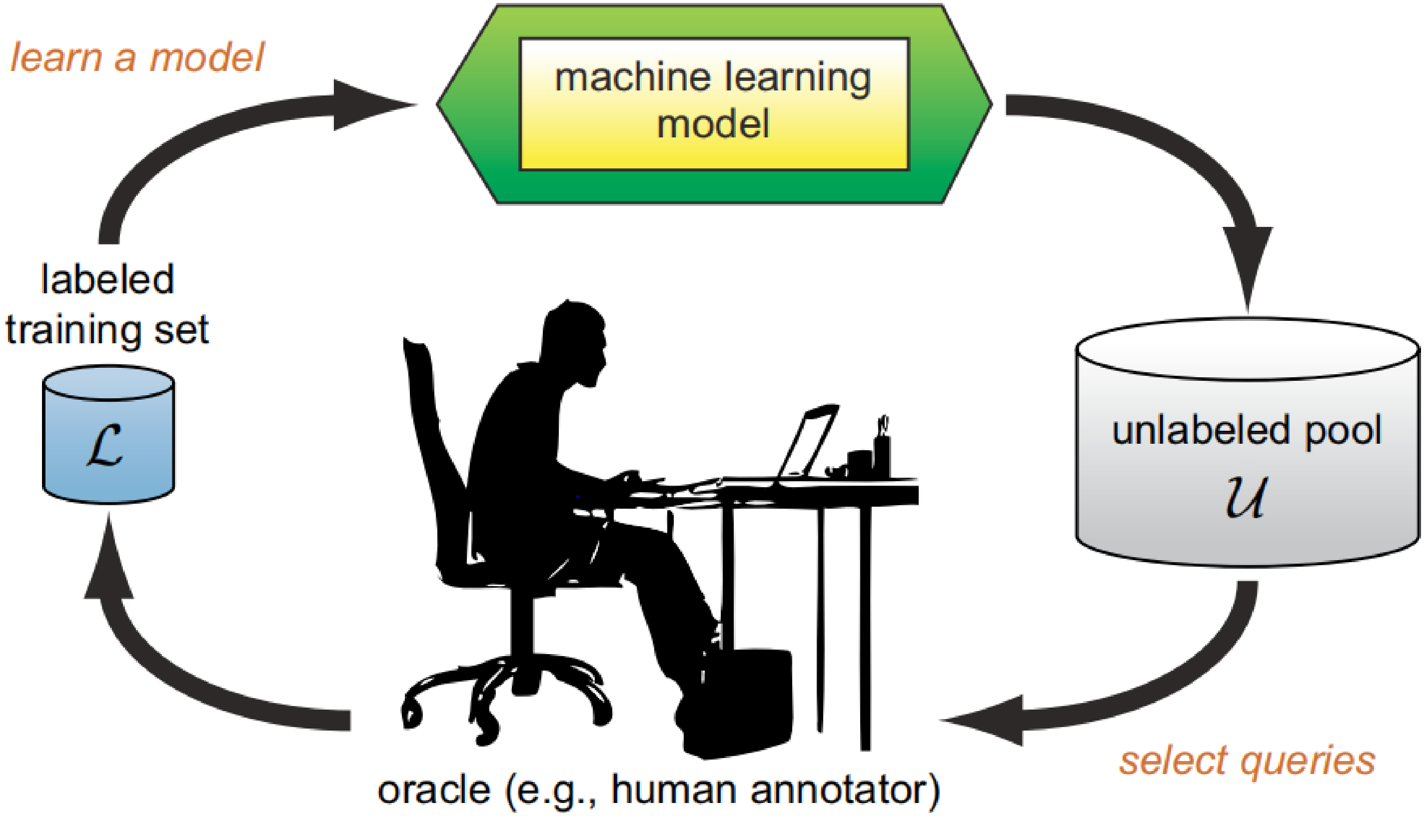
Meta-Learning Transferable Active Learning Policies by Deep Reinforcement Learning

arXiv:1806.04798

Outline

- Introduction
  - Active Learning
  - Reinforcement Learning
- Methods
  - Policy Network
  - Meta Network
  - Architecture
- Experiments Results

