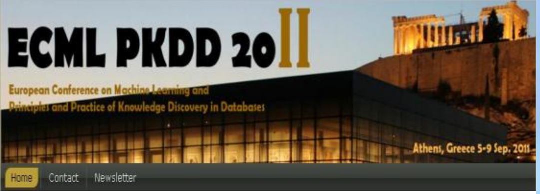
Multi-Label Ensemble Learning

Chuan Shi
Beijing University of Posts and Telecommunications

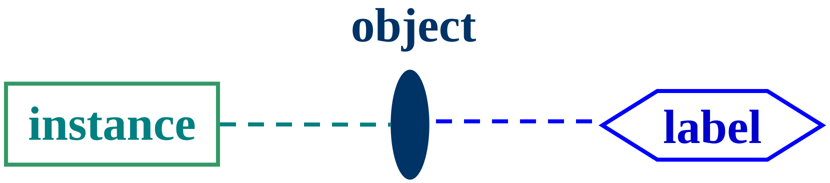
Xiangnan Kong   Philip S. Yu
University of Illinois at Chicago

Bai Wang
Beijing University of Posts and Telecommunications

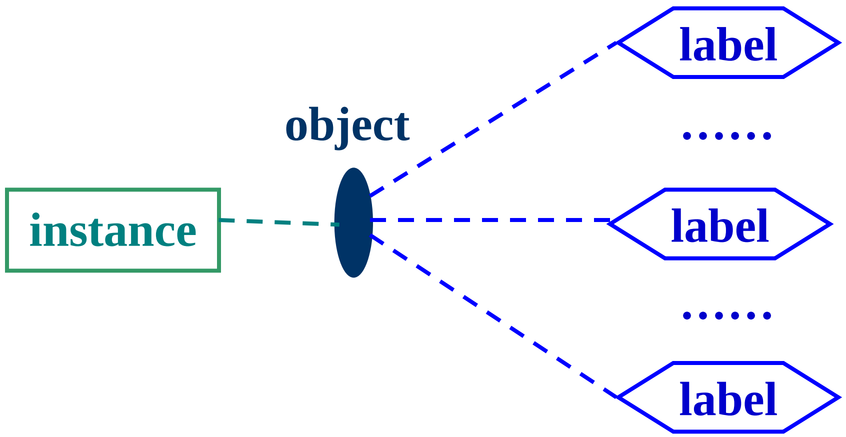


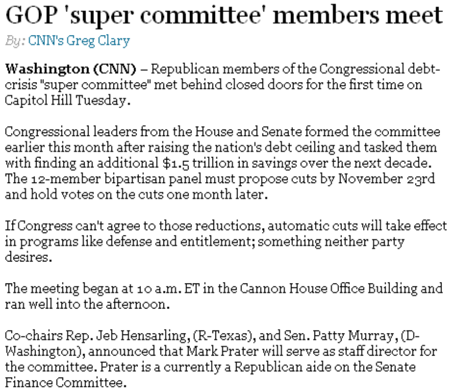
Multi-Label Learning

Traditional Supervised Learning



Multi-Label Learning (MLL)





GOP 'super committee' members meet

By: CNN's Greg Clary

Washington (CNN) – Republican members of the Congressional debt-crisis "super committee" met behind closed doors for the first time on Capitol Hill Tuesday.

Congressional leaders from the House and Senate formed the committee earlier this month after raising the nation's debt ceiling and tasked them with finding an additional $1.5 trillion in savings over the next decade. The 12-member bipartisan panel must propose cuts by November 23rd and hold votes on the cuts one month later.

If Congress can't agree to those reductions, automatic cuts will take effect in programs like defense and entitlement; something neither party desires.

The meeting began at 10 a.m. ET in the Cannon House Office Building and ran well into the afternoon.

