Supplementary material for the article "Operational Evaluation of Algorithms for Online Streaming Continual Multi-Label Classification"

Hugo Peuzet
Laboratory of Digital Sciences of Nantes
Nantes, France
hugo.peuzet@univ-nantes.fr

Pascale Kuntz
Laboratory of Digital Sciences of Nantes
Nantes, France
pascale.kuntz@univ-nantes.fr

Frank Meyer
Orange Labs
Lannion, France
frank.meyer@orange.com

Vincent Lemaire
Orange Labs
Lannion, France
vincent.lemaire@orange.com

Killian Le Mau
Orange Labs
Lannion, France
killian.lemau@orange.com

A. *Algorithms*

- Neural networks [1] (NN1): a neural network without hidden layers, equivalent to logistic regression, implemented using PyTorch. The output activation function is the sigmoid function. The loss function is the binary cross-entropy which is commonly applied in multi-label classification [1]. The optimization algorithm implemented is Adam. The time complexity for processing an input instance is $\mathcal{O}(d * l)$, where d is the number of input features and l the label number. Parameters: learning rate [0.1, 0.01, 0.001].
- Neural network with a targeted loss function (NN2_TL): a variation of the NN1 model that introduces a focused loss function computation. Unlike the traditional NN1, this loss function is here exclusively calculated on the labels belonging to the signature of the evaluated task[2]. The losses associated to the other labels are set to zero. The time complexity is $\mathcal{O}(d*l)$. Parameters: learning rate [0.1, 0.01, 0.001].
- Neural network with a targeted loss function and a hidden layer (NN3_TLH): a variation of the NN2_TL model with an additional hidden layer based on a ReLU activation function. The time complexity is $\mathcal{O}(d*h+h*l)$ where h is the size of the hidden layer. Parameters: learning rate [0.1, 0.01, 0.001], hidden layer size [200, 2000].
- Neural network with a targeted loss function, a hidden layer, and a FIFO memory (NN4_TLH_fifo): a variation of the NN3_TLH model with an online data replay based on a FIFO memory. Each instance of the data stream is stored in a memory with a deletion of the oldest stored instance, thereby maintaining a fixed capacity. During the training phase, a subset of n memorized instances is randomly sampled from the FIFO memory to train the online neural network [2]. The time complexity is $\mathcal{O}(r_f*(d*h+h*l))$ where r_f is the number of instances randomly sampled from the FIFO memory. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [200,2000], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10].
- Neural network with targeted loss function, a hidden layer, and reservoir sampling (NN5_TLH_sampling): a variation of the NN3_TLH model with an online data replay based on a reservoir sampling. Each instance of the data stream is stored in a reservoir with a deletion of a random stored instance. During the training phase, a subset of n memorized instances is randomly sampled from the reservoir to train the online neural network. The time complexity is $\mathcal{O}(r_s * (d * h + h * l))$ where r_s is the number of instances randomly sampled from the reservoir. Parameters: learning rate [0.1, 0.01, 0.001], size of the hidden layer [200,2000], size of the reservoir [100, 1000], number of instances randomly sampled from the reservoir [5,10].
- Neural network with targeted loss function, a hidden layer, FIFO memory, and reservoir sampling (NN6_TLH_memories): a variation of NN3_TLH with an online data replay based on a FIFO memory and a reservoir sampling. The time complexity is $\mathcal{O}((r_s+r_f)*(d*h+h*l))$. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [200,2000], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10], size of the reservoir [100,

[1] Normalization of the input feature vectors of the neural networks was considered. However, since the datasets did not require normalization and preliminary testing revealed negligible effects from its implementation, it was decided that normalization would not be employed in this study.

[2] In contrast to the other algorithms tested, neural networks with task-signature-targeted loss function computation can take advantage of the knowledge of the task in progress during learning.

1000], number of instances randomly sampled from the reservoir [5,10].

- Neural network with targeted loss function, a hidden layer, with data replay with FIFO memory and reservoir sampling in the form of mini-batches (NN7_TLH_mini_memories): a variation of NN6_TLH_memories with a single pass mini-batch learning for the current instance and the data replay. The time complexity is $\mathcal{O}((r_s + r_f) * (d * h + h * l))$. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [200,2000], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10], size of the reservoir [100, 1000], number of instances randomly sampled from the reservoir [5,10].
- Neural network with targeted loss function, using a ResNET10 architecture, FIFO memory, and reservoir sampling (NN8_TLH_mini_memories_deep): a variation of NN7_TLH_mini_memories with a single pass mini-batch learning for the current instance and the data replay, using a ResNET10 architecture. The time complexity is $\mathcal{O}((r_s + r_f) * d * (h + l))$. Parameters: learning rate [0.1, 0.01, 0.001], size of hidden layer [128,256,512], weight decay [1e-4, 1e-5], size of FIFO memory [100, 1000], number of instances randomly sampled from the FIFO memory [5,10], size of the reservoir [100, 1000], number of instances randomly sampled from the reservoir [5,10].
- T1_BR: Hoeffding tree classifier, as proposed by Hulten et al. [3], adapted for the multi-label problem using the binary relevance method [4]. The time complexity is $\mathcal{O}(l * d)$ where d is the number of input features and l the label number. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T2_LC: Hoeffding tree classifier, adapted for the multi-label problem using a pruned set method [5]. The time complexity is $\mathcal{O}(2^l * d)$. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T3_CC: Hoeffding tree classifier, adapted for the multi-label problem using a classifier chain [6]. The time complexity is $\mathcal{O}(l * d)$. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T4_ARF: Adaptive Random Forest algorithm, as proposed by Gomes et al. [7], adapted for the multi-label problem using the binary relevance method. The time complexity is $\mathcal{O}(k * l * d)$ where k is the number of classifiers. Parameters: number of models [5, 10, 15], grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- T5_SOUP: Incremental Structured Output Prediction Tree algorithm, developed for multi-target regression [8]. The time complexity is $\mathcal{O}(l * d)$. Parameters: grace period [100, 200], delta [1e-06,1e-07], tau [0.05, 0.1].
- B1_oracle: prediction of the correct label vector for each instance of the stream.
- B2_last_class: prediction of the label vector of the instance at $t-1$ for the instance at time t.
- B3_prior_distribution: prediction of the most frequent value of each label in the past stream. The time complexity is $\mathcal{O}(l)$ where l is the number of labels.
- B4_mean: prediction of the mean value of each label in the past stream. The time complexity is $\mathcal{O}(l)$.
- B5_1NN: prediction of the label vector of the instance in the FIFO memory closest to the input instance in the feature space. The time complexity is $\mathcal{O}(n * d)$ where n is the size of the FIFO memory and d the number of features. Parameters: FIFO memory size [0,10,100,1000], use of cosine or Euclidean similarity, similarity of two zero vectors to 1 or 0.

