Maintenance benchmark on real world data

Esteban Marquer

22 septembre

Goal

1. Get a results on real data
2. Get a benchmark on real data for CB maintenance
   $\rightarrow$ Reproduce the setting from:
   "Combining and Choosing Case Base Maintenance Algorithms"
   Ph.D. thesis of Lisa Cummins, 2013

Baseline

Benchmark experiment setup

Attempt to fix weights

Results

What next

Baselines Green: maintenance baseline avail. Red: not found

Dataset Name	Cases	Attributes	Classes	Accuracy (%)
Balance	625	4	3	85.12
Breast Cancer Diagnostic	569	30	2	96.90
Breast Cancer Prognostic	198	33	2	71.58
Breathalyser	127	5	2	71.60
Credit Approval	690	15	2	86.92
Dermatology	366	34	6	97.75
Glass Identification	214	9	7	69.05
Haberman's Survival	306	3	2	69.51
Heart Disease Cleveland	303	14	5	53.22
Hepatitis	155	19	2	80.63
Ionosphere	351	33	2	86.71
Iris	150	4	3	97.00
Lenses	24	4	3	72.50
Liver Disorders	345	6	2	64.20
Lung Cancer	32	56	3	48.00
Pima Indians Diabetes	768	8	2	70.78
Post-Operative Patient	90	8	3	64.71
Spam1	1000	699	2	93.35
Spam2	1000	699	2	94.3
Spam3	1000	699	2	98.25
Spam4	1000	699	2	97.05
Spam5	1000	699	2	94.8
Teaching Assistant Evaluation	151	5	3	55.33
Wine	178	13	3	96.67
Zoo	101	16	7	91.50
Average over twenty-five datasets	-	-	-	80.30

Baselines maintainance

	25 datasets		Breathalyser		Credit		Lenses	
Algorithm	Del	Acc	Del	Acc	Del	Acc	Del	Acc
ICF	78.76	73.58	77.53	74.00	84.02	83.38	44.38	52.50
RC	88.75	75.09	87.66	66.80	87.84	86.38	86.25	55.00
CBE	55.31	77.67	61.95	71.20	55.90	83.62	68.13	65.00

Table 2.7: Results for existing algorithms

- ICF: Iterative Case Filtering [9]
  H. Brighton and C. Mellish. "On the consistency of information filters for lazy learning algorithms", 1999
- RC: RC-CNN [58]
  E. McKenna and B. Smyth. "Competence-guided case-base editing techniques", 2000
- CBE (Case Base Editing) = BBNR (Blame-Based Noise Reduction) + CRR (Conservative Redundancy Reduction) [19]
  S.J. Delany and P. Cunningham. "An analysis of case-based editing in a spam filtering system", 2004.

We copy these values for our baseline.

Current datasets

verbose name	found	has maintenance baseline
Lenses	True	**True**
Credit Approval	True	**True**
Zoo	True	False
Wine	True	False
Teaching Assistant Evaluation	True	False
Post-Operative Patient	True	False
Pima indians Diabetes	True	False
Lung Cancer	True	False
Liver Disorders	True	False
Iris	True	False
Ionosphere	True	False
Hepatitis	True	False
Heart Disease Cleveland	True	False
Haberman's Survival	True	False
Glass Identification	True	False
Dermatology	True	False
Breast Cancer Pronostic	True	False
Breast Cancer Diagnostic	True	False
Balance	True	False

Current datasets

verbose name	found	has maintenance baseline
Breathalyser	False	**True**
Spam5	False	False
Spam4	False	False
Spam3	False	False
Spam2	False	False
Spam1	False	False

Process

train/dev/test split: 60%/20%/20%, as in the thesis

train: cases in the CB at the start
dev: cases used for maintenance decision
test: cases used to mesure performance

Similarity

Numeric attribute: $1-$ normalized absolute distance

$$sim(x, y) = 1 - \frac{x - y}{max(X) - min(X)}$$

with x, y the value of the att. for 2 cases,
X the values of the att. for all cases in the CB

Symbolic attribute:

$$sim(x, y) = 1 \text{ if } x = y \text{ else } 0$$

with x, y the value of the att. for 2 cases

Overall similarity: weighted similarity (how the weights are obtained not mentioned in the thesis)

Models

- MeATCube
- 1NN: 1-Nearest-Neighbor
- "KNN": "All"-Nearest-Neighbor
  outcome obtained by voting, vote weight inversely proportional to similarity of the case with the target
  Cummins mention this model without specifying the k of kNN, so we take all cases.

