Multi-Target Prediction and Applications in the Publishing, Energy & Retail Industries

GRIGORIOS TSOUMAKAS,
INTELLIGENT SYSTEMS GROUP,
SCHOOL OF INFORMATICS,
ARISTOTLE UNIVERSITY OF THESSALONIKI

Multi-Target Prediction

APPLICATIONS

Multimedia annotation/retrieval

- Video, image, audio, text

Gene function prediction

Recommending bid phrases for web pages (10 million labels)

Ensemble pruning/meta-learning

Ecological modeling

Demand forecasting

…

TASKS

Multi-label classification

Multivariate regression

Label ranking

Multi-task learning

Collaborative filtering

Dyadic prediction

CHALLENGES

Dealing with class imbalance

Exploiting dependencies between the targets

Scaling to extreme sizes of output spaces

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

Outline

THEORY

Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning
- with C. Papagiannopoulou, I. Tsamardinos
- at KDD, 2015

Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs
- with E. Spyromitros, W. Groves, I. Vlahavas
- at Machine Learning Journal, 2016

PRACTICE

Semantic Indexing of Scientific Publications
- Atypon Systems Inc.

Short-Term Forecasting of Natural Gas Demand
- Gas Supply Company of Thessaloniki & Thessaly

Customer Segmentation
- Diamantis Masoutis S.A.

Outline

THEORY

Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning

- with C. Papagiannopoulou, I. Tsamardinos
- at KDD, 2015

Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs

- with E. Spyromitros, W. Groves, I. Vlahavas
- at Machine Learning Journal, 2016

PRACTICE

Semantic Indexing of Scientific Publications

- Atypon Systems Inc.

Short-Term Forecasting of Natural Gas Demand

- Gas Supply Company of Thessaloniki & Thessaly

Customer Segmentation

- Diamantis Masoutis S.A.

Entailment Relationships

Let A and B be two labels with domain $\{false, true\}$

- Shortcut notation: $A = false \equiv \neg a, A = true \equiv a$

Entailment relationships

- Positive entailment
  - $a \rightarrow b$ and equivalent contrapositive $\neg b \rightarrow \neg a$
  - $b \rightarrow a$ and equivalent contrapositive $\neg a \rightarrow \neg b$
- Exclusion
  - $a \rightarrow \neg b$ and equivalent contrapositive $b \rightarrow \neg a$
- Co-exhaustion
  - $\neg a \rightarrow b$ and equivalent contrapositive $\neg b \rightarrow a$

pairwise positive entailments

This work!

Mutual exclusion among ≥ 2 labels

Extracting Relationships

Positive entailment

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

Mutual exclusion

- $a \to \neg b \land b \to \neg a$ 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 two labels

	$\boldsymbol{a}$	$\boldsymbol{\neg a}$
$\boldsymbol{b}$	S	T
$\boldsymbol{\neg b}$	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

- $a \to b$ (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

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

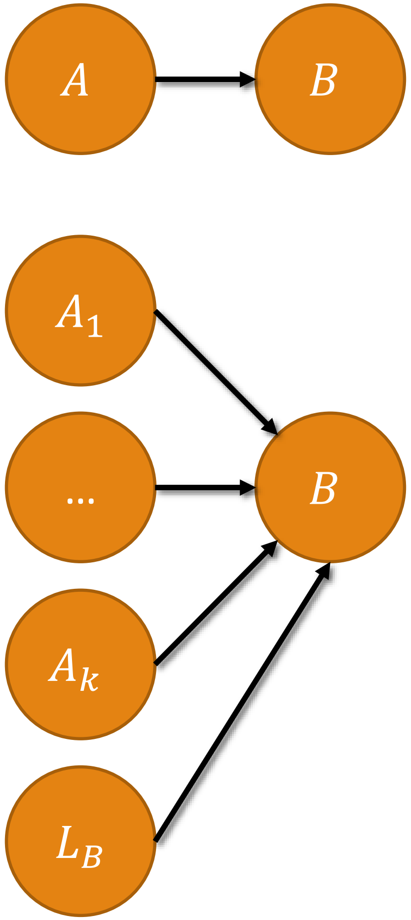
- $a \rightarrow b$

Generalization

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

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	$\neg b$
$A_1 \vee \cdots \vee A_k = true$	1	0
$A_1 \vee \cdots \vee A_k = false$	0	1

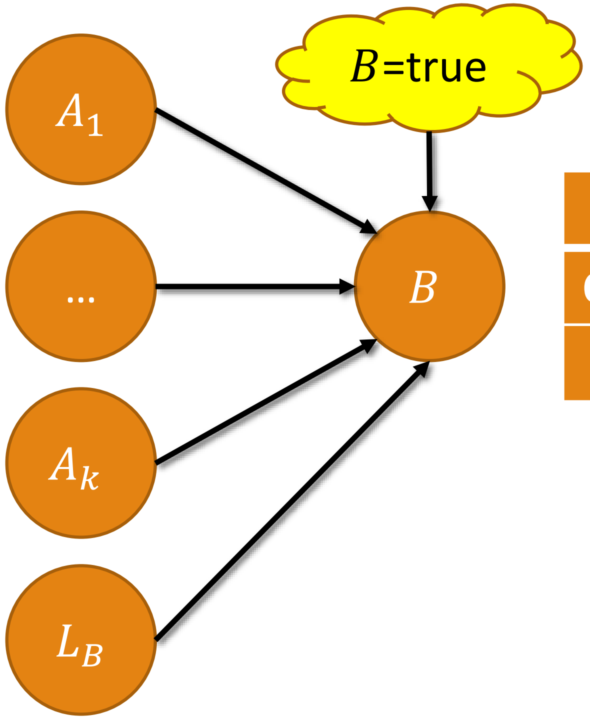
	b	$\neg b$
$A_1 \vee \cdots \vee A_k \vee L_B = true$	1	0
$A_1 \vee \cdots \vee A_k \vee L_B = false$	0	1

