Learning How to Active Learn by Dreaming

Thuy-Trang Vu
Faculty of Information Technology
Monash University, Australia
trang.vuthithuy@monash.edu

Ming Liu
School of Info Technology
Deakin University, Australia
m.liu@deakin.edu

Dinh Phung
Faculty of Information Technology
Monash University, Australia
dinh.phung@monash.edu

Gholamreza Haffari
Faculty of Information Technology
Monash University, Australia
gholamreza.haffari@monash.edu

ACL 2019

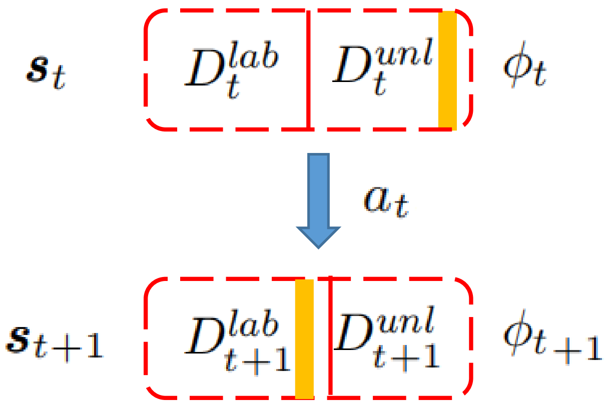
AL Query Strategy as a Sequential Decision Process

- A pool-based AL problem is a Markov decision process (MDP)

$$(S, A, Pr(\boldsymbol{s}_{t+1}|\boldsymbol{s}_t, a_t), R)$$

$\boldsymbol{s}_t \in S \longrightarrow (D_t^{lab}, D_t^{unl}, \phi_t)$

$a_t \in A \longrightarrow$ the selection of a query datapoint.

