Less is Better: An Energy-Based Approach to Case Base Competence

Esteban Marquer[1], Fadi Badra[2], Marie-Jeanne Lesot[3], Miguel Couceiro[1], David Leake[4]

[1] Université de Lorraine, CNRS, Loria, Nancy, France
[2]Université Sorbonne Paris Nord, LIMICS, Sorbonne Université, INSERM, F-93000, Bobigny, France
[3]Sorbonne Université, CNRS, LIP6, F-75005 Paris, France
[4]The Luddy School of Informatics, Computing, and Engineering, Indiana University, USA

July 17th 2023

Outline

Case-Based Reasoning and Analogical Transfer

Case-Based Reasoning (CBR):

- Situations $\mathcal{S}$
- Outcomes $\mathcal{R}$
- Case (s_i, r_i): situation and corresponding outcome
- Known cases: Case Base (CB)
- Task: given s_j, find its outcome

For 1-nearest neighbor, we take $(s_i, r_i) \in CB$ such that:

	s_j (new)	most similar to s_i
then	r_j (to predict)	most similar to r_i

$$s_i : s_j :: r_i : r_j$$

CoAT: Complexity-based Analogical Transfer

CoAT (Badra 2020)[1]: **Co**mplexity-based **A**nalogical **T**ransfer

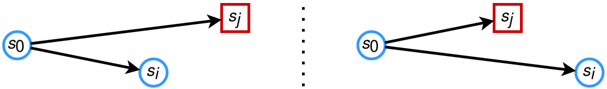
- (Dis)similar situations have (dis)similar outcomes
- Transfer similarity: situations $\rightsquigarrow$ outcomes
- We need compatible similarity measures

[1] Fadi Badra (2020). "A Dataset Complexity Measure for Analogical Transfer". In: *Proc. of the 29th Int. Joint Conf. on Artificial Intelligence IJCAI*, pp. 1601–1607

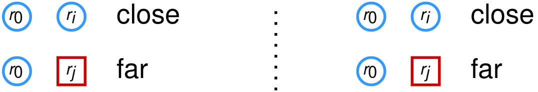
CoAT: **Co**mplexity-based **A**nalogical **T**ransfer

If s_0 closer to s_i than to s_j,
then r_0 closer to r_i than to r_j.

$\boldsymbol{\sigma_S}$: Euclidean similarity



$\boldsymbol{\sigma_R}$: class membership



$\boldsymbol{\sigma_S}$ and $\boldsymbol{\sigma_R}$ agree

$\boldsymbol{\sigma_S}$ and $\boldsymbol{\sigma_R}$ disagree

CoAT: Complexity-based Analogical Transfer

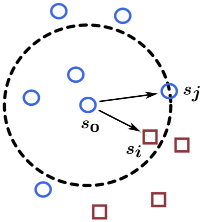
Continuity constraint: Ordinal compatibility between $\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}$

$$\text{if} \quad \sigma_{\mathcal{S}}(s_0, s_i) \geq \sigma_{\mathcal{S}}(s_0, s_j) \quad \text{then} \quad \sigma_{\mathcal{R}}(r_0, r_i) \geq \sigma_{\mathcal{R}}(r_0, r_j)$$

Inversion of similarity:

$$\sigma_{\mathcal{S}}(s_0, s_i) \geq \sigma_{\mathcal{S}}(s_0, s_j) \quad \text{but} \quad \sigma_{\mathcal{R}}(r_0, r_i) < \sigma_{\mathcal{R}}(r_0, r_j)$$

CoAT indicator $\Gamma(\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, CB)$: total number of inversions of similarity observed on a case base CB



CoAT: Complexity-based Analogical Transfer

Compatibility of $\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}$ with the CB: no inversion = perfect match

Predicting:

$$r_t = \arg\min_{r \in \mathcal{R}} \Gamma(\sigma_{\mathcal{S}}, \sigma_{\mathcal{R}}, CB \cup \{(s_t, r)\})$$

