Outline

- Intro: Multi-label Classification
- RM: The Ranking Method
- A Probability Distribution for RM
- END: Ensembles of Nested Dichotomies
- Related multi-label work: Pairwise Training
- Current work: M-END: Multi-label END
- Current work: EsPS: Ensembles of $s$plit Pruned Sets
- Summary

Intro: Multi-label Classification

Single-label (Multi-class) Classification:

- Set of instances D. Set of labels (classes) L.
- For each $d \in D$, select a label (class) $l \in L$
- Single-label representation: (d, l)

Multi-label Classification:

- Set of instances D. Set of labels L.
- For each $d \in D$, select a label subset $S \subseteq L$
- Multi-label representation: (d, S)

e.g. $L = \{Sports, Environment, Science, Politics\}$
(“Revealed: Polluting impact of humans on the oceans. . . ”,$\{Environment, Science\}$)

Intro: Multi-label Classification

Problem Transformation: Multi-label problems are transformed into one or more single-label problems.

Algorithm Adaption: Employs Problem Transformation internally to a single-label algorithm.

i.e. All multi-label classification involves **Problem Transformation**. There are three fundamental methods:

- BM (Binary Method)
- RM (Ranking Method)
- CM (Combination Method)

Intro: Multi-label Classification

Problem Transformation: Multi-label problems are transformed into one or more single-label problems.

Algorithm Adaption: Employs Problem Transformation internally to a single-label algorithm.

i.e. All multi-label classification involves **Problem Transformation**. There are three fundamental methods:

- BM (Binary Method)
- RM (Ranking Method)
- CM (Combination Method)

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

ML D_{train}; $(d, S \subseteq L)$
d_1,{$Sports$,$Politics$}
d_2,{$Science$,$Politics$}
d_3,{$Sports$}
d_4,{$Environment$,$Science$}

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

SL D_{train}; $(d, l \in L)$
d_1,$Sports$
d_1,$Politics$
d_2,$Science$
d_2,$Politics$
d_3,$Sports$
d_4,$Science$
d_4,$Environment$

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

SL D_{train}; $(d, l \in L)$
d_1,$Sports$
d_1,$Politics$
d_2,$Science$
d_2,$Politics$
d_3,$Sports$
d_4,$Science$
d_4,$Environment$

$(l \in L)$	$P(l\|d_x)$
$Sports$	λ
$Environment$	λ
$Science$	λ
$Sports$	λ

$d_x \in D_{test}$

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

SL D_{train}; $(d, l \in L)$
d_1,$Sports$
d_1,$Politics$
d_2,$Science$
d_2,$Politics$
d_3,$Sports$
d_4,$Science$
d_4,$Environment$

$(l \in L)$	$P(l\|d_x)$
$Sports$	0.03
$Environment$	0.48
$Science$	0.45
$Politics$	0.04

$d_x \in D_{test}$

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

SL D_{train}; $(d, l \in L)$
d_1,$Sports$
d_1,$Politics$
d_2,$Science$
d_2,$Politics$
d_3,$Sports$
d_4,$Science$
d_4,$Environment$

$(l \in L)$	$P(l \mid d_x)$
$Sports$	0.03
$Environment$	0.48
$Science$	0.45
$Politics$	0.04

$d_x \in D_{test}$, $t = 0.2$

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

SL D_{train}; $(d, l \in L)$
d_1,$Sports$
d_1,$Politics$
d_2,$Science$
d_2,$Politics$
d_3,$Sports$
d_4,$Science$
d_4,$Environment$

$(l \in L)$	$P(l\|d_x)$
$Sports$	0.03
$Environment$	0.48
$Science$	0.45
$Politics$	0.04

$d_x \in D_{test}$, $t = 0.2$: (d_x,{$Environment$,$Science$})

RM: The Ranking Method

Using the probability distribution from a single-label classifier, and a threshold (t), multi-labels are selected.

