A Dataset Complexity Measure for Analogical Transfer

Fadi Badra

HAL Id: hal-03029949
https://inria.hal.science/hal-03029949

Submitted on 29 Nov 2020

A Dataset Complexity Measure for Analogical Transfer

Fadi Badra

Université Sorbonne Paris Nord, Laboratoire d'Informatique Médicale et d'Ingénierie
des Connaissances en e-Santé – LIMICS, INSERM, UMR 1142, F-93000, Bobigny, France
badra@sorbonne-paris-nord.fr

Abstract

Analogical transfer consists in leveraging a measure of similarity between two situations to predict the amount of similarity between their outcomes. Acquiring a suitable similarity measure for analogical transfer may be difficult, especially when the data is sparse or when the domain knowledge is incomplete. To alleviate this problem, this paper presents a dataset complexity measure that can be used either to select an optimal similarity measure, or if the similarity measure is given, to perform analogical transfer: among the potential outcomes of a new situation, the most plausible is the one which minimizes the dataset complexity.

1 Introduction

Computational analogy is a subfield of computer science that aims at designing computational models of psychological and intellectual processes of analogical thinking [Gust *et al.*, 2008]. Contrary to machine learning approaches, no pre-trained model of the task at hand is required in computational analogy methods. Instead, the system solely relies on a structured memory and a proper similarity measure. Among the different tasks that computational analogy systems implement, the transfer task consists in using a measure of the similarity between two situations as a proxy to estimate the amount of similarity between their outcomes. Analogical transfer operates a special kind of plausible inference, according to which two situations that are judged similar are likely to have similar outcomes.

The choice of a similarity measure is critical for the success of a prediction task performed by analogical transfer. But designing a similarity measure for analogical transfer is not easy. Consider for example a simple regression task, the goal of which is to predict the price of an apartment given its characteristics. Suppose we have four apartments s_1, s_2, s_3, and s_4 of which the prices are known, and we want to predict the price of a new apartment t, knowing that it has two rooms, and is in downtown area (Tab. 1). Predicting the price of the new appartment t by analogical transfer consists in applying a *similarity principle*, according to which similar apartments have similar prices. The similarities between the situation t

Situation	nb_rooms	area	price
s_1	1	midtown	440
s_2	2	midtown	600
s_3	1	downtown	700
s_4	3	downtown	900
t	2	downtown	?

Table 1: Four apartments s_1, s_2, s_3, and s_4 for which the price is known, and a new apartment t which price is to be predicted.

and each of the situations s_i are thus taken as *proxies* to estimate the similarities between the price of t and the prices of each s_i. A price for t is predicted from these estimations. If the description of the apartment t were judged similar to the description of an apartment s_i, then the price of t should be similar to the price of s_i. But then, is t more similar to s_2 (which price is 600) than to s_4 (which price is 900)? As s_2, t has two rooms (nb_rooms = 2), but as s_4, the apartment t is located in downtown area (area = downtown). What is the importance of each criterion in the assessment of similarity? Which similarity metric should be used?

In computational analogy, the most common solution is to use a hand-crafted similarity measure, but it can be difficult and time-consuming [Jaiswal and Bach, 2019]. Another solution is to automate the construction of a similarity measure using machine learning [Mathisen *et al.*, 2019]. This strategy has some drawbacks. One of them is that it requires enough data to obtain a good predicting power. Similarity acquisition also depends on the algorithm used to perform the prediction, since the optimal similarity measure is chosen by comparing the predictive power of a particular prediction algorithm on the given task. Moreover, the obtained similarity measure is often a blackbox *i.e.*, no clue is given to a human observer on why two situations should be considered as similar [Gabel and Godehardt, 2015]. To be psychologically grounded, the learned similarity metric should verify some basic rationality principles [Coletti and Bouchon-Meunier, 2019], and metric learning methods learn distance functions from the data, whereas the mathematical notion of distance does not capture well psychological similarity measurements, since similarity judgements in practice often violate the basic distance axioms of minimality, symmetry, and triangular inequality [Tversky, 2013]. Many of these difficulties come
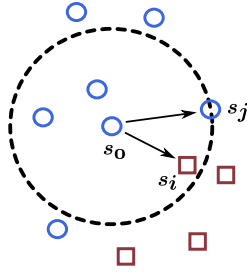
Figure 1: An inversion of similarity is observed in s_0: s_i is more similar to s_0 than s_j for the Euclidian distance (similarity on situations), but less similar for class membership (similarity on outcomes).

from the fact that similarity is a psychological notion that takes its root in human perception: similarity is the result of a judgement, it is highly dependent on the context and the knowledge and goals of the observer.

