Evaluating Hierarchical Document Categorisation

Qian Sun, Aili Shen, Hiyori Yoshikawa, Chunpeng Ma, Daniel Beck,
Tomoya Iwakura, Timothy Baldwin
The University of Melbourne & Fujitsu Limited

Motivation

- Hierarchical document categorisation is a special case of multi-label document categorisation, where there is a hierarchical relationship among the labels.
- There is no standard benchmark dataset, resulting in different methods being evaluated independently and there being no empirical consensus on what methods perform best.

Previous Approaches

- **Flat approach** simply ignores the label hierarchy.
- **Local approach** makes predictions in a top-down fashion, along paths in the label hierarchy. It can be divided into three groups: (1) a local classifier per node; (2) a local classifier per parent node; and (3) a local classifier per level.
- **Global approach** optimises across all labels simultaneously, taking label hierarchy into account.
- **Hybrid approach** combines methods mentioned above.

Datasets

- **RCV1**: a collection of news articles published by the Reuters News between 1996 and 1997.
- **SHINRA**: a collection of English Wikipedia articles annotated by a fine-grained named entity set.
- **WoS**: a collection of abstracts from academic papers across different research domains and areas.

Evaluation Metrics

- Micro-F_1: gives more weight to frequent labels.
- Macro-F_1: gives equal weight to all labels.

Text Encoders

- TextCNN: A CNN made up of convolutional and max-pooling layers.
- TextRNN: A single-layer Bi-LSTM with a cell size of 64 where the concatenated hidden state at the last timestep makes up the document representation.
- TextRCNN: A combination of TextCNN and TextRNN.
- BERT: The hidden state of "CLS" from BERT is used as the document representation, using the base-uncased version.

Experiments – Hierarchical Methods

- **Flat**, baseline method ignores hierarchical information.
- **Recursive Regularisation (RR)**, a hybrid method, utilising simple recursive regularisation to encourage parameter smoothness between linked nodes.
- **Hierarchical Multi-Label Classification Networks (HMCN)**, a hybrid local/global approach, where each level in the model corresponds to a level in the label hierarchy. The global model consists of multiple linear layers with ReLU as the activation function
- **Hierarchy-Aware Graph Networks (Hi-GCN)**, an end-to-end hierarchy-aware global model that extracts the label hierarchy information to achieve label-wise text features.

Experimental Findings

- The choice of text encoder is a strong determinant of performance than the choice of hierarchical methods.
- The global model Hi-GCN achieves superior performance on all three datasets, indicating the necessity of capturing the hierarchy label structure holistically.
- The structure of the label hierarchy and class distribution also affect performance