Towards Analogy-Based Decision - A Proposal

Richard Billingsley[2], Henri Prade[1,2(✉)], Gilles Richard[1], and Mary-Anne Williams[2]

[1] IRIT, CNRS and Toulouse University, Toulouse, France
{prade,richard}@irit.fr
[2] QCIS, University of Technology Sydney, Sydney, Australia
Richard.Billingsley@uts.edu.au, Mary-Anne@themagiclab.org

Abstract. This short paper outlines an analogy-based decision method. It takes advantage of analogical proportions between situations, i.e., a is to b as c is to d, for proposing plausibly good decisions that may be appropriate for a new situation at hand. It goes beyond case-based decision where the idea of graded similarity may hide some small but crucial differences between situations. The method relies on triples of known cases rather than on individual cases for making a prediction on the appropriateness of a potential decision, or for proposing a way of adapting a decision according to situations. The approach may be of interest in a variety of problems ranging from flexible querying systems to cooperative artificial agents.

1 Introduction

Making decision on a daily basis often relies on past experience [1], rather than being an explicit utility-maximization business based on the knowledge of a utility function. This observation has led decision theorists to propose an axiomatic modeling of case-based decision, relying on a repository of cases already experienced, where expected utility is replaced by a weighted average of the utilities of the results of this decision in various situations that are *similar* to the current situation, the sum being weighted in terms of this similarity [7].

A qualitative counterpart to case-based decision has been proposed [2,3]. Thus an optimistic (resp. pessimistic) decision criterion implements the idea that a decision is all the better as it led to a good decision in at least one (resp. all) situation(s) similar to the current situation, where the decision was made. Since both goodness and similarity are matters of degree, these two decision criteria estimate respectively to what extent the intersection of two fuzzy sets is not empty and to what extent a fuzzy set is included in another one.

While the idea of case-based decision is intuitively attractive, it may be difficult to implement it due to the difficulty of eliciting meaningful similarity relations between situations. Indeed, two situations may be quite similar in many respects and still have one crucial difference, in such a way that for a considered decision, we may obtain a good result in one case and a bad one in the other. Such a remark calls for an analysis in terms of analogical proportions. Analogical

H. Christiansen et al. (Eds.): FQAS 2017, LNAI 10333, pp. 28–35, 2017.
DOI: 10.1007/978-3-319-59692-1_3

proportions are statements of the form a is to b as c is to d, which compare a pair (a, b) with a pair (c, d) in terms of the similarities and of the differences between the elements of each pair [11].

In the following, we outline an approach to analogy-based decision where we take advantage of a repository of past experiences in terms of analogical proportions, suggesting the benefits that may be expected from such a view. We first provide a short background on case-based decision, and then on analogical proportions, before presenting the approach. The structural resemblances between case-based decision and flexible querying have been already discussed [4]; this leads to conjecture some applications of analogy-based decision to flexible querying. We also advocate the interest of analogy-based decision for artificial agents that have to cooperate with humans on the basis of past interactions.

2 Case-Based Decision

In case-based decision, we are supposed to have at our disposal a repository $\boldsymbol{R}$ of experienced decisions under the form of cases $c_i = (s_i, \delta_i, r_i), i \in [1, n]$. Case c_i means that decision $\delta_i \in \mathcal{D}$ ($\mathcal{D}$ is the set of potential decisions), applied in situation $s_i \in \mathcal{S}$ ($\mathcal{S}$ is the set of considered situations), has led to result $r_i \in \mathcal{R}$; it is assumed that r_i is uniquely determined by s_i and δ_i. Consider a new situation s_0, which may not be in $\mathcal{S}$, for which we have to take a decision δ_0. Let S be a similarity measure defined on $\mathcal{S} \times \mathcal{S}$ that associates each pair (s_i, s_j) of situations with a positive real number $S(s_i, s_j) \in [0, 1]$.

