



HAL
open science

Reasoning with data - A new challenge for AI?

Henri Prade

► **To cite this version:**

Henri Prade. Reasoning with data - A new challenge for AI?. International Conference on Scalable Uncertainty Management (SUM 2016), Sep 2016, Nice, France. pp.274-288. hal-02456759

HAL Id: hal-02456759

<https://hal.science/hal-02456759>

Submitted on 27 Jan 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Open Archive Toulouse Archive Ouverte

OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is an author's version published in:
<http://oatao.univ-toulouse.fr/24865>

Official URL

DOI : https://doi.org/10.1007/978-3-319-45856-4_19

To cite this version: Prade, Henri *Reasoning with data - A new challenge for AI?* (2016) In: International Conference on Scalable Uncertainty Management (SUM 2016), 21 September 2016 - 23 September 2016 (Nice, France).

Any correspondence concerning this service should be sent to the repository administrator: tech-oatao@listes-diff.inp-toulouse.fr

Reasoning with Data - A New Challenge for AI?

Henri Prade^{1,2}(✉)

¹ IRIT, Université Paul Sabatier, 118 route de Narbonne,
31062 Toulouse Cedex 09, France
prade@irit.fr

² QCIS, University of Technology, Sydney, Australia

Abstract. Artificial intelligence (AI) traditionally deals with knowledge rather than with data (with the noticeable exception of machine learning). The term “knowledge” refers here to information with a generic flavor, while “data” refers to information pertaining to (collections of) particular cases. The formalization of reasoning patterns with data has been much less studied until now than knowledge representation and its application to knowledge-based systems and reasoning, possibly in presence of imperfect information. Data are positive in nature by manifesting the possibility of what is observed or reported, and contrast with knowledge that delimit the extent of what is potentially possible by specifying what is impossible. Reasoning from knowledge and data goes much beyond the application of knowledge to data as in expert systems. Besides, the idea of similarity naturally applies to data and gives birth to specific forms of reasoning such as case-based reasoning, case-based decision, or even case-based argumentation, interpolation, extrapolation, and analogical reasoning. Moreover, the analysis, the interpretation of data sets raise original reasoning problems for making sense of data. This article is a manifesto in favor of the study of types of reasoning which have been somewhat neglected in AI, by showing that AI should contribute to (knowledge) and data sciences, not only in the machine learning and in the data mining areas.

Keywords: Knowledge · Data · Bipolarity · Similarity · Possibility theory

1 Introduction

Historically, knowledge representation and reasoning have played a central role in artificial intelligence (AI). AI has been more a “knowledge science” than a “data science” until now. This does not mean that AI fully ignores data. Certainly not. Machine learning, which has recently become the most prominent part of AI, works with data, and is extensively used for inducing knowledge from data. Note that in this paper we use the term “knowledge” for referring to information having a *generic* flavor, while “data” pertains to collections of *particular* cases or instances.

Originally, knowledge was mainly supposed to be provided by experts. Expert systems, which were popular from the mid-seventies to the mid-nineties of the

last century, are prototypical illustrations of this state of facts. However, the idea of extracting knowledge from data has been there for a long time as well. This is especially the case with numerical methods such as Bayesian nets, fuzzy rules-based systems, or neural nets. Besides, data mining methods look for pieces of knowledge of interest including association rules that may have not been thought of before.

The aim of this paper is neither to survey methods that extract knowledge from data, nor to discuss machine learning, but rather to provide a structured overview of problems and approaches where data are directly involved in reasoning processes. Note that in this paper we use the phrase “reasoning with data”, which may cover indifferently “reasoning about data”, or “reasoning from data”. Obviously, expert systems by operating the separation of the knowledge base from the factual part of the information pertaining to the case at end, were dealing with data. But these data, called facts there, were used in isolation and were specific of the case at hand to which the generic knowledge had to be applied. The factual base was gathering only facts pertaining to a single case at once, so it was not really a data base (which usually gathers data pertaining to collections of different cases). This remark applies as well to abductive reasoning in diagnosis, where symptoms are facts pertaining to a given situation.

As in knowledge-based systems, description logics also separate knowledge (in a “TBox”) and data (in a “ABox”, which is not restricted to a single case) [2]. Then they confront knowledge and data, but do not reason from data alone. In that respect, case-based reasoning by working from a repertory of cases, shows that AI has considered the problem of reasoning from data (more precisely from tuples of data, called cases) for a long time.

Interestingly enough, research on case-based reasoning has remained apart from the main knowledge representation and reasoning trend, and has been much less oriented towards formal theoretical studies. Indeed “reasoning with data” could not benefit much of the tradition of mathematical reasoning and mathematical logic interested in the deduction of universal theorems, while applied mathematics applies generic equations to (numerical) data, as it was the case for knowledge-based systems (up to the problem of handling exceptions and uncertainty, not much encountered in mathematics).

Reasoning with data, which certainly includes case-based reasoning, is however potentially much larger, as we are trying to suggest in this paper. Moreover, reasoning with data and reasoning with knowledge are not necessarily topics that should be considered separately. Indeed reasoning from both knowledge and data may be worth doing. This is already the case in machine learning, where background knowledge, when available, is jointly used with data, or in case-based reasoning where domain knowledge may be taken into account, but not so much in deductive reasoning.

