International Journal of Approximate Reasoning ••• (••••) •••–•••

Contents lists available at ScienceDirect

International Journal of Approximate Reasoning

www.elsevier.com/locate/ijar

Analogical proportions: From equality to inequality ☆

Henri Prade [a,b,*], Gilles Richard [a]

[a] IRIT, CNRS & Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 09, France
[b] QCIS, University of Technology, Sydney, Australia

ARTICLE INFO

Article history:
Received 22 December 2017
Received in revised form 1 June 2018
Accepted 15 July 2018
Available online xxxx

Keywords:
Analogical proportion
Similarity
Dissimilarity
Boolean logic
Multiple-valued logic
Algorithmic complexity

ABSTRACT

Analogical proportions are statements of the form a is to b as c is to d. Such expressions compare the pair (a, b) with the pair (c, d). Previous papers have developed logical modelings of such proportions both in Boolean and in multiple-valued settings. They emphasize a reading of the proportion as "the way a and b differ is the same as c and d differ". The ambition of this paper is twofold. The paper first provides a deeper understanding and further justifications of the Boolean modeling, before introducing analogical inequalities, where "as" is replaced by "as much as" in the comparison of two pairs. From an abstract viewpoint, analogical proportions are supposed to obey at least three postulates expressing reflexivity, symmetry, and stability under central permutation. Nevertheless these postulates are not enough to determine a single model and a minimality condition has to be added as shown in this paper. These models are organized in a complete lattice based on set inclusion. This leads us to discuss lower and upper approximations of the minimal model. Apart from being minimal, this model can also be evaluated in terms of Kolmogorov complexity via an expression reflecting the intended meaning of analogy. We show that the six Boolean patterns of the minimal model that make Boolean analogy true minimize this expression. Besides, analogical proportions extend to 4-tuples of Boolean vectors. This enables us to explain why analogical proportions also reads in terms of similarity (rather than difference, i.e., dissimilarity): a and d share the same presence or absence of features as b and c. Moreover, we establish a link between analogical proportion and Hamming distances between components of the proportion. We also emphasize that analogical proportions are pervasive in any comparison of two vectors a and d that implicitly induce the existence of "intermediary" vectors b and c forming together such a proportion. The similarity reading and the dissimilarity reading of a Boolean analogical proportion are no longer equivalent in the multiple-valued setting, where they give birth to two distinct options that are recalled. These options are also discussed with respect to their capability to handle so-called "continuous" logical proportions of the form a is to b as b is to c involving some idea of "betweenness". In all the previously investigated issues, the pairs involved in the 4-tuples were compared via equalities of similarities or equalities of dissimilarities. This observation suggests to also consider statements of the form "a is to b at least as much as c is to d", leading to the concept of "analogical inequalities". Thus, instead of expressing equality between differences or similarities, as it is the case for analogical proportions, it is also interesting to express inequalities between such differences or similarities. Starting from the modeling of analogical proportions, we investigate the logical modeling of analogical

☆ This paper is part of the Virtual special issue on the 14th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2017), Edited by Alessandro Antonucci, Laurence Cholvy and Odile Papini.

* Corresponding author at: IRIT, CNRS & Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex 09, France.
E-mail addresses: prade@irit.fr (H. Prade), richard@irit.fr (G. Richard).

https://doi.org/10.1016/j.ijar.2018.07.005

inequalities, both in the Boolean and in the multiple-valued cases, and discuss their potential use in relation with some recent related work in computer vision.

1. Introduction

Comparative thinking plays a key role in our assessment of reality. This has been recognized for a long time. Making comparison is closely related to similarity judgment [45] and analogy making [12,17]. Analogical proportions, i.e., statements of the form a is to b as c is to d, usually denoted $a : b :: c : d$, provides a well-known way for expressing a comparative judgment between two pairs (a, b) and (c, d); see, e.g., [14,15]. Such a statement suggests that the comparison (in terms of similarity and dissimilarity) of the elements of pair (a, b) yields the same kind of result as when comparing the elements of pair (c, d) [39].