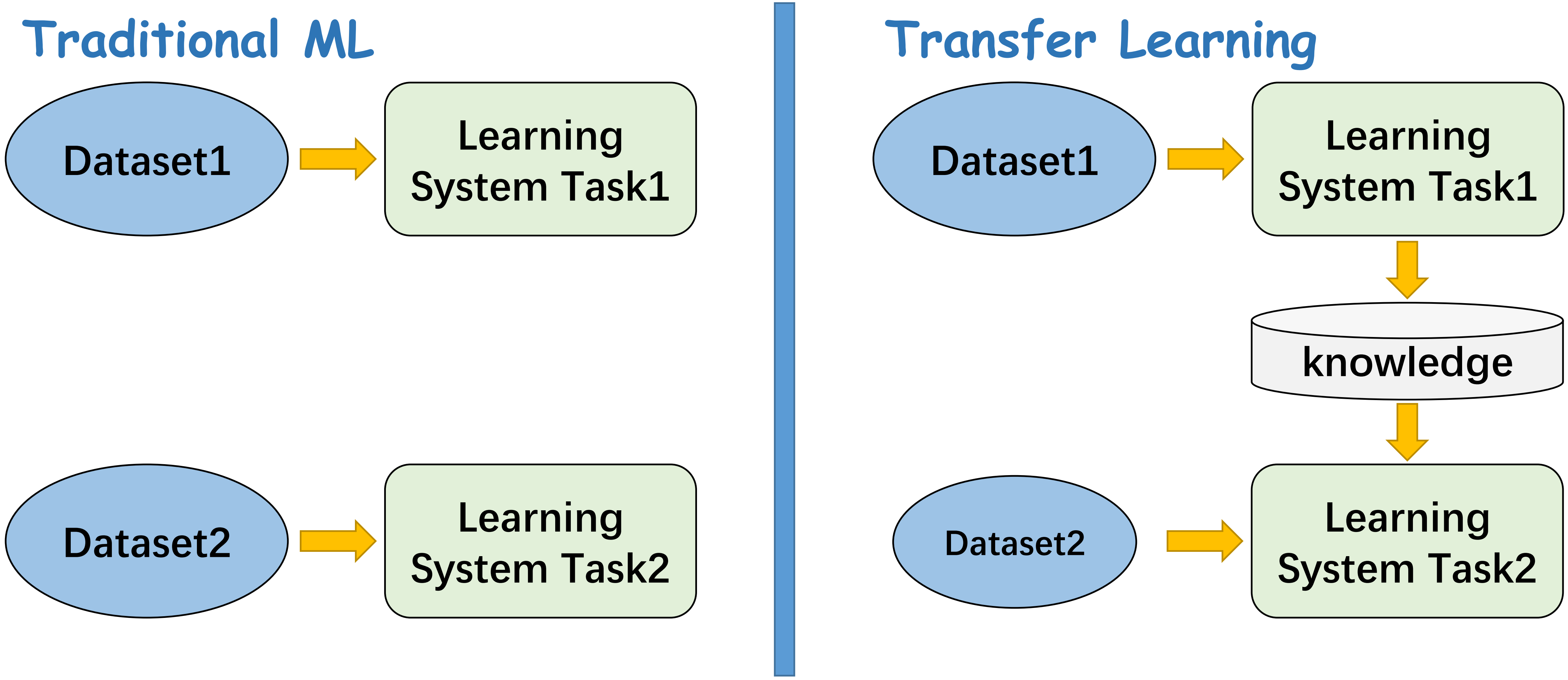


$$\mathrm{R}(\boldsymbol{s}_t, a_t, \boldsymbol{s}_{t+1}) = \mathrm{loss}(m_{\boldsymbol{\phi}_{t-1}}, D^{evl}) - \mathrm{loss}(m_{\boldsymbol{\phi}_t}, D^{evl})$$

Learn the Optimal Query Policy

$$\mathbb{E}_{(D^{lab}, D^{unl}, D^{evl}) \sim \mathcal{D}} \left[\mathbb{E}_{\pi_{\boldsymbol{\theta}}} \left[\sum_{t=1}^{\mathcal{B}} R(\boldsymbol{s}_t, a_t, \boldsymbol{s}_{t+1}) \right] \right]$$

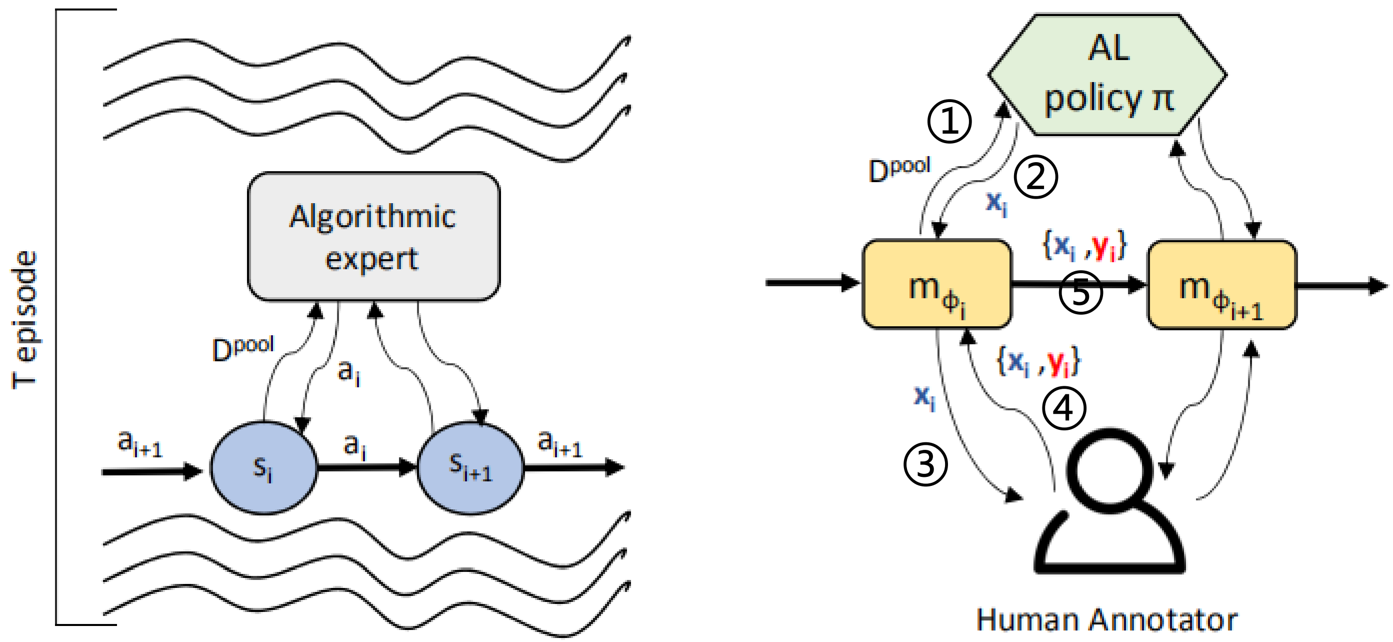
Introduction



The success of the transferred AL strategy highly depends on the relatedness of the source and target AL problems.

