Multilabel Learning for personalized Bookmark Classification

TECHNISCHE UNIVERSITÄT DARMSTADT

Outline

TECHNISCHE UNIVERSITÄT DARMSTADT

Part1: Web Resource Classification
- Tabbed Browsing
- Multilabel Classification

Part2: Multilabel Ranking
- Multilabel Classifier
- Dataset Decomposition
- Multilabel Classifier Index
- Approaches
- Reuters RCV1-V2

Part3: Evaluation
- Approach 1
- Approach 2
- Approach 3

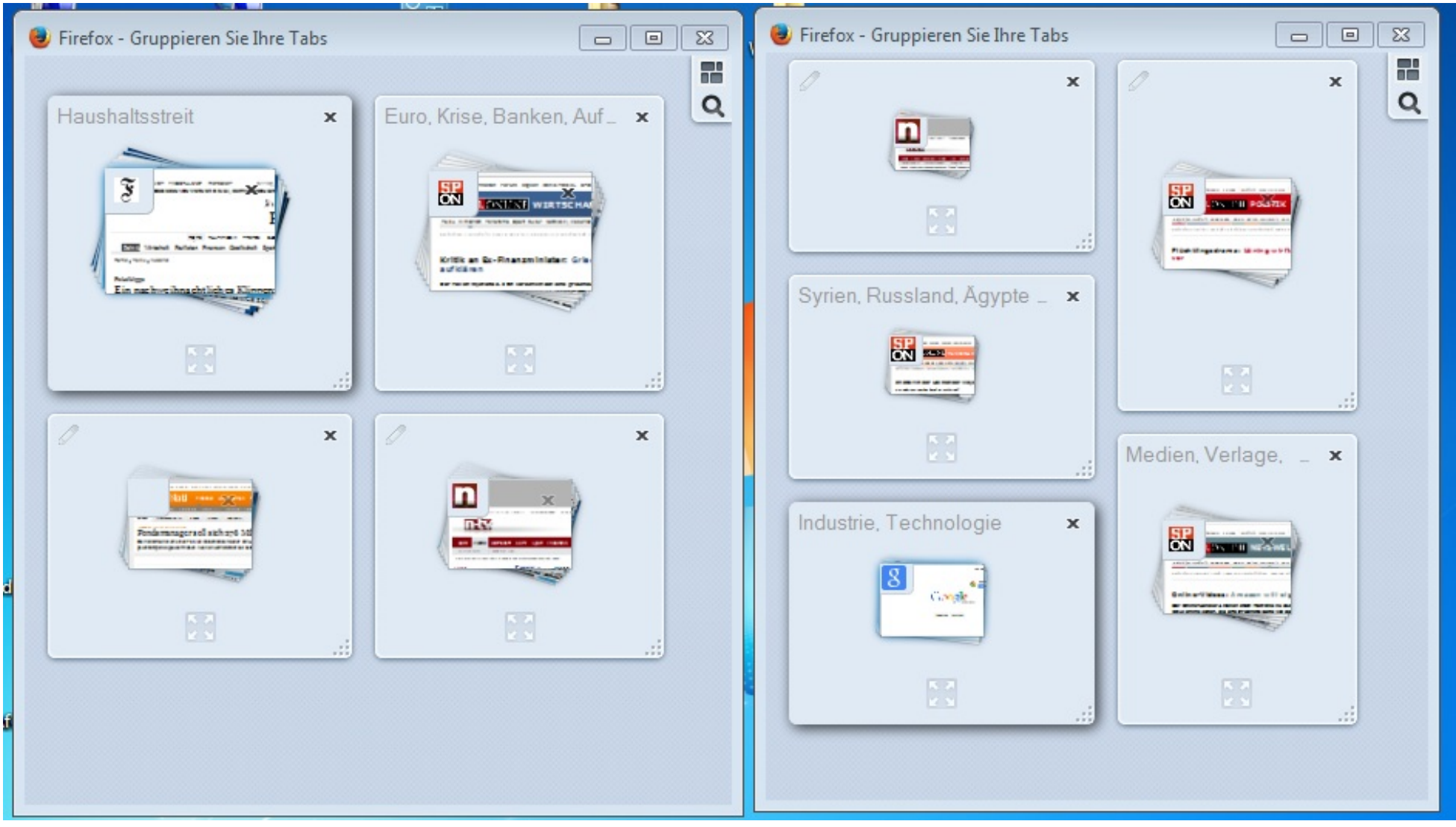
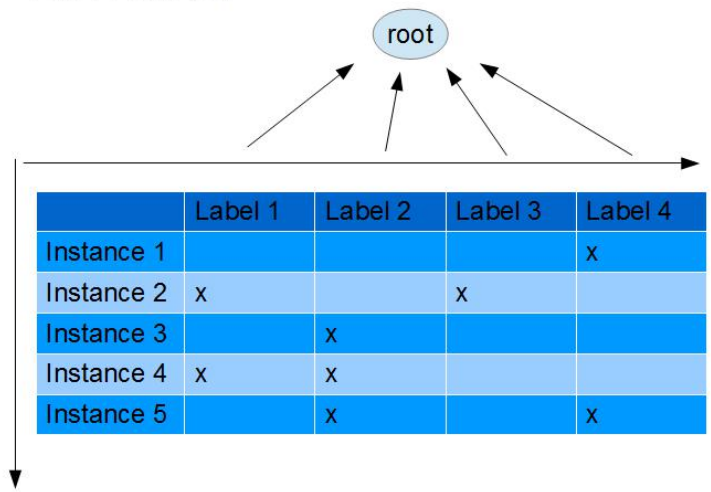