Knowledge and data are of different nature, and are not subject to have exactly the same kinds of defects, even if knowledge may be incomplete or pervaded with uncertainty, and data may be missing or suspect. Indeed knowledge may be inconsistent, or may be incoherent with respect to potential data, while

data can be accumulated without inconsistency, nor redundancy. Data cannot be inconsistent by themselves, but only with respect to integrity constraints. Still some discrepancy may exist between data, which is a matter of (dis)similarity, and may then suggest outliers among data (if the reliability of the data is not fully guaranteed). But this differs from logical inconsistency between pieces of knowledge.

Moreover, as just mentioned, the notion of similarity makes sense for data, and is often instrumental when reasoning with data, while knowledge is more concerned with subsumption. Besides, it is worth mentioning that data, even if there is a massive amount of them, may be scarce in some areas where we would need to have more information.

In fact, data are a positive form of information, while knowledge has rather a negative flavor. Indeed knowledge restricts possible worlds, and thus implicitly states what worlds are *impossible*. The more knowledge you have, the greater the restriction on the remaining possible worlds, and the closer the information is to be complete about the world of interest, provided that the knowledge you have remains consistent. This contrasts with data that by nature are diverse, but should be regarded as positive information expressing that some worlds are *really possible*, since they are observed, or reported (assuming that the sources are reliable).

The paper is a kind of manifesto¹ in favor of more unified thinking and researches on the diverse forms of use of data in reasoning, possibly jointly with knowledge. The rest of the paper is structured as follows. The next section recalls how the idea of bipolarity, which distinguishes between positive and negative information may provide a setting for reasoning with data, in a coherent manner with respect to knowledge-based reasoning. Section 3 deals with similarity-based reasoning, including case-based reasoning, case-based decision, analogical reasoning, interpolation and extrapolation; it also points out the role of data in argumentation. Section 4 surveys issues related to the need of making sense of data, of reasoning about data, rather than reasoning from data, as in the two previous sections.

2 Reasoning with Data vs. Reasoning with Knowledge: A Bipolar Issue

Pieces of knowledge are generally understood as constraints² on sets of possible worlds. “Humans are mortal” means that it is impossible to find a human who is not mortal. Thus generic information, generally referred to as knowledge, may be viewed as *negative*, in the sense that what is really stated is an *impossibility*, which by complementation defines a set of worlds that are not impossible, i.e. that are *potentially* possible. Generic knowledge may have exceptions, and thus

¹ A preliminary version of this paper exists in French [41].

² Such a view is sometime termed as being “intensional”; see Pearl [38] who opposes it, in the case of rules, to “extensional” approaches where a (decision) rule would then express the license (rather than the obligation) to do something.

the information is pervaded with uncertainty. In such a case, the situation is basically the same, but impossibility is no longer fully strong. “Generally birds fly” is to be understood as it is rather impossible, but maybe not completely impossible to find birds that cannot fly. The more knowledge we have, the more restricted the remaining set of possible worlds, by effect of the conjunctive combination of such *restrictive* pieces of information.

By contrast, if we consider the piece of data “Mary is 111 years old”, it is both a fact about Mary, and the indication that it is *possible for sure* (guaranteed possible) to live until 111 years, as long we regard the information as reliable. This type of information, based on observed, or reported cases, is not of the same nature as the claim that according to our understanding of our biological nature, it would be impossible to live more than 150 years in any case, where here living until 140 years remains just a potential possibility, as long as no case is reported. Observed facts give birth to what may be termed *positive* information. Positive information can be accumulated without any risk of inconsistency. For instance, if you want to know the price for a house having some specificities to let at a given time period, you may look to list of offers, select the ones that correspond to what you are looking for, and from them gather a collection of prices that can be regarded as possible for sure. But this does not mean that any other price would be impossible.

Possibility theory [19] (but also evidence theory [23], particular modal logics [15]) are suitable frameworks for representing both positive and negative information. Indeed the representation capabilities of possibilistic logic that extends classical logic by associating formulas with certainty levels, can be enlarged into a bipolar possibilistic setting [4, 15]. It allows the separate representation of both negative and positive information taken in the following sense. Negative information reflects what is not (fully) impossible and remains potentially possible. It induces (prioritized) constraints on where the real world is (when expressing knowledge), which can be encoded by necessity-based possibilistic logic formulas. Positive information expressing what is actually possible, is encoded by another type of formula based on a set function called guaranteed (or actual) possibility measure (which is to be distinguished from “standard” possibility measures that rather express potential possibility (as a matter of consistency with the available information)). This bipolar setting is thus of interest for representing both knowledge and reported data.

Positive information can be represented by formulas denoted $[\varphi, d]$, which expresses the constraint $\Delta(\varphi) \geq d$, where Δ denotes a measure of strong (actual) possibility [19] defined from a possibility distribution δ by $\Delta(\varphi) = \min_{\omega \models \varphi} \delta(\omega)$. Thus, the piece of positive information $[\varphi, d]$ expresses that any model of φ is at least possible with degree d (d reflects the minimal confidence in the reported observations gathered in the models of φ). More generally, let $D = \{[\varphi_j, d_j] \mid j = 1, \dots, k\}$ be a positive possibilistic logic base. Its semantics is given by the possibility distribution

$$\delta_D(\omega) = \max_{j=1, \dots, k} \delta_{[\varphi_j, d_j]}(\omega)$$

with $\delta_{[\varphi_j, d_j]}(\omega) = 0$ if $\omega \models \neg\varphi_j$, and $\delta_{[\varphi_j, d_j]}(\omega) = d_j$ if $\omega \models \varphi_j$. As can be seen, δ_D is obtained as the max-based *disjunctive* combination of the representation of each formula in D . This is in agreement with the idea that observations accumulate and are never in conflict with each other.