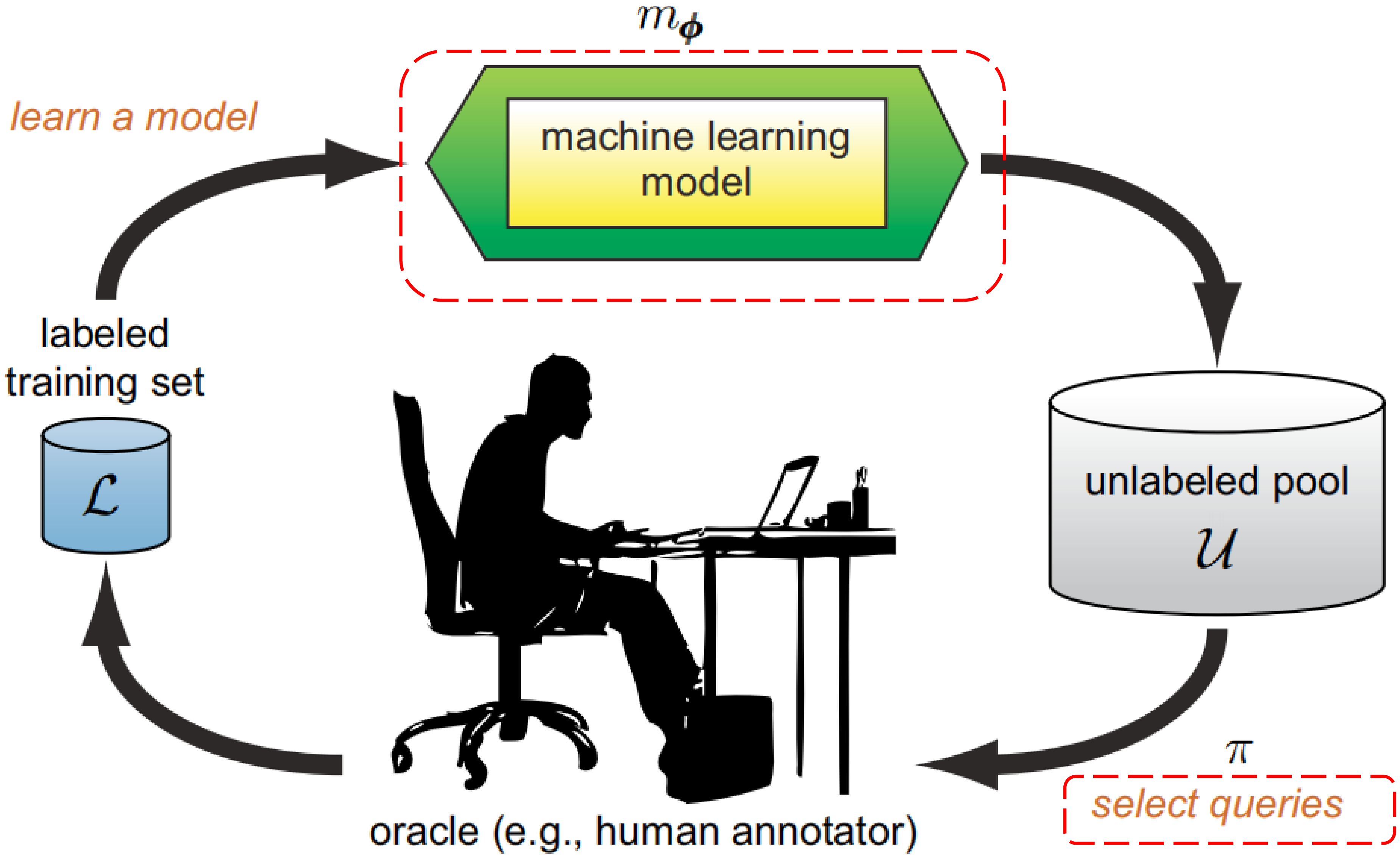
Method

Learn the AL policy directly on the target domain of interest by using wake and dream cycles.



(a) Train AL policy on source task (b) Wake (AL) phase (c) Dream phase

Method



m_{ϕ}
learn a model
machine learning model
labeled training set
$\mathcal{L}$
unlabeled pool
$\mathcal{U}$
π
select queries
oracle (e.g., human annotator)

Algorithm

Algorithm 1 Learning to AL by Dreaming

Input: labelled data D^{lab}, unlabelled pool D^{unl}, initial student model $\hat{\boldsymbol{\phi}}$, initial policy $\hat{\pi}$, dream episodes $\mathcal{E}$, dream length T_d , annotation budget $\mathcal{B}$, wake-dream cycles $\mathcal{W}$

Output: labelled dataset, trained model, policy

$\boldsymbol{\phi}_0 \leftarrow \hat{\boldsymbol{\phi}}$

$\pi_0 \leftarrow \hat{\pi}$

$T_w \leftarrow \frac{\mathcal{B}}{\mathcal{W}}$ ▷ length of the wake phase

for $t \in 1, \ldots, \mathcal{W}$ **do**

  $D^{lab}, \boldsymbol{\phi}_t \leftarrow \text{wakeLearn}(D^{lab}, D^{unl}, \boldsymbol{\phi}_{t-1}, \pi_{t-1}, T_w)$

