A MATHEMATICAL MODEL OF KNOWLEDGE

STEN KAIJSER

Abstract. In this paper we shall construct a mathematical model of knowledge. The model is based on the assumption that knowledge is the ability to use information in an adequate way by anticipating future development. Our main result is that very accurate short range predictions can be used for long range predictions.

1. Introduction

What is knowledge? This paper is based on the conviction that in the modern world where information is abundant, a reasonable characterization of knowledge is the ability to *use* information. Using information means that *new information changes activity*, and a first step towards a "mathematical description of knowledge" could be given by the relation

$$dA = KdI.$$

However at second thought it is not our knowledge but our conception of the world that makes us change our activity. We shall therefore start with the relation

$$dA = SdI,$$

where we shall think of S as a system which will later be called a shadow system. We shall then define *knowledge* as a property of the system S, which describes how adequate the change of activities is. This means that we shall need a measure of adequacy, a "reward function", which will be denoted by ρ. The purpose of the present paper is to give a *mathematical* model of knowledge based on this view of what knowledge should be.

2. General Discussion

2.1. **Knowledge in general.** We shall start by discussing knowledge in (mostly) non-mathematical terms in order to motivate the model which will be presented later.

As we all know the word knowledge is used in many different contexts and even if it is usually not (mathematically) defined most of us will in concrete cases agree on what knowledge should mean. The purpose of this paper is to construct a *mathematical model* of knowledge. In order to do this we shall first need a suitable *description* of what knowledge is.

Date: August 19, 1998.

Key words and phrases. Knowledge, Systems theory, learning.

Research supported by the Faculty of Science and Engineering at Uppsala University.

As stated in the introduction our starting point is that knowledge is the ability to use information, which we shall describe in terms of the relation

$$dA = SdI,$$

where dA represents a change of activity due to the change of information dI and S stands for the conception of the world (world image) which leads to a particular way of changing activity. (If we believe in it, then dA may very well be the reading of a particular spell of magic.)

Often, when thinking of knowledge, we have in mind some entity, say P having this knowledge. We could call P the "knower" and P may, besides being a human person, be e.g. a company or a community of scientists. However, as Laplace said in a similar situation

"je n'ai pas le besoin de cette hypothèse.

which in our case means that we don't need a "knower", since it is the world view S itself which is to have the knowledge.

Remark 1. "I don't need that hypothesis", is what Laplace is supposed to have said to Napoleon when the latter remarked that he had written a whole book about the world without mentioning its creator.

To continue our discussion we shall next have to consider the question of what we should have knowledge about, and continuing to be vague we shall introduce a system Σ, and we shall try to define what it means to have knowledge about Σ. Even if we intend to construct a *mathematical model* which means that it can not "be falsified by reality", as Karl Popper would have expressed it, we shall nevertheless give some examples of what kinds of systems that we have in mind.

One example of a system about which knowledge has existed for a long is the solar system. About this it is known (*sic!*) that several of the ancient high cultures such as the Babylonians (and perhaps already the Sumerians) could predict solar eclipses. It is also known that the pythagoreans believed that the earth was a perfect sphere and that Eudoxus constructed a geo-centric mathematical model of the solar system. This model was developed by (among others) Appolonius, Hipparchos and Ptolemaios, while Aristarchus even suggested a helio-centric model. In spite of the fact that all these models were intrinsically wrong, they certainly represented a form of knowledge, useful e.g. for navigators. It is perhaps also worth remarking that even if modern physicists describe the solar system in terms of Einstein's theory of general relativity, NASA relies on Newton for navigation. In order to allow even incorrect theories to "contain knowledge" we shall prefer a "phenomenological" description of knowledge.

Two other systems that we have in mind are e.g. the weather system and a financial market. A third, very interesting question is related to genetic knowledge. The cloning of the famous sheep *Dolly* showed both that all the necessary information for constructing a sheep was preserved in the DNA of a mature cell of a fully grown sheep, and that the knowledge of how to use the information was present in the combination of an egg cell and a matrix in another sheep. The interesting question that arises is if, assuming that we find an intact frozen cell from a mammoth, a modern Indian elephant has the necessary knowledge to produce a mammoth.

Two properties that these systems have in common are that they are dynamical systems, where knowledge is related to the ability of predicting future development, and also that we cannot influence them (unless we are a very big actor on some financial market). In the following we shall thus think of S as a conceptual image of a system Σ, and since S should "know Σ" we shall call S a *shadow system*. In order to have a definite mathematical system we shall assume that our system Σ is a dynamical system.

