Seminar aus Data Mining und Maschinellem Lernen
Extreme Classification
Seminar aus Data Mining und Maschinellem Lernen

- Time and place - Wednesdays, 17:10 - 18:50, Room E202
- First presentation: probably in three weeks
- Each Wednesday two talks
The students are expected to give a 30 minute talk on the material they are assigned, followed by 15 minutes of questions.

The talk and the slides are allowed to be both English or German, but we strongly encourage the students to give the talk in English.

It is expected of the students to participate in the discussions.

Important! The content of the talk should exceed the scope of the paper, and demonstrate that a thorough understanding of the material was achieved.

Follow the guidelines on the Seminar site and at https://www.ke.tu-darmstadt.de/lehre/arbeiten/giving-a-talk-at-a-ke-seminar-1
Extreme Classification

- Hot topic in the last one or two years
- Roughly: all types of classification problems where the target space, i.e. categories/classes/labels is large
- in practice often **multilabel classification problems**
  - the assignment of several classes instead of only one class
  - basic tutorial at https://www.ke.tu-darmstadt.de/staff/eneldo
Image annotation

\{Fall foliage, Field\} \quad \{Beach, Urban\}

*scene* dataset consists of 2407 images assigned to 6 labels

Title and reference

Classifications

**EUROVOC descriptor**
- data-processing law
- computer piracy
- copyright
- software
- approximation of laws

**Directory Code:**
- Law relating to undertakings/IPR Law

**Subject matter:**
- Internal market
- Industrial and commercial property

**Text**
COUNCIL DIRECTIVE of 14 May 1991 on the legal protection of computer programs (91/250/EEC)
THE COUNCIL OF THE EU,
Having Regard to the Treaty establishing the European Economic Community and in particular Article 100a thereof,
Having regard to the proposal of the Commission (1), ...
EUR-Lex repository

- 19328 (freely accessible) documents of the *Directory of Community legislation in force* of the European Union
  - documents available in several European languages
- multiple classifications of the same documents
- most challenging one: **EUROVOC** descriptors associated to each document
  - 3965 descriptors, on average 5.37 labels per document
  - descriptors are organized in a hierarchy with up to 7 levels
Formal definition

Given input:
- a set of training objects $x_1, \ldots, x_m, x_i$ vectors in $\mathbb{R}^a$
- a set of label mappings $y_1, \ldots, y_m$, each a subset of $Y=\{\lambda_1, \ldots, \lambda_n\}$

<table>
<thead>
<tr>
<th>i</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>...</th>
<th>$x_a$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0.1</td>
<td>${\lambda_1, \lambda_n}$</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>...</td>
<td>0.3</td>
<td>${\lambda_2}$</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>...</td>
<td>0.5</td>
<td>{}</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>4</td>
<td>1</td>
<td>...</td>
<td>0.6</td>
<td>${\lambda_1}$</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Objective:
- find a function $h: \mathbb{R}^a \rightarrow Y$ which maps $x_i$ to $y_i$
- as accurately as possible, as efficiently as possible
Formal definition

Alternative view:
- a set of training objects $x_1, \ldots, x_m, x_i$ vectors in $\mathbb{R}^a$
- a number of $n$ binary Target variables $y_i = \{0, 1\}$

<table>
<thead>
<tr>
<th>i</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>...</th>
<th>$x_a$</th>
<th>$y$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0.1</td>
<td>${\lambda_1, \lambda_n}$</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>...</td>
<td>0.3</td>
<td>${\lambda_2}$</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>...</td>
<td>0.5</td>
<td>{}</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>4</td>
<td>1</td>
<td>...</td>
<td>0.6</td>
<td>${\lambda_1}$</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>i</th>
<th>$x_1$</th>
<th>$x_2$</th>
<th>$x_3$</th>
<th>...</th>
<th>$x_a$</th>
<th>$y_1$</th>
<th>$y_2$</th>
<th>...</th>
<th>$y_n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0.1</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>2</td>
<td>1</td>
<td>...</td>
<td>0.3</td>
<td>0</td>
<td>1</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>C</td>
<td>3</td>
<td>0</td>
<td>...</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>D</td>
<td>4</td>
<td>1</td>
<td>...</td>
<td>0.6</td>
<td>1</td>
<td>0</td>
<td>...</td>
<td>0</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Objective:
- find a function $h: \mathbb{R}^a \rightarrow Y = \{0, 1\}^n$ which maps $x_i$ to a binary vector
- as accurately as possible, as efficiently as possible
Extreme Classification

Topics

Approaches can be roughly classified into

- Problem Transformation Approaches
  - Binary Relevance & simplifications
  - Hashing
- Decision Trees
- Landmark-based label selection
- Label Space Transformations
- Neural Networks and Embeddings
- Topic Models and Generative Approaches
Extreme Classification
Intro and Basics

- 1 Arturo Montejo Ráez, Luís Alfonso Ureña López, Ralf Steinberger. Adaptive Selection of Base Classifiers in One-Against-All Learning for Large Multi-labeled Collections
  - I. Katakis, G. Tsoumakas, and I. Vlahavas, Multilabel text classification for automated tag suggestion
- 2 G. Tsoumakas, I. Katakis, and I. Vlahavas, Effective and efficient multilabel classification in domains with large number of labels
  - S. Bengio, J. Weston, and D. Grangier. Label embedding trees for large multi-class task
Extreme Classification
Intro and Basics


- 4 Kilian Weinberger, Anirban Dasgupta, Josh Attenberg, John Langford, Alex Smola. Feature Hashing for Large Scale Multitask Learning. ICML, 2009
Extreme Classification
Intro and Basics

**Extreme Classification Basics/Decision Trees**


Extreme Classification
Label Space Transformations

- 14 Z. Lin, G. Ding, M. Hu, and J. Wang, Multi-label Classification via Feature-aware Implicit Label Space Encoding, in ICML, 2014.
Extreme Classification
Neural Networks and Embeddings

Extreme Classification
Landmark-based label selection

Extreme Classification
BR/ Topic Models


- 25 Piyush Rai, Changwei Hu, Ricardo Henao, Lawrence Carin. Large-Scale Bayesian Multi-Label Learning via Topic-Based Label Embeddings, NIPS; 2015