B. Synthetic datasets

To study the algorithm behavior, three synthetic datasets were generated following our protocol: synth_monolab, composed of tasks without common labels; synth_bilab, featuring tasks with a limited number of common labels for analyzing forward and backward transfers; and synth_rand, which presents a more complex scenario where the tasks all have the same label signature. Each dataset consists of four features and four labels, with features drawn randomly from a uniform distribution between 0 and 1. Three distinct $M = (M_{i,j})_{1 \leq i \leq 4, 1 \leq j \leq 4}$ matrices, where each row corresponds to a task and $M_{i,j}$ is the multiplier coefficient applied to feature j in task i, were defined. After the application of the multiplier coefficient, the corresponding label is equal to 1 if the feature value is greater than 0.5, and 0 otherwise.

- synth_monolab: dataset associated to a diagonal matrix.

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

- synth_bilab: dataset associated to a matrix with two multiplier coefficients per task.

$$M = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

- synth_rand: dataset associated to a randomly generated matrix of rank 4, to which a constant of 0.5 was added to mitigate the excessive occurrence of absent labels after thresholding at 0.5.

$$M = \begin{bmatrix} 1.1 & 1.3 & 1.4 & 1.3 \\ 0.9 & 1 & 1.5 & 1.1 \\ 1 & 1.3 & 0.6 & 1.2 \\ 0.7 & 0.6 & 0.6 & 1.4 \end{bmatrix}$$

C. Relationship between the negative forward transfer $FWT-$ and catastrophic forgetting

It is illustrated on a simplified scenario:

- Task 1. Labels: L_1, L_2, L_3, L_4.
- Task 2. Labels: L_3, L_4, L_5, L_6.
- Task 3. Labels: L_5, L_6, L_7, L_8.

- Task 4. Labels: L_1, L_2, L_7, L_8.

1) Task 1 appears in the stream: $FWT+$ is observed in task 2 (resp. 4) via the labels L_3 and L_4 (resp. L_1 and L_2).
2) Task 2 appears in the stream: $FWT+$ is observed in task 3 via the labels L_5 and L_6. However, an omission associated with labels L_1 and L_2 entails a negative forward transfer $FWT-$ for task 4.
3) etc.

This example illustrates the observations made with the synth_bilab dataset, which was developed to facilitate the examination of transfer phenomena.

D. Statistical tests

The Friedman test was successfully applied to all metrics, revealing significant differences among the algorithms in the experimental comparisons. The Nemenyi post hoc test did not provide conclusive results. A future analysis will require a new benchmark with a broader range of datasets.

E. Complementary results

Additional results, including measurements for RMSE and precision@p are provided below with comprehensive tables of the metrics along with results for the baseline.

References

[1] A. Maxwell, R. Li, B. Yang, H. Weng, A. Ou, H. Hong, Z. Zhou, P. Gong, and C. Zhang, "Deep learning architectures for multi-label classification of intelligent health risk prediction," *BMC bioinformatics*, vol. 18, pp. 121–131, 2017.

[2] P. Buzzega, M. Boschini, A. Porrello, and S. Calderara, "Rethinking experience replay: a bag of tricks for continual learning," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 2180–2187.

[3] G. Hulten, L. Spencer, and P. Domingos, "Mining time-changing data streams," in *SIGKDD international conference on Knowledge discovery and data mining*, 2001, pp. 97–106.

[4] M.-L. Zhang, Y.-K. Li, X.-Y. Liu, and X. Geng, "Binary relevance for multi-label learning: an overview," *Frontiers of Computer Science*, vol. 12, pp. 191–202, 2018.

[5] J. Read, B. Pfahringer, and G. Holmes, "Multi-label classification using ensembles of pruned sets," in *2008 eighth IEEE international conference on data mining*. IEEE, 2008, pp. 995–1000.

[6] J. Read, B. Pfahringer, G. Holmes, and E. Frank, "Classifier chains for multi-label classification," in *European Conference on Machine Learning*. Springer, 2009, pp. 254–269.

[7] H. M. Gomes, A. Bifet, J. Read, J. P. Barddal, F. Enembreck, B. Pfharinger, G. Holmes, and T. Abdessalem, "Adaptive random forests for evolving data stream classification," *Machine Learning*, vol. 106, pp. 1469–1495, 2017.

