On the Role of Similarity in Analogical Transfer

Fadi Badra, Karima Sedki, Adrien Ugon

HAL Id: hal-02459103

https://hal.inria.fr/hal-02459103

Submitted on 29 Jan 2020

On the Role of Similarity in Analogical Transfer

Fadi Badra[1], Karima Sedki[1], and Adrien Ugon[2]

[1] Université Paris 13, Sorbonne Université, Inserm, Laboratoire d'informatique médicale et d'ingénierie des connaissances en e-santé LIMICS, F-75006 Paris, France
{badra,karima.sedki}@univ-paris13.fr
[2] ESIEE-Paris, Noisy-le-Grand, France adrien.ugon@esiee.fr

Abstract. Analogical transfer consists in making the assumption that if two situations are alike in some respect, they may be alike in others. This paper studies analogical transfer from the viewpoint of measurement theory. We show that analogical transfer can be seen as a similarity-based reasoning on "contour lines" in a qualitative similarity model: if a contour line is included in another at a given point, then it may also be included at other points that are not too (analogically) dissimilar.
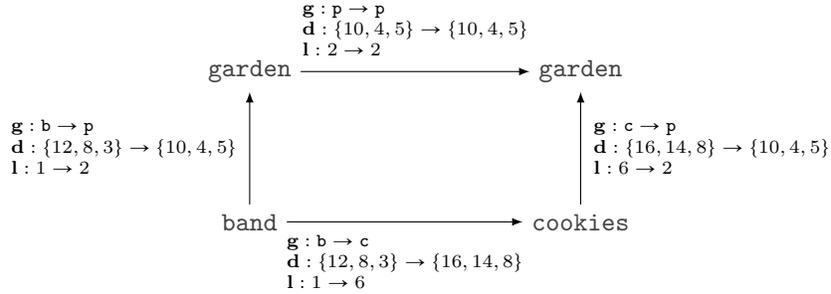
Keywords: Analogical Transfer · Qualitative Modeling · Similarity Measurement.

1 Introduction

Case-based reasoning (CBR) can be seen as the application of an analogical inference to problem solving. Analogical reasoning is a cognitive process in which some structural pattern identified in a source conceptualization is transferred to a target domain (possibly the same domain) in order to learn a target conceptualization [9]. Three main subtasks are identified: retrieval, mapping, and transfer. The aim of the retrieval phase is to use some memory to identify to which source situation the new (target) situation can be compared with. The mapping phase produces a mapping from the retrieved source situation to the target situation. The transfer phase uses this mapping to makes some prediction on how to complete the description of the target situation. Analogies in which mapping is followed with a transfer of information or explanations are called predictive analogies in [9].

Analogical transfer may be used to predict some *commonalities* between a source situation and a target situation. An example of predictive analogy that transfers a commonality is the k-Nearest Neighbor algorithm [19], an instance-based machine learning algorithm in which the class $cls(x_0)$ of a target instance x_0 is predicted from the class value $cls(x)$ of the k instances x that are most similar to x_0. But, analogical transfer is also used to predict *differences* between a target situation and the source situations retrieved from memory. Adaptation, *i.e.*, the cognitive ability to envision a target solution that is different from any previously encountered solution, is considered in cognitive psychology [20, 23], as one of the core components of analogical transfer. In the following example,

Fig. 1. The differences between the values of the three attributes goal (**g**), set of divisors (**d**), and left over (**l**) in the descriptions of some mathematical problems.

$$\begin{array}{l} \mathbf{g} : \mathtt{p} \to \mathtt{p} \\ \mathbf{d} : \{10, 4, 5\} \to \{10, 4, 5\} \\ \mathbf{l} : 2 \to 2 \end{array}$$

garden $\xrightarrow{\hspace{2cm}}$ garden

$$\begin{array}{l} \mathbf{g} : \mathtt{b} \to \mathtt{p} \\ \mathbf{d} : \{12, 8, 3\} \to \{10, 4, 5\} \\ \mathbf{l} : 1 \to 2 \end{array}$$

$$\begin{array}{l} \mathbf{g} : \mathtt{c} \to \mathtt{p} \\ \mathbf{d} : \{16, 14, 8\} \to \{10, 4, 5\} \\ \mathbf{l} : 6 \to 2 \end{array}$$

band $\xrightarrow{\hspace{2cm}}$ cookies

$$\begin{array}{l} \mathbf{g} : \mathtt{b} \to \mathtt{c} \\ \mathbf{d} : \{12, 8, 3\} \to \{16, 14, 8\} \\ \mathbf{l} : 1 \to 6 \end{array}$$

taken from [20], students adapt solution procedures to solve new mathematical problems. They are asked to adapt the solution procedure of a mathematical problem (called the *garden problem* [21]), in which the goal is to calculate the number of plants that two people have in their garden. The solution procedure consists in (i) computing the least common multiple (LCM) of a set of numbers (the LCM of 10, 4, and 5, is 20), (ii) generate the first multiples of the obtained number (20, 40, 60, etc.), (iii) add a left over to each multiple (the left over is 2, so we obtain 22, 42, 62, etc.), and (iv) find the smallest candidate solution that is divisible by a given number (the smallest divisible by 6 is 42). Two other mathematical problems, the *marching band* problem and the *cookies problem*, are structurally very similar to the the garden problem, but differ by the value given to three attributes: the goal (**g**), the set of divisors (**d**), and the left over (**l**). Fig. 1 represents the difference between these problems: `garden` denotes the description of the garden problem, `band` denotes the description of the marching band problem, and `cookies` denotes the description of the cookies problem. The cookies problem is solved by adapting the solution procedure of the garden problem. To do so, the student has to assess how to modify the solution procedure to account for the differences observed between the two problems. From the remembering of how the marching band problem has been previously solved by adapting the solution of the garden problem, most students were able to induce that a problem with the same structure as the garden problem may be solved by the same solution procedure, up to a substitution of attribute values. However, more complex differences between problems may require a deeper modification of the solution procedure, such as adding or removing steps. The authors give the example of the seashell problem, for which the number of multipliers is greater than the one of the garden problem. Adapting the solution of the garden problem to solve the seashell problem would require to add some reasoning steps to its solution procedure.

   Adaptation is recognized as being part of the case-based reasoning cycle. But surprisingly, a study of the CBR literature shows that the adaptation step

is not included in the case-based analogical inference [22]. The adaptation step is always performed *after* the analogical inference (*i.e.*, retrieval, mapping, and transfer) has taken place, and only aims at modifying its result. Some adaptation methods such as critique-based adaptation [10], or conservative adaptation [15] are used to resolve inconsistencies in the reused source case, whereas others such as differential adaptation [8], case-based adaptation [3] or adaptation by reformulation [16] modify the reused source case in order to fit the requirements on the target case. This singular status of adaptation in CBR research with respect to analogical transfer appears even more striking since many adaptation strategies apply to themselves an analogical process, by reusing adaptation knowledge retrieved from memory, in form of adaptation rules [4], substitutions [17], or adaptation cases [3]. One of the reasons why adaptation is left out of the case-based analogical inference is that adaptation essentially consists in reasoning on the *differences* that exist between two cases and the case-based inference is currently unable to transfer such differences to the target case. The case-based inference typically relies on a measure of similarity between cases, and while the importance of capturing and reusing case differences has long been acknowledged in adaptation research (see e.g.,[4, 8, 11, 14, 17], the link between case differences and a theory of similarity remains unclear. What is the role of similarity in analogical transfer?