Co-chairs Rep. Jeb Hensarling, (R-Texas), and Sen. Patty Murray, (D-Washington), announced that Mark Prater will serve as staff director for the committee. Prater is a currently a Republican aide on the Senate Finance Committee.

politics

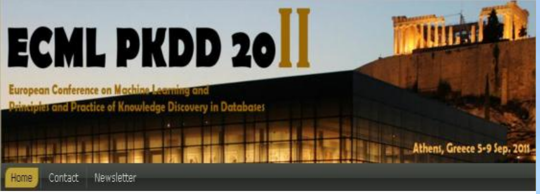
business

……

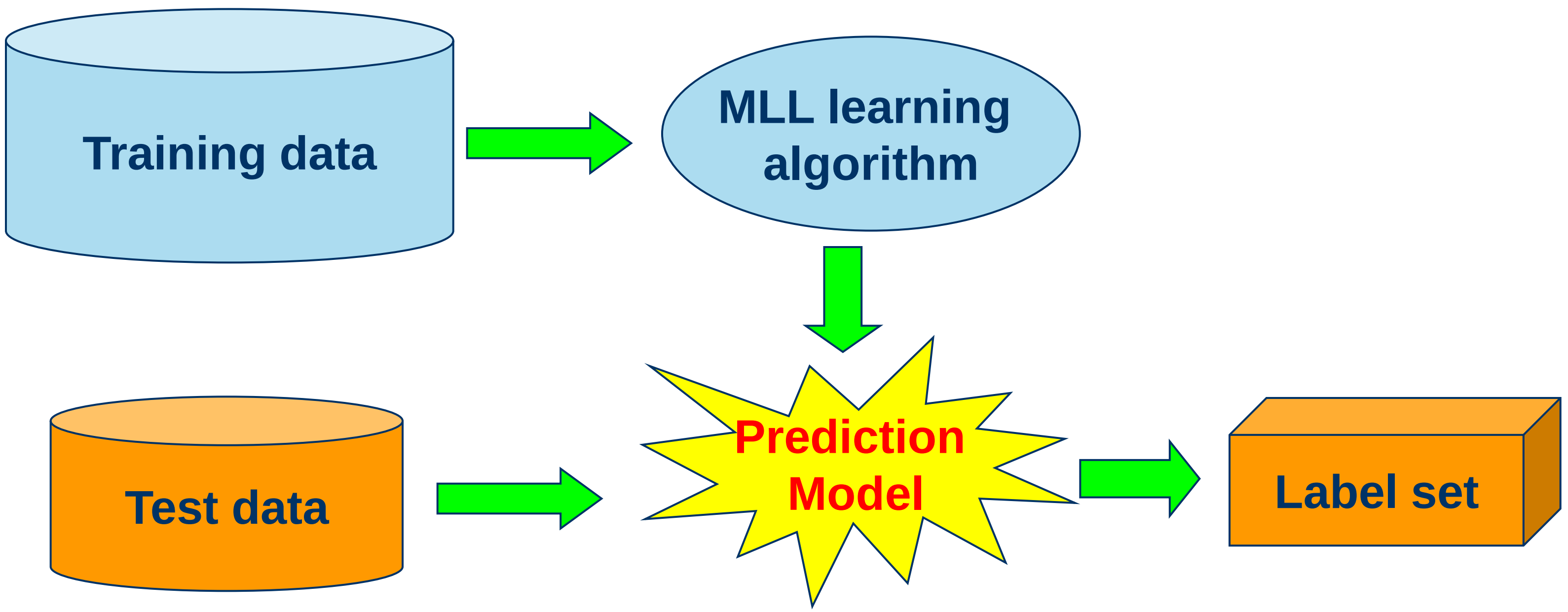
sea

boat

……



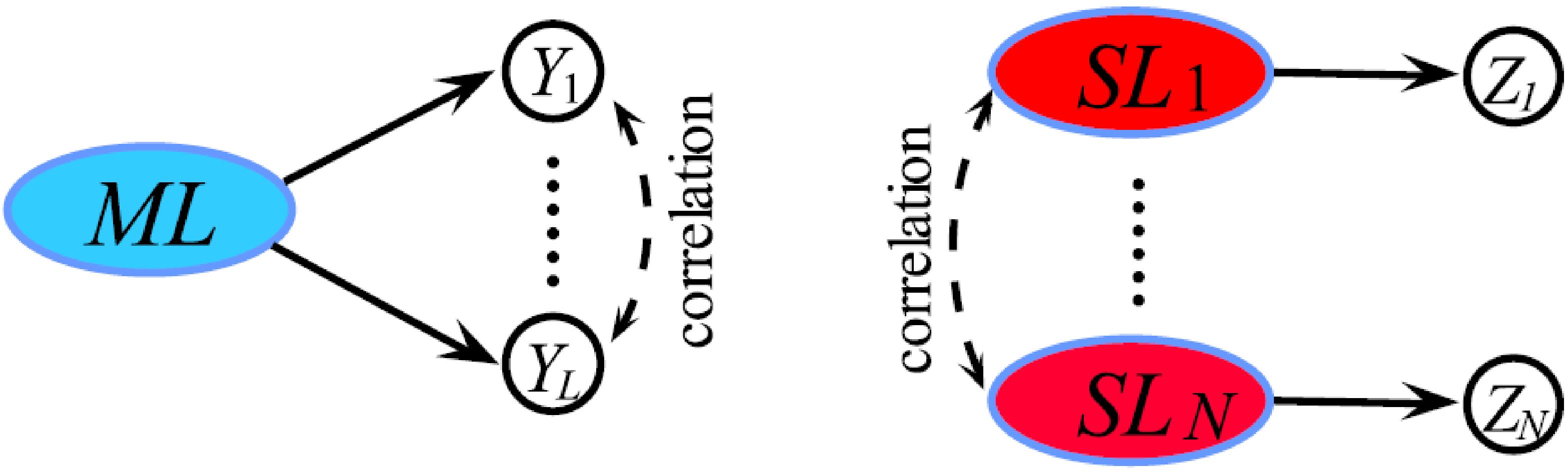
MLL Task and its Challenges

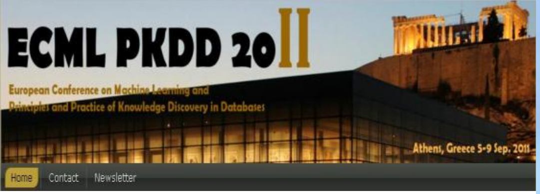


Exponential number of possible label sets ⇨ **Exploit label correlations to facilitate the learning process**

Related works

- Conventional approaches focus on exploiting the label correlations to improve the accuracy.
  - individual multi-label learner