Classical expected utility is then changed [7], for a candidate decision δ into

$$U(\delta) \triangleq \frac{\Sigma_{(s_i,\delta,r_i)\in \boldsymbol{R} \ s.t. \ \delta_i=\delta} \ S(s_0, s_i) \cdot u(r_i)}{\Sigma_{(s_i,\delta_i,r_i)\in \boldsymbol{R} \ s.t. \ \delta_i=\delta} \ S(s_0, s_i)},$$

where u is a utility function, here supposed to be valued in $[0, 1]$. Note that the value of u needs to be known only for the cases in the repository. Decision is made by choosing $\delta \in \mathcal{D}$ such that it maximizes $U(\delta)$.

A pessimistic and an optimistic qualitative counterparts [2] are given by

$$U_*(\delta) \triangleq \min_{(s_i,\delta_i,r_i)\in \boldsymbol{R} \ s.t. \ \delta_i=\delta} S(s_0, s_i) \rightarrow u(r_i) \text{ and}$$
$$U^*(\delta) \triangleq \max_{(s_i,\delta_i,r_i)\in \boldsymbol{R} \ s.t. \ \delta_i=\delta} \min(S(s_0, s_i), u(r_i)).$$

$U_*(\delta)$ expresses that decision δ is all the better as the fuzzy set of results associated with situations similar to s_0 where decision δ was experienced is included in the fuzzy set of good results. When $x \rightarrow y = \max(1 - x, y)$, $U_*(\delta) = 1$ only if the result obtained with decision δ in any known situation somewhat similar to s_0 was fully satisfactory. $U^*(\delta)$ expresses that a decision δ is all the better as it was already successfully experienced in a situation similar to s_0. See [3] for postulate-based justifications. The very pessimistic and optimistic nature of criteria U_* and U^* may be softened by introducing fuzzy quantifiers; for example, one might already be satisfied if a decision was a good choice in *most* similar cases, thus allowing for a few exceptions [4]. Note that these criteria may lead to prefer a rarely experienced decision that always led to good results (if any), rather to choose a decision with more feedbacks, but some bad ones. However, it may be desirable to avoid an accumulation effect if the same decision is chosen routinely for the same, frequently occurring, situation and thus is stored several times in the repertory of cases. In fact, what is addressed here is the possible variability of the output of a decision made in more or less similar situations rather than its possible non deterministic nature when repeatedly applied in (apparently) the same situation.

A situation s is usually described by means of several features, i.e., $s = (s^1, ..., s^m)$. Then the evaluation of the similarity between two situations s and $s' = (s'^1, ..., s'^m)$ amounts to estimating the similarity for each feature k according to a similarity relation S^k, and to combine these partial similarities using some aggregation operator agg, namely $S(s, s') = agg_{k=1,...,m} S^k(s^k, s'^k)$. 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 [4,5] for more details, references to generalizations to incompletely specified cases, or with discounting of untypical cases, and applications to flexible querying including examples (and counter-examples)-based querying[1].

3 Analogy: Brief Review

Analogical proportions provide another way to compare situations. Analogical proportions are statements of the form a is to b as c is to d. Let us assume that the four items a, b, c, d are represented by sets of binary features belonging to a universe U (i.e., an item is then viewed as the subset of the binary features in U that it satisfies). Then, the dissimilarity between a and b can be appreciated in terms of $a \cap \overline{b}$ and/or $\overline{a} \cap b$, where $\overline{a}$ denotes the complement of a in U. Indeed $a \cap \overline{b}$ (resp. $\overline{a} \cap b$) is the subset of properties possessed by a (resp. b) and not by b (resp. a). The similarity is estimated by means of $a \cap b$ (the subset of properties possessed by both a and b) and/or of $\overline{a} \cap \overline{b}$ (the subset of properties that both a and b do not possess). Then, an analogical proportion between subsets is formally defined [9] as a conjunction of equalities:

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

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, here denoted $a : b :: c : d$, where a, b, c, d now denote simple Boolean variables. In this logical setting, "are equated to" translates into "are equivalent to" ($\equiv$), $\overline{a}$ 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:

$$a : b :: c : d \triangleq (a \wedge \overline{b} \equiv c \wedge \overline{d}) \wedge (\overline{a} \wedge b \equiv \overline{c} \wedge d)$$

[1] 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).