In this article, analogical transfer is studied from the viewpoint of a qualitative measurement of similarity. We show that analogical transfer can be seen as a similarity-based reasoning on "contour lines" in a qualitative similarity model: if a contour line is included in another at a given point, then it may also be included at other points that are not too (analogically) dissimilar.

The paper is organized as follows. In the next section, we start with a formulation of analogical transfer as a rule of inference. To understand the role played by similarity in this rule of inference, Sec. 3 introduces qualitative measures of similarity. Sec. 4 shows that these measures can be used to measure analogical mappings. Sec. 5 defines analogical transfer as a reasoning on contour lines. Sec. 6 illustrates the approach on a use case, and Sec. 7 concludes the paper.

2   Analogical Transfer from a Logical Perspective

Analogical transfer is stated from a logical perspective in [5] as the following hypothetical rule of inference, referred to as the "analogical jump" (AJ):

$$\frac{P(s) \quad P(t) \quad Q(s)}{Q(t)} \tag{AJ}$$

The properties P and Q may represent commonalities. For example, Bob's car and John's car share the property P of being a 1982 Mustang GLX V6 hatchbacks, and Bob's car has the property Q of having a price of 3500 \$. The inference is that the price of John's car should also be around 3 500 \$. But P and Q may also represent differences. For example, in Fig. 2, adapted from [28], the pair $(\mathsf{a}, \mathsf{a}')$ can be mapped to the pair $(\mathsf{b}, \mathsf{b}')$ because b' differs from b in the same

Fig. 2. Six faces.
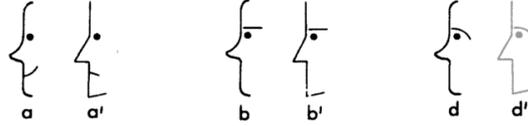
way that a' differs from a. Both pairs instanciate the two relational predicates $R_1=$" differs in profile shape from round to sharp" and $R_2=$" has the same eyebrow as". If we know that d' has a sharp profile (*i.e.*, $R_1(\mathsf{d},\mathsf{d}')$ holds), then assuming d' is to d what a' is to a, applying the rule (AJ) enables to make the hypothesis that d' has the same eyebrow (curved) as d, *i.e.*, that $R_2(\mathsf{d},\mathsf{d}')$ holds:

$$\frac{R_1(\mathsf{a},\mathsf{a}') \quad R_2(\mathsf{a},\mathsf{a}') \quad R_1(\mathsf{d},\mathsf{d}')}{R_2(\mathsf{d},\mathsf{d}')}$$

The authors of [5] claim that a sufficient condition for the inference (AJ) to be justified should: (i) be weaker than a generalization rule $\forall x P(x) \Rightarrow Q(x)$ (otherwise the inference is simply deductive), (ii) on the contrary, be stronger than an enumerative induction that consists in applying the rule $P \Rightarrow Q$ to t only because it was seen to hold on one example s, (iii) be in proportion to the *amount* of similarity between sources and targets, but similarity degrees are highly subjective and context-dependent. One criteria that seems satisfying is when a functional dependency holds between P and Q.

To be able to make an analogical jump, instance-based learning approaches apply a *similarity principle*, that states that the more similar two causes are, the more plausible it is that they have the same effect. In numerical settings, this property can be translated in the geometrical property that locally, a bounded similarity on problems results in a bounded similarity on solutions [13]. The inference schema (AJ) is applied on a neighborhood $\mathcal{N}(x)$ of the retrieved instance x:

$$\frac{x \in \mathcal{N}(x) \quad cls(x) = c \quad x_0 \in \mathcal{N}(x)}{cls(x_0) = c}$$

Another example of predictive analogy is the analogical proportion-based inference, used in analogical classification, which, as shown e.g., in [2], also consists in making an analogical jump.

3   Qualitative Similarity

Qualitative Similarity. A qualitative representation of similarity consists in dropping numerical values in similarity assessment, which is justified by the fact that actual values of distance between two cases are less significant than the ability to compare two distances. Let $\mathcal{U}$ be a finite, non-empty set, called the *universe*.

An element of the square product $\mathcal{U} \times \mathcal{U}$ is denoted (a, b), or simply ab. Following [29], a qualitative similarity relation $\succ$ is defined as a strict weak order on $\mathcal{U} \times \mathcal{U}$. For $a, b, c, d \in \mathcal{U}$, a strict ordering $ab \succ cd$ for a similarity relation $\succ$ is interpreted as "a and b are more similar than c and d", and represents a strict inequality. A strict weak order is a binary relation with the following two properties, for $a, b, c, d, e, f \in \mathcal{U}$: *asymetry* $(ab \succ cd \Rightarrow \neg (cd \succ ab))$, and *negative transitivity* $(\neg (ab \succ cd) \land \neg (cd \succ ef) \Rightarrow \neg (ab \succ ef))$. The axioms of asymetry and negative transitivity define a form of *rationality* of a similarity comparison [27]. For a strict weak order $\succ$, let $\sim$ be a binary relation, called the *indifference* relation, defined as: $ab \sim cd \Leftrightarrow \neg (ab \succ cd) \land \neg (cd \succ ab)$. An element $ab \sim cd$ is interpreted as "a and b are as similar as c and d". The indifference relation $\sim$ is an equivalence relation: it is reflexive, transitive, and symmetric. Each equivalence relation defines a partition, which is a collection of disjoint subsets of a set whose union is the whole set. The subset of $\sim$ that contains a pair $ab \in \mathcal{U} \times \mathcal{U}$ is written $[ab]_\sim = \{uv \in \mathcal{U} \times \mathcal{U} \mid ab \sim uv\}$, and is called the equivalence class of ab. The relation $\succeq$ represents a (non-strict) inequality, and is defined as $ab \succeq cd$ iff $ab \succ cd$ or $ab \sim cd$. Because $\succ$ is strict weak, there holds that for any $a, b, c, d \in \mathcal{U}$, exactly one of $ab \succ cd$, $cd \succ ab$, or $ab \sim cd$ hold. A direct consequence of the negative transitivity of $\succ$ is that an (in)equality $ab \succeq cd$ (resp., $ab \succ cd$, $ab \sim cd$) between two pairs ab and cd implies an (in)equality between the two equivalence classes $[ab]_\sim$ and $[cd]_\sim$:

$$[ab]_\sim \succeq [cd]_\sim \Leftrightarrow \forall uv \in [ab]_\sim, u'v' \in [cd]_\sim, uv \succeq u'v'$$
$$\Leftrightarrow ab \succeq cd$$

Scaling. A natural way to discretize a set is to apply a discrete-valued function. Such a function should preserve the underlying structure of the set. The set $(\mathcal{U} \times \mathcal{U}, \succeq)$ is a relational structure, *i.e.*, a set along with a collection of finitary operations and binary relations. A *scale* is an homomorphism between two relational structures [27].

Definition 1. *A* variation *is a scale* $\upsilon : (\mathcal{U} \times \mathcal{U}, \succeq) \longrightarrow (\mathcal{V}, \succeq)$.

A homomorphism is a mapping that preserves all relations and operations. Therefore, for $a, b, c, d \in \mathcal{U}$,

$$ab \succeq cd \Leftrightarrow \upsilon(ab) \succeq \upsilon(cd)$$

Variations can be used to describe a wide range of similarity relations, from the most specific one $(\upsilon(ab) = ab)$ to the most general $(\upsilon(ab) = 1)$. In this paper, we will restrict the discussion to variations that are constructed by scaling the set of pairs of values taken by a given feature[3]. A feature is a variable on the set $\mathcal{U}$, and can be modeled as a function $\varphi : \mathcal{U} \longrightarrow \mathcal{X}$. In the following, the term *feature* denotes either a binary variable $(\mathcal{X} = \{0, 1\})$, a nominal variable $(\mathcal{X}$ is

---

[3] The latter are a very common way to represent differences between two objects, although not the only one. Variations could for example represent complex rewriting rules, such as term reduction relations.