Figure: Tabbed Browsing: Groups of Tabs for Browsing Sessions

Multilabel Classification

TECHNISCHE UNIVERSITÄT DARMSTADT

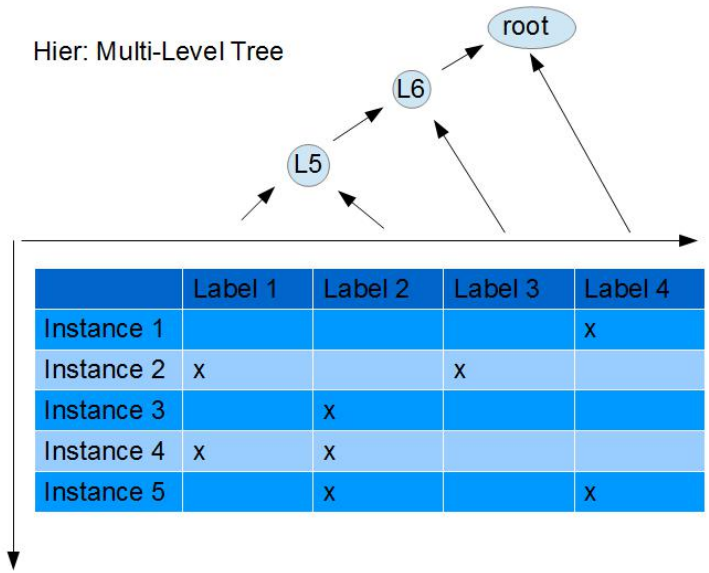
Flat: 1-Level Tree



	Label 1	Label 2	Label 3	Label 4
Instance 1				x
Instance 2	x		x	
Instance 3		x		
Instance 4	x	x		
Instance 5		x		x

Multilabel Classification

TECHNISCHE UNIVERSITÄT DARMSTADT



	Label 1	Label 2	Label 3	Label 4
Instance 1				x
Instance 2	x		x	
Instance 3		x		
Instance 4	x	x		
Instance 5		x		x

