COLLABORATIVE CONCEPT DRIFT DETECTION

Anonymous authors
Paper under double-blind review

ABSTRACT

This work draws a comparison with XYZ[1]. Rather than triggering drift detection based on similarities between models-to-features and model-to-model relationships, we propose Collaborative Concept Drift Detection (C2D2) in order to utilize diversity within a heterogeneous ensemble to detect invariant patterns that could relate to drift. C2D2 proposes mapping feature and target spaces into a correlation matrix calculated via FCBF. SVD is used to break down its invariant representations and post-hoc tests are applied to test significance and indicate the occurrence of drift.

1 NEW PERSPECTIVE FOR DRIFT DETECTION

In our earlier work (Doe & Doe)[2], we observed that many authors view **concept drift** (Widmer & Kubat (1996); Lu et al. (2019); Huyen (2022)) as an anomaly acting on a single stream (Gama et al. (2004); Blanco et al. (2015); Sethi & Kantardzic (2017); de Barros et al. (2016)). Such works fail to account for the fact that systems in practice support multiple kinds of models that predict more than one stream. Additionally, in such systems, models learn under a shared data pipeline with the possibility of having overlapping data. Although the models themselves operate independently of each other, and they hold information that describes the relationship between their predictors and respective targets, which may improve the understanding of concepts. Under this premise, our framework, XYZ (Doe & Doe), uses a combination of performance-based (Gama et al. (2004)) and data distribution-based drift detectors (dos Reis et al. (2016)) to act as local drift detectors (*i.e.,* detection per stream). A global concept drift detector is proposed by using collaborative filtering (Sarwar et al. (2001)) to form similarity scores between models via shapley values (Sundararajan & Najmi (2020)) based on their model-to-feature and model-to-model relationships. The advantage being that a drift detected via a local drift detector can pre-emptively detect or save the expense of running another local drift detector for a model recommended as being similar.

2 COLLABORATIVE CONCEPT DRIFT DETECTION (C2D2)

This work draws a comparison with XYZ. Similarly, our notion is to learn from multiple models in order to capture evolving concepts, but rather than triggering drift detection based on similarities between models-to-features and model-to-model relationships, we propose Collaborative Concept Drift Detection (C2D2) in order to utilize diversity (Brown et al. (2005); Tang et al. (2006); Nascimento et al. (2014)) within an ensemble to detect invariant patterns (des Combes et al. (2020); Stojanov et al. (2021)) that could relate to drift.

2.1 REASONING BEHIND DETECTING INVARIANTS

To fit a model is to take a function f onto a feature vector, $\boldsymbol{x}$ such that it provides a systematic estimate of the target variable, y (James et al. (2021)). ϵ is independent of $\boldsymbol{x}$ and represents random error, a combination of **irreducible error** and **reducible** error (1).

$$y = f(\boldsymbol{x}) + \epsilon \tag{1}$$

We speculate that ϵ of $f(\boldsymbol{x})$ can serve as a predictor for another model, $f'(\boldsymbol{x})$. In other words, ϵ is independent of $\boldsymbol{x}$, but is related to $\boldsymbol{x'}$ or y. The underlying assumption is that as a concept

[1]XYZ is the authors' previous work and is redacted to respect the blind reviewing
[2]All references under this is a placeholder to respect the blind reviewing

evolves, predictors that turn into confounding variables (i.e. predictors that affect both the target and other predictor(s)) are related to the source of drift. Additionally, there is the assumption that such confounding variables will be invariant across an ensemble of classifiers with heterogeneous training and prediction subspaces and can be used to indicate the occurrence of drift.

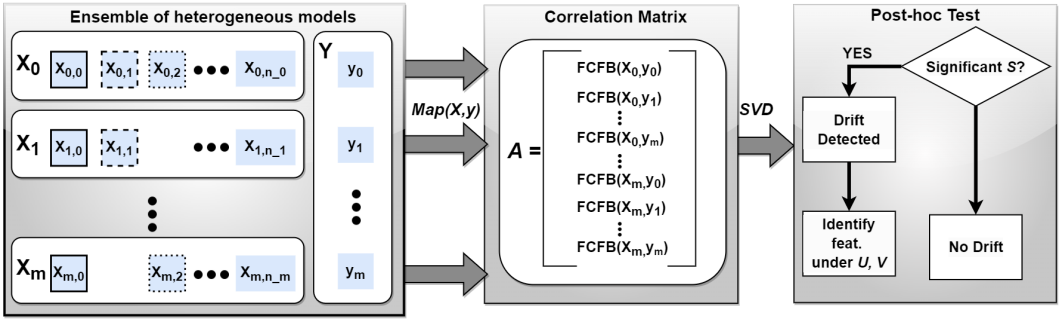

Figure 1: Proposed C2D2 framework. Each row $\boldsymbol{x}_i$ dictates a feature space to predict a respective y_j. Solid or dashed lines within the feature space indicate shared predictors while gaps indicate the absence of features. The feature and target spaces are mapped $m : n$ and their correlations are calculated via FCBF. SVD is used to break down its invariant representations and post-hoc tests are applied to test significance and indicate the occurrence of drift.

