A Mathematical Model of Information Theory: The Superiority of Collective Knowledge and Intelligence

Pedro G. Guillén(✉)

Knowdle Foundation & Research Institute, Málaga, Spain
pedrogguillen@gmail.com

Abstract. The mathematical model described here is an evolution of the one constructed within the studies [1–3] by De Santos and Villa, among others. This paper aims to formalize the concept of knowledge and its properties as a basis for creating generalized economic value functions [4], focusing on the business models of the current technological sector as the application environment. The main conclusions reached are focused on the improvement of the value of information and knowledge under the assumption of collective cooperation amongst information system agents, as well as the properties of the knowledge space derived from the model.

Keywords: Knowledge · Collective · Intelligence · Wisdom · Structure · Value · Economy · Swarm · Topology

1 Algebraic Structure of Knowledge

Definition 1: Let L be a natural language. The Universe of data (also called Universal Library) is defined as all possible combinations of characters of L of arbitrary length that have meaning, that is to say, whose Boolean function of L is not null [2]. In the future, this set will be denoted by D, and a subset $d \subseteq D$ will be called data. It will be assumed that the universe of data is dense, and any combination of elements of the universe of data is an element of the universe of data.

Definition 2: The set of Universe of Contexts or Universal Context is defined as a graph with global structure of thesaurus, and locally orientable. The set will be denoted by C, and a subset $c \subseteq C$ will be called context.

Proposition 1: Whatever the L language on which the previous structures D and C are defined, and whatever the structure of C, we have $C \subset D$.

Proof: Trivial.

Then the σ-algebras of the respective sets D and C are defined, and they will be called $\mathcal{D}$ and $\mathcal{C}$. The choice of σ-algebra is arbitrary; However, due to the structure of natural language and the results described in [1, 2], we can assume the choice of a "reasonable" non-trivial algebra.

Y. Tan et al. (Eds.): ICSI 2017, Part I, LNCS 10385, pp. 12–21, 2017.
DOI: 10.1007/978-3-319-61824-1_2

Definition 3: The following set is defined as the Universe of Knowledge of the natural language L:

$$\Omega := D \times C = \{(d, c) : d \in D, c \in C\}$$

A compact subset $\Delta \subseteq \Omega$ will be called knowledge domain or simply domain, and an element $\omega \in \Omega$ of the set will be called contextualized data.

In the model, we will call the subsets that are unions of path-connected components natural knowledge or simply knowledge. The motivation to conjecture this structure for natural knowledge comes from studies on cognitive structure and brain processes [5, 6], as well as studies of the Mathematical Theory of Information [7–9]. These studies present the natural structure in which an agent of an information system processes all the knowledge it acquires. Knowledge will be the basis for the topology of the Ω space.

Proposition 2: If any knowledge domain $\Delta \subseteq \Omega$ is defined, there is always an embedding function between the domain and the set that does not depend on the structure that is given to the space.

Proof: Is immediate, since having $\Delta \subseteq \Omega$, it is verified that the restriction is always defined and is always an embedding. Note that $C \subset D$.

Once the spaces (D, D) and (C, C) are defined, it is possible to define in the set Ω the inherited σ-algebra $\Xi = D\,x\,C$. Because of its algebraic properties, the fact that D, C have a structure of σ-algebra implies that Ξ has a structure of σ-algebra. The reason for presenting these sets by means of this algebraic structure is that a large number of useful operators can be built on it, as is defined below.

2 The Dynamic System

We define the continuous variable $t \in [0, \infty]$ which will be called time, and which must be given a starting point, t_0, in the model. Without loss of generality, we will assign $t_0 = 0$.

Analogously to the initial hypotheses, the sets of the Knowledge Universe are defined as functions of time:

$$D(t), C(t)$$

That enables the definition of a dynamic system in the σ-algebra of this universe, the Cartesian product:

$$\Omega(t) := D(t) \times C(t) = \{(d(t), c(t)) : d(t) \in D(t), c(t) \in C(t)\}$$

$$A(\Omega)(t) = A(t)$$

Regarding the model associated with the dynamic system that has been constructed, some additional assumptions must be made, because as the time variable takes a range of values or others, the set changes in a certain way.