Claiming that knowledge is the ability to use information, we next have to assume that we can get information about the system, e.g. as a list of measurements. Usually "knowledge" includes the ability of actively searching for information (e.g. by making experiments, or reading books), but to simplify our models we shall in the following not allow this. Instead information, $\mathcal{I}$, will simply be a string of symbols, and as every computer knows, we may simply assume that we have a string of binary digits (*bits*). New information will then simply be the reading of one more bit of information. Furthermore since all our information is about $\mathcal{I}$ it is really only about $\mathcal{I}$ that it is possible to have knowledge. We shall therefore identify Σ and $\mathcal{I}$ and since we wish to call it a system and denote it by Σ we shall interpret this as meaning that we have complete information (up to present time) about the system. This makes it natural to let Σ be a discrete time system, with the time determined by the reading of new information.

Since we want to interpret knowledge as adequate *reactions* on new information we first of all want the system S to react on new information. In real life the reactions may in some cases have an influence on the system itself while in others it does not. In order to make our model simple we shall only consider systems where the activities do not influence the system. To be precise we shall assume that there exists a (finite or infinite) set $\mathcal{A}$ of possible activities that we have at our disposal and we shall assume that each new bit of information will make it possible for us to choose one of these activities. This means that at each instant of time n we choose an action $a(n)$, based on the information up to time n.

The next thing is to describe when an activity is *adequate*, and for this we shall introduce a *reward function* ρ which may depend on both the action a and the state of the system Σ. If the function ρ should really reward knowledge it should in some sense be able to distinguish knowledge from No-knowledge. One way to achieve this is to use the theory of games, as invented by von Neumann. We thus assume that at each instant of time the systems Σ and S play a game. Without loss of generality we assume that the value of the game is 0, and we let the No-knowledge activity be to use the 0-strategy (see e.g. [von Neumann]).

A very simple example of this is the following game. At each instance of time both systems produce binary digits $\sigma(n)$ and $s(n)$, where $s(n)$ is allowed to depend on the observations $\sigma(k),\ 1 \leq k \leq n-1$. If the digits are equal then S wins one unit while if they are different S looses the same amount. A" pre-model" of knowledge could then consist of a finite or infinite binary string Σ, interpreted as a dynamical system, together with a shadow system S of the same form.

Knowledge $\mathcal{K}$ could then be described as the average gain i.e. as

$$\mathcal{K} = \mathbb{E}(\rho(\sigma, s)) = \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} \rho(\sigma, s)(k)$$

(where $\mathbb{E}$ means mathematical expectation), and one reason for calling this a pre-model is the fact that without further assumptions the limit need not exist.

The No-knowledge activity is in this model represented by a coin-tossing sequence (independent of Σ).

3. A mathematical model of knowledge

3.1. **Ingredients of the model.** In this section we are going to present a mathematical model of knowledge.

According to the discussion of the preceding section a model could consist of a 5-tuple $(\Sigma, \mathcal{I}, S, \mathcal{A}, \rho)$, where Σ is a dynamical system, $\mathcal{I}$ is information about Σ, S is the shadow system, which using information from $\mathcal{I}$ chooses activities from the set $\mathcal{A}$, and finally the reward function ρ measures the adequacy of the activities. It is furthermore desirable that there should exist a No-knowledge activity, and for normalization reasons ρ should be scaled so that the adequacy of the No-knowledge activity is evaluated as 0 and that the optimal activity, due to complete knowledge of Σ should have the value 1. In the preceding section it was observed that a simple model along these lines can be obtained if we assume that we have complete information about Σ, so that we may identify Σ with $\mathcal{I}$. This means that Σ simply consists of a string (finite or infinite) of *bits*. This assumption makes it natural to assume that the shadow system S is also a string of bits, and then we may as well identify each bit with an activity, so that we may avoid both the sets $\mathcal{I}$ and $\mathcal{A}$. A simple evaluation function is then given by the function

$$\rho(\sigma, s)(n) = (-1)^{\sigma(n)+s(n)}.$$

The No-knowledge activity in this case is to let S be a coin-tossing sequence, i.e. a Bernoulli sequence, independent of Σ and with

$$P(s(n) = 0) = \frac{1}{2} = P(s(n) = 1).$$

In this model we define the *displayed knowledge* $\mathcal{K}_S(n)$ of the shadow system S at time n as the average

$$\mathcal{K}_S(n) = \frac{1}{n}\sum_{k=1}^{n}(-1)^{\sigma(k)+s(k)}.$$

If the sequence Σ has finite length, say N, then the number $\mathcal{K}_S(N)$ is of course a measure of how much knowledge the system S has of the system Σ. If the sequence Σ has infinite length then we shall say that the system S has *knowledge* of σ if

$$\mathcal{K}_S = \underline{\lim}_{\,n\to\infty}\mathcal{K}_S(n) > 0.$$

Remark 2. It should be observed that this definition is truely mathematical for several reasons. The first is that infinite sequences only exist in mathematics and even if we may consider any sequence e.g. of measurements as the beginning of an infinite sequence, we can never know if the sequence is going to continue in the same way that it started. Furthermore also for mathematical sequences the definition requires some kind of absolute knowledge at least about the “primary” sequence Σ, since without this we can never know if a certain value really is the $\liminf$ of the sequence. Philosophically this means that knowledge can not be created from nothing – without knowing in advance that at least the sequence $\{\rho(sigma, s)(n)\}$ is not completely arbitrary, information of only a finite part of that sequence can never guarantee that it will not suddenly start to behave in a completely different way.