Table 1. Some examples of scales $\mathbf{o} : \mathcal{X} \times \mathcal{X} \longrightarrow \mathcal{V}$ defined on the square product of a feature space $\mathcal{X}$.

Feature space	Examples
All	$= (xy) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$
	$\mathbf{1}(xy) = \begin{cases} 1 & \text{if } x = y = 1 \\ 0 & \text{otherwise} \end{cases}$
	$Id(xy) = xy$
Ordinal	$\leq (xy) = \begin{cases} 1 & \text{if } x \leq y \\ 0 & \text{otherwise} \end{cases}$
Continuous	$- (xy) = y - x$
	$Sign(xy) = \begin{cases} + & \text{if } x < y \\ \text{o} & \text{if } x = y \\ - & \text{if } x > y \end{cases}$
	$d_p(xy) = \|y - x\|_p$
Sets	$\Delta(XY) = |X \backslash Y| + |Y \backslash X|$
	$AP(XY) = (X \backslash Y , Y \backslash X)$
	$s(XY) = (X \backslash Y , X \cap Y , Y \backslash X)$

an enumerated set, like $\{\mathtt{d}, \mathtt{m}, \mathtt{u}\}$), a quantity ($\mathcal{X}$ is a set of values on an ordinal, interval, or ratio scale), or a variable that takes its value in the powerset $\mathscr{P}(\mathcal{A})$ of a set $\mathcal{A}$. The term *feature space* denotes the set $\mathcal{X}$ of values taken by a feature.

Proposition 1. *Given a feature $\varphi : \mathcal{U} \to \mathcal{X}$, a scale $\mathbf{o} : \mathcal{X} \times \mathcal{X} \longrightarrow \mathcal{V}$, and a strict weak order $>$ on $\mathcal{V}$, the function $\upsilon_\mathbf{o}^\varphi : \mathcal{U} \times \mathcal{U} \longrightarrow \mathcal{V}$ defined by*

$$\upsilon_\mathbf{o}^\varphi(ab) = \mathbf{o}(\varphi(a), \varphi(b))$$

is a variation iff $ab > cd \Leftrightarrow \varphi(a)\varphi(b) >_{\mathcal{X} \times \mathcal{X}} \varphi(c)\varphi(d)$, where $>_{\mathcal{X} \times \mathcal{X}}$ is the (strict weak) order induced on $\mathcal{X} \times \mathcal{X}$ by the scale $\mathbf{o}$ (i.e., $\mathbf{o}(xy) > \mathbf{o}(zt) \Leftrightarrow xy >_{\mathcal{X} \times \mathcal{X}} zt$).

Proof. (Sketch) For two pairs ab and cd of $\mathcal{U} \times \mathcal{U}$, $\varphi(a)\varphi(b) \in \mathcal{X} \times \mathcal{X}$, and $\varphi(c)\varphi(d) \in \mathcal{X} \times \mathcal{X}$. The relation $>$ on $\mathcal{V}$ is strict weak, so exactly one of $\varphi(a)\varphi(b) >_{\mathcal{X} \times \mathcal{X}} \varphi(c)\varphi(d), \varphi(a)\varphi(b) <_{\mathcal{X} \times \mathcal{X}} \varphi(c)\varphi(d)$, or $\varphi(a)\varphi(b) \sim_{\mathcal{X} \times \mathcal{X}} \varphi(c)\varphi(d)$ holds.

Examples. Tab. 1 gives some examples of scales $\mathbf{o}$ defined on the set $\mathcal{X} \times \mathcal{X}$. The variation $\upsilon_{\underline{=}}^{\mathtt{eyebrow}}$ can be defined from the nominal feature $\varphi = \mathtt{eyebrow} : \mathcal{U} \longrightarrow \mathcal{X}$ and the scale $=: \mathcal{X} \times \mathcal{X} \longrightarrow \{0, 1\}$ as follows:

$$\upsilon_{\underline{=}}^{\mathtt{eyebrow}}(ab) = \begin{cases} 1 & \text{if } \mathtt{eyebrow}(a) = \mathtt{eyebrow}(b) \\ 0 & \text{if } \mathtt{eyebrow}(a) \neq \mathtt{eyebrow}(b) \end{cases}$$

The indicator function $\mathbf{1}_P$ of the class "sharing the property P" can be defined from the feature $\varphi = P$ and the scale $\mathbf{1} : \mathcal{X} \times \mathcal{X} \longrightarrow \{0, 1\}$ as follows:

$$\mathbf{1}_P(ab) = \upsilon_{\mathbf{1}}^P(ab) = \begin{cases} 1 & \text{if both } P(a) \text{ and } P(b) \text{ hold} \\ 0 & \text{otherwise} \end{cases}$$

The variation $\upsilon_{Id}^{\texttt{color}}(ab) = (\texttt{color}(a), \texttt{color}(b))$ can be defined from the nominal feature $\varphi = \texttt{color} : \mathcal{U} \longrightarrow \mathcal{X}$ with $\mathcal{X} = \{\texttt{r}, \texttt{g}, \texttt{b}\}$, and the scale $Id(xy) = xy$. The variation $\upsilon_{d_1}^{\texttt{age}}(xy) = |\texttt{age}(y) - \texttt{age}(x)|$ can be defined from the numerical feature $\varphi = \texttt{age} : \mathcal{U} \longrightarrow \mathbb{N}$ and the scale $d_1 : \mathbb{N} \times \mathbb{N} \longrightarrow \mathbb{N}$ where d_1 is the Manhattan distance. This function returns a natural number representing the (absolute) difference in value of the property $\texttt{age}$. Let $\varphi_1, \varphi_2, \ldots, \varphi_n$ be n numerical features, and $\mathbf{v}$ be the function such that $\mathbf{v}(a) = (\varphi_1(a), \varphi_2(a), \ldots, \varphi_n(a)) \in \mathbb{R}^n$ for $a \in \mathcal{U}$. The variation $\overrightarrow{ab} = \upsilon_{-}^{\mathbf{v}}(ab) = \mathbf{v}(b) - \mathbf{v}(a)$ gives the difference between two vectors. Let $\succeq_{a_\rightarrow}$ be the restriction of $\succeq$ to $\{a\} \times \mathcal{U}$. An inequality $ab \succeq_{a_\rightarrow} ac$ is interpreted as "b is at least as similar to a as c is to a". If the order $\succeq_{a_\rightarrow}$ is a complete preorder on $\{a\} \times \mathcal{U}$, there exists a function $u : \{a\} \times \mathcal{U} \longrightarrow \mathbb{R}$, called a *utility function*, such that for all $b, c \in \mathcal{U}$, $ab \succ ac \Leftrightarrow u(ab) > u(ac)$. The utility function can be defined for example using a distance measure $\|.\|_p$ on $\mathcal{U}$ by taking $u = \upsilon_{d_p}^{\mathbf{v}}$ where $\upsilon_{d_p}^{\mathbf{v}}(ab) = \|\overrightarrow{ab}\|_p$ gives the distance between a and b. Consider the function $\varphi = a_\rightarrow : \mathcal{U} \longrightarrow \mathbb{N}^+$ that gives for each $p \in \mathcal{U}$ the size of the principal ideal $\downarrow ap = \{ab \in \{a\} \times \mathcal{U} \mid ap \succeq ab\}$. The variation $\upsilon_{d_1}^{a_\rightarrow}(bc) = |size(\downarrow ac) - size(\downarrow ab)|$ gives a numerical measure of distance between two elements b and c when a is chosen as a frame of reference. In particular, if $b = a$, $\upsilon_{d_1}^{a_\rightarrow}(ac) = |size(\downarrow ac) - 1|$ gives a distance from a to c in the sense of the nearest neighbors: $\upsilon_{d_1}^{a_\rightarrow}(ac) = k$ iff c is the $k-$Nearest Neighbor of a. The scale $AP(XY) = (X \backslash Y, Y \backslash X)$ represents the difference from a set X to a set Y by "what is lost" from X to Y and "what is gained" from X to Y. The scale $\Delta(XY) = |X \backslash Y| + |Y \backslash X|$ measures the symmetric difference between X and Y. If the sets $X = (X_1, X_2, \ldots, X_n)$ and $Y = (Y_1, Y_2, \ldots, Y_n)$ are elements of a cartesian product $A_1 \times A_2 \times \ldots A_n$, the scale Δ is given by $\Delta(XY) = \sum_i (|X_i \backslash Y_i| + |Y_i \backslash X_i|)$. The function $s(X, Y) = (X \backslash Y, X \cap Y, Y \backslash X)$ corresponds to what Amos Tversky called a *matching function* in [28].

