Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT

Journal of Statistical Planning and Inference 115 (2003) 361–364

journal of statistical planning and inference

www.elsevier.com/locate/jspi

Discussion

Knowledge management and information superiority (a taxonomy)

Nozer D. Singpurwalla

The George Washington University, Washington, DC 20052, USA

Abstract

In this expository note we introduce a taxonomy for some recently introduced, but vaguely defined, terminology used in the military/industrial contexts. The material here is not intended to be conclusive. Its aim is to cast the new terminology in the context of notions that have been around for quite some time and which have proven to be of value.

Keywords: Command and control; Complexity; Dominance; Information; Likelihood

1. Introduction—preamble

Statisticians are often taken by surprise when technologies that are the mainstay of their profession appear on the horizon under new labels and then end up circumventing the entire statistical community. A case in point is "Bayesian Nets" and "Data Mining", new labels for conditional probabilities and data analysis, respectively. Indeed there is little, if any, that is truly Bayesian about Bayesian Nets, except that it sounded good to the computer scientists who coined the term. Whereas all this is a backhanded complement, and a re-affirmation of the power and the strength of statistical thinking, it is all the same important that statisticians be alert as to what else lurks on the horizon that is liable to diminish their ranks and their scientific standing. One such label is "Knowledge Management" or KM; another is "Information Dominance". The former has now become the lingua franca of Fortune 500 companies, whereas the latter is an often heard term in the strategic world of "Command and Control", henceforth C^2. Despite the lack of a clear explication of as to what these terms mean, not to mention the absence of a theory to support it, there is a large body of individuals who claim to be specialists in these, and the list seems to be growing. Indeed, there is also a professional society devoted to the practice of KM.

E-mail address: nozer@research.circ.gwu.edu (N.D. Singpurwalla).

0378-3758/02/$ - see front matter
PII: S0378-3758(02)00164-7

This note is addressed to both, statisticians (to include probabilists) and to those interested in KM (and C^2). The former will benefit from knowing what else lurks on the horizon that is liable to bypass them. The latter may appreciate our modest aim of giving some structure to the general captions they use in terms of notions that have been around for some time and have proven to be of value.

2. A taxonomy for KM and C^2

The following terms are often used by those to whom this note is addressed: *data*, *knowledge*, *complexity*, and *information*. Often, they are qualified by adjectives such as “management”, and “superiority”. Whereas these terms do contain some intuitive import, it is important that they be defined and distinguished. The following is an attempt at the above; by no means it is meant to be conclusive and definite, though I cannot think of alternatives.

2.1. What is data?

Data is that which is directly observable and therefore measurable; it is therefore called factual. Measurement is in terms of three dimensions; distance, mass, and time. To paraphrase Einstein, distance is what you measure with a ruler, mass is what you measure with a scale, and time is what you measure with a clock. Measurement is a means of quantification, and quantification is the essential ingredient of the scientific method. The scientific method is one that is based on logic; it *excludes* from consideration things such as tastes, feelings, emotions, etc. To quote Lord Kelvin “if it cannot be measured, it cannot be discussed”.

2.2. What is knowledge?

Knowledge is a statement about a hypothesis, be it scientific or sociological. The formulation of hypotheses could be based on perception or discovery. The laws of physics are hypotheses about the states of nature. Similarly, in the other sciences. The testing of hypotheses is one of the oldest scientific endeavors dating back to Newton, Laplace, and Mendel. In modern times the testing of hypotheses (scientific laws) have had much to do with work of De Finetti, Fisher, Neyman and Pearson, and Jeffreys. The testing of hypotheses is often based on data, suitably organized, and in some cases on both data and judgements (encapsulated by the likelihood function). Science is organized knowledge, i.e. a collection of one or more hypotheses in some logical manner.

2.3. What is information?

Information is a measure of uncertainty about a hypothesis. A simple example of a measure of uncertainty is the variance. A more elaborate example is $I[p(x)]$, where

$$I[p(x)] = \int_x p(x)\log(p(x))\,\mathrm{d}x,$$

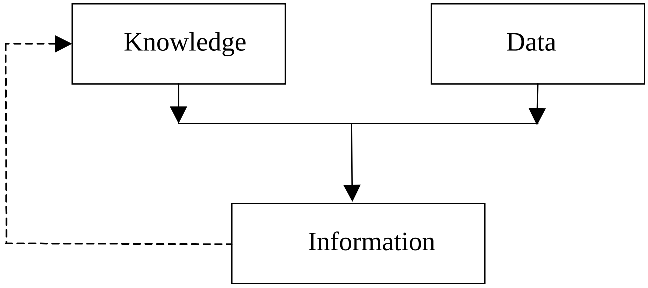

Fig. 1.

where $p(x)$ is the probability that a random (unknown) quantity, say X, takes the value x. In effect $p(x)$ is a measure of our belief about the hypothesis that $X = x$, for $x \in \mathbb{R}$. The quantity $-I[p(x)]$ is known as *entropy* in physics. It encapsulates the amount of disorder in some entity of interest.

The role of data is to change the amount of information or to decrease the entropy. Data by itself *cannot* formulate a hypothesis. Rather, it changes the odds in favor of (or against) a hypothesis. Indeed, data can increase our uncertainty about a hypothesis, e.g. a surprise.

Thus, we have as a hierarchy the illustration of Fig. 1 shown above.

3. Knowledge management and superiority: a proposal

3.1. What is knowledge management?

Knowledge management is the systematic use of data to test a hypothesis. De facto, it is the formulation of a *likelihood function*. Not all data is relevant to a hypothesis; when such is the case, the likelihood is flat. Otherwise it contains one or more models.

Thus if the term "knowledge management" is to have any scientific import, then it must be cast in the framework of the test of a hypothesis wherein the likelihood function plays a signal role. It is my view that knowledge management is the specification of a meaningful likelihood function, be it based on a notion of probability (objective, or subjective).

3.2. What is information superiority?

Information superiority is the acquisition of that data, *and* the manner in which it is utilized, so as to result in the lowering of uncertainty about a hypothesis (i.e. decreasing the entropy). Lowered uncertainty results in optimum decision making. For example, if two agents, say A and B possess data y_A and y_B, respectively, and the hypothesis in question is encapsulated by $p(x)$—where $p(x)$ reflects the uncertainty of both A and B about an X taking values x—then A possess a knowledge superiority over B, with respect to $p(x)$ if

$$\int_x p(x|y_A)\ln[p(x|y_A)]\,\mathrm{d}x < \int_x p(x|y_B)\ln[p(x|y_B)]\,\mathrm{d}x.$$

When several agents share a common $p(x)$, then for agent A to gain knowledge superiority over all others, A must obtain that y_A for which

$$\min_{\text{all}\, y_A} \int p(x|y_A)\ln[p(x|y_A)]\,\mathrm{d}x.$$

Acknowledgements

Research Supported by the Office of Naval Research, Grant N00014-99-1-0875, with The George Washington University.