EnergyCompress: A General Case Base Learning Strategy

Fadi Badra[1] , **Esteban Marquer**[2] , **Marie-Jeanne Lesot**[3] , **Miguel Couceiro**[4] and **David Leake**[5]
[1]Université Sorbonne Paris Nord, Sorbonne Université, INSERM, Limics, 93000, Bobigny, France
[2]CRIL CNRS, Université d'Artois, France
[3]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
[4]IST, University of Lisbon, INESC-ID, Lisbon, Portugal
[5]Luddy School, Indiana Unversity, Bloomington, IN, USA
badra@sorbonne-paris-nord.fr, esteban.marquer@gmail.com, Marie-Jeanne.Lesot@lip6.fr,
miguel.couceiro@inesc-id.pt, leake@iu.edu

Abstract

Case-based prediction (CBP) methods do not learn a model of the target decision function but instead perform an inference process that depends on two similarity measures and a reference case base. This paper proposes a strategy, called EnergyCompress, to learn an effective case base by selecting relevant cases from an initial set. Use of EnergyCompress decreases CBP inference time, through case base compression, and also increases prediction performance, for a wide variety of CBP algorithms. EnergyCompress relies on a general formulation of the CBP task in the framework of energy-based models, which leads to a new and valuable characterization of the notion of competence in case-based reasoning, in particular at the source case level. Extensive experimental results on 18 benchmarks comparing EnergyCompress to 5 reference algorithms for case base maintenance support the benefit of the proposed strategy.

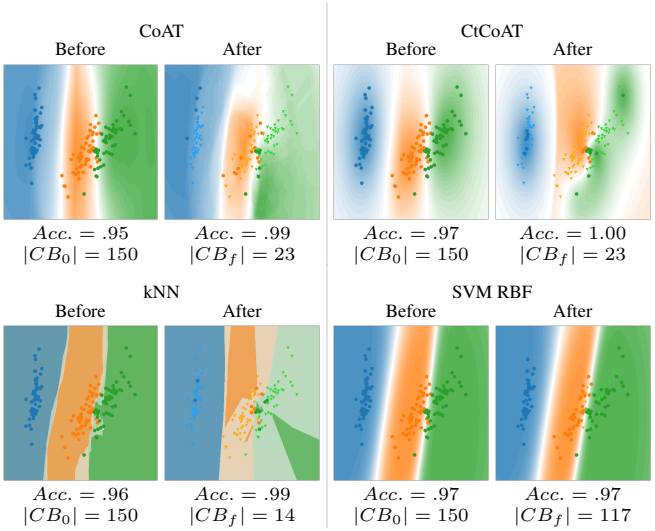

Figure 1: Sculpting the decision frontier by removing cases using EnergyCompress: improving both accuracy and inference time through case base compression (Iris dataset, 2D PCA projection).

1 Introduction

Case-based prediction (CBP) methods, such as the k nearest neighbor (k-NN) classifier, the AP-classifier [Bounhas *et al.*, 2017] or $CoAT$ [Badra, 2020], are kernel-based learning methods, insofar as they crucially depend on the choice of a similarity or distance function that takes a case base as a parameter. These methods do not learn a model of the entire decision function prior to prediction, but instead infer the decision function value for a new case by direct comparison with similar cases retrieved from the case base. This inference process depends on three parameters $\theta = (\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, CB)$: the similarity measures in the input (resp. output) space $\sigma_{\mathcal{S}}$, (resp. $\sigma_{\mathcal{R}}$), and the case base CB. The $\sigma_{\mathcal{R}}$ measure is usually fixed depending on the task at hand (e.g. classification or regression), but $\sigma_{\mathcal{S}}$ and CB have huge ranges of possibilities.

Traditionally, knowledge-based methods were used to derive similarity measures for a given case base [Kolodner and Leake, 1996]. In the machine learning community, the metric learning task [Ghojogh *et al.*, 2022] addresses a similar issue, although in a different context, and recent case-based reasoning (CBR) research applies machine learning methods to learning similarity [Mathisen *et al.*, 2019]. This paper proposes taking the reverse approach: rather than assuming the case base is given and generating a similarity measure for it, it introduces a method, EnergyCompress, that learns a case base by starting from a candidate case base and selecting an appropriate subset for a given similarity measure. This can be seen as relating to Richter's [2003] observation that CBR involves multiple *knowledge containers*, including similarity and case knowledge, interacting such that one can compensate for the other, *e.g.,* here, that a well-chosen case base can compensate for a suboptimal similarity measure.

The case base selection task has been studied extensively in case based reasoning research on case base maintenance by compression, which selects subsets of an original case base to retain (e.g. [Juarez *et al.*, 2018], see more details in Sec. 2). However, such work usually makes the implicit assumption that a "bigger is better" for prediction accuracy: research on case-base compression generally aims at other benefits, such

as increased efficiency. In contrast, EnergyCompress enables sculpting the decision frontier by removing cases to increase accuracy, in addition to compressing to reduce the inference time. Thus, in contrast to most existing approaches, it does not sacrifice one for the other. This is illustrated in Fig. 1, whick shows the decision regions obtained for the Iris dataset, on a 2D PCA projection, for three CBP algorithms (CoAT, CtCoAT and 3-NN, see their descriptions in Sec. 4, based on the Euclidean distance) and SVM with a radial basis function (width $= 6.4 \times 10^{-2}$). Both rows show the prediction regions and their associated accuracy computed on the whole case base, CB_0. For each CBP algorithm, the left figure shows the results when the inference algorithm uses the whole case base CB_0 while, on the right figure, it only uses CB_f, extracted from CB_0 by EnergyCompress. Here EnergyCompress leads to a different decision frontier that achieves a higher accuracy on CB_0, for all 4 algorithms, with smaller case bases CB_f, (size reduction up to 90% for the 3-NN) and thus decreases inference time.

The second main characteristic of EnergyCompress is that it is a general method: it can be applied to virtually any CBP algorithm, provided that the algorithm assigns a probability[1] to each potential outcome, to provide a tailored case base. Defining a single case base learning strategy applicable to different prediction algorithms remedies one of the main limitations of existing case base learning algorithms, which are dedicated to specific approaches (often to k-NN, see Sec. 2, or to CoAT [Marquer *et al.*, 2023]) .

EnergyCompress relies on an original formulation of the CBP task in the framework of energy-based models [LeCun *et al.*, 2006]. This formulation offers a new and valuable characterization of the notion of competence, in particular at the source case level. The case base learning method then consists of removing cases from a candidate set so as to learn the underlying energy function. This constitutes a generalization of the approach proposed by [Marquer *et al.*, 2023] for the specific case of the CoAT algorithm.