Processing

Processing (MeATCube):

1. set **unit weights** for each feature in the similarity for MeATCube
2. compress MeATCube using hinge competence

At each compression step step:

- Compute MeATCube prediction performance
- Recompute weights for 1NN and KNN and compute performance (does not change the weights of MeATCube)

Processing

Processing (1-NN and k-NN):

1. take current cases in the CB
2. find feature weights using Neighborhood Components Analysis (NCA):
   "learns a linear transformation in a supervised fashion to improve the classification accuracy of a stochastic nearest neighbors rule in the transformed space" (scikit-learn documentation)
3. **[NEW!]** find feature weights using:
   "Weighted Similarity Measure for k-Nearest Neighbors Algorithm" B. Karabulut, G. Arslan, H. M. Ünver, 2019, Celal Bayar University Journal of Science
4. compute accuracy

Result summary: conclusions (details after)

- using NCA:
  - Neither 1-NN nor k-NN **computed** match **reported** performance (before compression) except in rare cases.
- using Karabulut *et al.*'s method:
  - 1-NN much closer to **reported** perf., more stable, and sometimes some interesting effects along the deletion process: *e.g.*, k-NN and 1-NN perf. up at the very end of the process (more analysis needed).
  - */!\completely anachronistic: Cummins in 2013, Karabulut et al. in 2019.*

Result summary: conclusions (details after)

- Except for "Post-Operative Patient", MeATCube outperforms 1-NN and k-NN at its optimum for size & performance.
- MeATCube outperforms the baselines for maintenance in 1 of the 2 datasets for which we have the information.
  For the other we do not know if the performance difference is due to the method itself or to a bad similarity, *cf.* k-NN performance mismatch.

Attempt to fix weights

"Weighted Similarity Measure for k-Nearest Neighbors Algorithm"
B. Karabulut, G. Arslan, H. M. Ünver, 2019, Celal Bayar University Journal of Science

Computed on the CB

Using scikit-learn notation X, y for "situations, outcomes":

- X: situations for cases in the CB
- y: outcomes for cases in the CB
- $X[k] \in X$: situation of case k
- $X[k][a]$: value of attribute a for case k
- $y[k] \in classes$: outcome for case k
- $len(X) = len(y) = |CB|$

Attempt to fix weights

$C_i(a)$: set of values for attribute a belonging to class i

$$C_i(a) = \{X[k][a] : X[k] \in X \text{ and } y[k] = i\}$$

$A_i(a)$: set of cases with attribute a within values of class i

$$A_i(a) = \{X[k] \in X : min(C_i(a)) \leq X[k][a] \leq max(C_i(a))\}$$

$$A_i(a) = \{X[k] \in X : X[k][a] \in C_i(a)\} \text{ for nominal att. (defined by me)}$$

$B_i(a)$: set of cases with attribute a within values of class i but not any other class

$$B_i(a) = A_i(a) - \cup_{i \neq j, j \in classes} A_j(a)$$

w_a: weight for attribute a, average "ability to discriminate"

$$w_a = |\cup_{i \in classes} B_i(a)|/n, \;\; n : len(X)$$

$w_a* = w_a/(\sum_{a'} w_{a'})$: normalized w_a

(in paper it was $w_a = (\bigcup_{i \in classes} |B_i(a)|)/n$, but it makes no sense)

Attempt to fix weights

Karabulut *et al.*'s method not perfect

Issue: may not scale well for symbolic attributes, as this part is custom made by me

balance+scale: all $B_i(a) = \emptyset$, need to add a default value and weights end up all equal

Result format

“Steps”:

- initial: CB before maintenance
- best MeATCube F1: step where MeATCube reaches its highest F1
- best MeATCube acc.: step where MeATCube reaches its highest accuracy
- best KNN acc.: “All”-Nearest-Neighbor (vote inversely proportional to similarity) reaches its highest accuracy
- best 1NN acc.: 1-Nearest-Neighbor reaches its highest accuracy
- *if multiple with same score, take the one with the smallest CB*

