DW-ML-kNN: A Dual Weighted Multi-label kNN Algorithm

Duoqian Miao, Zhifei Zhang*

Tongji University

Department of Computer Science and Technology

22 September, 2012

Outline

- Multi-label Objects
- Multi-label Learning
- DW-ML-kNN Algorithm
- Experiments
- Conclusion

Outline

- Multi-label Objects
- Multi-label Learning
- DW-ML-kNN Algorithm
- Experiments
- Conclusion

Multi-label Objects

- Text

Diaoyu Islands issue

CNC report from Beijing

Added On September 14, 2012

The spokesman called the purchase 'illegal and invalid,' warning that China has made its strong opposition clear - and is now taking military measures, to safeguard its sovereignty.

An official with the Ministry of Agriculture said Thursday that China will conduct routine patrols near the Diaoyu Islands to preserve the country's territorial integrity, and protect its fishermen.

A vice minister of commerce said Thursday the move will inevitably have a negative impact on Sino-Japan economic and trade ties.

The official said some Japanese enterprises have already begun to feel the effects, after the government carried out its "nationalization" of the Diaoyu Islands.

Chinese tourists and travel agencies are also canceling trips to Japan, which they would otherwise have visited over the highly lucrative, week-long upcoming October National holiday...

Shandong Traffic and Communication Tourism Group, a travel consultancy, on Thursday announced it would suspend all trips to Japan, in protest against the country's "purchase" of the Diaoyu Islands.

Politics

Military

Tour

Economy

……

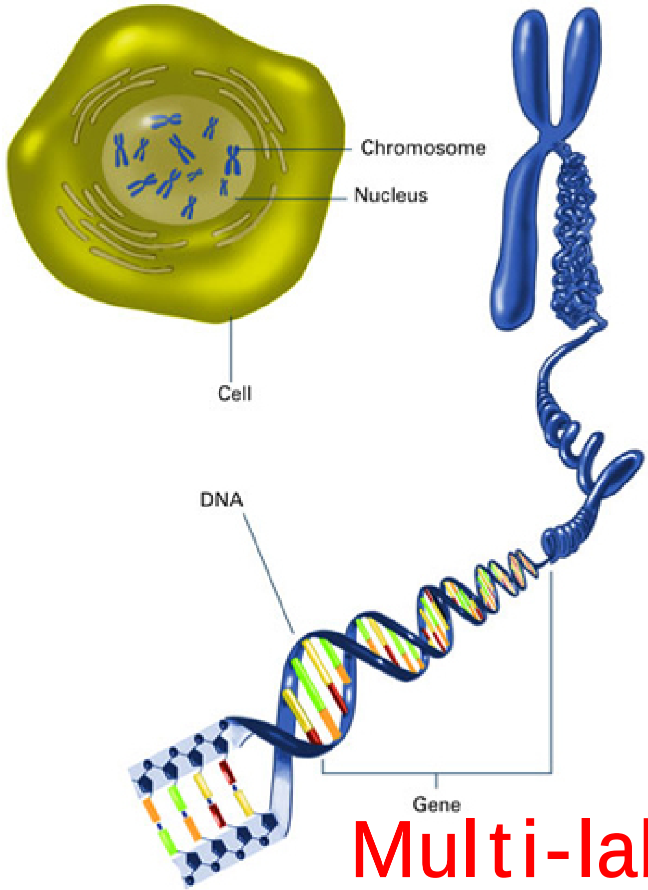
Multiple labels

Multi-label Objects (Cont')

- Image

Multi-label Objects (Cont')

- Genomics



Metabolism

Transcription

Protein Synthesis

……

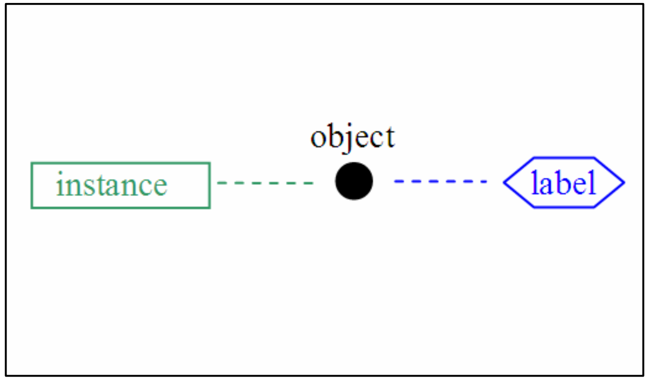
Multi-label objects are ubiquitous!