3.2. **Construction of the model.** In order to complete the construction of our model we have to put some restrictions on the systems Σ and S. A first observation is then that the model is symmetric in Σ and S, so that if S has knowledge of Σ then Σ has knowledge of S. In order that S should rightly be called the *shadow* system, and if it should really represent *knowledge* about the *primary* system Σ, there should be restrictions of what kind of systems that we should allow for S. The weakest reasonable assumption should be that $s(n)$ is completely determined from the observations $\sigma(k),\ 1 \leq k \leq n-1$ and this implies that there should exist a sequence, *specified in advance*, of Boolean functions φ_n, such that $s(n) = \varphi_n(\sigma(1), \sigma(2), \dots, \sigma(n-1)$.

For both practical and philosophical reasons this is really not restrictive enough as a condition of the shadow system. The practical reason if of course that not all such systems can be implemented into a computer, since some of them would require an infinite memory. The philosophical reason is similar since it is difficult to imagine any entity capable of remembering an infinite amount of functions.

A natural restriction would then be to require that the system S can be implemented on a computer with a finite memory. This would imply that S can only use a finite number, say N of the observations, and for most "real systems" (i.e. sequences intended to model a system coming from the "real world") the most relevant information is the latest, so that we should assume that $s(n)$ can be calculated as a function of the observations $\sigma(n-1), \sigma(n-2), \dots, \sigma(n-N)$.

An argument against this assumption could be that a reasonable kind of sequences modelling real life systems, would be given by choosing from a certain class depending ("continuously") on a certain set of real parameters. It would then be reasonable to use the early observations for determining the parameters, and then the rest of the sequence would perhaps be completely determined. However it is clear that if the parameters specify the sequence, then that information can be recovered also from the later observations. We shall therefore, whenever S is only allowed to remember a finite set of observations, asssume that those are always the most recent. We could then assume the existence of a Boolean function φ, such that

$$s(n) = \varphi(\sigma(n-1), \dots, \sigma(n-N)).$$

This assumption is on the other hand slightly too restrictive, since it means that we are unable to keep track of the time. With a finite memory we are of course also unable to represent arbitrarily large numbers, so that we cannot keep track of absolute time. Nevertheless it is possible to keep track of week-days, or the hour of the day, and in our context this means that it is possible to allow a periodicity of bounded length. This possibiliy is of vital importance for interpreting genetic information, because if we can't count to three then we will loose everything. A slightly more general model is therefore to assume that the function φ, depends not only on the last observations, but also on the residue class of $n \mod (p)$. We shall call a system that can be implemented on a computer with finite memory and which together with a string of binary digits produces another such string, a finitely represented system. If the system uses at most N observations, and if the period is of length p then we shall say that S has degree $D(S) = \max(N, p)$.

We have now finally the following mathematical model.

Definition 3.1. An *elementary knowledge model* K is a pair (Σ, S) where both the *primary system* and the *shadow system* S are infinite binary sequences. Furthermore the shadow system S is finitely represented. The *displayed knowledge* at time n is given by the average at time n, $\mathcal{K}_S(n)$, defined by the relation

$$\mathcal{K}_S(n) = \frac{1}{n}\sum_{k=1}^{n}(-1)^{\sigma(k)+s(k)}).$$

Definition 3.2. Let (Σ, S) be an elementary knowledge model. We shall then say that the system S has positive knowledge of the system Σ if

$$\mathcal{K}_S = \underline{\lim}_{n\to\infty}\mathcal{K}_S(n) = \alpha > 0.$$

The number α will be called the *amount of knowledge* possessed by S.

In terms of these definitions we shall say that a given system Σ is *finitely predictable* (or simply *predictable*) if there exists *some* finitely represented shadow system S that has (a positive amount of) knowledge of Σ.

Assuming that the system Σ is finitely predictable, we shall next define a number that tells how predictable it is.

Definition 3.3. Let Σ be finitely predictable. Then we shall say that Σ is α-predictable (or that it has predictability α) if

$$\sup_S \mathcal{K}_S = \alpha,$$

where the supremum is taken over all finitely representable shadow systems. If Σ is 1-predictable, then we shall say that it is *completely predictable*.

