MULTI-TARGET PREDICTION: CHALLENGES AND APPLICATIONS

with: *N. Bassiliades, W. Groves, M. Laliotis, N. Markantonatos, F. Markatopoulou, C. & E. Papagiannopoulou, Y. Papanikolaou, E. Spyromitros-Xioufis, I. Tsamardinos, I. Vlahavas, A. Vrekou*



Grigorios Tsoumakas,
School of informatics,
Aristotle university of Thessaloniki

MULTI-TARGET PREDICTION

Tasks

Multi-label learning

Multi-target regression

Label ranking

Multi-task learning

Collaborative filtering

Dyadic prediction

Challenges

Exploiting dependencies among the targets

Scaling to extreme sizes of output spaces

Dealing with class imbalance

Target heterogeneity

Applications

Multimedia annotation

- Video, image, audio, text

Gene function prediction

Ecological modelling

Demand forecasting

Ensemble pruning

Willem Waegeman, Krzysztof Dembczynski, Eyke Hüllermeier, Multi-Target Prediction, Tutorial @ ICML 2013

MULTI-TARGET PREDICTION

Tasks

Multi-label learning

Multi-target regression

Label ranking

Multi-task learning

Collaborative filtering

Dyadic prediction

Challenges

Exploiting dependencies among the targets

Scaling to extreme sizes of output spaces

Dealing with class imbalance

Target heterogeneity

Applications

Multimedia annotation

- Video, image, audio, text

Gene function prediction

Ecological modelling

Demand forecasting

Ensemble pruning

Willem Waegeman, Krzysztof Dembczynski, Eyke Hüllermeier, Multi-Target Prediction, Tutorial @ ICML 2013

OUTLINE

1. Deterministic label relationships
2. From multi-label classification to multi-target regression



3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning



OUTLINE




1. Deterministic label relationships
   - *Papagiannopoulou, C., Tsoumakas, G., Tsamardinos, I. (2015). Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*
   - *Papagiannopoulou, E., Tsoumakas, G., Bassiliades, N. (2015). On Discovering Relationships in Multi-Label Learning via Linked Open Data, In Proceedings of Know@LOD Workshop of ESWC*
2. From multi-label classification to multi-target regression
3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning

MULTI-LABEL LEARNING

X_1	X_2	…	X_p	Y_1	Y_2	…	Y_q
0.12	1	…	12	0	1	…	1
2.34	9	…	-5	1	1	…	0
1.22	3	…	40	1	0	…	1
2.18	2	…	8	?	?	…	?
1.76	7	…	23	?	?	…	?

training examples

unknown instances

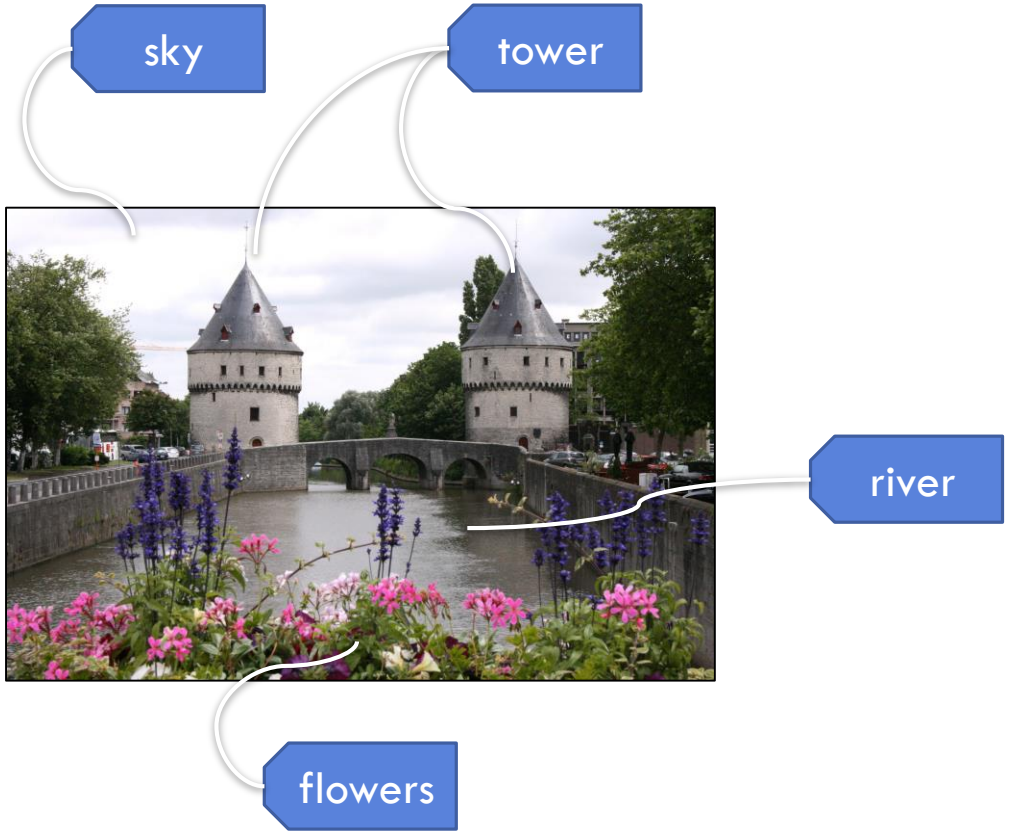
p input variables

q binary output variables

THE SEED

ImageCLEF 2011 challenge

- Automatic annotation of Flickr images
- JPG, EXIF information & user tags
- 99 concepts



THE QUESTION

Can we post-process the probabilities in a sound way so that they obey the relationships?

Label Relationships

Positive entailment

- River → Water
- Car → Vehicle

Mutual exclusion

- Autumn, Winter, Spring, Summer
- Single person, Small group, Big group, No persons

Sample Output

Label	Probability
River	0.7
Water	0.5
Autumn	0.6
Winter	0.4
Spring	0.2
Summer	0.1
…	…

EXTRACTING RELATIONSHIPS

Positive entailment

- $a \rightarrow b$ is extracted when $U = 0$
- $b \rightarrow a$ is extracted when $T = 0$
- The relationship's support is S

Mutual exclusion

- $a \rightarrow \neg b \wedge b \rightarrow \neg a$ is extracted when $S = 0$
- The relationship's support is $T + U$
- Higher order relationships are extracted following the Apriori algorithm paradigm

Contingency table for labels A and B

	a	$\neg a$
b	S	T
$\neg b$	U	V

TOY EXAMPLE

6 labels

10 training examples

A	B	C	D	E	F
1	1	1	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	1	1	0	1	0
1	1	1	0	0	0
0	1	1	1	0	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	0	0	0	0	1

TOY EXAMPLE

Positive entailments

- $a \rightarrow b$ (support 3)

A	B	C	D	E	F
1	1	1	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	1	1	0	1	0
1	1	1	0	0	0
0	1	1	1	0	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	0	0	0	0	1

TOY EXAMPLE

Positive entailments

- $a \rightarrow b$ (support 3)
- $a \rightarrow c$ (support 3)

A	B	C	D	E	F
1	1	1	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	1	1	0	1	0
1	1	1	0	0	0
0	1	1	1	0	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	0	0	0	0	1

TOY EXAMPLE

Positive entailments

- $a \rightarrow b$ (support 3)
- $a \rightarrow c$ (support 3)
- $b \rightarrow c$ (support 5)

A	B	C	D	E	F
1	1	1	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	1	1	0	1	0
1	1	1	0	0	0
0	1	1	1	0	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	0	0	0	0	1

TOY EXAMPLE

Positive entailments

- $a \rightarrow b$ (support 3)
- $a \rightarrow c$ (support 3)
- $b \rightarrow c$ (support 5)
- $d \rightarrow c$ (support 3)

A	B	C	D	E	F
1	1	1	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	1	1	0	1	0
1	1	1	0	0	0
0	1	1	1	0	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	0	0	0	0	1