  - a group of single-label learners





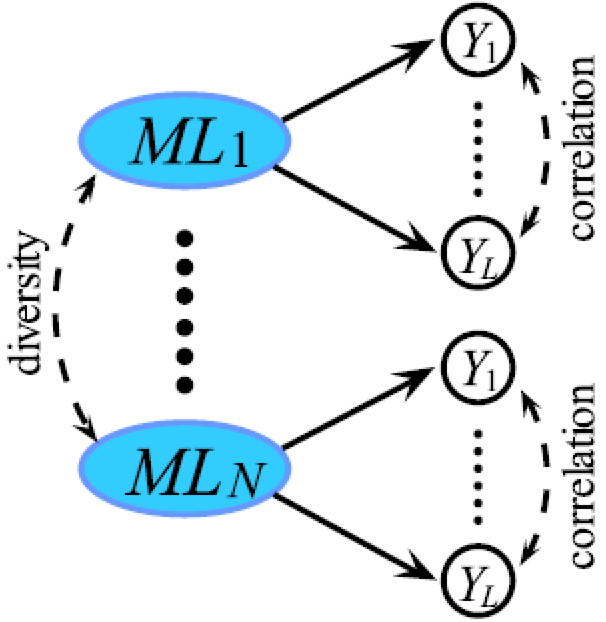
Motivation

- Conventional approaches focus on building ***one individual multi-label learner***.
- However, the generalization ability may be weak.
- Ensemble learning can improve the generalization ability of a learning system and reduce the overfitting risk.

Ensemble ***a group of multi-label base learners*** to make predictions on all labels

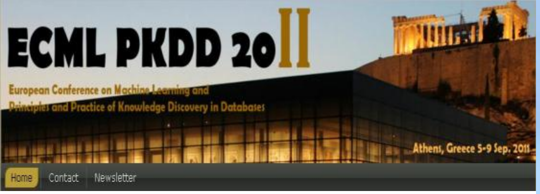
Multi-Label Ensemble Learning

- The generalization error of an ensemble
  - generalization error of the base learners
  - diversity among the base learners.
- Aim of multi-label ensemble learning
  - ***A group*** **of *accurate*** **and** ***diverse multi-label*** **base learners**