The contributions of the paper thus include a general case base learning strategy, that can be applied to a large variety of case-based prediction algorithms and that enables both decreasing inference time and increasing prediction performance, without sacrificing one for the other. Extensive experiments run on 18 UCI datasets demonstrate that EnergyCompress obtains very competitive compression ratios while leading to an accuracy (given in %) increase on average of $+7.4$ for k-NN, $+13.3$ for CoAT, $+26.5$ for CtCoAT, and $+3.7$ for SVM RBF. Although we focus in this paper on the selection of instances to include in the case base, the experimental results also support that the indicators we define lead to a valuable characterization of the competence notion (see Sec. 3.2 for a discussion).

The paper is organized as follows. Sec. 2 discusses related work, both in the domains of case-based prediction and energy-based machine learning. Sec. 3 describes the proposed EnergyCompress strategy, discussing the energy-based formulation of CBP and presenting the induced general competence model as well as the proposed case base learning strategy. Sec. 4 shows how this general case base learning strategy can be implemented for different CBP algorithms. Sec. 5 presents the experimental study of the approach. Finally, Sec. 6 concludes the paper and discusses future work.

[1]If a CBP algorithm assigns a support to each potential outcome instead of a probability, a probability can be computed as the ratio of the support for each outcome compared to the total support.

2 Related Work

This section discusses the position of the EnergyCompress method with respect to related works, both with respect to case-based prediction and energy-based learning.

Case-based prediction. Case-based reasoning (CBR) relates to the cognitive ability of analogical reasoning and models a special kind of plausible inference principle, according to which if two situations are judged similar, then it is plausible that they will also have similar outcomes, described in CBR by the phrase "similar problems have similar solutions" [Anthony and Ratsaby, 2015; Leake and Wilson, 1999, inter alia]. Case-based prediction (CBP) algorithms implement such a transfer step for prediction tasks, which corresponds to applying analogical reasoning to classification or regression tasks in machine learning. In CBP methods, analogical transfer is used to complete the description of a new case by direct comparison with similar cases. A recent review of the literature on case-based prediction [Badra and Lesot, 2023] shows that all surveyed implementations of the inference process share a common principle: they optimize a transfer of similarity knowledge, from the situation space to the outcome space. The predicted outcome is the one that optimizes a measure of compatibility between two similarity measures on a case base, following the plausible inference principle stated above. The compatibility measure can take the form of a joint similarity measure, a set of continuity constraints, a set of rules or a global indicator. In this paper, we focus on the approaches that optimize a global indicator computed on the whole case base, showing in Sec. 3.1 that their inference process can then be interpreted as an energy-based inference.

The notations used throughout the paper are as follows: $\mathcal{S}$ denotes the input space, and $\mathcal{R}$ the output space, respectively equipped with the similarity measures $\sigma_{\mathcal{S}}$ and $\sigma_{\mathcal{R}}$. An element of $\mathcal{S}$ is called a *situation*, and an element of $\mathcal{R}$ an *outcome*, or a result. A set $CB = \{(s_1, r_1), \ldots, (s_n, r_n)\}$ of elements in $\mathcal{S} \times \mathcal{R}$ is called a *case base*. An element $c_i = (s_i, r_i) \in CB$ is called a *source case*. Let $\mathcal{T}_{ref} \subset \mathcal{S} \times \mathcal{R}$ be a set of cases called a *reference set*, and $c_t = (s_t, r_t) \in \mathcal{T}_{ref}$ be a reference case. A *potential case* $(s_t, \hat{r})$ can be constructed from a reference case by associating to the same situation $s_t \in \mathcal{S}$, a different outcome $\hat{r} \in \mathcal{R}, \hat{r} \neq r_t$. Let us denote by $Pred_\theta$ a CBP algorithm and $\theta = (\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, CB)$ its parameters. The predicted outcome $r^* = Pred_\theta(s_t)$ for a new situation s_t is the one that makes the new similarity relations $\sigma_{\mathcal{R}}(r_i, r^*)$ most compatible with the new similarity relations $\sigma_{\mathcal{S}}(s_i, s_t)$.

Case base maintenance. Research in case-base maintenance has a long history (see [Juarez *et al.*, 2018] for a survey). The origins date back to early work on case selection in the context of the condensed nearest neighbor al-

gorithm (CNNR [Hart, 1968]), which aimed at selecting a subset of the initial set of cases still sufficient for k-NN to generate a correct categorization. This process is commonly referred to as compressing the case base. Focuses of case-base maintenance include increasing prediction accuracy, *e.g.,* by removing noisy cases, and removing unnecessary cases to increase the efficiency of source case retrieval [Cummins, 2013]. Much research concerns *competence models* and competence-based deletion, which prioritizes removal of cases that make the least contribution to overall coverage, considering both their coverage (cases for which they can provide a solution) and reachability (whether they could be solved using other cases) [Smyth and McKenna, 2001]. The aim is to compress the case base while minimizing competence loss; the "gold standard" for performance is the full case base. These methods depend on the "representativeness assumption" that existing cases are a good proxy for future cases, to assess their contributions; the proposed EnergyCompress method can accommodate this assumption, but does not depend on it[2]. Besides pioneering algorithms such as RENN [Tomek, 1976] or IB3 [Aha *et al.*, 1991], a recent literature review on instance selection for automatic text classification [Cunha *et al.*, 2023] shows that recent density-based algorithms such as LSSm [Leyva *et al.*, 2015] or XLDIS [Carbonera, 2017] are also effective in retaining accuracy while reducing the training set size and the training process time by discarding noisy or redundant instances.

Relationship to the knowledge containers perspective in CBR. The knowledge of CBR systems is often viewed in terms of a set of four "knowledge containers" [Richter, 2003]: the case base, the case vocabulary,[3] similarity knowledge and case adaptation knowledge, with strengths in one able to compensate for weaknesses in others. For example, having a better case base, with better coverage, may compensate for weak case adaptation knowledge. However, the value of knowledge cannot be judged in isolation; the harmonization of components such as similarity and adaptation knowledge strongly affects performance [Leake and Ye, 2021]. Much research focuses on acquiring similarity knowledge for a given case base (*e.g.,*[Mathisen *et al.*, 2019]). The work in this paper investigates the contrasting approach of starting from a given similarity measure and selecting a good case base for it.