TOY EXAMPLE

Positive entailments

- $a \rightarrow b$ (support 3)
- $a \rightarrow c$ (support 3)
- $b \rightarrow c$ (support 5)
- $d \rightarrow c$ (support 3)

Mutual exclusion

- $\{A, E, F\}$ (support 9)

A	B	C	D	E	F
1	1	1	0	0	0
1	1	1	1	0	0
0	0	0	0	1	0
0	1	1	0	1	0
1	1	1	0	0	0
0	1	1	1	0	1
0	0	1	1	1	0
0	0	0	0	0	1
0	0	1	0	0	0
0	0	0	0	0	1

EXPLOITING RELATIONSHIPS: POSITIVE ENTAILMENT

Label A entails label B

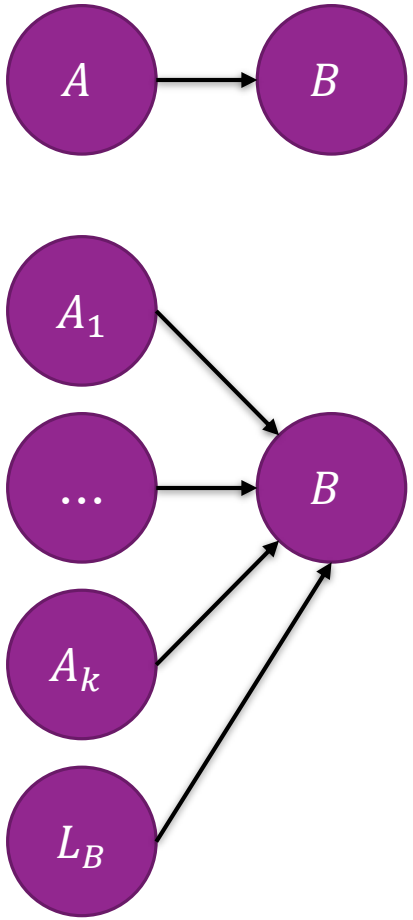
- $a \rightarrow b$

Generalization

- $a_1 \rightarrow b, \ldots, a_k \rightarrow b$

Leak node

- To consider other causes of B
- Virtual label equal to
  - True where B is true and all of its parents are false
  - False in all other training examples



	b	$\neg b$
a	1	0
$\neg a$	0	1

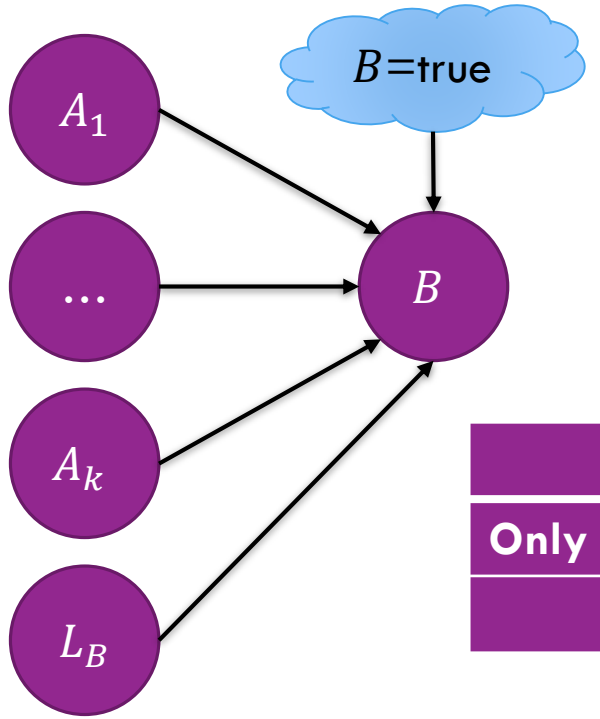
	b	$\neg b$
At least one parent of B true	1	0
otherwise	0	1

EXPLOITING RELATIONSHIPS: MUTUAL EXCLUSION

Among k labels $A_1, \ldots, A_k$

Leak node

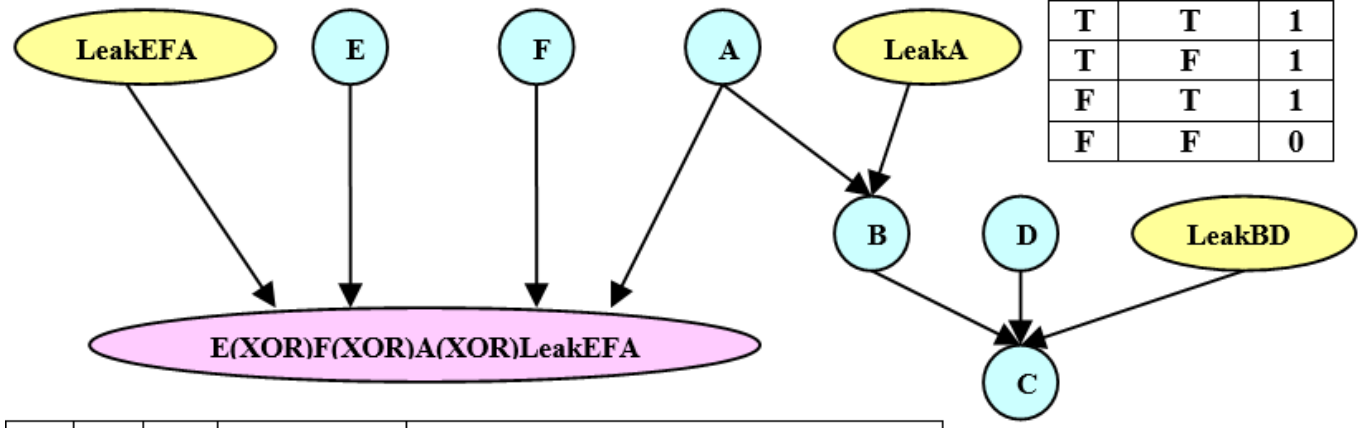
- To cover all training examples, i.e. to become exhaustive
- Virtual label equal to
  - True where all other parents of B are false
  - False in all other examples



	b	$\neg b$
Only one parent of B true	1	0
otherwise	0	1

TOY EXAMPLE

Node	Before	After
A	0.400	0.022
$LeakA$	0.350	0.082
B	0.250	0.096
D	0.600	0.031
$LeakBD$	0.010	0.050
C	0.200	0.345
F	0.300	0.064
E	0.850	0.850
$LeakEFA$	0.300	0.064



A	LeakA	B
T	T	1
T	F	1
F	T	1
F	F	0

E	F	A	LeakEFA	E(XOR)F(XOR)A(XOR)LeakEFA
T	T	T	T	0
T	T	T	F	0
T	T	F	T	0
T	T	F	F	0
T	F	T	T	0
T	F	T	F	0
T	F	F	T	0
T	F	F	F	1
F	T	T	T	0
F	T	T	F	0
F	T	F	T	0
F	T	F	F	1
F	F	T	T	0
F	F	T	F	1
F	F	F	T	1
F	F	F	F	0

B	D	LeakBD	C
T	T	T	1
T	T	F	1
T	F	T	1
T	F	F	1
F	T	T	1
F	T	F	1
F	F	T	1
F	F	F	0

EMPIRICAL STUDY

12 multi-label datasets

Relationship discovery

- Minimum support of 2 – increase exponentially in case of memory issues

Learning

- Binary Relevance + Random Forest with 10 trees
- Weka, Mulan

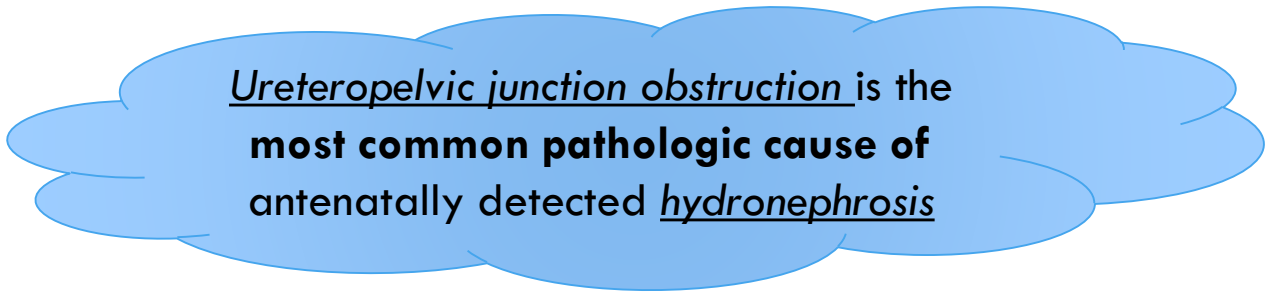
Inference

- Virtual evidence insertion, exact inference via clustering algorithm
- jSMILE library

