Elements of a Data-Driven Approach to Adaptation

Fadi Badra
Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé

Elements of a Data-Driven Approach to Adaptation

Fadi Badra

Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430
Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006
INSERM, U1142, LIMICS, F-75006, Paris, France
badra@univ-paris13.fr

Abstract. Adaptation is what enables a system to respond to perturbations in its environment. Even though case-based reasoning is commonly thought of as a data-driven problem-solving paradigm, adaptation methods often follow a model-driven pattern. This paper discusses what a purely data-driven approach to adaptation might look like. The proposed method uses co-variations to extrapolate to the target case some parts of the relational structure underlying the case base. One of the outcomes of this work is to show that adaptation can be (at least partially) included in the case-based inference process, which is known to be a special case of analogical reasoning.

Keywords: adaptation · extrapolation · data-driven · variations

1 Introduction

To make predictions over unobserved features, inference mechanisms usually require having an adequate abstraction of the domain. Such an abstraction is either built directly from a human expert (deduction/abduction), or extracted from data (induction). Two strategies can be distinguished to overcome the presence of a missing or inadequate abstraction. The *model-driven* strategy consists in using the faulty abstraction anyway: this is the basis underlying most non-monotonic [27] and approximate reasoning [29] methods. Another strategy, more *data-driven* [14], is to get rid of the abstraction and reason directly from data. Analogical classification [19] and commonsense reasoning [8] methods such as similarity-based reasoning, interpolation or extrapolation fall into this category.

Adaptation is what enables a system to respond to perturbations in its environment. Even though case-based reasoning (CBR) is commonly thought of as a data-driven problem-solving paradigm, where data takes the form of a collection of prior experiences (the cases), many adaptation strategies such as critique-based adaptation [11] or conservative adaptation [16] proposed in the CBR community follow a model-driven pattern. They simulate the corresponding "rule-based" inference process, perform an adaptation by copy, and then optionally repair solution inconsistencies.

This paper discusses what a purely data-driven approach to adaptation might look like. The proposed method uses co-variations to extrapolate to the target case some parts of the relational structure underlying the case base. The working hypothesis is that past experiences are more than just a collection of cases but are somehow related to each other and that it is this relational structure that drives the adaptation process. One of the outcomes of this work is to show that adaptation can be (at least partially) included in the case-based inference process, which is known to be a special case of analogical reasoning. This contradicts the view of [21], in which the authors argue that the main difference between CBR and analogy is that adaptation is not theoretically explained by computational methods of analogy.

The paper is organized as follows. Sec. 2 presents the main principles underlying the envisioned data-driven adaptation method. In Sec. 3, a modeling of similarity and of the case-based inference process is proposed. Co-variations are introduced in Sec. 4, and Sec. 5 shows that co-variations can be used to extrapolate to the target some parts of the relational structure underlying the case base. Sec. 6 gives some related work and Sec. 7 concludes the article and gives future work.

2 Idea of the Method

This section presents the main principles underlying the envisioned data-driven adaptation method.

> **Principle #1**: Adaptation is a transfer, guided by the target, of some relational structure present in the case base.

Adaptation is a relational concept since it consists in a change (in the system's knowledge) as a response to a change (in the system's environment). For the system to go beyond its initial problem-solving capabilities and propose original solutions to unseen target problems, the adaptation strategy consists in harnessing the relational structure of the case base.

> **Principle #2**: Similarity is a preorder on the set of pairs of cases.

The usual numerical distance between cases is generalized to a preorder $\preceq$ on the set of pairs of cases. Dropping numerical values in similarity assessment is justified by the fact that actual values of distance between two cases are less significant than the ability to compare two distances. This comparison is done here directly through the similarity preorder $\preceq$ and if for two pairs (x, y) and (z, t) the relation $(x, y) \preceq (z, t)$ holds, we would say that x is more similar to y than z is to t.

Principle #3: Regularity should drive the adaptation inference process.

Regularity is the idea that properties are implicitly preserved in the neighborhood of an object. Applied to knowledge representation, it can be seen as a hypothesis that similar objects have similar properties. This notion, that is central in commonsense reasoning methods [8], should also drive data-driven adaptation methods.

