南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

Evolutionary Strategy to perform Batch-Mode Active Learning on Multi-Label Data

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

Introduction

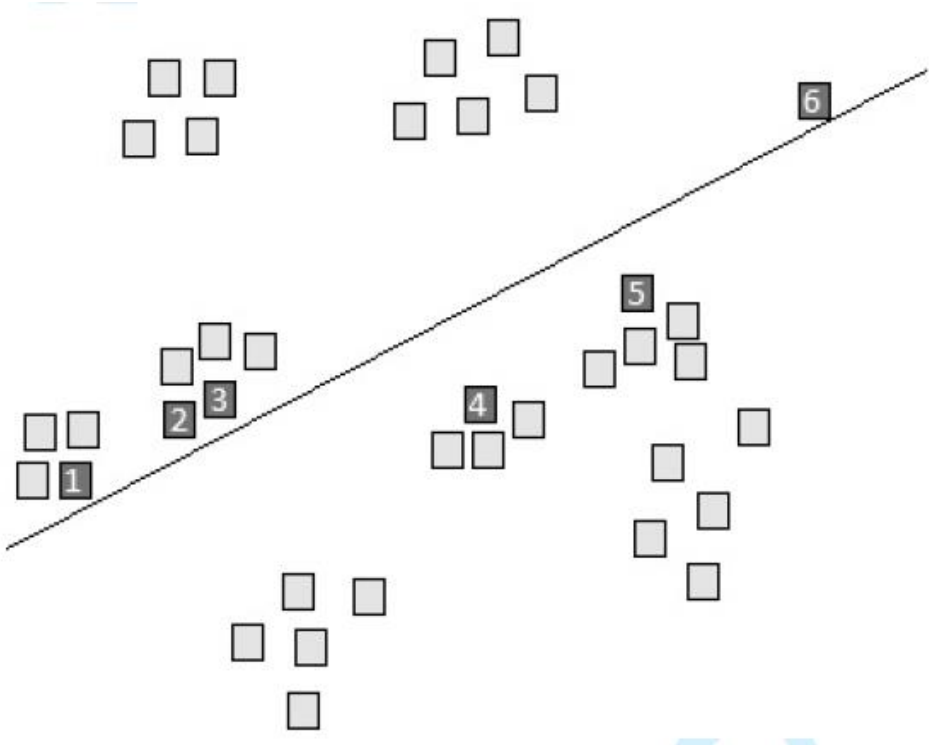
Batch-mode

- fast training procedure
- a parallel annotation environment is available

maximum information and minimal overlapping parts

AL

informativeness, representativeness and diversity



南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

AL

Informativeness

$$u_i = -\frac{1}{q}\sum_{\forall \ell \in \mathcal{Y}} [p_{i,\ell} \times \log p_{i,\ell} + (1 - p_{i,\ell}) \times \log(1 - p_{i,\ell})] \qquad u_S = \frac{1}{b}\sum_{\forall i \in \mathcal{S}} u_i$$

Diversity

$$d_i = \frac{1}{q \times |L|}\sum_{\forall l \in Y}\sum_{\forall j \in L}(p_{i,l} - p_{j,l})\ln\frac{p_{i,l}}{p_{j,l}} \qquad d_s = \frac{1}{b}\sum_{\forall i \in S} d_i + \frac{1}{q \times b^2}\sum_{\forall l \in Y}\sum_{\forall i,j \in S}(p_{i,l} - p_{j,l})\ln\frac{p_{i,l}}{p_{j,l}}$$

Representativeness

$$r_i = \frac{\sum_{\forall j \in \mathcal{L}} dx(i,j)}{\sum_{\forall j \neq i \in \mathcal{U}} dx(i,j)} \qquad r_S = \frac{1}{b}\sum_{\forall i \in S} r_i$$

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

Multi-objective problem

Multi-objective problem

$$\underset{S\in\Omega}{maximize}\, M(S)=(u_S,d_S,r_S)$$

Solution:

NSGA2 (Non-dominating Sorting Genetic Algorithm)

AL strategy :