POSITIVE ENTAILMENT IN "MEDICAL"

3 entailment relationships extracted from 978 radiologists' reports annotated with ICD-9 codes

			Sup.
Congenital obstruction of ureteropelvic junction	→	Hydronephrosis	4
Shortness of breath	→	Renal agenesis and dysgenesis	3
Vomiting alone	→	Renal agenesis and dysgenesis	3

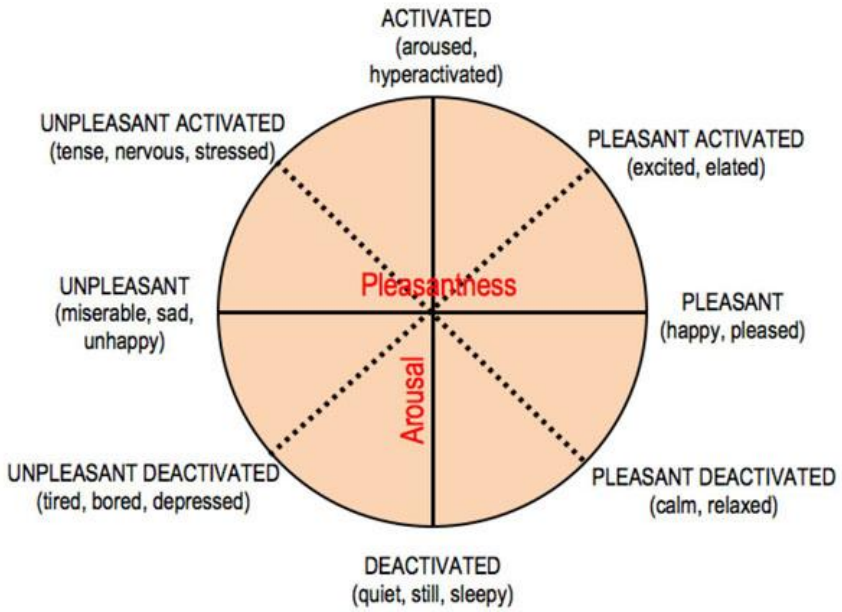




MUTUAL EXCLUSION

Emotions

quiet-still XOR amazed-surprised



Enron

"*Company Business, Strategy, etc.*"
XOR "friendship / affection"



In business, sir, one has no friends, only correspondents

~Alexandre Dumas

RESULTS: POSITIVE ENTAILMENT

Wilcoxon test
P-value 0.0156

Dataset	Minimum Support	Number of Labels	Number of Relations	% MAP Improvement
Bibtex	2	159	11	0.279
Bookmarks	2	208	4	0.068
Enron	2	53	4	0.391
ImageCLEF2011	2	99	28	2.977
ImageCLEF2012	2	94	1	0.168
Medical	2	45	6	2.284
Yeast	2	14	3	1.584

RESULTS: MUTUAL EXCLUSION (1/2)

Wilcoxon test
P-value 0.1099

Dataset	Minimum Support	Number of Labels	Number of Relations	% MAP Improvement
Bibtex	128	159	76	-1.626
Bookmarks	2048	208	1	-0.068
Emotions	8	6	1	1.424
Enron	2	53	481	-8.434
ImageCLEF2011	32	99	325	1.865
ImageCLEF2012	64	94	278	-2.862
IMDB	2	28	22	4.222
Medical	16	45	31	3.769
Scene	2	6	4	3.023
Slashdot	2	22	23	11.803
TMC2007	2	22	8	6.044
Yeast	2	14	2	1.760

RESULTS: MUTUAL EXCLUSION (2/2)

	Bibtex		Enron		ImageCLEF2011		ImageCLEF2012	
Minimum Support	128	256	2	32	32	128	64	256
Number of Relationships	76	3	481	22	325	56	278	40
% MAP Improvement	-1.63	0,60	-8.4	0.21	1.87	3.34	-2.87	0.63

OUTLINE




1. Deterministic label relationships
   - *Papagiannopoulou, C., Tsoumakas, G., Tsamardinos, I. (2015). Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*
   - *Papagiannopoulou, E., Tsoumakas, G., Bassiliades, N. (2015). On Discovering Relationships in Multi-Label Learning via Linked Open Data, In Proceedings of Know@LOD Workshop of ESWC*
2. From multi-label classification to multi-target regression
3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning

FINDING RELATIONSHIPS VIA LINKED OPEN DATA

Main question

- Can we discover label relationships by looking labels up at the LOD cloud?

Focus of this first study

- WordNet as the LOD source
- Discover common label ancestors
- Extend the output space with the common ancestors
- Apply algorithms that exploit label relationships aiming at improved prediction accuracy

OUR APPROACH

Steps

1. Preprocessing of label names
   - Tokenization
   - Removal of tokens that appear in all labels
2. Look up label names in WordNet
3. Follow hypernym-synsets up to root
4. Find common ancestors
5. Expand output space

Example

Dataset	Sample Label 1	Sample Label 2
Bibtex	**TAG**_physics	**TAG**_math
ImageCLEF	Winter_**attr**	Summer_**attr**

OUR APPROACH

Steps

1. Preprocessing of label names
2. Look up label names in WordNet
   - Assume correct sense is most frequent one
3. Follow hypernym-synsets up to root
4. Find common ancestors
5. Expand output space

Example

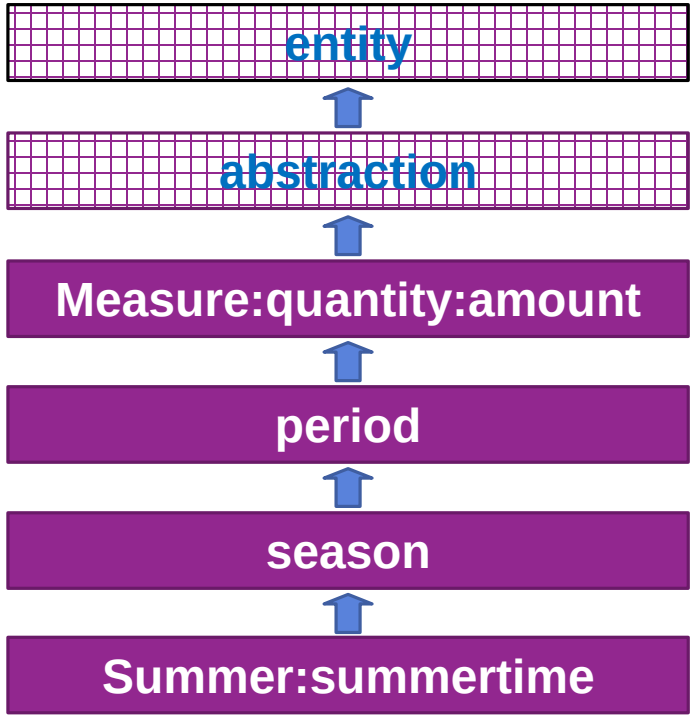
Label	Sense
Winter	**The coldest season of the year (24)**
-	Spend the winter (2)
Summer	**The warmest season of the year (58)**
-	The period of nest development, happiness or beauty
-	Spend the summer (0)