This paper explores another research direction, which consists in introducing some qualitative knowledge in the similarity acquisition process by making explicit some properties that the desired similarity measure should enforce, from the point of view of a rational agent. The most commonly used constraint in computational analogy, the similarity principle, is modeled as a qualitative continuity constraint. In qualitative terms, it says that a similarity measure on situations will be a good proxy for a similarity measure on outcomes if it orders the cases in the same manner: anytime a situation s_i is more similar to a situation s_0 than the situation s_j, this order should be preserved on outcomes. If it is not the case, we will say that there is an inversion of similarity. Consider for example the situation s_0 depicted in Fig. 1. The case base is divided into two sets of points, the red squares and the blue circles. The similarity on situations is measured by the (inverse of the) Euclidian distance, and the similarity on outcomes is measured by the class membership (two points of the same class being more similar than two points of different classes). In this example, there is an inversion of similarity in s_0 since the situation s_i is more similar to s_0 than the situation s_j for situations, but less similar for outcomes. The complexity of the dataset is measured by counting the number of inversions of similarity on the dataset.

Our contribution is two-fold: (i) we show that the proposed dataset complexity measure is an intrinsic indicator of the quality of the similarity measure that can be used to select an optimal similarity measure for a transfer task, and (ii) we show how the complexity measure can be used to actually perform the transfer: among the potential outcomes of a new situation, the most plausible is the one which minimizes the dataset complexity.

The paper is organised as follows. The next section introduces some notions and notations. Sec. 3 reviews some related work. Sec. 4 introduces the dataset complexity measure, which is the main contribution of the paper, along with three applications: selecting an optimal similarity measure, optimizing weights in a weighted sum, and performing analogical transfer. Sec. 5 presents some experiments to validate the approach, and Sec 6 concludes and presents future work.

2 Preliminaries

This section introduces some notions and notations that will be used throughout the paper.

Computational analogy. A typical design setting of a computational analogy system is as follows. Let $\mathcal{S}$ be a set of situations and $\mathcal{O}$ a set of outcomes. A function f associates to each situation s_i an outcome $f(s_i) \in \mathcal{O}$. A subset $CB = \{(s_1, f(s_1)), (s_2, f(s_2)), \ldots, (s_n, f(s_n))\}$ of $\mathcal{S} \times \mathcal{O}$ is called a case base. An element $(s_i, f(s_i)) \in CB$ is called a source case. An element of the square product $\mathcal{S} \times \mathcal{S}$ represents a pair of situations, and is denoted (s_i, s_j), or simply $s_i s_j$. Let us denote by σ_S a similarity measure on situations, and σ_R a similarity measure on outcomes. For a new situation t, three main tasks are commonly implemented by a computational analogy system [Gust *et al.*, 2008]:

- *Retrieval*: retrieve from CB a set of source cases $(s_i, f(s_i))$ to compare to the target situation t;
- *Mapping*: estimate the similarity $\sigma_S(t, s_i)$ between the target situation t and each of the retrieved situations s_i;
- *Transfer*: estimate the similarity $\sigma_R(f(t), f(s_i))$ on outcomes from the similarity $\sigma_S(t, s_i)$ on situations.

Analogical transfer can be used to (indirectly) predict the outcome $f(t)$ of a new situation t, since $f(t)$ can be predicted from the estimated similarity $\sigma_R(f(t), f(s_i))$. Most approaches of analogical transfer rely on a caracterization of the form of the function $\phi : \sigma_S \longrightarrow \sigma_R$ that relates the two functions σ_S and σ_R[1]. Such characterization is known as a *similarity principle*.

Qualitative modeling. Qualitative models [Forbus, 2011] capture how people reason about dynamical systems. 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 similarities. In this paper, we assume that the set of values of each similarity measure is totally ordered, which amounts to making a hypothesis of rationality regarding the way an agent makes a similarity judgement. A *scale* is a homomorphism, *i.e.*, a mapping that preserves all relations and operations. A *similarity scale* (called a variation in [Badra *et al.*, 2018]) is a scale on $\mathcal{S} \times \mathcal{S}$ which set of values is totally ordered. A similarity scale σ specifies a similarity relation $\succeq_\sigma$ on pairs of situations: for $s_i, s_j, s_k, s_\ell \in \mathcal{S}$, $\sigma(s_i s_j) \geq \sigma(s_k s_\ell) \Leftrightarrow s_i s_j \succeq_\sigma s_k s_\ell$. The similarity relation $\succeq_\sigma$ such that $ab \succeq_\sigma cd$ iff a and b are more similar than c and d for σ (or incomparable) is a total preorder (reflexive, transitive, and $ab \succeq_\sigma cd$ or $cd \succeq_\sigma ab$ for all $a, b, c, d \in \mathcal{S}$). The complementary relation $\prec_\sigma$ such that $ab \prec_\sigma cd$ iff a and b are strictly less similar than c and d for σ is then a strict partial order (transitive and asymmetric or areflexive) in which the relation of incomparability is transitive. Making the complementary relation $\prec_\sigma$

---

[1]As a misuse of language, we use the names of the similarity functions to denote their images.

be a strict partial order amounts to making a rationality hypothesis [Yao, 2000]. A human observer will judge that if two situations a and b are strictly less similar than two situations c and d, then c and d cannot be strictly less similar than a and b. Likewise, if a and b are not strictly less similar than c and d, and if c and d are not stricly less similar than e and f, then a and b can not be strictly less similar than e and f. Examples of similarity scales on a set of values include polynomial scales. Their form is as follows:

$$p_{n,L}(xy) = \frac{(L - |y - x|)^n}{L^n}$$

Polynomial scales associate to each pair of integers of the interval [0,L] a value in the interval [0,1]. Another useful similarity scale is the scale $=$ defined by:

$$= (xy) = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{if } x \neq y \end{cases}$$