Now the value function can be defined, which in the previous section was a vector associated with an element of σ-algebra. In the resulting dynamic system, the value function is a vector of n functions that depend of t.

$$V_t : A(t) \to K^n(t)$$

$$V_t(A(t)) := (m_1(A(t)), m_2(A(t)), \ldots, m_n(A(t)))$$

Or simplifying:

$$V_t : A(t) \to K^n(t)$$

$$V_t(t) := (m_1(t), m_2(t), \ldots, m_n(t))$$

3 The Value of Knowledge

Once the universe of knowledge is defined, the next step is to construct a function that associates each element of our σ-algebra with a numerical value. To this end the following results are introduced.

Definition 4: Any function of the form:

$$f(t) : A(\Omega)(t) \to K$$

that is a finite and bounded measure, and that it is dependent on a law of supply and demand, is defined as a value function.

Proposition 3: The value function is continuous.

Proof: Demonstration is trivial by construction, since an open set of the topology in the image space is the union of elementary images of the function, and its inverse image is precisely the union of basic openings of the topology of $\Omega(t)$. Therefore, the function is continuous.

Lemma 1: $(\Omega(t), T(A))$ is a Hausdorff topological space.

Proof: Take two elements of the space. There are two possible cases: that the points are in different components by paths or that they are in the same. In the first case, we have already found two open disjoints sets of the topology that contain both points and we are finished. In case of both points are in the sameset connected by paths, we consider all the possible paths that join both points. As the whole of the universe of data is dense, given two elements we can always find one that is in the middle, and therefore we can separate each path into two open disjoint paths, being the union of the halves the disjoint open sets containing each of the two starting points.

Lemma 2: Let A_i be a family of knowledge and f a function of value on $\Omega(t)$. Then it verifies that a $j \in N$ exists such that, if $i > j$:

$$\left|\frac{\partial f(\cap A_i)}{\partial t}\right| < 0$$

Proof: By Definition 4 we have that if f is a value function then it is dependent on a law of supply and demand. As by Definition 3 knowledge are sets belonging to agents of the information system, we then have the greater the number of agents possessing a certain knowledge (the intersection of all Ai) the greater the supply, and therefore the value of that set tends to go down in time for a sufficiently large index.

Corollary 1: A direct consequence of the previous lemma is that the temporary derivative of sufficiently infrequent knowledge value is inversely proportional to the temporal derivative of frequent knowledge. Thus, for a sufficiently large index under the hypotheses of Lemma 2:

$$\left|\frac{\partial f(\cap A_i)}{\partial t}\right| = -k\left|\frac{\partial f(\Omega \setminus \cap A_i)}{\partial t}\right|$$

is verified for some $\boldsymbol{k} > 0$.

Proof: Analogous to the previous case.

In general, the intention is not to establish a numerical value for each element of the σ-algebra, but a vector with n coordinates. For this, we have only to define n different measures on the given σ-algebra that gives to each set a value from the desired point of view, defining a value function as the composition of a finite set of measures defined on the universe of knowledge.

4 Collective Knowledge

One of the fundamental results of the model that has been described is the benefit of collective versus individual knowledge. Intuitively, we can think of profit as a positive difference of value, defined in the model by the homonymous function.

In the model, the differential is the derivative of the value function regarding the time variable, that is, in the growth of the functions

$$A : t \to A(t)$$

$$f_t : A(t) \to \mathbb{R}^n(t)$$

Theorem 1: Let Ai (t) be a family of knowledge and let f be a function of value. Then, for a sufficiently large index i:

$$\left|\frac{\partial f(\cup A_i)}{\partial t}\right| \geq \sum_{j=1}^{i}\left|\frac{\partial f(A_j)}{\partial t}\right|$$

Proof: The problem will be reduced to the case of two sets of knowledge. The general case is completely analogous. We know that any value function is a measure. To demonstrate the statement, we take two of the fundamental properties of our n-dimensional measure, namely

$$\forall A, B \in A | A \subseteq B \rightarrow f(A) \leq f(B)$$

$$f(A \cup B) = f(A) + f(B) - f(A \cap B)$$

Taking these properties into account, together with the linearity properties of the derivative:

$$\left|\frac{\partial V_t(A_1 \cup A_2)}{\partial t}\right| \geq \left|\frac{\partial V(A_1)}{\partial t}\right| + \left|\frac{\partial V(A_2)}{\partial t}\right| - \left|\frac{\partial V_t(A_1 \cap A_2)}{\partial t}\right|$$

is verified.