OUR APPROACH

Steps

1. Preprocessing of label names
2. Look up label names in WordNet
3. Follow hypernym-synsets up to root
   - Avoid general concepts
4. Find common ancestors
5. Expand output space

Example



OUR APPROACH

Steps

1. Preprocessing of label names
2. Look up label names in WordNet
3. Follow hypernym-synsets up to root
4. Find common ancestors
5. Expand output space

Example



OUR APPROACH

Steps

1. Preprocessing of label names
2. Look up label names in WordNet
3. Follow hypernym-synsets up to root
4. Find common ancestors
5. Expand output space

Example

x	summer	winter	…	season
$x^{(1)}$	0	1	…	1
$x^{(2)}$	1	0	…	1
…	…	…	…	…
$x^{(m)}$	0	0	…	0

EMPIRICAL STUDY: TWO VARIATIONS

LOD1

Hypernym-synsets up to 2 layers

Ignores the following 32 concepts

whole, unit, object, entity, abstraction, substance, content, message, theme, topic, subject, domain, activity, individual, someone, somebody, mortal, soul, organism, being, cause, go, locomote, formation, alter, modify, change, alteration, modification, happening, occurence, occurent

LOD2

Hypernym-synsets up to the root

Ignores the following 5 concepts

whole, unit, object, entity, abstraction

EMPIRICAL STUDY: SETUP AND RESULTS

Setup

6 multi-label datasets, without obscure label names

Labels Found and Ancestors Added

Dataset	Samples	Labels	Found	LOD1	LOD2
IC2011	8,000	99	75	17	69
bibtex	7,395	159	119	8	63
delicious	16,105	983	778	190	319
bookmarks	87,856	208	148	22	79
corel5k	5,000	374	367	84	214
IMDB-F	120,919	28	23	5	14

EMPIRICAL STUDY: SETUP AND RESULTS

Setup

6 multi-label datasets, without obscure label names

Calibrated Label Ranking

- Linear Support Vector Machines

Hold-out evaluation 70/30%

Mean Average Precision

Dataset	CLR	LOD1	LOD2
IC2011	.3275	.3263	.3245
bibtex	.3757	.3838	.3833
delicious	.1596	.1646	.1661
bookmarks	.2353	.2435	.2358
corel5k	.0612	.0584	.0580
IMDB-F	.1161	.1163	.1160

EMPIRICAL STUDY: SETUP AND RESULTS

Setup

6 multi-label datasets, without obscure label names

Calibrated Label Ranking

- Linear Support Vector Machines

Hold-out evaluation 70/30%

Logarithmic Loss

Dataset	CLR	LOD1	LOD2
IC2011	.7221	.7003	.7315
bibtex	.9288	.8840	.7881
delicious	.9273	.8305	.7901
bookmarks	.9401	.8842	.7945
corel5k	.9795	.8503	.7481
IMDB-F	.8256	.7109	.6411

OUTLINE





1. Deterministic label relationships
2. From multi-label classification to multi-target regression
   - *Tsoumakas, G., Spyromitros-Xioufis, E., Vrekou, A., Vlahavas, I. (2014). Multi-target Regression via Random Linear Target Combinations. In Proceedings of ECML PKDD 2014: 225-240*
   - *Spyromitros-Xioufis, E., Tsoumakas, G., Groves, W., Vlahavas, I. (2016) Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs. Machine Learning Journal 104(1), 55-98*
3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning

MULTI-TARGET (AKA MULTIVARIATE) REGRESSION

X_1	X_2	...	X_p	Y_1	Y_2	...	Y_q
0.12	1	…	12	0.14	10	…	-1.3
2.34	9	…	-5	4.15	12	…	-2.0
1.22	3	…	40	1.01	28	…	-5.3
2.18	2	…	8	?	?	…	?
1.76	7	…	23	?	?	…	?

training examples

unknown instances

p input variables

q continuous output variables

THE QUESTION

What is the equivalent of RAkEL[1] for multi-target regression?

- Take a random subsets of the labels
- Form new multi-class target, concerning all different output vectors of the selected subset

[1] *G. Tsoumakas, I. Vlahavas, (2007) Random k-Labelsets: An Ensemble Method for Multilabel Classification, Proc. ECML PKDD 2007, pp. 406-417, Warsaw, Poland, 2007*

THE QUESTION

What is the equivalent of RAkEL[1] for multi-target regression?

- Take a random subsets of the labels
- Form new multi-class target, concerning all different output vectors of the selected subset

Random Linear Combinations (RLC)

- Take a random subset of the targets
- Form a new continuous target, concerning a random linear combination of the selected subset

[1] *G. Tsoumakas, I. Vlahavas, (2007) Random k-Labelsets: An Ensemble Method for Multilabel Classification, Proc. ECML PKDD 2007, pp. 406-417, Warsaw, Poland, 2007*

SKETCHING RLC

q targets

	y_1	y_2	y_3
1	-0,5	-0,2	0,4
2	0,5	-0,3	-1
3	0,9	0,6	-0,5
4	-0,8	-0,5	-0,9
5	-0,5	0,6	0,7
6	0,4	0,1	0,8
7	-0,2	-0,3	0,8
8	-0,4	-0,4	-0,9

random linear combinations of the original targets



$r \gg q$ targets

	z_1	z_2	z_3	z_4	z_5	z_6
1						
2						
3						
4						
5						
6						
7						
8						

SKETCHING RLC

q targets

	y_1	y_2	y_3
1	-0,5	-0,2	0,4
2	0,5	-0,3	-1
3	0,9	0,6	-0,5
4	-0,8	-0,5	-0,9
5	-0,5	0,6	0,7
6	0,4	0,1	0,8
7	-0,2	-0,3	0,8
8	-0,4	-0,4	-0,9

random linear combinations of the original targets

$$Z = YC$$

0	0,7	-0,6	-0,6	0	-0,8
0,1	0	0,4	0	-0,4	-0,5
0,7	0,3	0	0,9	-0,2	0

$q \times r$ coefficient matrix C of standard uniform values

$r \gg q$ targets

	z_1	z_2	z_3	z_4	z_5	z_6
1	0,26	-0,23	0,22	0,66	0	0,5
2	-0,73	0,05	-0,42	-1,2	0,32	-0,25
3	-0,29	0,48	-0,3	-0,99	-0,14	-1,02
4	-0,68	-0,83	0,28	-0,33	0,38	0,89
5	0,55	-0,14	0,54	0,93	-0,38	0,1
6	0,57	0,52	-0,2	0,48	-0,2	-0,37
7	0,53	0,1	0	0,84	-0,04	0,31
8	-0,67	-0,55	0,08	-0,57	0,34	0,52

Multi-target regression model

solving a system of r linear equations with q unknowns

?	?	?

-0,2	1	0,2	-0,5	-0,4	0,5

SOME MORE DETAILS

q targets

	y_1	y_2	y_3
1	-0,5	-0,2	0,4
2	0,5	-0,3	-1
3	0,9	0,6	-0,5
4	-0,8	-0,5	-0,9
5	-0,5	0,6	0,7
6	0,4	0,1	0,8
7	-0,2	-0,3	0,8
8	-0,4	-0,4	-0,9

$$Z = YC$$

0	0,7	-0,6	-0,6	0	-0,8
0,1	0	0,4	0	-0,4	-0,5
0,7	0,3	0	0,9	-0,2	0

$q \times r$ coefficient matrix C of standard uniform values

Assumption: original targets take values from the same domain

Parameter $k \in \{2..q\}$ (number of targets being combined)