Energy-based models: macroscopic view on probability

"An **energy-based model** is a **probabilistic model** governed by an energy function that describes the probability of a certain state. [...] The Boltzmann distribution [(thermodynamics)] establishes a concrete relationship between energy and probability: **low-energy states are the most likely to be observed**."[1]

Energy function E_θ

Learning model: low energy for correct outcome, high for other

Predicting: given a situation, find outcome with the lowest energy

$$r_t = \arg\min_{r \in \mathcal{R}} E_\theta(s_t, r)$$

[1] The Physics of Energy-Based Models. Patrick Huembeli, Juan Miguel Arrazola, Nathan Killoran, Masoud Mohseni, Peter Wittek, Towards data science, 2021.

CoAT in terms of energy

Energy function & Γ: global view

low energy/Γ: more desirable outcome

Energy of a new case:

$$E_\theta(s_t, r) = \Gamma(\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB \cup \{(s_t, r)\})$$

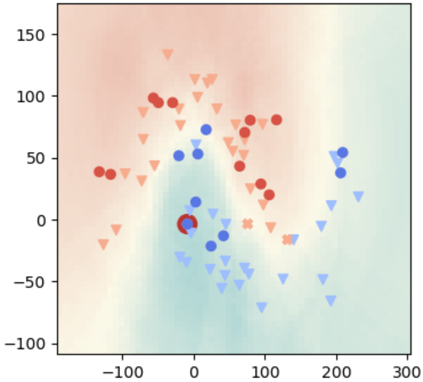
(s_t, r): new case added to the CB

$E_\theta : \mathcal{S} \times \mathcal{R} \longrightarrow \mathbb{R}$, with $\theta = (\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB)$

Predicting:

$$\begin{aligned} r_t &= \arg\min_{r \in \mathcal{R}} E_\theta(s_t, r) \\ &= \arg\min_{r \in \mathcal{R}} \Gamma(\sigma_\mathcal{S}, \sigma_\mathcal{R}, CB \cup \{(s_t, r)\}) \end{aligned}$$

CoAT in terms of energy: Half Moon Running Example



Energy map

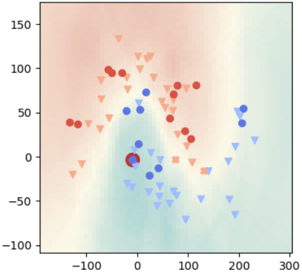
Blue VS **red** class

●▼X: explained later

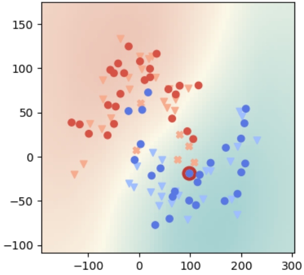
Case Base Competence

Competent CB: predicts the right outcome with high confidence

Less competent CB: wrong outcome or low confidence



Competent



Less competent

Case Base Competence

Energy-based pred. confidence[1]:

predicted outcome r_t − next best outcome $\hat{r}$

$$E_\theta(s_t, r_t) - \min_{\hat{r} \neq r_t} E_\theta(s_t, \hat{r})$$

High confidence: $|E_\theta(s_t, r_t) - \min_{\hat{r} \neq r_t} E_\theta(s_t, \hat{r})|$ large

Wrong prediction: expected instead of predicted outcome

$$E_\theta(s_t, r_t) - \min_{\hat{r} \neq r_t} E_\theta(s_t, \hat{r}) > 0$$

$$\iff \exists \hat{r} \neq r_t, E_\theta(s_t, \hat{r}) < E_\theta(s_t, r_t)$$

[1] Yann LeCun et al. (2006). "A Tutorial on Energy-Based Learning". In: *Predicting Structured Data*. MIT Press

Case Base Competence

Training loss of energy-based models[1]: prediction error of the CB *w.r.t.* new case $c_t = (s_t, r_t)$