4   Mapping

Analogical Equalities. An equivalence class of a variation can be seen as a *contour line* in the underlying similarity model. An analogical equality $ab \sim_\upsilon cd$ holds if d differs from c as b differs from a according to the scale υ. The equivalence class $[ab]_\upsilon = \{cd \in \mathcal{U} \times \mathcal{U} \mid \upsilon(ab) \sim_\upsilon \upsilon(cd)\}$ of a pair ab groups a set of pairs of $\mathcal{U} \times \mathcal{U}$ which are "equally different", *i.e.*, which differ from each other as much as b differs from a. Mapping a pair ab to a pair cd consists in establishing an analogical equality, *i.e.*, finding a variation υ for which the two pairs belong to the same contour line for υ:

$$\text{Find } \upsilon \text{ such that } ab \sim_\upsilon cd \qquad \text{(Mapping)}$$

If $\sim_\upsilon$ is the identity relation ($ab \sim_\upsilon cd \Leftrightarrow \upsilon(ab) = \upsilon(cd)$), then each value $\ell \in \mathcal{V}$ defines an equivalence class denoted by

$$(\upsilon : \ell) = \{ab \in \mathcal{U} \times \mathcal{U} \mid \upsilon(ab) = \ell\}$$

Sharing a difference defines a contour line. In Fig. 2, the pair $\mathtt{aa}'$ can be mapped to the pair $\mathtt{bb}'$ since $\upsilon_{\underline{\underline{\cong}}}^{\mathtt{eyebrow}}(\mathtt{aa}') = \upsilon_{\underline{\underline{\cong}}}^{\mathtt{eyebrow}}(\mathtt{bb}') = 1$, so they belong to the same equivalence class $(\upsilon_{\underline{\underline{\cong}}}^{\mathtt{eyebrow}} : 1)$ for the variation $\upsilon_{\underline{\underline{\cong}}}^{\mathtt{eyebrow}}$. Similarly, a contour line of the variation $\upsilon_{\underline{\vee}}^{\mathtt{v}}$ groups in the same equivalence class the pairs ab and cd when their difference is represented by the same vector (*i.e.*, such that $\overrightarrow{ab} = \overrightarrow{cd}$). Sharing a commonality (like having a property P) also defines a contour line. Let $s \in \mathcal{U}$ be an element that has a property P (*i.e.*, such that $P(s)$ holds). The variation $\mathbf{1}_P$ can be used to map the pair ss to all pairs $st \in [ss]_{\mathbf{1}_P} = (\mathbf{1}_P : 1)$ (and thus s to t). An analogical proportion also defines a contour line. The analogical proportion $X : Y :: Z : T$ is defined [25] by $X \backslash Y = Z \backslash T$ and $Y \backslash X = T \backslash Z$, which means that the analogical proportion $X : Y :: Z : T$ holds for two pairs XY and ZT iff the two pairs take the same value for the variation $AP(XY) = (X \backslash Y, Y \backslash X)$. Similarly, in preference theory, the pairs of transactions satisfying the same conditional preference rules constitute a contour line. Suppose $\mathcal{U}$ denotes a set of objects called *transactions*. Let $\mathcal{I}$ be a set of distinct literals (called *items*), and $\varphi : \mathcal{U} \longrightarrow \mathscr{P}(\mathcal{I})$ the function that associates to each transaction a set of items. A transaction t is said to be preferred to u according to a conditional preference rule $\pi : i^+ > i^- \mid C$ iff $(i^+, C, i^-) \subseteq s(\varphi(t), \varphi(u))$ with $s(XY) = (X \backslash Y, X \cap Y, Y \backslash X)$ [6]. As a result, two transactions that take the same value for this variation satisfy the same conditional preference rules.

Analogical Inequalities. The strict weak relation $>$ on $\mathcal{V}$ defines a linear order which acts as a *contour scale* on $\mathcal{U} \times \mathcal{U}$. Inequalities define a linear order on contour lines (since $ab \geq cd \Leftrightarrow [ab]_\sim \geq [cd]_\sim$). For example, in Fig. 2, two faces having the same eyebrow may be more similar than two faces having different eyebrows, *i.e.*, $1 \geq 0 \Leftrightarrow (\upsilon_{\underline{\underline{\cong}}}^{\mathtt{eyebrow}} : 1) \geq (\upsilon_{\underline{\underline{\cong}}}^{\mathtt{eyebrow}} : 0)$. Similarly, by taking the usual order $\leq$ on $\mathbb{N}$, one can define a contour scale: $n \leq m \Leftrightarrow (\upsilon_{d_1}^{\mathtt{age}} : n) \geq (\upsilon_{d_1}^{\mathtt{age}} : m)$. This scale expresses that "the lower the age difference, the more similar". By structuring the set $\mathscr{P}(\mathcal{A}) \times \mathscr{P}(\mathcal{A})$ with the inclusion relation $\subseteq$ defined by $(A, B) \subseteq (A', B')$ iff $A \subseteq A'$ and $B \subseteq B'$, a contour scale can be defined on analogical proportions: $\ell \subseteq k \Leftrightarrow (AP : \ell) \geq (AP : k)$. Such a scale formalizes that all things equal, the less properties are lost or gained when going from a set X to a set Y, the more similar X and Y are. For example, if $X = \{\mathtt{a}, \mathtt{b}\}$, $Y = \{\mathtt{b}, \mathtt{d}\}$, $Z = \{\mathtt{a}, \mathtt{c}\}$, and $T = \{\mathtt{c}\}$, then $ZT \geq XY$ since $AP(ZT) = \{\{\mathtt{a}\}, \{\}\} = \ell \subseteq AP(X, Y) = \{\{\mathtt{a}\}, \{\mathtt{d}\}\} = k$. The relation $[ab]_\sim \geq [cd]_\sim$ is noted $a : b \ll c : d$ in [26]. According to Tversky, a matching function should enforce a *monotony* property. This property is equivalent to the definition of a contour scale. By structuring the set $\mathscr{P}(\mathcal{A}) \times \mathscr{P}(\mathcal{A}) \times \mathscr{P}(\mathcal{A})$ with the inclusion relation $\subseteq$ defined by $(A, B, C) \subseteq (A', B', C')$ iff $A \subseteq A'$, $B \supseteq B'$, and $C \subseteq C'$ a contour scale on conditional preference rules can be defined as: $\ell \subseteq k \Leftrightarrow (s : \ell) \geq (s : k)$.