EMPIRICAL STUDY

Methods

ST

- A **single** regression model per **target**
- Stochastic gradient boosting

MORF

- **Multi-objective random forest**
- 100 trees

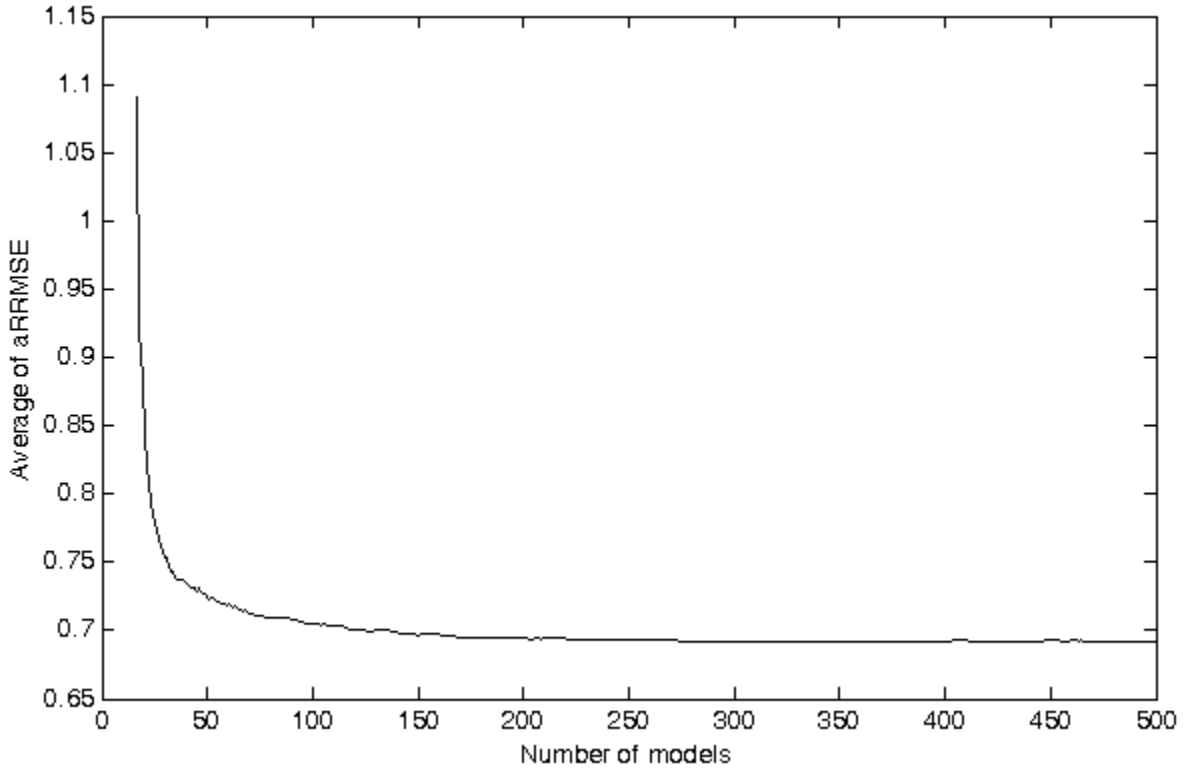
RLC

- Multi-target regression algorithm: ST
- Solving system of equations: least squares

Datasets

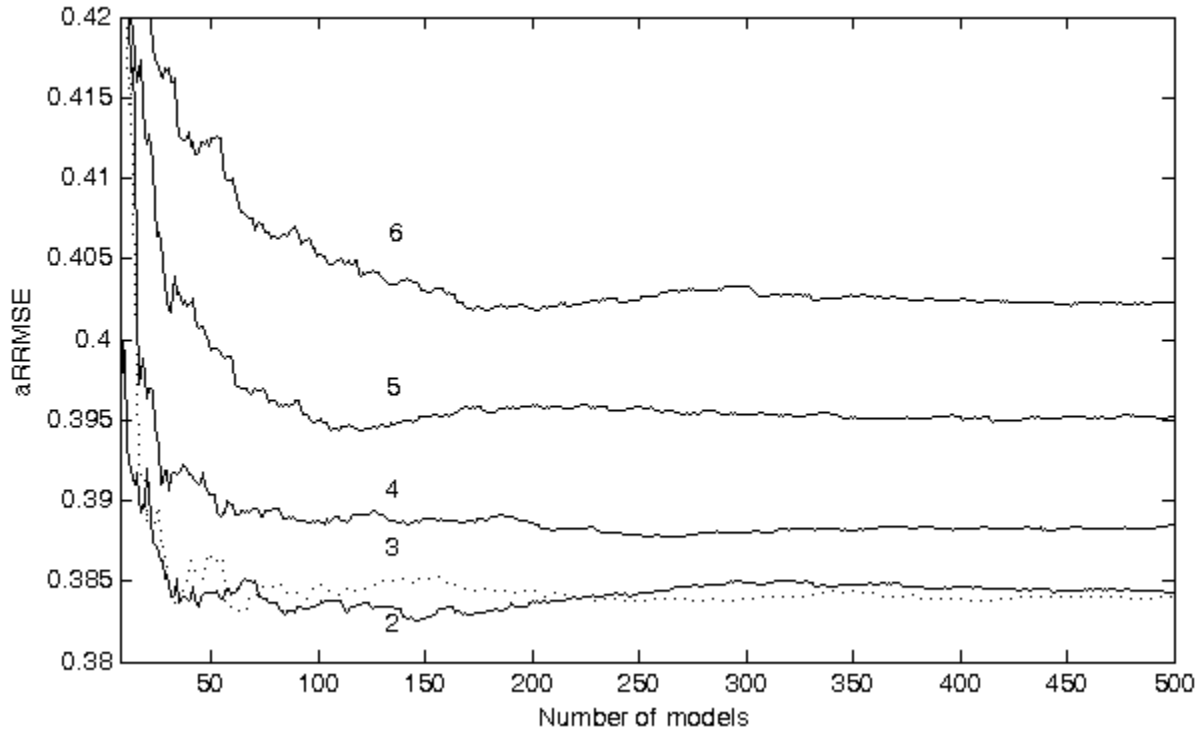
Id	Name	Targets
1,2	Airline Ticket Price 1 / 2	6
3	Electrical Discharge Machining	2
4,5	Occupational Employment Survey 1 / 2	16
6,7	River Flow 1 / 2	8
8,9	Solar Flare 1 / 2	3
10,11	Supply Chain Management 1 / 2	16
12	Water Quality	14

STUDYING THE r PARAMETER



Average of aRRMSE of our method (y-axis) with respect to r (x-axis) across all datasets and all k values

STUDYING THE k PARAMETER



aRRMSE of our method (y-axis) at the atp1d dataset with respect to r (x-axis) for $k \in \{2, 3, 4, 5, 6\}$

MAIN RESULTS

For $k=2/3$ and $r=500$ models

	RLC	ST	MORF
RLC	-	10:2	8:4
ST	2:10	-	7:5
MORF	4:8	5:7	-

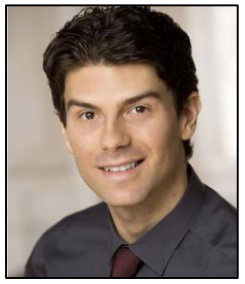
	RLC	ST	MORF
Avg. Rank	1.5	2.25	2.25

ST with stochastic gradient boosting appears to be a strong baseline

RLC is better than ST and MORF

Wilcoxon signed-rank test at 95% shows statistically significant difference between RLC and ST

OUTLINE


1. Deterministic label relationships
2. From multi-label classification to multi-target regression
   - *Tsoumakas, G., Spyromitros-Xioufis, E., Vrekou, A., Vlahavas, I. (2014). Multi-target Regression via Random Linear Target Combinations. In Proceedings of ECML PKDD 2014: 225-240*
   - *Spyromitros-Xioufis, E., Tsoumakas, G., Groves, W., Vlahavas, I. (2016) Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs. Machine Learning Journal 104(1), 55-98*
3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning