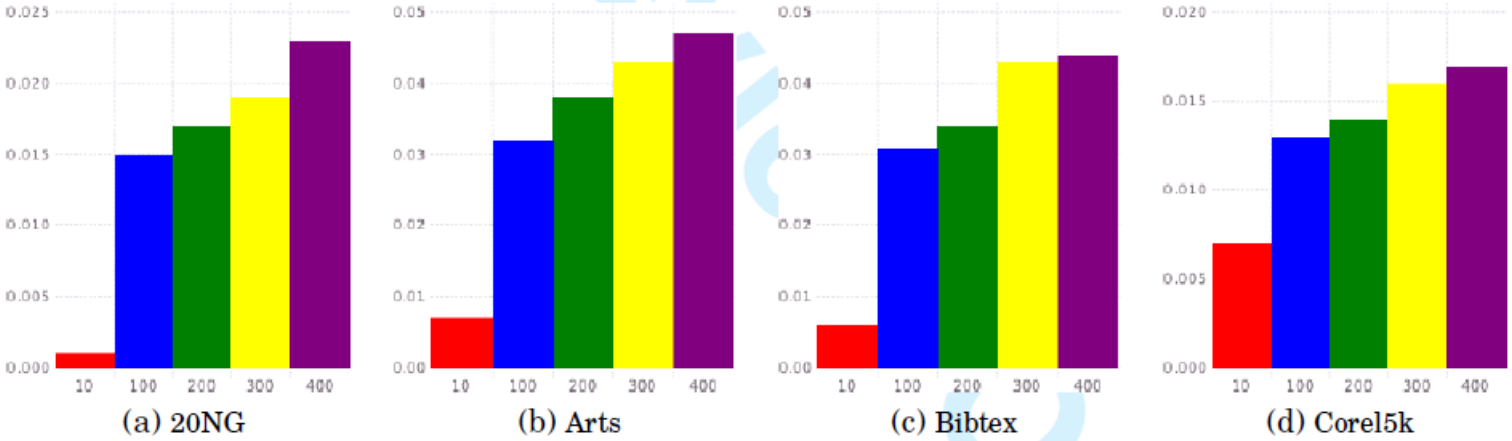
$$S=\underset{S\in P_\Omega}{\mathbf{argmax}}\ \frac{\sum_{\forall m\in\{u,d,r\}} pos(R_m,S)}{3\times|P_\Omega|}$$