Active Learning



Query Strategy:

Uncertainty Sampling

Query-By-Committee

Expected Model Change

Expected Error Reduction

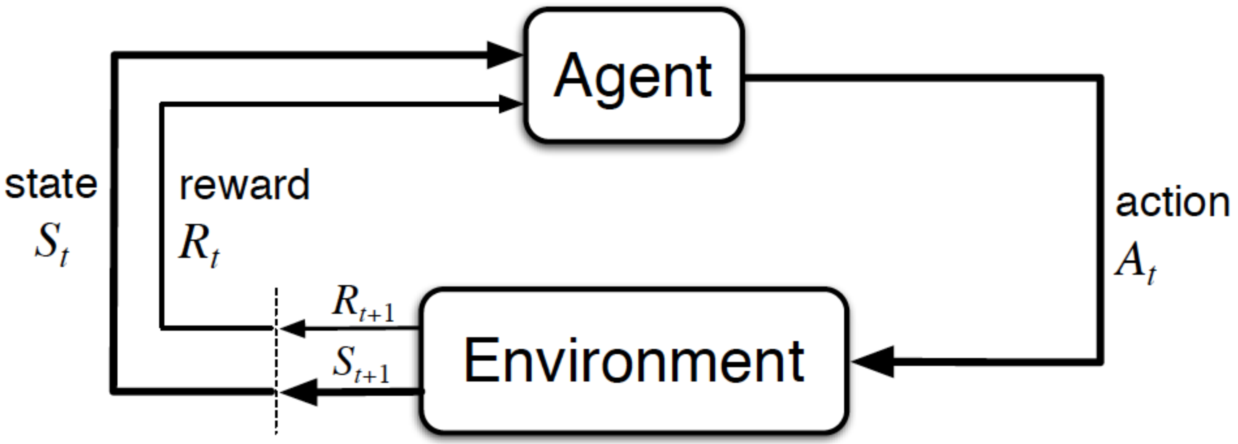
Variance Reduction

Density-Weighted Methods

QUIRE

random sampling

Reinforcement Learning



model-free reinforcement learning

State : $s_t = \{\mathcal{L}_t, \mathcal{U}_t, f_t\}$

Action : $\{1, \ldots, |\mathcal{U}_t|\}$

Policy : $\pi(a_i | s_t)$

Reward : $J = \sum_{t=1}^{\infty} \gamma^{t-1} r_t$

- It is natural to represent AL as a sequential decision making problem
- Discovery of the ideal criterion as an RL problem
- Model an AL criterion as a neural network
- Treat the underlying learner as part of the environment

Methods

Challenge:

1. How to learn a criterion that generalizes?
2. The test and training dataset statistics may differ , and moreover different datasets have different feature dimensionality d

Solution:

- Policy Network

  Input : $\boldsymbol{Z}_u \in \mathbb{R}^{N \times d}$(Z = [X, ξ(X)]) Output: N-way softmax

- Mate Network

  Input : $(\mathcal{L}_t, \mathcal{U}_t, f_t)$ Output: $\boldsymbol{W}_e \in \mathbb{R}^{d \times k}$ $\boldsymbol{W}_d \in \mathbb{R}^{k \times d}$

Meta Network

- **Function：** By synthesising dataset-conditional weight matricies based on dataset-embeddings of $\boldsymbol{Z}^T$