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 [9]. This could be used as a definition as well:

$$a : b :: c : d \triangleq (a \wedge d \equiv b \wedge c) \wedge (\overline{a} \wedge \overline{d} \equiv \overline{b} \wedge \overline{c}).$$

An analogical proportion is then a Boolean formula. $a : b :: c : d$ takes the truth value "1" only for 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).

A worth noticing property, beyond reflexivity $(a : b :: a : b)$, symmetry $(a : b :: c : d \Rightarrow c : d :: a : b)$, and central permutation $(a : b :: c : d \Rightarrow a : c :: b : d)$ is the fact that the analogical proportion remains true for the negation of the Boolean variables [11]. It expresses that the result does not depend on a positive or a negative encoding of the features describing the situations:

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

Moreover, 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 true, 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 $a \equiv (b \equiv c)$ [9]. The following example provides an illustration with nominal values, where the Boolean patterns are replaced by patterns of the forms XXYY, XYXY, and XXXX. Note that the 4th line (i.e., the description of `Queen`) can be calculated from the 3 first lines by solving here three analogical equations in terms of nominal values. This validates that "a man is to a king as a woman is to a queen".

	Sex	Position	Human
Man	*M*	*ordinary*	*yes*
King	*M*	*power*	*yes*
Woman	*F*	*ordinary*	*yes*
Queen	*F*	*power*	*yes*

The basic idea underlying the analogical proportion-based inference is as follows: if there is a proportion that holds between p components of four vectors, then this proportion may hold for the last remaining components as well. This inference principle [12] can be formally stated as below:

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

This is a generalized form of analogical reasoning, where we transfer knowledge from some components of our vectors to their remaining components, tacitly assuming that the values of the p first components determine the values of the others. Then analogical reasoning amounts to finding completely informed triples suitable for inferring the missing value(s) of an incompletely informed item. 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 where the inference is based on a collection of single cases (i.e., the nearest neighbors) rather than on a collection of triples. We may now move from case-based decision to analogical-proportion based decision.

4 Analogy-Based Decision

Let us consider a generic scenario where a decision δ was experienced in two different situations sit_1 and sit_2 in the presence or not of special circumstances, leading to *good* or *bad* results respectively depending on the absence or on the presence of these special circumstances. Suppose we have in our repository the first three lines of the following table (cases a, b, c), while we wonder if we should consider applying decision δ or not in sit_2 when no special circumstances are present (case d). The analogical inference leads here to the prediction that the result should be *good*.

Case	Situation	Special circumstances	Decision	Result
a	sit_1	*yes*	δ	*bad*
b	sit_1	*no*	δ	*good*
c	sit_2	*yes*	δ	*bad*
d	sit_2	*no*	$\boldsymbol{\delta}$	***good***

Further comments are in order here.

- First, note that if we apply a case-based decision view, case d might be found quite similar to case c, since they are identical on all the features used for describing situation sit_2, and differs only on the maybe unique feature describing the so-called "special circumstances"; this would lead to favor the idea that decision δ in case d would also lead to a *bad* result as in case c.
- However, a more careful examination of cases a, b, c may lead to an opposite conclusion. Indeed it is natural to implicitly assume here that the possibly many features gathered here under the labels "situation" and "special circumstances" are enough for describing the cases and for determining the quality of the result of decisions applied to the cases. Thus, the fact that in sit_1, the quality of the result of decision δ is *bad* (resp. *good*) is explained by the presence (resp. absence) of "special circumstances". Then the analogical inference enforces here that we should have the same behavior in sit_2.
- But, note that nothing forbids to also have in the repository a pair of cases pertaining to another situation, say sit_3, which is a counter-example to this behavior, namely one may have the two cases $a' = (sit_3, yes, \delta, good)$ and $c' = (sit_3, no, \delta, good)$ that states in sit_3 the "special circumstances" feature

has no influence when δ is applied; then from cases a', c, c', observing that $(sit_3, yes, \delta) : (sit_2, yes, \delta) :: (sit_3, no, \delta) : (sit_2, no, \delta)$, one concludes by analogical inference that the result of decision δ in case d is then *bad*. This would mean here that it is the situation itself that determines if the result is `good` or `bad`.