This contrasts with a standard possibilistic logic base $K = \{(\psi_i, c_i)\}_{i=1, \dots, m}$, which is associated with the possibility distribution π_K representing the weighted set of models of K :

$$\pi_K(\omega) = \min_{i=1, \dots, m} \max(\mu_{\|\psi_i\|}(\omega), 1 - c_i)$$

where an interpretation ω is all the less possible as it falsifies a formula ψ_i having a higher level of certainty c_i ($\mu_{\|\psi_i\|}$ is the characteristic function of the set of models of ψ_i). Each formula (ψ_i, c_i) corresponds to the semantic constraint $N(\psi_i) \geq c_i$, where N is a necessity measure, associated with a measure of (weak) possibility Π . Namely, we have $N(\psi) = 1 - \Pi(\neg\psi)$. Thus, the formula (ψ_i, c_i) expressed that the interpretations outside $\|\psi_i\|$ have a level of possibility upper bounded by $1 - c_i$, and are somewhat impossible (when ψ_i is fully certain, $c_i = 1$, and the possibility of any $\omega \notin \|\psi_i\|$ is 0, which means full impossibility).

A positive possibilistic knowledge base $D = \{[\varphi_j, d_j] | j = 1, k\}$ is inconsistent with a negative possibilistic knowledge base $K = \{(\varphi_i, a_i) | i = 1, m\}$ as soon as the following fuzzy set inclusion is violated:

$$\forall \omega, \delta_D(\omega) \leq \pi_K(\omega).$$

This violation occurs when something is observed or reported, while one is somewhat certain that the opposite should be true. Then a revision should take place by either questioning the generic knowledge represented by K , or what is reported, which is represented by D .

Reasoning with both negative and positive information is clearly an issue of interest, since one may have information of both type in the same time. For instance consider, a second-hand car; there may exist some rules stating that for a car of some trade mark having some mileage, the price should be in some range, but one may also have examples of similar cars recently sold. See [15, 51] for general settings allowing us to reason with knowledge and data in the same time. It is also worth mentioning that the setting of version space learning is bipolar in nature, since counter-examples play the role of negative information (counter-examples are by nature associated with the negation of generic rules), and examples are positive information [45].

3 Similarity-Based Forms of Reasoning

Similarity plays an important role when dealing with data. Two obvious examples are clustering data in unsupervised learning, and k -nearest neighbors methods in classification. Another example is provided by fuzzy rules in rule-based fuzzy controllers, where a rule is of the form “if the observed output x is in A ,

the command y should be chosen in B ", and A and B are fuzzy sets [32]³. These fuzzy sets, which have unimodal membership functions, may be understood as expressing closeness to the mode of the membership function. If a (resp. b) is the unique value having a membership grade to A (resp. B) equal to 1, then the rule means "the closer x is to a , the closer y is to b ". This a gradual rule [6, 18]. This is the basis for an *interpolation* mechanism [39], as soon as we have a second rule "the closer x is to a' , the closer y is to b' ", and an input $x = a_0$, such that $a_0 \in [a, a']$. This can be also related to the representation of co-variations [46].

3.1 Case-Based Reasoning

Case-based reasoning (CBR) is the main form of reasoning with data studied in AI. An attempt at formalizing it has been proposed in the setting of fuzzy sets and possibility theory [29]. Viewing a case as a pair (\langle situation \rangle , \langle associated result \rangle), it relies on the modeling of a CBR principle that relates the similarity of situations to the similarity of associated results. Let us state the idea more formally. Let \mathcal{C} be a repertory of n cases $c_i = (s_i, r_i)$ with $i = 1, \dots, n$, where $s_i \in \mathcal{S}$ (resp. $r_i \in \mathcal{R}$) denotes a situation (resp. a result). Let S and R be two graded similarity relations (assumed to be reflexive and symmetrical) defined on $\mathcal{S} \times \mathcal{S}$ and $\mathcal{R} \times \mathcal{R}$ respectively, where $S(s, s') \in [0, 1]$ and $R(r, r') \in [0, 1]$. Let us assume that we use a CBR principle based on the gradual rule "the more similar s_0 to s_i , the more similar r_0 to r_i " [1], where s_0 denotes the situation under consideration, and r_0 the unknown associated result. Then, it leads to the following expression for the fuzzy set \tilde{r}_0 of possible values for the unknown value y of r_0 :

$$\tilde{r}_0(y) = \min_{(s_j, r_j) \in \mathcal{C}} S(s_0, s_j) \rightarrow R(y, r_j) \quad (1)$$

where \rightarrow denotes Gödel implication $a \rightarrow b = 1$ if $a \leq b$ and $a \rightarrow b = b$ if $a > b$. It is worth noticing that the above expression underlies an *interpolation* mechanism. For instance, if a second hand car s_0 is identical to two other cars s and s' , except that its mileage is between the ones of s and s' , then the estimated price r_0 will be between r and r' , and may be quite precise due to the min-based combination in (1). Thus, the estimation of r_0 is not just based on the closest similar case, but takes advantage of the "position" of s_0 among the s_i 's such as $(s_i, r_i) \in \mathcal{C}$. In order to ensure the normalization of the fuzzy set \tilde{r}_0 in (1), it is necessary for the repertory of cases to be "consistent" with the CBR principle used (see [13] for details), which means, informally speaking, that *the cases in the repertory* should themselves obey the principle "the more similar two case situations, the more similar the case results". In particular, letting $s_0 = s_i$, if we want to ensure $\tilde{r}_i(r_i) = 1$ (i.e., one retrieves the case (s_i, r_i) as a solution) for any i , we should have $\forall i \forall j S(s_i, s_j) \leq R(r_i, r_j)$.