Table 2. Four subsets X, Y, Z, and T.

	a	b	c	d
X	1	1	0	0
Y	0	0	1	0
Z	0	1	1	0
T	1	0	1	1

Analogical Dissimilarity. Analogical dissimilarity measures the distance between two contour lines. Analogical dissimilarity is given by a scale $\mathtt{ad}$ on $\mathcal{V}$, which is coherent with analogy, *i.e.*, $\mathtt{ad}(ab, cd) = 0 \Leftrightarrow ab \sim_\upsilon cd$. As a consequence, if $ab \sim_\upsilon a'b'$ and $cd \sim_\upsilon c'd'$, then $\mathtt{ad}(ab, cd) = \mathtt{ad}(a'b', c'd')$. If $\sim_\upsilon$ is the identity relation, the previous observation implies that $\mathtt{ad}(ab, cd)$ gives the same value for all $ab \in (\upsilon : \ell)$ and $cd \in (\upsilon : \ell')$ taken in two different contour lines. Therefore, we will write:

$$\mathtt{ad}(\ell, \ell') = \mathtt{ad}(ab, cd) \text{ for } ab \in (\upsilon : \ell) \text{ and } cd \in (\upsilon : \ell')$$

The variation $\upsilon_{d_1}^{a\rightarrow}(bc) = |size(\downarrow ac) - size(\downarrow ab)|$ gives a numerical measure of distance between b and c with respect to a. The analogical dissimilarity is given by

$$\mathtt{ad}(ab, cd) = \upsilon_{d_1}^{\upsilon_{d_1}^{a\rightarrow}}(ab, cd) \qquad \text{for } a, b, c, d \in \mathcal{U}$$

$$\textit{i.e.,} \qquad \mathtt{ad}(n, m) = |n - m| \qquad \text{for } n, m \in \mathbb{N}$$

The analogical dissimilarity between two vectors [18] gives the difference between two contour lines:

$$\mathtt{ad}(ab, cd) = \upsilon_{d_p}^{\upsilon^{\vee}}(ab, cd) = \|\overrightarrow{ab} - \overrightarrow{cd}\|_p$$

$$\textit{i.e.,} \qquad \mathtt{ad}(\overrightarrow{u}, \overrightarrow{v}) = \|\overrightarrow{v} - \overrightarrow{u}\|_p$$

If the feature $\varphi : \mathcal{U} \longrightarrow \mathscr{P}(\mathcal{A})$ associates to each element of $\mathcal{U}$ a subset of a set $\mathcal{A}$, the analogical dissimilarity between analogical proportions is given by a symmetric difference:

$$\mathtt{ad}(ab, cd) = \upsilon_{\Delta}^{\upsilon_{AP}^{\varphi}}(ab, cd)$$
$$= |U \backslash W| + |W \backslash U| + |V \backslash Q| + |Q \backslash V|$$
$$\text{with } U = X \backslash Y,\ V = Y \backslash X,\ W = Z \backslash T,\ \text{ and } Q = T \backslash Z$$
$$\text{where } X = \varphi(a),\ Y = \varphi(b),\ Z = \varphi(c),\ \text{ and } T = \varphi(d)$$

To illustrate, take $\mathcal{A} = \{\mathsf{a}, \mathsf{b}, \mathsf{c}, \mathsf{d}\}$, and $a, b, c, d \in \mathcal{U}$ such that $X = \varphi(a) = \{\mathsf{a}, \mathsf{b}\}$, $Y = \varphi(b) = \{\mathsf{c}\}$, $Z = \varphi(c) = \{\mathsf{b}, \mathsf{c}\}$, and $T = \varphi(d) = \{\mathsf{a}, \mathsf{c}, \mathsf{d}\}$ (Tab. 2). The scale AP gives $AP(XY) = UV$ with

- $U = X\backslash Y = \{\mathsf{a,b}\}$ is what is lost from X to Y
- $V = Y\backslash X = \{\mathsf{c}\}$ is what is gained from X to Y

and $AP(ZT) = WQ$ with

- $W = Z\backslash T = \{\mathsf{b}\}$ is what is lost from Z to T
- $Q = T\backslash Z = \{\mathsf{a,d}\}$ is what is gained from Z to T

The analogical dissimilarity between ab and cd is $\mathtt{ad}(ab,cd) = 1+0+1+2 = 4$. If the feature $\varphi : \mathcal{U} \longrightarrow \mathscr{P}(\mathcal{I})$ associates to each transaction a set of items, the analogical dissimilarity between two pairs of transactions is given by the symmetric difference:

$$
\begin{aligned}
\mathtt{ad}(ab,cd) &= \upsilon_{\Delta^s}^{\upsilon^\varphi}(ab,cd)\\
&= |U\backslash W| + |W\backslash U| + |R\backslash S| + |S\backslash R|\\
&\quad + |V\backslash Q| + |Q\backslash V|\\
&\text{with } U = X\backslash Y,\ V = Y\backslash X,\ R = X\cap Y,\ S = Z\cap T,\\
&W = Z\backslash T,\ \text{ and } Q = T\backslash Z\\
&\text{where } X = \varphi(a),\ Y = \varphi(b),\ Z = \varphi(c),\ \text{ and } T = \varphi(d)
\end{aligned}
$$

For example, if $\mathcal{I} = \{\mathsf{a,b,c,d}\}$, and $a,b,c,d \in \mathcal{U}$ are defined as above by $X = \varphi(a) = \{\mathsf{a,b}\}$, $Y = \varphi(b) = \{\mathsf{c}\}$, $Z = \varphi(c) = \{\mathsf{b,c}\}$, and $T = \varphi(d) = \{\mathsf{a,c,d}\}$, then

- $R = X \cap Y = \varnothing$ is what is preserved X to Y
- $S = Z \cap T = \{\mathsf{c}\}$ is what is preserved from Z to T

The analogical dissimilarity between ab and cd is $\mathtt{ad}(ab,cd) = 1+0+0+1+1+2 = 5$. Therefore, we can say that the distance between the two conditional preference rules $\pi_1 : i^+{}_1 > i^-{}_1 \mid C_1$ and $\pi_2 : i^+{}_2 > i^-{}_2 \mid C_2$ is

$$
\mathtt{ad}(\pi_1,\pi_2) = \left|i^+{}_1\ \Delta\ i^+{}_2\right| + \left|i^-{}_1\ \Delta\ i^-{}_2\right| + |C_1\ \Delta\ C_2|
$$

For example, the distance between $\pi_1 : \mathsf{c} > \mathsf{ab} \mid \varnothing$ and $\pi_2 : \mathsf{ad} > \mathsf{b} \mid \mathsf{c}$ is 5. Such measure could be useful e.g., to predict preferences in a transductive way [24].

