HAL
open science

An approach to case-based reasoning based on local enrichment of the case base

Yves Lepage, Jean Lieber

To cite this version:

Yves Lepage, Jean Lieber. An approach to case-based reasoning based on local enrichment of the case base. ICCBR 2019 - 27th International Conference on Case-Based Reasoning Research and Development, Sep 2019, Otzenhausen, Germany. hal-02370721

HAL Id: hal-02370721

https://inria.hal.science/hal-02370721v1

Submitted on 19 Nov 2019

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

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

An approach to case-based reasoning based on local enrichment of the case base

Yves Lepage and Jean Lieber*

Waseda University, IPS, 2-7 Hibikino, 808-0135 Kitakyushu, Japan
yves.lepage@waseda.jp
Université de Lorraine, CNRS, Inria, LORIA, F-54000 Nancy, France
jean.lieber@loria.fr

Abstract. This paper describes an approach to case-based reasoning by which the case base is enriched at reasoning time. Enrichment results from the local application of variations to seed cases: new hypothetical cases are created which get closer and closer to the target problem. The creation of these hypothetical cases is based on structures associated to the problem and solution spaces, called variation spaces, that enable to define a language of adaptation rules. Ultimately reaching the target problem (exactly or nearly) allows the system to deliver a solution. A realistic application of the proposed approach to machine translation between French and English shows behind state-of-the-art, but promising results.

Keywords: analogical reasoning, case base enrichment, case-based machine translation, case-based reasoning

1 Introduction

In machine learning, some techniques are used to enrich the training set in order to improve the accuracy of a learning system. This is called data augmentation. It can be done using general transformations (e.g., flipping, rotating, etc. images when this does not affect the class in image classification, see for example [9] or [16]), adding some controlled noise (like Gaussian noise on images [7]), or by analogical reasoning (see, e.g. [2]).

While case-based reasoning (CBR [15]) is usually less greedy than most current machine learning techniques, the enrichment of the case base can be useful. In this paper, we propose to perform the enrichment of the case base in a "case-based way": the case base is enriched with new source cases that are "around" the target problem. This constitutes a local enrichment of the case base, computed on-line, at the time of problem-solving. The case base is enriched with the help of learned adaptation rules: given a source case, another source case closer to the target problem is generated thanks to such a rule. The process can be repeated, though each application of a rule may degrade the generated case, in the sense that it is less and less likely to be a licit case.

* The first author is supported by JSPS Grant-In-Aid 18K11447: "Self-explainable and fast-to-train example-based machine translation using neural networks."

To deal with this issue, the notion of penalized case is introduced: this is a hypothetical case, whose likelihood to be licit is characterized by a number that penalizes the re-usability of the case. Therefore, the enriched case base is a set of penalized cases: the cases to be reused are the ones which offer the best compromise between the similarity to the target problem (e.g., in the sense of a distance function) and their penalties.

This paper is organized as follows. Section 2 presents general definitions, notations and assumptions. The notions of penalized cases and penalized case bases, mentioned above, are presented with details in Section 3. The notions of variations between problems and between solutions, and some related notions useful for presenting the approach, are defined in Section 4, with some strong assumptions that make the general ideas easier to understand. Section 5 is the core of the paper: it presents the approach for local enrichment of the case base. Two applications are presented in Section 6, where the strong assumptions mentioned above about the representation of variations do not necessary hold. However, the principle of the approach still holds. These applications show how case-based translation [10] can be actually performed thanks to the approach of local enrichment of the case base. The results are demonstrative and promising.

2 Preliminaries

This section presents some general notions and assumptions about case-based reasoning and analogies, with the notations used throughout the paper.

2.1 Notations and assumptions on case-based reasoning

Let $\mathcal{P}$ and $\mathcal{S}$ be two given sets, called the *problem space* and the *solution space*. A *problem* is by definition an element $\mathtt{x}$ of $\mathcal{P}$ and a *solution*, an element $\mathtt{y}$ of $\mathcal{S}$. A relation $\rightsquigarrow$ on $\mathcal{P} \times \mathcal{S}$ is assumed to exist and $\mathtt{x} \rightsquigarrow \mathtt{y}$ is read "$\mathtt{x}$ has for solution $\mathtt{y}$" or "$\mathtt{y}$ is a solution to $\mathtt{x}$". A case is a pair $(\mathtt{x}, \mathtt{y}) \in \mathcal{P} \times \mathcal{S}$ such that $\mathtt{x} \rightsquigarrow \mathtt{y}$. The *case base* $\mathtt{CB}$ is a finite and nonempty set of cases. A *source case* is an element $(\mathtt{x}^s, \mathtt{y}^s)$ of $\mathtt{CB}$. Case-based reasoning (CBR) aims at solving a new problem $\mathtt{x}^t$ called the *target problem* with the help of the case base. It usually consists in the following steps:

- Retrieval (aka "retrieve" in [1]) selects a subset of $\mathtt{CB}$;
- Adaptation (aka "reuse") proposes a *plausible* solution $\mathtt{y}^t$ to $\mathtt{x}^t$, using the retrieved source cases;
- Learning (aka as "revise" and "retain") consists in validating/correcting $\mathtt{y}^t$ (for example, with the help of a human expert) and in storing the newly formed case $(\mathtt{x}^t, \mathtt{y}^t)$ in the case base, if this storage is judged appropriate.

It is worth noting that, for many applications, $\mathtt{y}^t$ is only a plausible solution: CBR often appears as a hypothetical reasoning whose use is motivated by the incompleteness of the knowledge of the relation $\rightsquigarrow$. The term "hypothetical case"

stands for any pair $(\mathtt{x}, \mathtt{y}) \in \mathcal{P} \times \mathcal{S}$, though this notion is generally used when a solution $\mathtt{y}$ to $\mathtt{x}$ is plausibly inferred.

In some applications, other knowledge containers are used [15]: the domain knowledge (DK, aka domain ontology), the retrieval knowledge (RK), and the adaptation knowledge (AK). These four knowledge containers (CB, DK, RK and AK) are interrelated. In particular, there are studies on learning AK using CB presented further.