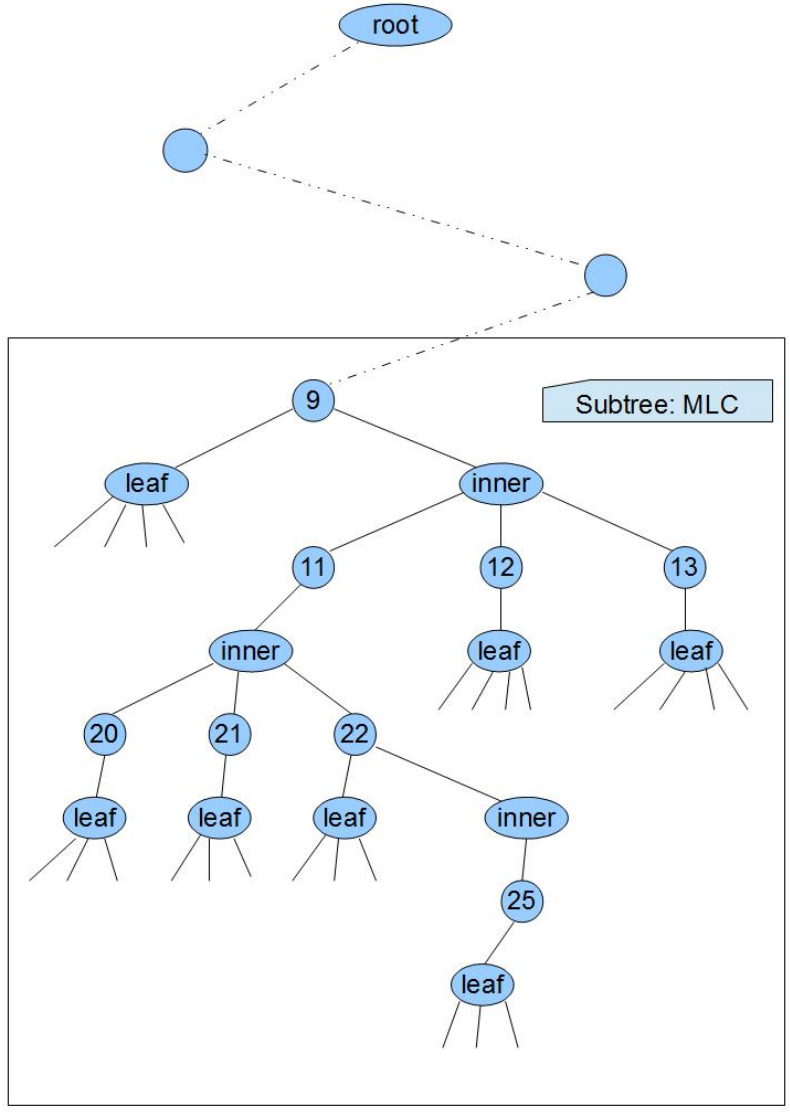


Figure: Subtree for Multilabel Classifier

Dataset Decomposition

TECHNISCHE UNIVERSITÄT DARMSTADT

1. For our multilabel dataset we will apply the widely known technique of binary relevance decomposition in order to build binary classification tasks
2. We will use SVM classifiers for each binary classification task
3. For a multi-level tree we will build for each subtree a multi-label classifier, representing a flat tree.

Multilabel Classifier Index

TECHNISCHE UNIVERSITÄT DARMSTADT

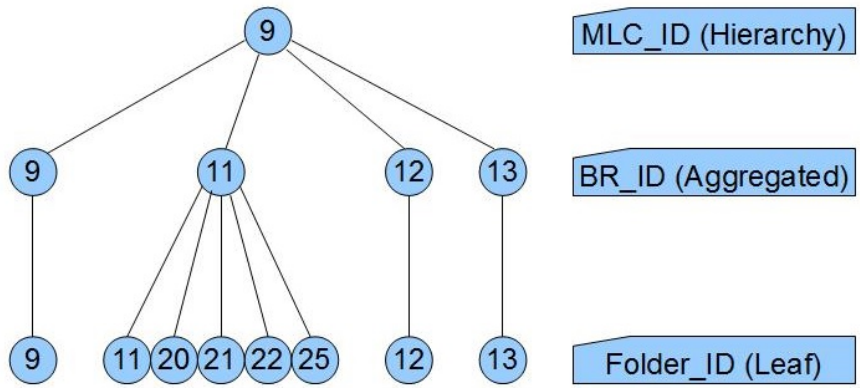


Figure: Multilabel Classifier Index

Approach 1

TECHNISCHE UNIVERSITÄT DARMSTADT

1. 1-Level, Flat Tree (Binary Relevance)
2. Only one subtree considered
3. Training-Phase: SVM-Training-Models are generated
4. Prediction-Phase: For a new query instance, just make a prediction for each model
5. Positive Class Prediction Probability (relative Probability)
6. If PCPP > 0.5, the label is assigned to the bipartition-positive
7. Else, the label is assigned to the bipartition-negative
8. Label Ranking: sort bipartition(positive, negative) by probability value

Approach 2

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Multi-Level, hierarchical Tree (HOMER)
2. Training-Phase: Multiple subtrees considered in hierarchical decomposition
3. We build the index structure for each subtree and generate training models
4. Here we are applying an aggregated binary relevance technique on each subtree
5. Prediction-Phase: Hierarchical exploration of the trained SVM models
6. Label Ranking: Absolute Probability / Path Probability