5   Analogical Transfer

A situation of special interest is when two contour lines intersect. Let $\Upsilon = \{\upsilon_1, \upsilon_2, \dots, \upsilon_n\}$ be a set of n variations with $\upsilon_i : \mathcal{U} \times \mathcal{U} \longrightarrow \mathcal{V}_i$, and $n \geq 2$. For $\upsilon_i, \upsilon_j \in \Upsilon$, the inclusion of two contour lines at a point ab is called a *co-variation* and is written $\upsilon_i \overset{ab}{\rightharpoonup} \upsilon_j$. We say that the variation υ_i *co-varies* with the variation υ_j at ab. The notion of co-variation at a point $ab \in \mathcal{U} \times \mathcal{U}$ can be extended to a subset $S \subseteq \mathcal{U} \times \mathcal{U}$ by stating that:

$$
\upsilon_i \overset{S}{\rightharpoonup} \upsilon_j \text{ iff } \upsilon_i \overset{ab}{\rightharpoonup} \upsilon_j \text{ for all } ab \in S
$$

Fig. 3. The co-variation $(\upsilon_i : \alpha) \overset{ab}{\rightharpoonup} (\upsilon_j : \gamma)$ can be interpreted with different semantics on the set $\Upsilon = \{\upsilon_i, \upsilon_j, \upsilon_k, \upsilon_m\}$. Hatched areas on the figure represent the resulting domains of validity of the co-variation. **a)** Bayesian semantics: the co-variation is verified on the whole equivalence class $(\upsilon_i : \alpha)$ **b)** Semi-bayesian semantics: the co-variation is verified only on a subset of the equivalence class where some other variations (here, υ_m) keep the same value **c)** Ceteris paribus semantics: the co-variation is verified on the subset of the equivalence class where *all* other variations keep the same value.
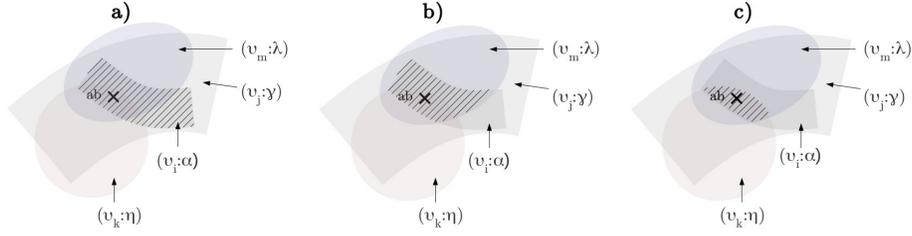
If each $\sim_{\upsilon_i}$ is the identity relation, then a co-variation is a functional dependency [1] between the two values $\ell = \upsilon_i(ab)$ and $k = \upsilon_j(ab)$, and the co-variation $\upsilon_i \overset{ab}{\rightharpoonup} \upsilon_j$ can be written $(\upsilon_i : \ell) \overset{ab}{\rightharpoonup} (\upsilon_j : k)$. For example, the co-variation $(\upsilon_{Id}^{\texttt{profile}} : (\texttt{round}, \texttt{sharp})) \overset{\{aa', bb'\}}{\rightharpoonup} (\upsilon_{=}^{\texttt{eyebrow}} : 1)$ represents the fact that when the profile changes from round to sharp, the eyebrow stays the same.

Different semantics can be given to a co-variation, depending on how it is interpreted with respect to the inclusion of equivalence classes $[ab]_{\upsilon_i} \subseteq [ab]_{\upsilon_j}$. Such an inclusion of equivalence classes is known as the *similarity principle*, and states that the equivalence relations $\sim_{\upsilon_i}$ and $\sim_{\upsilon_j}$ are locally co-monotone, *i.e.*, for all $a, b, c, d \in \mathcal{U}$, $ab \sim_{\upsilon_i} cd \Rightarrow ab \sim_{\upsilon_j} cd$. This principle expresses that if a and b are as similar as c and d for υ_i, then a and b are also as similar as c and d for υ_j. Therefore, different semantics can be given to a co-variation depending on which subset $R \subseteq [ab]_{\upsilon_i}$ of the equivalence class $[ab]_{\upsilon_i}$ the similarity principle is to be verified (Fig. 3):

- *Bayesian semantics*: $R = [ab]_{\upsilon_i}$, *i.e.*, the co-variation is verified on the whole equivalence class ;
- *Semi-bayesian semantics*: $R = \bigcap_{k \in \{i\} \cup D} [ab]_{\upsilon_k}$, *i.e.*, the co-variation is verified *all things equal*, but only in the subset of $[ab]_{\upsilon_i}$ where a set of variations υ_k with $k \in D$ take the same value [30] ;
- *Ceteris paribus semantics*: $R = \bigcap_{k \neq j} [ab]_{\upsilon_k}$, *i.e.*, the co-variation is verified *all things equal*, in the subset of $[ab]_{\upsilon_i}$ where all other variations take the same value.

Performing an analogical transfer consists in taking the similarity principle as a local inference rule, in order to make the hypothesis that if a contour line $[ab]_{\upsilon_i}$ is included in a contour line $[ab]_{\upsilon_j}$ at a point ab, then $[ab]_{\upsilon_i}$ should also be included in $[ab]_{\upsilon_j}$ at (at least some of) other points cd of $[ab]_{\upsilon_i}$ (*i.e.*, in some

points cd for which $\mathtt{ad}(ab, cd) = 0$), and even maybe at points that are not too dissimilar (*i.e.*, in some points cd for which $\mathtt{ad}(ab, cd) \leq k$):

$$\frac{\upsilon_i \overset{ab}{\leadsto} \upsilon_j,\ \mathtt{ad}(ab, cd) \leq k}{\upsilon_i \overset{cd}{\leadsto} \upsilon_j} \qquad\qquad \text{(Transfer)}$$

If each $\sim_{\upsilon_i}$ is the identity relation, then (Transfer) gives:

$$\frac{(\upsilon_i : \ell_0) \overset{ab}{\leadsto} (\upsilon_j : \gamma),\ cd \in (\upsilon_i : \ell) \text{ such that } \mathtt{ad}(\ell_0, \ell) \leq k}{(\upsilon_i : \ell) \overset{cd}{\leadsto} (\upsilon_j : \gamma)}$$

The analogical "jump" (AJ) is a the expression of an analogical transfer where $\upsilon_i = \mathbf{1}_P$, and $\upsilon_j = \mathbf{1}_Q$:

$$\frac{(\mathbf{1}_P : 1) \overset{ss}{\leadsto} (\mathbf{1}_Q : 1),\ st \in (\mathbf{1}_P : 1),\ \text{with } \mathtt{ad}(1, 1) = 0}{(\mathbf{1}_P : 1) \overset{st}{\leadsto} (\mathbf{1}_Q : 1)}$$

By applying the co-variation $\mathbf{1}_P \overset{st}{\leadsto} \mathbf{1}_Q$, the state of sharing the property Q is transferred to st. The co-variation is a functional dependency, which is consistent with the observation of [5] that a sufficient condition for an analogical jump (AJ) to be justified is that there exists a functional dependency holding between P and Q. Differences may also be used in the transfer. In Fig. 2, if we know that $\mathtt{d}'$ has a sharp profile, we can make the hypothesis that $\mathtt{d}'$ has the same eyebrow as $\mathtt{d}$ (curved), since $\mathtt{d}$ has a round profile and $\mathtt{d}'$ has a sharp profile:

$$\frac{(\upsilon_{Id}^{\mathtt{profile}} : (\mathtt{round}, \mathtt{sharp})) \overset{\{\mathtt{aa}', \mathtt{bb}'\}}{\leadsto} (\upsilon_{\equiv}^{\mathtt{eyebrow}} : 1)}{\mathtt{dd}' \in (\upsilon_{Id}^{\mathtt{profile}} : (\mathtt{round}, \mathtt{sharp}))}$$
$$\overline{(\upsilon_{Id}^{\mathtt{profile}} : (\mathtt{round}, \mathtt{sharp})) \overset{\mathtt{dd}'}{\leadsto} (\upsilon_{\equiv}^{\mathtt{eyebrow}} : 1)}$$