Challenges in Multi-Label Ensemble

Challenges		Solution	
Accuracy evaluation	⇒	ML-HSIC	EnML
Diversity evaluation	⇒	ML-NCL	
Considering both simultaneously	⇒	Evolutionary Multi-objective Optimization	



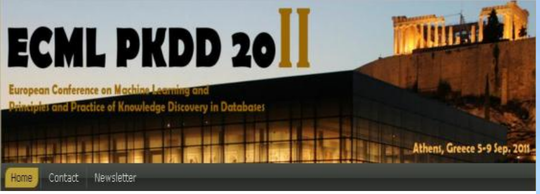
ML-HSIC Measure

- The accuracy of a learner *h* can be considered as the similarity of the true label set (*TL*) and the predicted label set (*PL*).
- The similarity can be evaluated with the dependence between them.
- Based on Hilbert-Schmidt Independence Criterion (HSIC), the accuracy of a learner *h* is:

$$ML\text{-}HSIC(h) = tr(PHQH)$$

- $H = [h_{ij}]_{m\times m},\ h_{ij} = \delta_{ij} - 1/m$ $\quad \delta_{ij}$ indicator function
  $P = [p_{ij}]_{m\times m}$ kernel on *TL* $\quad Q = [q_{ij}]_{m\times m}$ kernel on *PL*

RBF kernel is used in *P* and *Q*



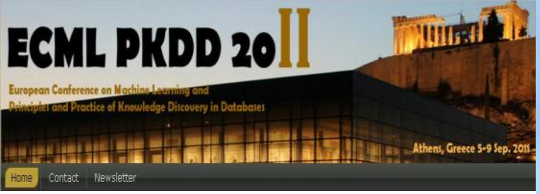
ML-NCL Measure

- In multi-label learning, the output are a set of labels, instead of a single label.
- Inspired by Negative Correlation Learning (NCL), *ML-NCL* is proposed to evaluate the negative correlation of each base learner's error with the error for the rest of ensemble.

$$ML\text{-}NCL(h_j) = -\sum_{i=1}^{m}\{(h_j(\mathbf{x}_i) - h_{ens}(\mathbf{x}_i))^T \sum_{k\neq j}(h_k(\mathbf{x}_i) - h_{ens}(\mathbf{x}_i))\}$$

$$= \sum_{i=1}^{m} \|h_j(\mathbf{x}_i) - h_{ens}(\mathbf{x}_i)\|^2$$

$$h_{ens}(\mathbf{x}_i) = \frac{1}{N}\sum_{j=1}^{N} h_j(\mathbf{x}_i)$$



Multi-objective Optimization (1)

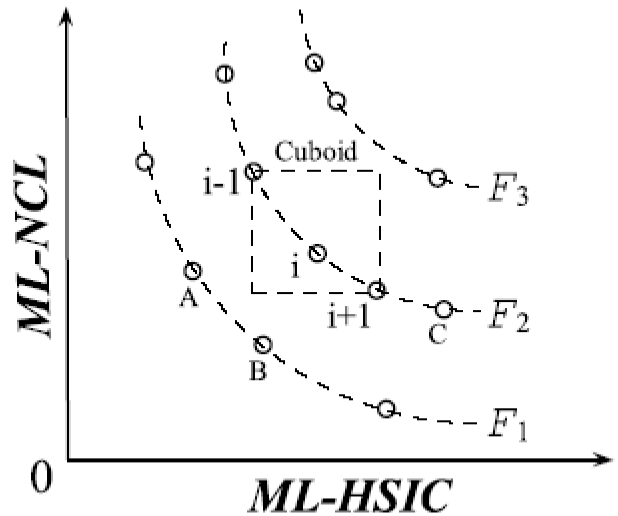
- EnML simultaneously optimizes two objectives:

  Max {*ML-HSIC*, *ML-NCL*}

- Convert it into a single objective optimization by weight sum method
  - suffers from the weights setting
- Evolutionary Multi-objective Optimization (EMO) can balance the trade-off
  - Solutions converge to optimal front and maintain diversity
  - Generate promising solutions in each generation

Multi-objective Optimization (2)

- Multi-objective Optimization Mechanism
  - *non-dominated-sort:* sorts solutions according to their raw fitness (i.e. *ML-HSIC* and *ML-NCL*)
  - *Density-assignment:* estimate the density of solutions