Energy-based models. This section briefly summarizes the basics of the energy-based framework the paper uses to provide an original solution to the above-mentioned case-base maintenance issues. Inspired by statistical physics, energy-based models [LeCun *et al.*, 2006] specify a probability distribution $p(x;\theta) = e^{-E_\theta(x)} / \int e^{-E_\theta(x)} dx$ via a parameterized scalar-valued function $E_\theta(x)$ called an *energy function*. In its conditional version, the function $E_\theta : \mathcal{X} \times \mathcal{Y} \longrightarrow \mathbb{R}$ associates each pair $(x, y) \in \mathcal{X} \times \mathcal{Y}$ with a scalar value $E_\theta(x, y)$ that represents the compatibility between the input x and the output y under the set of parameters θ. The energy function E_θ takes low values when y is compatible with x, and higher values when y and x are less compatible. The goal of the energy-based *inference* is to find, among a set of outputs $\mathcal{Y}$, the output $y^* \in \mathcal{Y}$ that minimizes the value of the energy function: $y^* = \arg\min_{y \in \mathcal{Y}} E_\theta(x, y)$. Given a family of energy functions $E_\theta(x, y)$ indexed by a set of parameters θ, the goal of *learning* the energy function is to optimize the θ parameters in order to "push down" (*i.e.*, assign lower energy values to) the points on the energy surface that are around the training samples, and to "pull up" all other points. Contrastive divergence [Hinton *et al.*, 2006] is a common learning strategy that, given a numerical hyperparameter λ, optimizes a contrastive loss function such as the hinge loss, which is defined, for a training sample (x_k, y_k) and a generated out of distribution sample $(x_k, \hat{y})$ by: $\ell(\theta, x_k, y_k, \hat{y}) = \max(0, \lambda + E_\theta(x_k, y_k) - E_\theta(x_k, \hat{y}))$. The hinge loss associates a loss value to a training sample (x_k, y_k) whenever its energy is not lower by at least a margin λ than the energy of the incorrect sample $(x_k, \hat{y})$. In CBR terms, the inputs $\mathcal{X}$ correspond to the situations $\mathcal{S}$, and the outputs $\mathcal{Y}$ to the outcomes $\mathcal{R}$. If we associate CBP algorithms with an energy function E_θ (we explain how in Sec. 4), the hinge loss associated with a case (s_k, r_k) and E_θ can be measured. This serves as the foundation for the competence measure we propose in Sec. 3.2.

3 EnergyCompress

In order to solve the problem of learning an appropriate case base for a given CBP algorithm with chosen similarity measures $\sigma_\mathcal{S}, \sigma_\mathcal{R}$, the proposed EnergyCompress strategy relies on an energy-based model of CBP. The general model is described in Sec. 3.1 below, its implementation for four reference specific algorithms is detailed in Sec. 4. This section then describes the induced competence model, which leads to the proposed case base learning strategy.

3.1 An Energy-Based Model

A core idea proposed in [Marquer *et al.*, 2023] is that an inference process that optimizes a global indicator of compatibility between two similarity measures on a case base can be interpreted as an energy-based inference.

The input space $\mathcal{X}$ is the situation space $\mathcal{S}$ and the output space $\mathcal{Y}$ is the outcome space $\mathcal{R}$. The energy function $E_\theta^{Pred} : \mathcal{S} \times \mathcal{R} \longrightarrow \mathbb{R}$ measures, for $Pred_\theta$, the compatibility of the outcome similarities with the added situation similarities when a potential new case $\hat{c}_t = (s_t, \hat{r})$ is added to the case base. We propose to extend this definition to compute the energy of a set $A \subset \mathcal{S} \times \mathcal{R}$ of cases by taking

$$E_\theta^{Pred}(A) = \sum_{c_t = (s_t, r_t) \in A} E_\theta^{Pred}(s_t, r_t).$$

The goal of the energy-based inference is to find, among the set of potential outcomes $\hat{r} \in \mathcal{R}$, the outcome that minimizes the energy function, *i.e.*,

$$r^* = Pred_\theta(s_t) = \arg\min_{\hat{r} \in \mathcal{R}} E_\theta^{Pred}(s_t, \hat{r}). \quad (1)$$

[2] In Sec. 3.2 we define the competence used by EnergyCompress, measured on a set $\mathcal{T}_{ref}$ of cases distinct from the case base CB. Under the representativeness assumption, we can have $\mathcal{T}_{ref} = CB$ as CB is representative of the distribution of cases.

[3] Richter defines the *vocabulary* of "a knowledge representation system [as] which data structures and which elements of the structures are used to represent primitive notions." Generally (and here) the attribute-value representation is used, typically in tabular form.

The energy function E_θ^{Pred} is learned by minimizing on a reference set $\mathcal{T}_{ref}$ the loss $L = \sum_{c_t=(s_t,r_t)\in\mathcal{T}_{ref}} \ell(\theta, c_t)$, where ℓ is defined by:

$$\ell(\theta, c_t) = \max(0, \lambda + E_\theta^{Pred}(s_t, r_t) - \min_{\hat{r}\neq r_t} E_\theta^{Pred}(s_t, \hat{r})).$$

3.2 A General Competence Model

The notions of competence of a case or of a case base, and of the influence of a case on the prediction, can be captured in an energy-based framework. Let us denote by $D \subseteq \mathcal{S} \times \mathcal{R}$ a set of instances, and $(x, y) \in D$ an instance of D.

Let $Pred_\theta$ with $\theta = (\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB)$ be a CBP algorithm that can be expressed as an energy-based inference, *i.e.*, that satisfies Eq. 1 for some energy function E_θ^{Pred}. The competence of a case base CB is measured w. r. t. a given reference set $\mathcal{T}_{ref}$ by taking the (opposite of) the hinge loss on $\mathcal{T}_{ref}$:

$$C(\theta, \mathcal{T}_{ref}) = -\frac{1}{|\mathcal{T}_{ref}|} \sum_{c_t\in\mathcal{T}_{ref}} \ell(\theta, c_t). \tag{2}$$

The competence of a source case $c = (s, r) \in CB$ w. r. t. a reference set $\mathcal{T}_{ref}$ can be defined as the loss of competence that would happen if this source case was deleted from CB:

$$C(c, \theta, \mathcal{T}_{ref}) = C(\theta, \mathcal{T}_{ref}) - C((\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB \setminus \{c\}), \mathcal{T}_{ref}).$$

This definition also allows defining a notion of competence locally as the contribution of c on each individual reference case $c_t \in \mathcal{T}_{ref}$ of the reference set:

$$influence_\theta(c, c_t) = \ell(\theta, c_t) - \ell((\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB \setminus \{c\}), c_t).$$

This notion of influence of a case on the prediction can be seen as a continuous counterpart of the popular notion of case coverage introduced in [Smyth and McKenna, 2001] for k-NN. The influence of a case is computed here with respect to a reference set $\mathcal{T}_{ref}$, and not to the case base CB. The case influences are aggregated to measure the competence of a case base on the whole reference set, as well as the competence of each source case. In this paper, we focus on the application of this competence model to learn the most competent case base for a given prediction algorithm $Pred_\theta$.