Exploiting Relationships: Mutual Exclusion

Among k labels $A_1, \dots, A_k$

Leak node

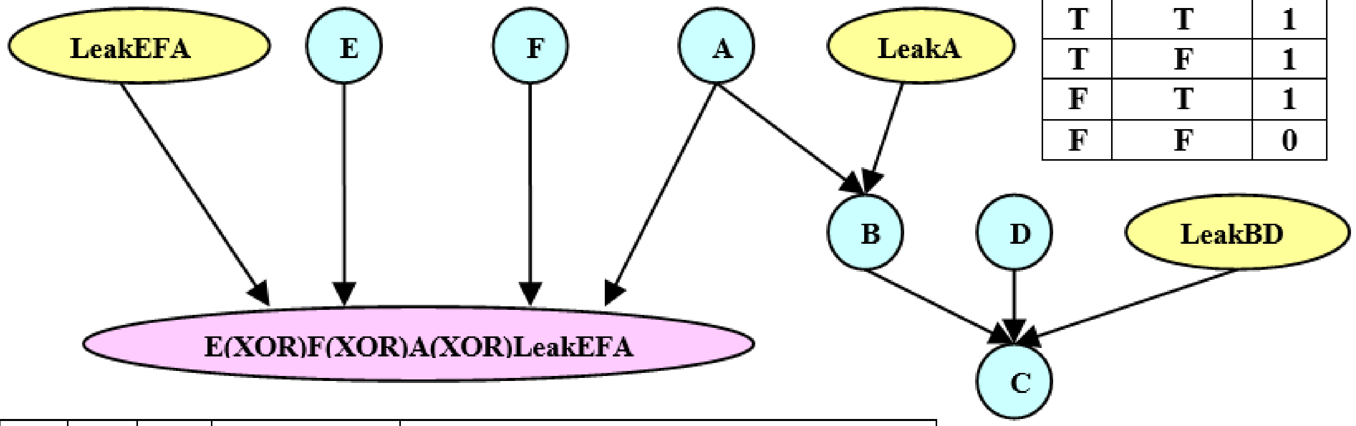
- To cover all training examples, i.e. to become exhaustive
- Virtual label equal to
  - True where all other parents of B are false
  - False in all other examples



	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

Experimental Setup

12 datasets

Relationship discovery

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

Binary Relevance + Random Forest with 10 trees per (virtual) label

- Weka, Mulan

Clustering for inference in Bayesian networks

- jSMILE

10-fold cross-validation

- Relationship discovery and exploitation based on training set
- Actual relation discussion based on presence on all folds

Positive Entailment + Mutual Exclusion

		Minimum Support		Improvement %		
Dataset	MAP of Standard BR	Positive Entailment	Mutual Exclusion	Positive Entailment	Mutual Exclusion	Positive + Exclusion
Bibtex	0.2152	2	256	0.279	0.604	1.347
Bookmarks	0.1474	2	2048	0.068	-0.068	0
Enron	0.2810	2	32	0.391	0.214	0.214
ImageCLEF2011	0.2788	2	32	2.977	1.865	2.080
ImageCLEF2012	0.2376	2	256	0.168	0.631	0.716
Medical	0.5997	2	16	2.284	3.769	4.436
Yeast	0.4545	2	2	1.584	1.760	2.904

Medical

3 entailment relationships extracted from 978 examples

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

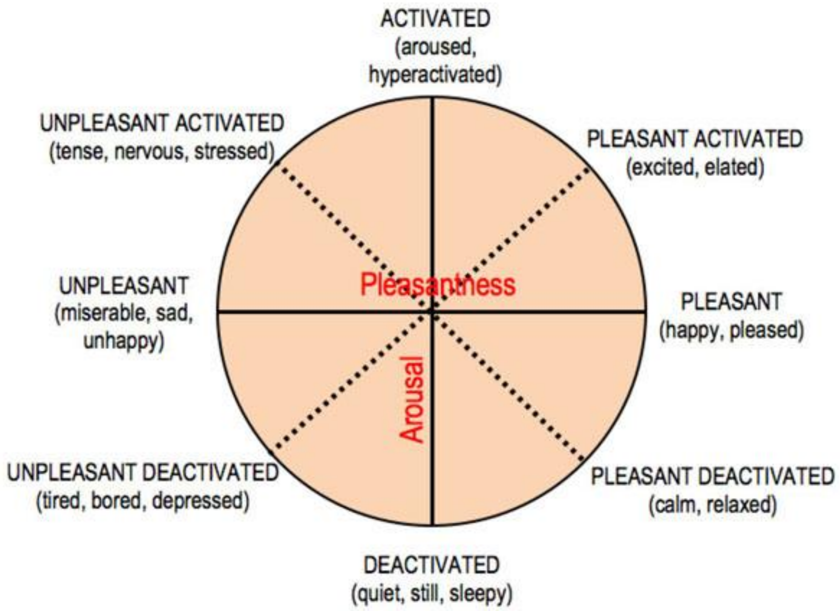
Ureteropelvic junction obstruction is the **most common pathologic cause of** antenatally detected *hydronephrosis*

Google

Emotions and Enron

EMOTIONS

{quiet-still, amazed-surprised}



ENRON

{"*Company Business, Strategy, etc.*",
"friendship / affection"}

Outline

THEORY

Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning

- with C. Papagiannopoulou, I. Tsamardinos
- at KDD, 2015

Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs

- with E. Spyromitros, W. Groves, I. Vlahavas
- at Machine Learning Journal, 2016

PRACTICE

Semantic Indexing of Scientific Publications

- Atypon Systems Inc.

Short-Term Forecasting of Natural Gas Demand

- Gas Supply Company of Thessaloniki & Thessaly

Customer Segmentation

- Diamantis Masoutis S.A.

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

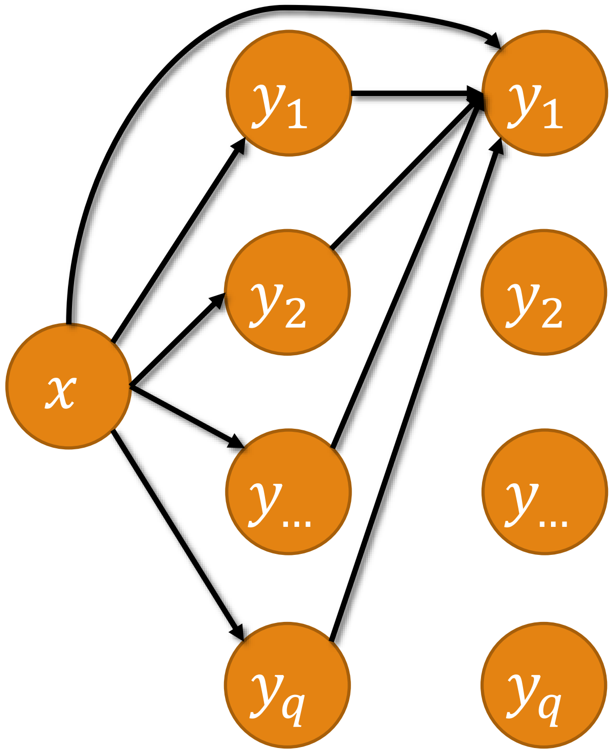
- Ensemble of classifier chains
  - *Read J, Pfahringer B, Holmes G, Frank E (2011) Classifier chains for multi-label classification. MLJ 85(3):333–359*
- Stacking Binary Relevance
  - Godbole S, Sarawagi S (2004) Discriminative methods for multi-labeled classification. In Proceedings of the 8th Pacific-Asia Conference, PAKDD 2004, Sydney, Australia, May 26-28, 2004, Proceedings, pp 22–30