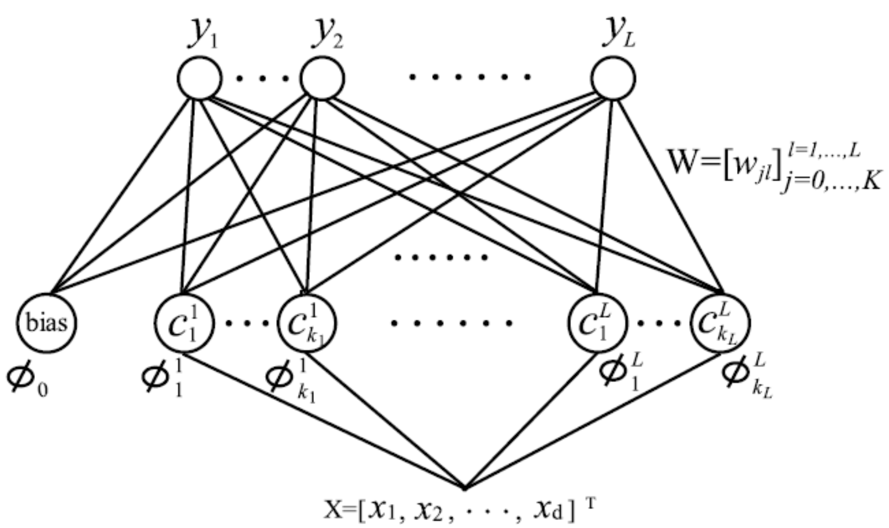


- Select-population: selects top solutions
  - For h_i :non-dominated rank h_i^{rank} density estimation $h_i^{density}$

$$h_i \prec h_j \Longleftrightarrow h_i^{rank} < h_j^{rank} \vee (h_i^{rank} = h_j^{rank} \wedge h_i^{density} > h_j^{density})$$

Multi-objective Optimization (3)

- Multi-label base learner: modified RBF
- Genetic representation: sequence of prototypes
- Initialization: a set of random RBF learners
- Generate-offspring: do crossover and mutation operation based on the roulette wheel selection.



bias	c_1^1	$\cdots$	$c_{k_1}^1$	$\cdots\cdots$	c_1^L	$\cdots$	$c_{k_L}^L$

Algorithm 1 EnML

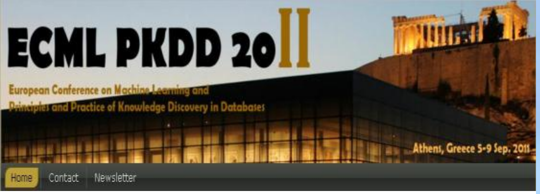
Input:
$\mathcal{D}$: training data $\mathcal{U}$: testing data h: base learner
N: # of base learners G: # of generations
output:
$Y(\mathbf{x})$: predicted labels for instance $\mathbf{x} \in \mathcal{U}$

procedure TRAINING
  generate $P_0 = \{h_1, h_2, \cdots, h_N\}$ on $\mathcal{D}$ at random
  $P_1 = (\mathcal{F}_1, \mathcal{F}_2, \cdots) = \textit{non-dominated-sort}(P_0)$
  for $t = 1 : G$ **do**
    $Q_t = \textit{generate-offspring}(P_t)$
    $R_t = P_t \bigcup Q_t$
    $F = (\mathcal{F}_1, \mathcal{F}_2, \cdots) = \textit{non-dominated-sort}(R_t)$
    $\textit{density-assignment}(F)$
    $P_{t+1} = \textit{select-population}(F)$
    $t = t + 1$
  end for
end procedure

procedure TESTING
  For $\mathbf{x} \in \mathcal{U}$, label set $Y(\mathbf{x}) = \{l \mid \frac{1}{N}\sum_{i=1}^{N} h_i(\mathbf{x}, l) > 0; h_i \in P_t, l \in \mathcal{L}\}$
end procedure

EMO framework

majority voting



Experiment Setup (1)

- **Data Collections:**
  - *Yeast* in biology, predict the gene functional classes.
  - *Image,* automatic image annotation for scene images.
  - Five datasets are from Yahoo, predict topic categories of documents.
- **Evaluation Metrics:**
  - *hamming loss*
  - *ranking loss*
  - *one-error*
  - *coverage*
  - *average precision*

Experiment Setup (2)

- Compared Methods
  - EnML: our approach, optimizes *ML-HSIC* and *ML-NCL*.
  - EnML_{HSIC} : only optimizes *ML-HSIC*.
  - EnML_{NCL} : only optimizes *ML-NCL*.
  - ML-RBF: the base learner in EnML.
  - ECC: an ensemble method for multi-label learning based on the bagging of classifier chains.
  - RAKEL: an ensemble method based on random forest.