2.2 Analogies

An analogy is a quaternary relation on a set $\mathcal{U}$ denoted by $A : B :: C : D$ for $(A, B, C, D) \in \mathcal{U}^4$, that is to be read "$A$ is to B as C is to D". It satisfies the following postulates (for any $(A, B, C, D) \in \mathcal{U}^4$):[1]

(reflexivity of conformity) $A : B :: A : B$;
(symmetry of conformity) if $A : B :: C : D$ then $C : D :: A : B$;
(exchange of the means) if $A : B :: C : D$ then $A : C :: B : D$.

A *ratio* is an expression of the form $P : Q$ ("P is to Q"), the relation $::$ ("as") is called conformity. Thus, analogy is a conformity of ratios.

Classical examples of analogies are as follows:

(geometrical analogy) Here the ratio is division, conformity is equality, and $\mathcal{U} = \mathbb{R} \setminus \{0\}$. This analogy is defined by $A : B :: C : D$ if $\frac{A}{B} = \frac{C}{D}$.
(arithmetic analogy) Here the ratio is subtraction, conformity is equality, and $\mathcal{U} = \mathbb{R}$. This analogy is defined by $A : B :: C : D$ if $A - B = C - D$.
(analogy on tuples) If an analogy is defined on each set $\mathcal{U}_i$ $(1 \leq i \leq n)$ and $\mathcal{U} = \mathcal{U}_1 \times \mathcal{U}_2 \times \ldots \times \mathcal{U}_n$ then the following analogy can be defined on $\mathcal{U}$: $A : B :: C : D$ if for every i, $A_i : B_i :: C_i : D_i$.
(analogies on strings) Let dist be the LCS distance.[2] The Parikh vector of a string is the tuple of the number of occurrences of each character in the string. A ratio $P : Q$ between two strings P and Q can be defined as the difference of their Parikh vectors (arithmetic analogies on tuples) plus the LCS distance between them. However, this ratio does not entail the exchange of the means because $\mathtt{dist}(A, B) = \mathtt{dist}(C, D)$ does not imply $\mathtt{dist}(A, C) = \mathtt{dist}(B, D)$ in general. To define an analogy it is necessary to state: $A : B :: C : D$ if $A : B = C : D$ and $A : C = B : D$.

An analogical equation is an expression of the form $A : B :: C : y$, where y is the unknown. Solving it amounts to find the set of y such that the analogy holds.

[1] Some authors consider that analogy requires additional postulates [11]. However, only these three postulates are taken into account in this paper.

[2] The LCS ("longest common subsequence") distance is an edit distance based on the character insertion and character deletion edit operations, with a cost of 1 for both. In other terms, if P and Q are two strings and L is the LCS of P and Q, then $\mathtt{dist}(P, Q) = (|P| - |L|) + (|Q| - |L|)$.

It may have 0, 1 or several solutions, depending on the type of analogies: for geometrical analogies on $\mathbb{R} \setminus \{0\}$ and arithmetic analogies on $\mathbb{R}$, every analogical equation has exactly one solution. By contrast, with the analogy on strings defined above, an analogical equation may have 0, 1 or several solutions.

3 Penalized cases and penalized case bases

The enrichment of the case base presented in this paper is based on a hypothetical reasoning: it generates hypothetical cases $(\mathtt{x}, \mathtt{y})$, which means that the assertion "$\mathtt{x} \rightsquigarrow \mathtt{y}$" is uncertain. Thus, a hypothetical case is less trustworthy than a licit case (such as a source case), and thus, the former has to be penalized in the reasoning, in comparison to the latter. Two hypothetical cases should have different penalties if one of them is more uncertain than the other. The notion of penalty as a way to model uncertainty is introduced for this purpose. The penalties are associated to hypothetical inferences: the more uncertain is the inference, the higher is the inference cost, that is the additional penalty associated with the inference. Finally, the notion of penalized case is introduced: they are triples $(\mathtt{x}, \mathtt{y}, \pi)$ where $\mathtt{x} \in \mathcal{P}$, $\mathtt{y} \in \mathcal{S}$ and π is a penalty. An enriched case base is actually a set of penalized cases, i.e., a penalized case base.

3.1 Uncertainty and Penalties

A hypothetical reasoning leads to an uncertain result. In this paper, uncertainty of an event is measured by an *uncertainty penalty* (or, simply, a *penalty*) $\pi \in [0, \infty]$ such that the higher π is, the less certain the event is. The penalty of an event that is certain is $\pi = 0$. The penalty of an impossible event is $\pi = \infty$.[3]

Remark 1: If two penalties π^1 and π^2 are associated to the same event, with $\pi^1 < \pi^2$, the lower penalty —associated to the higher certainty— is kept. In other words, if π is associated to an event e, then every $\pi' \geq \pi$ can also be associated to e.

3.2 Cost associated to a hypothetical inference

Let φ^0 be a piece of knowledge whose uncertainty (to be consistent with the real world) has an uncertainty penalty π^0. From φ^0, a new piece of knowledge φ^1 can be produced by a hypothetical inference `hypo`. Since `hypo` is hypothetical, it adds some uncertainty, thus an uncertainty penalty π^1 associated to φ^1 can be computed as $\pi^1 = \pi^0 + \mathtt{c}$ where $\mathtt{c} > 0$ measures the additional uncertainty

[3] If uncertainty is modeled thanks to a probability measure, it is possible to associate to a probability $P \in [0; 1]$ a penalty $\pi = -\log P \in [0; \infty]$. However, the representation of uncertainty by penalties is preferred in this paper, for practical reasons.

of $\mathtt{hypo}$. $\mathtt{c}$ is assumed to be computed on the basis on $\mathtt{hypo}$ by a function called $\mathtt{cost}$: $\mathtt{c} = \mathtt{cost}(\mathtt{hypo})$.[4] In summary:

$$\begin{aligned} &\textbf{if } \pi^0 \text{ is an uncertainty penalty associated to } \varphi^0 \\ &\textbf{and } \varphi^1 \text{ is inferred from } \varphi^0 \text{ by the hypothetical inference } \mathtt{hypo} \\ &\textbf{then } \pi^1 = \pi^0 + \mathtt{cost}(\mathtt{hypo}) \text{ is an uncertainty penalty associated to } \varphi^1 \end{aligned} \tag{1}$$

Now, if φ^0, φ^1, ..., φ^n are pieces of knowledge, φ^0 being certain (it can be associated to an uncertainty penalty $\pi^0 = 0$) and φ^i being produced by a hypothetical inference $\mathtt{hypo}_i$ ($1 \leq i \leq n$), then, according to (1), the uncertainty penalty $\pi^n = \sum_{i=1}^{n} \mathtt{cost}(\mathtt{hypo}_i)$ can be associated to φ^n.