TREATING TARGETS AS INPUTS

For each target variable, treat (some of) the rest as inputs

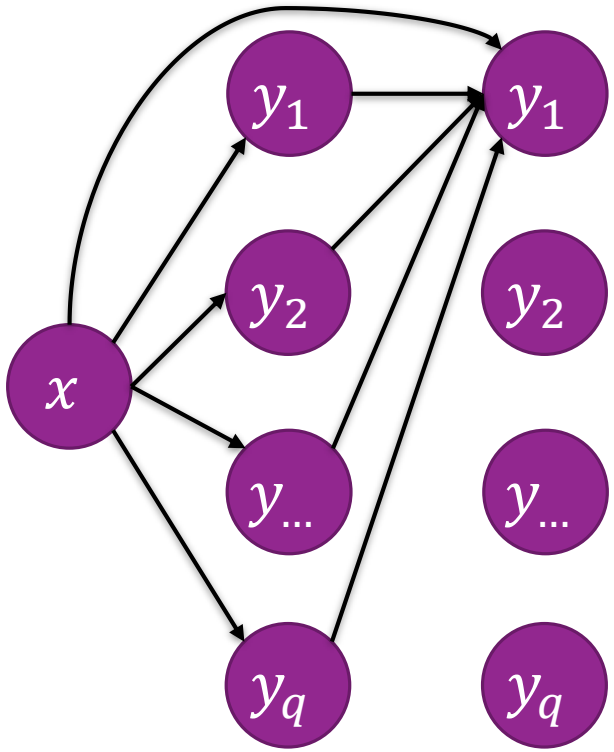
- A conceptually simple way of exploiting target dependencies

Already done in multi-label classification

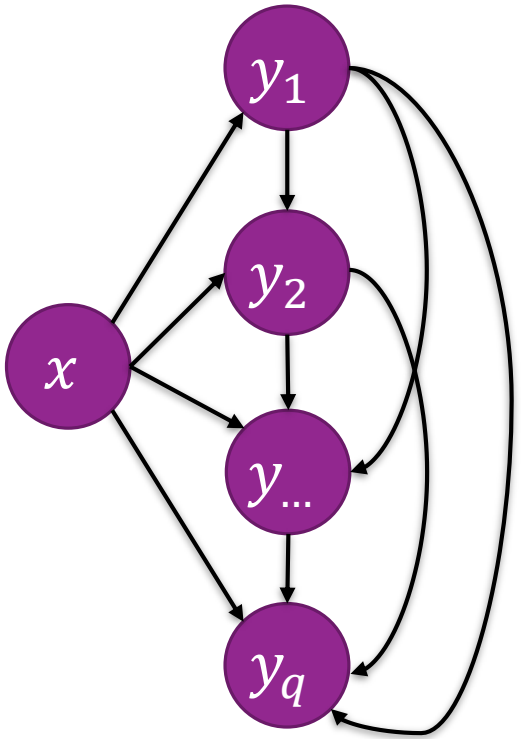
- *Read J, Pfahringer B, Holmes G, Frank E (2011) Classifier chains for multi-label classification. Machine Learning Journal 85(3):333–359*
- *Godbole S, Sarawagi S (2004) Discriminative methods for multi-labeled classification. In Proceedings of PAKDD 2004, Sydney, Australia, May 26-28, 2004, pp 22–30*

TWO "NEW" ALGORITHMS

Stacked Single Target (SST)



Ensemble of Regressor Chains (ERC)



GENERATING META-INPUT DATA

What values are used for the targets when they are used as inputs?

- SST uses in-sample estimates (*train*)
- ERC uses actual true values instead of estimates (*true*)
- In practice these will be based on out-of-sample estimates!

Proposal

- Use out-of-sample estimates based on cross-validation (*cv*)

EMPIRICAL STUDY

18 datasets (several of them first used in this paper)

RRMSE, statistical testing per dataset and per target

Multi-task learning algorithms: TNR, Dirty

Multi-target learning algorithms: MORF, ST, RLC, SST and ERC variants

Several base learning algorithms: (bagged) regression trees, ridge regression, stochastic gradient boosting, support vector regression

ON WEAK AND STRONG BASE LEARNERS

Base learners mean rank

Learner	Dataset	Target
Bagging	1.83 (1)	1.67 (1)
SGB	2.30 (2)	2.81 (2)
Ridge	3.22 (3)	2.97 (3)
SVR	3.56 (4)	3.75 (4)
Trees	3.89 (5)	3.80 (5)

Strong learners can make a difference

Multi-task learning approaches fail

State-of-the-art mean rank

	Dataset		Target	
Learner	Bagging	Ridge	Bagging	Ridge
SST-train	2.44 (1)	4.11 (5)	2.62 (1)	3.92 (5)
ST	2.94 (2)	3.94 (4)	3.27 (3)	3.35 (4)
RLC	3.11 (3)	2.89 (2)	2.81 (2)	3.18 (3)
ERC-true	3.39 (4)	3.33 (3)	3.57 (5)	3.04 (2)
MORF	4.06 (5)	2.39 (1)	3.41 (4)	2.40 (1)
Dirty	5.94 (6)	5.61 (6)	6.13 (6)	5.91 (6)
TNR	6.11 (7)	5.72 (7)	6.19 (7)	6.20 (7)

ON VARIANTS OF META-INPUT GENERATION

Variants mean rank

Learner	Dataset	Target
ERC-train	1.67 (1)	1.78 (2)
ERC-cv	1.72 (2)	1.64 (1)
ERC-true	2.61 (3)	2.59 (3)

Learner	Dataset	Target
SST-cv	1.56 (1)	1.58 (1)
SST-train	1.72 (2)	1.87 (2)
SST-true	2.72 (3)	2.55 (3)

Best variants vs state-of-the-art

Learner	Dataset	Target
ERC-cv	2.28 (1)	2.46 (1)
ERC-train	2.33 (2)	2.57 (2)
SST-cv	3.11 (3)	3.24 (3)
RLC	3.67 (4)	3.48 (4)
MORF	4.44 (5)	3.91 (5)
Dirty	6.00 (6)	6.16 (6)
TNR	6.17 (7)	6.18 (7)

OUTLINE






1. Deterministic label relationships
2. From multi-label classification to multi-target regression
3. Semantic indexing of biomedical literature
   - *Papanikolaou, Y., Tsoumakas, G., Laliotis, M., Markantonatos, N., Vlahavas, I. (to appear) Large-Scale Online Semantic Indexing of Biomedical Articles via an Ensemble of Multi-Label Classification Models, Journal of Biomedical Semantics*
4. Multi-label classification for instance-based ensemble pruning

SEMANTIC INDEXING OF SCIENTIFIC PUBLICATIONS

Collaboration with Atypon Inc.



- Literatum is Atypon's online content hosting and management platform
- Atypon is home to more than one-third of the world's English-language professional and scholarly journals—more than any other technology company
- Atypon's clients include Elsevier, IEEE, MIT Press, Oxford University Press, Taylor & Francis, …
- Literatum provides rapid UI/UX development tools, access management, SEO and discovery, content targeting, subscription modeling, automated semantic indexing, eCommerce and analytics
- Atypon was acquired by John Wiley & Sons earlier this year for $120,000,000

Automated semantic indexing

- LDA models that extract latent topics
- Multi-label models that classify articles to given ontologies

LARGE-SCALE ONLINE BIOMEDICAL INDEXING

Corpus (training abstracts)