[8] A. Osojnik, P. Panov, and S. Džeroski, "Tree-based methods for online multi-target regression," *Journal of Intelligent Information Systems*, vol. 50, pp. 315–339, 2018.

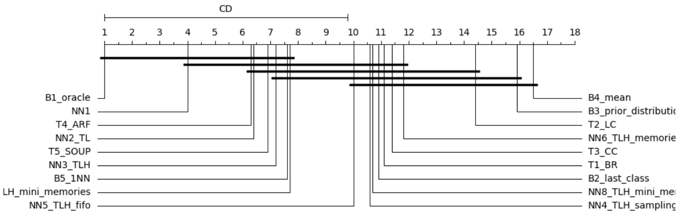

Fig. 1. BA_{macro}

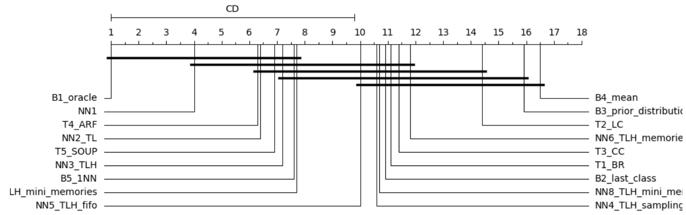

Fig. 2. Precision@p

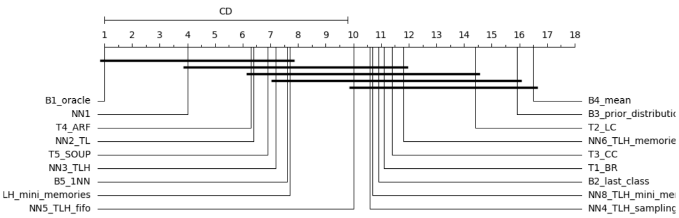

Fig. 3. RMSE

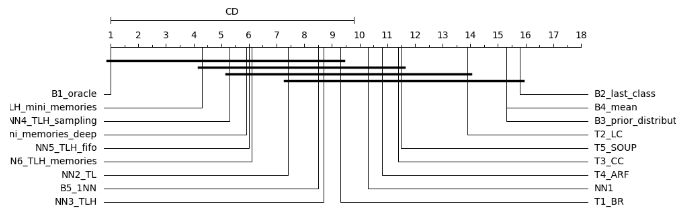

Fig. 4. Average accuracy

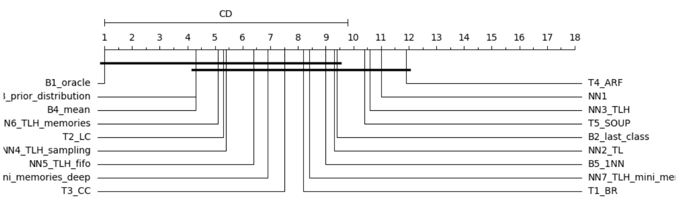

Fig. 5. $BWT-$

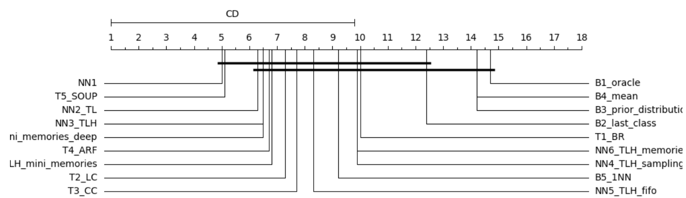

Fig. 6. $BWT+$

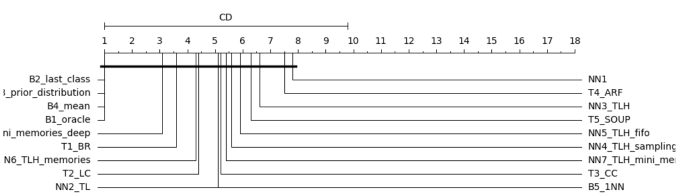

Fig. 7. $FWT-$

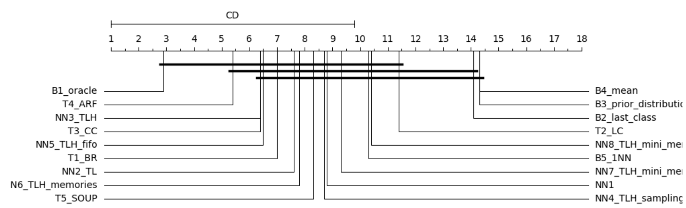

Fig. 8. $FWT+$

TABLE I
FINAL BA_{macro} OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.93	0.94	0.94	0.86	0.68	0.67	0.51	0.87	0.52	0.77 ± 0.17	4.00 ± 2.99
NN2_TL	0.74	0.82	0.94	0.68	0.67	0.67	0.56	0.85	0.52	0.72 ± 0.14	6.40 ± 3.36
NN3_TLH	0.77	0.83	0.95	0.70	0.69	0.65	0.50	0.84	0.51	0.72 ± 0.15	7.20 ± 4.28
NN4_TLH_sampling	0.70	0.81	0.89	0.66	0.62	0.65	0.50	0.84	0.50	0.68 ± 0.14	10.60 ± 3.71
NN5_TLH_fifo	0.78	0.80	0.95	0.64	0.62	0.65	0.50	0.81	0.49	0.69 ± 0.15	10.00 ± 5.58
NN6_TLH_memories	0.79	0.78	0.93	0.69	0.62	0.65	0.50	0.51	0.49	0.66 ± 0.15	11.80 ± 4.29