Minimum Classification Error (MCE) loss:

$$\ell_{MCE}(CB, c_t) = E_\theta(s_t, r_t) - \min_{\hat{r} \neq r_t} E_\theta(s_t, \hat{r})$$

Hinge loss:

$$\ell_{hinge}(CB, c_t) = \max(0, \lambda + \ell_{MCE}(CB, c_t))$$

[1] Yann LeCun et al. (2006). "A Tutorial on Energy-Based Learning". In: *Predicting Structured Data*. MIT Press

Case Base Competence

Energy-inspired CB competence, on reference set $\mathcal{T}$:

$$C(CB, \mathcal{T}) = -\frac{1}{|\mathcal{T}|} \sum_{c_t \in \mathcal{T}} \ell(CB, c_t)$$

Case Competence

Contribution of c to the CB competence:

$$C(c, CB, \mathcal{T}) = C(CB, \mathcal{T}) - C(CB \setminus \{c\}, \mathcal{T})$$

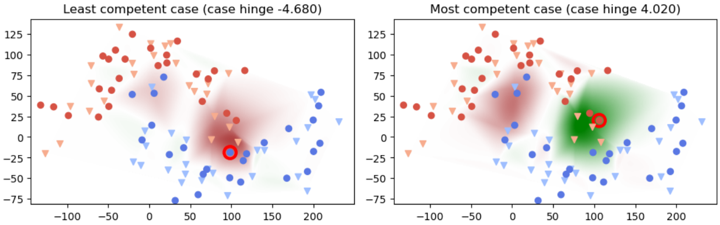
I.e., inversions involving case c

Case Influence & Expertise Area

Case competence *w.r.t.* a situation:

$$influence_{CB}(c, c_t) = \ell(CB, c_t) - \ell(CB \setminus \{c\}, c_t)$$
$$C(c, CB, \mathcal{T}) = \frac{1}{|\mathcal{T}|} \sum_{c_t \in \mathcal{T}} influence_{CB}(c, c_t)$$

Positive and negative influence of 2 cases (ℓ_{hinge}):



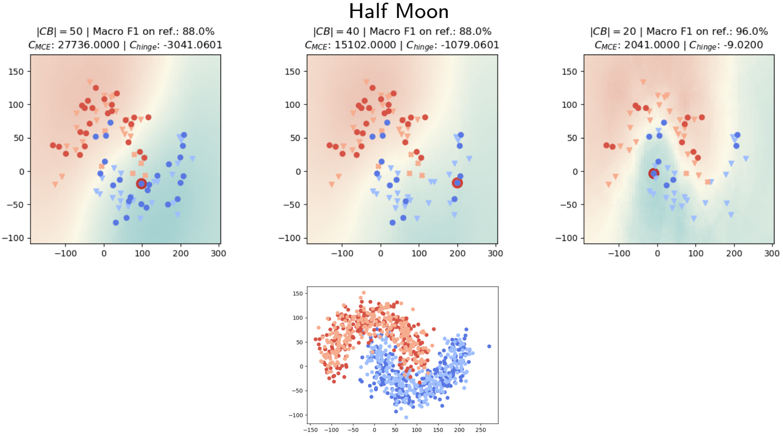
CB Maintenance: Case Deletion

Algorithm 1 Case deletion procedure

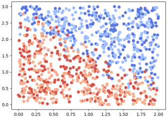
Require: An initial case base CB and a reference set $\mathcal{T}$
while $|CB| > 0$ **and** stopping condition not fulfilled **do**
$c_{worse} = \arg\min_{c \in CB} \; C(c, CB, \mathcal{T})$
$CB = CB \setminus \{c_{worse}\}$
end while

- $\sigma_{\mathcal{S}} = \exp(-||x-y||_2^2)$: Euclidean sim.
- $\sigma_{\mathcal{R}} = 1$ if same class else 0: Class membership
- 10 CB $\times 10$ reference set (100 pairs), each 50 cases