– Thus different triples may lead to different predictions for the case d under consideration. Thus, one may associate each potential decision δ with the multiset $\mathcal{E}(\delta)$ of the evaluations expected for its result in the current case. For instance, $\mathcal{E}(\delta) = \{good, good, bad\}$, or more generally in case of a non binary evaluation, e.g., $\mathcal{E}(\delta) = \{very\ good, good, good, bad, very\ bad\}$. The difficulty is then to compare multisets of different sizes, e.g.,$\{good\}$ and $\{good, good, bad\}$. Indeed what is better? To choose a rarely experienced decision that is always associated with a good prediction? Or to choose a decision with many good predictions, but some bad ones? Then, several ways of handling these multiple evaluations are possible:
  - a simple majority vote for summarizing the evaluation of each candidate decision δ_i;
  - a more sophisticated weighted sum; for instance, in the above example, sit_2 may be much closer to sit_1 than to sit_3 (in terms of Hamming distance), which leads to somewhat discount the evaluation *bad* and consider that we are closer to *good*.

Generally speaking, this approach can be split in two steps:

step 1: For each candidate decision δ, one first looks for triples (a, b, c) of available cases pertaining to δ whose utilities $u(a), u(b), u(c)$ are known, such that the utility equation $u(a) : u(b) :: u(c) : x$ is solvable (an analogical proportion that holds involves one or two distinct utilities). This task which reduces the space of candidate triples can be performed offline.

step 2: When one wonders what to do in a case $d = (d_1, \ldots, d_p)$, one looks for candidate triples $a = (a_1, \ldots, a_p), b = (b_1, \ldots, b_p)$ and $c = (c_1, \ldots, c_p)$ such that analogical proportions hold component-wise for the feature values, i.e., the proportion holds on all features (if they are all regarded as relevant), or at least on a maximum number of features (analogical inference tolerates irrelevant features, but then requiring that the proportion holds on all features would diminish the number of triples available). Then, the solution of the utility equation $u(a) : u(b) :: u(c) : x$ for any such triple (a, b, c) is a possible value of the utility $u(d)$ of the result of the application of δ in case d.

This approach extends from the Boolean features case to nominal and numerical features, since the definition of analogical proportion can be extended to these cases [5,6].

5 Adaptive Decision with Analogy

Rather than analogically predicting the evaluation of the output of a potential decision in a new situation, one may suppose that we start with a repertory of recommended actions in a variety of circumstances, and then one may also think of trying to take advantage of the creative capabilities of analogy for adapting a decision to the new situation. Indeed, from three patterns, say, e.g., 00, 01, 10, by solving equations $0 : 0 :: 1 : x$ and $0 : 1 :: 0 : y$, one may obtain a fourth pattern, here 11, which is new.

This may be useful when the final decision has diverse options. Such as *Serve a tea* with or without sugar, with or without milk. Let us consider this example to illustrate the idea. As stored in the table below, in situation sit_1 with contraindication ($c\ i$), it is recommended to serve tea only, in situation sit_1 with no $c\ i$, tea with sugar, while in situation sit_2 with $c\ i$ one serves tea with milk. What to do in situation sit_2 with no $c\ i$? Common sense suggests tea with sugar and milk, maybe. It is what analogical proportion equations says: indeed $\delta : \delta :: \delta : x$, $0 : 1 :: 0 : y$ and $0 : 0 :: 1 : z$ yield $xyz = \delta 11$ as in the table below.

Case	Situation	Contraindication	Decision	Option 1	Option 2
a	sit_1	*yes*	δ	0	0
b	sit_1	*no*	δ	1	0
c	sit_2	*yes*	δ	0	1
d	sit_2	*no*	$\boldsymbol{\delta}$	**1**	**1**

We have only outlined how to adapt a decision, viewed as a compound act, depending on circumstances, by means of an analogical proportion-based reasoning. Further investigations are needed for developing the idea.