NN7_TLH_mini_memories	0.72	0.84	0.93	0.70	0.59	0.65	0.59	0.86	0.51	0.71 ± 0.14	7.70 ± 3.73
NN8_TLH_mini_memories_deep	0.65	0.70	0.81	0.61	0.62	0.63	0.56	0.80	0.53	0.66 ± 0.10	10.70 ± 4.67
T1_BR	0.70	0.80	0.89	0.78	0.58	0.53	0.52	0.65	0.52	0.66 ± 0.14	11.10 ± 2.71
T2_LC	0.60	0.63	0.77	0.74	0.56	0.49	0.50	0.65	0.50	0.60 ± 0.10	14.40 ± 2.73
T3_CC	0.67	0.68	0.89	0.81	0.58	0.54	0.52	0.64	0.52	0.65 ± 0.13	11.40 ± 3.44
T4_ARF	0.96	0.93	0.94	0.85	0.63	0.55	0.52	0.61	0.53	0.72 ± 0.19	6.30 ± 4.30
T5_SOUP	0.82	0.88	0.93	0.86	0.63	0.60	0.52	0.75	0.52	0.72 ± 0.16	6.90 ± 3.19
B1_oracle	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00 ± 0.00	1.00 ± 0.00
B2_last_class	0.70	0.67	0.56	0.77	0.64	0.57	0.52	0.58	0.53	0.62 ± 0.09	10.90 ± 4.39
B3_prior_distribution	0.58	0.58	0.52	0.64	0.53	0.51	0.51	0.50	0.50	0.54 ± 0.05	15.90 ± 1.92
B4_mean	0.58	0.58	0.52	0.64	0.53	0.51	0.51	0.50	0.50	0.54 ± 0.05	16.50 ± 2.12
B5_1NN	0.59	0.68	0.86	0.83	0.65	0.66	0.61	0.81	0.63	0.70 ± 0.10	7.60 ± 5.22

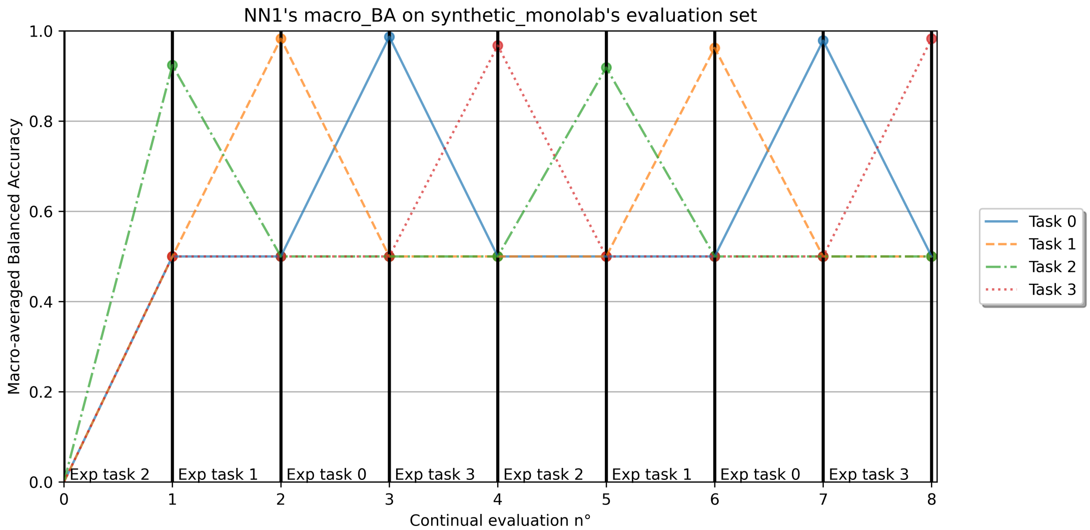

Fig. 9. BA_{macro} evolution of NN1 over time on synth_monolab.

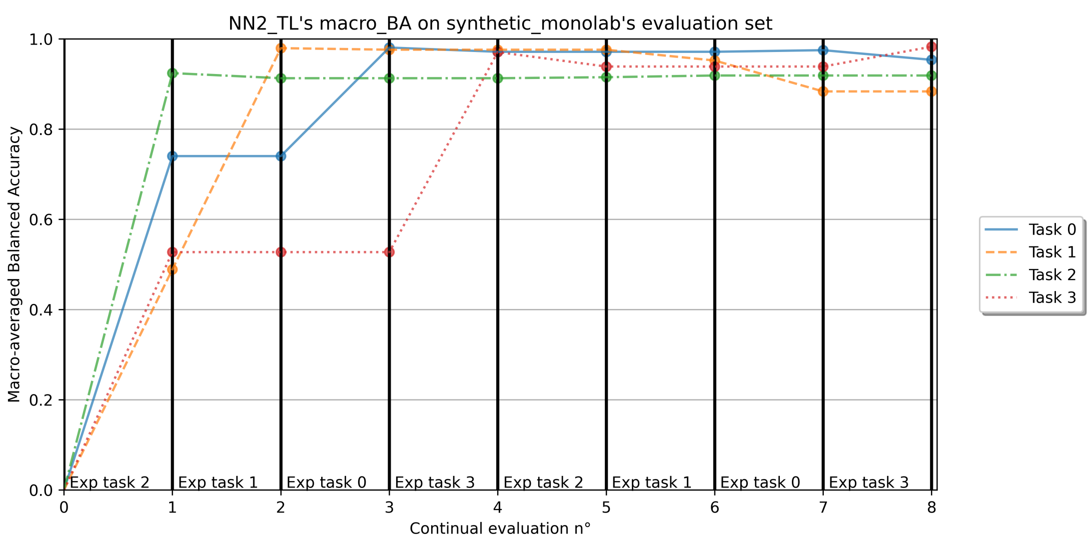

Fig. 10. BA_{macro} evolution of NN2_TL over time on synth_monolab.

