Supplementary Document for ”XMAP: eXplainable Mapping Analytical Process”

Su Nguyen, Binh Tran
Center for Data Analytics and Cognition, La Trobe University

Algorithm 1 Adaptive Topological Learning (ATL)

Input: transformed data $Z = \{z_1, \dots, z_N\}$
Output: ATL network $\mathcal{N}$
1: set $epoch \leftarrow 0$, $t \leftarrow 0$, $a_{max} \leftarrow \infty$
2: randomly initialise $\mathcal{N}$ with two nodes
3: **repeat**
4: randomly shuffle Z
5: **for** each input $z \in Z$ **do**
6: calculate the distance $d_{z,i}$ between z and node i
7: identify the nearest (winning) node s_1 and the second-nearest node s_2 for input z, and their respective weights w_{s_1} and w_{s_2}
8: update the similarity thresholds for T_{s_1} and T_{s_2}
9: **if** $||z - w_{s_1}|| > T_{s_1}$ or $||z - w_{s_2}|| > T_{s_2}$ **then**
10: insert a new node s into $\mathcal{N}$ with $w_s \leftarrow z$
11: create an edge between s_1 and s_2 and set its age to zero if the edge does not exist; otherwise set the edge's age to zero
12: $NB_{s_1} \leftarrow$ get the neighbours of s_1
13: $\overline{d_{s_1}} \leftarrow \left(\sum_{n \in NB_{s_1}} d_{z,n}\right) / |NB|_{s_1}$
14: **for** each $n \in NB_{s_1}$ **do**
15: $age_n \leftarrow age_n + 1 + |NB_n| \frac{d_{z,n}}{\overline{d_{s_1}}}$
16: updating node s_1: $w_{s_1} \leftarrow w_{s_1} + \epsilon_1(z - w_{s_1})$
17: updating neighbour n of s_1: $w_n \leftarrow w_n + \epsilon_2(z - w_n)$
18: removing edges with age larger than age_{max} and nodes with no emanating edges
19: **if** t is an integer multiple of parameter λ **then**
20: delete nodes with no neighbour or only one neighbour
21: $age_{max} \leftarrow (\lambda + |\mathcal{N}|)/2$
22: $t \leftarrow t + 1$
23: **until** $epoch = maximum\_epoch$

I. TOPOLOGICAL LEARNING ALGORITHM

We propose Adaptive Topological Learning (ATL) algorithm, a variant of adjusted self-organising incremental neural network (ASOINN) [1] to learn the topological structure of data sets. The pseudo code of ATL is presented in Algorithm 1.

In this algorithm, the similarity thresholds are used as the condition to decide when the network needs to grow. The thresholds can be calculated as follows:

- If the node i has a non-empty set of neighbours NB_i:

$$T_i = \max_{c \in NB_i} ||w_i - w_c|| \tag{1}$$

- If node i has no neighbour:

$$T_i = \min_{c \in \mathcal{N} \setminus \{i\}} ||w_i - w_c|| \tag{2}$$

The adaptive learning rates ϵ_1 and ϵ_2 for a node i can be calculated as follows:

$$\epsilon_1 = \frac{1}{M_i} \tag{3}$$

$$\epsilon_2 = \frac{1}{100 M_i} \tag{4}$$

where M_i is the times for node i to be the winner.

As compared to ASOINN and SOINN, ATL adopts a new adaptive heuristic to update the connections between nodes in $\mathcal{N}$. First, besides increasing the ages of all edges emanating from s_1 by 1, ATL also adds the adaptive term $|NB_n| \frac{d_{z,n}}{\overline{d_{s_1}}}$ (line 15). By using this new updating rule, it is easy to see that the edge connecting s_1 and a node n with more neighbours or with a further distance $d_{z,n}$ will age faster. This trick will help to reduce the complexity of the learned network by eliminating useless edges and nodes. Furthermore, we make age_{max} adaptive to the need to represent the distribution and topological relations of input data (line 21). As a result, ATL only relies on one user-designed parameter λ.

II. DATA SETS AND EXPERIMENTAL SETTINGS

This section provide details of the data sets and the parameters settings for XMAP used in our experiments.

A. Data sets

Table II-A shows 12 data sets used in our experiments with different numbers of examples and numbers of features. These data sets are selected as they have been widely used in the literature for both conventional ML algorithms and IML. As noted in the table, the number of features d reported here are those after the discretisation step. For continuous features such as monthly income in the IBM-HR data set or the customer age in the German Credit Risk data set, we apply k-bins discretiser and one-hot encoding [2] with $k = 4$.

B. Parameter settings and performance metrics

Parameters settings for algorithms used in XMAP are presented in Table II-A. For UMAP, we used the default parameters which produce very robust performance across different data sets. For ATL, we use $\lambda = 200$, as it gives a good balance between speed and performance. For CDA, we

TABLE I
DATASETS USED IN THE EXPERIMENTS

Dataset	N	d **(discretised)**	**Conditions for** $y_i = 1$	$Pr(y_i = 1)$	**Reference**
Heart Disease	303	42	patient has heart disease	54.50%	[3]
Breast Cancer	683	36	diagnosis is malignant	35.00%	[4]
Australian Credit Approval	690	56	credit is approved	44.50%	[5]
Mammo	961	14	person has breast cancer	46.30%	[6]
German Credit Risk	1,000	77	customer has bad credit	30.00%	[7]
Human Resource-IBM	1,470	121	employee attrites	16.10%	[8]
Spambase	4,601	228	e-mail is spam	39.40%	[9]
Customer Churn	7,032	49	customer churns	26.60%	[10]
Mushroom	8,124	113	mushroom is poisonous	48.20%	[11]
Intensive Care Unit	11,773	53	patient is dead	10.70%	[12]
Adult	32,561	36	person in 1994 US census earns over $50	24.10%	[13]
Bank	41,188	57	person opens bank account after marketing call	11.30%	[14]