Abstracting key ideas from solutions tailored to related problems

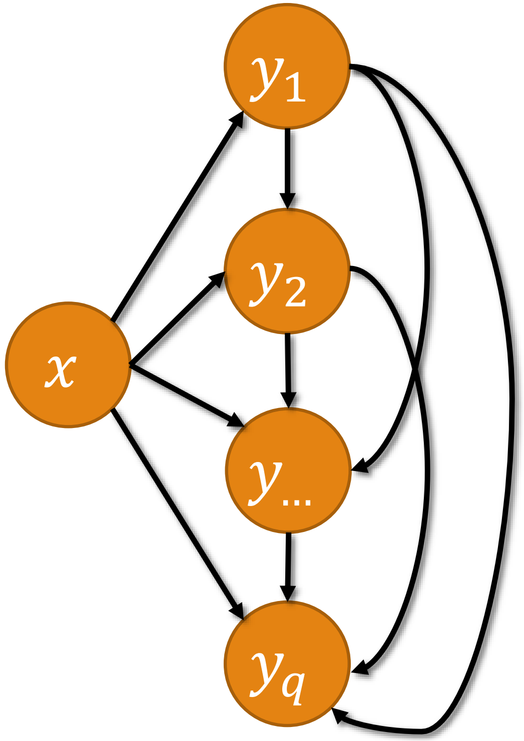
- Improves the modularity and conceptual simplicity of learning techniques
- Avoids reinvention of the same solutions
- NIPS11 WS “*relations between machine learning problems – an approach to unify the field*”

Two “New” Algorithms

STACKED SINGLE TARGET (SST)



ENSEMBLE OF REGRESSOR CHAINS (ERC)



Generating Meta-Input Data

Discrepancy of train/test distributions

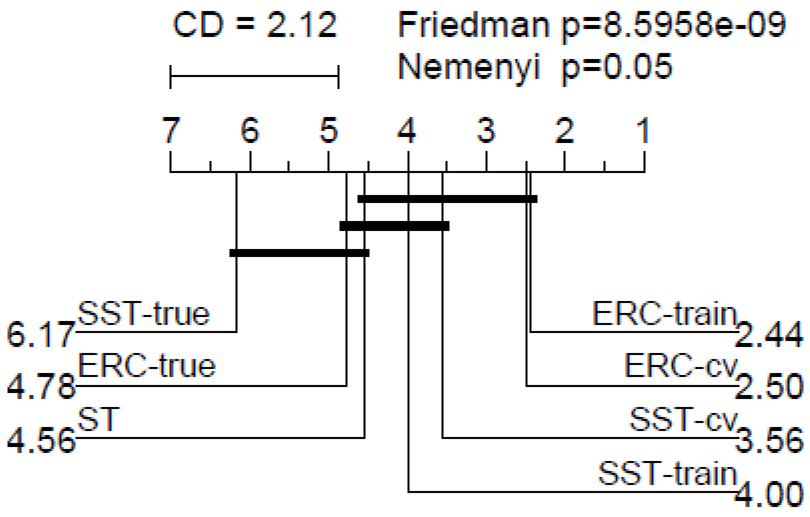
- Test distributions are based on out-of-sample estimates
- SST uses in-sample estimates
- ERC uses actual true values instead of estimates

Proposal

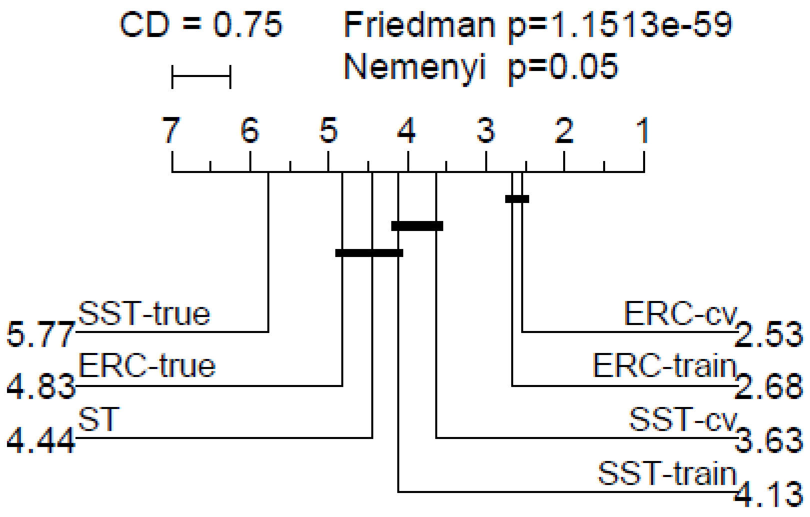
- Use out-of-sample estimates based on cross-validation