TABLE II
AVERAGE ACCURACY OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.68	0.77	0.88	0.64	0.54	0.61	0.64	0.66	0.57	0.66 ± 0.10	10.30 ± 2.28
NN2_TL	0.94	0.94	0.88	0.68	0.54	0.61	0.66	0.67	0.57	0.72 ± 0.16	7.40 ± 2.30
NN3_TLH	0.92	0.95	0.89	0.65	0.54	0.61	0.64	0.67	0.56	0.72 ± 0.16	8.70 ± 1.90
NN4_TLH_sampling	0.97	0.97	0.91	0.70	0.58	0.63	0.64	0.69	0.56	0.74 ± 0.16	5.30 ± 2.70
NN5_TLH_fifo	0.98	0.95	0.90	0.68	0.58	0.64	0.64	0.70	0.56	0.74 ± 0.16	6.00 ± 4.32
NN6_TLH_memories	0.98	0.97	0.91	0.70	0.58	0.63	0.64	0.63	0.56	0.73 ± 0.17	6.10 ± 3.60
NN7_TLH_mini_memories	0.97	0.98	0.91	0.71	0.57	0.63	0.68	0.68	0.57	0.74 ± 0.17	4.30 ± 2.19
NN8_TLH_mini_memories_deep	0.84	0.84	0.84	0.73	0.56	0.63	0.67	0.71	0.57	0.71 ± 0.11	5.90 ± 4.00
T1_BR	0.63	0.87	0.92	0.64	0.55	0.58	0.64	0.62	0.57	0.67 ± 0.13	9.30 ± 4.49
T2_LC	0.64	0.68	0.80	0.66	0.56	0.43	0.63	0.59	0.44	0.60 ± 0.12	13.90 ± 3.60
T3_CC	0.64	0.78	0.92	0.63	0.54	0.57	0.64	0.62	0.57	0.66 ± 0.12	11.40 ± 3.60
T4_ARF	0.69	0.79	0.91	0.62	0.53	0.57	0.64	0.63	0.57	0.66 ± 0.12	10.80 ± 3.24
T5_SOUP	0.65	0.77	0.88	0.58	0.52	0.60	0.64	0.64	0.56	0.65 ± 0.11	11.50 ± 2.77
B1_oracle	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00 ± 0.00	1.00 ± 0.00
B2_last_class	0.50	0.50	0.50	0.52	0.50	0.56	0.64	0.61	0.56	0.54 ± 0.05	15.80 ± 1.00
B3_prior_distribution	0.50	0.50	0.50	0.51	0.50	0.57	0.64	0.63	0.56	0.55 ± 0.06	15.30 ± 1.72
B4_mean	0.50	0.50	0.50	0.51	0.50	0.57	0.64	0.63	0.56	0.55 ± 0.06	15.30 ± 1.72
B5_1NN	0.67	0.76	0.89	0.68	0.53	0.58	0.65	0.72	0.61	0.68 ± 0.11	8.50 ± 4.56

TABLE III
AVERAGE NEGATIVE BACKWARD TRANSFER (BWT-) OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	−0.46	−0.29	−0.08	−0.19	−0.07	−0.04	−0.00	−0.09	−0.00	−0.13 ± 0.15	11.00 ± 4.66
NN2_TL	−0.02	−0.02	−0.08	−0.13	−0.07	−0.04	−0.02	−0.08	−0.00	−0.05 ± 0.04	9.30 ± 3.57
NN3_TLH	−0.05	−0.02	−0.08	−0.11	−0.08	−0.05	−0.01	−0.09	−0.00	−0.05 ± 0.04	10.60 ± 4.13
NN4_TLH_sampling	−0.01	−0.01	−0.02	−0.02	−0.03	−0.03	−0.01	−0.11	−0.00	−0.03 ± 0.03	5.40 ± 4.44
NN5_TLH_fifo	−0.01	−0.03	−0.09	−0.05	−0.02	−0.03	−0.01	−0.08	−0.00	−0.03 ± 0.03	6.40 ± 3.91
NN6_TLH_memories	−0.01	−0.01	−0.06	−0.06	−0.02	−0.04	−0.01	−0.01	−0.00	−0.02 ± 0.02	5.10 ± 2.92
NN7_TLH_mini_memories	−0.02	−0.01	−0.06	−0.04	−0.02	−0.05	−0.01	−0.11	−0.01	−0.04 ± 0.03	8.40 ± 4.13
NN8_TLH_mini_memories_deep	−0.04	−0.02	−0.01	−0.03	−0.02	−0.04	−0.01	−0.05	−0.01	−0.03 ± 0.01	6.90 ± 3.77
T1_BR	−0.32	−0.14	−0.04	−0.08	−0.03	−0.02	−0.00	−0.05	−0.01	−0.08 ± 0.10	8.20 ± 2.96
T2_LC	−0.24	−0.14	−0.03	−0.04	−0.01	−0.00	−0.00	−0.01	−0.00	−0.05 ± 0.08	5.30 ± 3.97
T3_CC	−0.22	−0.19	−0.04	−0.07	−0.02	−0.02	−0.01	−0.06	−0.00	−0.07 ± 0.08	7.50 ± 3.18
T4_ARF	−0.48	−0.28	−0.09	−0.15	−0.05	−0.00	−0.01	−0.11	−0.01	−0.13 ± 0.16	11.90 ± 4.45
T5_SOUP	−0.39	−0.28	−0.08	−0.11	−0.04	−0.04	−0.01	−0.10	−0.01	−0.12 ± 0.13	10.40 ± 3.02
B1_oracle	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00	−0.00 ± 0.00	1.00 ± 0.00
B2_last_class	−0.00	−0.00	−0.00	−0.20	−0.08	−0.06	−0.03	−0.08	−0.01	−0.05 ± 0.07	9.40 ± 6.54
B3_prior_distribution	−0.00	−0.00	−0.00	−0.20	−0.08	−0.00	−0.00	−0.00	−0.00	−0.03 ± 0.07	4.30 ± 6.40
B4_mean	−0.00	−0.00	−0.00	−0.20	−0.08	−0.00	−0.00	−0.00	−0.00	−0.03 ± 0.07	4.30 ± 6.40
B5_1NN	−0.06	−0.05	−0.01	−0.02	−0.08	−0.08	−0.03	−0.03	−0.01	−0.04 ± 0.03	9.00 ± 5.25