Feature scales are similarity scales that order the pairs of values taken by a given feature. A feature is a variable on the set $\mathcal{S}$ that is defined for a particular dimension of interest (e.g., price, or number of rooms). It can be modeled as a function $\varphi : \mathcal{S} \longrightarrow \mathcal{V}$. As shown in [Badra *et al.*, 2018], assuming that there exists a total preorder on the set L of values taken by a feature $\varphi : \mathcal{S} \longrightarrow \mathcal{V}$, one can define, for any scale $\mathbf{o} : \mathcal{V} \times \mathcal{V} \longrightarrow L$ on the values of φ, the similarity scale

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

For example, the similarity scale $\sigma_{p_{n,L}}^{\texttt{nb\_rooms}}$ can be defined from the nominal feature $\varphi = \texttt{nb\_rooms}$ and the scale $p_{n,L}$ as follows: $\sigma_{p_{n,L}}^{\texttt{nb\_rooms}}(ab) = p_{n,L}(\texttt{nb\_rooms}(a), \texttt{nb\_rooms}(b))$. The scale $\sigma_{=}^{\texttt{area}}(ab)$ can be defined from the nominal feature $\varphi = \texttt{area}$ and the scale $=$ as follows:

$$\sigma_{=}^{\texttt{area}}(ab) = \begin{cases} 1 & \text{if } \texttt{area}(a) = \texttt{area}(b) \\ 0 & \text{if } \texttt{area}(a) \neq \texttt{area}(b) \end{cases}$$

3 Related Work

This section reviews some related work in the literature.

Analogical transfer. Since the pioneering work of [Davis and Russell, 1987], which introduces a logical formulation of analogical transfer, many formalizations were proposed. A first type of approach consists in searching where the two similarity measures σ_S and σ_R align locally, and reason by similarity on the found alignments [Miclet *et al.*, 2008; Ontañón and Plaza, 2012; Hug *et al.*, 2016; Badra *et al.*, 2018]. The main difficulty with this kind of approach is to establish which relations must be generalized. A method was proposed in [Badra, 2016] to learn generalizations (called co-variations), but finding a general rule is rare. For example, apartment prices usually increase with the number of rooms, but not always: it might not be the case for two apartments located in different areas. A strategy could be to compare only identical situations *ceteris paribus*, but then a very large training set would be required to find valid rules. Besides, learning generalization rules is sensitive to noise: the presence of outliers in the data may prevent from learning the rules. Another type of approach consists in expressing the similarity principle as a continuity constraint [Hullermeier, 2007], which is a negative constraint on the set of possible states according to which it is not plausible to observe situations very dissimilar for σ_R when they are similar for σ_S. This principle can be expressed in probability theory by stating that if two situations are above a similarity level α for σ_S, it is likely that their similarity for σ_R is greater or equal than a value β. The function $\beta : [0,1] \longrightarrow [0,1]$ which associates to each similarity level α for σ_S a similarity level β for σ_R is called a *similarity profile*. It is defined as $\beta(\alpha) = \inf_{s,s':\sigma_S(ss')=\alpha} \sigma_R(f(s)f(s'))$. Assuming that β is known, one can compute for a new situation s_0 a set $C(s_0) = \bigcap_s \{f(s) : \sigma_R(f(s_0), f(s)) \geq \beta(\sigma_S(s_0 s))\}$ of *credible* solutions by taking, for each retrieved source case s, the set of solutions $f(s)$ of the situations s in the neighborhood of s_0 that satisfy the similarity principle. The problem is then to learn β. In [Hullermeier, 2007], this function is approximated by a step function, which is learned from the data. In [Anthony and Ratsaby, 2015], each hypothesis for β is a multi-category classifier. Determining the levels β of a similarity profile from the data is a task that is very sensitive to outliers. Besides, enough data is needed to learn the profiles, and two global similarity measures σ_S and σ_R must be available. When the similarity principle is expressed as a continuity constraint, a complementary approach consists in measuring the evidential support of each potential outcome $f(s_0)$, *i.e.*, the extent to which the rule $\sigma_S \longrightarrow \sigma_R$ is supported by the data when s_0 is compared to other cases [Dubois *et al.*, 2000; Hüllermeier, 2003; Beringer and Hullermeier, 2008]. This rule is usually expressed as a possibility rule, and combined for all $s \in CB$ with a principle of maximum informativeness, leading to the possibility distribution $\delta(s_0, f(s_0)) = \max_{(s, f(s)) \in CB} \min\{\sigma_S(s_0 s), \sigma_R(f(s_0)f(s))\}$.