  $\pi_t \leftarrow \text{dreamLearn}(D^{lab}, D^{unl}, \boldsymbol{\phi}_{t-1}, \pi_{t-1}, \mathcal{E}, T_d)$

end for

return $\boldsymbol{\phi}_{\mathcal{W}}$

Algorithm

Algorithm 2 Wake Learn

Input: labelled data D^{lab}, unlabelled pool D^{unl}, student model $\boldsymbol{\phi}$, query policy π, wake length T_w,

Output: labelled dataset and trained model

for $t \in 1, \ldots, T_w$ **do**

  $\boldsymbol{s}_t \leftarrow (D^{lab}, D^{unl}, \boldsymbol{\phi})$

  $\boldsymbol{x}_t \leftarrow \arg\max_{\boldsymbol{x}' \in D^{unl}} \pi(\boldsymbol{x}'; \boldsymbol{s}_t)$

  $\boldsymbol{y}_t \leftarrow \text{askHumanAnnotation}(\boldsymbol{x}_t)$

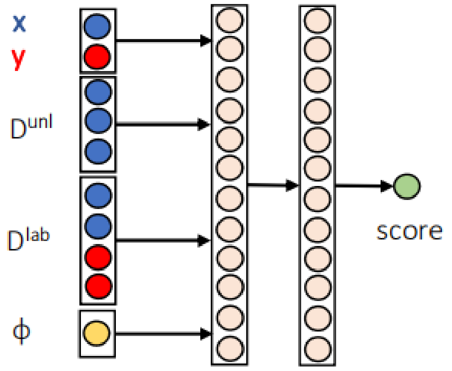
  $D^{lab} \leftarrow D^{lab} + \{(\boldsymbol{x}_t, \boldsymbol{y}_t)\}$

  $D^{unl} \leftarrow D^{unl} - \{\boldsymbol{x}_t\}$

  $\phi \leftarrow \text{retrainModel}(\boldsymbol{\phi}, D^{lab})$

end for

return D^{lab} and $\boldsymbol{\phi}$



(d) AL Policy network

Algorithm 3 Dream Learn

Input: labelled data D^{lab}, unlabelled data D^{unl}, student model $\boldsymbol{\phi}$, policy $\hat{\pi}$, dream episodes $\mathcal{E}$, dream length T_d

Output: The learned policy

$M \leftarrow \emptyset$ ▷ the aggregated dreamt AL trajectories

$\hat{\pi}_0 \leftarrow \pi$

$D^{pool} \leftarrow \text{labelGen}(m_{\boldsymbol{\phi}}, D^{unl})$

for $\tau \in 1, \ldots, \mathcal{E}$ **do**

  $D^{trn}, D^{evl} \leftarrow \text{dataPartition}(D^{lab})$

  $M += \text{trajectoryGen}(D^{trn}, D^{evl}, D^{pool}, \pi_{\tau-1}, T_d)$

  $\pi_\tau \leftarrow \text{retrainPolicy}(\pi_{\tau-1}, M)$

end for

return $\hat{\pi}_{\mathcal{E}}$

Algorithm

Algorithm 3 Dream Learn

Input: labelled data D^{lab}, unlabelled data D^{unl}, student model ϕ, policy $\hat{\pi}$, dream episodes $\mathcal{E}$, dream length T_d
Output: The learned policy

1: $M \leftarrow \emptyset$ ▷ the aggregated dreamt AL trajectories
2: $\hat{\pi}_0 \leftarrow \pi$
3: $D^{pool} \leftarrow \text{labelGen}(m_{\phi}, D^{unl})$
4: **for** $\tau \in 1, \ldots, \mathcal{E}$ **do**
5: $\quad D^{trn}, D^{evl} \leftarrow \text{dataPartition}(D^{lab})$
6: $\quad M+ = \text{trajectoryGen}(D^{trn}, D^{evl}, D^{pool}, \pi_{\tau-1}, T_d)$
7: $\quad \pi_{\tau} \leftarrow \text{retrainPolicy}(\pi_{\tau-1}, M)$
8: **end for**
9: **return** $\hat{\pi}_{\mathcal{E}}$

$$\boldsymbol{b}_t = \arg\max_{\boldsymbol{x} \in D_t^{pool}} \pi(\boldsymbol{s}_t; \boldsymbol{x}', \hat{y}')$$

$$\boldsymbol{a}_t = \arg\max_{(\boldsymbol{x}', \hat{y}') \in D_t^{pool}} - \text{loss}(m_{\boldsymbol{\phi}_t'}, D^{evl})$$

where $m_{\boldsymbol{\phi}_t'} = \text{retrainModel}(\boldsymbol{\phi}_{t-1}, (\boldsymbol{x}', y'))$

