A Definition of Entropy based on Qualitative Descriptions

Llorenç Roselló and **Francesc Prats** and **Mónica Sánchez**
Polytechnical University of Catalonia, Barcelona, Spain
e-mail: {llorenc.rosello, francesc.prats, monica.sanchez}@upc.edu

Núria Agell *
Esade, Ramon Llull University, Barcelona, Spain
e-mail: nuria.agell@esade.edu

Abstract

A new concept of generalized absolute orders of magnitude qualitative spaces is introduced in this paper. The new structure makes it possible to define sets of qualitative labels of any cardinality, and is consistent with the classical structure of qualitative spaces of absolute orders of magnitude and with the classical interval algebra. In addition, the algebraic structure of these spaces ensures initial conditions for adapting measure theory to a qualitative environment. This theory provides the appropriate framework in which to introduce the concept of entropy and, consequently, the opportunity to measure the gain or loss of information when working within qualitative spaces. The results obtained are significant in terms of situations which arise naturally in many real applications when dealing with different levels of precision.

INTRODUCTION

Qualitative Reasoning (QR) is a subarea of Artificial Intelligence that seeks to understand and explain human beings' ability for qualitative reasoning (Forbus 1996), (Kuipers 2004). The main objective is to develop systems that permit operating in conditions of insufficient numerical data or in the absence of such data. As indicated in (Travè-Massuyès and Dague 2003), this could be due to both a lack of information as well as to an information overload.

A main goal of Qualitative Reasoning is to tackle problems in such a way that the principle of relevance is preserved; that is to say each variable has to be valued with the level of precision required (Forbus 1984). It is not unusual for a situation to arise in which it is necessary to work simultaneously with different levels of precision, depending on the available information, in order to ensure interpretability of the obtained results. To this end, the mathematical structures of Orders of Magnitude Qualitative Spaces (OM) were introduced.

*This work has been partly funded by MEC (Spanish Ministry of Education and Science) AURA project (TIN2005-08873-C02). Authors would like to thank their colleagues of GREC research group of knowledge engineering for helpful discussions and suggestions.

The word *information* appears constantly in QR. However, its meaning is as yet undefined within a qualitative context. The implicit and explicit use of the term and concept addresses the need to define and, perhaps paradoxically, to quantify them.

In this work it is presented a way of measuring the amount of information of a system when using orders of magnitude descriptions to represent it. Taking into account that the entropy can be used to measure the information, this work is intended to be a first step towards this measure by means of orders of magnitude qualitative spaces.

The concept of entropy has its origins in the nineteenth century, particularly in thermodynamics and statistics. This theory has been developed from two aspects: the macroscopic, as introduced by Carnot, Clausius, Gibbs, Planck and Caratheodory and the microscopic, developed by Maxwell and Boltzmann (Rokhlin 1967). The statistical concept of Shannon's entropy, related to the microscopic aspect, is a measure of the amount of information (Shannon 1948),(Cover and Thomas 1991).

In order to define the concept of information within the QR framework, this paper adapts the basic principles of Measure Theory (Halmos 1974), (Folland 1999) to give OM a structure in which to define the concept of entropy, and, consequently, the concept of information.

Section 2 defines the concept of generalized absolute orders of magnitude qualitative spaces. In Section 3, the algebraic structure of these spaces is analyzed in order to ensure initial conditions in which to adapt the Measure Theory. A measure and the concept of entropy in the generalized absolute orders of magnitude spaces are given in section 4 and 5 respectively. The paper ends with several conclusions and outlines some proposals for future research.

GENERALIZED ABSOLUTE ORDERS OF MAGNITUDE QUALITATIVE SPACES $\mathbb{S}_g^*$

The classical version of the qualitative orders of magnitude that appears in (Travè-Massuyès and Dague 2003) is an abstraction of intuitive concepts of “very small ”,“small”,“big”, or “very hot”, “hot”, etc., i.e. an abstraction of concepts with which human beings reason. This abstraction is done through the introduction of *qualitative labels* in a way that defines a finite and discrete set of labels representing the above concepts. This paper proposes a further step towards the generalization of qualitative orders of magnitude. This generalization makes it possible to define orders of magnitude as either a discrete or continuous set of labels, providing the theoretical basis on which to develop a Measure Theory in this context.