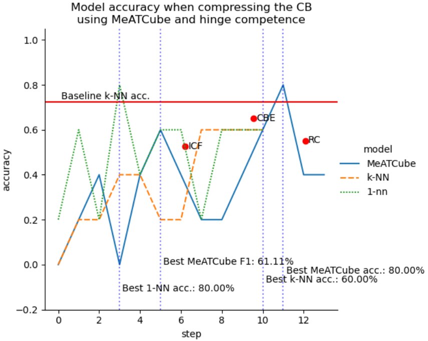
Result format

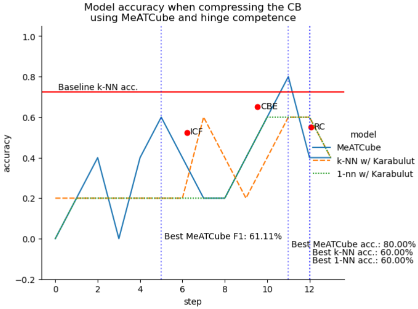
Available metrics:

- cb size
- deletion rate $= 1 - (\text{cb size}/\text{initial size})$
- step $= \text{initial size} - \text{cb size}$
- MeATCube F1 (macro) & acc.
- KNN & 1NN acc.

Result: MeATCube outperforms all but RC

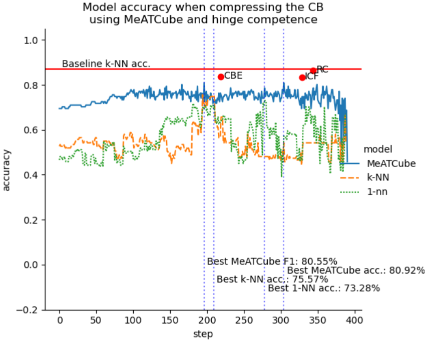
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

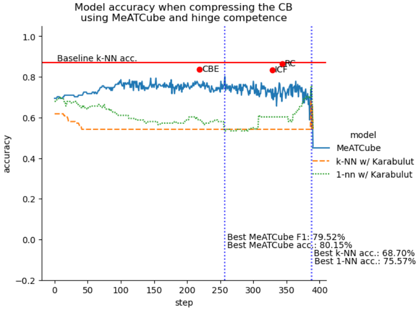




Result: MeATCube outperforms k-NN, not the baseline

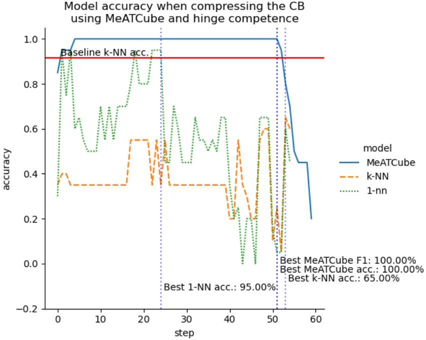
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

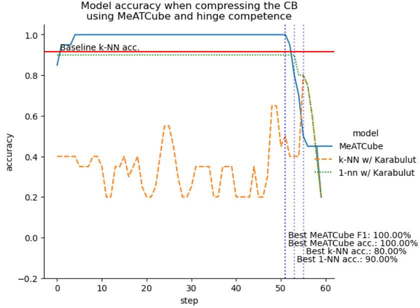




Result: MeATCube outperforms baseline and reaches perfect scores

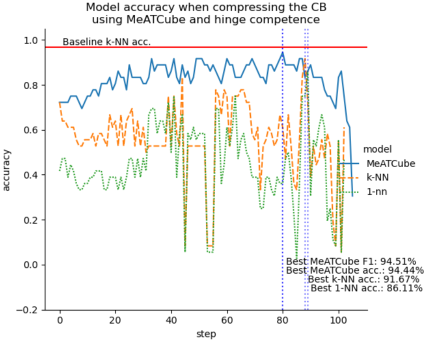
Lenses
Credit Approval
<u>Zoo</u>
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