3 METHODOLOGY

Our work incorporates the methods of domain adaptation (Nguyen et al. (2021); des Combes et al. (2020)) to engender a context of $\boldsymbol{x}$ on $y \in \mathbb{Y}$ and homogeneous ensembles to provide the context of $\boldsymbol{x} \in \mathbb{X}$ on a specific target y. Incorporating both creates heterogeneous ensembling, where models are trained on a mix of $\mathbb{X}$ and $\mathbb{Y}$ subspaces (2.1)

The correlation of every combination of $\boldsymbol{x}$ within $\mathbb{X}$ mapped to every $\boldsymbol{y}$ in $\mathbb{Y}$ is calculated by Fast Correlation Based Filtering (FCBF) (Nguyen et al. (2012)) to form a correlation matrix, $\boldsymbol{A}$ for every data instance. Singular Value Decomposition (SVD) (2) is then used to reduce the matrix to its invariant representation, representing features that may have drifted. At windows of 30 instances, a post-hoc test on $\boldsymbol{S}$ (Demsar (2006)) is to be followed to indicate *if* drift has occurred with $\boldsymbol{U}$ and $\boldsymbol{V}$ indicating *where* it occurred.

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^T \tag{2}$$

4 ANALYSIS AND CONCLUSION

Both frameworks rely on overlapping feature spaces and predictors for systems with many models. A limitation of XYZ is that its similarity scores are calculated via shapley values and the complexity explodes in complexity to $O(2^k)$, where k is the entire feature size across all models. Additionally, it has an inherent trade-off between incurring high false alarms and long delays in detection. With C2D2, however, the complexity is $O(mn)$ and its constraint lies in its ability to identify how diverse an ensemble needs to be in order to isolate the invariants to be related only to drift.

C2D2 is arguably orthogonal in reasoning to XYZ. Instead of basing its detection on similarities between the relationship of models and their shared features, it relies on the diversity of models so that performing decomposition transformations could reveal invariant representations of confounding variables related to concept drift. The idea being that if the source of drift is related to predictors evolving into confounding variables, then it should be capturable within the ensemble feature space.

ACKNOWLEDGEMENTS

This work was supported by Graduate Assistance in Areas of National Need (GAANN) under award [redacted] and by the National Science Foundation under award [redacted].

URM STATEMENT

The authors acknowledge that at least one key author of this work meets the URM criteria of ICLR 2023 Tiny Papers Track.

REFERENCES

Isvani Inocencio Frías Blanco, José del Campo-Ávila, Gonzalo Ramos-Jiménez, Rafael Morales Bueno, Agustín Alejandro Ortiz Díaz, and Yailé Caballero Mota. Online and non-parametric drift detection methods based on hoeffding's bounds. *IEEE Trans. Knowl. Data Eng.*, 27(3):810–823, 2015. doi: 10.1109/TKDE.2014.2345382. URL https://doi.org/10.1109/TKDE.2014.2345382.

Gavin Brown, Jeremy L. Wyatt, Rachel Harris, and Xin Yao. Diversity creation methods: a survey and categorisation. *Inf. Fusion*, 6(1):5–20, 2005. doi: 10.1016/j.inffus.2004.04.004. URL https://doi.org/10.1016/j.inffus.2004.04.004.

Roberto Souto Maior de Barros, Silas Garrido Teixeira de Carvalho Santos, and Paulo Mauricio Gonçalves Júnior. A boosting-like online learning ensemble. In *2016 International Joint Conference on Neural Networks, IJCNN 2016, Vancouver, BC, Canada, July 24-29, 2016*, pp. 1871–1878. IEEE, 2016. doi: 10.1109/IJCNN.2016.7727427. URL https://doi.org/10.1109/IJCNN.2016.7727427.

Janez Demsar. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.*, 7:1–30, 2006. URL http://jmlr.org/papers/v7/demsar06a.html.

Remi Tachet des Combes, Han Zhao, Yu-Xiang Wang, and Geoffrey J. Gordon. Domain adaptation with conditional distribution matching and generalized label shift. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/dfbfa7ddcfffeb581f50edcf9a0204bb-Abstract.html.

John Doe and Jane Doe. Reference removed to respect blind reviewing. In John Doe (ed.), *Reference removed to respect blind reviewing,*.