Results

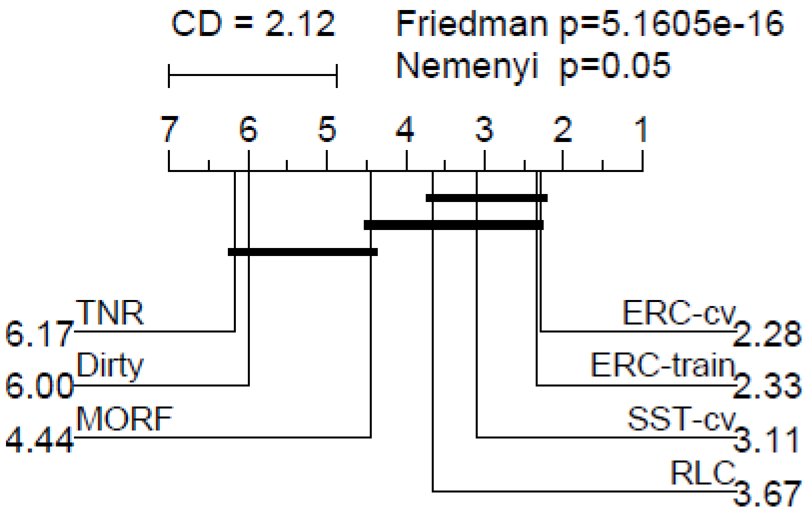
- 18 datasets (several of them first used in this paper)
- Strong learner: bagging of regression trees
- RRMSE, statistical testing per target and per dataset
- Significant improvement compared to standard versions and the state of the art



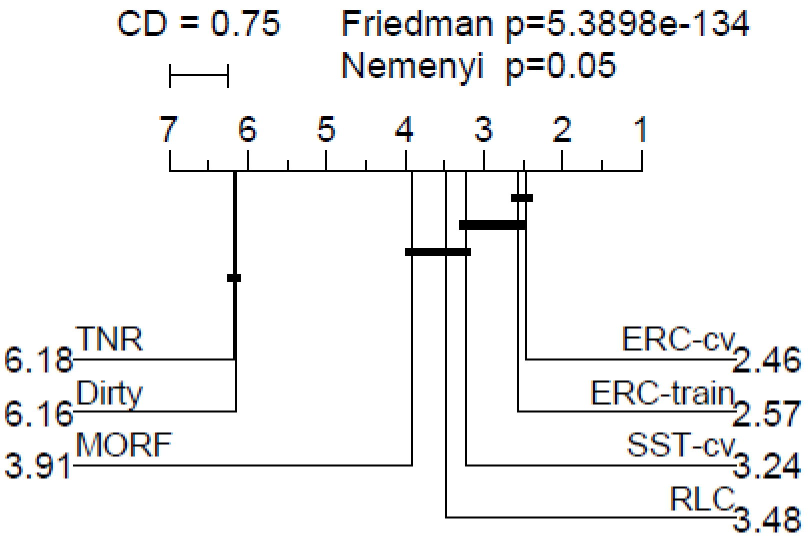
(a) Per dataset analysis.



(b) Per target analysis.



(a) Per dataset analysis.



(b) Per target analysis.

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

tsoumakas / **mulan**

Unwatch 1 Star 0 Fork 0

Code Issues 3 Pull requests 0 Wiki Pulse Graphs Settings

Filters is:issue is:open Labels Milestones New issue

3 Open ✓ 0 Closed Author Labels Milestones Assignee Sort

Create an R wrapper for Mulan enhancement help wanted
#3 opened 20 seconds ago by tsoumakas

Parallelize BR, while avoiding redundant memory allocation enhancement help wanted
#2 opened 10 days ago by tsoumakas

Publish Mulan to the Central Repository enhancement help wanted
#1 opened 11 days ago by tsoumakas

Outline

THEORY

Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning

- with C. Papagiannopoulou, I. Tsamardinos
- at KDD, 2015

Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs

- with E. Spyromitros, W. Groves, I. Vlahavas
- at Machine Learning Journal, 2016

PRACTICE

Semantic Indexing of Scientific Publications

- Atypon Systems Inc.

Short-Term Forecasting of Natural Gas Demand

- Gas Supply Company of Thessaloniki & Thessaly

Customer Segmentation

- Diamantis Masoutis S.A.

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 tagging, eCommerce, and analytics

Automated semantic tagging

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

The BioASQ Project

ICT-2011.4.4 (d)

- Intelligent Information Management
  - Targeted Competition Framework

Two challenges on the domain of biomedicine

A. Large-scale online biomedical semantic indexing

B. Biomedical question answering

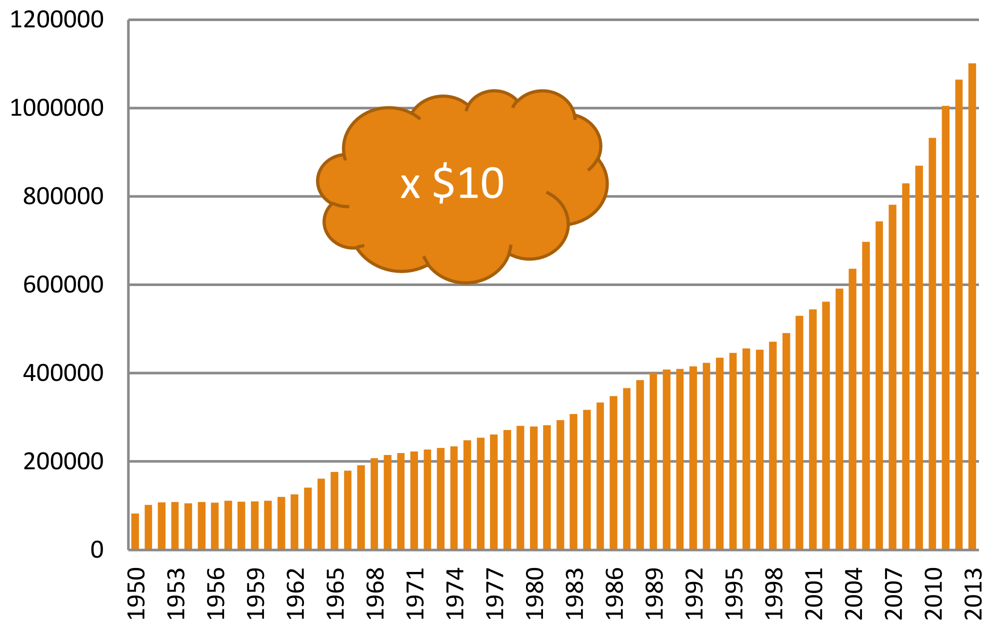
Duration: October 2012 – September 2014

Large-Scale Online Biomedical Indexing

Corpus (training abstracts)
- 2013: 10,876,004 (18Gb)
- 2014: 12,628,968 (20Gb)