Definition 1 *Let X be a non-empty set, I a subset of $\mathbb{R}$, and $B : I \to \mathcal{P}(X)$ an injective function. Then each $B(t) = B_t \subset X$ is a* generalized basic label on X *and the set $\mathcal{S}$ of* generalized basic labels on X *is*

$$\mathcal{S} = \{B_t \mid t \in I\}.$$

Note that if $t \neq t'$, then $B_t \neq B_{t'}$.

Definition 2 *If $i, j \in I$, with $i < j$, the* generalized non-basic label $[B_i, B_j)$ *is defined by*

$$[B_i, B_j) = \{B_t \mid t \in I, i \leq t < j\}.$$

In the case $i = j \in I$, the convention $[B_i, B_i) = \{B_i\}$ will be used. If necessary, $[B_i, B_i) = \{B_i\}$ can be identified with the basic label B_i.

Definition 3 *If $i \in I$, the* generalized non-basic label $[B_i, B_\infty)$ *is defined by*

$$[B_i, B_\infty) = \{B_t \mid t \in I, i \leq t\}.$$

Note that B_∞ is a symbol, not a basic label.

Definition 4 *The set of* Generalized Orders of Magnitude $\mathbb{S}_g^*$ *is:*

$$\mathbb{S}_g^* = \{\emptyset\} \cup \{[B_i, B_j) \mid i, j \in I,\ i \leq j\} \cup \{[B_i, B_\infty) \mid i \in I\}.$$

In this definition of $\mathbb{S}_g^*$ the basic label B_i has been identified with the singleton $\{B_i\}$.

It is important to remark that the function $B : I \to \mathcal{P}(X)$ determines the elements of $\mathcal{S}$ and $\mathbb{S}_g^*$, and the cardinal of the set $I \subset \mathbb{R}$ determines the cardinal of $\mathcal{S}$ and therefore the cardinal of $\mathbb{S}_g^*$.

The *classical orders of magnitude qualitative spaces* (Travè-Massuyès and Dague 2003) verifies the conditions of the generalized model that has just been introduced. This model are build from a set of ordered basic qualitative labels determined by a partition of the real line.

Let X be the real interval $[a_1, a_n)$, and a partition of this set given by $\{a_2, \ldots, a_{n-1}\}$, with $a_1 < a_2 < \ldots < a_{n-1} < a_n$. The set of basic labels is

$$\mathcal{S} = \{B_1, \ldots, B_{n-1}\},$$

where, for $1 \leq i \leq n-1$, B_i is the real interval $[a_i, a_{i+1})$. The set of indexes is $I = \{1, 2, \ldots, n-1\}$.

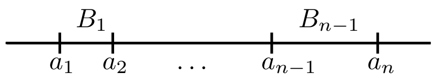

Figure 1. Classical aualitative labels $\mathbb{S}_n$

For $1 \leq i < j \leq n-1$ the non-basic label $[B_i, B_j)$ is:

$$[B_i, B_j) = \{B_i, B_{i+1}, \ldots, B_{j-1}\},$$

and it is interpreted as the real interval $[a_i, a_j)$.

For $1 \leq i \leq n-1$ the non-basic label $[B_i, B_\infty)$ is:

$$[B_i, B_\infty) = \{B_i, B_{i+1}, \ldots, B_{n-1}\},$$

and it is interpreted as the real interval $[a_i, a_n)$.

The complete universe of description for the Orders of Magnitude Space is the set

$$\mathbb{S}_n = \{\, [B_i, B_j) \mid B_i, B_j \in \mathcal{S}, i \leq j\} \cup \{\, [B_i, B_\infty) \mid B_i \in \mathcal{S}\},$$

which is called the absolute orders of magnitude qualitative space with granularity n, also denoted $OM(n)$. In this case, $\mathbb{S}_g^* = \{\emptyset\} \cup \mathbb{S}_n$.