Dataset complexity. Dataset complexity measures were originally introduced to evaluate the intrinsic difficulty of a classification problem [Ho and Basu, 2000]. The goal of a binary classification problem is to accurately predict on which side of the class boundary a new instance lies, so the complexity of the class boundary has a direct impact on the difficulty of the classification problem: the more complex the actual boundary is, the more difficult it is to learn a representation of this boundary from the data. Complexity measures have extensively been used for case-based maintenance [Cummins, 2013] or to optimize retrieval (see for example [Smyth and Keane, 1995; Lamontagne, 2006; Lieber *et al.*, 2019]), but none of these measures are designed to find an optimal similarity measure or to perform the transfer task. Some measures evaluate the extent to which the similarity principle is verified locally. The Friend/Enemy ratio [Massie *et al.*, 2007] compares the average similarity of a case to the k nearest cases of the same class with its average similarity to the k nearest cases of a different class. It detects at most one similarity inversion for each case, and uses this information to identify potential noisy cases. In [Leake and Wilson, 1999], the quality of the retrieval phase is measured by evaluating if the similarity principle is enforced

in the neighborhood of each source case. In [Dileep and Chakraborti, 2014], fractals are used in the context of binary classification to detect which cases are the closest to the decision boundary, by noticing that those are the ones for which the similarity principle is less enforced.

4 A Dataset Complexity Measure

This section shows how to measure the complexity of a case base with respect to the transfer task.

Definition of the measure. Let σ_S be a similarity scale on situations, σ_R be a similarity scale on outcomes, and $CB \subseteq \mathcal{S} \times \mathcal{O}$ a case base. The values of the similarity measures σ_S and σ_R are assumed to be totally ordered. To measure the extent to which the similarity principle is enforced in a case base CB, we use a qualitative version of the continuity constraint that is commonly used in the literature. This constraint formalizes the idea that the function $\phi : \sigma_S \longrightarrow \sigma_R$ should preserve the orderings on similarity values: for three situations s_0, s_i and s_j, if s_i is judged more similar to s_0 than s_j for σ_S, then it should also be more similar to s_0 than s_j for σ_R. More formally, let $\succeq_{\sigma_S}$ (resp., $\succeq_{\sigma_R}$) be the similarity relation specified by σ_S (resp., σ_R) . The continuity constraint expresses that $\forall s_i, s_j \in \mathcal{S}, s_0 s_i \succeq_{\sigma_S} s_0 s_j \Rightarrow s_0 s_i \succeq_{\sigma_R} s_0 s_j$. Let us denote by $\mathtt{Inv}(s_0)$ the set of pairs $s_i s_j$ for which the constraint is *not* verified. The set $\mathtt{Inv}(s_0)$ is defined by:

$$\mathtt{Inv}(s_0) = \{s_i s_j \mid \sigma_S(s_0 s_i) \geq \sigma_S(s_0 s_j) \text{ and } \sigma_R(s_0 s_i) < \sigma_R(s_0 s_j)\}$$

If $s_i s_j \in \mathtt{Inv}(s_0)$, we will say that there is an *inversion of similarity* in s_0. In the example of Fig. 1, let $\mathcal{S} = \mathbb{R}^2$, $\sigma_S(uv) = (1 + \|\vec{uv}\|_2)^{-1}$ be the similarity scale that returns a value inversely monotonic to the norm of the vector $\vec{uv}$, and $\sigma_R = \sigma_{=}^{\mathtt{color}}$ be the scale constructed from the nominal attribute $\mathtt{color} : \mathcal{S} \longrightarrow \{\mathtt{blue}, \mathtt{red}\}$. The situations s_0, s_i and s_j verify $\sigma_S(s_0 s_i) \geq \sigma_S(s_0 s_j)$ and $\sigma_R(s_0 s_i) < \sigma_R(s_0 s_j)$, showing an inversion of similarity in s_0. Let $\gamma(s_0)$ denote the complexity of the source case $(s_0, f(s_0))$. It measures the extent to which the similarity principle is locally enforced in s_0, and is defined as the number of inversions of similarity that can be observed in s_0:

$$\gamma(s_0) = |\mathtt{Inv}(s_0)|$$

The complexity $\Gamma(\sigma_S, \sigma_R, CB)$ of the case base CB is defined as the sum of the complexities of all source cases:

$$\Gamma(\sigma_S, \sigma_R, CB) = \sum_{(s_0, f(s_0)) \in CB} \gamma(s_0)$$

Selecting an optimal similarity measure. From the point of view of a rational agent, the optimal similarity measure σ_S is the one that minimizes the case base complexity Γ. The key idea here is that similarity acquisition does not depend on the algorithm used to perform the transfer. Instead, it relies on an indicator (the complexity measure Γ) that quantifies the extent to which the similarity principle would be verified for each source case of the case base. In usual similarity evaluation methods, the optimal similarity measure is chosen by comparing the predictive power of a particular prediction algorithm for different similarity measures. One potential limitation however, is that the similarity principle may not always
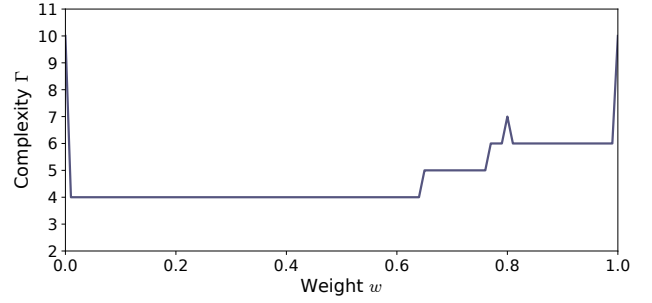