³ Strictly speaking, such a rule was usually modeled as meaning "if x is in A , then y can be chosen in B ", implicitly taking the view that it was reflecting commands already observed as being successful, and thus echoing positive information, or "extensional" rules [38]; see footnote 2.

If, on the contrary, there exist i and j such that $S(s_i, s_j) > R(r_i, r_j)$, i.e., the situations are more similar than the results, then another *weaker* CBR principle should be used. Namely, the fuzzy CBR principle reads “the more similar s_0 to s_i , the more *possible* the similarity r_0 and r_i ”, and then we obtain [16]

$$\tilde{r}_0(y) = \max_{(s_j, r_j) \in \mathcal{C}} \min(S(s_0, s_j), R(y, r_j)) \quad (2)$$

As can be seen, we now take the *union* (rather than the intersection) of the fuzzy sets of values close to the r_i 's weighted by the similarity of s_0 with s_i , for all $(s_j, r_j) \in \mathcal{C}$. For instance, if a second hand car s_0 is quite similar to two other cars s and s' , thus themselves quite similar, but having quite different prices r and r' , then the estimated price r_0 will be the union of the fuzzy sets of values that are close to r or close to r' (the union may be replaced here by the convex hull, for taking into account that here the price domain is a continuum). Generally speaking, the result may be quite imprecise due to the max-based combination in (2). Still, it is a weighted union of all the possibilities that are supported by known cases. Note also that (2) is fully in the spirit of reasoning with data as discussed in the previous section: each result of a reported case is all the more guaranteed to be possible as the case is similar to the situation at hand, and all these conclusions are combined disjunctively.

One might also think of using a fuzzy rule of the form “the more similar s_0 to s_i , the more *certain* the similarity r_0 and r_i ”, leading to an expression similar to (1) where Gödel implication is replaced by Dienes implication (i.e., $a \rightarrow b = \max(1 - a, b)$). However, such a rule would be less appropriate here, even if it leaves room for exceptions, since we observe that $\tilde{r}_i(r_i) = 1$ holds for any i , only if $\forall i \forall j S(s_i, s_j) > 0 \Rightarrow R(r_i, r_j) = 1$, which is a condition stronger than the one for (1) with Gödel implication.

A thorough study of the formalization of CBR principles linking the similarity of solutions to the one of problems is presented in the research monograph [29].

3.2 Case-Based Decision

This approach can be readily extended to *case-based decision*, where we have a repertory \mathcal{D} of experienced decisions under the form of cases $c_i = (s_i, d, r_i)$, which means that decision d in situation s_i has led to result r_i (it is assumed that r_i is uniquely determined by s_i and d). Classical expected utility is then changed into $U(d) = \frac{\sum_{(s_i, d, r_i) \in \mathcal{D}} S(s_0, s_i) \cdot u(r_i)}{\sum_{(s_i, d, r_i) \in \mathcal{D}} S(s_0, s_i)}$, where u is a utility function, here supposed to be valued in $[0, 1]$ [28]. Besides, counterparts to (1) and (2) are

$$U_*(d) = \min_{(s_i, d, r_i) \in \mathcal{D}} S(s_0, s_i) \rightarrow u(r_i)$$

and

$$U^*(d) = \max_{(s_i, d, r_i) \in \mathcal{D}} \min(S(s_0, s_i), u(r_i)).$$

$U_*(d)$ is a pessimistic qualitative utility that expresses that a decision d is all the better as the fuzzy set of results associated with situations similar to s_0 where decision d was experienced is included in the fuzzy set of good results. When \rightarrow is Dienes implication, $U_*(d) = 1$ only if the result obtained with decision d in any known situation somewhat similar to s_0 was fully satisfactory. $U^*(d)$ is an optimistic qualitative utility since it expresses that a decision d is all the better as it was already successfully experienced in a situation similar to s_0 . See [14] for postulate-based justifications. Another idea would be to use the approach of the previous subsection for estimating the more or less possible results of each decision, and then to compute the possible values of the utility function for each of them, which would then lead to compare fuzzy utilities.

A situation s is usually described by means of several features, i.e., $s = (s^1, \dots, s^m)$. Then the evaluation of the similarity between two situations s and $s' = (s'^1, \dots, s'^m)$ amounts to estimating the similarity according to each feature k according to a similarity relation S^k , and to combine these partial similarities using some aggregation operator agg , namely $S(s, s') = agg_{k=1, \dots, m} S^k(s^k, s'^k)$. A classical choice for agg is the conjunction operator \min , which retains the smallest similarity value as the global evaluation. But one may also think, for instance, of using some weighted aggregation if all the features have not the same importance. See [12] for a detailed example (with \min).

Besides, the approach can be extended to prediction about some imprecisely or fuzzily specified cases (e.g., one has to estimate the price of a car with precisely specified features except that the horse power is between 90 and 110). A further generalization is necessary in order to accommodate incompletely specified cases in the repertory. See [16] for these extensions in the case of possibility rules (thus corresponding to (2)), and [31] for the discussion of several other generalizations (including the discounting of untypical cases and the flexible handling and adequate adaptation of different similarity relations, which provides a way of incorporating domain-specific (expert) knowledge). A comparative discussion with instance-based learning algorithms, a form of transduction, is in [30]. Applications to flexible querying [9], including examples (and counter-examples)-based querying⁴, and to recommendation systems [17] have been also proposed.