β $\downarrow$ $t = 1$ **to** T_d

trajectory $\Rightarrow$ $M = \{(\boldsymbol{s}_i, \boldsymbol{a}_i)\}$

Objective

$$\sum_{i=1}^{|M|} \log Pr(\boldsymbol{a}_i | D_i^{pool})$$

where $$Pr(\boldsymbol{a}_i | D_i^{pool}) = \frac{\exp \pi(\boldsymbol{a}_i; \boldsymbol{s}_i)}{\sum_{\boldsymbol{x} \in D_i^{pool}} \exp \pi(\boldsymbol{x}; \boldsymbol{s}_i)}$$

DAGGER: A Reduction of Imitation Learning and Structured PredictionALIL: to No-Regret Online Learning

Experiments: Text Classification

- Datasets

src	tgt	doc. (src/tgt) number	doc. (src/tgt) avg. len. (tokens)
elec.	music dev.	27k/1k	35/20
book	movie	24k/2k	140/150
en	sp	3.6k/4.2k	1.15k/1.35k
en	pt	3.6k/1.2k	1.15k/1.03k

Table 1: The data sets used in sentiment classification (top part) and gender profiling (bottom part).

- The state representation

(i) $\boldsymbol{h}(\boldsymbol{x})$

the candidate document represented by a CNN;

(ii) $m_{\phi}(\boldsymbol{x})$

the distribution over the document's class labels;

(iii) $\sum_{\boldsymbol{x}' \in D^{lab}} \boldsymbol{h}(\boldsymbol{x}')$

the sum of all document vector representations in the labelled set;

(iV) $\sum_{\boldsymbol{x}' \in D^{pool}_{rnd}} \boldsymbol{h}(\boldsymbol{x}')$

the sum of all document vectors in the sample unlabelled pool;

(V)the empirical distribution of class labels in the labelled dataset.

Experiments: Text Classification

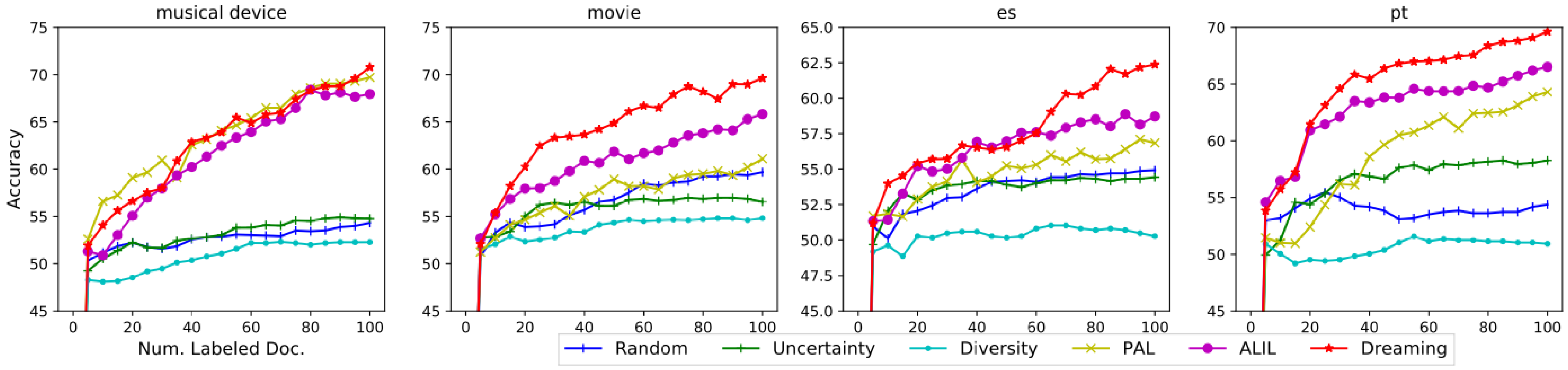


Figure 2: Accuracy of different active learning methods for cross domain sentiment classification (left two plots) and cross lingual authorship profiling (right two plots).

PAL: Learning how to Active Learn:A Deep Reinforcement Learning Approach
ALIL: Learning How to Actively Learn: A Deep Imitation Learning Approach

Experiments:Text Classification

	musical	movie	es	pt
CL-warm-transfer (ALIL)	67.95	65.82	58.72	66.52
CL-cold-dream	62.32	64.86	57.43	58.30
CL-warm-dream	**70.80**	**69.60**	**62.37**	**69.62**
WL-warm-transfer (ALIL)	76.79	80.81	64.35	69.05
WL-cold-dream	76.00	80.07	63.78	68.56
WL-warm-dream	**77.92**	**81.62**	**67.57**	**70.70**

Table 2: Classifiers performance under different initialization settings of underlying classifier and AL policy. CL, WL denotes cold-start and warm-start classifier.