In particular, let us consider the simple approach to CBR consisting in retrieving a source case $(\mathtt{x}^s, \mathtt{y}^s)$ and in reusing (without modification) $\mathtt{y}^s$ as a plausible solution to the target problem $\mathtt{x}^t$ ($\mathtt{y}^t = \mathtt{y}^s$). If retrieval is based on a distance function $\mathtt{dist}$, then the higher $\mathtt{dist}(\mathtt{x}^s, \mathtt{x}^t)$ is, the more the assertion $\mathtt{x}^t \rightsquigarrow \mathtt{y}^t$ is uncertain, and the higher the cost of this inference has to be. For this reason, $\mathtt{dist}(\mathtt{x}^s, \mathtt{x}^t)$ can be used to measure the cost of this inference: this is how the distance function is interpreted and used in the rest of the paper.

3.3 Penalized cases

In this paper, every hypothetical case $(\mathtt{x}, \mathtt{y}) \in \mathcal{P} \times \mathcal{S}$ is either a licit case ($\mathtt{x} \rightsquigarrow \mathtt{y}$) or not ($\mathtt{x} \not\rightsquigarrow \mathtt{y}$): there is no gradual distinction between licit and illicit cases. By contrast, a hypothetical case is more or less certain to be a licit case. So, a hypothetical case should be preferred to another one on the basis of their respective chances of being licit. A *penalized case* is a triple $(\mathtt{x}, \mathtt{y}, \pi)$ with $(\mathtt{x}, \mathtt{y})$, a hypothetical case and π, a penalty measuring the uncertainty that $\mathtt{x} \rightsquigarrow \mathtt{y}$ is licit. The estimation of π is made on the basis of an inference that has conducted to the hypothetical case $(\mathtt{x}, \mathtt{y})$. If $\pi = 0$, then $(\mathtt{x}, \mathtt{y}, \pi) = (\mathtt{x}, \mathtt{y}, 0)$ is assimilated to the case $(\mathtt{x}, \mathtt{y})$.

A *penalized case base* $\mathtt{PCB}$ is a finite set of penalized cases $(\mathtt{x}^s, \mathtt{y}^s, \pi^s) \in \mathcal{P} \times \mathcal{S} \times [0, \infty]$. In particular, $\mathtt{CB}$ is a penalized case base with every penalty set to 0.

When a penalized case base $\mathtt{PCB}$ is used, instead of a classical case base, how does it affect the CBR process? An answer is to take into account the penalties of the case by adding them to the cost of the inference. For example, since $\mathtt{dist}$ is interpreted as a cost of the simple CBR inference based on the retrieval of a single source case and reusing it as such (cf. the last paragraph of Section 3.2), this approach to CBR consists in selecting the $(\mathtt{x}^s, \mathtt{y}^s, \pi^s) \in \mathtt{PCB}$ which minimizes $\mathtt{dist}(\mathtt{x}^s, \mathtt{x}^t) + \pi^s$.

[4] Once again, costs could be associated to probabilities: $\mathtt{cost}(\mathtt{hypo})$ could be defined by $-\log P(\varphi^2 \mid \varphi^1)$, where $P(\varphi^2 \mid \varphi^1)$ is the probability of φ^2 being true given that φ^1 is.

4 Problem and solution variations

This section introduces the notions of problem variations and solution variations that are useful to present our approach. It also relates these notions to the issue of adaptation knowledge learning and to analogies.

4.1 Definitions

Intuitively, the variation from a problem $\mathtt{x}^i$ to a problem $\mathtt{x}^j$, denoted by $\overrightarrow{\mathtt{x}^i\mathtt{x}^j}$ in this paper, encodes the information necessary to transform $\mathtt{x}^i$ into $\mathtt{x}^j$. More formally, a triple $(\Delta\mathcal{P}, +, \overrightarrow{\cdot})$ called the *problem variation space* and verifying the following postulates is assumed to exist:

(i) $(\Delta\mathcal{P}, +)$ is a commutative group. Its neutral element is denoted by $\overrightarrow{0}$; the inverse of $\overrightarrow{u} \in \Delta\mathcal{P}$ is denoted by $-\overrightarrow{u}$.[5]

(ii) $\overrightarrow{\cdot}$ is an onto mapping $(\mathtt{x}^i, \mathtt{x}^j) \in \mathcal{P}^2 \mapsto \overrightarrow{\mathtt{x}^i\mathtt{x}^j} \in \Delta\mathcal{P}$: for each $\overrightarrow{u} \in \Delta\mathcal{P}$, there exists $(\mathtt{x}^i, \mathtt{x}^j) \in \mathcal{P}^2$ such that $\overrightarrow{u} = \overrightarrow{\mathtt{x}^i\mathtt{x}^j}$.

(iii) For every $\mathtt{x}^i, \mathtt{x}^j, \mathtt{x}^k \in \mathcal{P}$, $\overrightarrow{\mathtt{x}^i\mathtt{x}^j} + \overrightarrow{\mathtt{x}^j\mathtt{x}^k} = \overrightarrow{\mathtt{x}^i\mathtt{x}^k}$.

(iv) For each $\mathtt{x}^i \in \mathcal{P}$ and $\overrightarrow{u} \in \Delta\mathcal{P}$, there exists *at most* one $\mathtt{x}^j \in \mathcal{P}$ such that $\overrightarrow{\mathtt{x}^i\mathtt{x}^j} = \overrightarrow{u}$. This $\mathtt{x}^j$, *when it exists*, is denoted by $tr_{\overrightarrow{u}}(\mathtt{x}^i)$ (tr stands for "translation", borrowing the term from the field of vector spaces).