Lastly, one may think of cases that provide an argumentative support in favor of a claim as positive examples of it, or more strongly of cases used as a counter-example to a rule used in an argument; see a brief outline of this idea in [40], when discussing an argumentative view of case-based decision.

3.3 Analogical Reasoning

The notion of similarity is as essential to CBR as it is to the idea of analogy, and in particular, to analogical proportions. The core idea underlying analogical proportions comes from the numerical field where proportions express an equality of ratios, e.g. $\frac{1}{2} = \frac{5}{10}$, which could be read “1 is to 2 as 5 is to 10”. It is also

⁴ An item is all the more a solution as it resembles to some example(s) in all important aspects, and is dissimilar from all counter-examples in some important aspect(s).

agreed that “read is to reader as lecture is to lecturer” is a natural language analogical proportion, and the notation $read : reader :: lecture : lecturer$ is then preferred. More generally, an analogical proportion is an expression usually denoted $a : b :: c : d$ involving 4 terms a, b, c, d , which reads “ a is to b as c is to d ”. It clearly involves comparisons between the pairs (a, b) and (c, d) . Recent works have led to a logical formalization of analogical proportions, where similarities/dissimilarities existing between a and b are equated to similarities/dissimilarities existing between c and d .

Let us assume that the items a, b, c, d represent sets of binary features belonging to an universe U (i.e. an item is then viewed as the set of binary features in U that it satisfies). Then, the dissimilarity between a and b can be appreciated in terms of $a \cap \bar{b}$ and/or $\bar{a} \cap b$, where \bar{a} denotes the complement of a in U , while the similarity is estimated by means of $a \cap b$ and/or of $\bar{a} \cap \bar{b}$. Then, an analogical proportion between subsets is formally defined as [35]:

$$a \cap \bar{b} = c \cap \bar{d} \text{ and } \bar{a} \cap b = \bar{c} \cap d$$

This expresses that “ a differs from b as c differs from d ” and that “ b differs from a as d differs from c ”. It can be viewed as the expression of a co-variation. It has an easy counterpart in Boolean logic, where a, b, c, d now denote simple Boolean variables. In this logical setting, “are equated to” translates into “are equivalent to” (\equiv), \bar{a} is now the negation of a , and \cap is changed into a conjunction (\wedge), and we get the logical condition expressing that 4 Boolean variables make an analogical proportion:

$$(a \wedge \bar{b} \equiv c \wedge \bar{d}) \wedge (\bar{a} \wedge b \equiv \bar{c} \wedge d)$$

It is logically equivalent to the following condition that expresses that the pairs made by the extremes and the means, namely (a, d) and (b, c) , are (positively and negatively) similar [35]:

$$(a \wedge d \equiv b \wedge c) \wedge (\bar{a} \wedge \bar{d} \equiv \bar{b} \wedge \bar{c}).$$

An analogical proportion is then a Boolean formula. It takes the truth value “1” only for any of the 6 following patterns for $abcd$: 1111, 0000, 1100, 0011, 1010, 0101. For the 10 other lines of its truth table, it is false (i.e., equal to 0). As expected, it satisfies the following remarkable properties:

$$\begin{aligned} & a : b :: a : b \text{ (reflexivity),} \\ & \text{(and thus } a : a :: a : a \text{ (identity));} \\ & a : b :: c : d \Rightarrow c : d :: a : b \text{ (symmetry);} \\ & a : b :: c : d \Rightarrow a : c :: b : d \text{ (central permutation).} \end{aligned}$$

Another worth noticing property [42] is the fact that the analogical proportion remains true for the negation of the Boolean variables. It expresses that the result does not depend on a positive or a negative⁵ encoding of the features:

⁵ The use of these words here just refers to the application of a negation, and should not be confused with their use in other parts of the paper.

$$a : b :: c : d \Rightarrow \bar{a} : \bar{b} :: \bar{c} : \bar{d} \text{ (code independency).}$$

Finally, analogical proportions satisfy a unique solution property, which means that, 3 Boolean values a, b, c being given, when we have to find a fourth one x such that $a : b :: c : x$ holds, we have either no solution (as in the cases of $011x$ or $100x$), or a unique one (as, e.g., in the case of $110x$). More formally, the analogical equation $a : b :: c : x$ is solvable iff $((a \equiv b) \vee (a \equiv c)) = 1$. In that case, the unique solution x is $a \equiv (b \equiv c)$ [35]. This allows us to deal with Boolean analogical proportions in a simple way.

The basic idea underlying the analogical proportion-based inference is as follows: *if there is a proportion that holds between the first p components of four vectors, then this proportion should hold for the last remaining components as well.* This inference principle [50] can be formally stated as below:

$$\frac{\forall i \in \{1, \dots, p\}, \quad a_i : b_i :: c_i : d_i \text{ holds}}{\forall j \in \{p + 1, \dots, n\}, \quad a_j : b_j :: c_j : d_j \text{ holds}}$$

This is a generalized form of analogical reasoning, where we transfer knowledge from some components of our vectors to their remaining components.