P_Ω : Pareto front

$pos(R_m,S)$: position of batch S in the ranking R_m

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

Batch size influence

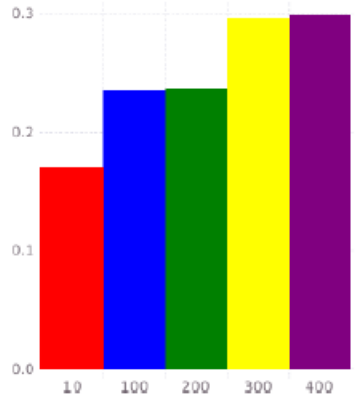


(a) 20NG

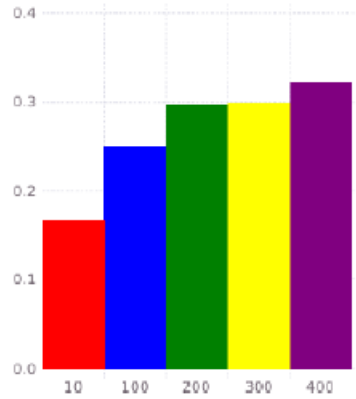
(b) Arts

(c) Bibtex

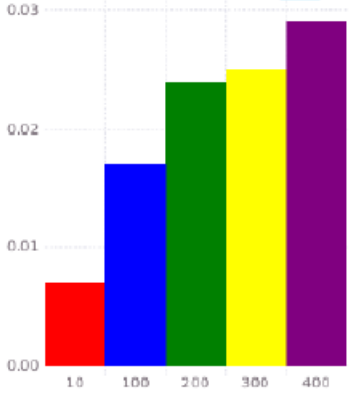
(d) Corel5k



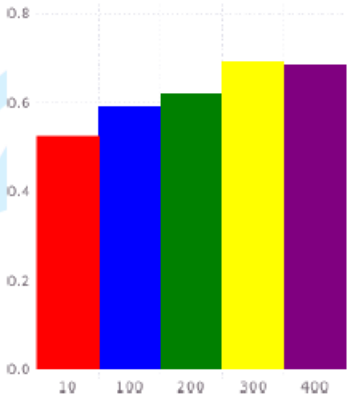
(e) Enron



(f) Health

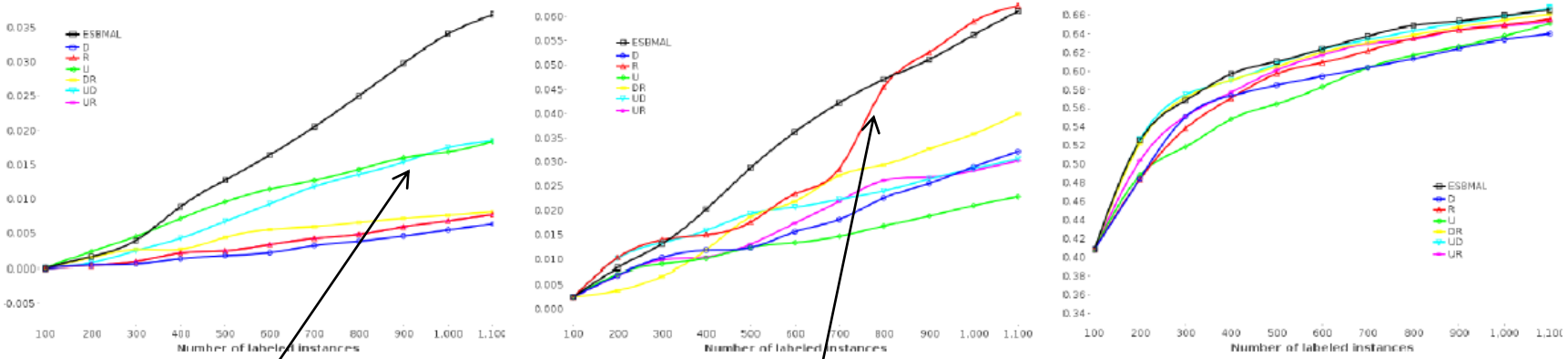


(g) Recreation



(h) Scene

Benefit of combining informativeness (U), diversity (D) and representativeness (R) criteria