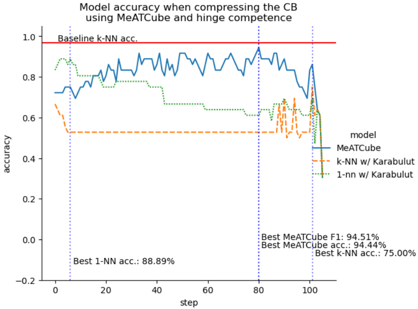




Result

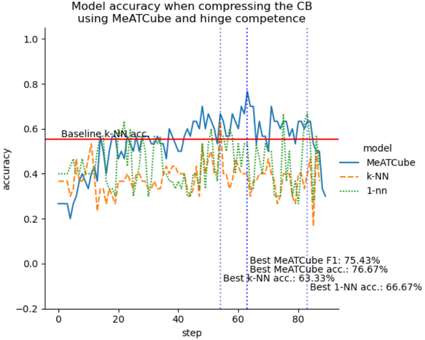
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

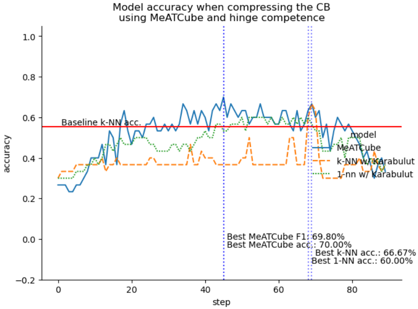




Result: MeATCube outperforms baseline

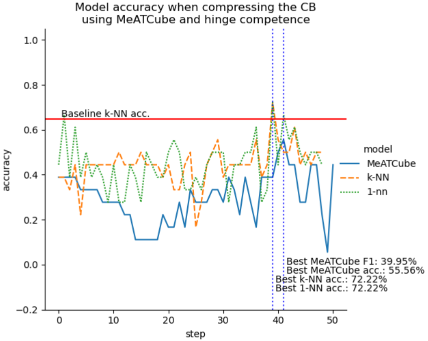
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

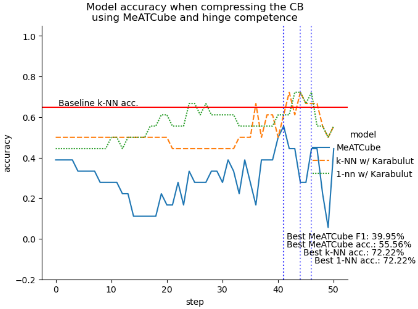




Result: MeATCube outperformed

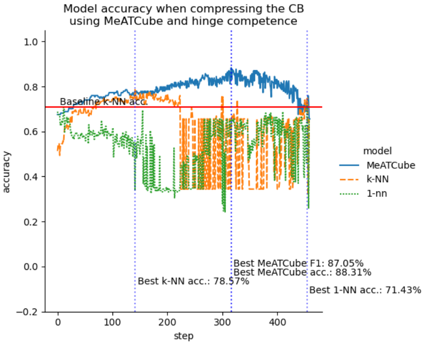
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

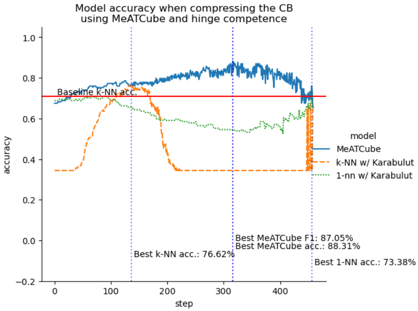




Result

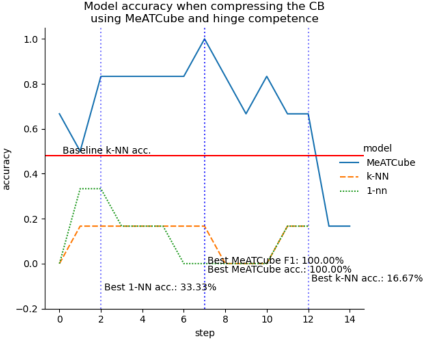
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