Experiment Results

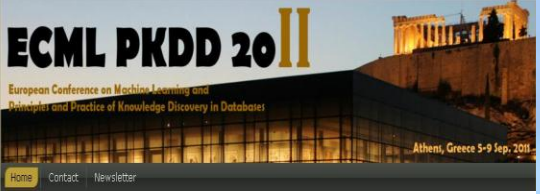
- Two examples

Table 2. Performance (mean±std.(rank)) of each algorithm in terms of *hamming loss*. Ave. Rank represents the mean and standard deviation of the rank values of each algorithm in all datasets.

	Algorithm					
Dataset	EnML	ML-RBF	ECC	RAKEL	EnML_{HSIC}	EnML_{NCL}
Image	0.1603±0.0058(2)	0.1653±0.0067(3)	0.1786±0.0108(6)	0.1724±0.0117(5)	0.1586±0.0065(1)	0.1665±0.0051(4)
Yeast	0.1889±0.0052(2)	0.1935±0.0058(4)	0.2056±0.0082(5)	0.2287±0.0105(6)	0.1887±0.0064(1)	0.1894±0.0059(3)
Arts	0.0531±0.0014(2)	0.0542±0.0016(4)	0.0754±0.0045(6)	0.0612±0.0013(5)	0.0528±0.0014(1)	0.0538±0.0015(3)
Health	0.0316±0.0016(2)	0.0331±0.0016(4)	0.0361±0.0021(5)	0.0373±0.0016(6)	0.0314±0.0017(1)	0.0322±0.0017(3)
Science	0.0317±0.0008(2)	0.0324±0.0009(4)	0.0424±0.0054(6)	0.0360±0.0016(5)	0.0313±0.0010(1)	0.0320±0.0008(3)
Recreation	0.0543±0.0023(2)	0.0555±0.0022(4)	0.0688±0.0055(6)	0.0589±0.0028(5)	0.0539±0.0021(1)	0.0553±0.0025(3)
Entertain.	0.0502±0.0016(2)	0.0512±0.0016(3)	0.0654±0.0053(6)	0.0587±0.0030(5)	0.0496±0.0013(1)	0.0514±0.0012(4)
Ave. Rank	2.00±0.00	3.71±0.49	5.71±0.49	5.29±0.49	1.00±0.00	3.29±0.49

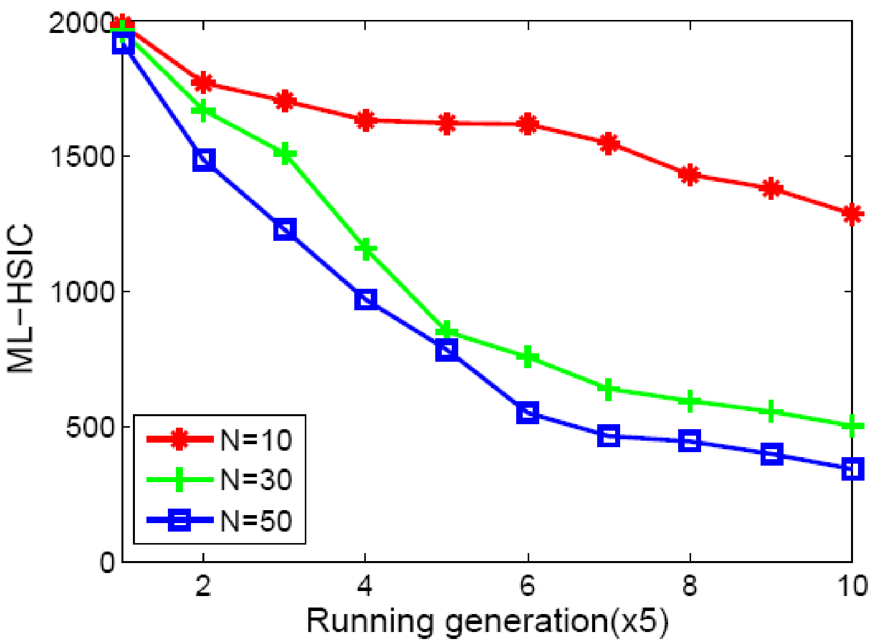
Table 3. Performance (mean±std.(rank)) of each algorithm in terms of *ranking loss*.