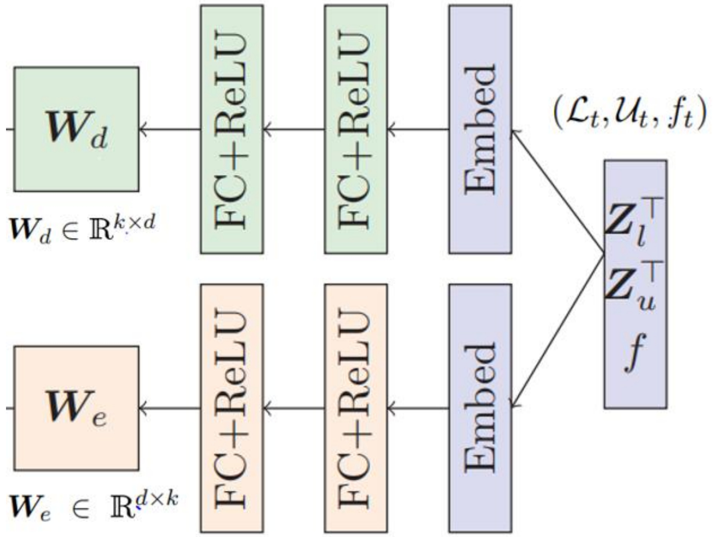
$$\boldsymbol{W}_e \in \mathbb{R}^{d\times k} \qquad \Psi_{\theta_m^e} : \{(\mathcal{L}_t, \mathcal{U}_t, f_t) \rightarrow \boldsymbol{W}_e; \theta_m^e\}$$

$$\boldsymbol{W}_d \in \mathbb{R}^{k\times d} \qquad \Psi_{\theta_m^d} : \{(\mathcal{L}_t, \mathcal{U}_t, f_t) \rightarrow \boldsymbol{W}_d; \theta_m^d\}$$

'representative' and 'discriminative' histogram embedding

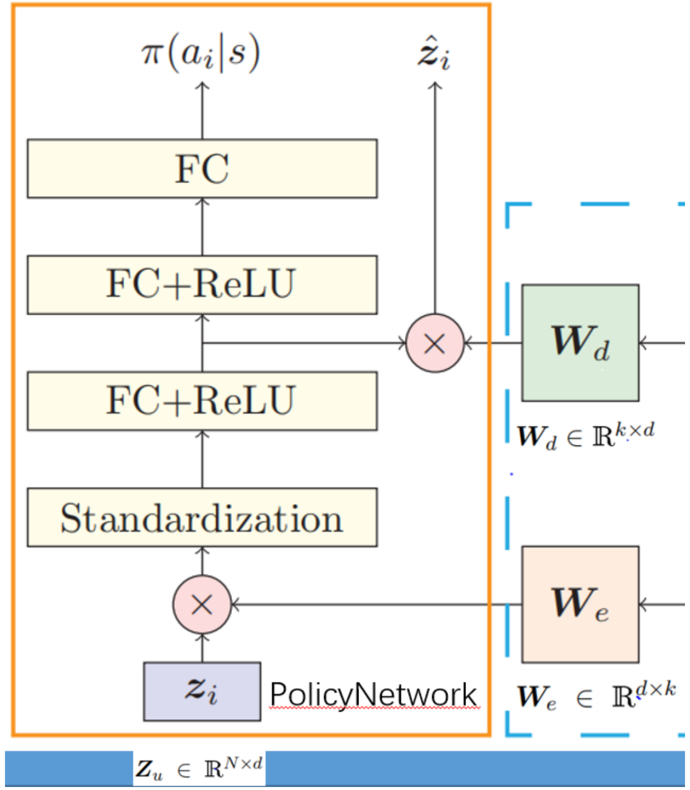
$$(\boldsymbol{e}_j^1(\boldsymbol{Z}_u^T) \text{ and } \boldsymbol{e}_j^1(\boldsymbol{Z}_l^T)) \qquad (\boldsymbol{e}_j^2([\boldsymbol{Z}_u^T, \boldsymbol{Z}_l^T], f_t)):$$

$$(\boldsymbol{W}_e)_j = \Psi\Big([\boldsymbol{e}_j^1(\boldsymbol{Z}_u^T), \boldsymbol{e}_j^1(\boldsymbol{Z}_l^T), \boldsymbol{e}_j^2([\boldsymbol{Z}_u^T, \boldsymbol{Z}_l^T], f_t)]\Big).$$



Policy Network

- Input: the N currently unlabeled instances $\boldsymbol{Z}_u \in \mathbb{R}^{N \times d}$
- Output: an N-way softmax for selecting the instance to query



- selects actions via the softmax $\pi(a_i|s_t) \propto \exp^{\Phi_{\theta_p}(\boldsymbol{W}_e^T \boldsymbol{z}_i)}$