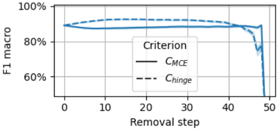
CB Maintenance: Case Deletion Example

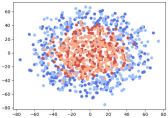


CB Maintenance: Performance and Competence

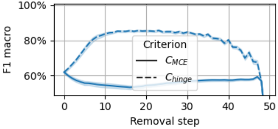


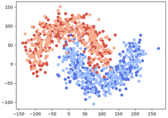
Line



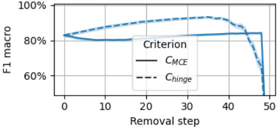


Ring





Half Moon



Performance after removing (using C_{hinge}):

1. $\nearrow$: "noisy" case
2. $\rightarrow$: "redundant" case
3. $\searrow$: "useful" case

Main Findings

- Global approach VS usually local

- C_{hinge} more suitable than C_{MCE}
- C_{hinge} linked to CB performance, matches intuition of competence

- Can compress by 60% or 80% while increasing performance
- Robust *w.r.t.* initial CB & reference set
  - Properly distributed reference set $\not\rightarrowtail$ bad bias
  - Enough initial cases in CB $\not\rightarrowtail$ holes
- Appears to fit any distrib. (*c.f.* example Half Moon) even if $\sigma_{\mathcal{S}}$ not really suitable

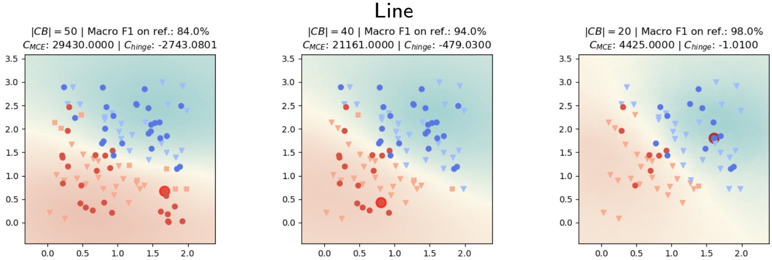
(Some) Ongoing Work

- Competence linked with CB performance using other CBR methods than CoAT
- Real-world data
- More than 2 classes
- Dynamic aspect: impact of order of removal
- Robustness to errors, small differences, ... in similarities
- ...

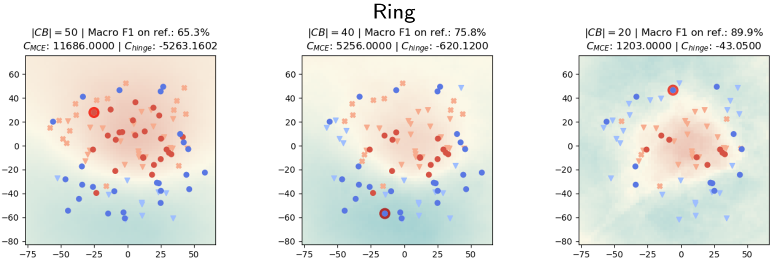
Questions

Thank you for your attention!

Case Deletion: Example



Case Deletion: Example



Case Deletion Run Time

Current implementation: quite optimized for CPU
Could go much faster with GPU

CB: 50 cases $\rightsquigarrow$ 1 case
50 reference cases

- Similarities computed once at the start $< 0.05s$
- Step 1: $\approx 2s$
- Step 25: $\approx 1s$
- Step 49: $< 0.1s$
- Total: $\approx 50s$

Bibliography

Badra, Fadi (2020). "A Dataset Complexity Measure for Analogical Transfer". In: *Proc. of the 29th Int. Joint Conf. on Artificial Intelligence IJCAI*, pp. 1601–1607.

LeCun, Yann et al. (2006). "A Tutorial on Energy-Based Learning". In: *Predicting Structured Data*. MIT Press.