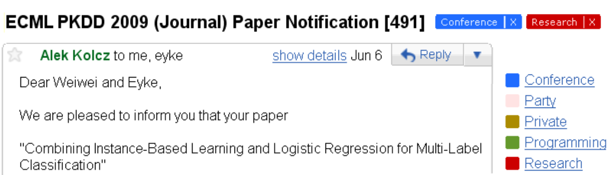
Combining Instance-Based Learning and Logistic Regression for Multilabel Classification

Weiwei Cheng and Eyke Hüllermeier, University of Marburg, Germany

Introduction

Multilabel Classification: Given an instance space $\mathbb{X}$, a finite set of labels $\mathcal{L}$, and training data $\{(\mathbf{x}_i, L_i)\} \subset \mathbb{X} \times 2^{\mathcal{L}}$, the goal is to learn a classifier $h : \mathbb{X} \rightarrow 2^{\mathcal{L}}$.



ECML PKDD 2009 (Journal) Paper Notification [491] Conference Research

Alek Kolcz to me, eyke show details Jun 6 Reply

Dear Weiwei and Eyke,

We are pleased to inform you that your paper

"Combining Instance-Based Learning and Logistic Regression for Multi-Label Classification"

Conference
Party
Private
Programming
Research

- most current multilabel classifiers are model-based
- instance-based classifiers (e.g. *MLKNN*) showed potential but have not been fully exploited yet (ignorance of correlation between labels)

Instance-Based Learning as Logistic Regression

Key idea: Consider the labels of its neighbors as "*extra features*" of an instance.

age	weight	height	sex	w.child	#	#	#			
26	62	1.83	male	no	1/3	0	1	1	0	1
16	45	1.65	female	no	0	1	1/3	0	1	0
28	85	1.90	male	yes	2/3	0	1	1	0	0
	...		...		...		...		...	
27	50	1.63	male	yes	2/3	1/3	1/3	?	?	?

Binary case

Consider query instance $\mathbf{x}_0$, distance $\delta_i \overset{\mathrm{df}}{=} \Delta(\mathbf{x}_0, \mathbf{x}_i)$, posterior probability $\pi_0 \overset{\mathrm{df}}{=} \mathbf{P}(y_0 = +1 \mid y_i)$:

$$\frac{\pi_0}{1-\pi_0} = \frac{\mathbf{P}(y_i \mid y_0 = +1)}{\mathbf{P}(y_i \mid y_0 = -1)} \cdot \frac{p_0}{1-p_0} = \rho \cdot \frac{p_0}{1-p_0}$$

$$\log\left(\frac{\pi_0}{1-\pi_0}\right) = \log(\rho) + \underbrace{\log(p_0) - \log(1-p_0)}_{\omega_0}$$

For example, we can define $\rho = \rho(\delta) \overset{\mathrm{df}}{=} \exp\left(y_i \cdot \frac{\alpha}{\delta}\right)$.

Now consider the whole neighborhood of $\mathbf{x}_0$:

$$\log\left(\frac{\pi_0}{1-\pi_0}\right) = \omega_0 + \alpha \sum_{\mathbf{x}_i \in \mathcal{N}(\mathbf{x}_0)} \frac{y_i}{\delta_i} = \omega_0 + \alpha \cdot \omega_+(\mathbf{x}_0)$$

Multilabel case

We solve one logistic regression problem for each label.

$$\log\left(\frac{\pi_{\text{football}}}{1-\pi_{\text{football}}}\right) = \omega_0 + \alpha_{\text{football}} \cdot \omega_{+\text{football}}(\mathbf{x}_0) + \alpha_{\text{basketball}} \cdot \omega_{+\text{basketball}}(\mathbf{x}_0) + \alpha_{\text{table tennis}} \cdot \omega_{+\text{table tennis}}(\mathbf{x}_0)$$

bias term (prior probability)

To what extent does the presence of label basketball in the neighborhood increase the probability that football is relevant for the query?

Results

dataset	domain	#inst.	#attr.	#labels	card.
emotions	music	593	72	6	1,87
image	vision	2000	135	5	1,24
genbase	biology	662	1186(n)	27	1,25
mediamill	multimedia	5000	120	101	4,27
reuters	text	7119	243	7	1,24
scene	vision	2407	294	6	1,07
yeast	biology	2417	103	14	4,24

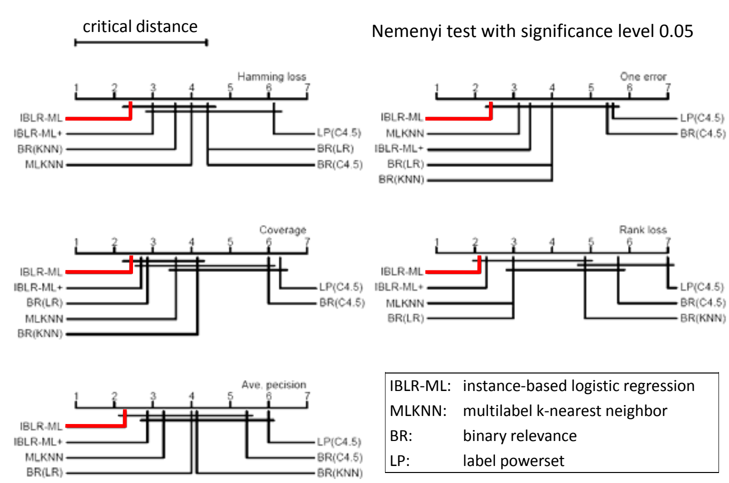
Evaluation metrics

- Hamming loss $= \frac{1}{|\mathcal{L}|} |h(\mathbf{x}) \, \Delta \, L_{\mathbf{x}}|$
- one error $= \begin{cases} 1 & \text{if } \arg\max_{\lambda \in \mathcal{L}} f(\mathbf{x}, \lambda) \notin L_{\mathbf{x}} \\ 0 & \text{otherwise} \end{cases}$
- coverage $= \max_{\lambda \in L_{\mathbf{x}}} \mathrm{rank}_f(\mathbf{x}, \lambda) - 1$
- rank loss $= \frac{\#\{(\lambda, \lambda') \mid f(\mathbf{x}, \lambda) \leq f(\mathbf{x}, \lambda'), (\lambda, \lambda') \in L_{\mathbf{x}} \times \overline{L_{\mathbf{x}}}\}}{|L_{\mathbf{x}}||\overline{L_{\mathbf{x}}}|}$
- average precision

$$= \frac{1}{|L_{\mathbf{x}}|} \sum_{\lambda \in L_{\mathbf{x}}} \frac{|\{\lambda' \mid \mathrm{rank}_f(\mathbf{x}, \lambda') \leq \mathrm{rank}_f(\mathbf{x}, \lambda), \lambda' \in L_{\mathbf{x}}\}|}{\mathrm{rank}_f(\mathbf{x}, \lambda)}$$

Two-step statistical test

1. Test if all methods perform equally (Friedman test) ✓
2. If not, compare learners in a pairwise way (Nemenyi test)



Our Contribution

- A new instance-based multilabel learning method,
- which is based on a formalization of instance-based classification as logistic regression (combination of model-based and instance-based learning),
- takes the correlation between labels into account and represents it in an easily interpretable way.

Poster presented at ECMLPKDD 2009, Bled, Slovenia, September 2009