Denis Moreira dos Reis, Peter A. Flach, Stan Matwin, and Gustavo E. A. P. A. Batista. Fast unsupervised online drift detection using incremental kolmogorov-smirnov test. In Balaji Krishnapuram, Mohak Shah, Alexander J. Smola, Charu C. Aggarwal, Dou Shen, and Rajeev Rastogi (eds.), *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 13-17, 2016*, pp. 1545–1554. ACM, 2016. doi: 10.1145/2939672.2939836. URL https://doi.org/10.1145/2939672.2939836.

João Gama, Pedro Medas, Gladys Castillo, and Pedro Pereira Rodrigues. Learning with drift detection. In Ana L. C. Bazzan and Sofiane Labidi (eds.), *Advances in Artificial Intelligence - SBIA 2004, 17th Brazilian Symposium on Artificial Intelligence, São Luis, Maranhão, Brazil, September 29 - October 1, 2004, Proceedings*, volume 3171 of *Lecture Notes in Computer Science*, pp. 286–295. Springer, 2004. doi: 10.1007/978-3-540-28645-5_29. URL https://doi.org/10.1007/978-3-540-28645-5_29.

Chip Huyen. Data distribution shifts and monitoring. In *Designing machine learning systems*. O'Reilly, 2022.

Gareth James, Daniela Witten, Trevor Hastie, and Robert Tibshirani. An introduction to statistical learning: With applications in r. Springer, 2021.

Jie Lu, Anjin Liu, Fan Dong, Feng Gu, João Gama, and Guangquan Zhang. Learning under concept drift: A review. *IEEE Trans. Knowl. Data Eng.*, 31(12):2346–2363, 2019. doi: 10.1109/TKDE.2018.2876857. URL https://doi.org/10.1109/TKDE.2018.2876857.

Diego Silveira Costa Nascimento, André L. V. Coelho, and Anne M. P. Canuto. Integrating complementary techniques for promoting diversity in classifier ensembles: A systematic study. *Neurocomputing*, 138:347–357, 2014. doi: 10.1016/j.neucom.2014.01.027. URL https://doi.org/10.1016/j.neucom.2014.01.027.

A. Tuan Nguyen, Toan Tran, Yarin Gal, and Atilim Gunes Baydin. Domain invariant representation learning with domain density transformations. In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 5264–5275, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/2a2717956118b4d223ceca17ce3865e2-Abstract.html.

Hai-Long Nguyen, Yew-Kwong Woon, Wee Keong Ng, and Li Wan. Heterogeneous ensemble for feature drifts in data streams. In Pang-Ning Tan, Sanjay Chawla, Chin Kuan Ho, and James Bailey (eds.), *Advances in Knowledge Discovery and Data Mining - 16th Pacific-Asia Conference, PAKDD 2012, Kuala Lumpur, Malaysia, May 29 - June 1, 2012, Proceedings, Part II*, volume 7302 of *Lecture Notes in Computer Science*, pp. 1–12. Springer, 2012. doi: 10.1007/978-3-642-30220-6_1. URL https://doi.org/10.1007/978-3-642-30220-6_1.

Badrul Munir Sarwar, George Karypis, Joseph A. Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In Vincent Y. Shen, Nobuo Saito, Michael R. Lyu, and Mary Ellen Zurko (eds.), *Proceedings of the Tenth International World Wide Web Conference, WWW 10, Hong Kong, China, May 1-5, 2001*, pp. 285–295. ACM, 2001. doi: 10.1145/371920.372071. URL https://doi.org/10.1145/371920.372071.

Tegjyot Singh Sethi and Mehmed M. Kantardzic. On the reliable detection of concept drift from streaming unlabeled data. *Expert Syst. Appl.*, 82:77–99, 2017. doi: 10.1016/j.eswa.2017.04.008. URL https://doi.org/10.1016/j.eswa.2017.04.008.

Petar Stojanov, Zijian Li, Mingming Gong, Ruichu Cai, Jaime G. Carbonell, and Kun Zhang. Domain adaptation with invariant representation learning: What transformations to learn? In Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pp. 24791–24803, 2021. URL https://proceedings.neurips.cc/paper/2021/hash/cfc5d9422f0c8f8ad796711102dbe32b-Abstract.html.

Mukund Sundararajan and Amir Najmi. The many shapley values for model explanation. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pp. 9269–9278. PMLR, 2020. URL http://proceedings.mlr.press/v119/sundararajan20b.html.

E. Ke Tang, Ponnuthurai N. Suganthan, and Xin Yao. An analysis of diversity measures. *Mach. Learn.*, 65(1):247–271, 2006. doi: 10.1007/s10994-006-9449-2. URL https://doi.org/10.1007/s10994-006-9449-2.

Gerhard Widmer and Miroslav Kubat. Learning in the presence of concept drift and hidden contexts. *Mach. Learn.*, 23(1):69–101, 1996. doi: 10.1007/BF00116900. URL https://doi.org/10.1007/BF00116900.