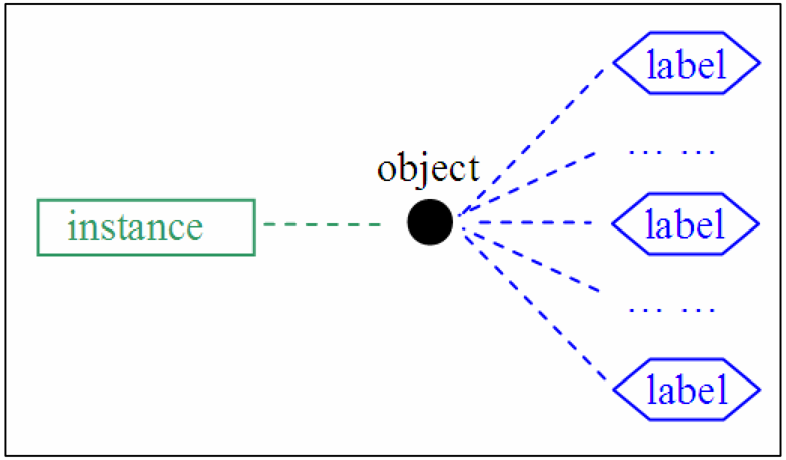
Outline

- Multi-label Objects
- Multi-label Learning
- DW-ML-kNN Algorithm
- Experiments
- Conclusion

Multi-label Learning

- MLL VS SLL (Single-label Learning)





Multi-label Learning

- Formal Definition
  - d-dimensional feature space $\mathcal{X} = \mathbb{R}^d$
  - label space with q labels $\mathcal{Y} = \{l_1, l_2, \ldots, l_q\}$
  - Inputs: training set with m examples

    $S = \{(x_i, Y_i) \mid x_i \in \mathcal{X}, Y_i \subseteq \mathcal{Y}\}(i = 1, 2, \ldots, m)$
  - Outputs:
    - multi-label predictor $h : \mathcal{X} \rightarrow 2^{\mathcal{Y}}$
    - a ranking function $f : \mathcal{X} \times \mathcal{Y} \rightarrow \mathbb{R}$

Multi-label Learning (Cont')

- Applications
  - Text classification
  - Image annotation
  - Functional genomics
  - ……
- Methods
  - Problem transformation methods — Fit data to algorithm
  - Algorithm adaptation methods — Fit algorithm to data

Outline

- Multi-label Objects
- Multi-label Learning
- DW-ML-kNN Algorithm
- Experiments
- Conclusion

ML-kNN [Zhang&Zhou, PRJ07]

- Basic idea
  - kNN + MAP with neighbors' labeling information
- Settings
  - the k nearest neighbors of x identified in the training set $N(x)$
  - the number of examples in $N(x)$ having the l-th label $C_x(l)$
  - the hypothesis that an example have (not) the l-th label $H_1^l(H_0^l)$
  - the event that there are exactly j examples in $N(x)$ having the l-th label E_j^l

同濟大學 TONGJI UNIVERSITY

ML-kNN (Cont')

- MAP

$$l \in y \text{ iff } P(H_1^l \mid E_{C_x(l)}^l) > P(H_0^l \mid E_{C_x(l)}^l)$$

$$P(H_b^l \mid E_{C_x(l)}^l) \propto P(H_b^l)P(E_{C_x(l)}^l \mid H_b^l)$$

Probabilities needed:

$$P(H_b^l)\ (1 \le l \le q, b \in \{0,1\})$$

$$P(E_j^l \mid H_b^l)\ (0 \le j \le k)$$

directly estimated from the training set based on frequency counting

DW-ML-kNN

- Why?
  - ML-kNN performs poor when dealing with imbalanced data.