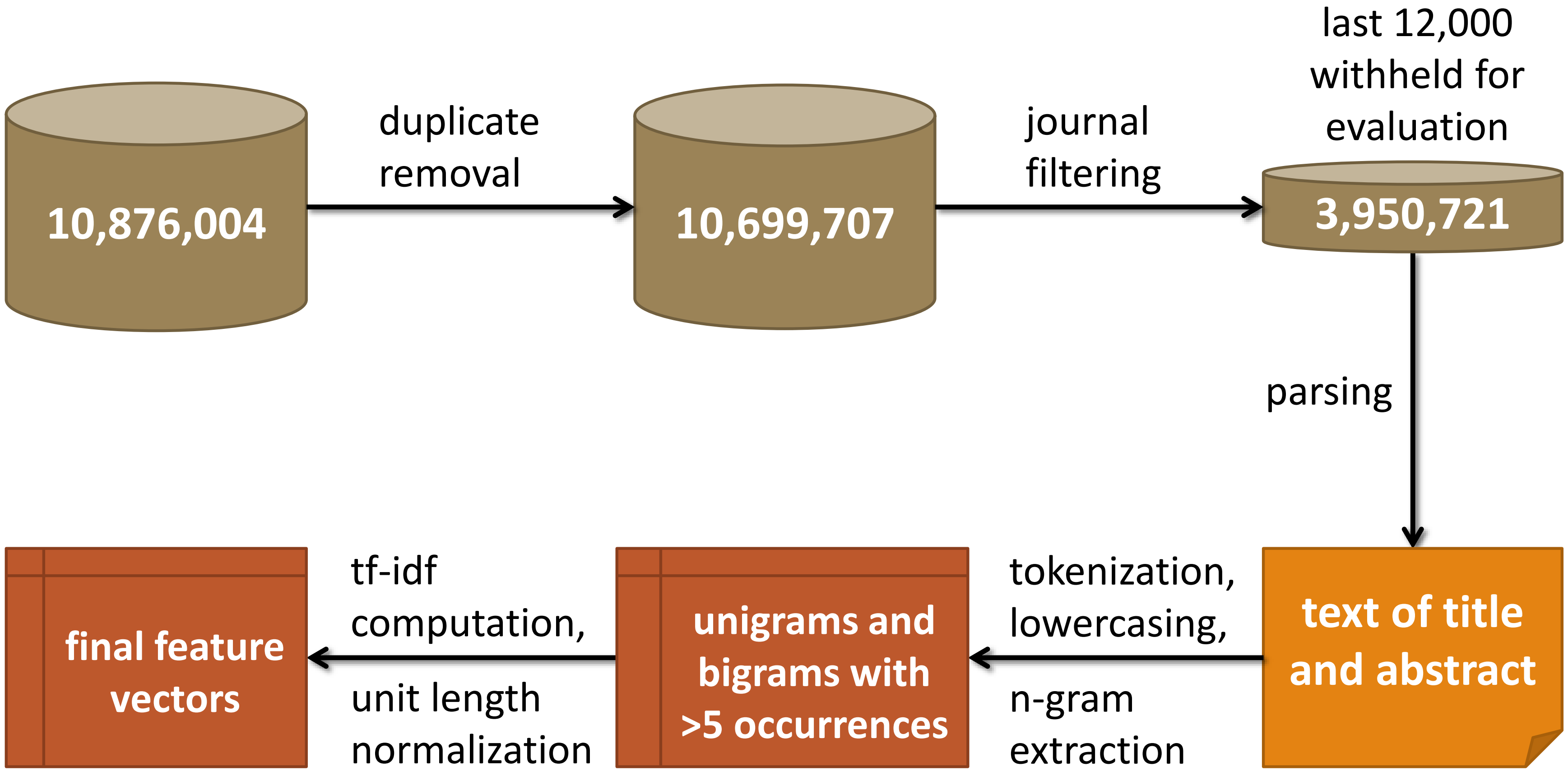
MeSH terms
- 2013: 26,563 / 2014: 26,583

Online test setting
- 22/4/13 - 19/8/13, 3 phases, 6w per phase
- 3/2/14 – 12/5/14, 3 phases, 5w per phase
- 2013: 87,080 (average of 4,838 docs per week)
- BioASQ requested MeSH terms for 790 to 10,139 abstracts within 21 hours



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

Our Approach: Preprocessing



Our Approach: Learning

- Any multi-label learning algorithm that can output a ranking of the labels
- We used a linear SVM per label and considered their unthresholded output

Biomedical Document

Label Ranker

kinetics	prognosis	liposomes	polymers
4	1	2	3

Meta Labeler

Number of relevant labels: 2

Output is: {*prognosis*, *liposomes*}

- Regression or (ordinal) classification using original features or label scores/ranks
- We used linear SVM regression based on the original features

Tang, L., Rajan, S., Narayanan, V.K. "Large scale multi-label classification via metalabeler. In Proc. 18th Int. Conf. on World Wide Web (WWW '09)

Results & Implementation Aspects

RESULTS

2013

- 1st position

2014 – 2016

- 2nd position

2013 – 2016

- Consistently better than the production system of the National Library of Medicine

IMPLEMENTATION ASPECTS

Hardware

- 4 10-core processors at 2.26 GHz,
- 1 Tb RAM and 2.4 Tb storage (6 x 600 Gb SAS 10k disks in RAID 5)

Parallel learning/use of binary SVMs

- With 40 threads training & saving takes 36h
- With 20 threads loading & prediction takes 45m

Serialization

- Storing the binary SVM models required 406 Gb
- 10x compression due to sparsity

Outline

THEORY

Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning

- with C. Papagiannopoulou, I. Tsamardinos
- at KDD, 2015

Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs

- with E. Spyromitros, W. Groves, I. Vlahavas
- at Machine Learning Journal, 2016

PRACTICE

Semantic Indexing of Scientific Publications

- Atypon Systems Inc.

Short-Term Forecasting of Natural Gas Demand

- Gas Supply Company of Thessaloniki & Thessaly

Customer Segmentation

- Diamantis Masoutis S.A.