$L = \{Sports, Environment, Science, Politics\}$

SL D_{train}; $(d, l \in L)$
d_1,$Sports$
d_1,$Politics$
d_2,$Science$
d_2,$Politics$
d_3,$Sports$
d_4,$Science$
d_4,$Environment$

$(l \in L)$	$P(l\|d_x)$
$Sports$	0.03
$Environment$	0.48
$Science$	0.45
$Politics$	0.04

$d_x \in D_{test}$, $t = 0.2$: (d_x,{$Environment$,$Science$})

- Assumes that all labels are independent
- Issues with threshold selection / classifier selection

RM: SL Classifier Selection

Each PT method must be supplied a single-label (SL) classifier

```
weka.classifiers.multilabel.RM -t MEDC.arff
```

e.g. SMO:

```
  -W weka.classifiers.functions.SMO
```

e.g. SMO $-M$ (probabilistic outputs):

```
  -W weka.classifiers.functions.SMO -M
```

e.g. SMO under Bagging:

```
  -W weka.classifiers.meta.Bagging --
     -W weka.classifiers.functions.SMO
```

e.g. Ensembles of Nested Dichotomies:

```
  -W weka.classifiers.meta.END --
     -W weka....meta.nestedDichotomies.ND --
        -W weka.classifiers.functions.SMO
```

RM: SL Classifier Selection

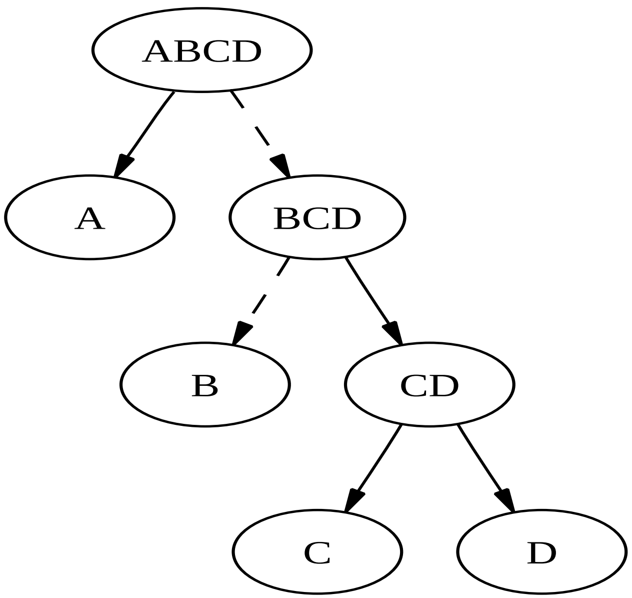
RM: Accuracy on the *Medical* and *Enron* datasets.

				END		
			Bag.	ND	CBND	DNBND
D	SMO	SMO $-M$	SMO	SMO	SMO	SMO
MED	74.53	75.43	73.47	77.74	80.31	79.55
ENR	20.90	38.49	23.78	42.86	40.47	42.53

- *Medical*: average of 1.25 labels / instance
- *Enron*: average of 3.38 labels / instance
- In each case SMO is the SL classifier. Therefore the difference is in the probability distribution and the comparison between labels.

END: Ensembles of Nested Dichotomies

- [a]: Single-label; Binary Tree; Dichotomy; Ensembles
- Can supply RM-transformed multi-label data and use prob. distribtion and threshold to gather multi-labels

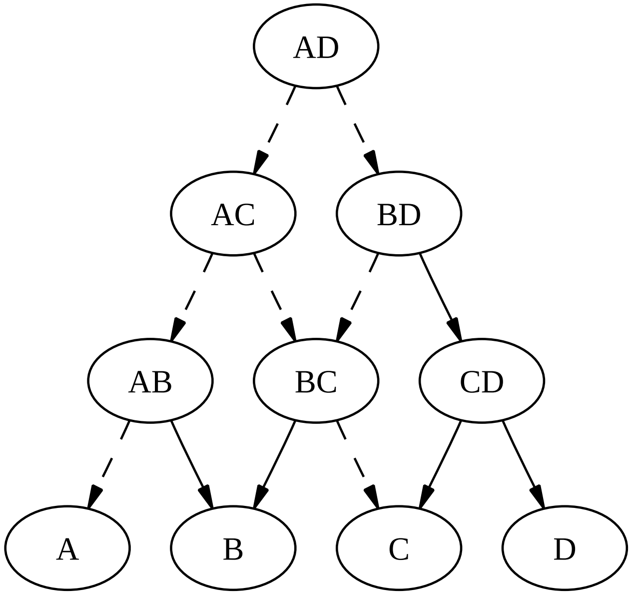


an example dichotomy, $L = \{A, B, C, D\}$. (d_x, B)

[a] Eibe F. and Kramer S. Ensembles of nested dichotomies for multi-class problems. Proc. of the 21st International conference of Machine Learning, 2004.