Principle #4: The quality of the solution should be maximized.

Among possible adaptations, the preferred one(s) are the ones that optimize a quality measure on solutions. This quality measure measures the consistency of the solution with respect to the underlying process. For example, in the medical domain, a health care trajectory adaptation should keep the health care quality maximal. In the cooking domain, an adapted recipe should be as tasty as possible for the end-user.

3 Modelling the Case-Based Inference Process

In this section, the case-based inference (CBI) process is modeled by a self map on the set of pairs of cases.

Agent's memory and knowledge. Let $\mathcal{U}$ be a set (called the *case universe*). An element of $\mathcal{U}$ is called a *case* and represents a possible experience. Among the cases of $\mathcal{U}$, some actually happened, were witnessed by the agent and retained in memory. These cases, the *case base*, constitute the memory of the agent. An element `srce` of the case base is called a *source case*. As in [21], a case is a single description, and the problem and the solution are two parts of this single description. The CBI aims at making more precise the description of a target case `tgt` for which a partial description $\mathcal{E}_{\texttt{tgt}}$ is already known (Fig. 1).

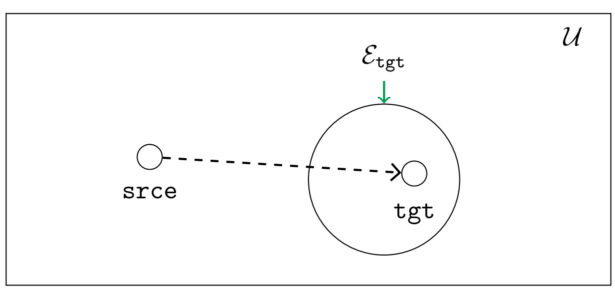

Fig. 1: A schematic view of the agent's memory and knowledge, with (U) the case universe, (`srce`) a source case, ($\mathcal{E}_{\texttt{tgt}}$) a partial description of the desired target case, and (`tgt`) the constructed target case.

Similarity. Similarity is modeled by a preorder $(\mathcal{U} \times \mathcal{U}, \preceq)$ on the set of pairs of possible cases. A preorder is a binary relation that is reflexive ($\forall e \in \mathcal{U} \times \mathcal{U}, e \preceq e$) and transitive ($\forall e, f, g \in \mathcal{U} \times \mathcal{U}$, if $e \preceq f$ and $f \preceq g$ then $e \preceq g$). For two pairs of cases (x, y) and (z, t) of $\mathcal{U} \times \mathcal{U}$, $(x, y) \preceq (z, t)$ means that x is more similar to y than z is to t.

Case-Based Inference. The CBI is a self map $\phi : \mathcal{U} \times \mathcal{U} \longrightarrow \mathcal{U} \times \mathcal{U}$. This function associates to each pair of possible cases of $\mathcal{U}$ (representing a possible change in the system's knowledge) to another pair of possible cases of $\mathcal{U}$ (representing a change in the system's environment). It should verify the following properties, for all $e, f \in \mathcal{U} \times \mathcal{U}$, which makes it a kernel operator (or projection) on $(\mathcal{U} \times \mathcal{U}, \preceq)$:

1. (*Contraction*) $\phi(e) \preceq e$
2. (*Monotonicity*) if $e \preceq f$ then $\phi(e) \preceq \phi(f)$
3. (*Idempotency*) $\phi(\phi(e)) = \phi(e)$

The (*Contraction*) property states that the proposed solution should be consistent with the constraints or observations made on tgt (that is, $\mathtt{tgt} \subseteq \mathcal{E}_{\mathtt{tgt}}$). The (*Monotonicity*) property expresses the common hypothesis in CBR that "similar problems have similar solutions". It states that the more similar two problems are, the more similar their two solutions should be. The (*Idempotency*) property states that re-adapting an adapted case does not lead to any improvement.

Note that here, the CBI function does not produce a representation of the target case tgt. It is rather used to determine a region of the relational structure $(\mathcal{U} \times \mathcal{U}, \preceq)$ where the pair (srce, tgt) is to be found.

4 Formalization using Variations

Variations provide relational structures that can be transported to the set of binary relations between cases and then provide local approximations of the adaptation function.

