Multi-Label Classification with Meta Labels

Jesse Read[1], Antti Puurula[2], Albert Bifet[3]

(1) Aalto University and HIIT, Finland
(2) University of Waikato, New Zealand
(3) Huawei Noah's Ark Lab, Hong Kong

17 December 2014

Overview

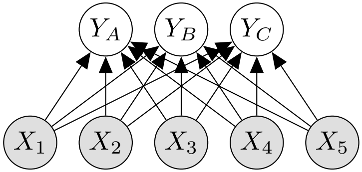
Multi-label classification:

$\subseteq \{\texttt{beach}, \texttt{sunset}, \textit{foliage}, \texttt{field}, \textit{mountain}, \texttt{urban}\}$

- Most multi-label classification methods can be expressed in a general framework of meta-label classification
- Our work combines labels into meta-labels, so as to learn dependence efficiently and effectively.

Multi-label Learning

With input variables X, produce predictions for *multiple* output variables Y. The basic binary relevance approach,

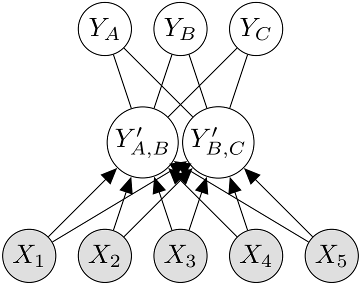


- does not capture label dependence (among Y-variables)
- does not scale to large number of labels

The label powerset approach models label combinations as class values in a multi-class problem.

Meta Labels

We introduce a layer of meta-labels,



- captures label dependence
- meta-labels can be fewer than the original number of labels
- and are deterministically decodable into the labels

Producing Meta Labels

dataset		binary relevance			label powerset	meta labels	
instance	labels	Y_A	Y_B	Y_C	$Y_{A,B,C}$	$Y_{A,C}$	$Y_{B,C}$
1	B	0	1	0	B	$\emptyset$	B
2	B,C	0	1	1	BC	$\emptyset$	BC
3	C	0	0	1	C	$\emptyset$	C
4	B	0	1	0	B	$\emptyset$	B
5	A,C	1	0	1	AC	AC	C
6	A,C	1	0	1	AC	AC	C
7	A,C	1	0	1	AC	AC	C
8	A,B,C	1	1	1	ABC	$\emptyset$	BC
9	C	0	0	1	C	$\emptyset$	C

- *binary relevance*: 9 exs, 3×2 binary classes
- *label powerset*: 9 exs, 1×5 multi-class

Producing Meta Labels

dataset		binary relevance			label powerset	meta labels	
instance	labels	Y_A	Y_B	Y_C	$Y_{A,B,C}$	$Y_{A,C}$	$Y_{B,C}$
1	B	0	1	0	B	$\emptyset$	B
2	B,C	0	1	1	BC	$\emptyset$	BC
3	C	0	0	1	C	$\emptyset$	C
4	B	0	1	0	B	$\emptyset$	B
5	A,C	1	0	1	AC	AC	C
6	A,C	1	0	1	AC	AC	C
7	A,C	1	0	1	AC	AC	C
8	A,B,C	1	1	1	ABC	$\emptyset$	BC
9	C	0	0	1	C	$\emptyset$	C

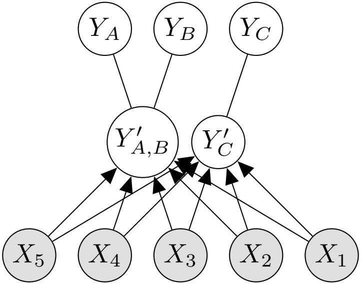
- *binary relevance*: 9 exs, 3×2 binary classes
- *label powerset*: 9 exs, 1×5 multi-class
- *pruned meta labels*: 9 exs, 2 meta labels, of 2 and 3 values

$$Y'_{AC} \in \{\emptyset, AC\}, Y'_{BC} \in \{B, C, BC\}$$

(one possible formulation)

Example 2

There is no need to model labels together if there is no strong dependence between them,



$$Y'_{A,B} \in \{\emptyset, B, AB\},\ Y'_C \in \{\emptyset, C\}$$

(e.g., no strong relation between C and the other labels)

General process for classification with meta-labels

1. Make a partition (either overlapping or disjoint) of the *label set*
2. Relabel the meta-labels, deciding on how many values each label can take (i.e., possibly pruning some)
3. Train classifiers to predict meta-labels from the input instances
4. Make predictions into the meta-label space
5. Recombine predictions into the label space

Voting Using Meta Labels

Table : Meta-label Vote, e.g., for $Y'_{AB} \in \{\emptyset, B, AB\}$

v	A	B	$p(Y'_{AB} = v \mid \tilde{\mathbf{x}})$
$\emptyset$	0	0	0.0
B	0	1	0.9
AB	1	1	0.1
P_{AB}	0.1	1.0	

Table : Labelset Voting: From Meta-labels to Labels

	A	B	C
P_{AB}	0.1	1.0	
P_{BC}		0.7	0.3
$\sum_k P_{k,j}$	0.1	1.7	0.3
$\hat{\mathbf{y}}$ (> 0.5)	0	1	0

Results

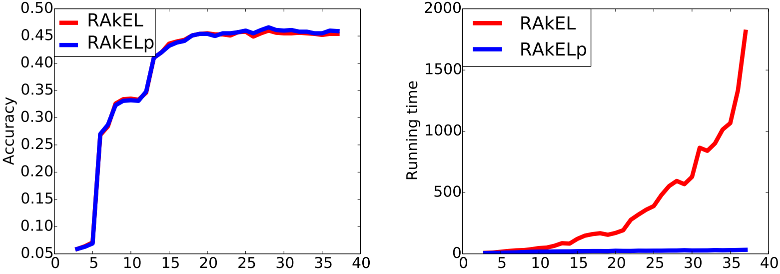


Figure : On Enron (1700 emails, 53 labels). RAkEL: random ensembles of 'label-powerset method' on subsets of size k (horizontal axis) vs RAkELp: with pruned meta-labels

Summary

- General framework of meta-labels for multi-label classification
- Unifies various approaches from the literature
- New models RAkELp and EpRd have large improvement in running time
- Part of solution for LSHTC4 (1st place) and WISE (2nd Place) Kaggle challenges
- Code available at

`http://meka.sourceforge.net`

Multi-Label Classification with Meta Labels

Jesse Read(1), Antti Puurula(2), Albert Bifet(3)

(1) Aalto University and HIIT, Finland
(2) University of Waikato, New Zealand
(3) Huawei Noah's Ark Lab, Hong Kong

17 December 2014

Aalto University

THE UNIVERSITY OF WAIKATO
Te Whare Wānanga o Waikato

HUAWEI