Experiments:Named Entity Recognition

☐ The state representation

(i) the representation of the candidate sentence using the sentence convolution network cnn_{sent} (Kim, 2014)

(ii) the representation of the labelling marginals using the label-level convolution network $\text{cnn}_{\text{lab}}(\mathbb{E}_{m_{\boldsymbol{\phi}}(\boldsymbol{y}|\boldsymbol{x})}[\boldsymbol{y}])$ (Fang et al., 2017)

(iii) the bag-of-word representation of sentences in the sample pool of unlabelled data $\sum_{\boldsymbol{x}' \in D^{pool}_{rnd}} \sum_{w \in \boldsymbol{x}'} \boldsymbol{e}(w)$ where $\boldsymbol{e}(w)$ is embedding of word w

(iv) the representation of ground-truth labels in the labelled data $\sum_{(\boldsymbol{x}',\boldsymbol{y}') \in D^{lab}} \text{cnn}_{\text{lab}}(\boldsymbol{y}')$ using the empirical distributions

(v) the confidence of the sequential prediction $\sqrt[|\boldsymbol{x}|]{\max_{\boldsymbol{y}} m_{\boldsymbol{\phi}}(\boldsymbol{y}|\boldsymbol{x})}$

(vi) the representation of the entropy sequences for each word label in the sentence using another convolution network cnn_{ent}

(vii) entropy statistics includes max entropy, average entropy and sum entropy

Experiments:Named Entity Recognition

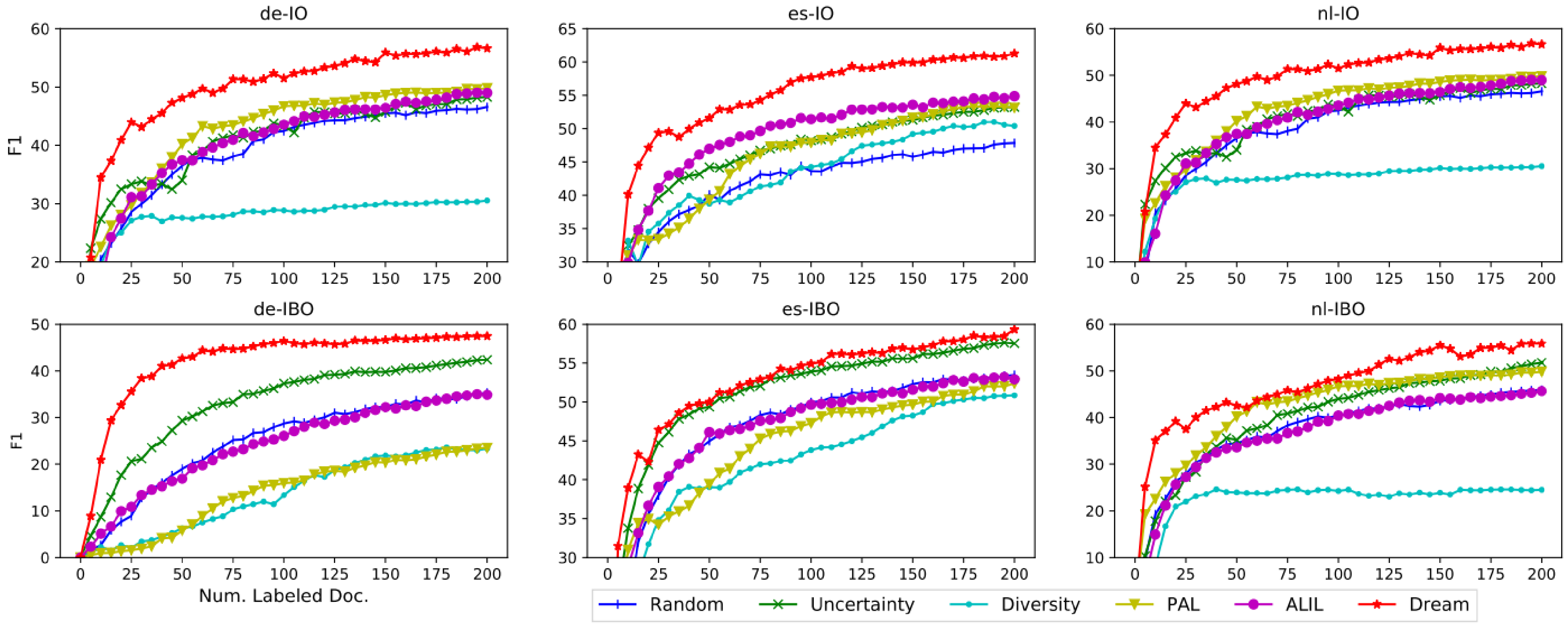


Figure 3: The performance of dreaming methods on bilingual settings under IO and IBO annotation scheme for three target languages: German (de), Spanish (es) and Dutch (nl).