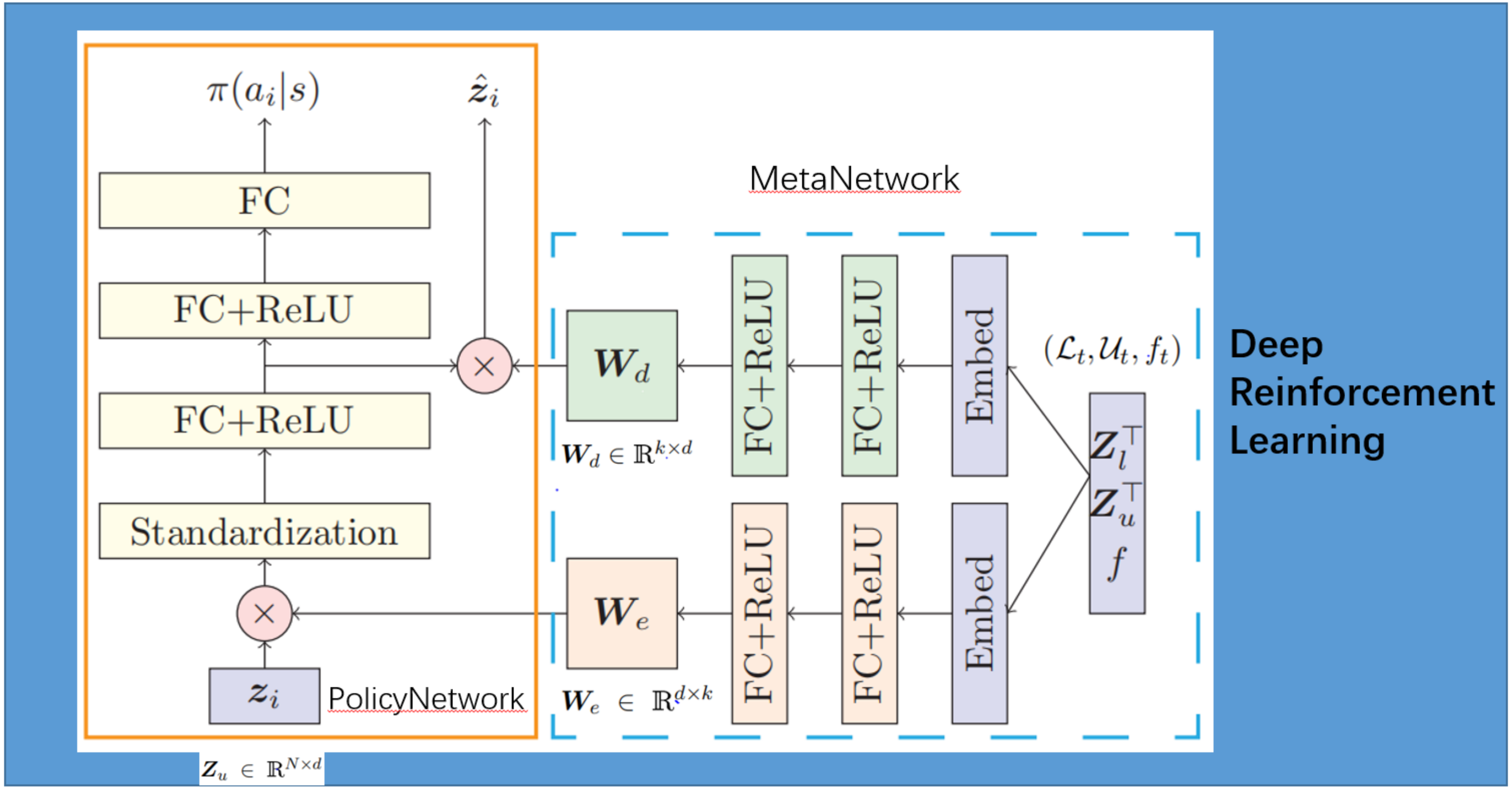
$\boldsymbol{z}_i \in \mathbb{R}^d$ is the ith unlabelled instance in $\boldsymbol{Z}_u$