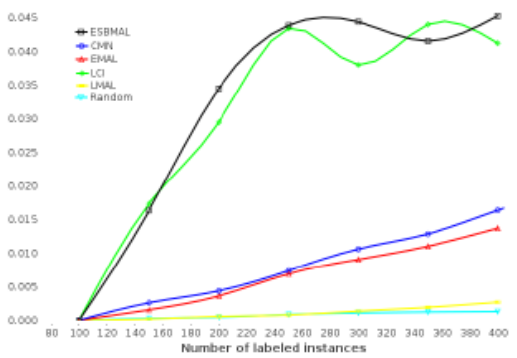
(a) 20NG (b) Bibtex (c) Scene

U,UD

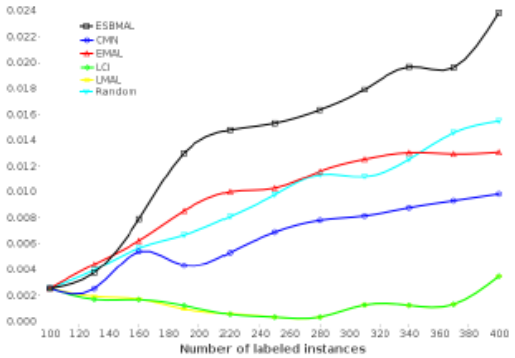
R

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

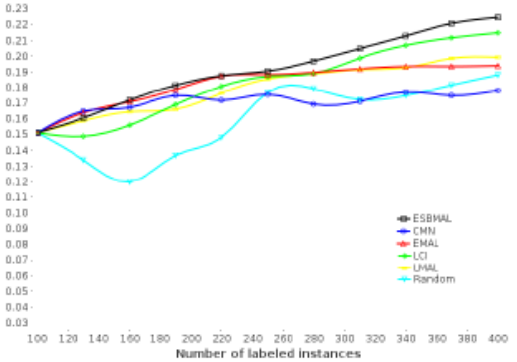
Comparison with myopic AL strategies



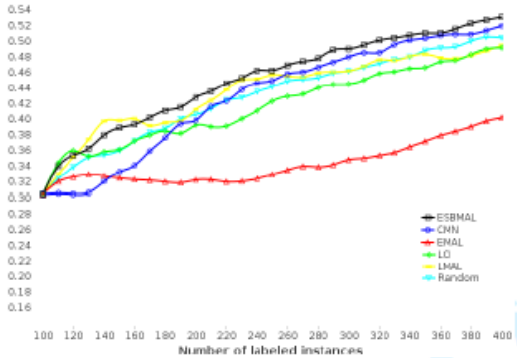
(a) 20NG



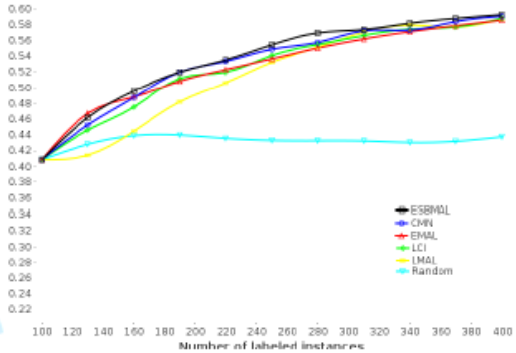
(b) Bibtex



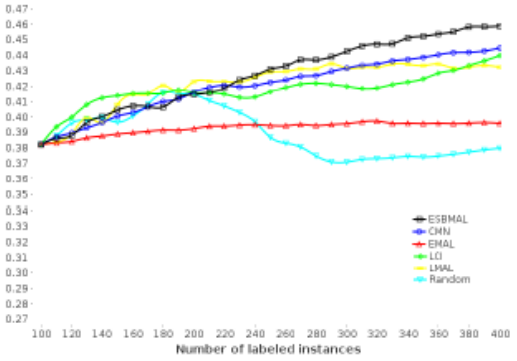
(c) Enron



(d) Medical



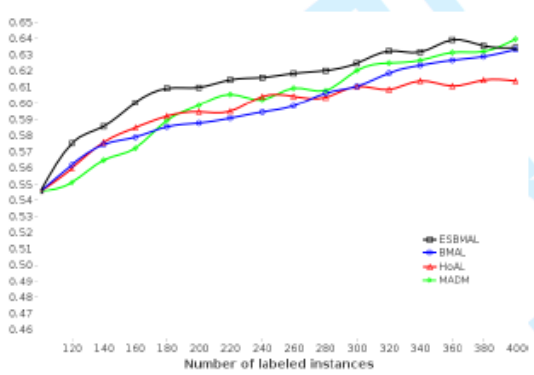
(e) Scene



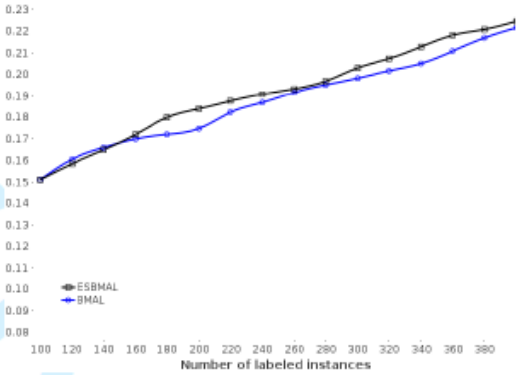
(f) TMC2007-500

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

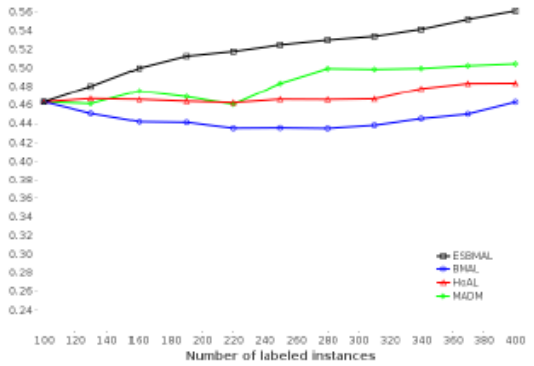
Comparison with batch-mode AL strategies