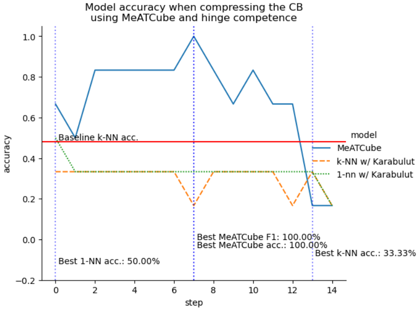




Result: MeATCube significantly outperforms everything

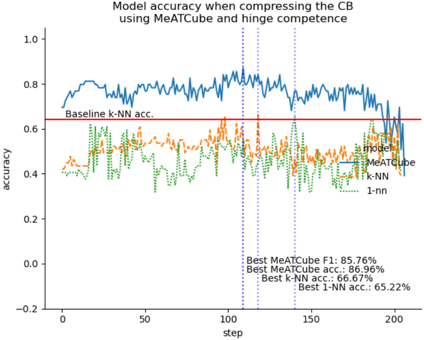
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

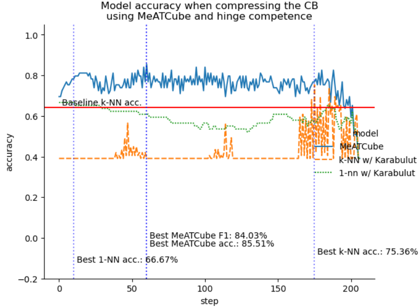




Result: MeATCube significantly outperforms everything

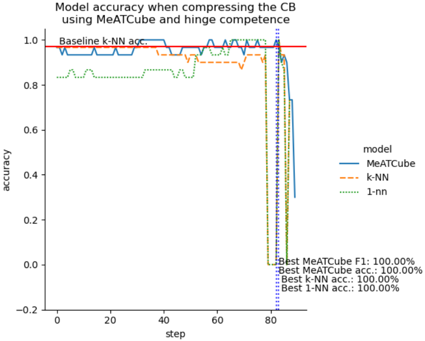
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
<u>Liver Disorders</u>
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

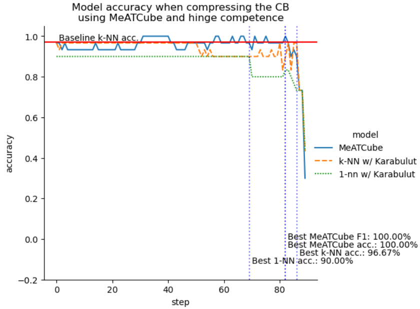




Result: all models comparable

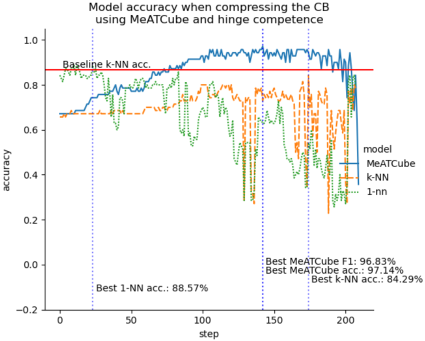
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
<u>Iris</u>
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

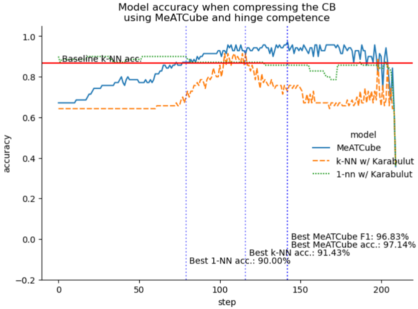




Result

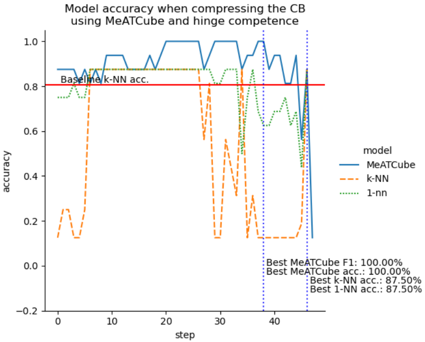
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

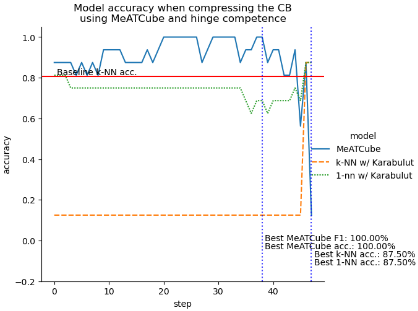




Result

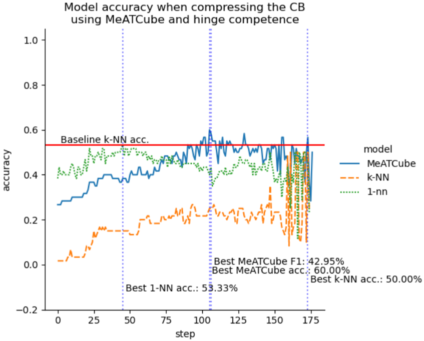
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