For further investigations of predictable sequences it will probably be of interest to have a relation between the needed degree of a shadow system and the amount of knowledge it may contain. Using the preceding definition we have next the following

Definition 3.4. Let Σ be finitely predictable. Then we shall define the *knowledge function*, $\kappa_\Sigma(D)$, to be the maximal amount of knowledge, that any shadow system of degree D can have. Its inverse function $d_\Sigma(\alpha)$ is then the smallest degree possible for a shadow system having at least the amount α of knowledge about Σ. Since we are speaking of knowledge we shall call d_Σ the *difficulty* function. This means that

$$\kappa_\Sigma(d) = \sup_{D(S)\le d} \mathcal{K}_S$$

and

$$d_\Sigma(\alpha) = \min_{\mathcal{K}_S\ge\alpha} D(S).$$

4. Predictable systems

4.1. **Long range predictions.** In this section we shall consider one implication of the model we have presented, and we shall discuss some problems that they give rise to.

The first problem we shall consider is the question of whether predictability gives rise to long range predictability.

Definition 4.1. Let Σ be finitely predictable. Then we shall say that Σ is n-step predictable if there is a finitely represented system S, having positive knowledge of Σ, such that $s(k)$ is determined from the observations only up to time $k - n$.

Remark 3. One reason for considering n-step predictability is that when numerical information is digitalized it is so to say easier to use *bytes* than *bits*. This means that information comes in packages and it should then be possible to predict at least the next package of information. Another reason is that it is related to chaos. As we know it is very difficult to make long time weather fore-casts, a fact that is usually explained by saying that the weather system is chaotic. On the other hand it seems conceivable that it should be possible to obtain more and more precise short range predictions. As the following theorem indicates this highly unlikely, since drastically improved short range predictions will lead to better and better long range predictions.

It is from examples easy to see that 1-step predictability does not in general imply n-step predictability. On the other hand it is reasonable to expect that very good 1-step predictability should make it possible to replace future observations by their predicted values. That this is in fact true is the content of the following

Theorem 4.1. *Let Σ be completely predictable. Then Σ is n-step predictable for all n.*

To prove this theorem we shall prove the following

Lemma 4.1. *Let Σ be predictable and let S be a finitely represented shadow system such that $\kappa_\Sigma(S) \geq (1-\beta)$. Then there exists a shadow system S', which computes $s(k)$ using only observations of Σ up to time $(k-n)$, such that $\kappa_\Sigma(S') \geq (1-n\beta)$.*

Proof. Consider the strings $\{\sigma(i)\}_{i=m}^{m+L}$ and $\{s(i)\}_{i=m}^{m+L}$. By assumption they coincide "most of the time" and if we also consider the sequence $\{\rho(i)\}_{i=m}^{m+L}$ we know that at most $L \cdot \frac{\beta}{2}$ of these are negative. If we now define s' in the same way as we would have defined s from observations of σ but with the last n σ-s replaced by the corresponding s-es, then we are sure to get the same value for s' if there was no mistake along the way. However each mistake can lead to at most n mistakes of s', so the total number of mistakes is at most $Ln\beta/2$ and then the average of ρ is at least $(1-n\beta)$ and this proves the lemma. □

It is clear that the theorem follows immediately.

Remark 4. Observe that if we make a mistake, then after n steps we have to recalculate all the s-es.

4.2. **Concluding remarks.** We believe that the most interesting concept arising out of our model is the notion of predictability and concerning this notion we wish to point out that our model is a mathematical model with all the limitations and advantages of those. The main limitation is that it is not a model for all that we usually call knowledge. In fact the definition of predictability is in some sense both to restrictive and to general. An example of a predictable sequence is a coded random sequence of letters, where the code is self-correcting. For such sequences every n-th bit may be predicted which implies that it is $(\frac{1}{n})$-predictable. Whether this should be called knowledge or not is a matter of judgement. It is however similar to knowledge of the genetic code.

An important aspect of the model is that it only accepts *predictions* as knowledge. The ability to *explain* afterwards why something happened is not considered as knowledge unless it can be used for e.g. avoiding the same thing to happen again.

One aspect of knowledge that lies completely outside of this model is *learning* and other ways of improving knowledge. On the other hand it is quite clear that *examinations* fit well into the general frame-work of the model (i.e. a model containing also $\mathcal{I}, \mathcal{A}$ and ρ, so that in this sense our model really models how *testing* of knowledge is usually performed.

Another example indicating that our model really captures what many of us consider as specialized knowledge is the advertisement of financial institutes that proudly announce that their analysts are *beating index*. The definition of knowledge given by our model should tell us that their performance during a limited period does not prove that they have knowledge.

Reference: von Neumann, A theory of Games.

Department of Mathematics, Uppsala University, P.O.Box 480, S-751 06 Uppsala, Sweden

E-mail address: sten.kaijser@@math.uu.se