There is a partial order relation $\leq_P$ in $\mathbb{S}_n$ “to be more precise than”, given by:

$$L_1 \leq_P L_2 \Longleftrightarrow L_1 \subset L_2.$$

The least precise label is denoted by ? and it is the label $[B_1, B_\infty)$, which corresponds to the interval $[a_1, a_n)$.

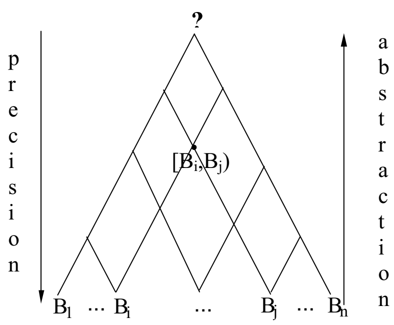

Figure 2. The space $\mathbb{S}_n$

This structure permits working with all different levels of precision from the label ? to the basic labels.

In some theoretical works, orders of magnitude qualitative spaces are constructed by partitioning the whole real line $(-\infty, +\infty)$ instead of a finite real interval $[a_1, a_n)$. However, in most real world applications involved variables do have a lower bound a_1 and an upper bound a_n, and then values less than a_1 or greater than a_n are considered as outliers and they are not treated like any other.

The classical sign algebra $\mathcal{S} = \{-, 0, +\}$ was the first absolute orders of magnitude space considered by the QR community. It corresponds to the case $\mathcal{S} = \{B_{-1} = (-\infty, 0), B_0 = \{0\}, B_1 = (0, +\infty)\}$. The sign algebra is obtained via a partition of the real line given by an unique landmark 0. The classical orders of magnitude qualitative spaces are built from partitions via a set of landmarks $\{a_2, \ldots, a_{n-1}\}$, and the classical interval algebra is built from the finest partition of the real line whose landmarks are all real numbers.

It is important to remark the significance of the presented mathematical formalism in the sense that it permits to lump together a family of $\mathbb{S}_g^*$ forming a continuum from the sign algebra $\mathcal{S} = \{-, 0, +\}$ to the interval algebra corresponding to $\mathcal{S} = \mathbb{R}$.

THE MEASURE SPACE $(\mathcal{P}(X), \Sigma(\mathbb{S}_g^*), \mu^*)$

To introduce the classical concept of entropy by means of qualitative orders of magnitude spaces, Measure Theory is required. This theory seeks to generalize the concept of "length", "area"and "volume", understanding that these quantities need not necessarily correspond to their physical counterparts, but may in fact represent others. The main use of the measure is to define the concept of integration for orders of magnitude spaces. First, it is necessary to define the algebraic structure on which to define a measure.

Definition 5 *A class of sets $\Im$ is called a* semi-ring *if the following properties are satisfied:*

1. $\emptyset \in \Im$.
2. *If $A, B \in \Im$, then $A \cap B \in \Im$.*
3. *If $A, B \in \Im$, $A \subset B$, then $\exists n \in \mathbb{N}, n \geq 1$ and $\exists D_1, D_2, \ldots, D_n$ such that $A = D_0 \subset D_1 \subset \ldots \subset D_n = B$, with $D_k - D_{k-1} \in \Im, \forall k \in \{1, \ldots, n\}$.*

Proposition 1 *$\mathbb{S}_g^*$ is a semi-ring.*

Proof:

1. $\emptyset \in \mathbb{S}_g^*$ by definition.
2. If $[B_i, B_j), [B_k, B_l) \in \mathbb{S}_g^*$, it is trivial to check that $[B_i, B_j) \cap [B_k, B_l) \in \mathbb{S}_g^*$, taking into account the relative position between the real intervals $[i, j)$ and $[k, l)$. Analogously, in the case of intersections $[B_i, B_j) \cap [B_k, B_\infty)$ or $[B_i, B_\infty) \cap [B_k, B_\infty)$.
3. If $[B_i, B_j), [B_k, B_l) \in \mathbb{S}_g^*$ such that $[B_i, B_j) \subset [B_k, B_l)$, then two cases are considered:
   (a) If $B_k = B_i$ or $B_l = B_j$, it suffices to take $D_0 = [B_i, B_j)$ and $D_1 = [B_k, B_l)$.
   (b) Otherwise, take $D_0 = [B_i, B_j), D_1 = [B_i, B_l)$ and $D_2 = [B_k, B_l)$.
   The cases $[B_i, B_j) \subset [B_k, B_\infty)$ and $[B_i, B_\infty) \subset [B_k, B_\infty)$ are proved in a similar way.

