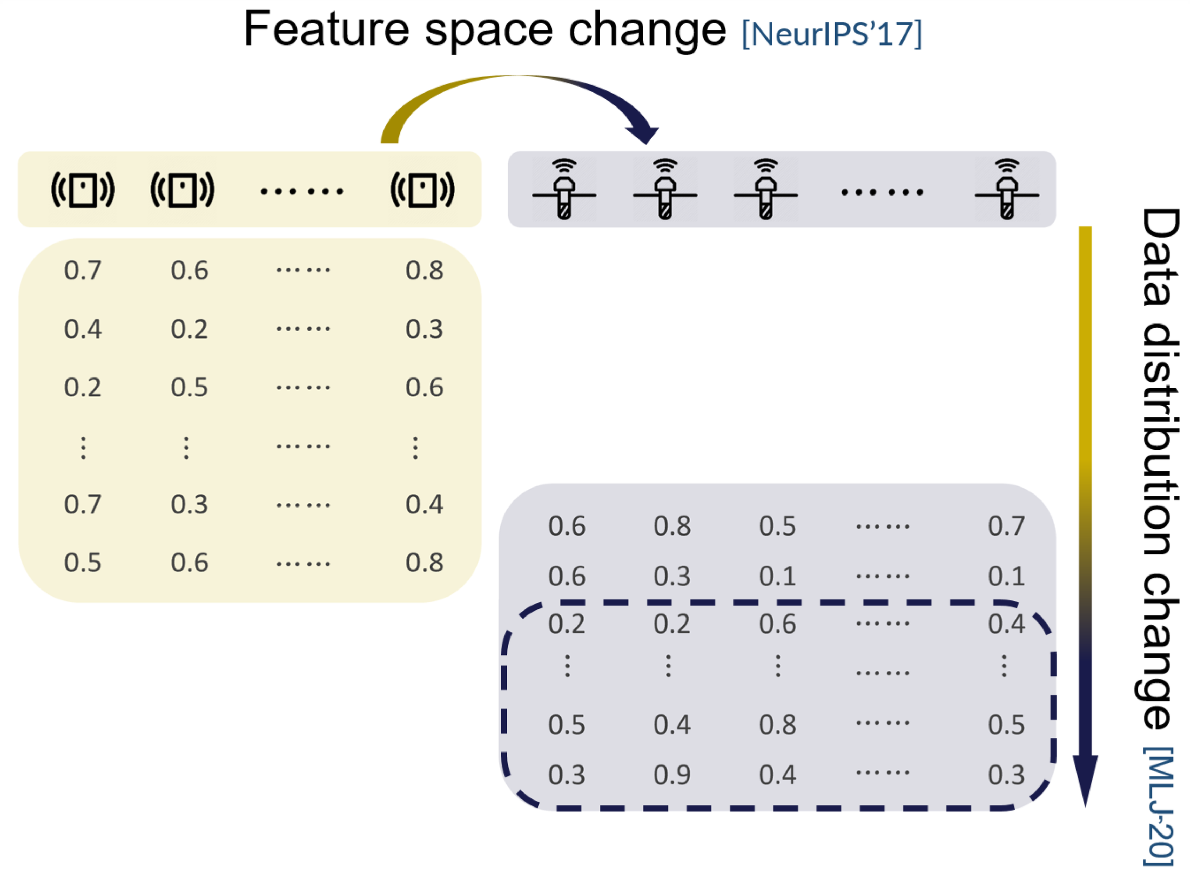
Learning with Feature and Distribution Evolvable Streams