4.1 Variations

A *variation* between cases [3] is a function $\upsilon : \mathcal{U} \times \mathcal{U} \longrightarrow \mathcal{V}$ which associates a value (or, more generally, a description) taken in a set $\mathcal{V}$ to some pairs of possible cases of the case universe. An example of variation is the function $\mathtt{age}^<$, which returns true for a pair of cases (x, y) if the value of the numerical attribute age is strictly smaller for x than for y:

$$\mathtt{age}^<((x, y)) = \begin{cases} \mathtt{true} & \text{if } \mathtt{age}(x) < \mathtt{age}(y) \\ \mathtt{false} & \text{otherwise} \end{cases}$$

When cases are represented by sets of binary attributes, let us say, through a mapping $\varphi : \mathcal{U} \longrightarrow 2^{\mathtt{M}}$, with $\mathtt{M} = \{\mathtt{a}, \mathtt{b}, \mathtt{c}, \mathtt{d}, \mathtt{e}\}$, the variation

$$\upsilon_{ap}(x, y) = \{\, \varphi(x) \cap \overline{\varphi(y)} \,,\, \overline{\varphi(x)} \cap \varphi(y) \,\}$$

associates to a pair of cases (x, y) the sets of attributes that are lost $(\varphi(x) \cap \overline{\varphi(y)})$ and the sets of attributes that are gained $(\overline{\varphi(x)} \cap \varphi(y))$ when going from x to y. For example, if $\varphi(x) = \{\mathtt{a}, \mathtt{b}\}$ and $\varphi(y) = \{\mathtt{b}, \mathtt{d}\}$, then $\upsilon_{ap}(x, y) = \{\{\mathtt{a}\}, \{\mathtt{d}\}\}$ since a is lost and d is gained when going from x to y.

It can be noted that the set of pairs that share the same value for a variation form a binary relation on $\mathcal{U}$. For example, the set $\{(x, y) \mid \mathtt{age}^{<}((x, y)) = \mathtt{true}\}$ is a binary relation which contain the pairs of cases for which there is an age increase. Similarly, it follows from the definition of analogical proportions [22] that $x : y :: z : t$ holds for any two pairs (x,y) and (z,t) if and only if they take the same value for υ_{ap}.

4.2 Similarity

Any preorder on the values $\mathcal{V}$ of a variation can be interpreted as a similarity relation on $\mathcal{U} \times \mathcal{U}$. For example, let the variation $\mathtt{age}^{-}(x, y) = |\mathtt{age}(x) - \mathtt{age}(y)|$ be the function that, for two cases, returns a natural number representing the (absolute) difference in value of the property age. Then, the preorder $(\mathbb{N}, \leq)$ on natural numbers induces a preorder on the set $\mathcal{U} \times \mathcal{U}$ of pairs of cases. Taking this preorder as the similarity relation (*i.e.*, defining $(x, y) \preceq (z, t)$ iff $\mathtt{age}^{-}(x, y) \leq \mathtt{age}^{-}(z, t)$) amounts to supposing that the lower the age difference, the more similar two cases are. For the variation υ_{ap}, the inclusion relation $\subseteq$ can be used as a preorder on $2^{\mathtt{M}} \times 2^{\mathtt{M}}$ ($\{A, B\} \subseteq \{A', B'\}$ iff $A \subseteq A'$ and $B \subseteq B'$). Taking this preorder as the similarity relation (*i.e.*, defining $(x, y) \preceq (z, t)$ iff $\upsilon_{ap}(x, y) < \upsilon_{ap}(z, t)$) amounts to supposing that the less properties are lost or gained when going from a case x to a case y, the more similar x and y are. For example, if $\varphi(x) = \{\mathtt{a}, \mathtt{b}\}$, $\varphi(y) = \{\mathtt{b}, \mathtt{d}\}$, $\varphi(z) = \{\mathtt{a}, \mathtt{c}\}$, and $\varphi(t) = \{\mathtt{c}\}$, then $(z, t) \preceq (x, y)$ since $\upsilon_{ap}(z, t) = \{\{\mathtt{a}\}, \{\}\} \subseteq \upsilon_{ap}(x, y) = \{\{\mathtt{a}\}, \{\mathtt{d}\}\}$.