Analogical proportions constitute a key notion for formalizing analogical inference by relying on the following principle: if such proportions hold on a noticeable subset of known features used for describing the four items, the proportion may still hold on other features as well, which may help guessing the unknown values of d on these other features from their values on a, b, and c. The interest of such inference mechanism has been recently pointed out in machine learning for classification problems [2,28,4], and in visual multiple-class categorization tasks for handling pieces of knowledge about semantic relationships between classes. More precisely in this latter case, analogical proportions are used for expressing analogies between pairs of concrete objects in the same semantic universe and with similar abstraction level, and then this gives birth to constraints that serve regularization purposes [18]. Besides, the power of analogical proportion-based inference has been also illustrated on the solving of IQ tests [6].

Different formal modelings of analogical proportions have been proposed in the last decades. Quite early, a theory of analogical reasoning, where elements are represented as points in multidimensional Euclidean spaces, and analogical proportions are represented by parallelograms in such spaces, has been proposed in [41]. This geometric view is at work in the above-mentioned reference in visual categorization. An empirical modeling of analogy making, where the fourth term d of an analogical proportion $a : b :: c : d$ to be completed is obtained by minimization of the difference of changes between a and b and between c and d is at work in the programs ANALOGY [10] and COPYCAT [16]. Later on, a machine learning-oriented view where analogical proportions are represented in terms of Kolmogorov algorithmic complexity has been presented in [5]. A similar, but simplified modeling, still expressing that a and b differ as c and d differ, can be found in [1], where the complexities of the target and source universes have not to be taken into account, since they are identical in this latter case. Quite more recently, a set of various algebraic modelings of analogical proportions have been introduced and discussed in [25,29,30,46]. Following these works, a logical modeling has been proposed [31,32]. This logical modeling makes clear that the analogical proportion holds if and only if a differs from b as c differs from d and vice-versa. This fits quite well with what is suggested by the usual reading of the proportion that states that "a is to b as c is to d", where "a is to b" (resp. "c is to d") refers to an implicit pairwise comparison, and the central "as" to an identity. This leads to a Boolean truth table for $a : b :: c : d$ which makes the expression true for six patterns of values of the 4-tuple a, b, c, d among $2^4 = 16$ possible patterns. It can easily be checked that the expected postulates (reflexivity, symmetry, formal permutation) are satisfied by the modeling. However, one may wonder if other modelings would make sense for an analogical proportion, and if other justifications could be found for the above-mentioned modeling. This is one of the goals of this paper.

The paper first investigates new justifications of the Boolean expression of an analogical proportion. First, starting from the core postulates supposed to be satisfied by an analogical proportion, and agreed by everybody for a long time, we exhibit all the Boolean models compatible with them. There are several ones, but the smallest model is the basic Boolean expression of an analogical proportion previously proposed. This smallest model is indeed characterized by the six expected Boolean patterns. Another understanding of analogical proportion, in terms of *similarity*, can be stated as "what a and d have in common (positively and negatively), b and c have it also". It corresponds to a Boolean formula that turns to be equivalent to the one induced by the difference-based reading, since the same truth-table is obtained in both cases, as observed for about ten years now [32]. We also provide a direct proof and an intuitive explanation of this fact.

Moreover, we try to evaluate the cognitive significance of the proposed Boolean modeling of an analogical proportion in terms of algorithmic complexity (i.e., Kolmogorov complexity) and show that it is also minimal among all Boolean patterns with respect to an algorithmic complexity-based definition of an analogical proportion. Indeed algorithmic complexity measures a kind of universal information content of a Boolean string. Despite its inherent uncomputability, there exist powerful tools for computing good approximations. Kolmogorov complexity has been proved to be of great value in diverse applications: for example, in distance measures [3] and classification methods, plagiarism detection, network intrusion detection [13], and in numerous other applications [27].

As already said, analogical proportions express the identity of the results of the comparisons of two pairs. We may wonder if an inequality instead of an equality would make sense as well and would be useful for expressing constraints of the form "a is to b as much as c is to d". In fact, constraints of the same kind, but stated in terms of distances, have been shown to be useful for categorization tasks in computer vision for representing pieces of knowledge stating *relative* comparisons between quadruplets of images, feature by feature [23,24]. Interestingly enough, it has been also recently

Please cite this article in press as: H. Prade, G. Richard, Analogical proportions: From equality to inequality, Int. J. Approx. Reason. (2018), https://doi.org/10.1016/j.ijar.2018.07.005