Figure 2: Case base complexity Γ according to weight w.

be enforced in a real dataset, and when it is not the case, such similarity acquisition method may lead to suboptimal solutions. As an example, suppose we want to determine an optimal global similarity measure σ_S when the case base CB is the set of apartment descriptions given in the introduction (Tab.1). We choose $\sigma_R = \sigma_{p_{2,1000}}^{\mathtt{price}}$ and the similarity measure σ_S is assumed to be a weighted sum of the similarity $\sigma_{p_{2,6}}^{\mathtt{nb\_rooms}}$ according to the number of rooms and the similarity $\sigma_{=}^{\mathtt{area}}$ according to the location area:

$$\sigma_S(uv) = w \times \sigma_{p_{2,6}}^{\mathtt{nb\_rooms}}(uv) + (1 - w) \times \sigma_{=}^{\mathtt{area}}(uv)$$

The goal is to determine the weight $w \in [0, 1]$ to set in order to obtain for σ_S an optimal similarity measure. Fig. 2 shows the case base complexity Γ for different values of w. The optimal similarity measure σ_S is obtained for $w \in [0.01, 0.64]$.

Optimizing weights in a weighted sum. Complexity can also be used for feature scoring when constructing a similarity measure. A popular similarity scale σ_S is the weighted sum of n similarity scales $\sigma_1, \sigma_2, \ldots, \sigma_n$:

$$\sigma_S(s_0 s_i) = \sum_k w_k \times \sigma_k(s_0 s_i)$$

To determine optimal values for the weights w_k, we apply the method of [Jaiswal and Bach, 2019], which consists in ranking the similarity scales σ_k according to an indicator, and then setting each w_k to be the rank of the corresponding scale σ_k in the ordered set of indicator values. The chosen indicator is the complexity measure Γ. The similarity scales σ_k are ranked in decreasing order of complexity, and the weights w_k are set accordingly. In the example of apartments, the method gives equal weights, which amounts to setting $w = 0.5$.

Algorithm 1 Optimize weights in a weighted sum

Inputs : $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_n\}, \sigma_R, CB$
Output : a list of weights $[w_1, w_2, \ldots, w_n]$
$\ell \leftarrow [\,]$
for $\sigma_k \in \Sigma$ **do**
    $\ell.\mathtt{append}(\Gamma(\sigma_k, \sigma_R, CB))$
end for
$\ell \leftarrow \ell.\mathtt{sort\_in\_decreasing\_order}()$
return $[1 + \ell.\mathtt{index}(\mathtt{e})$ for $\mathtt{e}$ in $\ell]$

Algorithm 2 Complexity-based analogical transfer

Inputs : $t, \mathcal{O}, \sigma_S, \sigma_R, CB$
Output : a predicted outcome for the new situation t
$\ell \leftarrow [\,]$
for $o \in \mathcal{O}$ **do**
   `change_outcome`(t, o)
   $\ell.\text{append}(\Gamma(\sigma_S, \sigma_R, CB \cup \{t\}))$
end for
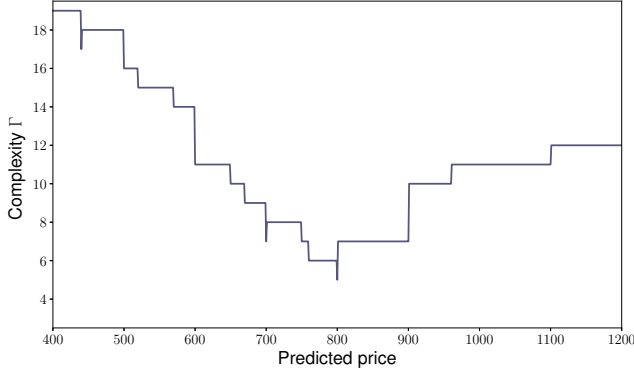**return** $\mathcal{O}[\ell.\text{index}(\min(\ell))]$

Figure 3: Complexity for each potential apartment price.

Complexity-based analogical transfer. The complexity measure can be used to perform analogical transfer: the most plausible outcome of a new situation t is the one which leads to the lowest case base complexity Γ. This approach to transfer will be referred to as **Co**mplexity-based **A**nalogical **T**ransfer (CoAT). Suppose that we want to predict the price of the apartment t presented in the introduction (Tab.1). The similarity measure σ_S is a weighted sum of the two feature scales $\sigma_{p_{2,6}}^{\text{nb\_rooms}}$ and $\sigma_{=}^{\text{area}}$, and weights are determined using the method presented in the previous paragraph. Fig. 3 shows the case base complexity computed for different values of the potential price of apartment t. Results show that the optimal price for t is 800 ($\Gamma = 5$), which is consistent since 800 is the value given by a simple regression on this dataset. But we also obtain an estimation of the plausibility of other prices in the interval $[400, 1200]$.