Classification by Pairwise Training

- [a]: Multi-label; DAG; Trichotomy



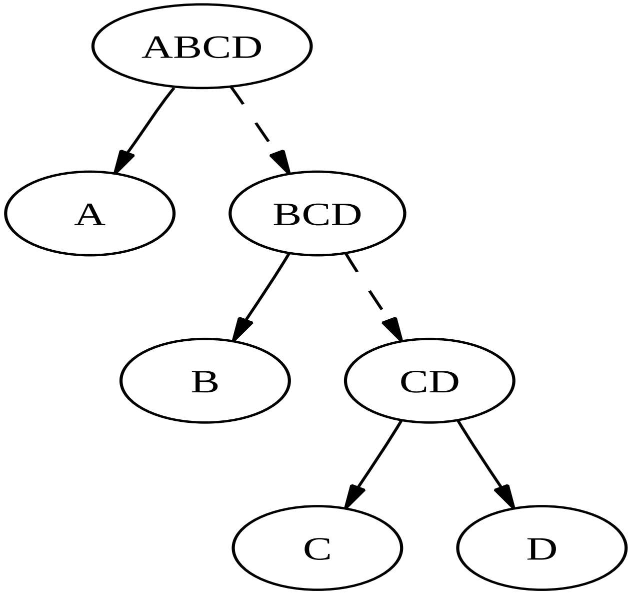
an example DAG, $L = \{A, B, C, D\}$. $(d_x, \{A, C\})$

- Error propagation
- Classification sensitive to DAG arrangement (Ensembles?)

[a] M. Chang, C. Lin and R. Weng. Multi-label Classification by Pairwise Training

M-END: Multi-label END

- Multi-label; Binary Tree; Dichotomy; Ensembles



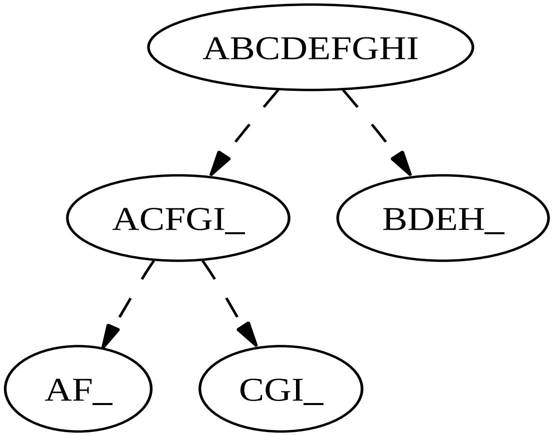
an example dichotomy, $L = \{A, B, C, D\}$. (d_x, C, D)

- Can return multiple labels if these labels co-occur often
- Creates splits according to label co-occurrences

A work in progress…

EsPS: Ensembles of $s$plit Pruned Sets

- A somewhat related method inspired by work on M-END



$L = \{A, \cdots, I, \_\}$ ($\_ = null$)

- Some param. to determine when branching stops
- Multi-label method (i.e. PS aka PPT) train on each leaf subset
- Classifications of all leaves are combined

A work in progress. . .

EsPS: Ensembles of $s$plit Pruned Sets

Compared to state-of-the-art RAKEL:

$Dataset$	RAKEL	CM	BM	RM	EsPS
Scene	71.58±0.89	71.81±1.22	58.28±0.92 •	71.72±0.98	73.51±0.56 ◦
Medical	72.55±2.32	74.71±1.32	73.00±1.08	72.71±1.56	76.86±1.60
Yeast	54.49±0.98	51.98±0.93 •	49.64±0.88 •	51.95±0.62 •	54.21±0.95
Enron	42.98±0.63	41.02±1.08	38.64±1.05 •	27.22±0.31 •	47.55±0.80 ◦
Reuters	31.80±0.29	49.17±0.67 ◦	31.91±0.76	49.08±0.59 ◦	46.52±1.03 ◦

◦, • statistically significant improvement or degradation

$Dataset$	PS.END.DNBND
Medical*	75.82±1.52

5 x 2 fold CV; all ensemble methods 10 iterations, all other parameters tuned via 5 fold internal CV. SMO used as internal single-label classifier in each case.

Summary

- Good probability distributions are important to threshold methods like RM
- Ensembles are a way to achieve this (plus the additional benefits of an ensemble)
  - END
- In multi-label classification it is very important to take into account label-correlations
  - M-END
  - EsPS