most complex theory) uses a polynomial $y = \sum_{k=0}^{N-1} a_k t^k$ of order N-1 and, therefore, fits the data without error. The relative cost of transmitting each theory is, in a first-order approximation, given by the number of parameters of the model (the order of the polynomials) times the number of bits used to encode each coefficient. The cost of transmitting the time series is proportional to the error bars (difference between the data and the model) divided by the required level of quantization, which remains fixed across models. There obviously exists a trade-off between the naïve approach of transmitting N raw data points and the overkill of transmitting N coefficients (Theory C). Depending on the number of data points, Theory A or Theory B will become the optimal method for transmitting the time series.

EQUIVALENCE WITH BAYESIAN APPROACH

Interestingly, the MDL principle has strong connections with the Bayesian formalism for model selection [MITC97]. The Bayes-optimal hypothesis h_{MAP} is the one that maximizes the posterior $P(h|D)$: the probability that a given hypothesis h is true after D has been observed. This Maximum A Posteriori (MAP) principle can be stated as:

$$h_{MAP} = \arg\max_{h\in H} \{P(h \mid D)\}$$

Applying Bayes' theorem and taking logarithms:

$$\begin{aligned} h_{MAP} &= \arg\max_{h\in H} \left\{ \frac{P(D \mid h)P(h)}{P(D)} \right\} = \\ &= \arg\max_{h\in H} \{P(D \mid h)P(h)\} = \\ &= \arg\max_{h\in H} \{\log_2 P(D \mid h) + \log_2 P(h)\} \end{aligned}$$

From Information Theory we know that the optimal coding of a grammarless string of messages $\{m_1, m_2, m_3, \ldots\}$ is achieved by assigning codes length $L(m_i) = -\log_2 P(m_i)$, where $P(m)$ is the probability mass function governing the generation of messages. In other words, events that are more likely to occur are encoded with fewer bits so that the overall message string length is minimized. If such an encoding is used to transmit hypothesis and data, MAP and MDL are equivalent since maximizing the log posterior equates to minimizing description length.

ABSTRACTION LEVEL SELECTION

The same MDL model selection principle can be adopted to determine an appropriate level of abstraction to explain a given phenomenon, relative to the perspective of a user or intelligent agent. In this case, hypotheses or theories "explaining" the model become different levels of abstraction. The user/agent makes observations and tries to interpret/understand them by asking questions to an oracle. The oracle's task is then to find an appropriate level of abstraction that yields a simple mental model for the agent, yet accurate enough to explain the phenomena. A very abstract model will not be able to accurately explain the observations, resulting in a large number of special cases to handle the intricacies of the observations. On the other hand, a low level (of abstraction) model will accurately explain the observations, but will be too complex to be of any use to the agent.

CONCLUSION

We have shown an approach that allows an optimized choice of abstraction level given a complexity measure and a perspective of a complex system.

ACKNOWLEDGEMENT

This research was supported by awards from NSF/CISE 9601670, 9906340 and NSF/CAREER 9984426.

REFERENCES

[GARC98] Garcia, Oscar N., "An Eclectic Approach to Complexity from a Human-Centered Perspective," in the Proceedings of the Human Interaction with Complex Systems Symposium, pp. 4-14, March 21-3, Dayton, OH, 1998.

[GARC99] Garcia, Oscar N., "An Approach to Complexity from a Human-Centered Artificial Intelligence Perspective" in Encyclopedia of Computer Science and Technology, Vol. 40 (A. Kent and J. G. Williams, eds.), Marcel Dekker, New York, 1999.

[MITC97] Mitchell, Tom, Machine Learning, McGraw-Hill, 1997.

[RISS 89] Rissanen, Jorma, Stochastic Complexity in Statistical Inquiry (Series in Computer Science, Vol. 15), World Scientific, 1989.