5 Experiments

Three experiments were run. The first one tests the hypothesis that the complexity measure Γ is an indicator of the quality of the similarity measure σ_S. The second one evaluates the performance of CoAT on a regression task, and the third one evaluates the performance of CoAT on classification tasks.

Experiment #1. The PIMA indian diabetes dataset[2] includes data about 768 native American women. The prediction task consists in predicting if a person suffers from diabetes (taken as a binary class) from the value of 8 continuous attributes. The similarity scale σ_{man} (taken from [Jaiswal

---
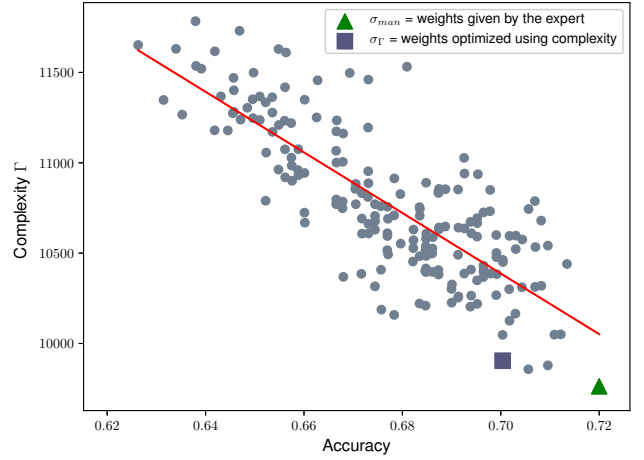[2]https://kaggle.com/uciml/pima-indians-diabetes-database

Figure 4: Relation between complexity and performance.

and Bach, 2019]) is the weighted sum of a set of polynomial similarity scales, with feature weights given by a domain expert, computed by including all features. From this similarity scale, a set of similarity scales was generated by assigning for each weight of σ_{man} a random value between 0 and 8, and then renormalizing the weights so that they sum up to 1. The similarity scale σ_{Γ} was constructed from the same set of polynomial similarity scales using the feature scoring method proposed in Sec. 4. The quality of each similarity scale is estimated by the accuracy of the k-Nearest Neighbor (k-NN) algorithm, with $k = 5$, computed using 10-fold cross validation[3]. Fig. 4 shows the relation between complexity and accuracy. Each grey point represents a similarity scale. The green triangle represents the σ_{man} similarity scale given by the expert. The blue square represents the σ_{Γ} similarity scale. Results show a strong correlation between the dataset complexity Γ and the quality (performance) of the similarity measure. Pearson's correlation coefficient is -0.8 for the 188 randomly generated similarity scales. The similarity measure σ_{man} given by the expert shows both the highest performance (72% accuracy) and the lowest complexity (9763). The σ_{Γ} similarity scale is nearly optimal: its complexity is among the lowest (9905), and its performance nearly as high as σ_{man} (70%). These results validate the hypothesis that for a given algorithm, complexity is a good indicator of the quality of the similarity measure σ_S.

Experiment #2. The Automobile dataset[4] includes data about 205 automobiles, from which were kept only the 159 instances that contain no missing values. The task consists in predicting the price of a car. Following [Dubois *et al.*, 2000], in each source case $(s_0, f(s_0))$ the source situation s_0 describes a car with the 2 continuous attribues `horsepower` and `engine_size`, and the associated outcome $f(s_0)$ is given by the `price` attribute. Car similarity σ_S is computed using

---
[3]Due to current limitations in our implementation, the training sets of each pass were restricted to 50 randomly chosen instances, and the test sets to 77 instances.
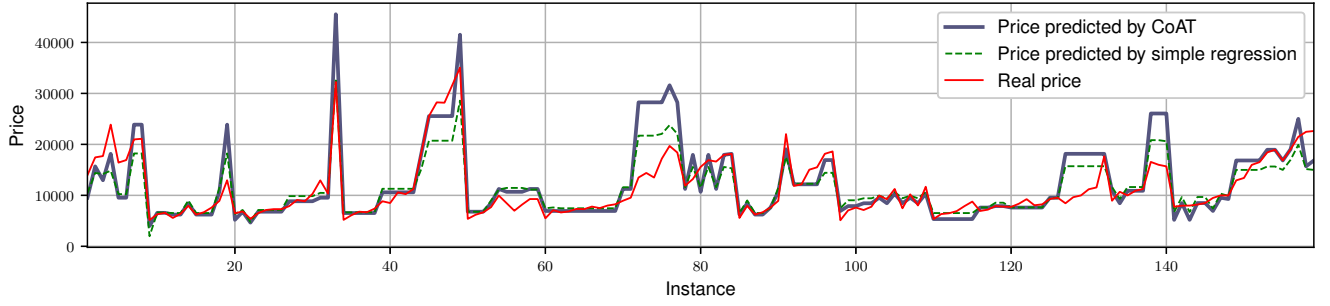
[4]https://archive.ics.uci.edu/ml/datasets/Automobile