compute probabilities by frequency counting

tend to assign labels with high frequency

DW-ML-kNN (Cont')

- How?
  - convert distances to weights
  - find neighbors with or without a certain label
- Idea
  - assign higher weights to closer neighbors, and lower weights to farther neighbors

    Dual + Weighted

  - higher probability with a label but lower without it, assign the example this label to a large extent

DW-ML-kNN (Cont')

- "Weighted"
  - the probability that example x with the same label of its one neighbor a

$$P(a \mid x) = \frac{1}{\sqrt{2p}} e^{-\frac{d^2(x,a)}{2}}$$

  - the weighted value of example x with the label l

$$w_x(l) = \frac{\sum_{a \in N(x) \wedge y_a(l)=1} P(a \mid x)}{\sum_{a \in N(x)} P(a \mid x)}$$

replace

~~frequency counting~~

DW-ML-kNN (Cont’)

- “Dual”
  - two posterior probabilities

$$y_t(l) = \arg\max_{b\in\{0,1\}}\{P(H_b^l \mid E_{C_t(l)}^l) + P(H_b^l \mid \boxed{E_{K-C_t(l)}^{\sim l}})\}$$

the event that there are exactly K-$C_t(l)$ examples in $N(t)$ without the l-th label

$$P(H_b^l)[P(E_{C_t(l)}^l \mid H_b^l) + P(E_{K-C_t(l)}^{\sim l} \mid H_b^l)]$$

Probabilities needed:

$$P(H_b^l)$$

$$P(E_{C_t(l)}^l \mid H_b^l)$$

$$P(E_{K-C_t(l)}^{\sim l} \mid H_b^l)$$

directly estimated from the training set based on distance weighting

DW-ML-kNN (Cont')

- Process
  - Step1: compute the prior probabilities from the training set
  - Step2: compute the posterior probabilities from the training set based on **distance weighting**
  - Step3: compute the label set and label ranking for the unseen example based on **dual posteriors**

Outline

- Multi-label Objects
- Multi-label Learning
- DW-ML-kNN Algorithm
- Experiments
- Conclusion

Experiments

- Datasets
  - Yeast: biological dataset about protein function classification
  - Scene: image dataset about semantic indexing of still scenes
  - Emotions: music dataset about song classification by emotions

Dataset	Domain	$\|S\|$	$D(S)$	$L(S)$	$LC(S)$	$LD(S)$
Yeast	Biology	2417	103	14	4.237	0.303
Scene	Multimedia	2712	294	6	1.074	0.179
Emotions	Music	593	72	6	1.869	0.311

Experiments (Cont’)

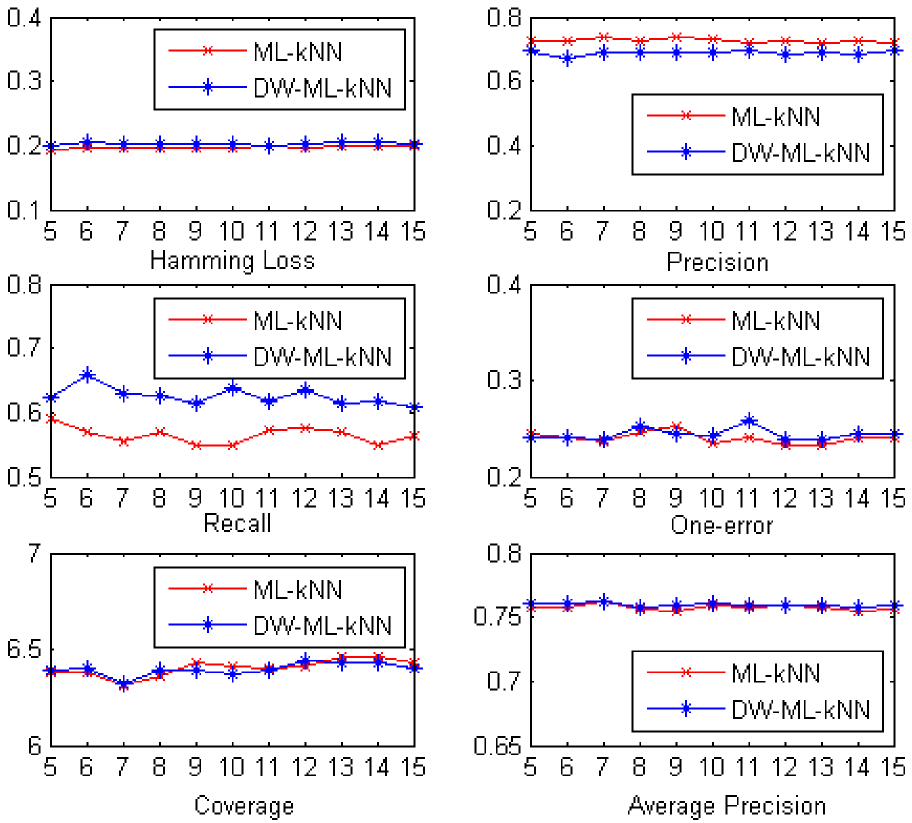
- Evaluation
  - One-error ↓ : Average times the top-ranked label is not in the set of proper labels of the example
  - Coverage ↓ : Average steps are needed to go down the label list to cover the true label set
  - Average precision ↑ : Average fraction of labels ranked above a proper label in the true label set
  - Hamming loss ↓ : Average times an example-label pair is misclassified
  - Precision ↑ : Average fraction of truly predicted labels of the predicted labels
  - Recall ↑ : Average fraction of truly predicted labels of the true labels