$\boldsymbol{W}_e \in \mathbb{R}^{d \times k}$ are dataset dependent weights

$$\boldsymbol{u}_i = \boldsymbol{W}_e^T \boldsymbol{z}_i \in \mathbb{R}^k$$

$\pi(a_i|s_t) \propto \exp^{\Phi_{\theta_p}(\boldsymbol{u}_i)}$ is independent of d

Architecture



$$\mathcal{L}_{t+1} = \mathcal{L}_t \cup \{(\boldsymbol{x}^{(t)}, y^{(t)})\}, \mathcal{U}_{t+1} = \mathcal{U}_t \setminus \{\boldsymbol{x}^{(t)}\}$$

$$F = J_\theta(\Phi) - \lambda_1 A_{\theta_m^d}(\boldsymbol{Z}_u) + \lambda_2 \mathcal{H}(\pi_\theta(\boldsymbol{a}|\boldsymbol{Z}_u))$$

Objective Function

- Reward： the improvement in test split accuracy

$$J(\theta) = \mathbb{E}[\sum_{t=1}^{\infty} \gamma^{t-1} r_t(\dot{s}, \pi_\theta(\cdot, s))] \qquad r_t = Acc_t - Acc_{t-1}$$

- Auxiliary Regularisation Losses :

Reconstruction： $A(\boldsymbol{Z}_u) = |\boldsymbol{Z}_u - \hat{\boldsymbol{Z}}_u|_F$

Entropy： $\mathcal{H}(\pi_\theta(\boldsymbol{a}|\boldsymbol{Z}_u))$

- Train both networks where end-to-end, maximising $\theta = \{\theta_p, \theta_m^e\}$ in:

$$F = J_\theta(\Phi) - \lambda_1 A_{\theta_m^d}(\boldsymbol{Z}_u) + \lambda_2 \mathcal{H}(\pi_\theta(\boldsymbol{a}|\boldsymbol{Z}_u))$$

Experiments Results

Table 1: Comparison of AL algorithms, leave one dataset out setting. Linear SVM base learner. AUC averages (%) over 100 trials (and 13 training occurrences for MLP-GAL (Tr)).

Liinear	MLP-GAL (Tr)	MLP-GAL (Te)	SingleRL (Te)	Uncertainty	DFF	RAND	ALBL	T-LSA	QUIRE	QBB
austra	80.14	78.09	75.72	78.24	75.63	75.87	75.31	72.98	64.46	**78.58**
breast	96.67	95.95	94.78	95.41	95.76	94.71	95.67	**96.21**	95.60	95.73
diabetes	67.53	**65.99**	64.78	64.18	57.31	64.05	61.35	57.34	53.75	64.46
fertility	78.26	75.09	**77.86**	75.79	70.44	71.28	66.92	71.18	54.93	73.87
fourclass	74.79	**74.11**	71.83	69.55	71.26	69.08	68.69	69.98	64.48	70.81
haberman	67.31	**65.61**	64.91	60.16	60.26	57.40	52.49	59.67	45.89	60.58
heart	76.68	72.77	72.84	73.38	**73.99**	73.06	71.78	71.52	67.07	73.36
german	68.01	**64.68**	63.35	63.34	61.78	62.77	61.74	58.75	51.82	64.16
ILPD	62.48	59.30	**61.08**	57.60	50.97	57.62	52.91	53.15	48.57	56.77
ionospheres	74.96	**71.46**	69.78	70.47	59.64	69.81	68.44	58.95	57.84	70.40
liver	55.66	55.51	**55.62**	53.45	52.87	52.87	51.25	51.36	48.11	52.13
pima	67.64	**67.01**	64.67	64.18	57.31	63.69	61.27	57.03	53.75	64.24
planning	60.74	**58.63**	56.75	55.09	52.77	54.17	49.46	52.04	39.90	55.43
wdbc	90.90	90.09	88.72	**90.93**	87.55	88.52	88.41	85.15	82.17	90.68
Avg	72.98	**70.94**	70.19	69.41	66.25	68.21	66.12	65.38	59.17	69.37
Num Wins	-	**7**	3	1	1	0	0	1	0	1

Experiments Results