It is worth pointing out that properties such as *full identity* or *code independency* are especially relevant in that perspective. Indeed, it is expected that in the case where d is such that it exists a case a in the repertory with $\forall i \in \{1, \dots, p\}, d_i = a_i$, then $a_i : a_i :: a_i : d_i$ holds. Thus, the approach includes the extreme particular case where we have to classify (or to predict components of) an item whose representation (in the input space) is completely similar to the one of a completely known item. The code independency property, which expresses independence with respect to the encoding, seems also very desirable since it ensures that whatever the convention used for the positive or the negative encodings of the value of each feature and of the class, one shall obtain the same result for features in $\{p + 1, \dots, n\}$. Then analogical reasoning amounts to finding completely informed triples suitable for inferring the missing value(s) of an incompletely informed item as in the following example. In case of the existence of several possible triples leading to possibly distinct plausible conclusions, a voting procedure may be used, as in case-based reasoning.

Let us consider for instance a database of homes to let, containing houses (1) and flats (0), which are well equipped or not (1/0), which are cheap or expensive (1/0), where you have to pay a tax or not (1/0). Then a house, well equipped, expensive and taxable is represented by the vector $a = (1, 1, 0, 1)$. Having 2 other cases $b = (1, 0, 1, 1)$, $c = (0, 1, 0, 1)$, we can predict the price and taxation status of a new case d which is a flat not well equipped, i.e. $d = (0, 0, x, y)$ where 2 values are unknown. Applying the above approach, and noticing that an analogical proportion $a : b :: c : d$ holds for the 2 first components of each vector, we “infer” that such a proportion should hold for the 2 last components as well, yielding $x = 1$ and $y = 1$ (i.e. cheap and taxable).

This approach, using Boolean analogical proportions, has been extended to numerical features using multiple-valued connectives [43]. It has been successfully applied to classification problems [3, 34, 44], where the attribute to be predicted

is the class of the new item. Analogical proportions may be also applied to interpolation and extrapolation reasoning between if-then rules [10,48,49], but this is beyond their direct application to data.

4 Making Sense of Data

Making sense of data may cover a large range of situations where we reason about data. By reasoning about data, we mean reasoning from a (possibly dynamic) set of data, without the purpose of drawing a conclusion on a particular attribute in a given situation, as in deductive, abductive, case-based, or analogical reasoning. The issue is then to understand the whole set of data in a way or another. Reasoning about data covers a variety of problems as briefly reviewed now.

A first class of problems is when receiving a flux of information to figure out what is going on. We are close to the recognition of temporal scenarii [52]. We may need to identify what causes what (see, e.g., [7]). In such problems, we have to check if data fits with knowledge describing an abnormal, or the normal course of things.

Another important class of problems deals with the structuring of the data. We may start from a table of data, as in formal concept analysis [25], where a formal context R indicates what Boolean attribute(s) is/are true for a given object. Then, a formal concept is a maximal pair (X, Y) , such as $X \times Y \subseteq R$ where X is a set of objects and Y is a set of properties; each object in X has all properties in Y , and each property in Y is possessed by all objects in X . A formal context is associated with a lattice of formal concepts, from which association rules can be extracted [24,36]. This is the theoretical basis for data mining.

Interestingly enough, the operator which is at the basis of the definition of formal concepts is analogous to the guaranteed possibility measure mentioned in Sect. 2; indeed, in a formal concept (X, Y) , the properties in Y are guaranteed for any object in X . Note also that $(x, y) \in R$ is understood here as a positive fact, while $(x', y') \notin R$ is not viewed as a negative fact, it rather means that the piece of information $(x', y') \in R$ is not available (at least if there is no closed world assumption underlying the formal context R). Moreover, other possibility theory operators have been imported in formal concept analysis, and enables us to consider other forms of reasoning, still to be investigated in detail, including case-based reasoning, see [20]. Moreover, formal concept analysis can be related [21] to other theoretical frameworks such as rough sets [37] or extensional fuzzy sets, in the general setting of granular computing [53], where the idea of clustering is implicitly at work. Closely related is the summarization of data which exploits ideas of similarity and clustering (e.g., [5,27,33]).

Classification or estimation methods are usually black box devices. They may be learnt from data. It is clearly of interest to lay bare the contents of these black boxes in understandable terms. There have been a number of attempts in that directions; let us mention a few examples like a non-monotonic inference view [26] or a fuzzy rule-based interpretation [8] of neural nets, or more recently a weighted logic view of Sugeno integrals [22] laying bare the rules underlying the global estimation.

5 Concluding Remarks

Taking machine learning and data mining apart, reasoning with data has remained confined in few specialized works (at least if we restrict ourselves to formalized approaches), or in particular areas such as fuzzy logic, or rough sets [37]. This overview has emphasized two important points: (i) data and knowledge being of different nature, they should be handled differently, and handling both knowledge and data requires a bipolar setting; (ii) similarity (and dissimilarity) play an important role when reasoning with data.

It becomes timely to recognize reasoning with data as a general research trend in AI, to identify all the facets and issues raised by the handling of data in various forms of reasoning, and to develop a unified view of these problems. It may also contribute to a better interfacing between reasoning and learning research areas [11, 33, 47].