6 Some Potential Uses

The approach has been presented in general decision terms, and may be applied to different problems. We suggest here two examples of potential uses we may think of.

In querying systems it is important to try to avoid empty sets of answers. Replacing *good* (resp. *bad*) by *non empty set* (resp. *empty set*) in the kind of generic example of the Sect. 4, one may predict when a query may lead to an empty set of answers based on the results of past queries that are not too far from the query under consideration (i.e. differing on few points on which the risk of empty set of answers is sensitive). Then one may also identify the special circumstances that favor empty sets of answers.

Robots must learn utilities from sparse feedback. If the context of the situations is to be considered when learning preferences then the parameter set may be far larger than the training set resulting in over-fitting. The ability to make generalizations between alternative preferences allows the robot to learn more from less data. Just as linear regression does not need the power set of alternatives to learn corresponding weights, the ability for a robot to learn relationships in the data leads to an acceleration of the learning process with fewer mistakes. It also enables small but crucial situational differences in circumstance to be more quickly identified which may otherwise be hard to untangle.

7 Concluding Remarks

This short paper only intends to suggest the potentialities of analogical proportion-based decision. In that respect it is preliminary and much remains to be done for its effective use in practice. Moreover, the proposed approach belongs to a general trend of research that amounts to reasoning with data [10]. Another benefit of this approach relies in its explanation capabilities, i.e., that the robot will be able to explain what it is doing and why. Besides, interestingly enough, the use of quaternary relation where the variation (or the dissimilarity) from a to b is greater (or smaller) than the variation from c to d, rather than being equal to as in analogical proportion, has been also recently introduced in deep learning [8]. This might lead to fruitful developments.

References

1. Amgoud, L., D'Inverno, M., Osman, N., Prade, H., Sierra, C.: Experiences, a forgotten component of epistemic states. In: Ossowski, S., Toni, F., Vouros, G. (eds.) Proceedings of the 1st International Conference on Agreement Technologies (AT'), Dubrovnik, 15–16 October, CEUR Workshop Proceedings, vol. 918, pp. 220–230 (2012)
2. Dubois, D., Esteva, F., Garcia, P., Godo, L., López de Màntaras, R., Prade, H.: Fuzzy set modelling in cased-based reasoning. Int. J. Intell. Syst. **13**(4), 345–373 (1998)
3. 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)
4. Dubois, D., Hüllermeier, E., Prade, H.: Fuzzy methods for case-based recommendation and decision support. J. Intell. Inf. Syst. **27**, 95–115 (2006)
5. Dubois, D., Prade, H., Richard, G.: Case-based and analogical reasoning. In: Festschrift for Ramon Lopez de Mantaras' 60 years' birthday, pp. 59–77 (2012)
6. Dubois, D., Prade, H., Richard, G.: Multiple-valued extensions of analogical proportions. Fuzzy Sets Syst. **292**, 193–202 (2016)
7. Gilboa, I., Schmeidler, D.: Case-based decision theory. Q. J. Econ. **110**, 605–639 (1995)
8. Law, M.T., Thome, N., Cord, M.: Quadruplet-wise image similarity learning. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV), Sydney, 1–8 December, pp. 249–256 (2013)
9. Miclet, L., Prade, H.: Handling analogical proportions in classical logic and fuzzy logics settings. In: Sossai, C., Chemello, G. (eds.) ECSQARU 2009. LNCS (LNAI), vol. 5590, pp. 638–650. Springer, Heidelberg (2009). doi:10.1007/978-3-642-02906-6_55
10. Prade, H.: Reasoning with data - a new challenge for AI? In: Schockaert, S., Senellart, P. (eds.) SUM 2016. LNCS (LNAI), vol. 9858, pp. 274–288. Springer, Cham (2016). doi:10.1007/978-3-319-45856-4_19
11. Prade, H., Richard, G.: From analogical proportion to logical proportions. Logica Universalis **7**(4), 441–505 (2013)
12. Stroppa, N., Yvon, F.: Analogical learning and formal proportions: definitions and methodological issues. Technical report ENST D-2005-004, Paris, June 2005