3.3 The Case Base Learning Strategy

The case base learning strategy consists in iteratively deleting the least competent source case from a candidate set CB_0. As the competence of a case base is defined directly from the loss function of the underlying energy-based model (Eq. 2), selecting the most competent case base enables learning the energy function E_θ^{Pred} by optimizing the CB parameter. The EnergyCompress case deletion procedure is described by Algorithm 1: at each iteration, the source case that contributes least to the competence of the case base CB w.r.t. the reference set $\mathcal{T}_{ref}$ is deleted from the case base (line 11). In this algorithm, the returned parameter $\theta_f = (\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB_f)$ is the one that maximizes accuracy (function Acc, line 6) on the reference set $\mathcal{T}_{ref}$. The computational complexity of EnergyCompress is in $O(|CB| \times |\mathcal{T}_{ref}| \times |\mathcal{R}|) \times O(E)$, where E is the computational complexity of the energy computation.

Algorithm 1 EnergyCompress case deletion procedure. $C(c, \theta, \mathcal{T}_{ref})$ depends on $Pred$ and the associated E_θ^{Pred}.

1: **function** ENERGY_COMPRESS($\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB_0, \mathcal{T}_{ref}$)
2: $CB = CB_0$
3: $max\_acc = -1$
4: **repeat**
5: $\theta = (\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB)$
6: $acc = \text{Acc}(Pred_\theta(\mathcal{T}_{ref}))$
7: **if** $acc \geq max\_acc$ **then**
8: $max\_acc = acc$
9: $\theta_f = \theta$
10: **end if**
11: $CB = CB \setminus \{\arg\min_{c\in CB}\ C(c, \theta, \mathcal{T}_{ref})\}$
12: **until** $|CB| = 0$
13: **return** θ_f
14: **end function**

4 Application to Various CBP Algorithms

For EnergyCompress to be applicable to a given CBP algorithm $Pred_\theta$, the only requirement is that $Pred_\theta$ can be expressed as an energy-based inference, i.e., as an algorithm that predicts the outcome that minimizes an energy function E_θ^{Pred} (Eq. 1). We show in this section that each of the four CBP algorithms CoAT, CtCoAT, k-NN, and the AP-Classifier can be expressed as minimizing an energy function. Tab. 1 gives examples of such energy functions.

CoAT. The CoAT case-based prediction algorithm was introduced in [Badra, 2020] and formulated in an energy-based model in [Marquer *et al.*, 2023]. Its energy function E_θ^{CoAT}, reported in Tab. 1, simply counts the number of new triplets that are not ordered in the same way by $\sigma_\mathcal{S}$ and by $\sigma_\mathcal{R}$.

CtCoAT. The CtCoAT algorithm is a variant of the CoAT case-based prediction algorithm that we introduce in this paper. Its energy function E_θ^{CtCoAT} is a continuous version of the CoAT energy function where in a classification setting, each positive triplet (*i.e.*, a triplet (i, j, k) such that $\sigma_\mathcal{R}(r_i,\ r_j) = \sigma_\mathcal{R}(r_i, r_k)$) contributes $\frac{1}{2}$ and each negative triplet (such that $\sigma_\mathcal{R}(r_i,\ r_j) \neq \sigma_\mathcal{R}(r_i, r_k)$) contributes to a value between 0 (easy negatives) and 1 (hard negatives). When $\sigma_\mathcal{R}(r_i,\ r_j) < \sigma_\mathcal{R}(r_i, r_k)$, if $\sigma_\mathcal{S}(s_i, s_j) \geq \sigma_\mathcal{S}(s_i, s_k)$ (hard negative), then $\frac{1}{2}(1 - [\sigma_\mathcal{S}(s_i, s_j) - \sigma_\mathcal{S}(s_i, s_k)][\sigma_\mathcal{R}(r_i, r_j) - \sigma_\mathcal{R}(r_i, r_k)]) \in [\frac{1}{2}, 1]$ and the triplet greatly increases the energy, as in E_θ^{CoAT} under the same conditions. If $\sigma_\mathcal{S}(s_i, s_j) < \sigma_\mathcal{S}(s_i, s_k)$ then the contribution of (i, j, k) is in $[0, \frac{1}{2}[$ (easy negative), and the triplet has a low contribution to the energy.

k nearest neighbors (k-NN) alg. The k-NN decision function can be written as $r^* = \arg\max_{\hat{r}\in\mathcal{R}} \chi_\theta(s_t, \hat{r})$, where

$$\chi_\theta(s_t, \hat{r}) = \sum_{(s_i,r_i)\in CB} \mathbb{1}^k_{\sigma_\mathcal{S}(s_t,\cdot)}(s_i) \times \sigma_\mathcal{R}(r_i, \hat{r})$$

is a joint similarity measure that measures the compatibility of $\sigma_\mathcal{S}$ with $\sigma_\mathcal{R}$ on CB. In this expression, $\mathbb{1}^k_{\sigma_\mathcal{S}(s_i,\cdot)}(s)$ returns 1 if s belongs to the set of k nearest neighbors of s_t. The energy function E_θ^{kNN} measures the incompatibility of $\sigma_\mathcal{S}$