Zhen-Yu Zhang, Peng Zhao, Yuan Jiang, Zhi-Hua Zhou

{zhangzy, zhaop, jiangy, zhouzh}@lamda.nju.edu.cn

ICML | 2020

Feature and Distribution Evolvable Streams



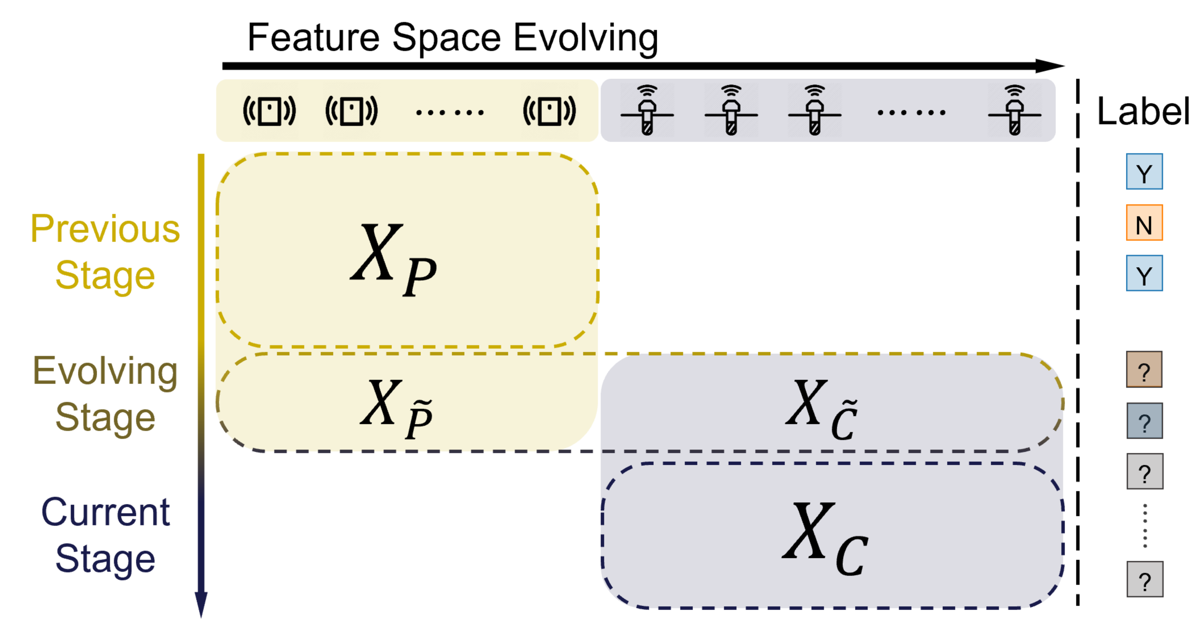
[NeurIPS'17] Hou, B.-J., Zhang, L., & Zhou, Z.-H. (2017). Learning with feature evolvable streams.

[MLJ'20] Zhao, P., Cai, L.-W., & Zhou, Z.-H. (2020). Handling concept drift via model reuse.

Formulation

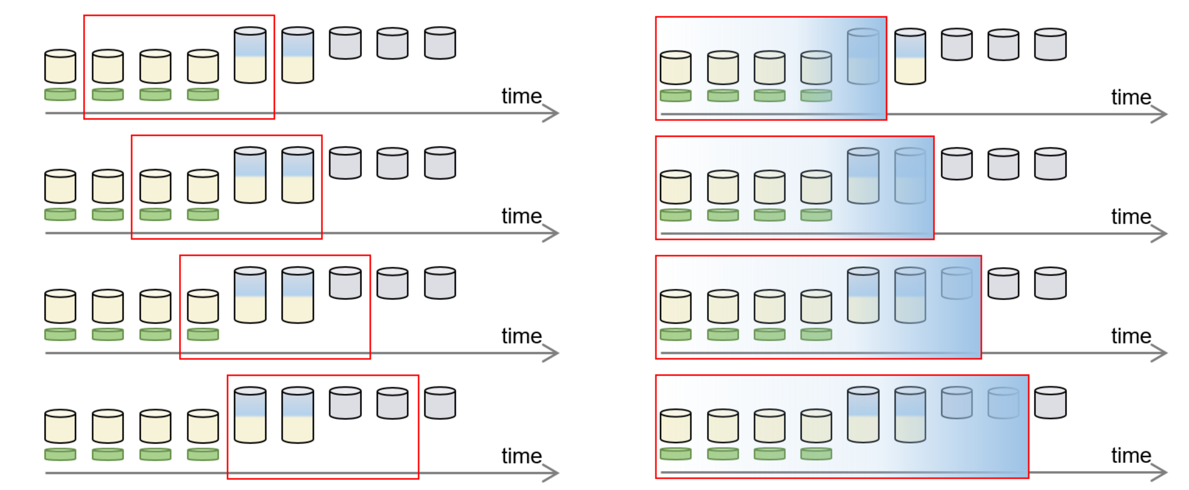
Feature and Distribution Evolving Stream Learning (FDESL)