From these postulates, the following properties can be deduced:

$$\overrightarrow{\mathtt{x}^i\mathtt{x}^j} = \overrightarrow{0} \qquad \text{iff} \qquad \mathtt{x}^i = \mathtt{x}^j \tag{2}$$

$$-\overrightarrow{\mathtt{x}^i\mathtt{x}^j} = \overrightarrow{\mathtt{x}^j\mathtt{x}^i} \tag{3}$$

$$\text{if } \overrightarrow{\mathtt{x}^1\mathtt{x}^2} = \overrightarrow{\mathtt{x}^3\mathtt{x}^4} \text{ then } \overrightarrow{\mathtt{x}^1\mathtt{x}^3} = \overrightarrow{\mathtt{x}^2\mathtt{x}^4} \tag{4}$$

for every $\mathtt{x}^i, \mathtt{x}^j, \mathtt{x}^1, \mathtt{x}^2, \mathtt{x}^3, \mathtt{x}^4 \in \mathcal{P}$.

There are many ways of defining $\Delta\mathcal{P}$ and the mapping $\overrightarrow{\cdot}$. They depend partly upon the problem space $\mathcal{P}$. For example:

- If $\mathcal{P}$ is an affine space of dimension n on $\mathbb{R}$, $\Delta\mathcal{P}$ can be the vector space $\mathbb{R}^n$ associated with $\mathcal{P}$: $\overrightarrow{\mathtt{x}^i\mathtt{x}^j} = (\mathtt{x}^j_1 - \mathtt{x}^i_1, \mathtt{x}^j_2 - \mathtt{x}^i_2, \ldots, \mathtt{x}^j_n - \mathtt{x}^i_n)$ and $tr_{\overrightarrow{u}}$ is the translation operator of vector $\overrightarrow{u}$. This example explains the notations chosen in this paper.
- More generally, if $\mathcal{P}$ is defined by attribute-value pairs, the problem of defining $\overrightarrow{\mathtt{x}^i\mathtt{x}^j}$ can be reduced to the problem of defining the variation from $\mathtt{x}^i$ to $\mathtt{x}^j$ for each attribute. This is considered in particular in [4].

In the same way and with the same notations, a solution variation space $(\Delta\mathcal{S}, +, \overrightarrow{\cdot})$ can be defined.

[5] $(\Delta\mathcal{P}, +)$ being a commutative group means that $\Delta\mathcal{P}$ is a set, that $+$ is an associative and commutative operation on $\Delta\mathcal{P}$, and that every $\overrightarrow{u} \in \Delta\mathcal{P}$ has an inverse element $-\overrightarrow{u}$ (meaning $\overrightarrow{u} + (-\overrightarrow{u}) = \overrightarrow{0}$).

4.2 Adaptation knowledge learning expressed in terms of variations

The seminal paper [6] presents the main principles of the AK learning issue. They are reformulated below, thanks to the notions of variations introduced above.

From the case base, the multiset[6]

$$\mathtt{TS} = \left\{\!\!\left\{ \left(\overrightarrow{\mathtt{x}^i\mathtt{x}^j}, \overrightarrow{\mathtt{y}^i\mathtt{y}^j}\right) \;\middle|\; (\mathtt{x}^i, \mathtt{y}^i), (\mathtt{x}^j, \mathtt{y}^j) \in \mathtt{CB},\ \text{with } \mathtt{x}^i \neq \mathtt{x}^j \right\}\!\!\right\}$$

is computed. This multiset is used in the training of a supervised learning process (the inputs of the examples are the $\overrightarrow{\mathtt{x}^i\mathtt{x}^j}$, the outputs are the $\overrightarrow{\mathtt{y}^i\mathtt{y}^j}$). The learned model is used as adaptation knowledge.

Several studies have followed this scheme, and a few examples are given below. In [3], a variety of learning techniques are used, in particular decision tree induction and ensemble learning. In [4], frequent closed itemset extraction in used. The expert interpretation enables to produce adaptation rules to be added to AK. In [8], an ensemble approach provides adaptation rules for a nominal representation (feature-value pairs, where values are categories).

An example of adaptation learning approach suited for discrete representations is as follows. First, a triple $(\overrightarrow{u}, \overrightarrow{v}, \mathtt{c}) \in \Delta\mathcal{P} \times \Delta\mathcal{S} \times [0, \infty[$ such that $\overrightarrow{u} \neq \overrightarrow{0}$ can be seen as an adaptation rule (for $(\mathtt{x}, \mathtt{y}, \pi)$, a penalized case and $\mathtt{x}^t \in \mathcal{P}$):

if $\overrightarrow{\mathtt{x}\mathtt{x}^t} = \overrightarrow{u}$ **then** $\mathtt{y}^t = tr_{\overrightarrow{v}}(\mathtt{y})$ is a plausible solution of $\mathtt{x}^t$, with penalty $\pi + \mathtt{c}$

(recall that $\mathtt{y}^t = tr_{\overrightarrow{v}}(\mathtt{y})$ iff $\overrightarrow{\mathtt{y}\mathtt{y}^t} = \overrightarrow{v}$). Among the $(\overrightarrow{u}, \overrightarrow{v}, \mathtt{c}) \in \Delta\mathcal{P} \times \Delta\mathcal{S} \times [0, \infty[$ ($\overrightarrow{u} \neq \overrightarrow{0}$), the ones that are selected are the ones which are the most supported by the training set. More formally, let $\mathtt{supp}\,(\overrightarrow{u}, \overrightarrow{v})$ (the *support* of the ordered pair $(\overrightarrow{u}, \overrightarrow{v})$) be the multiplicity of $(\overrightarrow{u}, \overrightarrow{v})$ in TS. In other terms:

$$\mathtt{supp}\,(\overrightarrow{u}, \overrightarrow{v}) = \left|\left\{ ((\mathtt{x}^i, \mathtt{y}^i), (\mathtt{x}^j, \mathtt{y}^j)) \in \mathtt{CB}^2 \;\middle|\; \overrightarrow{\mathtt{x}^i\mathtt{x}^j} = \overrightarrow{u},\ \overrightarrow{\mathtt{y}^i\mathtt{y}^j} = \overrightarrow{v} \right\}\right|$$

Hence, the adaptation knowledge learning process consists in computing the pairs $(\overrightarrow{u}, \overrightarrow{v})$ such that their support is above a given threshold $\tau_{\mathtt{supp}}$. It is assumed in this paper that $\tau_{\mathtt{supp}} \geq 2$. The value of the support is used on the basis of the following heuristics: the higher the support is, the less the application of the adaptation rule adds uncertainty. Therefore, a value c is computed thanks to a decreasing function $f : \mathbb{N} \setminus \{0, 1\} \to \mathbb{R}$ by $\mathtt{c} = f(\mathtt{supp}\,(\overrightarrow{u}, \overrightarrow{v}))$. For our experiments, we have chosen $f(n) = 1/n$. Finally, the rule $(\overrightarrow{u}, \overrightarrow{v}, \mathtt{c})$ is added to AK.