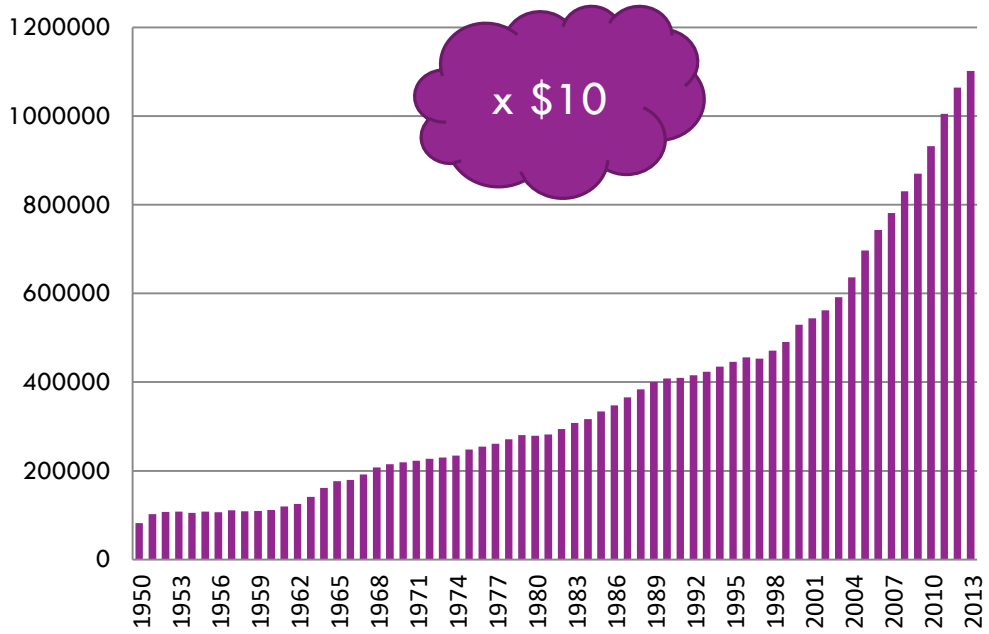
- 10,876,004 (18Gb)

MeSH terms

- 26,563

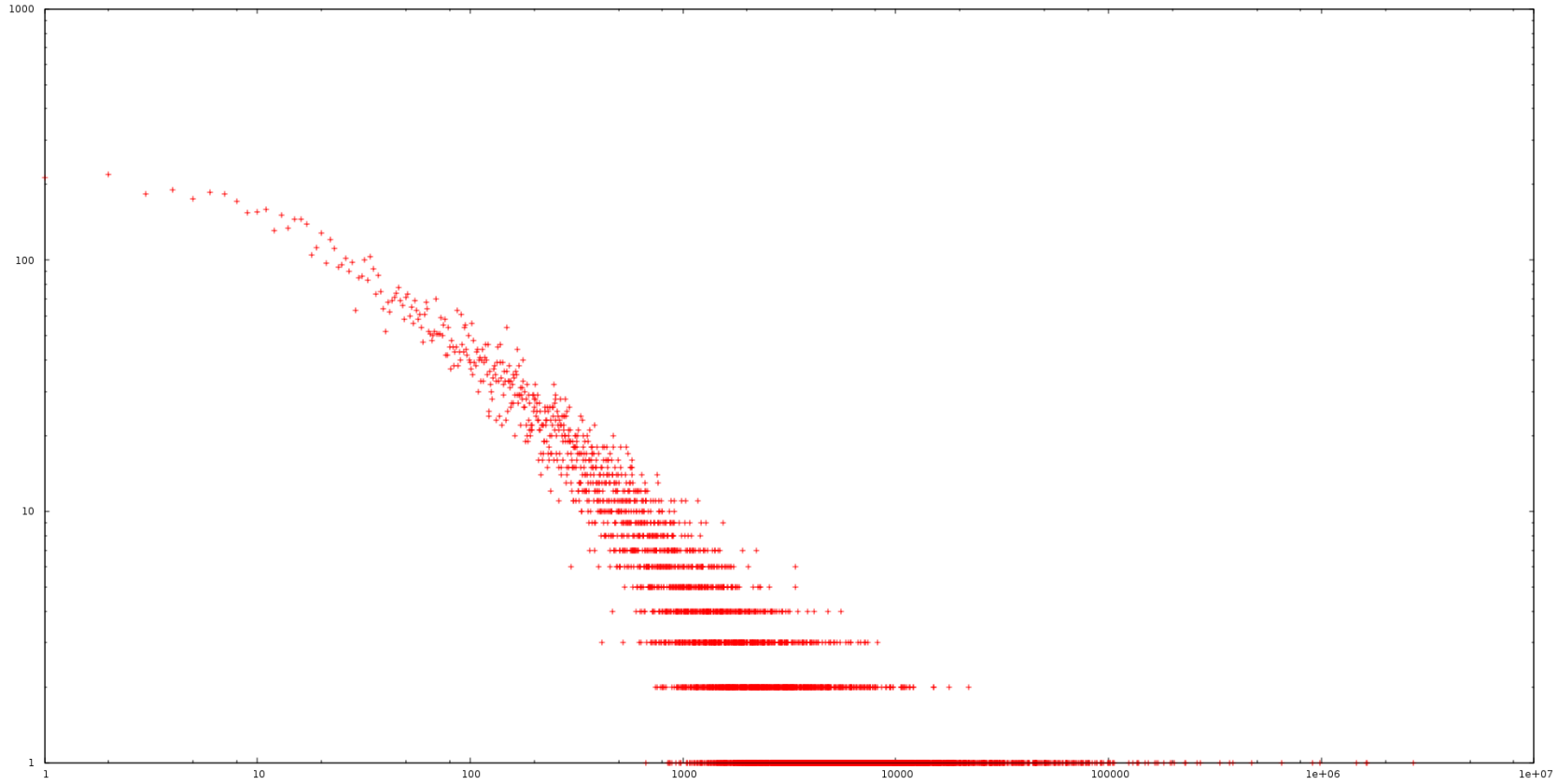
Online test setting

- 3 phases, 6 weeks per phase
- 87,080 docs (4,838 docs per week avg)
- BioASQ requested MeSH terms for 790 to 10,139 abstracts within 21 hours



1 million docs / year ≅ 2,740 docs / day

LABEL FREQUENCIES



4.3 million abstracts

213 labels with 1 example

1,680 labels with less than 10 examples

4 labels with more than 1 million examples

ENSEMBLE APPROACHES

For each label select the model that improves the corresponding F-measure most

- *Jimeno-Yepes, A., Mork, J.G., Demner-Fushman, D., Aronson, A.R.: A one-size-fits-all indexing method does not exist: Automatic selection based on meta-learning. JCSE 6(2), 151-160 (2012)*
- It can't be used for optimizing a global non-decomposable evaluation measure, such as micro-F

Iteratively select the model that improves micro-F

- *Fan, R.E., Lin, C.J.: A study on threshold selection for multi-label classification. Technical report, National Taiwan University (2007)*
- Can we trust the evaluation based on only a few positive samples?



Based on e-mail discussion with Prof. Charles Elkan, UCSD, Amazon

Our Aproach

- For each label use a McNemar test to assess the significance in micro-F improvement compared to the model that is globally best across all labels

MULE: MULTI-LABEL ENSEMBLE

1. Apply all models on a validation set and determine the globally best one h^*
2. Determine for each label which model(s) would lead to an improvement of the global evaluation measure compared to h^*
3. Compare the differences in predictions of each one of these models against the predictions of h^* using a McNemar test
   - Globally best model has been determined by much more positive samples (all labels)
   - Robustness to uncertainty due to label rarity
4. If the null hypothesis is rejected for more than one model, select the one for which the null hypothesis has lowest probability

EMPIRICAL RESULTS IN BIOASQ

Model	Micro-F	Macro-F
Vanilla SVMs	0.5568	0.4789
Weighted SVMs	0.5665	0.5102
MetaLabeler	**0.5855**	**0.5488**
Labeled LDA	0.3698	0.3010

Ensemble	Micro-F	Macro-F
Improve F	0.5584 (all)	0.5339 (MetaLabeler, Weighted SVM)
Improve Micro-F	0.5867 (all)	-
MULE	**0.5892 (all)**	**0.5492 (MetaLabeler, Labeled LDA)**

In the BioASQ challenge

- 2013 – 2016: consistently better than the production system of the National Library of Medicine
- 2013: 1st position, 2014 – 2016 2nd position

OUTLINE




1. Deterministic label relationships
2. From multi-label classification to multi-target regression
3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning
   - *Markatopoulou, F., Tsoumakas G., Vlahavas I., (2015) Dynamic ensemble pruning based on multi-label classification, Neurocomputing, Volume 150, Part B, 20 February 2015, Pages 501-512.*