- previous stage (labeled)
- evolving stage (feature space and distribution changing)
- current stage (unlabeled)



Challenge

Can we use Sliding Window or Forgetting Factor?



Sliding Window — Forgetting factor

Drawbacks

- hard to handle distribution change
- hard to analyze generalization ability

Evolving Discrepancy & Generalization Bound

Idea: Measure Discrepancy of data from different feature spaces

Bridge the gap via the evolving stage

- data in the evolving stage share the same labels
- admissible loss aligns the hypotheses via the evolving stage

$$disc''_E(S_P, S_C) = \sup_{g\in\mathcal{G}, h\in\mathcal{H}} \underbrace{\left|\hat{R}_{S_P}(g, y_P) - \hat{R}_{S_{\tilde{P}}}(g, y_{\tilde{P}})\right|}_{\text{Previous Stage to Evolving Stage}} + \underbrace{\left|\hat{R}_{S_{\tilde{C}}}(h, y_{\tilde{C}}) - \hat{R}_{S_C}(h, y_C)\right|}_{\text{Evolving Stage to Current Stage}} + \underbrace{\sigma d_1(g, h)}_{\text{Hypotheses Alignment}}$$

Generalization bound

- The expected risk on the current data batch is upper bounded by empirical risk & evolving discrepancy

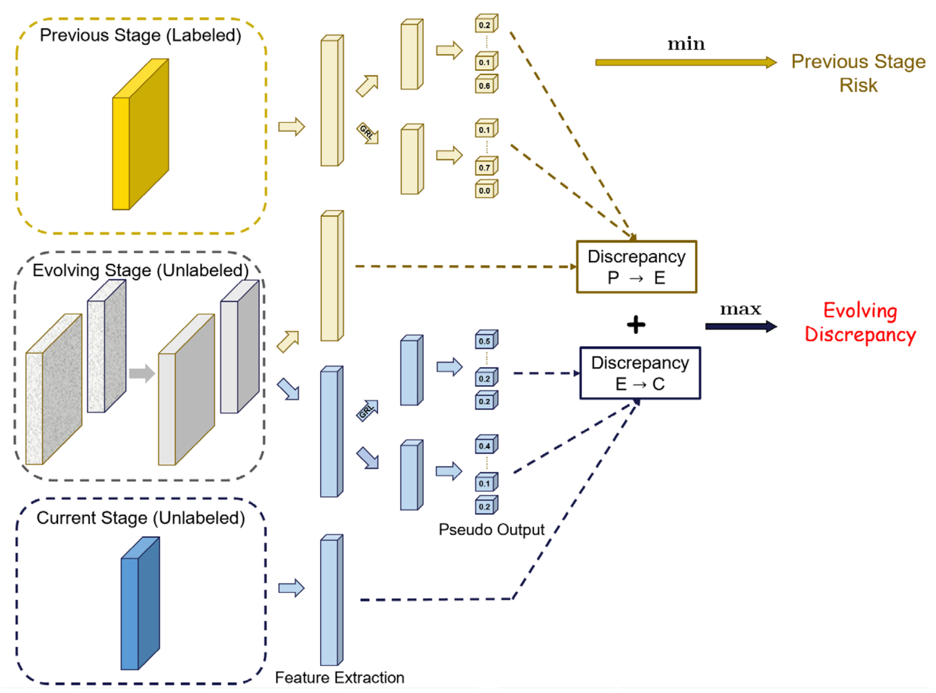
$$R_{\mathcal{D}_C}(h, f_C) \le \boxed{\hat{R}_{S_{P_\alpha}}(g, y_P)} + \boxed{disc_E(S_P, S_C)} + 2L\mathfrak{R}_n(\mathcal{H}) + M_C\sqrt{\frac{\log(1/\delta)}{2n}}$$