With this adaptation knowledge learning approach, it is noteworthy that for each learned adaptation rule $(\overrightarrow{u}, \overrightarrow{v}, \mathtt{c})$ there is another learned adaptation rule $(-\overrightarrow{u}, -\overrightarrow{v}, \mathtt{c})$. Indeed, $\left(\overrightarrow{\mathtt{x}^i\mathtt{x}^j}, \overrightarrow{\mathtt{y}^i\mathtt{y}^j}\right)$ occurs in the multiset TS with the same multiplicity as $\left(\overrightarrow{\mathtt{x}^j\mathtt{x}^i}, \overrightarrow{\mathtt{y}^j\mathtt{y}^i}\right)$.

[6] A multiset is denoted with double braces; for example $M = \{\!\!\{a, a, b, c, c, c\}\!\!\}$ contains a with multiplicity 2, b with multiplicity 1 and c with multiplicity 3. Thus the cardinality of M is $|M| = 2 + 1 + 3 = 6$.

4.3 Variation-based analogies

Let us consider the relation on problems defined, for $A, B, C, D \in \mathcal{P}$, as follows:

$$A : B :: C : D \quad \text{if} \quad \overrightarrow{AB} = \overrightarrow{CD}$$

It satisfies the postulates of analogy. An analogy on $\mathcal{S}$ can be defined likewise.

Therefore, using the problem variation space and the solution variation space, an analogy on $\mathcal{P}$ and an analogy on $\mathcal{S}$ can be built, and thus, the approach to CBR based on the following principle (called *extrapolation* in [13] and used in [12] and [10]) can be applied:

- Given a target problem $\mathtt{x}^t$, a triple $((\mathtt{x}^1, \mathtt{y}^1), (\mathtt{x}^2, \mathtt{y}^2), (\mathtt{x}^3, \mathtt{y}^3)) \in \mathtt{CB}^3$ is retrieved such that $\mathtt{x}^1 : \mathtt{x}^2 :: \mathtt{x}^3 : \mathtt{x}^t$ in the problem space (i.e., $\overrightarrow{\mathtt{x}^1\mathtt{x}^2} = \overrightarrow{\mathtt{x}^3\mathtt{x}^t}$).
- Then, the analogical equation $\mathtt{y}^1 : \mathtt{y}^2 :: \mathtt{y}^3 : y$ in the solution space is solved, and the solution of this equation is given as a plausible solution to $\mathtt{x}^t$ (in the framework of the postulates given below, this solution, when it exists, is unique, and verifies $y = tr_{\overrightarrow{v}}(\mathtt{y}^3)$ with $\overrightarrow{v} = \overrightarrow{\mathtt{y}^1\mathtt{y}^2}$).

4.4 Analogy-based variations

In the domain of strings, it is possible to define a vector corresponding to a ratio as follows (using the notion of ratios introduced in Section 2.2):

$$\overrightarrow{AB} = A : B$$

However, as mentioned in Section 2.2, the only definition of a ratio does not make an analogy. In such a domain, we impose for $(A, B, C, D) \in \mathcal{P}^4$:

$$\overrightarrow{AB} = \overrightarrow{CD} \quad \text{if} \quad A : B :: C : D, \text{ i.e., if } A : B = C : D \text{ and } A : C = B : D$$

In other words, in such a domain, we implement the extrapolation approach mentioned at the end of Section 4.3 by restraining ourselves to the use of variations such that both $\overrightarrow{AB} = \overrightarrow{CD}$ and $\overrightarrow{AC} = \overrightarrow{BD}$ hold at the same time. This ensures that the postulates of analogy are verified for the used variations.

5 CBR by local enrichment of the case base

The enrichment of the case base consists in adding to the original case base $\mathtt{CB}$ some (penalized) cases inferred from $\mathtt{CB}$. The inferences considered in this paper consist in applying the learned adaptation rules $(\overrightarrow{u}, \overrightarrow{v}, \mathtt{c}) \in \mathtt{AK}$. In theory, all the penalized cases that can be so inferred can enrich $\mathtt{CB}$: this is considered in Section 5.1. However, this leads usually to a penalized case base that is too large. In Section 5.2, a local enrichment is proposed that consists in adding to $\mathtt{CB}$ penalized cases that are "around" the target problem.

5.1 Theoretical view: global enrichment of CB

The principle of global enrichment of CB is simple: it consists in computing all the penalized cases $(\mathtt{x}, \mathtt{y}, \pi)$ that can be inferred from CB by application of one or several adaptation rules, with the constraint $\pi \leq \tau_{\mathtt{penalty}}$, where $\tau_{\mathtt{penalty}}$ is a given threshold.

The size of the enriched case base, PCB, can be estimated as follows. Let $p = |\mathtt{AK}|$ and $d = \lfloor \tau_{\mathtt{penalty}} / \min\{\mathtt{c} \mid (\overrightarrow{u}, \overrightarrow{v}, \mathtt{c}) \in \mathtt{AK}\} \rfloor$. If no hypothetical case $(\mathtt{x}, \mathtt{y})$ is generated twice in the process, then $|\mathtt{PCB}| = |\mathtt{CB}| \times \dfrac{p^{d+1} - 1}{p - 1}$, assuming $p \neq 1$. Therefore, $|\mathtt{PCB}| \,/\, |\mathtt{CB}| = \mathcal{O}(p^d)$. For example, using $p = 10$ adaptation rules having the same cost $\mathtt{c} = 1$, if $\tau_{\mathtt{penalty}} = 3$, the size of PCB is about a thousand times the size of CB. This illustrates the fact that this global enrichment of the case base approach produces a case base whose size is, for most CBR applications, too large, which motivates the local enrichment of the case base.