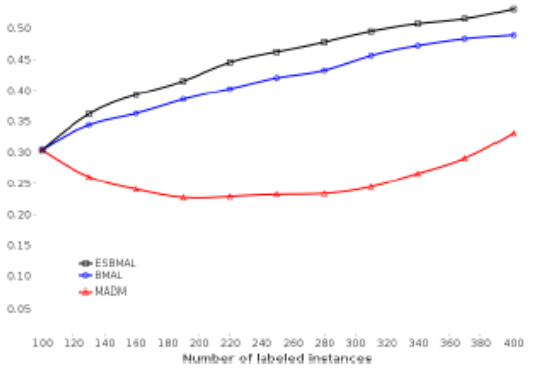
(a) Emotions



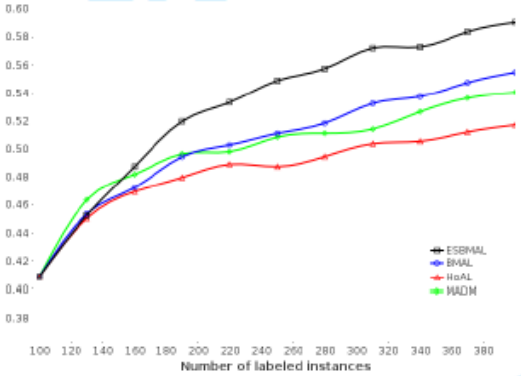
(b) Enron



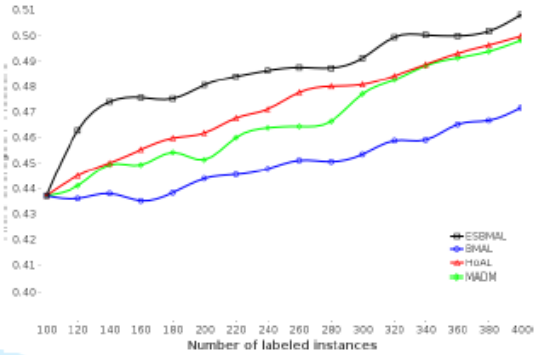
(c) GnegativePseAAC



(d) Medical



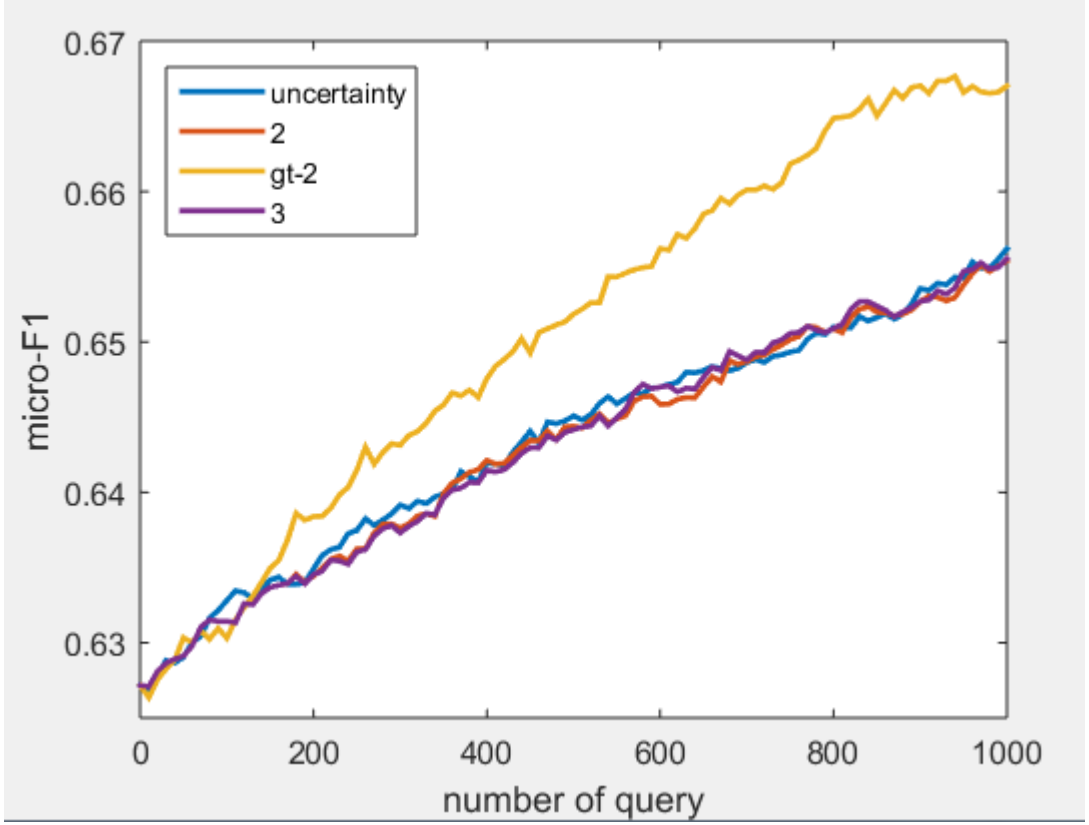
(e) Scene



(f) Water-quality

南京航空航天大学
NANJING UNIVERSITY OF AERONAUTICS & ASTRONAUTICS

Hierarchical Active Learning Classification



Thank you