Prediction algorithm ($Pred_\theta$)	**Energy function** $E_\theta^{Pred}(s_t, \hat{r})$
CoAT	$E_\theta^{CoAT}(s_t,\hat{r}) = \Gamma(\sigma_\mathcal{S},\sigma_\mathcal{R},CB\cup\{(s_t,\hat{r})\}) - \Gamma(\sigma_\mathcal{S},\sigma_\mathcal{R},CB)$, where $\Gamma(\sigma_\mathcal{S},\sigma_\mathcal{R},CB) = \mid \{((s_0,r_0),(s_i,r_i),(s_j,r_j)) \in CB^3$ such that $\sigma_\mathcal{S}(s_0,s_i) \geq \sigma_\mathcal{S}(s_0,s_j)$ and $\sigma_\mathcal{R}(r_0,\ r_i) < \sigma_\mathcal{R}(r_0,r_j)\} \mid$
CtCoAT	$E_\theta^{CtCoAT}(s_t,\hat{r}) = \Gamma_{Ct}(\sigma_\mathcal{S},\sigma_\mathcal{R},CB\cup\{(s_t,\hat{r})\}) - \Gamma_{Ct}(\sigma_\mathcal{S},\sigma_\mathcal{R},CB)$, where $\Gamma_{Ct}(\sigma_\mathcal{S},\sigma_\mathcal{R},CB) = \frac{1}{2}\sum_{i,j,k} 1 - [\sigma_\mathcal{S}(s_i,s_j) - \sigma_\mathcal{S}(s_i,s_k)][\sigma_\mathcal{R}(r_i,r_j) - \sigma_\mathcal{R}(r_i,r_k)]$
k-NN	$E_\theta^{kNN}(s_t,\hat{r}) = 1 - \frac{1}{\lvert CB\rvert}\sum_{(s_i,r_i)\in CB} \mathbb{1}^k_{\sigma_\mathcal{S}(s_t,\cdot)}(s_i) \times \sigma_\mathcal{R}(r_i,\hat{r})$
AP-Classifier	$E_\theta^{AP}(s_t,\hat{r}) = 1 - \frac{1}{\lvert CB\rvert}\sum_{(s_i,r_i)\in CB} \sigma_\mathcal{S}(s_t,s_i) \times \sigma_\mathcal{R}(r_i,\hat{r})$
Probabilistic model: $\arg\max_{\hat{r}\in\mathcal{R}} P(r\mid s_t)$	$E_\theta^{P}(s_t,\hat{r}) = 1 - P(\hat{r}\mid s_t)$

Table 1: Examples of energy functions that can be learned by EnergyCompress.

with $\sigma_\mathcal{R}$ on CB according to χ_θ: its value is 1 if $\sigma_\mathcal{S}$ and $\sigma_\mathcal{R}$ are incompatible, and then decreases as χ_θ increases.

Analogical proportion-based classifiers. Analogical proportion-based classifiers (see e.g. [Bounhas and Prade, 2024; Couceiro *et al.*, 2017; Couceiro *et al.*, 2018]) can be modeled as special CBP algorithms that reason on similar *differences* between instances [Badra and Lesot, 2022]. The case base $CB = \{c_i = (s_i, r_i)\}$ is constructed from all pairs $(\mathbf{a}, \mathbf{b})$ of instances of a set of instances D, *i.e.*, $c_i = (s_i, r_i) = (\mathbf{a} - \mathbf{b}, f(\mathbf{a}) - f(\mathbf{b}))$ where $f(a)$ is the outcome of a. A new instance $(x, \hat{y})$ added to D leads to a set of $|D|$ potential new cases $\hat{c}_t = (s_t, \hat{r})$ with $\hat{c}_t = (s_t, \hat{r}) = (\mathbf{c} - x, f(\mathbf{c}) - \hat{y})$, formed by taking all possible triples $(\mathbf{a}, \mathbf{b}, \mathbf{c})$ from D. The similarity measures $\sigma_\mathcal{S}$ and $\sigma_\mathcal{R}$ are chosen such that $\sigma_\mathcal{S}(\mathbf{a} - \mathbf{b}, \mathbf{c} - x) = 1$ iff $\mathbf{a} : \mathbf{b} :: \mathbf{c} : x$ holds, and $\sigma_\mathcal{R}(f(\mathbf{a}) - f(\mathbf{b}), f(\mathbf{c}) - \hat{y}) = 1$ iff $f(\mathbf{a}) : f(\mathbf{b}) :: f(\mathbf{c}) : \hat{y}$ holds. The transfer strategy is rule-based, and consists in predicting for a new x the value y^* that maximizes the number of times that the rule $(\sigma_\mathcal{S} \approx 1) \rightarrow (\sigma_\mathcal{R} = 1)$ can be triggered to derive y^*. This means that the prediction y^* is the one that maximizes a compatibility measure $\chi_\theta(s_t, \hat{r})$ that counts for each potential outcome $\hat{r}$ the number of new pairs of cases $(c_i, \hat{c}_t)$ for which $\sigma_\mathcal{R}(r_i, \hat{r}) = 1$ and $\sigma_\mathcal{S}(s_i, s_t) \approx 1$. The AP-Classifier [Bounhas *et al.*, 2017] implements this transfer strategy by adding up, for each potential new case $\hat{c}_t = (s_t, \hat{r})$, the similarity values $\sigma_\mathcal{S}(s_i, s_t)$ for all source cases $c_i = (s_i, r_i)$ such that $\sigma_\mathcal{R}(r_i, \hat{r}) = 1$. The obtained energy function E_θ^{AP} resembles the one defined for k-NN (see Tab. 1), but needs to be aggregated on the $|D|$ new cases $\hat{c}_t$ resulting from the addition of the instance $(x, \hat{y})$. The decision function of the AP-Classifier can be written as:

$$y^* = \arg\min_{\hat{y}} E_\theta^{AP}(\{\hat{c}_t = (\mathbf{c} - x, f(\mathbf{c}) - \hat{y}), \mathbf{c} \in D\}).$$

Although these algorithms also maximize a global compatibility measure χ_θ and can be interpreted as implementing an energy-based inference, they are not included in the following experiments, due to their high computational cost.

Extension to any probability-based machine learning model. At a very general level, predictive models that provide probabilities for each possible outcome can be interpreted as energy-based models. For our purposes, given a conditional probability model $P(\hat{r}|s)$ that predicts $r^* = \arg\max_{\hat{r}\in\mathcal{R}} P(\hat{r}|s_t)$, we define the energy function of the model as $E_\theta^P(s_t, \hat{r}) = 1 - P(\hat{r}|s_t)$. This definition follows the intuitions of energy-based models: more desirable outcomes are identified by higher probabilities and lower energies.

Support Vector Machines (SVMs). SVMs can be seen as a specific case of the previous probability-based machine learning model. To do so, their posterior can be approximated by Platt scaling [Platt, 1999].

5 Experiments

This section describes the experiments run to validate that (*i*) EnergyCompress substantially increases the performance of CBP algorithms for a fixed similarity measure $\sigma_\mathcal{S}$, (*ii*) the obtained compression ratio is competitive w.r.t. the state of the art, (*iii*) the method is general, in that it both allows capturing competence for a variety of CBP algorithms and learning a case base tailored for each prediction algorithm[4]

5.1 Protocol

The case base learning methods CNNR, ENN, XLDIS, LSSm, IB3, and EnergyCompress were tested on 18 UCI datasets against the prediction algorithms CoAT, CtCoAT, k-NN (with $k = 7$), as well as SVMs with linear, polynomial, and radial basis function kernels.