Remarks

- Evolving discrepancy measures the discrepancy between two consecutive batches (with different feature space & distribution) via the evolving stage
- So we can analyze the generalization ability for the FDESL problem

Deep Neural Network Implementation

- estimate weights $\alpha \in \mathbb{R}^m, \beta \in \mathbb{R}^n$ [TCS'14]
- solve the minimax optimization



Adversarial network [JMLR'16] based EDM Framework

[TCS'14] Cortes, C., & Mohri, M. (2014). Domain adaptation and sample bias correction theory and algorithm for regression.

[JMLR'16] Ganin, Yaroslav, et al. (2016). Domain-adversarial training of neural networks.

Experiments

Synthetic data

- the evolving discrepancy reflects the relation

Real-world applications

- the EDM algorithm solves the FDESL problem

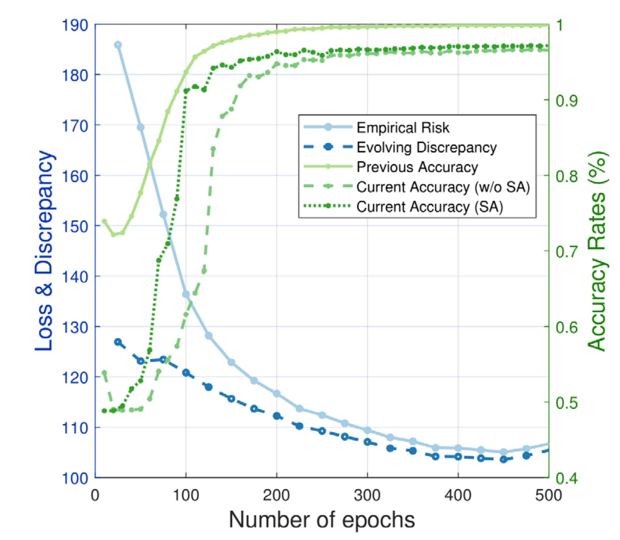
Evolving discrepancy verification

- minimizing the evolving discrepancy improves the average accuracy

Methods	RFID	A.Books	A.Movies	A.CDs
FESL	77.39 ± 2.5	70.53 ± 4.7	67.30 ± 3.6	61.79 ± 3.4
TSIW	91.34 ± 1.1	73.83 ± 2.1	72.61 ± 2.0	63.93 ± 0.7
EDM	**93.32 ± 1.2**	**77.97 ± 5.2**	**76.16 ± 1.8**	**69.47 ± 2.5**

Methods	EN-FR	FR-SP	GR-IT	IT-GR
FESL	78.51 ± 1.9	73.64 ± 2.6	75.12 ± 1.4	77.96 ± 0.9
TSIW	84.42 ± 1.8	79.43 ± 2.3	81.92 ± 4.4	82.30 ± 2.7
EDM	**86.74 ± 0.7**	**80.72 ± 1.4**	**85.40 ± 3.9**	**84.84 ± 2.7**

Experiments on real-world applications



Evolving discrepancy verification

Conclusion

- ♠ We formulate the Feature and Distribution Evolving Stream Learning (FDESL) problem, which accommodates a variety of real-world applications
- ♠ We characterize the FDESL problem by evolving discrepancy and derive the generalization ability analysis
- ♠ We propose the Evolving Discrepancy Minimization (EDM) algorithm and validates the effectiveness on synthetic and real-world data