4.3 Co-variations

Co-variations are defined [4] as functional dependencies between variations.

Any general-to-specific ordering $\leq_g$ on variations generates a co-variation. Such co-variation plays the role of some "inclusion axioms", but for a relational setting. A general-to-specific ordering on variations can be obtained whenever a Boolean-value function h_υ can be associated to each variation υ. We would say that a variation υ' is more general than a variation υ, denoted by $h_\upsilon \leq_g h_{\upsilon'}$, iff $h_{\upsilon'}(x)$ is true whenever $h_\upsilon(x)$ is true. For Boolean-value variations, the variation υ itself can be chosen as the function h_υ, by setting *i.e.*, $h_\upsilon = \upsilon$. For example, if the variation $\mathtt{age}^{\neq}$ is defined by:

$$\mathtt{age}^{\neq}((x, y)) = \begin{cases} \mathtt{true} & \text{if } \mathtt{age}(x) \neq \mathtt{age}(y) \\ \mathtt{false} & \text{otherwise} \end{cases}$$

then the ordering $\mathtt{age}^{<} \leq_g \mathtt{age}^{\neq}$ is interpreted as the co-variation $\mathtt{age}^{<} \rightsquigarrow \mathtt{age}^{\neq}$.

Co-variations can also be learnt from the data. We showed in [4] that the co-variation learning task can be reduced to association rule learning by choosing a suitable pattern structure.

5 Adaptation as Extrapolation

Co-variations constitute local approximations of the CBI function. Therefore, adaptation can be performed by transferring co-variations from the case base to a target case, in a form of extrapolation.

5.1 Example

Suppose that the goal is to predict the price of an apartment given its characteristics. The case base consists in three apartments $\texttt{srce}_1$, $\texttt{srce}_2$, and $\texttt{srce}_3$, of which the prices are known, and we want to predict the price of a `tgt` apartment, knowing that it has two rooms, it is on 1^{st} floor, and in downtown area (Tab. 1).

	Rooms (nb_rooms)	Floor (floor)	Area (area)	Price (price)
$\texttt{srce}_1$	3	0	downtown	440
$\texttt{srce}_2$	1	15	midtown	290
$\texttt{srce}_3$	4	3	midtown	900
tgt	2	1	downtown	?

Table 1: Descriptions of three source cases $\texttt{srce}_1$, $\texttt{srce}_2$, and $\texttt{srce}_3$, and of the target case `tgt` in the apartment price prediction scenario.

The co-variation $\texttt{nb\_rooms}^{\leq} \rightsquigarrow \texttt{price}^{\leq}$ may be extracted from the case base and transferred to `tgt`. This co-variation expresses that, in previous experiences, an increase (resp., decrease) in the number of rooms results in an increase (resp., decrease) in price. Applying this co-variation enables to make the hypothesis that the price of `tgt` is between 290 $ and 440 $.

5.2 Some Remarks

Several remarks can be made when considering such an approach to adaptation.

1 - Crisp *vs* Fuzzy. Co-variations give only a partial account of the (*Monotonicity*) property introduced in Sec. 3. A co-variation expresses only that for some $e, f \in \mathcal{U} \times \mathcal{U}$, if $e = f$ then $a(e) = a(f)$. Having a more strict account of the (*Monotonicity*) property may be achieved by considering a more fuzzy version of variations, for example in the spirit of what is done in [1].

2 - Relevance. Co-variations map some problem variations to some solution variations. Choosing which co-variations to apply for a given target amounts to solving the relevance problem [10], which consists in identifying which relation on the sources cases are relevant and should be transferred to the target case. If more than one co-variation apply, one may also need to choose which one to favor, in accordance with the quality criteria outlined in Sec. 2.