However, we had by Lemma 1 that for a large enough family of knowledge (we suppose the result true for $i = 2$)

$$\left|\frac{\partial(\cap A_i)}{\partial t}\right| < 0$$

Therefore, it is verified that

$$\left|\frac{\partial V_t(A_1 \cup A_2)}{\partial t}\right| \geq \left|\frac{\partial V(A_1)}{\partial t}\right| + \left|\frac{\partial V(A_2)}{\partial t}\right| - \left|\frac{\partial V_t(A_1 \cap A_2)}{\partial t}\right| \geq$$

$$\left|\frac{\partial V(A_1)}{\partial t}\right| + \left|\frac{\partial V(A_2)}{\partial t}\right|$$

That proves the result. Basically, the interpretation of this theorem is that the knowledge which agents bring exclusively to others who do not possess it, generates more valuable knowledge than that is common to all agents.

5 Maximal Knowledge: The Hypersurface

Proposition 4: The image of $\Omega(t)$ by k value functions is a hypersurface of dimension k.

Proof: By hypothesis, $\Omega(t)$ is a union of sets connected by paths. Thus, locally it is a set connected by paths. Applying Lemma 1, we have that the space $(\Omega(t), T(A))$ is a

Hausdorff space, and because it is locally connected by paths it is also locally arc-connected. Therefore, by taking a point in space $\omega \in \Omega(t)$, and an open one containing it, there are $I \subseteq R$ paths that are homeomorphic to R, and whose union builds a local homeomorphism between a subspace of R^m and $\Omega(t)$. Therefore, $\Omega(t)$ is a hypersurface and since the k value functions are continuous, and it has to bethat $m = k$.

Hereinafter, the notation $f(\Omega(t)) = H(t)$ will be used.

Proposition 5: Value functions form a vector space of dimension n > k.

Proof: Amongst the properties of measures is that the sum of measures is a measure that preserves algebraic properties, and the same thing happens with the product of scalars. Therefore, any function dependent on a law of supply and demand continues to be such after an operation with another function or with a scalar.

Corollary 2: H (t) is contained in a vector space of infinite dimension.

Proof: The result is immediate, since the product of measurable functions is measurable and gives rise to a higher degree function that is not expressible as an element of the starting vector space. By repeating the process indefinitely, we obtain a vector space of infinite dimension containing $H(t)$.

From now on, this space will be called parameterization space.

Proposition 6: Once a domain of knowledge is established, there is knowledge that has extreme values (maximum or minimum).

Proof: By definition, the constraint of $H(t)$ to a domain is a compact set, since the domains are compact and the value function is continuous. Applying the Weierstrass theorem, a continuous function defined in a compact has at least one extreme value, and the result is concluded.

6 Through Knowledge: Intelligence

Once the existence of maximal knowledge in the σ-algebra/parametrization/hypersurface structure is guaranteed, it is logical to ask what structure has the maximal knowledge set or if the value of knowledge can be treated as knowledge in itself to generalize in a higher abstraction the concept of value function.

Definition 5: An intelligence function is defined as a function dependent on t of the form:

$$I(t) = \int_{H(t)} F(x)dx$$

Being F a vector field function, such that the nature of the intelligence function is equivalent to the concept of divergence in fluid mechanics [10], i.e., that F would be a function that would represent the velocity field of a fluid and the function of

Intelligence measures how intelligence flows and expands through the hypersurface of knowledge.

Being a primitive of a function, it is evident that functions of intelligence are differentiable. Assuming the conditions in the previous section, and applying the fact that in this case H (t) is a compact set, both the hypersurface and differential maximal curves at the points of the hypersurface border can be calculated.