□

Definition 6 *A class $\mathcal{A}$ of subsets of a non-empty set X is called an* algebra *when it contains the finite unions and the complements of its elements. If finite unions are replaced by countable unions, it is called a σ-algebra.*

The smallest σ-algebra that contains $\mathbb{S}_g^* \subset \mathcal{P}(X)$ is called the σ-algebra generated by $\mathbb{S}_g^*$, denoted by $\Sigma(\mathbb{S}_g^*)$.

Definition 7 *Let X be a non-empty set and $\mathcal{C} \subset \mathcal{P}(X)$, with $\emptyset \in \mathcal{C}$. A* measure *on $\mathcal{C}$ is an application $\mu : \mathcal{C} \to [0, +\infty]$ satisfying the following properties:*

1. $\mu(\emptyset) = 0$.
2. *For any sequence $(E_n)_{n=1}^{\infty}$ of disjoint sets of $\mathcal{C}$ such that $\cup_{n=1}^{+\infty} E_n \in \mathcal{C}$, then*
$$\mu(\bigcup_{n=1}^{+\infty} E_n) = \sum_{n=1}^{+\infty} \mu(E_n).$$

Any measure μ on the whole $\mathcal{P}(X)$, when it is restricted to $\mathbb{S}_g^*$, gives a measure on $\mathbb{S}_g^*$.

Definition 8 *Let μ be a measure on $\mathbb{S}_g^*$. The* outer measure *on an arbitrary subset A of X is defined by:*
$$\mu^*(A) = \inf\{\sum_{k\in\mathbb{N}} \mu([B_{s_k}, B_{t_k})),\ A \subset \bigcup_{k\in\mathbb{N}} [B_{s_k}, B_{t_k})\}.$$

Carathéodory theorem (Halmos 1974) assures μ^* of definition 7 is a measure on $\Sigma(\mathbb{S}_g^*)$, and $(\mathcal{P}(X), \Sigma(\mathbb{S}_g^*), \mu^*)$ is called a measure space. It is proved that, since $\mathbb{S}_g^*$ is a semi-ring, $\mu^*_{|\mathbb{S}_g^*} = \mu$.

In this measure space an integration with respect μ^* can be defined. Because of the fact that $\mu^*_{|\mathbb{S}_g^*} = \mu$, in any integration on $\mathbb{S}_g^*$ the measure μ^* can be replaced by μ.

ENTROPY BY MEANS OF $\mathbb{S}_g^*$

Once the integration in $\mathbb{S}_g^*$ has been defined, entropy can then be considered. To introduce the concept of entropy by means of qualitative orders of magnitude, it is necessary to consider the qualitativization function between the set to be qualitatively described and the space of qualitative labels, $\mathbb{S}_g^*$.

To simplify the notation, let us express with a calligraphic letter the elements in $\mathbb{S}_g^*$; thus, for example, elements $[B_i, B_j)$ or $[B_i, B_\infty)$ shall be denoted as $\mathcal{E}$.
Let Λ be the set that represents a magnitude or a feature that is qualitatively described by means of the labels of $\mathbb{S}_g^*$. Since Λ can represent both a continuous magnitude such as position and temperature, etc., and a discrete feature such as salary and colour, etc., Λ could be considered as the range of a function

$$a : I \subset \mathbb{R} \to Y,$$

where Y is a convenient set. For instance, if a is a room temperature during a period of time $I = [t_0, t_1]$, Λ is the range of temperatures during this period of time. Another example can be considered when $I = \{1, \ldots, n\}$ and $\Lambda = \{a(1), \ldots, a(n)\}$ are n number of people whose eye colour we aim to describe. In general, $\Lambda = \{a(t) = a_t \mid t \in I\}$.