3 - Context. Related to relevance is the notion of context: some co-variations may be highly contextual and our modeling of variations does not allow to render it easily. For example, the floor may influence the price differently depending on the area, because in some parts of town the buildings are smaller and have no elevator, so beyond the 4^{th} floor the price of the apartments would decrease in those areas. This knowledge could be approximated by co-variations such as $\mathtt{area}^{\mathtt{downtown}=} \wedge \mathtt{nb\_rooms}^{=} \wedge \mathtt{floor}^{\leq} \rightsquigarrow \mathtt{price}^{\leq}$ but many rules would be needed to approach the complexity of the relation between the attributes `floor` and `price`. In particular, it would be interesting to have the notion of *ceteris paribus* co-variations just as what is done for preference representation [28], to represent for example that "everything else being equal", the price increases when the floor increases.

4 - (De-)Composition. Each co-variation may only suggest partial modifications of a source case. To obtain a complete target case, this adapation strategy could be coupled with a strategy of case decomposition [26], of combinaison of solutions [23], or with a planning strategy [2].

6 Related Work

The importance for adaptation to assess the differences between cases has long been recognized in the CBR literature. Several adaptation methods make use of a representation of the differences between problems to generate a difference between solutions: [7] encodes these differences in symbolic properties, whereas other methods [6, 9, 18] compute numerical values. In [17], a reformulation is a pair $(\mathtt{r}, \mathtt{A_r})$ of binary relations such that if `pb r pb'` for two problems `pb` and `pb'`, then any solution `sol(pb)` of `pb` can be adapted in a solution `sol(pb')` of `pb'`. According to this definition, a reformulation is simply a co-variation that constitutes a valid approximation of the case-based inference function. Different methods [6, 7, 12, 18] were also proposed to derive adaptation knowledge from differences between cases.

In transfer learning [15], the difference between source and target problems is what characterizes the transfer distance. The proposed case-based inference process can be considered as a kind of transfer learning method since it consists in transferring knowledge learned from the case base to help learn a description of the target. Our proposal is to transfer relational knowledge from the case base to the target, following the principles of [24]. This can be contrasted with the approach of [21], where the CBI exploits a hierarchical structure of the case base to transfer knowledge from the source case to the target.

Our approach can be seen as a symbolic account of the credible case-based inference [13], where we adopt a logical approach to similarity [25] by modeling the similarity relation as a preorder on the set of pairs of cases. The CBR hypothesis ("similar problems have similar solutions") is a regularity principle, which is expressed by requiring that the CBI function obey the (*Monotonicity*) property. Regularity can be linked to the idea of minimum description length in analogy [20].

The idea that the target case should optimize a quality measure on the solution space was introduced in [5].

7 Conclusion and Future Work

This article gives the main principles of a data-driven method to adaptation. Co-variations are used to extrapolate to the target case some parts of the relational structure underlying the case base. Such an approach relies on a modeling the case-based inference process as a self map on the set of pairs of cases. By doing so, we show that adaptation can be (at least partially) included in the case-based inference process.

Future work includes confronting the model to a more complex example in order to determine in particular what choices should be made to tackle the relevance and context challenges outlined in Sec. 5. We are currently implementing the co-variation learning method proposed in [4]. This learning method could be used to learn co-variations on the fly from the case base so that they can be transferred to the target during the reasoning.

Acknowledgements. The author wishes to thank the reviewers for their constructive remarks.

References

1. Arrazola, I., Plainfossé, A., Prade, H., Testemale, C.: Extrapolation of Fuzzy Values from Incomplete Data Bases. Inf. Syst. 14(6) (1989)
2. Badra, F.: Case Adaptation with Modal Logic: The Modal Adaptation. In: Proc. Int. Conf. Case-Based Reason. ICCBR'14 (2014)
3. Badra, F.: Representing and Learning Variations. In: Int. Conf. Tools Artif. Intell. pp. 950–957. IEEE, Vietri sul Mare (2015)
4. Badra, F.: Reasoning with Co-variations. In: AIMSA. Varna, Bulgaria (2016)
5. Bergmann, R., Wilke, W.: Towards a New Formal Model of Transformational Adaptation in Case-Based Reasoning. Thirteen. Eur. Conf. Artif. Intell. pp. 53–57 (1998)
6. Craw, S., Wiratunga, N., Rowe, R.C.: Learning adaptation knowledge to improve case-based reasoning. Artif. Intell. 170(16-17), 1175–1192 (nov 2006)
7. D'Aquin, M., Badra, F., Lafrogne, S., Lieber, J., Napoli, A., Szathmary, L.: Case base mining for adaptation knowledge acquisition. In: IJCAI Int. Jt. Conf. Artif. Intell. pp. 750–755. Hyderabad, India (2007)
8. Derrac, J., Schockaert, S.: Inducing semantic relations from conceptual spaces: A data-driven approach to plausible reasoning. Artif. Intell. 228, 66–94 (2015)
9. Fuchs, B., Lieber, J., Mille, A., Napoli, A.: Differential adaptation: An operational approach to adaptation for solving numerical problems with CBR. Knowledge-Based Syst. 68, 103–114 (apr 2014)
10. Gust, H., Krumnack, U., Kühnberger, K., Schwering, A.: Analogical Reasoning: A Core of Cognition. KI - Künstliche Intelligenz 22(1), 8–12 (2008)
11. Hammond, K.J.: CHEF: A Model of Case-based Planning. In: AAAI Proc. pp. 267–271 (1986)