Experiments:Biomedical Named Entity Recognition

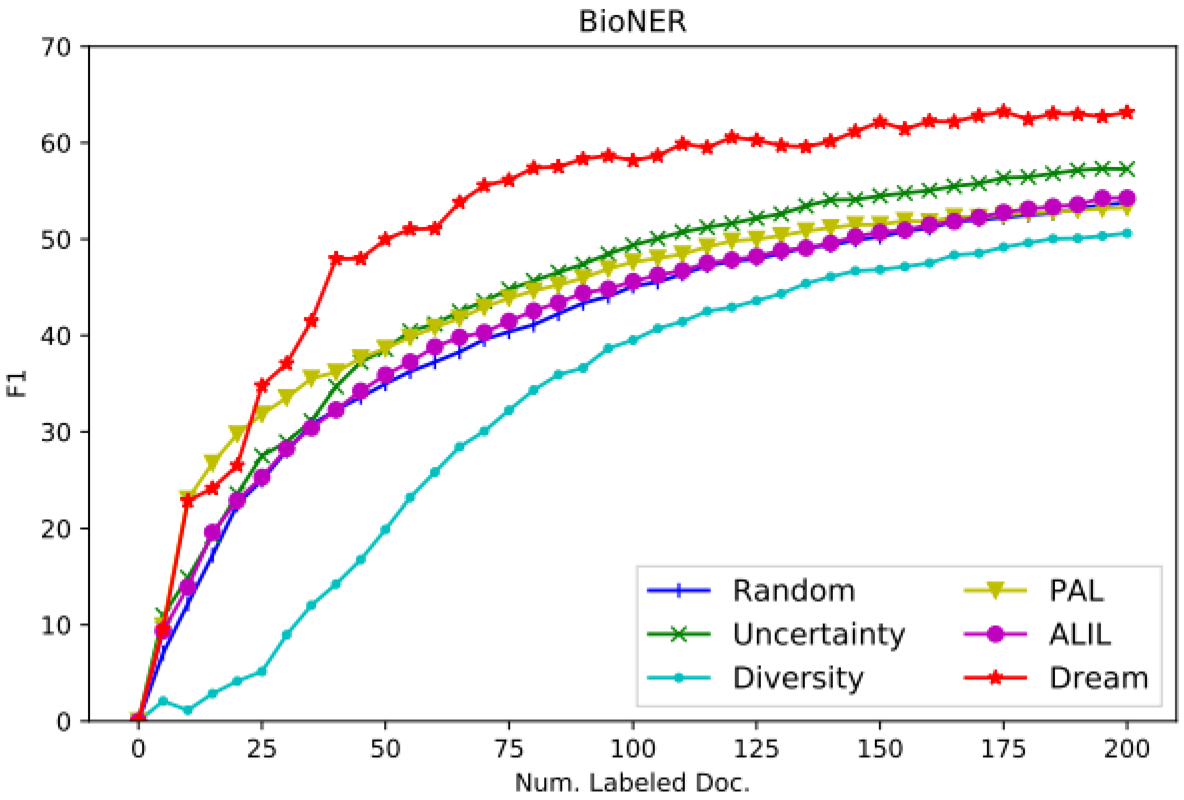


Figure 4: The performance of transferring trained policy on English NER to BioNER task.