Ranking -based

Classification -based

Experiments (Cont')

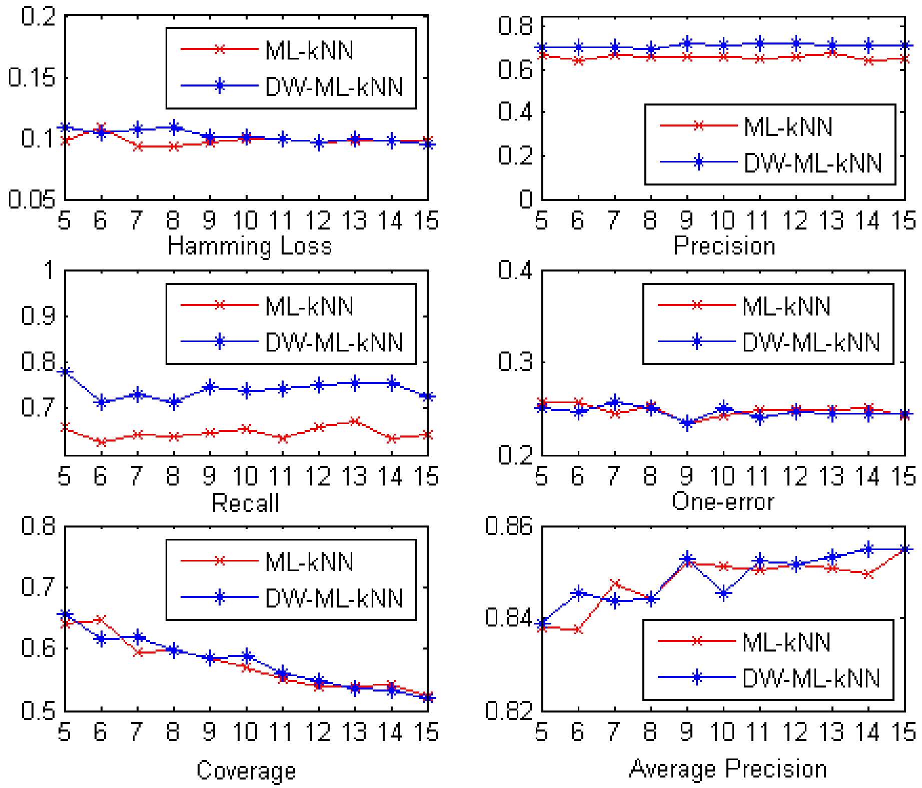
- Yeast



k=7

Experiments (Cont’)