In this case the maximal curves should represent the optimal curves of knowledge through a given local domain, and provide, through the differential of the hypersurface itself (which is differentiable) the most valuable knowledge in the future, under a Domain and given time series. The formalization of the calculation would be given by the formula:

$$I_\gamma(t) = \int_{H(t)} F(\gamma(x)) \cdot |\gamma'(x)| dx$$

The usefulness of frontier points is given by how they will relate knowledge of a given domain with other knowledge outside. Therefore, we must analyze intelligence functions from a temporal perspective, obtaining a formula dependent on the parameter selected for the auxiliary function:

$$I(x) = \int_{H(t)} F(x) dt$$

That is why the definition and choice of the auxiliary function is so important when modeling these behaviors in a dynamic system.

7 Collective Intelligence

Definition 7: For each agent i of the information system, an intelligence function with an auxiliary function is defined as:

$$I_i(t) = \int_{H(t)} F_i(x) dx$$

Proposition 7: Any agent of the system can know all the intelligence functions of all the other agents setting a sufficiently restricted domain.

Proof: Since any function of intelligence is obviously differentiable (since it is a primitive) and locally bijective, we can apply the inverse function theorem to obtain, given the starting hypersurface and results in a local domain, a local inverse, and with it an explicit local equation of $I_i(t)$. Thus, the result is concluded.

It now makes sense to define the concept of Collective Intelligence.

Definition 8: In the above conditions, the function:

$$I_c(t) = sup \int_{H(t)} |\{F_i(x)\}|dx$$

x defined as Collective Intelligence. The reason for taking the absolute value of the auxiliary function is that the positive and negative values of the field integral cancel out, and it is necessary to consider the global flow through the hypersurface, regardless of its sign.

That is, the function of intelligence that has as auxiliary function the supreme function of all the intelligence functions of the system agents. The most important property that Cognitive Intelligence verifies is the following inequality:

Theorem 2: For any set of agents, with any set of intelligence functions and any domain:

$$|I_i(t)| \leq |I_c(t)|$$

is verified.

Proof: Because of the supreme function's properties, the following has to be

$$|I_i(t)| = \left| \int_{H(t)} F_i(x)dx \right| \leq sup|I_i(t)|$$

And now, applying a measure theory result [15]:

$$sup|I_i(t)| = sup\left| \int_{H(t)} F_i(x)dx \right| \leq sup \int_{H(t)} |\{F_i(x)\}|dx = |I_c(t)|$$

thus proving our own result.

The interpretation of the previous theorem is: Collective Intelligence always obtains superior performance of the knowledge processed through value criteria than any function of intelligence of a single agent (or of a restricted part of the set of agents).

8 Extending Intelligence: Wisdom

The last level of cognitive inference of the mathematical model is a result that applies to the functions of intelligence to obtain a generalization of the model into non-parameterized places, even places of the parameterization space that are not an image of any element of the σ-algebra.

Theorem 3: Every intelligence function defined in a domain, in particular Collective Intelligence, supports a continuous extension to all the parameterization space containing H(t).

Proof: To begin with, an intelligence function is a linear form with respect to its auxiliary function, since in the space of all the continuous functions defined on a compact set (the fixed domain), the integral defined in that compact is a linear form.

In these conditions, it makes sense to apply the Hahn-Banach Theorem [11], which states that a linear form defined in a vector subspace on a field and bounded by a sublinear function (which it fulfills, since the functions are bounded by construction) then there exists a linear extension $\hat{\mathrm{I}} : P \to \mathbb{R}$ from I to the whole space P, i.e. there is a linear function Î such that:

$$\hat{\mathrm{I}}(x) = I(x) \forall x \in P$$

This concludes the result.

Corollary 3: It is possible to know properties and values associated with knowledge or elements of space $\Omega(t)$ that do not yet exist.

Proof: In the case that $F(x)$ is the identity function and $x = t$, the function to be extended by Theorem 3 would be the hypersurface itself as a function of time, since

$$I(x) = I(t) = \int_{H(t)} 1.dt = H(t)$$

Therefore, cognitive and intelligence factors of elements of (t) that do not yet exist could be known.