A fortiori reasoning is a type of inference that exploits the co-monotony of two partial orders to estimate the value of an attribute. In [7], the authors give the following example: "If whiskey is stronger than beer, and buying beer is illegal under the age of 18, then we can plausibly derive that buying whiskey is illegal under the age of 18". This inference consists in applying the co-variation $(\upsilon_{\leq}^{\mathtt{degree}} : 1) \overset{\mathtt{bw}}{\leadsto} (\upsilon_{\leq}^{\mathtt{legal\_age}} : 1)$, where $\mathtt{b}$ stands for beer, $\mathtt{w}$ stands for whiskey, and $\mathtt{degree}$ and $\mathtt{legal\_age}$ are two numerical features that associate to alcohols their degree and minimum legal age of consumption. Suppose now that we want to estimate the minimum legal age for cider (denoted by $\mathtt{c}$). If we know that cider is less strong than beer ($\mathtt{degree}(\mathtt{c}) \leq \mathtt{degree}(\mathtt{b})$), then one can use analogical transfer to infer make the hypothesis that the minimum legal age for cider is less than the one for beer (18):

$$\frac{(\upsilon_{\leq}^{\mathtt{degree}} : 1) \overset{\mathtt{bw}}{\leadsto} (\upsilon_{\leq}^{\mathtt{legal\_age}} : 1),\ \mathtt{cb} \in (\upsilon_{\leq}^{\mathtt{degree}} : 1)}{(\upsilon_{\leq}^{\mathtt{degree}} : 1) \overset{\mathtt{cb}}{\leadsto} (\upsilon_{\leq}^{\mathtt{legal\_age}} : 1)}$$

Analogical classifiers and instance-based-learning methods apply the similarity principle to the *nearest analogical neighbors* [12]. In instance-based learning, the relation "belonging to the same class" is transferred from an element x to an element x_0 whenever x_0 is found to be in the neighborhood of x. For example, in the k-Nearest Neighbor algorithm, $(\upsilon_{d_p}^{x\rightarrow} : n)$ is the equivalence relation that relates the pairs xy that are the nth most distant from x according to $\|.\|_p$, and $(\upsilon_{=}^{cls} : 1)$ is the equivalence relation that relates all elements that share the same class. The rule (Transfer) writes:

$$\frac{(\upsilon_{d_p}^{x\rightarrow} : 0) \overset{xx}{\rightsquigarrow} (\upsilon_{=}^{cls} : 1), xx_0 \in (\upsilon_{d_p}^{x\rightarrow} : n) \text{ with } \mathtt{ad}(0,n) \leq k}{(\upsilon_{d_p}^{x\rightarrow} : n) \overset{xx_0}{\rightsquigarrow} (\upsilon_{=}^{cls} : 1)}$$

6 Experiments

In this experiment, an expert formulated some knowledge the he/she uses in the interpretation of polysomnographies, and we want to check to which extent this knowledge fits the actual data. The knowledge acquired from the expert expresses the fact that the transition between wakefulness (stage W) and light sleep stage N_1 is characterized by a decrease of α waves on the occipital lobe, and/or a decrease of the quantity of eye movements (left or right) and/or a decrease of muscle tonus.

Materials. Ten subjects showing no clinical sign of sleep disorders spent three consecutive nights at the hospital to monitor their sleep with a polysomnography. Afterwards, an experienced sleep expert scored all the curves to generate a hypnogram, which is the succession of sleep stages during the night by 30-second epochs. Our work focused on sleep transitions, which are more difficult to describe since they are not similar to any given sleep stage, but curves are composed of waveforms and sleep events varying from one state to another, which makes their study of interest. Let $\mathcal{U}$ be the set of epochs of a given patient. Each epoch e_i is described by 18 numerical attributes $\varphi_1, \varphi_2, \ldots, \varphi_{18}$, described by a categorical attribute $\mathtt{s} : \mathcal{U} \longrightarrow \{W, N_1, N_2, N_3, R\}$ giving the associated sleep stage: W (Wakefulness), N_1, N_2 (light sleep), N_3 (deep sleep), and R (REM, Rapid Eye Movements Sleep).

Method. Among the attributes $\varphi_1, \varphi_2, \ldots, \varphi_{18}$, the quantity of α waves on the occipital lobe is given by the numerical attribute $\varphi_1 = \alpha.\mathtt{occipital}$. The quantity of left (resp., right) eye movements is given by the numerical attribute $\varphi_2 = \mathtt{left\_eye}$ (resp., $\varphi_3 = \mathtt{right\_eye}$). The muscle tonus is given by the numerical attribute $\varphi_4 = \mathtt{tonus}$. For each attribute, the square product of its feature space is discretized using the scale $Sign_\epsilon$ (for a real value $\epsilon \geq 0$), such that:

$$\upsilon_{Sign_\epsilon}^{\varphi_i}(xy) = \begin{cases} + & \text{if } \varphi_i(y) > \varphi_i(x) + \epsilon/2 \\ - & \text{if } \varphi_i(y) < \varphi_i(x) - \epsilon/2 \\ \mathrm{o} & \text{if } \varphi_i(y) \in [\varphi_i(x) - \epsilon/2, \varphi_i(x) + \epsilon/2] \end{cases}$$

The expert knowledge is formalized by the two rules $A \curvearrowright B$ and $B \curvearrowright A$ with:

$$A : (\upsilon_{Sign_\epsilon}^{\alpha.\texttt{occipital}} : -) \vee (\upsilon_{Sign_\epsilon}^{\texttt{left\_eye}} : -) \vee (\upsilon_{Sign_\epsilon}^{\texttt{right\_eye}} : -) \vee (\upsilon_{Sign_\epsilon}^{\texttt{tonus}} : -)$$
$$B : (\upsilon_{Id}^{\texttt{s}} : (W, N_1))$$

The goal is to test two hypothesis (labelled $\mathbf{H}_0$ and $\mathbf{H}_1$ hereafter) on the set of pairs of epochs $E = \{(e_{n-2}, e_n) \mid e_{n-2}, e_n \in \mathcal{U}\} \subseteq \mathcal{U} \times \mathcal{U}$ that are distant of one minute, for every patient and every night of the dataset.

$\mathbf{H}_0$ The rule $A \curvearrowright B$ is verified
$\mathbf{H}_1$ The rule $B \curvearrowright A$ is verified

Analogical dissimilarity is defined for the rule $A \curvearrowright B$ by the scale $\texttt{ad}(ab, cd) = \mathbf{1}_A(ab, cd) = 1$ if $A(ab)$ and $A(cd)$, and 0 otherwise. Similarly, analogical dissimilarity is defined for the rule $B \curvearrowright A$ by the scale $\texttt{ad}(ab, cd) = \mathbf{1}_B(ab, cd) = 1$ if $B(ab)$ and $B(cd)$, and 0 otherwise. As a result, for two pairs $ab, cd \in A$, if we know that $ab \in B$, we estimate the probability $P(cd \in B)$ that the transfer rule

$$\frac{A \overset{ab}{\curvearrowright} B,\ cd \in A \text{ such that } \texttt{ad}(ab, cd) = 0}{A \overset{cd}{\curvearrowright} B}$$

is successfully applied to cd using the confidence measure $P(cd \in B) = \frac{|A \cap B|}{|A|}$. Similarly, we estimate the probability that the reverse rule $B \curvearrowright A$ is verified using the confidence measure $P(cd \in A) = \frac{|A \cap B|}{|B|}$. These two indicators are used to test the two hypothesis $\mathbf{H}_0$ and $\mathbf{H}_1$.

Table 3. Confidence indicators for the rules $A \curvearrowright B$ and $B \curvearrowright A$ according to the value given to the discretization parameter ϵ.