Forecasting of Natural Gas Demand

Collaboration with Gas Supply Company of Thessaloniki & Thessaly

The problem

- Daily statements of one day ahead demand must be submitted to the regulatory entity
- Actual consumption must lie within a percentage of the statement (e.g. 10%), otherwise economic fines are imposed

Similar framework in the electricity domain

- Demand per hour renders the problem multi-target

The Solution

VARIABLES

Internal
- Contracts
- Historical demand

Temporal
- Holiday, day, month

Weather (API)
- Temperature
- Wind speed and direction
- Humidity
- …

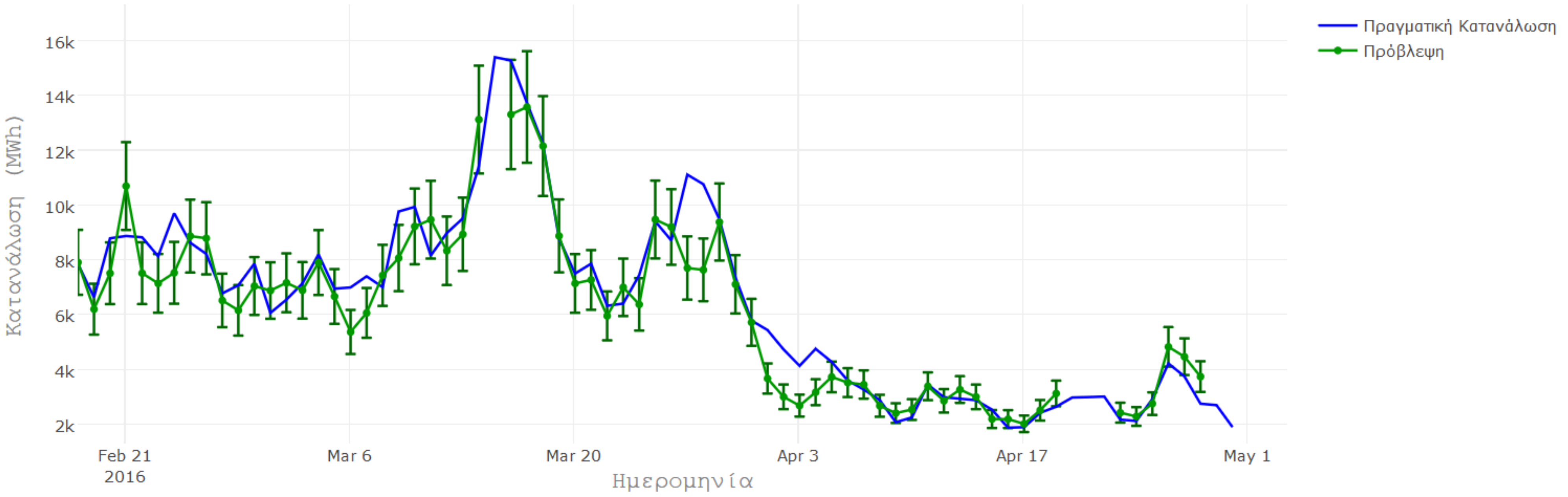
MODEL & TOOLS

Support Vector Regression
- Kernel = Radial Basis Function

Gas Prediction | Γραφική παράσταση | Ιστορικά δεδομένα | Προβλεψεις | Ιστορικό προβλέψεων

Διαθέσιμες Πόλεις

- ◉ Θεσσαλονίκη
- ○ Λάρισα
- ○ Βόλος
- ○ Τρίκαλα
- ○ Καρδίτσα



Αξιολόγηση απόδοσης

Μέσος όρος της διαφοράς Πραγματικής κατανάλωσης από την Πρόβλεψη(απόλυτη τιμή):

732.3582

Μέσος όρος μεγαβατώρων εκτός του 10%:

308.4134

Outline

THEORY

Discovering and Exploiting Deterministic Label Relationships in Multi-Label Learning
- with C. Papagiannopoulou, I. Tsamardinos
- at KDD, 2015

Multi-Target Regression via Input Space Expansion: Treating Targets as Inputs
- with E. Spyromitros, W. Groves, I. Vlahavas
- at Machine Learning Journal, 2016

PRACTICE

Semantic Indexing of Scientific Publications
- Atypon Systems Inc.

Short-Term Forecasting of Natural Gas Demand
- Gas Supply Company of Thessaloniki & Thessaly

Customer Segmentation
- Diamantis Masoutis S.A.

http://segmento.csd.auth.gr

Collaboration with Diamantis Masoutis S.A.

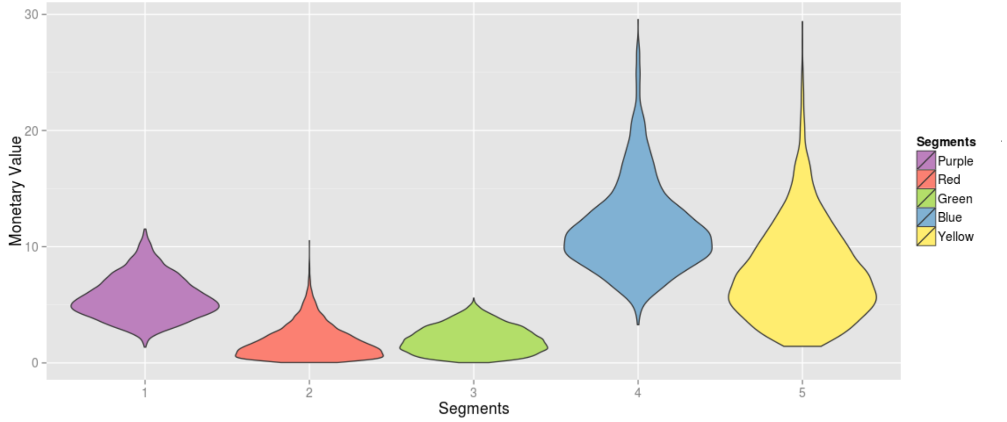
- 260 supermarket stores
- 6000 personnel
- € 750m sales

Functionality

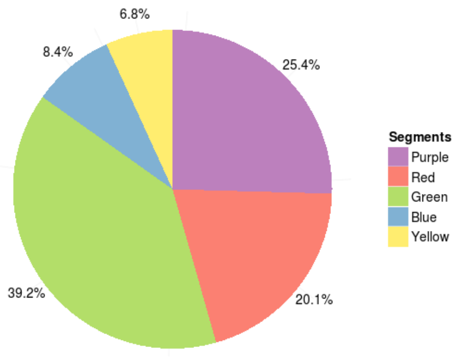
- Customer segmentation
- Comparative analysis of clustering results at two different time steps
- Directed marketing campaigns