5.2 Practical view: local enrichment of CB

The principle of local enrichment is based first on the choice of *seed cases*, i.e., cases from CB that are chosen to produce penalized cases to be added to the case base. If $(\mathtt{x}^s, \mathtt{y}^s)$ is a seed case, then a penalized case $(\mathtt{x}, \mathtt{y}, \pi)$ is produced by a gradient descent starting from $(\mathtt{x}^s, \mathtt{y}^s, 0)$, by decreasing $\mathtt{dist}(\mathtt{x}, \mathtt{x}^t) + \pi$, each step corresponding to the application of an adaptation rule. The penalized case $(\mathtt{x}, \mathtt{y}, \pi)$ to be added to the case base thus constitutes a local optimum of the set of cases generated from the seed case $(\mathtt{x}^s, \mathtt{y}^s)$.

The selection of the set of seed cases SC can be done following several strategies, such as the following ones:

- The simplest strategy consists in taking all the source cases: SC = CB. This has the advantage of simplicity, but may lead to an important growth of the case base (the enriched case base size, $|\mathtt{PCB}|$, will be between $|\mathtt{CB}|$ and $2|\mathtt{CB}|$).
- If the size of the case base is too large already, only a few additional cases should be added and the following solutions can be proposed:
  - Choose SC by a random sampling from CB;
  - Choose SC as the k nearest neighbors of $\mathtt{x}^t$, e.g. according to dist.

6 Applications

In the sequel of the paper, we present two applications of local enrichment of the case base during case-based reasoning. Both examples create strings in a second domain (the solution space) that correspond to strings in a first domain (the problem space). The first application is a theoretical example: the problem space and the solution space are formal languages. The second application is actual machine translation: the problem space and the solution space are actual natural languages: French and English.

6.1 Machine translation of formal languages

The first example shows how local enrichment of the case base can be used to translate from a regular language into a context-free language. The languages we use are the prototypical examples of these families of formal languages, i.e., the problem space and the solution space are $\mathcal{P} = \{(ab)^n \mid n \in \mathbb{N}\}$ and $\mathcal{S} = \{A^n B^n \mid n \in \mathbb{N}\}$, respectively. Let us suppose that our case base contains only the three smallest nonempty members of each of these languages: $\texttt{CB} = \{(ab, AB), (abab, AABB), (ababab, AAABBB)\}$.

From such a case base, in the problem space, one variation with a support greater than 1 is extracted. It corresponds to the ratio $ab : abab = abab : ababab$.[7] This variation in the problem space corresponds to a variation in the solution space: $AB : AABB = AABB : AAABBB$. Of course, these variations have their corresponding inverse variations in the problem and solution spaces.

An actual trace of the system is given in Figure 1 for the translation of the string $(ab)^6$. We choose to select all cases in the case base as seed cases, i.e., $\texttt{SC} = \texttt{CB}$. The seed cases are sorted by distance to the target problem. Their LCS distance to the target problem is given by δ in Figure 1. Starting from the problems in the seed cases, applying the variation has the effect of enriching the case base with cases of the form $((ab)^n, A^n B^n)$ from $n = 4$ to 6, one after another. This is indeed an induction over n for $(ab)^n$ and $A^n B^n$ simultaneously in both spaces.

During enrichment, the distance from the new source problems to the target problem decreases down to 2. The distance of 0 is not mentioned, as it means that the new source problem is indeed the target problem, for which a solution has been found.

Such an example can be easily amended to translate from a regular language into a context-sensitive language (like $\{A^n B^n C^n \mid n \in \mathbb{N}\}$), or a context-free language into a context-sensitive language. Changing the direction of translation is also possible: from context-free to regular, etc.

6.2 Machine translation of natural languages

The second application deals with machine translation of natural languages. We use French–English as the language pair and data from the Tatoeba Corpus[8] as our bilingual corpus. There are important remarks to make on this domain.

Nature of the data: strings of characters. This implies again that variations are defined as in Section 4.4. The case base consists of sentence pairs which are in translation relation. We retain sentences of less than 10 words in length and select 90 % of them for training and the other 10 % for testing. This makes 109,390 sentence pairs in total in the training set. The average length of a sentence in French is 6.9 ± 1.8 words and 6.6 ± 1.6 in English. Such sentence pairs are illustrated in Figure 2. Notice that the sentences are lowercased and tokenized.

[7] Note that this is the equality of two ratios. Of course, it is also an analogy by itself ($ab : abab :: abab : ababab$), but this is not what is meant here.

[8] `https://tatoeba.org/` and `http://www.manythings.org/anki/`

PROBLEM SPACE	SOLUTION SPACE	CASE BASE	
		Initial case base = Seed cases	
		$(ab)^1$	A^1B^1
		$(ab)^2$	A^2B^2
		$(ab)^3$	A^3B^3
Source problems in seed cases			
$\delta = 6$ $(ab)^3$			
$\delta = 8$ $(ab)^2$			
$\delta = 10$ $(ab)^1$		Enrichment	
$(ab)^1 : (ab)^2 :: (ab)^3 : \mathbf{(ab)^4}$	$A^1B^1 : A^2B^2 :: A^3B^3 : \mathbf{A^4B^4}$	$\mathbf{(ab)^4}$	$\mathbf{A^4B^4}$
$(ab)^2 : (ab)^3 :: (ab)^3 : \mathbf{(ab)^4}$	$A^2B^2 : A^3B^3 :: A^3B^3 : \mathbf{A^4B^4}$	$\mathbf{(ab)^4}$	$\mathbf{A^4B^4}$
$(ab)^1 : (ab)^2 :: (ab)^2 : (ab)^3$	$A^1B^1 : A^2B^2 :: A^2B^2 : A^3B^3$	$(ab)^3$	A^3B^3
$(ab)^2 : (ab)^3 :: (ab)^2 : (ab)^3$	$A^2B^2 : A^3B^3 :: A^2B^2 : A^3B^3$	$(ab)^3$	A^3B^3
$(ab)^1 : (ab)^2 :: (ab)^1 : (ab)^2$	$A^1B^1 : A^2B^2 :: A^1B^1 : A^2B^2$	$(ab)^2$	A^2B^2
$(ab)^2 : (ab)^3 :: (ab)^1 : (ab)^2$	$A^2B^2 : A^3B^3 :: A^1B^1 : A^2B^2$	$(ab)^2$	A^2B^2
New source problem			
$\delta = 4$ $\mathbf{(ab)^4}$			
$(ab)^1 : (ab)^2 :: \mathbf{(ab)^4} : \mathbf{(ab)^5}$	$A^1B^1 : A^2B^2 :: \mathbf{A^4B^4} : \mathbf{A^5B^5}$	$\mathbf{(ab)^5}$	$\mathbf{A^5B^5}$
$(ab)^2 : (ab)^3 :: \mathbf{(ab)^4} : \mathbf{(ab)^5}$	$A^2B^2 : A^3B^3 :: \mathbf{A^4B^4} : \mathbf{A^5B^5}$	$\mathbf{(ab)^5}$	$\mathbf{A^5B^5}$
New source problem			
$\delta = 2$ $\mathbf{(ab)^5}$			
$(ab)^1 : (ab)^2 :: \mathbf{(ab)^5} : \mathbf{(ab)^6}$	$A^1B^1 : A^2B^2 :: \mathbf{A^5B^5} : \mathbf{A^6B^6}$	$\mathbf{(ab)^6}$	$\mathbf{A^6B^6}$
$(ab)^2 : (ab)^3 :: \mathbf{(ab)^5} : \mathbf{(ab)^6}$	$A^2B^2 : A^3B^3 :: \mathbf{A^5B^5} : \mathbf{A^6B^6}$	$\mathbf{(ab)^6}$	$\mathbf{A^6B^6}$
		Final case base	
		$(ab)^1$	A^1B^1
		$(ab)^2$	A^2B^2
		$(ab)^3$	A^3B^3
		$(ab)^4$	A^4B^4
		$(ab)^5$	A^5B^5
		$(ab)^6$	A^6B^6