The process of qualitativization is given by a function

$$Q : \Lambda \to \mathbb{S}_g^*,$$

where $a_t \mapsto Q(a_t) = \mathcal{E}_t =$ minimum label (with respect to the inclusion $\subset$) which describes a_t, i.e. the most precise qualitative label describing a_t. All the elements of the set $Q^{-1}(\mathcal{E}_t)$ are "representatives" of the label $\mathcal{E}_t$ or "are qualitatively described" by $\mathcal{E}_t$. They can be considered qualitatively equal.

The function Q induces a partition in Λ by means of the equivalence relation:

$$a \sim_Q b \Longleftrightarrow Q(a) = Q(b).$$

This partition will be denoted by $\Lambda/\sim_Q$, and its equivalence classes are the sets $Q^{-1}(Q(a_j)) = Q^{-1}(\mathcal{E}_j)$, $\forall j \in J \subset I$. Each of these classes contains all the elements of Λ which are described by the same qualitative label.

Definition 9 *Let μ be a measure on $\mathbb{S}_g^*$ such that*

$$\int_{\bigcup_{i \in I} \{B_i\}} d\mu = 1.$$

The entropy H with respect the partition $\Lambda/\sim_Q$ is the integral:

$$H(\Lambda/\sim_Q) = -\int_{Q(\Lambda)} \log \mu \, d\mu, \tag{1}$$

where $Q(\Lambda)$ is the set of labels mapped by Q (logarithms are to the base 2).
The expression (1) can be written as:

$$H(\Lambda/\sim_Q) = -\sum_{j \in J} \log(\mu(\mathcal{E}_j))\mu(\mathcal{E}_j). \tag{2}$$

As in most definitions of entropy, it gives a measure of the amount of information. In Definition 9 entropy can be interpreted as the measure of the amount of information that provides the knowledge of Λ by means of Q.

Nevertheless, the inner features of the orders of magnitude structure considered introduce some differences between the entropy defined in (1) and the entropy defined by Rokhlin (Rokhlin 1967) and Shannon (Shannon 1948), as can be seen in the following example:

Example 1 Suppose that Q maps each element of Λ to the same label $\mathcal{E} \in \mathbb{S}_g^*$; then the induced partition $\Lambda/\sim_Q$ contains only one class equal to Λ and the entropy defined in equation (1) is $H(\Lambda/\sim_Q) = -\mu(\mathcal{E}) \log \mu(\mathcal{E})$. In the classical interpretation of the entropy, the knowledge about Λ induced by this particular Q will lead to an entropy equal to zero, because in the given situation it is understood that this trivial partition of Λ provides no information at all. On the contrary, in the approach that has been presented in this paper, although Q map the whole set to the same label it could give a certain information about Λ: the intrinsic information provided by the measure of the label itself.

Two different measures that show this fact are considered in the following examples. On the one hand, the first differs from Shannon's classical interpretation of entropy as noted in Example 1: although Q map each element of Λ to the same label $\mathcal{E} \in \mathbb{S}_g^*$ entropy is not equal to zero . On the other, the entropy corresponding to Example 3 behaves like the classical interpretation of Shannon and Rokhlin, in the sense just discussed. Example 2 takes into account the lengths of the intervals corresponding to the labels, and Example 3 is related to the cardinality of the set of representatives of each label.

Example 2 Let us define a particular measure μ on $\{\emptyset\} \cup \mathbb{S}_n$ as follows:
For the basic labels $B_i = [a_i, a_{i+1})$, whith $i = 1, \ldots, n-1$, let

$$\mu(B_i) = \frac{a_{i+1} - a_i}{a_n - a_1}.$$

This measure is proportional to the knowledge of imprecision about the magnitude and it is normalized with respect to the "basic" known range given by the length $a_n - a_1$. For non-basic labels the measure is, for $i, j = 1, \ldots, n-1$, $i < j$:

$$\mu([B_i, B_j)) = \sum_{k=i}^{j-1} \mu(B_k) = \frac{a_j - a_i}{a_n - a_1},$$

and for $i = 1, \ldots, n-1$:

$$\mu([B_i, B_\infty)) = \sum_{k=i}^{n-1} \mu(B_k) = \frac{a_n - a_i}{a_n - a_1}.$$

Elements of Λ represented by quite precise labels will provide a bigger contribution to entropy H than those who are represented by less precise labels. Considering the particular case in which Q maps all the elements of Λ to the same label: $Q(\Lambda) = \{\mathcal{E}\}$, then $\Lambda/\sim_Q = \Lambda$ and $H(\Lambda/\sim_Q) = -\mu(\mathcal{E}) \log(\mu(\mathcal{E})) \neq 0$.