TABLE IV
AVERAGE POSITIVE BACKWARD TRANSFER (BWT+) OF 18 STRATÉGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.00	0.12	0.07	0.08	0.03	0.02	0.00	0.10	0.00	0.05 ± 0.05	5.00 ± 4.48
NN2_TL	0.00	0.01	0.07	0.01	0.03	0.03	0.01	0.05	0.00	0.02 ± 0.02	6.30 ± 3.26
NN3_TLH	0.06	0.02	0.07	0.00	0.02	0.04	0.01	0.02	0.00	0.03 ± 0.03	6.50 ± 4.09
NN4_TLH_sampling	0.00	0.01	0.02	0.03	0.02	0.03	0.00	0.02	0.00	0.01 ± 0.01	9.90 ± 2.35
NN5_TLH_fifo	0.01	0.03	0.07	0.07	0.01	0.02	0.00	0.02	0.00	0.03 ± 0.03	8.30 ± 4.50
NN6_TLH_memories	0.01	0.01	0.05	0.06	0.02	0.03	0.00	0.01	0.00	0.02 ± 0.02	9.90 ± 3.90
NN7_TLH_mini_memories	0.02	0.01	0.06	0.04	0.02	0.03	0.01	0.02	0.00	0.02 ± 0.02	6.80 ± 2.79
NN8_TLH_mini_memories_deep	0.02	0.02	0.02	0.03	0.02	0.03	0.01	0.03	0.01	0.02 ± 0.01	6.50 ± 3.41
T1_BR	0.00	0.07	0.04	0.01	0.01	0.01	0.01	0.01	0.01	0.02 ± 0.02	10.00 ± 4.02
T2_LC	0.09	0.05	0.01	0.02	0.04	0.00	0.02	0.01	0.00	0.03 ± 0.03	7.30 ± 5.57
T3_CC	0.00	0.08	0.03	0.01	0.01	0.02	0.01	0.06	0.00	0.03 ± 0.03	7.70 ± 4.41
T4_ARF	0.00	0.12	0.06	0.01	0.02	0.00	0.01	0.03	0.01	0.03 ± 0.04	6.70 ± 3.76
T5_SOUP	0.00	0.14	0.06	0.01	0.03	0.03	0.01	0.03	0.01	0.04 ± 0.05	5.10 ± 4.25
B1_oracle	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00 ± 0.00	14.70 ± 2.46
B2_last_class	0.00	0.00	0.00	0.00	0.02	0.02	0.00	0.01	0.00	0.01 ± 0.01	12.40 ± 2.42
B3_prior_distribution	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00 ± 0.00	14.20 ± 2.29
B4_mean	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00 ± 0.00	14.20 ± 2.29
B5_1NN	0.00	0.00	0.01	0.02	0.03	0.05	0.01	0.00	0.00	0.01 ± 0.02	9.20 ± 5.05