The 18 datasets, listed in Tab. 2, are attribute-value datasets with 3 to 56 attributes and 2 to 7 classes, used in a classification setting. For each classification task, the initial parameters $\theta_0 = (\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB_0)$ are chosen as follows. The outcome space $\mathcal{R}$ is the set of class labels r_i, and $\sigma_\mathcal{R}$ is the class membership similarity measure, such that $\sigma_\mathcal{R}(r_i, r_j) = 1$ if $r_i = r_j$, and 0 otherwise. For CoAT, CtCoAT, and k-NN, the

[4]Code for reproducing the experiments is available at: https://github.com/EMarquer/MeATCube/tree/maintenance_benchmark

Dataset	# instances	# attributes	# classes
Balance Scale	625	4	3
Br. Can. Wisc. Diag.	569	30	2
Br. Can. Wisc. Prog.	198	33	2
Credit Approval	690	15	2
Dermatology	366	34	6
Glass Identification	214	9	7
Haberman's Survival	306	3	2
Heart Disease	303	14	5
Hepatitis	155	19	2
Ionosphere	351	33	2
Iris	150	4	3
Pima	768	8	2
Teach. Ass. Eval.	151	5	3
Lenses	24	4	3
Liver Disorders	345	6	2
Lung Cancer	32	56	3
Wine	178	13	3
Zoo	101	16	7

Table 2: The 18 UCI datasets used in the experiments.

similarity measure $\sigma_{\mathcal{S}}$ is chosen to be the decreasing function $\sigma_{\mathcal{S}}(s_i, s_j) = e^{-d(s_i,s_j)}$ of the Euclidean distance d. The SVMs use the implementation and default parameters from scikit-learn[5] to fit the kernel and to estimate the probabilities.

All pairs $(Learn, Pred)$ are formed, where $Learn \in \{$CNNR, ENN, $\ldots\}$ is a case base learning method and $Pred \in \{$CoAT, CtCoAT,$\ldots\}$ is a CBP algorithm. For each pair $(Learn, Pred)$, the learning algorithm $Learn$ is applied to learn CB_f from the candidate base CB_0, with a hinge margin $\lambda = 0.1$. The similarity measures $\sigma_{\mathcal{S}}$ and $\sigma_{\mathcal{R}}$ remain fixed in the process, only the case base CB_f is learned.

We apply 10-fold cross validation with stratified splitting. For each dataset D, each fold is constructed by sampling three distinct subsets CB_0, $\mathcal{T}_{ref}$, and $\mathcal{T}_{test}$ from D. The candidate base CB_0 serves as the initial case base. The reference set $\mathcal{T}_{ref}$ is used by EnergyCompress to learn CB_f from CB_0. The test set $\mathcal{T}_{test}$ is not used for learning, but only to measure the accuracy of each CBP algorithm before and after learning took place. The sizes for these sets are taken to be $|\mathcal{T}_{ref}| = |\mathcal{T}_{test}| = \min(100, .2 \times |D|)$, and $|CB_0| = \min(50, |D| - (|\mathcal{T}_{ref}| + |\mathcal{T}_{test}|))$.

Two quality criteria are considered. The accuracy increase $\text{Acc}(Pred_{\theta_f}(\mathcal{T}_{test})) - \text{Acc}(Pred_{\theta_0}(\mathcal{T}_{test}))$ measures the difference between the initial accuracy $\text{Acc}(Pred_{\theta_0}(\mathcal{T}_{test}))$ and the accuracy $\text{Acc}(Pred_{\theta_f}(\mathcal{T}_{test}))$ computed with the final parameters $\theta_f = (\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, CB_f)$. The relative size $\frac{|CB_f|}{|CB_0|}$ of the learned case base is the ratio between the size of the learned case base CB_f and the size of the initial case base CB_0.

In order to study the extent to which EnergyCompress provides case bases that are tailored for the prediction algorithm using them, we apply the following protocol: for each pair $(pred_1, pred_2) \in Pred^2$ (*i*) we learn a case base using EnergyCompress on $pred_1$ and $\mathcal{T}_{ref}$, and (*ii*) we use the learned case base to train $pred_2$ and observe its performance on $\mathcal{T}_{test}$. If the performance with the pair $(pred_1, pred_2)$ $(pred_1 \neq pred_2)$ is significantly lower than with the pairs $(pred_1, pred_1)$ and $(pred_2, pred_2)$, the case base learned with $pred_1$ (resp. $pred_2$) is more suitable for prediction with $pred_1$ (resp. $pred_2$).

[5]https://scikit-learn.org/stable/modules/generated/sklearn.svm.SVC.html

Learn	*Pred*	Acc. increase (%)	$\frac{\lvert CB_f\rvert}{\lvert CB_0\rvert}$
CNNR	CoAT	+6.3 ±7.9	53.6% ±12.7
ENN		+5.4 ±9.1	85.7% ±18.4
XLDIS		+3.8 ±5.3	62.6% ±21.5
LSSm		+3.6 ±6.0	89.5% ±13.6
IB3		+1.6 ±1.8	75.7% ±13.8
EC		**+13.3** ±11.4	**43.7%** ±13.4
CNNR	CtCoAT	+12.3 ±17.0	**43.9%** ±24.3
ENN		+4.2 ±6.2	90.3% ±19.4
XLDIS		+7.8 ±12.6	69.0% ±30.3
LSSm		+2.4 ±4.1	92.9% ±12.6
IB3		+1.6 ±2.1	87.0% ±11.1
EC		**+26.5** ±18.4	72.7% ±17.3
CNNR	kNN	+2.1 ±5.7	**60.2%** ±12.1
ENN		+1.8 ±4.0	90.3% ±15.5
XLDIS		+1.4 ±4.1	86.7% ±26.4
LSSm		+1.3 ±1.9	92.7% ±10.4
IB3		+1.3 ±2.4	85.8% ±15.6
EC		**+7.4** ±4.7	75.0% ±15.9
CNNR	SVM-linear	+3.1 ±3.6	**63.0%** ±9.5
ENN		+2.2 ±2.3	90.2% ±10.5
XLDIS		+2.0 ±3.8	78.5% ±18.6
LSSm		+2.5 ±3.2	90.0% ±11.4
IB3		+2.1 ±3.0	79.9% ±9.7
EC		**+5.3** ±7.1	75.8% ±17.4
CNNR	SVM-poly	+2.0 ±5.1	66.2% ±9.1
ENN		+1.0 ±1.2	89.7% ±8.0
XLDIS		+1.4 ±2.4	77.5% ±19.4
LSSm		+1.6 ±3.5	90.6% ±12.0
IB3		+1.4 ±2.6	84.2% ±13.1
EC		**+3.8** ±5.6	**45.2%** ±18.0
CNNR	SVM-rbf	+1.6 ±3.7	**62.3%** ±10.6
ENN		+0.8 ±1.0	93.1% ±7.6
XLDIS		+1.3 ±2.8	83.7% ±22.4
LSSm		+1.1 ±2.3	93.0% ±12.0
IB3		+1.0 ±2.4	89.1% ±13.2
EC		**+3.7** ±6.0	72.9% ±17.0

Table 3: Average accuracy increase and relative case base size.