	Algorithm					
Dataset	EnML	ML-RBF	ECC	RAKEL	EnML_{HSIC}	EnML_{NCL}
Image	0.1478±0.0112(1)	0.1558±0.0121(4)	0.2411±0.0153(6)	0.1765±0.0200(5)	0.1485±0.0112(2)	0.1536±0.0106(3)
Yeast	0.1597±0.0083(1)	0.1621±0.0073(4)	0.2776±0.0223(6)	0.2179±0.0156(5)	0.1603±0.0087(2)	0.1619±0.0073(3)
Arts	0.1119±0.0099(1)	0.1131±0.0098(3)	0.3814±0.0251(6)	0.2589±0.0106(5)	0.1150±0.0104(4)	0.1124±0.0093(2)
Health	0.0482±0.0057(1)	0.0496±0.0051(3)	0.2401±0.0130(6)	0.1822±0.0125(5)	0.0505±0.0056(4)	0.0490±0.0054(2)
Science	0.0957±0.0072(1)	0.1002±0.0071(3)	0.3840±0.0238(6)	0.2854±0.0138(5)	0.1017±0.0079(4)	0.0992±0.0072(2)
Recreation	0.1216±0.0101(1)	0.1253±0.0099(3)	0.3434±0.0203(6)	0.2874±0.0227(5)	0.1257±0.0118(4)	0.1229±0.0095(2)
Entertain.	0.0913±0.0070(1)	0.0946±0.0073(3)	0.2926±0.0193(6)	0.2874±0.0221(5)	0.0949±0.0073(4)	0.0933±0.0062(2)
Ave. Rank	1.00±0.00	3.29±0.49	6.00±0.00	5.00±0.00	3.43±0.98	2.29±0.49

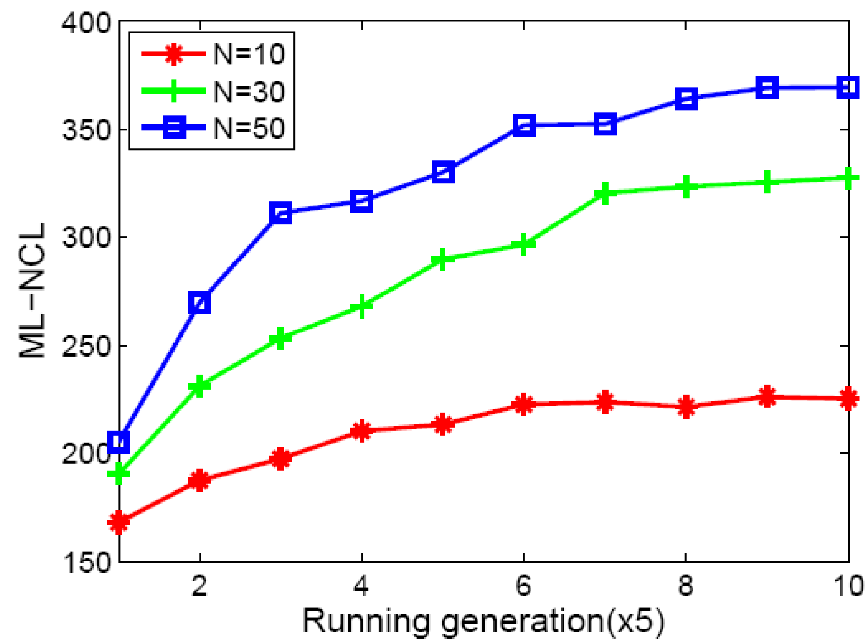


Parameter Experiments (1)

- Change population size and running generation. Observe objective values, running time and weights.
- The different trend indicates that objectives have the intrinsic conflict, which helps to find a good trade-off.