TABLE II
PARAMETER SETTINGS

Step	**Algorithm**	**Parameters**	**Value**
Mapping	UMAP	neighbour size	15
		negative sample rate	5
		distance metric	Euclidean
Topological Learning	ATL	λ	200
Extracting interpretable contexts	CDA	description size	5
		θ	0.01
Interpretable prediction	Logistic Regression (LR)	inverse regularization	0.1
Other ML algorithms	Decision Tree (DT)	max depth	5
		min split	2
	Artificial Neural Network (ANN)	hidden layer size	15
		max iteration	100
	Extreme Boosting Machine (XGBoost)	max depth	3
		number of estimators	100

do not see a big difference when changing the description size; therefore, we keep the size $S = 5$ for better interpretations. The very small $\theta = 0.01$ is selected to ensure that the context description can efficiently cover the corresponding cluster. In the interpretable prediction step, we only focus on logistic regression with $L1$ regularisation as it is fairly easily to interpret. We also compare the predicion performance of XMAP with other popular algorithms in the literature such as decision trees (DT), artificial neural network (ANN) [15], and XGBoost [16]. The parameters settings for DT, ANN, and XGBoost in Table II-A are selected as these parameters provide a good generalisation in most data sets investigated in this paper based on our pilot experiments. To compare the performance in the binary classification task, we use area under the curve (AUC) and the accuracy (ACC) with the 10-fold cross validation.

REFERENCES

[1] F. Shen and O. Hasegawa, "A fast nearest neighbor classifier based on self-organizing incremental neural network," *Neural Netw.*, vol. 21, no. 10, pp. 1537–1547, Dec. 2008. [Online]. Available: http://dx.doi.org/10.1016/j.neunet.2008.07.001

[2] L. Buitinck, G. Louppe, M. Blondel, F. Pedregosa, A. Mueller, O. Grisel, V. Niculae, P. Prettenhofer, A. Gramfort, J. Grobler, R. Layton, J. VanderPlas, A. Joly, B. Holt, and G. Varoquaux, "API design for machine learning software: experiences from the scikit-learn project," in *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 2013, pp. 108–122.

[3] R. Detrano, A. Jánosi, W. Steinbrunn, M. Pfisterer, J. Schmid, S. K. Sandhu, K. H. Guppy, S. W. Lee, and V. Froelicher, "International application of a new probability algorithm for the diagnosis of coronary artery disease." *The American journal of cardiology*, vol. 64 5, pp. 304–10, 1989.

[4] O. L. Mangasarian, W. N. Street, and W. H. Wolberg, "Breast cancer diagnosis and prognosis via linear programming," *Operations Research*, vol. 43, no. 4, pp. 570–577, 1995.

[5] J. R. Quinlan, "Simplifying decision trees," *Int. J. Man-Mach. Stud.*, vol. 27, no. 3, pp. 221–234, Sep. 1987. [Online]. Available: http://dx.doi.org/10.1016/S0020-7373(87)80053-6

[6] M. Elter, R. Schulz-Wendtland, and T. Wittenberg, "The prediction of breast cancer biopsy outcomes using two cad approaches that both emphasize an intelligible decision process," *Medical Physics*, vol. 34, no. 11, pp. 4164–4172, 2007.

[7] A. Khashman, "Neural networks for credit risk evaluation: Investigation of different neural models and learning schemes," *Expert Syst. Appl.*, vol. 37, no. 9, pp. 6233–6239, Sep. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.eswa.2010.02.101

[8] Kaggle, *IBM HR Analytics Employee Attrition & Performance*, 2019 (accessed November 11, 2019), https://www.kaggle.com/pavansubhasht/ibm-hr-analytics-attrition-dataset.

[9] L. F. Cranor and B. A. LaMacchia, "Spam!" *Commun. ACM*, vol. 41, no. 8, pp. 74–83, Aug. 1998. [Online]. Available: http://doi.acm.org/10.1145/280324.280336

[10] Kaggle, *Telco Customer Churn*, 2019 (accessed November 11, 2019), https://www.kaggle.com/blastchar/telco-customer-churn.

[11] J. C. Schlimmer, "Concept acquisition through representational adjustment," Ph.D. dissertation, 1987, aAI8724747.

[12] A. E. Johnson, J. Aboab, J. Raffa, T. Pollard, R. Deliberato, L. Celi, and D. Stone, "A comparative analysis of sepsis identification methods in an electronic database," *Critical Care Medicine*, vol. 46, no. 4, p. 494–499,

2018.

[13] R. Kohavi, "Scaling up the accuracy of naive-bayes classifiers: A decision-tree hybrid," in *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, ser. KDD'96. AAAI Press, 1996, pp. 202–207. [Online]. Available: http://dl.acm.org/citation.cfm?id=3001460.3001502

[14] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems*, vol. 62, pp. 22–31, 2014.

[15] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, 2nd ed. Springer, 2009. [Online]. Available: http://www-stat.stanford.edu/ tibs/ElemStatLearn/

[16] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: ACM, 2016, pp. 785–794. [Online]. Available: http://doi.acm.org/10.1145/2939672.2939785