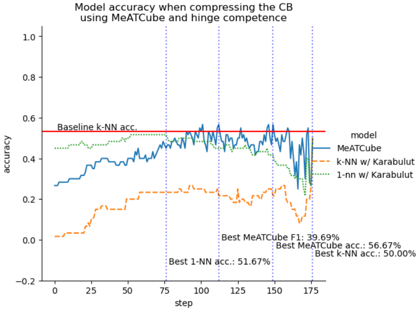




Result: k-NN and MeATCube **appear** correlated

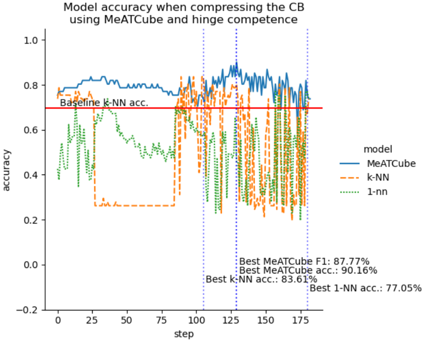
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
<u>Heart Disease Cleveland</u>
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

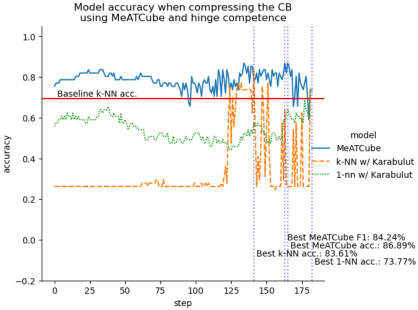




Result

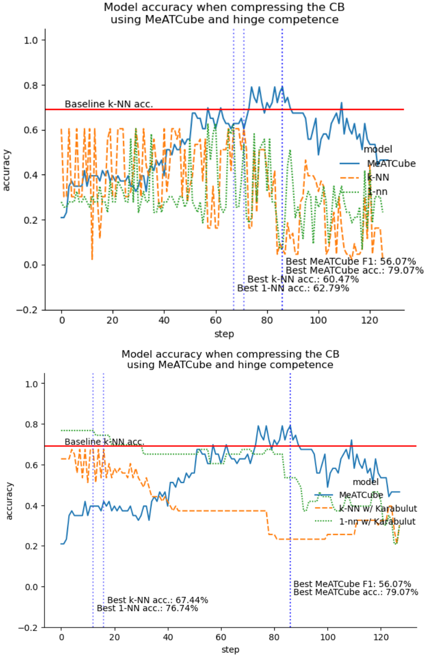
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance





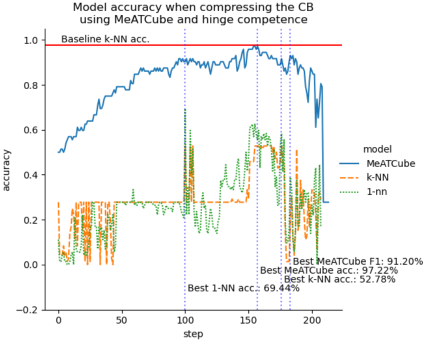
Result

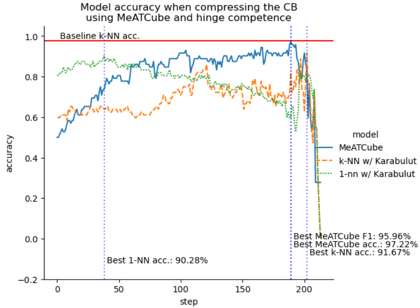
- **Lenses**
- **Credit Approval**
- Zoo
- Wine
- Teaching Assistant Evaluation
- Post-Operative Patient
- Pima indians Diabetes
- Lung Cancer
- Liver Disorders
- Iris
- Ionosphere
- Hepatitis
- Heart Disease Cleveland
- Haberman's Survival
- Glass Identification
- Dermatology
- Breast Cancer Pronostic
- Breast Cancer Diagnostic
- Balance