References

1. Arrazola, I., Plainfossé, A., Prade, H., Testemale, C.: Extrapolation of fuzzy values from incomplete data bases. *Inf. Syst.* **14**(6), 487–492 (1989)
2. Baader, F., Horrocks, I., Sattler, U.: Description logics. In: van Harmelen, F., Lifschitz, V., Porter, B. (eds.) *Handbook of Knowledge Representation*, chapter 3, pp. 135–180. Elsevier (2007)
3. Bayouhd, S., Miclet, L., Delhay, A.: Learning by analogy: a classification rule for binary and nominal data. In: *Proceedings of International Conference on Artificial Intelligence (IJCAI 2007)*, pp. 678–683 (2007)
4. Benferhat, S., Dubois, D., Kaci, S., Prade, H.: Modeling positive and negative information in possibility theory. *Int. J. Intell. Syst.* **23**, 1094–1118 (2008)
5. Bosc, P., Dubois, D., Pivert, O., Prade, H., De Calmès, M.: Fuzzy summarization of data using fuzzy cardinalities. In: *Proceedings of 9th International Conference Information Processing and Management of Uncertainty in Knowledge-Based Systems (IPMU 2002)*, Annecy, pp. 1553–1559, 1–5 July 2002
6. Bouchon-Meunier, B., Laurent, A., Lesot, M.-J., Rifqi, M.: Strengthening fuzzy gradual rules through “all the more” clauses. In: *Proceedings of IEEE International Conference on Fuzzy Systems (Fuzz-IEEE 2010)*, Barcelona, July 2010
7. Chassy, P., de Calmès, M., Prade, H.: Making sense as a process emerging from perception-memory interaction: a model. *Int. J. Intell. Syst.* **27**, 757–775 (2012)
8. d’Alché-Buc, F., Andrés, V., Nadal, J.-P.: Rule extraction with fuzzy neural network. *Int. J. Neural Syst.* **5**(1), 1–11 (1994)
9. De Calmès, M., Dubois, D., Hüllermeier, E., Prade, H., Sédes, F.: Flexibility, fuzzy case-based evaluation in querying: an illustration in an experimental setting. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **11**(1), 43–66 (2003)
10. Derrac, J., Schockaert, S.: Inducing semantic relations from conceptual spaces: a data-driven approach to plausible reasoning. *Artif. Intell.* **228**, 66–94 (2015)
11. Domingos, P., Kok, S., Lowd, D., Poon, H., Richardson, M., Singla, P.: Markov logic. In: Raedt, L., Frasconi, P., Kersting, K., Muggleton, S.H. (eds.) *Probabilistic Inductive Logic Programming*. LNCS (LNAI), vol. 4911, pp. 92–117. Springer, Heidelberg (2008)

12. Dubois, D., Esteva, F., Garcia, P., Godo, L., López de Mántaras, R., Prade, H.: Fuzzy modelling of case-based reasoning and decision. In: Leake, D.B., Plaza, E. (eds.) ICCBR 1997. LNCS, vol. 1266, pp. 599–610. Springer, Heidelberg (1997)
13. Dubois, D., Esteva, F., Garcia, P., Godo, L., López de Mántaras, R., Prade, H.: Fuzzy set modelling in case-based reasoning. *Int. J. Intell. Syst.* **13**, 345–373 (1998)
14. Dubois, D., Godo, L., Prade, H., Zapico, A.: On the possibilistic decision model: from decision under uncertainty to case-based decision. *Int. J. Uncertain. Fuzziness Knowl. Based Syst.* **7**(6), 631–670 (1999)
15. Dubois, D., Hajek, P., Prade, H.: Knowledge-driven versus data-driven logics. *J. Log. Lang. Inf.* **9**, 65–89 (2000)
16. Dubois, D., Hüllermeier, E., Prade, H.: Fuzzy set-based methods in instance-based reasoning. *IEEE Trans. Fuzzy Syst.* **10**(3), 322–332 (2002)
17. Dubois, D., Hüllermeier, E., Prade, H.: Fuzzy methods for case-based recommendation and decision support. *J. Intell. Inf. Syst.* **27**, 95–115 (2006)
18. Dubois, D., Prade, H.: Gradual inference rules in approximate reasoning. *Inf. Sci.* **61**(1–2), 103–122 (1992)
19. Dubois, D., Prade, H.: Possibility theory: qualitative and quantitative aspects. In: Gabbay, D., Smets, P. (eds.) *Quantified Representation of Uncertainty and Imprecision. Handbook of Defeasible Reasoning and Uncertainty Management Systems Series*, vol. 1, pp. 169–226. Kluwer Academic Publishers (1998)
20. Dubois, D., Prade, H.: Possibility theory and formal concept analysis: characterizing independent sub-contexts and handling approximations. *Fuzzy Sets Syst.* **196**, 4–16 (2012)
21. Dubois, D., Prade, H.: Bridging gaps between several forms of granular computing. *Granul. Comput.* **1**(2), 115–126 (2016)
22. Dubois, D., Prade, H., Rico, A.: The logical encoding of Sugeno integrals. *Fuzzy Sets Syst.* **241**, 61–75 (2014)
23. Dubois, D., Prade, H., Smets, P.: “Not impossible” vs. “guaranteed possible” in fusion and revision. In: Benferhat, S., Besnard, P. (eds.) *ECSQARU 2001. LNCS (LNAI)*, vol. 2143, pp. 522–531. Springer, Heidelberg (2001)
24. Duquenne, V., Guigues, J.-L.: Famille minimale d’implications informatives résultant d’un tableau de données binaires. *Math. Sci. Hum.* **24**(95), 5–18 (1986)
25. Ganter, B., Wille, R.: *Formal Concept Analysis*. Springer, New York (1999)
26. Gärdenfors, P.: Nonmonotonic inference, expectations, and neural networks. In: Kruse, R., Siegel, P. (eds.) *Symbolic and Quantitative Approaches to Uncertainty. LNCS*, vol. 548, pp. 12–27. Springer, Heidelberg (1991)
27. Gaume, B., Navarro, E., Prade, H.: Clustering bipartite graphs in terms of approximate formal concepts and sub-contexts. *Int. J. Comput. Intell. Syst.* **6**(6), 1125–1142 (2013)
28. Gilboa, I., Schmeidler, D.: Case-based decision theory. *Q. J. Econ.* **110**, 605–639 (1995)
29. Hüllermeier, E.: *Case-Based Approximate Reasoning. Theory and Decision Library*. Springer, New York (2007)
30. Hüllermeier, E., Dubois, D., Prade, H.: Model adaptation in possibilistic instance-based reasoning. *IEEE Trans. Fuzzy Syst.* **10**(3), 333–339 (2002)
31. Hüllermeier, E., Dubois, D., Prade, H.: Knowledge-based extrapolation of cases: a possibilistic approach. In: Bouchon-Meunier, B., Gutiérrez-Ríos, J., Magdalena, L., Yager, R.R. (eds.) *Technologies for Constructing Intelligent Systems 1*, pp. 377–390. Springer, Heidelberg (2002)
32. Mamdani, E.H., Assilian, S.: An experiment in linguistic synthesis with a fuzzy logic controller. *Int. J. Man Mach. Stud.* **7**, 1–13 (1975)