Reuters RCV1-V2

TECHNISCHE UNIVERSITÄT DARMSTADT

Dataset Statistics

The dataset (200.000 test-instances) is specified by the following numbers:

Level	# of Labels	Label-Cardinality	Label-Density	# Instance-Assignm.
0	101	3.2429	0.0321	1478091
1	0	0.0	0.0	0
2	4	1.1781	0.2945	945334
3	54	1.4623	0.0271	1128784
4	42	1.1354	0.0270	509546
5	1	1.0	1.0	23211

Approach 1: Bipartition-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Micro Average			Hamming-Loss
Test-Instances	Level	Precision	Recall	F1	Average
200000	0	0.9529	0.5208	0.6735	0.0157
200000	1	1.0000	1.0000	1.0000	0.0000
200000	2	0.9487	0.8224	0.8810	0.0654
	3	0.9560	0.3230	0.4829	0.0178
	4	0.9643	0.4096	0.5750	0.0090
200000	5	0.7923	0.1227	0.2125	0.0092

Approach 1: Ranking-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Ranking-Loss	Ranking-Precision
Test-Instances	Level	Average	Average
200000	0	0.0245	0.8556
200000	1	0.0000	NaN
200000	2	0.0301	0.9679
	3	0.0316	0.8403
	4	0.0193	0.8540
200000	5	0.0000	1.0000

1. Ordering Bipartition-Positive: Apriori-Value
2. Ordering Bipartition-Negative: Apriori-Value

Approach 2: Bipartition-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Micro Average			Hamming-Loss
Test-Instances	Level	Precision	Recall	F1	Average
200000	0	0.4511	0.6201	0.5223	0.0354
200000	1	1.0000	1.0000	1.0000	0.0000
200000	2	0.9487	0.8224	0.8810	0.0654
	3	0.2293	0.4813	0.3107	0.0551
	4	0.7794	0.5625	0.6534	0.0088
200000	5	0.9462	0.1479	0.2559	0.0087

Approach 2.1: Ranking-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Ranking-Loss	Ranking-Precision
Test-Instances	Level	Average	Average
200000	0	0.0940	0.4622
200000	1	0.0000	NaN
200000	2	0.0848	0.9347
	3	0.1411	0.3557
	4	0.0412	0.6550
200000	5	0.0000	1.0000

1. Ordering Bipartition-Positive: Apriori-Value
2. Ordering Bipartition-Negative: Apriori-Value

Approach 2.2: Ranking-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Ranking-Loss	Ranking-Precision
Test-Instances	Level	Average	Average
200000	0	0.0403	0.8023
200000	1	0.0000	NaN
200000	2	0.0306	0.9675
	3	0.0727	0.7218
	4	0.0230	0.7897
200000	5	0.0000	1.0000

1. Ordering Bipartition-Positive: Path-Value
2. Ordering Bipartition-Negative: Apriori-Value

Approach 3: Bipartition-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Micro Average	Hamming-Loss		
Test-Instances	Level	Precision	Recall	F1	Average
200000	0	0.4558	0.7871	0.5773	0.0359
200000	1	1.0000	1.0000	1.0000	0.0000
200000	2	0.9227	0.8982	0.9103	0.0522
	3	0.2660	0.7143	0.3876	0.0582
	4	0.7444	0.7458	0.7451	0.0076
200000	5	0.8653	0.6487	0.7415	0.0046

Approach 3: Ranking-Measures

TECHNISCHE UNIVERSITÄT DARMSTADT

		Ranking-Loss	Ranking-Precision
Test-Instances	Level	Average	Average
200000	0	0.0236	0.8802
200000	1	0.0000	NaN
200000	2	0.0236	0.9751
	3	0.0403	0.8414
	4	0.0129	0.8726
200000	5	0.0000	1.0000

1. Ordering Bipartition-Positive: Path-Value
2. Ordering Bipartition-Negative: Path-Value

Summary

TECHNISCHE UNIVERSITÄT DARMSTADT

1. Improved rankings (recall) with higher cost value
2. Different threshold-parameter leads to more observations
3. Users can build groups of labels

TECHNISCHE
UNIVERSITÄT
DARMSTADT

Thanks for your attention !