TABLE V

AVERAGE NEGATIVE FORWARD TRANSFER (FWT-) OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	-0.00	-0.23	-0.05	-0.06	-0.03	-0.06	-0.00	-0.00	-0.01	-0.05 ± 0.07	7.80 ± 5.11
NN2_TL	-0.00	-0.01	-0.05	-0.01	-0.03	-0.02	-0.00	-0.01	-0.02	-0.02 ± 0.02	5.10 ± 4.61
NN3_TLH	-0.12	-0.01	-0.05	-0.13	-0.02	-0.00	-0.00	-0.04	-0.00	-0.04 ± 0.05	6.60 ± 4.12
NN4_TLH_sampling	-0.05	-0.00	-0.01	-0.04	-0.06	-0.00	-0.00	-0.08	-0.00	-0.03 ± 0.03	5.60 ± 4.41
NN5_TLH_fifo	-0.05	-0.04	-0.04	-0.08	-0.01	-0.00	-0.00	-0.04	-0.00	-0.03 ± 0.03	5.90 ± 3.43
NN6_TLH_memories	-0.04	-0.03	-0.01	-0.04	-0.03	-0.00	-0.00	-0.03	-0.00	-0.02 ± 0.02	4.30 ± 3.77
NN7_TLH_mini_memories	-0.08	-0.00	-0.01	-0.00	-0.06	-0.00	-0.02	-0.03	-0.00	-0.02 ± 0.03	5.40 ± 4.62
NN8_TLH_mini_memories_deep	-0.00	-0.01	-0.02	-0.00	-0.01	-0.00	-0.00	-0.02	-0.01	-0.01 ± 0.01	3.10 ± 2.44
T1_BR	-0.00	-0.13	-0.01	-0.03	-0.00	-0.00	-0.00	-0.00	-0.00	-0.02 ± 0.04	3.60 ± 2.79
T2_LC	-0.03	-0.24	-0.02	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.03 ± 0.08	4.40 ± 4.18
T3_CC	-0.00	-0.08	-0.02	-0.03	-0.00	-0.00	-0.01	-0.05	-0.00	-0.02 ± 0.03	5.20 ± 3.10
T4_ARF	-0.00	-0.13	-0.03	-0.18	-0.01	-0.00	-0.01	-0.10	-0.00	-0.05 ± 0.07	7.50 ± 3.97
T5_SOUP	-0.00	-0.21	-0.00	-0.12	-0.03	-0.01	-0.00	-0.03	-0.01	-0.05 ± 0.07	6.30 ± 3.97
B1_oracle	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00 ± 0.00	1.00 ± 0.00
B2_last_class	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00 ± 0.00	1.00 ± 0.00
B3_prior_distribution	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00 ± 0.00	1.00 ± 0.00
B4_mean	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00 ± 0.00	1.00 ± 0.00
B5_1NN	-0.00	-0.11	-0.01	-0.00	-0.00	-0.04	-0.02	-0.00	-0.02	-0.02 ± 0.04	5.10 ± 3.52

TABLE VI

AVERAGE POSITIVE FORWARD TRANSFER (FWT+) OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.17	0.29	0.31	0.51	0.18	0.32	0.20	0.22	0.29	0.28 ± 0.10	8.80 ± 3.81
NN2_TL	0.20	0.36	0.31	0.51	0.18	0.30	0.19	0.27	0.29	0.29 ± 0.10	7.60 ± 4.41
NN3_TLH	0.25	0.38	0.31	0.53	0.27	0.20	0.19	0.29	0.29	0.30 ± 0.10	6.40 ± 4.36
NN4_TLH_sampling	0.15	0.34	0.31	0.53	0.20	0.32	0.19	0.28	0.19	0.28 ± 0.12	8.70 ± 4.03
NN5_TLH_fifo	0.17	0.38	0.32	0.53	0.20	0.21	0.29	0.31	0.19	0.29 ± 0.12	6.50 ± 2.83
NN6_TLH_memories	0.25	0.38	0.31	0.50	0.27	0.21	0.19	0.27	0.19	0.29 ± 0.10	7.80 ± 4.06
NN7_TLH_mini_memories	0.16	0.34	0.31	0.31	0.20	0.20	0.29	0.30	0.19	0.26 ± 0.07	9.30 ± 2.40
NN8_TLH_mini_memories_deep	0.17	0.32	0.29	0.31	0.18	0.22	0.28	0.26	0.29	0.25 ± 0.05	10.40 ± 3.06
T1_BR	0.17	0.33	0.46	0.50	0.25	0.19	0.30	0.20	0.57	0.33 ± 0.15	7.00 ± 5.36
T2_LC	0.25	0.29	0.44	0.28	0.20	0.14	0.23	0.19	0.14	0.24 ± 0.09	11.40 ± 5.57
T3_CC	0.17	0.39	0.31	0.50	0.25	0.20	0.30	0.31	0.29	0.30 ± 0.10	6.40 ± 3.71
T4_ARF	0.17	0.46	0.33	0.56	0.25	0.19	0.30	0.32	0.19	0.31 ± 0.13	5.40 ± 4.90
T5_SOUP	0.17	0.29	0.30	0.51	0.25	0.22	0.20	0.31	0.29	0.28 ± 0.10	8.30 ± 3.52
B1_oracle	0.33	0.33	0.33	0.50	0.33	0.33	0.33	0.33	0.33	0.35 ± 0.06	2.90 ± 3.41
B2_last_class	0.17	0.17	0.17	0.25	0.16	0.19	0.20	0.21	0.19	0.19 ± 0.03	14.10 ± 3.02
B3_prior_distribution	0.17	0.17	0.17	0.25	0.16	0.19	0.20	0.20	0.19	0.19 ± 0.03	14.30 ± 3.04
B4_mean	0.17	0.17	0.17	0.25	0.16	0.19	0.20	0.20	0.19	0.19 ± 0.03	14.30 ± 3.04
B5_1NN	0.17	0.27	0.29	0.25	0.27	0.31	0.21	0.22	0.28	0.25 ± 0.05	10.30 ± 4.31