Example 3 Another interpretation of the entropy defined in equation (1) is obtained by defining another measure μ over $\{\emptyset \cup \mathbb{S}_n$ as follows: For each $\mathcal{E}_t \in \{\emptyset\} \cup \mathbb{S}_n$,

$$\mu(\emptyset) = 0,\ \mu(\mathcal{E}_t) = \mathrm{card}(Q^{-1}(\mathcal{E}_t))/\mathrm{card}(\Lambda).$$

This case recovers the classical interpretation of Shannon and Rokhlin in the sense that if Q maps all the elements of Λ to the same label, then the partition does not give information of Λ because the entropy is $H(\Lambda/\sim_Q\ ) = -1 \cdot \log 1 = 0$. Moreover, the entropy reaches its maximum when different elements of Λ are mapped to different labels $\mathcal{E}_t \in \mathbb{S}_n$, i.e., when Q is an injective map from Λ onto $\mathbb{S}_n$. This maximum is $H(\Lambda/\sim_Q\ ) = \log(\mathrm{card}\,\Lambda)$.

CONCLUSION AND FUTURE WORK

This paper introduces the concept of entropy by means of absolute orders of magnitude qualitative spaces. This entropy measures the amount of information of a system when using orders of magnitude descriptions to represent it.

In order to define the concept of entropy within Qualitative Reasoning framework, this paper adapts the basic principles of Measure Theory to give the space of absolute orders of magnitude the necessary structure. With the presented structure, we obtain a family of qualitative spaces forming a continuum from the sign algebra to the classical interval algebra.

From a theoretical point of view, future research could focus on two lines. On the one hand, it could focus on the comparison of the given entropy with the macroscopic concept of Caratheodory entropy. On the other hand, the adaptation of Measure Theory provides the theoretical framework in which developing a rigorous analytical study of functions between orders of magnitude spaces. The continuity and differentiability of these functions will allow the dynamical study of qualitatively described processes.

Within the framework of applications, this work and its related methodology will be orientated towards the modelization and the resolution of financial and marketing problems. Regarding financial problems, the concept of entropy will facilitate the study of the evolution and variation of the financial ratings. On the other hand, entropy as a measurement of coherence and reliability is useful in group decision-making problems arising from retail marketing applications.

Moreover, the introduced entropy will allow defining a conditional entropy in this framework, which in turn will allow considering the Rokhlin distance to be used in decision-making problems of ranking and selection of alternatives.

References

Cover, T. M., and Thomas, J. A. 1991. *Elements of Information Theory*. Wiley Series in Telecomunications.

Folland, G. 1999. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley-Interscience Series of Texts, Monographs, and Tracks. John Wiley & Sons, Inc.

Forbus, K. 1984. Qualitative process theory. *Artificial Intelligence* 24:85–158.

Forbus, K. 1996. *Qualitative Reasoning*. CRC Hand-book of Computer Science and Engineering. CRC Press.

Halmos, P. R. 1974. *Measure Theory*. Springer-Verlag.

Kuipers, B. 2004. Making sense of common sense knowledge. *Ubiquity* 4(45).

Rokhlin, V. 1967. Lectures on the entropy of eeasure ereserving eransformations. *Russian Math. Surveys* 22:1 – 52.

Shannon, C. E. 1948. A mathematical theory of communication. *The Bell System Technical Journal* 27:379 – 423.

Travè-Massuyès, L., and Dague, P., eds. 2003. *Modèles et raisonaments qualitatifs*. Hermes Science (París).