Definition 9: A wisdom function is defined as the Hahn - Banach extension of an intelligence function.

9 Conclusions

Given the results of the model described, it makes sense to establish some conjectures that arise.

Conjecture 1: Is it possible to define an explicit method to obtain a function of wisdom from a given intelligence function?

It must be said that in the proof of Hahn Banach's Theorem the extension Î in general is not unique and the proof, which uses the Zorn lemma, gives no method to find it. In fact, even if it were possible to find explicitly a function of wisdom from a given intelligence function, it would be highly improbable to find a general method.

Conjecture 2: Is it possible to obtain values of non-measurable sets?

This conjecture corresponds to a vacuum in model theory, related to non-measurable sets. For any non-trivial measure there are sets within the σ-algebra that can not be measured under any measure construction. These sets, such as those described in

Vitali's Theorem [12, 13], can generate distortions in the valuation of certain transformations, as with Banach-Tarski Paradox. Fortunately, these situations only occur when the elements of σ-algebra are not explicitly constructed, which in our model does not occur, since a set of Vitalican not be constructed explicitly. However, it makes sense to wonder if these sets can receive values by a function of wisdom that extends the hypersurface of values to non-measurable sets.

Conjecture 3: In an environment of a massive amount of system agents, can Collective Intelligence be subject to indetermination?

The motivation of this conjecture comes from the existence of artificial agents. Thus, an agent that is not limited to making changes per second in intelligence functions can lead the formulation of Collective Intelligence to include two different expressions of the same intelligence function, and even two simultaneous expressions of Collective Intelligence, which could have unpredictable consequences for the model in a hypothetical practical implementation [14].

References

1. Guillén, P.G.: Topological proof of the computability of the algorithm based on the morphosyntactic distance. In: Conference Series of the 9th Annual International Conference on Computer Science and Information Systems, COM2013-0595 (2013)
2. Serradilla, F., Villa, E., De Santos, A., Guillén, P.G.: Semantic construction of an univocal language. ITHEA: Inf. Theor. Appl. **19**(3), 211 (2012)
3. De Santos, A., Villa, E., Serradilla, F., Guillén, P.G.: Construction of morphosyntactic distance on semantic structures. ITHEA: Inf. Theor. Appl. **19**(4), 336 (2012)
4. Menger, C.: Principles of Economics. Ludwig von Mises Institute, Auburn (1976). 120 p.
5. Frankland, P.W., Bontempi, B.: The organization of recent and remote memories. Nat. Rev. Neurosci. **6**(2), 119–130 (2005)
6. Rugg, M., Yonelinas, A.P.: Human recognition memory: a cognitive neuroscience perspective. Trends Cogn. Sci. **7**(7), 313–319 (2003)
7. Shannon, C.E., Weaver, W.: The Mathematical Theory of Communication. University of Illinois Press, Urbana (1964). 125 p.
8. Moreiro González, J.A.: Aplicaciones al análisis automático del contenido provenientes de la teoría matemática de la información. Anales de documentación **5**, 273–286 (2002)
9. Chung, F.: Graph theory in the information age. Not. AMS **57**(6), 726–732 (2010)
10. Bourbaki, N.: Integration I. Chap. III–IV. Springer, Heidelberg (2004)
11. Narici, L., Beckenstein, E.: The Hahn-Banach theorem: the life and times. Topol. Appl. **77**, 193–211 (1997)
12. Foreman, M., Wehrung, F.: The Hahn-Banach theorem implies the existence of a non-Lebesgue measurable set. Fundamenta Mathematicae **138**, 13–19 (1991)
13. Herrlich, H.: Axiom of Choice. Springer, Heidelberg (2006). 120 p.
14. Bosyk, G.M.: Más allá de Heisenberg. Relaciones de incerteza tipo Landau-Pollak y tipo entrópicas (2014). 113 p.
15. Hawkins, T.: The Lebesgue's Theory of Integration. University of Wisconsin Press, Madison (1970)