Result

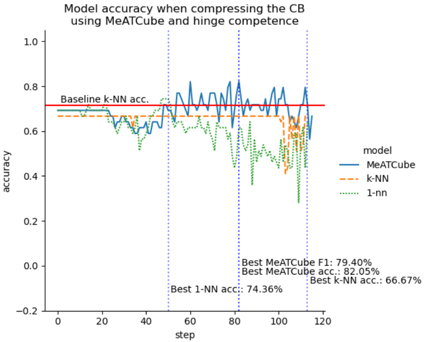
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance

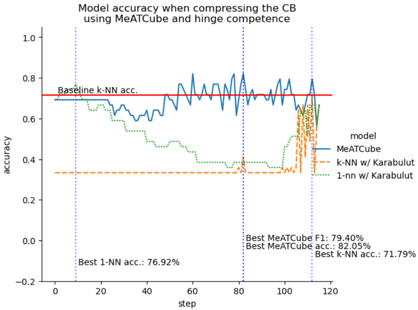




Result

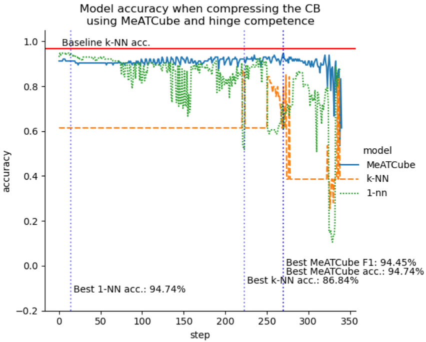
Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
<u>Breast Cancer Pronostic</u>
Breast Cancer Diagnostic
Balance

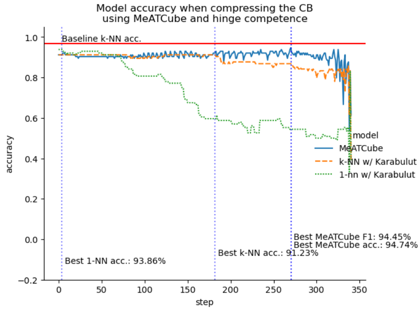




Result

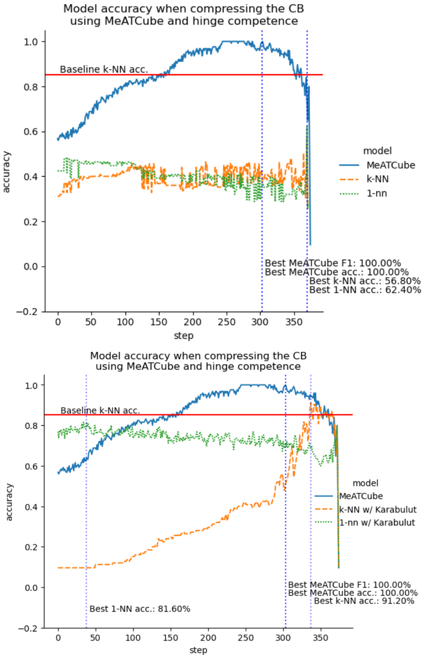
- **Lenses**
- **Credit Approval**
- Zoo
- Wine
- Teaching Assistant Evaluation
- Post-Operative Patient
- Pima indians Diabetes
- Lung Cancer
- Liver Disorders
- Iris
- Ionosphere
- Hepatitis
- Heart Disease Cleveland
- Haberman's Survival
- Glass Identification
- Dermatology
- Breast Cancer Pronostic
- Breast Cancer Diagnostic
- Balance





Result

Lenses
Credit Approval
Zoo
Wine
Teaching Assistant Evaluation
Post-Operative Patient
Pima indians Diabetes
Lung Cancer
Liver Disorders
Iris
Ionosphere
Hepatitis
Heart Disease Cleveland
Haberman's Survival
Glass Identification
Dermatology
Breast Cancer Pronostic
Breast Cancer Diagnostic
Balance



What next

Datasets to add?

Fixing the weights: how?

Apply the baselines instead of copying them
Apply more recent baselines

Cross-validation