Figure 5: Price predicted by taking the value that results in the lowest dataset complexity, *vs* real price on the autos dataset.

a weighted sum of $\Sigma = \{\sigma_{P2,240}^{\text{horsepower}}, \sigma_{P2,265}^{\text{engine\_size}}\}$, with weights optimized as in Sec. 4, and $\sigma_R = \sigma_{P2,40282}^{\text{price}}$. Fig. 5 compares for each instance the price predicted by `CoAT` (thick blue line) to the price obtained by a simple regression (dashed green line), as well as the real price (red line). Although our first implementation does not strictly surpass the performance of linear regression in this example (the root mean squared error is 4035 for `CoAT`, and 3166 for simple regression), the results demonstrate the pertinence of using complexity minimization for analogical transfer. In this example, no model of the data was learned prior to achieving the prediction task, and similarity learning was reduced to the acquisition of a few similarity scales. Each outcome is recovered from all others from the similarity measures alone, by taking the value that results in the lowest dataset complexity.

Experiment #3. The performance of `CoAT` was compared to well-known machine learning algorithms on 6 classical datasets of the UCI repository[5] (Tab. 2): the Monks datasets (`monks1`, `monks2`, and `monks3`), the User Modeling dataset (`user`), the Iris dataset (`iris`), and the Zoo dataset (`zoo`). The similarity measure σ_S is a weighted sum of n feature scales. The feature scale $\sigma_{\doteq}^{\varphi}$ was used for each binary feature φ, and a polynomial scale was used for each continuous feature. The weights are set by the method proposed in Sec. 4. Results show excellent prediction performance (Tab. 3): `CoAT` equals or outperforms state of the art algorithms on the `monks2`, the `iris`, and the `zoo` datasets. On `monks1`, both classes often resulted in the same complexity value, and could not be discriminated, hence the poor results.

Dataset	# instances	# features	# classes
monks1	432	6	2
monks2	432	6	2
monks3	432	6	2
user	258	5	4
iris	150	4	3
zoo	101	18	7

Table 2: The datasets used for the classification tasks.

Dataset	SVM	IB1	JRip	C4.5	CoAT
monks1	75%	**99%**	75%	93%	69%
monks2	**67%**	60%	66%	**67%**	**67%**
monks3	**100%**	99%	99%	**100%**	97%
user	75%	81%	89%	**90%**	84%
iris	**97%**	95%	95%	94%	**97%**
zoo	91%	96%	88%	92%	**99%**

Table 3: Accuracy results of `CoAT` on classification tasks.

6 Conclusion and Future Work

This paper opens a new research direction for analogical transfer, which consists in making explicit a set of qualitative constraints that a rational agent would want enforced. Two constraints were applied: a rationality hypothesis (similarity values are totally ordered) and a continuity constraint (anytime a situation s_i is more similar to a situation s_0 than a situation s_j, this order should be preserved on outcomes). A dataset complexity measure was proposed to measure the extent to which the continuity constraint is enforced on the whole dataset. This constraint expresses the well-known similarity principle. We demonstrated the usefulness of such measurement, either (i) to optimize the choice of the similarity measure, by choosing a measure that best enforces the constraint, or (ii) to find the most plausible transfer, by choosing the potential outcome which best preserves the constraint on the augmented dataset. First results are very encouraging. Complexity appears to be a good indicator of the quality of a similarity measure (similarity measures of lower complexity are in average of better quality), and we showed that complexity minimization can be used to perform analogical transfer on regression and classification tasks.

Future work includes providing more theoretical justifications of the approach, and improving the implementation in order to make it scale to datasets that are larger than a few hundred instances. Complexity measurement is currently in $\mathcal{O}(n^3)$, but huge improvements could be expected since computation could be run in parallel for each instance of the case base, thus reducing it to $\mathcal{O}(n^2)$. More qualitative constraints could also be added to refine the method, such as attribute independence or monotonicity.

References

[Anthony and Ratsaby, 2015] Martin Anthony and Joel Ratsaby. A probabilistic approach to case-based inference. *Theor. Comput. Sci.*, 589:61–75, 2015.

[Badra *et al.*, 2018] Fadi Badra, Karima Sedki, and Adrien Ugon. On the Role of Similarity in Analogical Transfer. In *ICCBR, Int. Conf. Case-Based Reason.*, 2018.

[Badra, 2016] Fadi Badra. Reasoning with Co-variations. In *Artif. Intell. Methodol. Syst. Appl. - 17th Int. Conf. AIMSA*, Varna, Bulgaria, 2016.

[Beringer and Hullermeier, 2008] Jürgen Beringer and Eyke Hullermeier. Case-Based Learning in a Bipolar Possibilistic Framework. *Int. J. Intell. Syst.*, 23:1119–1134, 2008.

[Coletti and Bouchon-Meunier, 2019] Giulianella Coletti and Bernadette Bouchon-Meunier. A study of similarity measures through the paradigm of measurement theory: the classic case. *Soft Comput.*, pages 1–19, 2019.