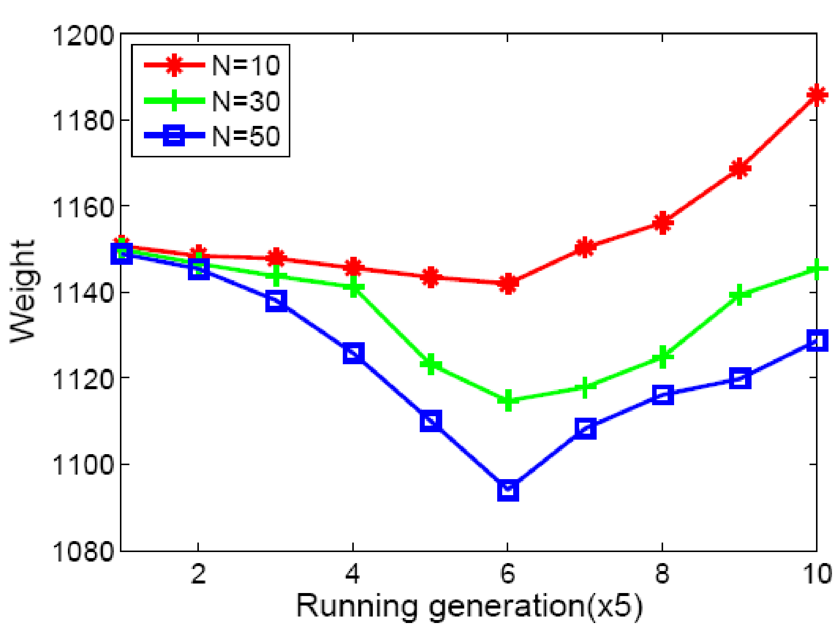
(a) *ML-HSIC*



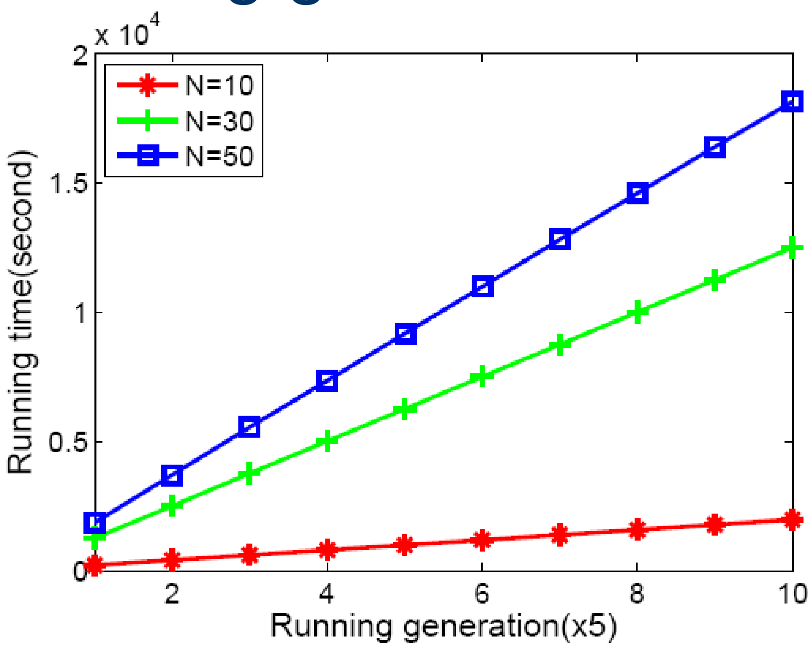
(b) *ML-NCL*

Parameter Experiments (2)

- The regularization term in modified RBF helps to control the model complexity.
- The running time of EnML increases linearly with the population size *N* and running generation *G*.



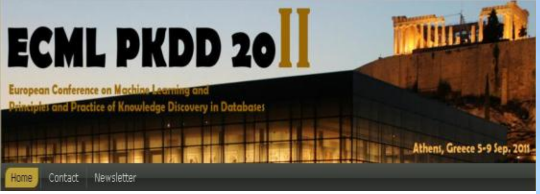
(c) Weight



(d) Running time

Conclusion

- We first study the multi-label ensemble learning problem, which aims at building a set of accurate and diverse multi-label base learners.
- We propose a solution EnML, which optimizes two novel measures with evolutionary multi-objective optimization.
- Experiments show that EnML can effectively boost the predictive performance for multi-label classification.



- Thanks

- Questions?