5.2 Results

Increasing performance. EnergyCompress (denoted EC in the result table) always increased the performance of the CBP algorithms while substantially reducing the case base size. Tab. 3 shows the average accuracy increase and relative size after compression obtained on the 18 UCI datasets for each case base learning method and each CBP algorithm. The obtained accuracy increase was strictly positive for k-NN, CoAT, and CtCoAT on all datasets. The average accuracy increase is of $+7.4$ for k-NN (min 2.5, max 22.5), $+13.3$ for CoAT (min 3, max 39.7), and $+26.5$ for CtCoAT (min 8, max 65.8). The accuracy increase is less significant for SVMs but still always positive or null on all datasets, between 3.7% and 5.3% in average. The reason why CtCoAT exhibits the

CBP algorithm	Acc. before	Acc. after
CoAT	$65.8\%_{\pm 24.9}$	$79.9\%_{\pm 15.6}$
CtCoAT	$51.5\%_{\pm 25.3}$	$78.3\%_{\pm 15.7}$
kNN	$70.9\%_{\pm 19.1}$	$78.4\%_{\pm 16.0}$
SVM-linear	$72.1\%_{\pm 23.4}$	$77.4\%_{\pm 17.3}$
SVM-poly	$65.3\%_{\pm 21.5}$	$69.1\%_{\pm 16.8}$
SVM-rbf	$72.5\%_{\pm 22.7}$	$76.2\%_{\pm 17.6}$

Table 4: Average accuracy before and after EnergyCompress.

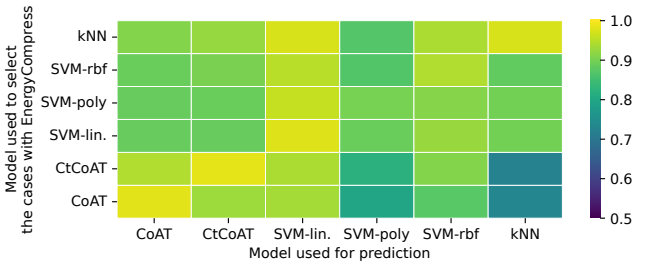

Figure 2: Mean accuracy over the 10 fold of cross validation on Iris, when predicting with a different model (horizontal axis) than the one used with EnergyCompress (vertical axis).

highest increase is probably due to the fact that for the same similarity measure $\sigma_{\mathcal{S}}$, the average performance of CtCoAT is the lowest before case base learning (Tab. 4), which means that CtCoAT starts with the highest margin of improvement. Nonetheless, after learning, the algorithm exhibits a performance comparable to an algorithm such as k-NN.

With respect to its objective to increase accuracy, EnergyCompress far outperforms all methods of the state of the art for k-NN, CoAT, and CtCoAT: the accuracy increase is doubled compared to other methods on average. All accuracy increases provided in Tab. 3 and the more detailed results given in Tab. 4 are significant, as for each row (180 experiments each) the paired t-tests p-values are of the order of 10^{-4} or smaller, except for 5 cases where it is of the order of 10^{-3} (IB3 with SVM-poly, SVM-rbf, and CoAT, and XLDIS with SVM-rbf and kNN). The gain in case base size is also very competitive compared to the state of the art: on average EnergyCompress gives the lowest relative case base size for CoAT and SVM-poly, and is among the three lowest case base sizes for the other algorithms, even though in these experiments, the stopping criterion for compression was chosen to optimize accuracy on the reference set, not compression.

The results also demonstrate the generality of the approach. EnergyCompress can be applied to a variety of prediction algorithms and allows reducing the case base size (and thus, the inference time) for all tested algorithms. The accuracy increase is less significant for SVMs, however, perhaps because SVMs are not relying that much on a complex inference process involving the case base CB at prediction time, so tuning the CB parameter may have less effect on their performance.

Additionally, the success of the compression process and the increase in performance validate the proposed general competence model. Despite not using accuracy when computing the competence, for all the considered CBP approaches the cases selected for removal have a similar impact on performance as they have on competence. This follows what was observed for synthetic data with CoAT in [Marquer *et al.*, 2023], and this alignment between predictive performance and the measure of case competence substantiates the intuitions behind using the hinge loss to measure competence. Additionally, the success of the compression process and the increase in performance validates the proposed general competence model.

Case base tailored learning. The case base learned by EnergyCompress is tailored for a given CBP algorithm. Fig. 2 illustrates this idea by showing for the Iris dataset the accuracies obtained according to the model used to select the cases and the model used for prediction. The best accuracy is obtained on the diagonal, when these two models coincide. The performance drop between this diagonal and the other pairs is significant (according to paired Student t-tests), with, for instance, for Iris, more than 50% of pairs with p-values under 0.031, and 25% under 0.001.

6 Conclusion and Future Work

EnergyCompress is a general case-base learning strategy that enables increasing the performance of a wide variety of case-based prediction algorithms while reducing the size of the case base. The obtained case base is not only smaller, but it is also tailored for the prediction algorithm for which it has been learned. The method is general: it can be applied to any case base prediction algorithm, and even to training data selection for any classifier that assigns a probability (normalized or not) to each potential outcome. The underlying energy-based model can be leveraged to capture various competence notions such as the competence of a case or of a case base, or the influence of a particular case on the prediction.