Descriptives

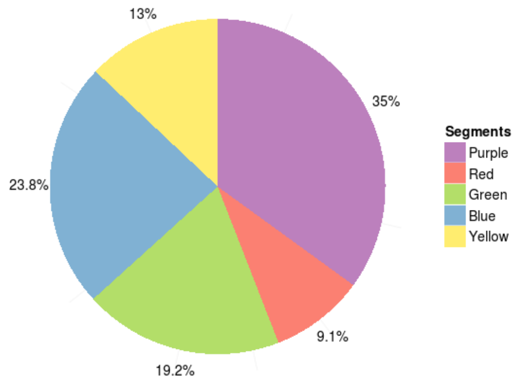
Recency | Frequency | Monetary Value | Transaction Value (Basket) | Private Label Value
Daily Reward Points | Points Redemption

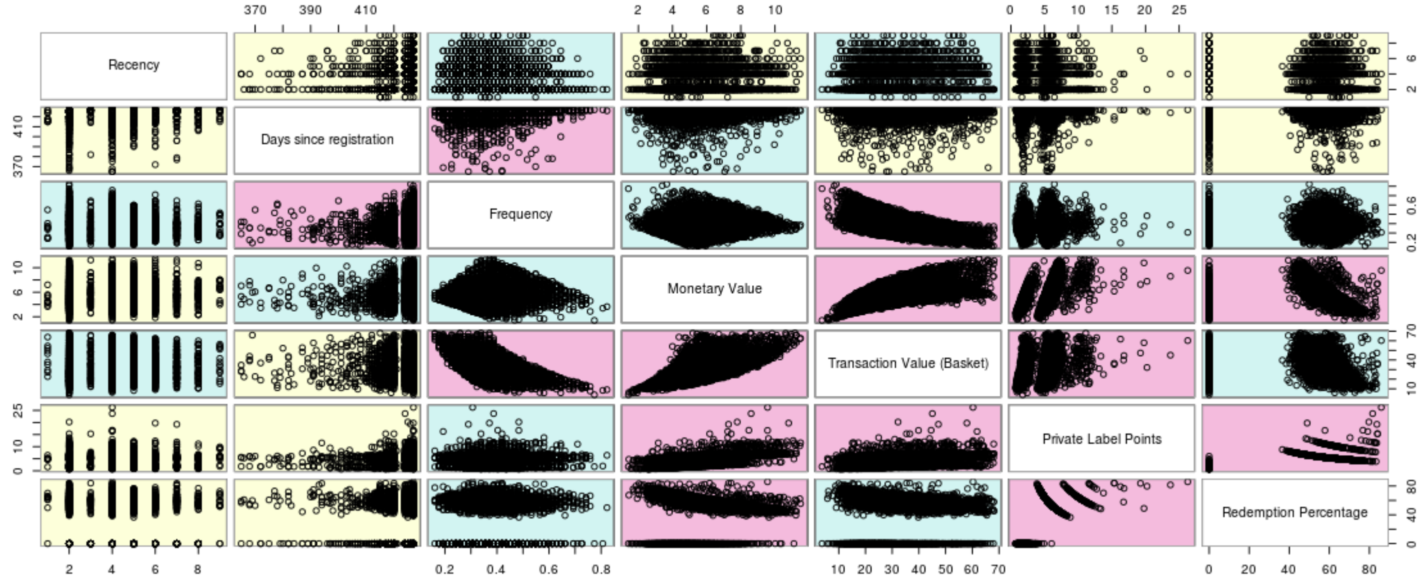


Segments Size in a total of 26.462 customers

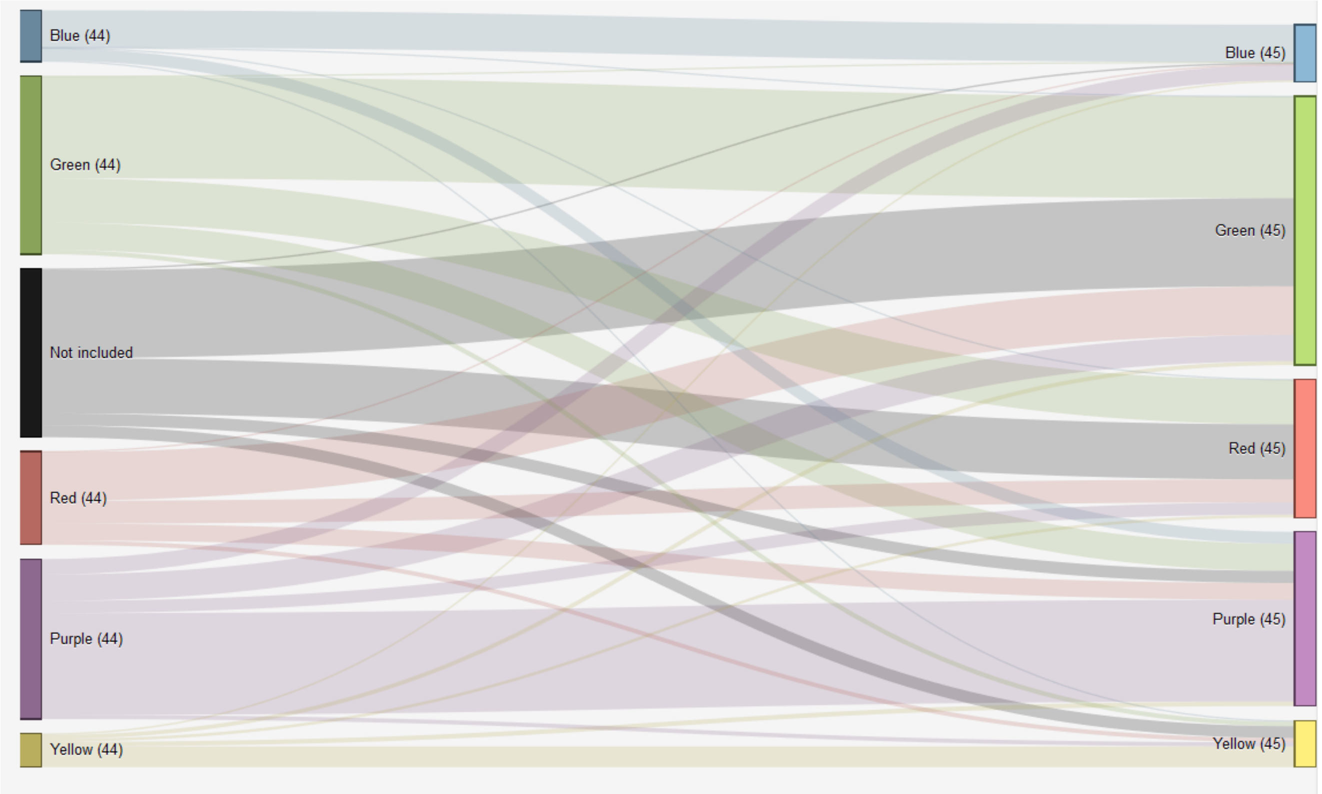


Total transactions value: 108.558,80 €





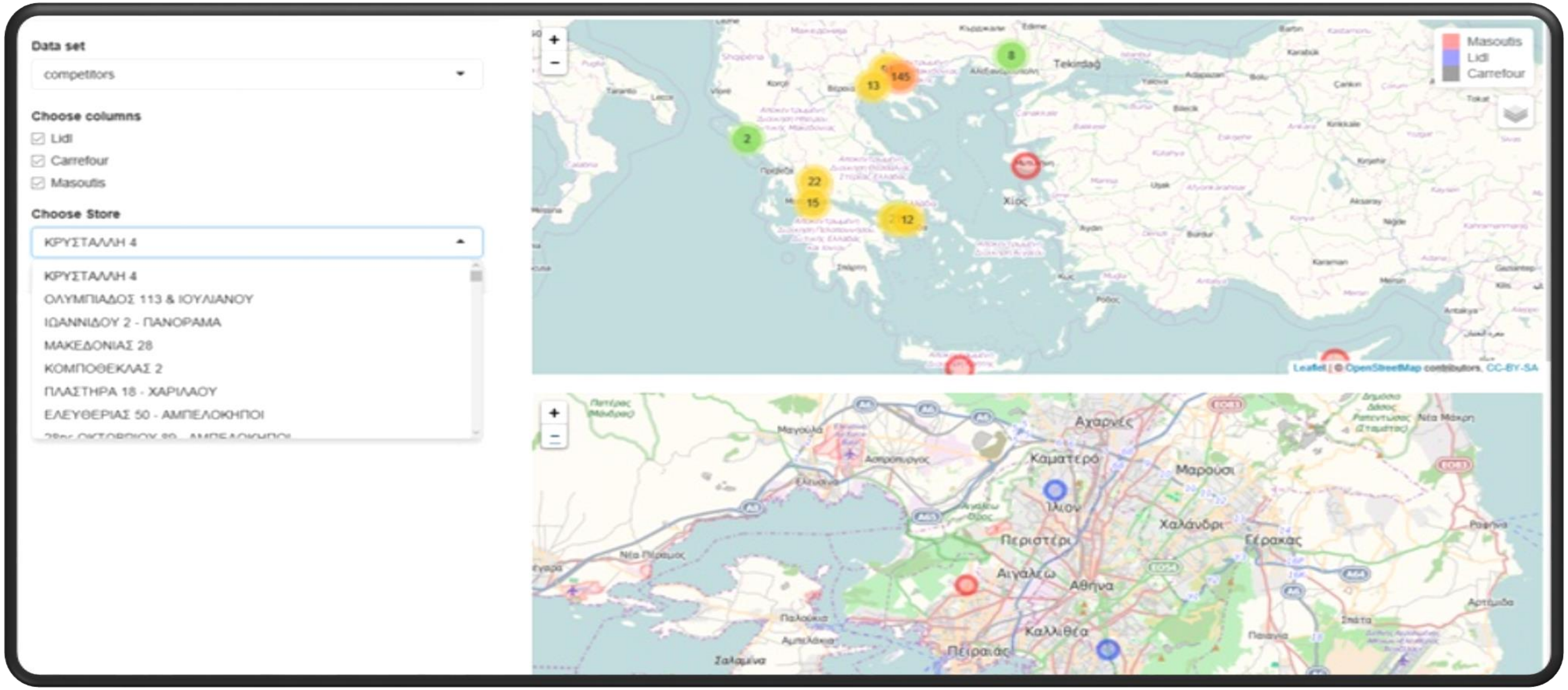
Transitions Among Segments



Profit/Loss Estimation

	Segmentation44	Segmentation45	Volume	Euros/Day /Customer	Daily Transition Cost
1	Purple	Green	1014	-3.51	-3559.14
2	Blue	Purple	463	-5.82	-2694.66
3	Purple	Red	475	-3.67	-1743.25
4	Yellow	Green	153	-5.8	-887.4
5	Yellow	Red	109	-5.96	-649.64
6	Yellow	Purple	176	-2.16	-380.16
7	Red	Red	885	-0.28	-247.8
8	Red	Green	1879	-0.12	-225.48
9	Blue	Yellow	51	-3.67	-187.17
10	Blue	Red	17	-9.62	-163.54
11	Green	Red	1716	-0.09	-154.44
12	Blue	Green	11	-9.46	-104.06
13	Yellow	Yellow	800	-0.01	-8
14	Green	Blue	20	9.73	194.6
15	Red	Blue	23	9.55	219.65
16	Yellow	Blue	63	3.86	243.18
17	Green	Green	3927	0.07	274.89
18	Blue	Blue	1443	0.2	288.6
19	Purple	Yellow	160	2.28	364.8
20	Purple	Purple	3933	0.13	511.29

Location Intelligence



Applied Machine Learning that Matters

CARLA BRODLEY ET AL. (2012) *CHALLENGES & OPPORTUNITIES IN APPLIED MACHINE LEARNING*, AI MAGAZINE 33(1)

"Machine-learning research is often conducted in vitro, divorced from motivating practical applications."

"… being involved in an application can highlight shortcomings of existing methodologies and lead to new insights into previously unaddressed research issues."

"Contact with real problems, real data, real experts, and real users can generate the creative friction that leads to new directions in machine learning."

KIRI WAGSTAFF (2012) *MACHINE LEARNING THAT MATTERS*, INTERNATIONAL CONFERENCE ON MACHINE LEARNING

"Much of current machine learning (ML) research has lost its connection to problems of import to the larger world of science and society."

"Aiming for real impact does not just increase our job satisfaction (though it may well do that); it is the only way to get the rest of the world to notice, recognize, value, and adopt ML solutions."