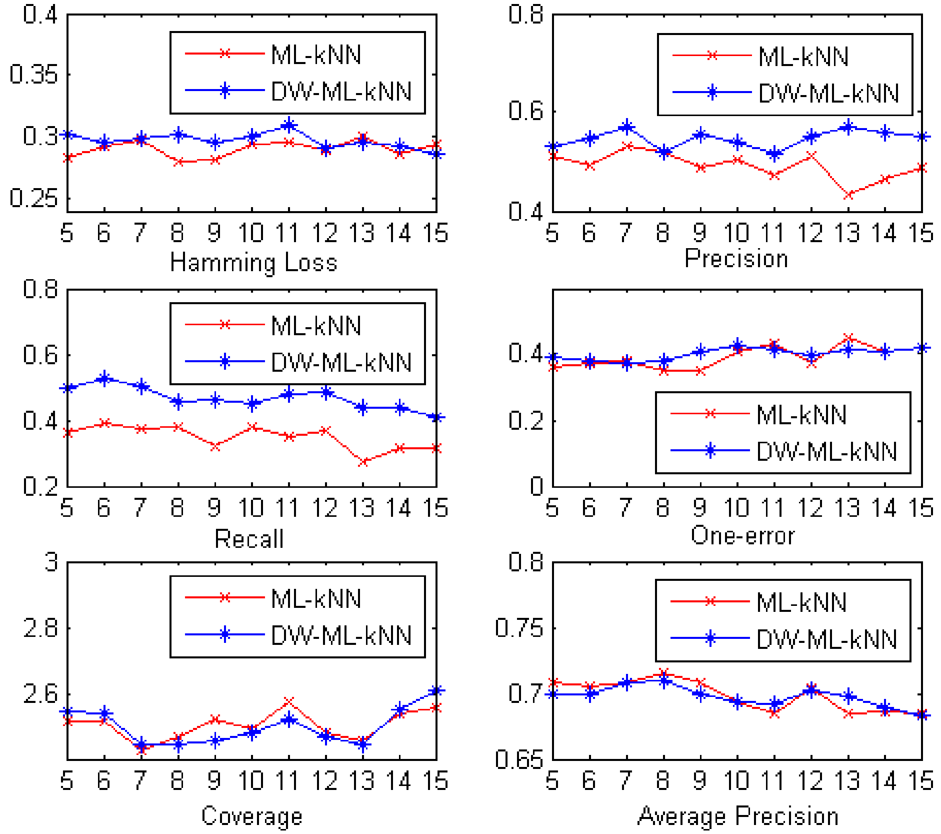
- Scene



k=9

Experiments (Cont’)

- Emotions



k=12

Experiments (Cont')

- DW-ML-kNN VS ML-kNN

Metric	Yeast		Scene		Emotions	
	ML-kNN	DW-ML-kNN	ML-kNN	DW-ML-kNN	ML-kNN	DW-ML-kNN
Hamming loss	**0.1973**	0.2028	**0.0978**	0.1013	**0.2900**	0.2981
Precision	**0.7273**	0.6880	0.6583	**0.7106**	0.4928	**0.5468**
Recall	0.5653	**0.6275**	0.6471	**0.7625**	0.3510	**0.4781**
One-error	0.2405	**0.2398**	0.2481	**0.2462**	**0.3896**	0.4012
Coverage	6.4015	**6.3956**	**0.5804**	0.5828	2.4980	**2.4888**
Average precision	0.7576	**0.7595**	0.8472	**0.8490**	0.6898	**0.6994**
Win(s)	2(6)	4(6)	2(6)	4(6)	2(6)	4(6)

- DW-ML-kNN achieves better performance
- Recall is obviously improved
- Hamming loss isn't improved, but close

Outline

- Multi-label Objects
- Multi-label Learning
- DW-ML-kNN Algorithm
- Experiments
- Conclusion

Conclusion

- The proposed algorithm
  - DW-ML-kNN is better than ML-kNN on the whole
  - Deal with imbalanced data better to some extent
- Further exploiting label correlations
  - a document labeled as *politics* would be unlikely labeled as *entertainment*
  - an image labeled as *trees* and *lake* would be likely labeled as *clouds*

Thank you!

Q & A?

Email: zzf_tj01@126.com

Weibo: @同济志飞