Task: Hierarchical Document Classification

- Sub-Task A: classify a book’s blurb into one or multiple general genres with a total of 8 classes, normal multi-label classification task.

- Sub-Task B: an hierarchical classification task where a book’s blurb can be associated at each level to several labels in a 3-level hierarchy: 8 labels on the root level, 93 on the 1st level, and 242 on the 2nd level.

  - Train one multi-label classifier for each parent node, target are the children labels.
  - Forces an hierarchical label constrain, no need for a post-processing step.
  - Training several classifiers: Root Node: 1 classifier; Level 1: 8 classifiers; Level 2: 42 classifiers.

Global Classifier

- Uses a flattened hierarchy and learns how to predict a vector of 343 dimensions.
- Advantages: single classifier to tune.
- Disadvantages: needs post-processing to enforce the hierarchical structure, large and sparse label space.
- IDEA: leverage on label co-occurrence to initialize the weight matrix of the hidden layer – not explored due to time constrains.

Results Sub-Task B - Dev Test (using Logit for prediction for Root Node)

<table>
<thead>
<tr>
<th>Filter Size</th>
<th>Filter Maps</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conf1</td>
<td>1, 2, 3</td>
<td>0.7163</td>
<td>0.7484</td>
<td>0.7320</td>
</tr>
<tr>
<td>Conf2</td>
<td>3, 5, 7, 10</td>
<td>0.5257</td>
<td>0.4603</td>
<td>0.4909</td>
</tr>
<tr>
<td>Conf3</td>
<td>1, 2, 3, 5, 7, 10</td>
<td>0.7353</td>
<td>0.7686</td>
<td>0.7516</td>
</tr>
</tbody>
</table>

Future Work

- Explore more features: based on author’s name and publication date.
- Local Classifier: explore different architectures and parameters for the different classifiers on Level 1 and 2, depending on size of training data.
- Global Classifier: initialization of the weight matrix in the last hidden layer based on the label co-occurrence.
- Properly set the prediction threshold for different classifiers experimentally.