Fig. 1. Trace for the translation from a regular language into a context-free language. The target problem is $(ab)^6$. It is correctly translated into A^6B^6 after enrichment of the case base. The problem space is on the left, the solution space in the middle. The case base and its enrichment are shown on the right. The distance to the target problem is denoted by δ. New source problems, solutions and cases are boldfaced; old ones are grayed out.

French	English
regardez comment je le fais !	*watch how i do it .*
vous m' avez oublié , n' est-ce pas ?	*you 've forgotten me , haven 't you ?*
je suis finlandais , mais je parle aussi suédois .	*i am finnish , but i speak also swedish .*
pensez-vous sérieusement à vendre cela sur internet ?	*are you seriously thinking about selling this online ?*
les travailleurs se plaignent de leurs conditions de travail .	*the workers are complaining about their working conditions .*

Fig. 2. French and English example sentences, i.e., problems and solutions, in the case base for case-based machine translation

get up . : *don 't get up .*	*he runs .* : *she runs .*	*i can 't believe this .* : *i just can 't believe this .*
tell me . : *don' t tell me .*	*he is young .* : *she is young .*	*i need a little more time .* : *i just need a little more time .*
leave us . : *don 't leave us .*	*he helps us .* : *she helps us .*	*i want this to be over .* : *i just want this to be over .*
leave me . : *don 't leave me .*	*he loves singing .* : *she loves singing .*	*i want someone to talk to .* : *i just want someone to talk to .*
	he just left . : *she just left .*	*i felt a drop of rain .* : *i just felt a drop of rain .*

Fig. 3. Three analogical clusters which stand for variations in the solution space

Size of the the case base. For case-based reasoning, the case base here is quite large: 109,390 cases. However, in the field of machine translation, on the contrary, it is considered rather small.

Nature of the ratios: they are defined as in Section 2.2 (analogies on strings) as we deal with strings. A ratio is a vector made of the difference between the Parikh vectors of the two strings considered, plus an extra dimension with the LCS distance between the two strings. Notice again that the equality between ratios does not imply the existence of an analogy, on the contrary to arithmetic or geometric analogies on numbers or tuples: $\texttt{dist}(A, B) = \texttt{dist}(C, D)$ does not imply $\texttt{dist}(A, C) = \texttt{dist}(B, D)$. Also, conformity is not transitive, so it is not an equivalence relation.

Nature of the variations: analogical clusters. Because of the nature of the ratios and the nature of conformity, variations are defined as groups of ratios among which an analogy is verified between any two pairs of ratios. We use

$\delta = 12$	*vous êtes vraiment très bons .*	*you are really very good .*	(in CB)
$\delta = 13$	*vous êtes venu trop tôt .*	*you 've come too early .*	"
$\delta = 14$	*vous êtes venu trop tard .*	*you came too late .*	"
$\delta = 14$	**vous êtes tatillon .**	**you 're finicky .**	"
$\delta = 14$	*vous êtes venue trop tôt .*	*you 've come too early .*	"
$\delta = 4$	*vous êtes vraiment très tatillon .*	*you are really very finicky .*	(new case)
$\delta = 18$	*ne vous vraiment pas êtes tatillon .*	*you really not 're finicky .*	"
$\delta = 6$	*vous êtes vraiment fort tatillon .*	*you are really very finicky .*	"
$\delta = 0$	*vous êtes vraiment trop tatillon .*	*you are really too finicky .*	"

Fig. 4. Translation process of a French sentence into English

vous êtes bons .	:	*vous êtes vraiment très bons .*	:: **vous êtes tatillon .**	:	*vous êtes vraiment très tatillon .*
you 're good .	:	*you are really very good .*	:: **you 're finicky .**	:	*you are really very finicky .*

Fig. 5. Variation in the problem and solution spaces resulting in the enrichment of the case base

the tools[9] described in [5] to extract all analogical clusters containing at least 2 ratios from the case base in the problem and the solution spaces.

Number of variations: It is rather large: almost 8 million analogical clusters were extracted in French, more than half a million in English. The extraction of such variations from an actual corpus is time-consuming. For efficiency, we retain only the first 3,000 largest analogical clusters in number of ratios. Three examples of analogical clusters are given in Figure 3. Typically, variations reflect grammatical oppositions. In the examples of Figure 3, affirmative / negative, masculine / feminine and insertion of the adverb *just*.

As an example, the translation process of the tokenized sentence *vous êtes vraiment trop tatillon .* from French into English is given in Figure 4. The seed problems are the 5 most similar French sentences in the case base. The translation process takes three steps corresponding to each block of sentences in the table below. The distance to the target problem is given on the left. Notice the infelicitous enrichment of the case base with an invalid sentence: *you really not 're finicky .*

An example of a variation applied in the problem and solution spaces during the above translation process is shown in Figure 5. It corresponds to a variation applied on the fourth seed case (boldfaced) in Figure 4. As a result, the case base is enriched with the first case marked as a new case in the above table ($\delta = 4$).