33. Memory, A., Kimmig, A., Bach, S.H., Raschid, L., Getoor, L.: Graph summarization in annotated data using probabilistic soft logic. In: Bobillo, F., et al. (eds.) Proceedings of 8th International Workshop on Uncertainty Reasoning for the Semantic Web (URSW 2012), Boston, November 2011, vol. 900, pp. 75–86. CEUR Workshop Proceedings (2012)
34. Miclet, L., Bayouhd, S., Delhay, A.: Analogical dissimilarity: definition, algorithms and two experiments in machine learning. *J. Artif. Intell. Res. (JAIR)* **32**, 793–824 (2008)
35. Miclet, L., Prade, H.: Handling analogical proportions in classical logic and fuzzy logics settings. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS, vol. 5590, pp. 638–650. Springer, Heidelberg (2009)
36. Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Efficient mining of association rules using closed itemset lattices. *Inf. Syst.* **24**(1), 25–46 (1999)
37. Pawlak, Z.: *Rough Sets: Theoretical Aspects of Reasoning About Data*. Kluwer Academic Publishers, Dordrecht (1991)
38. Pearl, J.: *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers, San Mateo (1988)
39. Perfilieva, I., Dubois, D., Prade, H., Esteva, F., Godo, L., Hodáková, P.: Interpolation of fuzzy data: analytical approach and overview. *Fuzzy Sets Syst.* **192**, 134–158 (2012)
40. Prade, H.: Qualitative evaluation of decisions in an argumentative manner - a general discussion in a unified setting. In: Proceedings of 4th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT-LFA 2005 Joint Conference), Barcelona, 7–9 September, pp. 1003–1008 (2005)
41. Prade, H.: Raisonner avec des données. Un nouveau chantier pour l'IA? Actes 10èmes Jour. Intellig. Artif. Fondamentale (JIAF), Montpellier, 15–17 June 2016. https://www.supagro.fr/jfpc.jiaf_2016/Articles.IAF.2016/Actes.IAF.2016.pdf
42. Prade, H., Richard, G.: From analogical proportion to logical proportions. *Log. Univers.* **7**(4), 441–505 (2013)
43. Prade, H., Richard, G.: Analogical proportions and multiple-valued logics. In: van der Gaag, L.C. (ed.) ECSQARU 2013. LNCS, vol. 7958, pp. 497–509. Springer, Heidelberg (2013)
44. Prade, H., Richard, G., Yao, B.: Enforcing regularity by means of analogy-related proportions - a new approach to classification. *Int. J. Comput. Inf. Syst. Ind. Manag. Appl.* **4**, 648–658 (2012)
45. Prade, H., Serrurier, M.: Bipolar version space learning. *Int. J. Intell. Syst.* **23**(10), 1135–1152 (2008)
46. Raccah, P.Y. (ed.): *Topoi et Gestion des Connaissances*. Masson, Paris (1996)
47. Russell, S.J.: Unifying logic and probability. *Commun. ACM* **58**(7), 88–97 (2015)
48. Schockaert, S., Prade, H.: Interpolation and extrapolation in conceptual spaces: a case study in the music domain. In: Rudolph, S., Gutierrez, C. (eds.) RR 2011. LNCS, vol. 6902, pp. 217–231. Springer, Heidelberg (2011)
49. Schockaert, S., Prade, H.: Completing symbolic rule bases using betweenness and analogical proportion. In: Prade, H., Richard, G. (eds.) *Computational Approaches to Analogical Reasoning: Current Trends*, pp. 195–215. Springer, Heidelberg (2014)
50. Stroppa, N., Yvon, F.: Analogical learning and formal proportions: definitions and methodological issues. Technical report, ENST D-2005-004, Paris, June 2005
51. Ughetto, L., Dubois, D., Prade, H.: Implicative and conjunctive fuzzy rules - a tool for reasoning from knowledge and examples. In: Proceedings of 15th National Conference in Artificial Intelligence (AAAI 1999), Orlando, pp. 214–219, July 1999

52. Vu, V.-T., Brémond, F., Thonnat, M.: Automatic video interpretation: a novel algorithm for temporal scenario recognition. In: Gottlob, G., Walsh, T. (eds.) Proceedings of 18th International Joint Conference Artificial Intelligence (IJCAI 2003), Acapulco, 9–15 August, pp. 1295–1302 (2003)
53. Zadeh, L.A.: Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. *Fuzzy Sets Syst.* **90**, 111–128 (1997)