Experiments:Sensitivity analysis

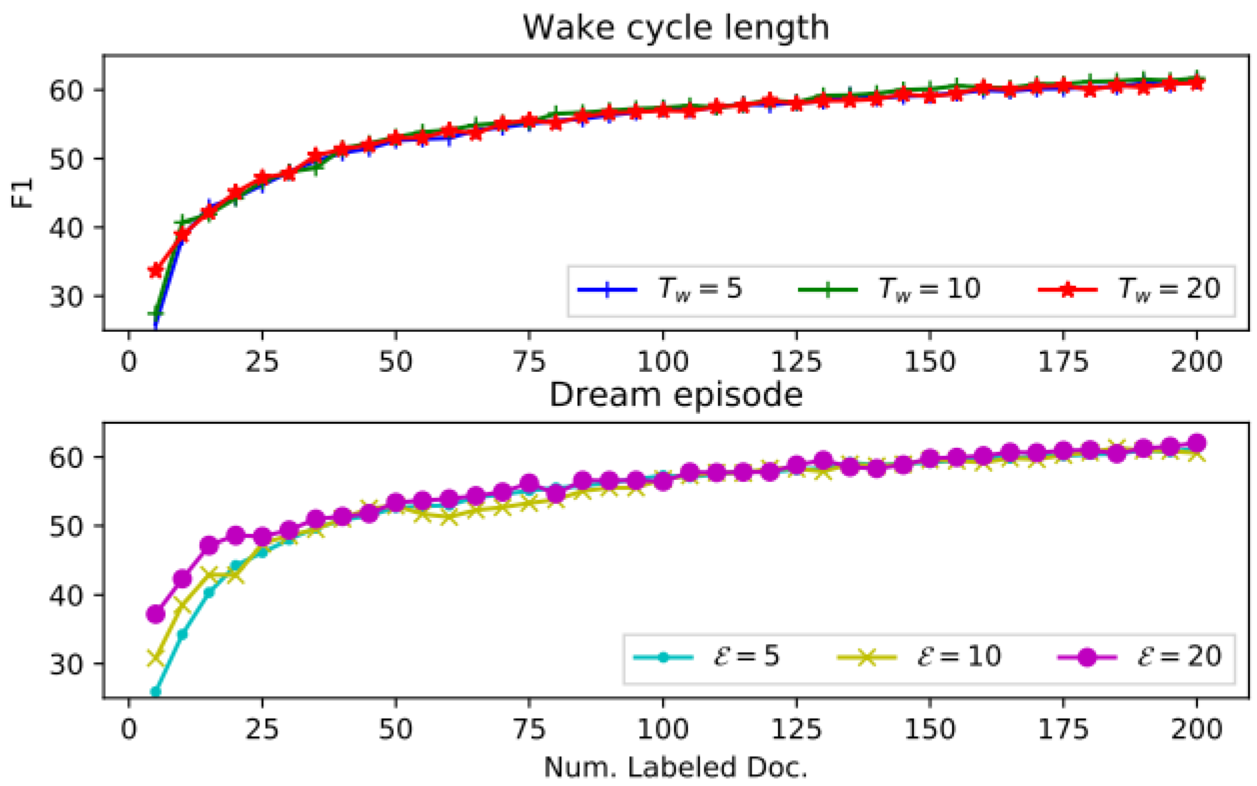


Figure 5: The performance of Spanish NER taggers respect to different wake phase length T_w and number of dream episode $\mathcal{E}$.

Experiments:Candidate selection strategy

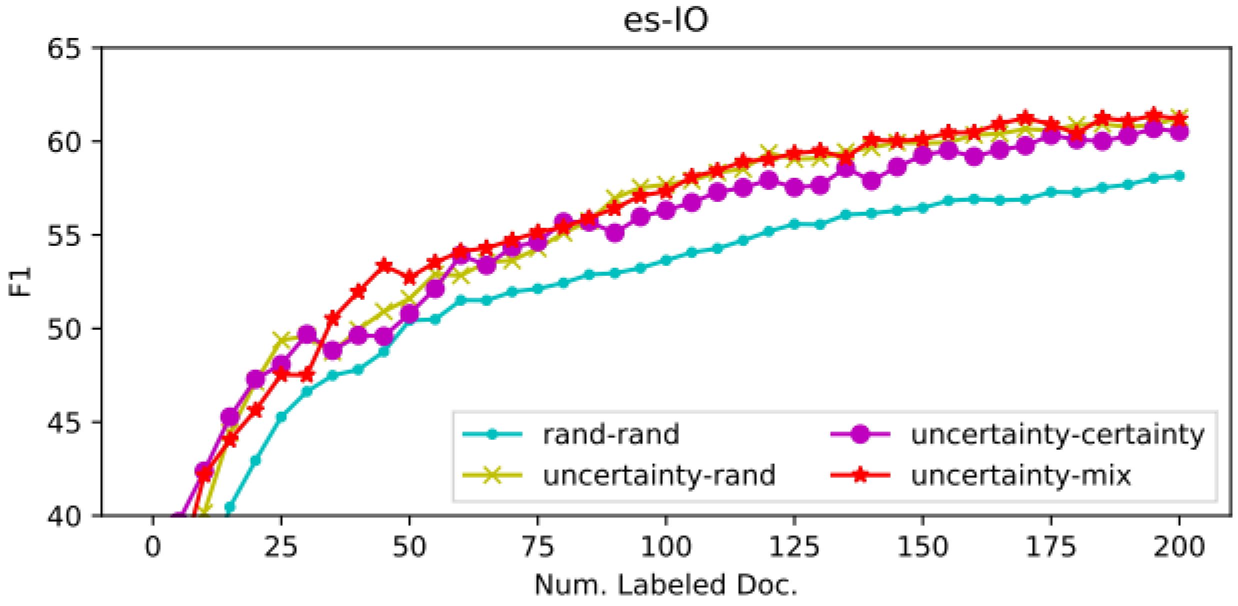


Figure 6: The performance of Spanish NER taggers under different candidate selection strategies.

Thanks