Rule	Indicator	Value for ϵ							
		0	0.1	0.5	1	5	10	50	100
$A \curvearrowright B$	$\frac{|A \cap B|}{|A|}$ (in %)	2.5	2.8	3.2	3.8	4.8	6.2	12	19.5
$B \curvearrowright A$	$\frac{|A \cap B|}{|B|}$ (in %)	97.4	94.6	89.5	84.5	62.3	45	7.8	1.5

Results. First results (Tab. 3) show that whatever the discretization, the presence in polysomnographies of property A is not discriminant to predict the sleep transition. However, the sleep transition from wakefulness (W) to light sleep stage $N1$ can be characterized by a signal having the property A. When signals are discretized with $\epsilon = 0$, a very large proportion (97%) of such sleep transitions are associated with a polysomnography in which the form described by the expert is verified.

7 Conclusion

There is an intimate link between the qualitative measurement of similarity and computational analogy. Some essential notions of formal models of analogy, such as analogical equalities/inequalities, or analogical dissimilarity, and the related inferences (mapping and transfer) can be formulated as operations on ordinal similarity relations. Qualitative similarity measurement thus provides a bridge between qualitative models and analogical methods. One direct outcome is that it should allow to inject more psychologically grounded similarity measures in learning methods.

References

1. Badra, F.: Reasoning with Co-variations. In: Artificial Intelligence: Methodology, Systems, and Applications - 17th International Conference, AIMSA. Varna, Bulgaria (2016). https://doi.org/10.1007/7854
2. Bounhas, M., Prade, H., Richard, G.: Analogy-based classifiers for nominal or numerical data. International Journal of Approximate Reasoning **91**, 36–55 (2017). https://doi.org/10.1016/j.ijar.2017.08.010
3. Craw, S., Wiratunga, N., Rowe, R.C.: Learning adaptation knowledge to improve case-based reasoning. Artificial Intelligence **170**(16-17), 1175–1192 (nov 2006). https://doi.org/10.1016/j.artint.2006.09.001
4. D'Aquin, M., Badra, F., Lafrogne, S., Lieber, J., Napoli, A., Szathmary, L.: Case base mining for adaptation knowledge acquisition. In: IJCAI International Joint Conference on Artificial Intelligence. pp. 750–755. Hyderabad, India (2007)
5. Davis, T.R., Russell, S.J.: A Logical Approach to Reasoning by Analogy. In: IJCAI International Joint Conference on Artificial Intelligence (1987). https://doi.org/10.1017/CBO9781107415324.004
6. De Amo, S., Diallo, M.S., Diop, C.T., Giacometti, A., Li, D., Soulet, A.: Contextual preference mining for user profile construction. Information Systems **49**(April), 182–199 (2015). https://doi.org/10.1016/j.is.2014.11.009
7. Derrac, J., Schockaert, S.: Inducing semantic relations from conceptual spaces: A data-driven approach to plausible reasoning. Artificial Intelligence **228**, 66–94 (2015). https://doi.org/10.1016/j.artint.2015.07.002
8. Fuchs, B., Lieber, J., Mille, A., Napoli, A.: Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR. Knowledge-Based Systems **68**, 103–114 (apr 2014). https://doi.org/10.1016/j.knosys.2014.03.009
9. Gust, H., Krumnack, U., Kühnberger, K., Schwering, A.: Analogical Reasoning: A Core of Cognition. KI - Künstliche Intelligenz **22**(1), 8–12 (2008)
10. Hammond, K.J.: CHEF: A Model of Case-based Planning. In: AAAI Proceedings. pp. 267–271 (1986)
11. Hanney, K., Keane, M.T.: The adaptation knowledge bottleneck: How to ease it by learning from cases. In: International Conference on Case-Based Reasoning. vol. 1266, pp. 359–370 (1997)
12. Hug, N., Prade, H., Richard, G., Serrurier, M.: Analogical classifiers: A theoretical perspective. In: 22nd European Conference on Artificial Intelligence – ECAI. vol. 285, pp. 689–697 (2016). https://doi.org/10.3233/978-1-61499-672-9-689

13. Hüllermeier, E.: Possibilistic instance-based learning. Artificial Intelligence **148**(1-2), 335–383 (2003). https://doi.org/10.1016/S0004-3702(03)00019-5
14. Jalali, V., Leake, D., Forouzandehmehr, N.: Learning and applying case adaptation rules for classification: An ensemble approach. IJCAI International Joint Conference on Artificial Intelligence pp. 4874–4878 (2017)
15. Lieber, J.: Application of the Revision Theory to Adaptation in Case-Based Reasoning: The Conservative Adaptation. In: ICCBR proceedings (2007)
16. Lieber, J., Napoli, A.: Correct and Complete Retrieval for Case-Based Problem-Solving. In: ECAI. pp. 68–72 (1998)
17. McSherry, D.: Demand-driven discovery of adaptation knowledge. IJCAI pp. 222–227 (1999)
18. Miclet, L., Bayoudh, S., Delhay, A.: Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning. Journal of Artificial Intelligence Research **32**, 793–824 (2008)
19. Mitchell, T.M.: Machine Learning. Elsevier (1983)
20. Novick, L.R., Holyoak, K.J.: Mathematical problem solving by analogy. Journal of experimental psychology. Learning, memory, and cognition **17**(3), 398–415 (1991). https://doi.org/10.1037/0278-7393.17.3.398, `http://www.ncbi.nlm.nih.gov/pubmed/1829473`
21. Novick, L.R.: Analogical Transfer, Problem Similarity, and Expertise. Journal of Experimental Psychology: Learning, Memory, and Cognition **14**(3), 510–529 (1988)
22. Ontañón, S., Plaza, E.: On Knowledge Transfer in Case-Based Inference. In: Díaz-Agudo, B., Watson, I. (eds.) ICCBR proceedings. pp. 312–326. Springer-Verlag Berlin Heidelberg (2012)
23. Paritosh, P.K., Klenk, M.E.: Cognitive processes in quantitative estimation: analogical anchors and causal adjustment. In: Proceedings of the 28th Annual Conference of the Cognitive Science Society (CogSci-06) (2006), `http://www.cs.northwestern.edu/{~}mek802/papers/knack-formatted.pdf`
24. Pirlot, M., Prade, H., Richard, G.: Completing Preferences by Means of Analogical Proportions. In: Modeling Decisions for Artificial Intelligence. vol. 4617, pp. 318–329 (2016). https://doi.org/10.1007/978-3-540-73729-2
25. Prade, H., Richard, G.: Reasoning with Logical Proportions. In: Principles of Knowledge Representation and Reasoning: Proceedings of the Twelfth International Conference, (KR). pp. 545–555. Toronto, Canada (2010)
26. Prade, H., Richard, G.: Analogical inequalities. In: Antonucci, A., , Cholvy, L., , Papini, O. (eds.) Symbolic and Quantitative Approaches to Reasoning with Uncertainty: 14th European Conference, ECSQARU'17, pp. 3–9. Springer International Publishing, Lugano, Switzerland (2017)
27. Robert, F.S.: Measurement Theory (1985)
28. Tversky, A.: Features of Similarity. In: Readings in Cognitive Science: A Perspective from Psychology and Artificial Intelligence, pp. 290–302. Elsevier Inc. (oct 2013)
29. Yao, Y.Y.: Qualitative Similarity. In: Suzuki, Y., Ovaska, S., Furuhashi, T., Roy, R., Dote, Y. (eds.) Soft Computing in Industrial Applications, pp. 339–348. Springer London (2000)
30. Žabkar, J., Bratko, I., Demšar, J.: Extracting qualitative relations from categorical data. Artificial Intelligence **239**, 54–69 (2016). https://doi.org/10.1016/j.artint.2016.06.007