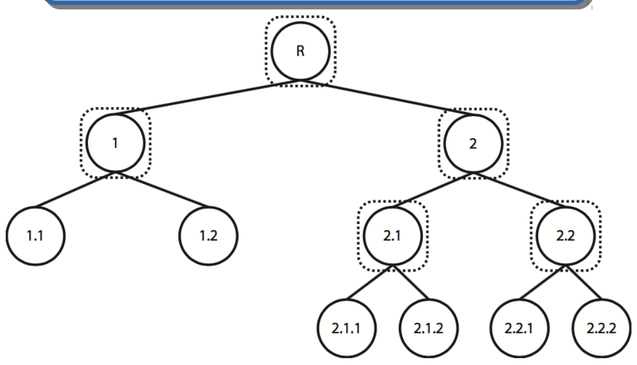
COMTRAVO-DS team at GermEval'19 Task 1 on Hierarchical Classification of Blurbs

David S. Batista and Matti Lyra
Comtravo GmbH
{david.batista, matti.lyra}@comtravo.com

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.

Local Classifier



- 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.

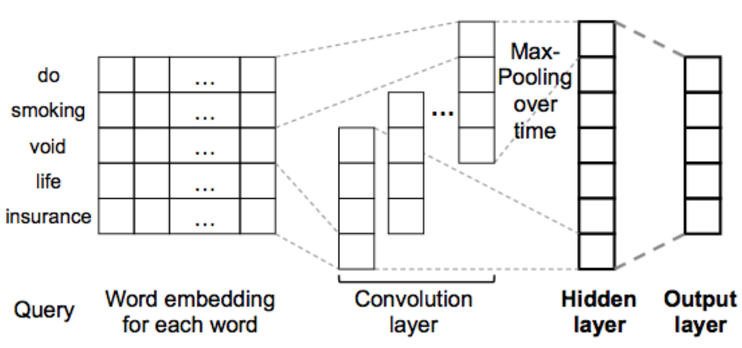
Root Node

- Logistic Regression (TF-IDF)
- Bi-LSTM
- Bag-of-Tricks

Level 1 and 2

- Convolutional Neural Networks

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.

Threshold adjustment on prediction time: 0.5 by default, if no predictions lowered to 0.4, and again to 0.3 if no predictions.

Local Classifier

Results Sub-Task A - Dev Test

Method	Precision	Recall	F_1
Logit (TF-IDF)	0.8211	**0.8359**	**0.8284**
CNN	**0.8542**	0.7879	0.8197
bi-LSTM	0.8062	0.7987	0.8024
Bag-of-Tricks	0.3787	0.6717	0.4843

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

	Filter Size	Filter Maps	Precision	Recall	F_1
$Conf_1$	1,2	300	0.7151	**0.5330**	**0.6108**
$Conf_2$	1,2,3	200	0.7144	0.5303	0.6087
$Conf_3$	1,2,3,5,7	300	0.7219	0.5235	0.6069
$Conf_4$	3,5,6,10	256	**0.7274**	0.5085	0.5986

Global Classifier

Results Sub-Task A and B - Dev Test

	Filter Windows	Filter Maps
$Conf_1$	1, 2, 3	300
$Conf_2$	3, 5, 7, 10	256
$Conf_3$	**1, 2, 3, 5, 7, 10**	**256**

	Precision	Recall	F_1
$Conf_1$			
Sub-Task A	0.7163	0.7484	0.7320
Sub-Task B	0.5257	0.4603	0.4909
$Conf_2$			
Sub-Task A	0.7353	**0.7686**	0.7516
Sub-Task B	0.5470	0.4717	0.5066
$Conf_3$			
Sub-Task A	**0.8389**	0.7659	**0.8008**
Sub-Task B	**0.6733**	**0.5032**	**0.5760**

Results on Test Set

Local Classifier

Task	Precision	Recall	F_1
Sub-Task A	0.8144	0.8255	0.8199
Sub-Task B	0.7042	0.5274	0.6031

Global Classifier

Task	Precision	Recall	F_1
Sub-Task A	0.7761	0.7839	0.7839
Sub-Task B	0.5672	0.5185	0.5418

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.

COMTRAVO

https://github.com/davidsbatista/GermEval-2019-Task_1