12. Hanney, K., Keane, M.T.: The Adaptation Knowledge Bottleneck: How to Ease it by Learning from Cases
13. Hüllermeier, E.: Credible case-based inference using similarity profiles. IEEE Trans. Knowl. Data Eng. 19(6), 847–858 (2007)
14. Hüllermeier, E.: From knowledge-based to data-driven fuzzy modeling - Development, criticism, and alternative directions. Inform. Spektrum 38(6), 500–509 (2015)
15. Klenk, M., Aha, D.W., Molineaux, M.: The Case for Case-Based Transfer Learning pp. 54–69 (2011)
16. Lieber, J.: Application of the Revision Theory to Adaptation in Case-Based Reasoning: The Conservative Adaptation. In: ICCBR Proc. (2007)
17. Lieber, J., Napoli, A.: Correct and Complete Retrieval for Case-Based Problem-Solving. In: ECAI. pp. 68–72 (1998)
18. McSherry, D.: Demand-driven discovery of adaptation knowledge. IJCAI pp. 222–227 (1999)
19. Miclet, L., Bayoudh, S., Delhay, A.: Analogical Dissimilarity: Definition, Algorithms and Two Experiments in Machine Learning 32, 793–824 (2008)
20. Murena, P.a., Cornuéjols, A.: Minimum Description Length Principle applied to Structure Adaptation for Classification under Concept Drift. In: Proc. IJCNN-2016 (Int. Jt. Conf. Neural Networks (2016)
21. Ontañón, S., Plaza, E.: On Knowledge Transfer in Case-Based Inference. In: Díaz-Agudo, B., Watson, I. (eds.) ICCBR Proc. pp. 312–326. Springer-Verlag Berlin Heidelberg (2012)
22. Prade, H., Richard, G.: Reasoning with Logical Proportions. In: Princ. Knowl. Represent. Reason. Proc. Twelfth Int. Conf. pp. 545–555. Toronto, Canada (2010)
23. Purvis, L., Pu, P.: COMPOSER: A Case Based Reasoning System for Engineering Design. Robotica 16(3), 285–295 (1998)
24. Ragni, M., Strube, G.: Cognitive Complexity and Analogies in Transfer Learning. KI - Künstliche Intelligenz 28(1), 39–43 (2014), http://dx.doi.org/10.1007/s13218-013-0288-6
25. Richter, M.M.: Classification and Learning of Similarity Measures. In: Inf. Classif., pp. 323–334 (1993)
26. Smyth, B., Cunningham, P.: Déjà Vu: A Hierarchical Case-Based Reasoning System for Software Design. In: ECAI. vol. 92, pp. 587–589 (1992)
27. Sombé, L.: Reasoning under incomplete information in artificial intelligence: A comparison of formalisms using a single example. Int. J. Intell. Syst. 5(4), 323–372 (1990)
28. Van Benthem, J., Girard, P., Roy, O.: Everything else being equal: A modal logic for ceteris paribus preferences. J. Philos. Log. 38(1), 83–125 (2009)
29. Zadeh, L.: Fuzzy Logic and Approximate Reasoning. J. Chem. Inf. Model. 30(i), 407–428 (1975)