INSTANCE-BASED ENSEMBLE PRUNING

Number of Models

Classifier fusion

- Combine all models, e.g. with voting

Classifier selection

- Select one model, e.g. based on accuracy

Ensemble selection/pruning

- Select a subset of the available models, e.g. based on accuracy

Test Instance Awareness

Static

- Fixed weights or model or subset of models for all test instances

Dynamic or Instance-based

- Different weights or model or subset of models per test instance

MULTI-LABEL LEARNING FORMULATION

Consider as example the segment dataset of the European project Statlog, where the target attribute has 6 values: *brickface*, *sky*, *foliage*, *cement*, *window*, *path*, *grass*.

x	y
$x^{(1)}$	sky
$x^{(2)}$	window
…	…
$x^{(m)}$	foliage

training set

o_1	o_2	…	o_t
path	sky	…	cement
window	grass	…	window
…	…	…	…
foliage	foliage	…	path

classifier prediction

x	o_1	o_2	…	o_t
$x^{(1)}$	0	1	…	0
$x^{(2)}$	1	0	…	1
…	…	…	…	…
$x^{(m)}$	1	1	…	0

multi-label training set

OPERATION AND IMPLICATIONS

Consider a set of t models $M = \{h_1, h_2, \ldots, h_t\}$

Given a new instance x, the multi-label classifier outputs a set of models $Z \subset M$

- $Z = \emptyset$ is a generally acceptable multi-label output
- We should augment the multi-label classifier with the constrain $Z \neq \emptyset$

OPERATION AND IMPLICATIONS

Consider a set of t models $M = \{h_1, h_2, \ldots, h_t\}$

Given a new instance x, the multi-label classifier outputs a set of models $Z \subset M$

Then, each of the models in Z is queried and their decisions are combined via plurality voting aka simple majority voting

If $W \subset M$ is the set of models that correctly predict the class of x,
then, assuming binary classification, the voting process is correct when $\frac{|Z \cap W|}{|Z|} > 0.5$

- The left-hand side is equal to the example-based precision of the multi-label classifier
- Precision could form the target loss function for multi-label learners in this task

OPERATION AND IMPLICATIONS

Consider a set of t models $M = \{h_1, h_2, \ldots, h_t\}$

Given a new instance x, the multi-label classifier outputs a set of models $Z \subset M$

Then, each of the models in Z is queried and their decisions are combined via plurality voting aka simple majority voting

If $W \subset M$ is the set of models that correctly predict the class of x,
then, assuming binary classification, the voting process is correct when $\frac{|Z \cap W|}{|Z|} > 0.5$

More often than not, multi-label models output a score for each label, typically a value between 0 and 1, instead of a set of relevant labels

- An intuitive threshold that is used to obtain the set of relevant labels is 0.5
- Increasing this threshold would increase the precision of the multi-label model

EMPIRICAL STUDY

Setup

40 datasets from the UCI repository

200 models

- 40 MLPs, 60 kNNs, 80 SVMs, 20 DTs

Approaches

- Majority voting (MV)
- OLA, LCA, MCB, DV – DVS, KNORA
- ML-*k*NN with auto thresholding
- CLR with auto thresholding

Main Results

Approach	Avg. Rank	Avg. Size
CLR	3.1	36
ML-*k*NN	4.1	113
OLA	4.2	1
MCB	4.8	1
KNORA	4.9	85
DS	5.1	1
LCA	5.9	1
DV	6.9	1
DVS	7.4	36
MV	8.6	200

Wilcoxon test for
CLR vs OLA returns
a p-value of 0.0022

Model composition in CLR

1. Decicion trees
2. Good version of SVMs
3. Good versions of MLPs
4. kNNs (lower k better)

OUTLINE

1. Deterministic label relationships
2. From multi-label classification to multi-target regression



3. Semantic indexing of biomedical literature
4. Multi-label classification for instance-based ensemble pruning



Google mulan library

Scholar About 3,770 results (**0.08** sec)

Articles
Case law
My library

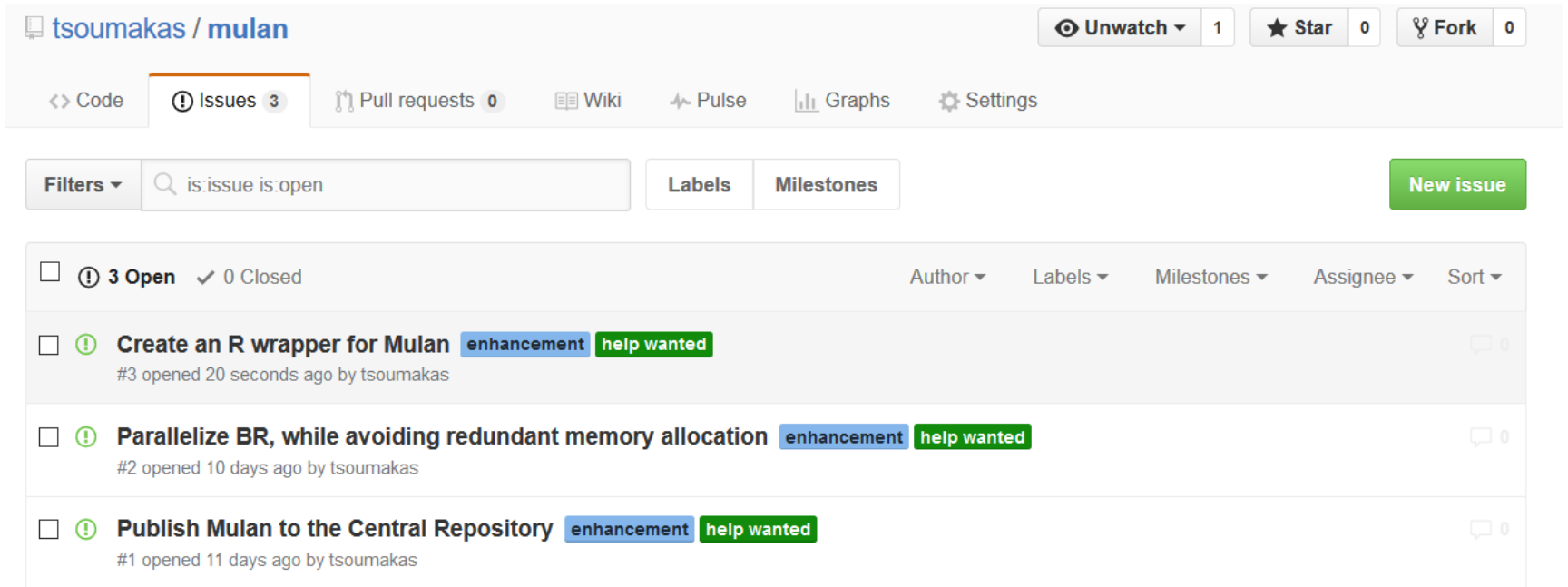
Any time

[HTML] **Mulan**: A java **library** for multi-label learning [HTML] from jmlr.org

G Tsoumakas, E Spyromitros-Xioufis, J Vilcek... - Journal of Machine ..., 2011 - jmlr.org

Abstract **MULAN** is a Java **library** for learning from multi-label data. It offers a variety of classification, ranking, thresholding and dimensionality reduction algorithms, as well as algorithms for learning from hierarchically structured labels. In addition, it contains an ...

Cited by 279 Related articles All 15 versions Cite Save



MULTI-TARGET PREDICTION: CHALLENGES AND APPLICATIONS

with: *N. Bassiliades, W. Groves, M. Laliotis, N. Markantonatos, F. Markatopoulou, C. & E. Papagiannopoulou, Y. Papanikolaou, E. Spyromitros-Xioufis, I. Tsamardinos, I. Vlahavas, A. Vrekou*



Grigorios Tsoumakas,
School of informatics,
Aristotle university of Thessaloniki