We believe that the obtained results (*e.g.,*$+7.4\%$ accuracy in average for the k-NN algorithm) constitute a breakthrough because they demonstrate that the role of the case base is not only to provide valuable knowledge on the task at hand, but a parameter that can be learned to tune the prediction algorithm to the task. These improvements were made possible by the recent progress in the modeling of the case base inference process, but work is needed to better understand this family of algorithms (*e.g.,*understand why the CtCoAT algorithm relies so heavily on the quality of the case base). In this paper, $\sigma_{\mathcal{S}}$ was fixed, so the next steps will include exploring how case base compression interacts with similarity learning for CBP. Other perspectives include testing on regression scenarios, exploring alternative stopping criteria in Algorithm 1 (*e.g.,*to favor the compression ratio instead of accuracy), developing and enriching the competence model, running a qualitative analysis of the obtained case bases.

Acknowledgements

This work has been supported by the French National Agency for Research (ANR), namely, projects ANR-22-CE23-0002 (ERIANA), ANR-22-CE23-0023 (AT2TA), and ANR-22-CE23-0032 (SMeLT).

References

[Aha *et al.*, 1991] David W. Aha, Dennis F. Kibler, and Marc K. Albert. Instance-based learning algorithms. *Machine Learning*, 6:37–66, 1991.

[Anthony and Ratsaby, 2015] Martin Anthony and Joel Ratsaby. A probabilistic approach to case-based inference. *Theoretical Computer Science*, 589:61–75, July 2015.

[Badra and Lesot, 2022] Fadi Badra and Marie-Jeanne Lesot. CoAT-APC: When Analogical Proportion-based Classification Meets Case-Based Prediction. In *Workshop on Analogies: From Theory to Applications, ATA@ICCBR*, 2022.

[Badra and Lesot, 2023] Fadi Badra and Marie-Jeanne Lesot. Case-based prediction – A survey. *Int. Journal of Approximate Reasoning*, 158, 2023.

[Badra, 2020] Fadi Badra. A Dataset Complexity Measure for Analogical Transfer. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'20*, pages 1601–1607, 2020.

[Bounhas and Prade, 2024] Myriam Bounhas and Henri Prade. Revisiting analogical proportions and analogical inference. *Int. Journal of Approximate Reasoning*, 171, 2024.

[Bounhas *et al.*, 2017] Myriam Bounhas, Henri Prade, and Gilles Richard. Analogy-based classifiers for nominal or numerical data. *Int. Journal of Approximate Reasoning*, 91:36–55, 2017.

[Carbonera, 2017] Joel Luis Carbonera. An efficient approach for instance selection. In *DaWaK*, volume 10440 of *Lecture Notes in Computer Science*, pages 228–243. Springer, 2017.

[Couceiro *et al.*, 2017] Miguel Couceiro, Nicolas Hug, Henri Prade, and Gilles Richard. Analogy-preserving functions: A way to extend boolean samples. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'17*, pages 1575–1581, 2017.

[Couceiro *et al.*, 2018] Miguel Couceiro, Nicolas Hug, Henri Prade, and Gilles Richard. Behavior of analogical inference w.r.t. boolean functions. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI'18*, pages 2057–2063, 2018.

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

[Cunha *et al.*, 2023] Washington Cunha, Felipe Viegas, Celso França, Thierson Rosa, Leonardo Rocha, and Marcos André Gonçalves. A comparative survey of instance selection methods applied to nonneural and transformer-based text classification. *ACM Computing Surveys*, 2023.

[Ghojogh *et al.*, 2022] Benyamin Ghojogh, Ali Ghodsi, Fakhri Karray, and Mark Crowley. Spectral, probabilistic, and deep metric learning: Tutorial and survey. In *arXiv:2201.09267*, 2022.

[Hart, 1968] Peter Hart. The condensed nearest neighbor rule. *IEEE Transactions on Information Theory*, 14, 1968.

[Hinton *et al.*, 2006] Geoffrey Hinton, Simon Osindero, Max Welling, and Yee-Whye Teh. Unsupervised Discovery of Nonlinear Structure Using Contrastive Backpropagation. *Cognitive Science*, 30(4):725–731, 2006.

[Juarez *et al.*, 2018] Jose M. Juarez, Susan Craw, J. Ricardo Lopez-Delgado, and Manuel Campos. Maintenance of Case Bases: Current Algorithms after Fifty Years. In *Proc. of the Int. Joint Conf. on Artificial Intelligence, IJCAI*, pages 5457–5463, 2018.

[Kolodner and Leake, 1996] Janet Kolodner and David Leake. A tutorial introduction to case-based reasoning. In *Case-Based Reasoning: Experiences, Lessons, and Future Directions*, pages 31–65. AAAI Press, 1996.

[Leake and Wilson, 1999] D. Leake and D. Wilson. When experience is wrong: Examining CBR for changing tasks and environments. In *Case-Based Reasoning Research and Development, ICCBR 1999*, pages 218–232, Berlin, 1999. Springer.

[Leake and Ye, 2021] David Leake and Xiaomeng Ye. Harmonizing case retrieval and adaptation with alternating optimization. In *Case-Based Reasoning Research and Development, ICCBR 2021*, pages 125–139. Springer, 2021.

[LeCun *et al.*, 2006] Yann LeCun, Sumit Chopra, Raia Hadsell, Marc'Aurelio Ranzato, and Fu Jie Huang. A Tutorial on Energy-Based Learning. In *Predicting Strutured Data*, page 59. MIT Press, 2006.

[Leyva *et al.*, 2015] Enrique Leyva, Antonio González Muñoz, and Raúl Pérez. Three new instance selection methods based on local sets: A comparative study with several approaches from a bi-objective perspective. *Pattern Recognition*, 48(4):1523–1537, 2015.

[Marquer *et al.*, 2023] Esteban Marquer, Fadi Badra, Marie-Jeanne Lesot, Miguel Couceiro, and David Leake. Less is Better: An Energy-Based Approach to Case Base Competence. In *Workshop on Analogies: From Theory to Applications, ATA@ICCBR*, 2023.

[Mathisen *et al.*, 2019] Bjørn Magnus Mathisen, Agnar Aamodt, Kerstin Bach, and Helge Langseth. Learning similarity measures from data. *Progress in Artificial Intelligence*, 10 2019.

[Platt, 1999] John Platt. Probabilities for SV Machines. In *Advances in Large Margin Classifiers*, volume 10 (3), pages 61–74, January 1999.

[Richter, 2003] Michael M Richter. Knowledge Containers. In *Readings in Case-Based Reasoning*, 2003.

[Smyth and McKenna, 2001] Barry Smyth and Elizabeth McKenna. Competence Models and the Maintenance Problem. *Computational Intelligence*, 17(2):235–249, 2001.

[Tomek, 1976] Ivan Tomek. An Experiment with the Edited Nearest-Neighbor Rule. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-6(6):448–452, 1976.