[9] `https://lepage-lab.ips.waseda.ac.jp/`

The standard metric BLEU [14] is used for the evaluation of the accuracy of a machine translation output against a given reference set. BLEU scores range from 0 to 1; the higher, the better. The system described above achieves a BLEU score of 0.51 in translating the 10,998 sentences in the test set. This is a reasonable score when compared with the scores of two much more elaborated systems, a neural system (OpenNMT[10]), and a statistical system (GIZA++, Moses, KenLM, MERT[11]), which achieve 0.60 and 0.65 respectively, on exactly the same data.

7 Conclusion, discussion and related work

In this paper, we proposed a new approach to case-based reasoning which consists in enriching the case base while performing reasoning. Enrichment results from the application of adaptation rules to seed cases, i.e., cases taken from the case base as starting points. New cases are created, which get closer and closer to the target problem, but they get penalties characteristic of the uncertainty brought by the application of the adaptation rules. Adaptation rules are given by variations in the problem space and variations in the solution space. The last variations should approach the target problem itself, so that corresponding variations in the solution space will produce (hypothetical) solutions to the target problem. We implemented such a new approach and illustrated it with two applications which shared the fact that the solution and problem spaces were spaces of strings of characters: formal and natural languages.

The general framework can be adapted to various scenarios. Several points can be adapted to the specificity of the domains at hand. For instance, the selection of the seed cases can be performed in various ways suggested in Section 5.2, at random or according to some selection method specific to the domain.

Our approach to case-based reasoning can be seen as a variant of gradient descent or hill climbing. Similarly, our approach exhibits the risk of reaching local minima (or maxima) instead of global minima (or maxima). Here, the landscape is shaped by the variations observed between the cases present in the initial case base. This issue of local optimality can be partially addressed by considering several branches generated from each seed case: instead of a single path approaching the target problem, a tree can be generated rooted at this seed case, whose breadth should be controlled to avoid an explosion of the enriched case base size. This way, several new cases can be generated from a single seed case. The precise study of this idea remains to be done.

From a theoretical viewpoint, the approach is presented in a very constrained framework, in particular for the definition of variations. This makes the explanations simpler, but, in particular for the considered applications in machine translation, some of these constraints do not hold. Thus, a theoretical study on less constrained variation spaces must be carried on.

[10] `http://opennmt.net`
[11] `http://www.statmt.org`

References

1. Aamodt, A., Plaza, E.: Case-based reasoning: Foundational issues, methodological variations, and system approaches. AI Commun. **7**(1), 39–59 (1994)
2. Couceiro, M., Hug, N., Prade, H., Richard, G.: Analogy-preserving functions: a way to extend Boolean samples. In: Proc. of the 26th Int. Joint Conf. on Artificial Intelligence (IJCAI'17). pp. 1575–1581. Morgan Kaufmann, Inc. (2017)
3. Craw, S., Wiratunga, N., Rowe, R.C.: Learning adaptation knowledge to improve case-based reasoning. Artificial Intelligence **170**(16-17), 1175–1192 (2006)
4. d'Aquin, M., Badra, F., Lafrogne, S., Lieber, J., Napoli, A., Szathmary, L.: Case Base Mining for Adaptation Knowledge Acquisition. In: Veloso, M.M. (ed.) Proceedings of the 20th International Joint Conference on Artificial Intelligence (IJCAI'07). pp. 750–755. Morgan Kaufman (2007)
5. Fam, R., Lepage, Y.: Tools for the production of analogical grids and a resource of n-gram analogical grids in 11 languages. In: LREC 2018. pp. 1060–1066. ELRA (May 2018)
6. Hanney, K., Keane, M.: Learning adaptation rules from a case-base. In: Smith, I., Faltings, B. (eds.) Advances in Case-based reasoning, pp. 179–192. No. 1168 in Lecture notes in AI, Springer Verlag, Berlin (1996)
7. Hussain, Z., Gimenez, F., Yi, D., Rubin, D.: Differential data augmentation techniques for medical imaging classification tasks. In: Annual symposium proceedings. pp. 979–984 (April 2017), `https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5977656/`
8. Jalali, V., Leake, D., Forouzandehmehr, N.: Learning and applying adaptation rules for categorical features: An ensemble approach. AI Communications **30**(3-4), 193–205 (2017)
9. Kim, E.J., Brunner, R.J.: Star-galaxy classification using deep convolutional neural networks. Monthly Notices of the Royal Astronomical Society p. stw2672 (2016)
10. Lepage, Y., Lieber, J.: Case-Based Translation: First Steps from a Knowledge-Light Approach Based on Analogy to a Knowledge-Intensive One. In: ICCBR 2018 - 26th International Conference on Case-Based Reasoning. Stockholm, Sweden (Jul 2018), `https://hal.inria.fr/hal-01906528`
11. Lepage, Y.: Proportional analogy in written language data. In: Gala, N., Rapp, R., Bel-Enguix, G. (eds.) Language, Production, Cognition and the Lexicon, pp. 151–173. Text, Speech and Language Technology 48, Springer International Publishing Switzerland (2014)
12. Lepage, Y., Denoual, E.: Purest ever example-based machine translation: detailed presentation and assessment. Machine Translation **19**, 251–282 (2005)
13. Lieber, J., Nauer, E., Prade, H., Richard, G.: Making the Best of Cases by Approximation, Interpolation and Extrapolation. In: ICCBR 2018 - 26th International Conference on Case-Based Reasoning. Stockholm, Sweden (Jul 2018), `https://hal.inria.fr/hal-01905058`
14. Papineni, K., Roukos, S., Ward, T., Zhu, W.J.: BLEU: a method for automatic evaluation of machine translation. Research report RC22176, IBM (Sep 2001)
15. Richter, M.M., Weber, R.: Case-based reasoning: a textbook. Springer Science & Business Media (2013)
16. Taylor, L., Nitschke, G.: Improving deep learning using generic data augmentation. CoRR **abs/1708.06020** (2017), `http://arxiv.org/abs/1708.06020`