[Cummins, 2013] Lisa Cummins. *Combining and Choosing Case Base Maintenance Algorithms*. PhD thesis, University College Cork, 2013.

[Davis and Russell, 1987] Todd R. Davis and Stuart J. Russell. A Logical Approach to Reasoning by Analogy. In *IJCAI Int. Jt. Conf. Artif. Intell.*, 1987.

[Dileep and Chakraborti, 2014] K V S Dileep and Sutanu Chakraborti. Estimating Case Base Complexity Using Fractal Dimension. In Luc Lamontagne and Enric Plaza, editors, *Case-Based Reason. Res. Dev.*, pages 230–244, Cham, 2014. Springer International Publishing.

[Dubois *et al.*, 2000] Didier Dubois, Eyke Hüllermeier, and Henri Prade. Flexible Control of Case-Based Prediction in the Framework of Possibility Theory. In *EWCBR*, pages 61–73. Springer-Verlag, 2000.

[Forbus, 2011] Kenneth D Forbus. Qualitative modeling. *Wiley Interdiscip. Rev. Cogn. Sci.*, 2(4):374–391, 2011.

[Gabel and Godehardt, 2015] Thomas Gabel and Eicke Godehardt. Top-Down Induction of Similarity Measures Using Similarity Clouds. In *23rd Int. Conf. iccbr 2015 frankfurt am main, Ger. Sept. 28–30, 2015 Proc.*, volume 9343, pages 1–16, 2015.

[Gust *et al.*, 2008] Helmar Gust, Ulf Krumnack, KU Kühnberger, and Angela Schwering. Analogical Reasoning: A Core of Cognition. *KI - Künstliche Intelligenz*, 22(1):8–12, 2008.

[Ho and Basu, 2000] Tin Kam Ho and Mitra Basu. Measuring the Complexity of Classification Problems. In *15th Int. Conf. Pattern Recognit.*, pages 43–47, 2000.

[Hug *et al.*, 2016] Nicolas Hug, Henri Prade, Gilles Richard, and Mathieu Serrurier. Analogical classifiers: A theoretical perspective. In *22nd Eur. Conf. Artif. Intell. – ECAI*, volume 285, pages 689–697, 2016.

[Hüllermeier, 2003] Eyke Hüllermeier. Possibilistic instance-based learning. *Artif. Intell.*, 148(1-2):335–383, 2003.

[Hullermeier, 2007] Eyke Hullermeier. *Case-Based Approximate Reasoning*. Theory and Decision Library. Springer, 2007.

[Jaiswal and Bach, 2019] Amar Jaiswal and Kerstin Bach. A Data-Driven Approach for Determining Weights in Global Similarity Functions. In Kerstin Bach and Cinty Marling, editors, *Case-Based Reason. Res. Dev. ICCBR 2019. Lect. Notes Comput. Sci.*, volume 11680 LNAI, pages 125–139, 2019.

[Lamontagne, 2006] Luc Lamontagne. Textual CBR authoring using case cohesion. *3rd TCBR*, pages 33–43, 2006.

[Leake and Wilson, 1999] David B. Leake and David C. Wilson. When Experience is Wrong : Examining CBR for Changing Tasks and Environments. In *Int. Conf. Case-Based Reason. ICCBR*, volume 1650, pages 218–232, 1999.

[Lieber *et al.*, 2019] Jean Lieber, Emmanuel Nauer, and Prade. Improving Analogical Extrapolation Using Case Pair Competence. In *ICCBR, Int. Conf. Case-Based Reason.*, 2019.

[Massie *et al.*, 2007] Stewart Massie, Susan Craw, and Nirmalie Wiratunga. When similar problems don't have similar solutions. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, 4626 LNAI(May 2014):92–106, 2007.

[Mathisen *et al.*, 2019] Bjørn Magnus Mathisen, Agnar Aamodt, Kerstin Bach, and Helge Langseth. Learning similarity measures from data. *Prog. Artif. Intell.*, 2019.

[Miclet *et al.*, 2008] Laurent Miclet, Sabri Bayoudh, and Arnaud Delhay. Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning. *J. Artif. Intell. Res.*, 32:793–824, 2008.

[Ontañón and Plaza, 2012] Santiago Ontañón and Enric Plaza. On Knowledge Transfer in Case-Based Inference. In B. Díaz-Agudo and I. Watson, editors, *ICCBR Proc.*, pages 312–326. Springer-Verlag Berlin Heidelberg, 2012.

[Smyth and Keane, 1995] Barry Smyth and Mark T. Keane. Remembering To Forget. In *IJCAI Int. Jt. Conf. Artif. Intell.*, 1995.

[Tversky, 2013] Amos Tversky. Features of Similarity. In *Readings Cogn. Sci. A Perspect. from Psychol. Artif. Intell.*, pages 290–302. Elsevier Inc., oct 2013.

[Yao, 2000] Y. Y. Yao. Qualitative Similarity. In Yukinori Suzuki, Seppo Ovaska, Takeshi Furuhashi, Rajkumar Roy, and Yasuhiko Dote, editors, *Soft Comput. Ind. Appl.*, pages 339–348. Springer London, London, 2000.