TABLE VII

RMSE OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.93	0.94	0.94	0.86	0.68	0.67	0.51	0.87	0.52	0.77 ± 0.17	4.00 ± 2.99
NN2_TL	0.74	0.82	0.94	0.68	0.67	0.67	0.56	0.85	0.52	0.72 ± 0.14	6.40 ± 3.36
NN3_TLH	0.77	0.83	0.95	0.70	0.69	0.65	0.50	0.84	0.51	0.72 ± 0.15	7.20 ± 4.28
NN4_TLH_sampling	0.70	0.81	0.89	0.66	0.62	0.65	0.50	0.84	0.50	0.68 ± 0.14	10.60 ± 3.71
NN5_TLH_fifo	0.78	0.80	0.95	0.64	0.62	0.65	0.50	0.81	0.49	0.69 ± 0.15	10.00 ± 5.58
NN6_TLH_memories	0.79	0.78	0.93	0.69	0.62	0.65	0.50	0.51	0.49	0.66 ± 0.15	11.80 ± 4.29
NN7_TLH_mini_memories	0.72	0.84	0.93	0.70	0.59	0.65	0.59	0.86	0.51	0.71 ± 0.14	7.70 ± 3.73
NN8_TLH_mini_memories_deep	0.65	0.70	0.81	0.61	0.62	0.63	0.56	0.80	0.53	0.66 ± 0.10	10.70 ± 4.67
T1_BR	0.70	0.80	0.89	0.78	0.58	0.53	0.52	0.65	0.52	0.66 ± 0.14	11.10 ± 2.71
T2_LC	0.60	0.63	0.77	0.74	0.56	0.49	0.50	0.65	0.50	0.60 ± 0.10	14.40 ± 2.73
T3_CC	0.67	0.68	0.89	0.81	0.58	0.54	0.52	0.64	0.52	0.65 ± 0.13	11.40 ± 3.44
T4_ARF	0.96	0.93	0.94	0.85	0.63	0.55	0.52	0.61	0.53	0.72 ± 0.19	6.30 ± 4.30
T5_SOUP	0.82	0.88	0.93	0.86	0.63	0.60	0.52	0.75	0.52	0.72 ± 0.16	6.90 ± 3.19
B1_oracle	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00 ± 0.00	1.00 ± 0.00
B2_last_class	0.70	0.67	0.56	0.77	0.64	0.57	0.52	0.58	0.53	0.62 ± 0.09	10.90 ± 4.39
B3_prior_distribution	0.58	0.58	0.52	0.64	0.53	0.51	0.51	0.50	0.50	0.54 ± 0.05	15.90 ± 1.92
B4_mean	0.58	0.58	0.52	0.64	0.53	0.51	0.51	0.50	0.50	0.54 ± 0.05	16.50 ± 2.12
B5_1NN	0.59	0.68	0.86	0.83	0.65	0.66	0.61	0.81	0.63	0.70 ± 0.10	7.60 ± 5.22

TABLE VIII

PRECISION@P OF 18 STRATEGIES ON 9 DATASETS.

	synth_monolab	synth_bilab	synth_rand	Scene	Yeast	Slashdot	Reuters-K500	20NG	Mediamill	Avg. value	Avg. Rank
NN1	0.93	0.94	0.94	0.86	0.68	0.67	0.51	0.87	0.52	0.77 ± 0.17	4.00 ± 2.99
NN2_TL	0.74	0.82	0.94	0.68	0.67	0.67	0.56	0.85	0.52	0.72 ± 0.14	6.40 ± 3.36
NN3_TLH	0.77	0.83	0.95	0.70	0.69	0.65	0.50	0.84	0.51	0.72 ± 0.15	7.20 ± 4.28
NN4_TLH_sampling	0.70	0.81	0.89	0.66	0.62	0.65	0.50	0.84	0.50	0.68 ± 0.14	10.60 ± 3.71
NN5_TLH_fifo	0.78	0.80	0.95	0.64	0.62	0.65	0.50	0.81	0.49	0.69 ± 0.15	10.00 ± 5.58
NN6_TLH_memories	0.79	0.78	0.93	0.69	0.62	0.65	0.50	0.51	0.49	0.66 ± 0.15	11.80 ± 4.29
NN7_TLH_mini_memories	0.72	0.84	0.93	0.70	0.59	0.65	0.59	0.86	0.51	0.71 ± 0.14	7.70 ± 3.73
NN8_TLH_mini_memories_deep	0.65	0.70	0.81	0.61	0.62	0.63	0.56	0.80	0.53	0.66 ± 0.10	10.70 ± 4.67
T1_BR	0.70	0.80	0.89	0.78	0.58	0.53	0.52	0.65	0.52	0.66 ± 0.14	11.10 ± 2.71
T2_LC	0.60	0.63	0.77	0.74	0.56	0.49	0.50	0.65	0.50	0.60 ± 0.10	14.40 ± 2.73
T3_CC	0.67	0.68	0.89	0.81	0.58	0.54	0.52	0.64	0.52	0.65 ± 0.13	11.40 ± 3.44
T4_ARF	0.96	0.93	0.94	0.85	0.63	0.55	0.52	0.61	0.53	0.72 ± 0.19	6.30 ± 4.30
T5_SOUP	0.82	0.88	0.93	0.86	0.63	0.60	0.52	0.75	0.52	0.72 ± 0.16	6.90 ± 3.19
B1_oracle	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00 ± 0.00	1.00 ± 0.00
B2_last_class	0.70	0.67	0.56	0.77	0.64	0.57	0.52	0.58	0.53	0.62 ± 0.09	10.90 ± 4.39
B3_prior_distribution	0.58	0.58	0.52	0.64	0.53	0.51	0.51	0.50	0.50	0.54 ± 0.05	15.90 ± 1.92
B4_mean	0.58	0.58	0.52	0.64	0.53	0.51	0.51	0.50	0.50	0.54 ± 0.05	16.50 ± 2.12
B5_1NN	0.59	0.68	0.86	0.83	0.65	0.66	0.61	0.81	0.63	0.70 ± 0.10	7.60 ± 5.22