End-to-End Learning to Index and Search in Large Output Spaces

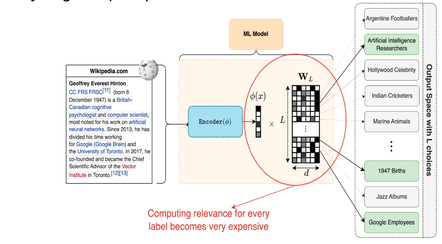
Nilesh Gupta, Patrick H. Chen, Hsiang-Fu Yu, Cho-Jui Hsieh, Inderjit S. Dhillon

TEXAS The University of Texas at Austin

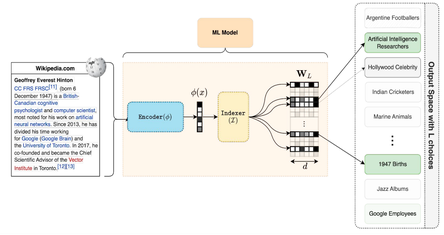
UCLA

Classification in Large Output Space

Many real-world scenarios (recommendation, openQA, etc) have very large output space i.e. L in millions/billions

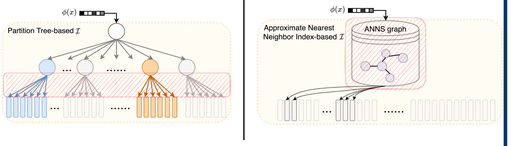


Efficient Search Index



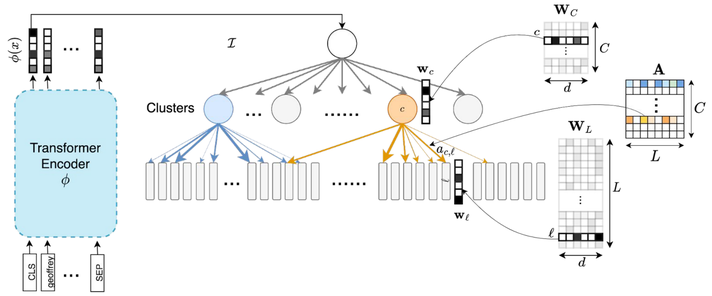
Existing Approaches

Existing approaches fix their index structure before training



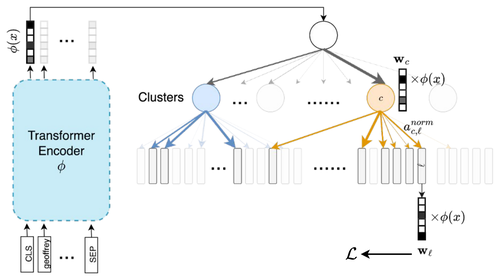
ELIAS : End-to-End Learning to Index and Search

- Relaxes partition tree-based index to weighted graph-based index
- Parameterize cluster-to-label edges as learnable adjacency matrix
- Learn A end-to-end with rest of the model parameters (encoder, classifiers)



Model parameters - $\phi, \mathbf{W}_C, \mathbf{A}, \mathbf{W}_L$

ELIAS forward



Step 1
Score all clusters → $\hat{s}_C \sim \mathbf{W}_C \times \phi(x)$

Step 2
Select top clusters → $\text{argtop-}b(\hat{s}_C)$

Step 3
Score all potential paths through top clusters
$\hat{s}_c * a^{norm}_{c,\ell}$

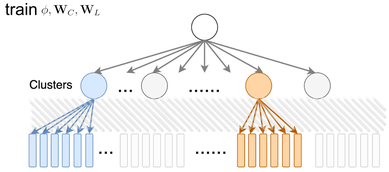
Step 4
Select top K labels based on path scores

Step 5
Evaluate label classifier for all top K labels
Final score of ℓ : $\hat{s}_c * a^{norm}_{c,\ell} * \sigma(\mathbf{w}_\ell^T \phi(x))$

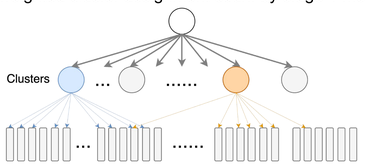
ELIAS staged training

- **Computational challenge** - operating on full adjacency matrix can be very expensive for web-scale datasets
  - Learn a row-wise sparse adjacency matrix
- **Optimization challenge** - because of flexibility in the model to assign a label to various clusters, hard for a label to get confidently assigned to only a few relevant clusters
  - Train in two stages

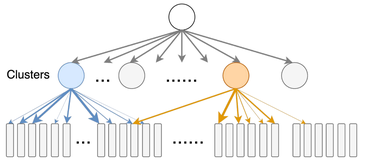
Stage 1: fix A as traditional partition clusters and train $\phi, \mathbf{W}_C, \mathbf{W}_L$



Initialize approximate row-wise sparse A based on weighted cluster assignment count by stage 1 model



Stage 2: train full model i.e. $\phi, \mathbf{W}_C, \mathbf{W}_L$, and non-zero entries of A



Experimental Results

State-of-the-art on several large-scale extreme classification benchmarks

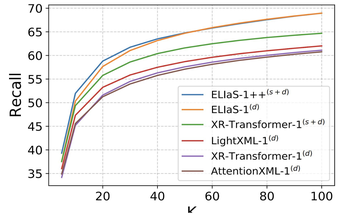
Amazon-670K

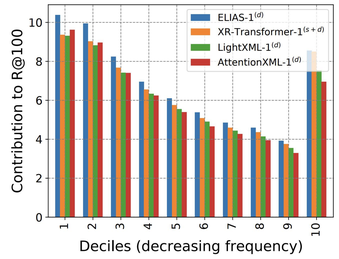
Method	P@1	P@3	P@5
AttentionXML	47.58	42.61	38.92
LightXML	49.10	43.83	39.85
XR-Transformer	50.11	44.56	40.64
ELIAS	50.63	45.49	41.60
ELIAS++	**53.02**	**47.18**	**42.97**

Wikipedia-500K

Method	P@1	P@3	P@5
AttentionXML	76.95	58.42	46.14
LightXML	77.78	58.85	45.57
XR-Transformer	79.40	59.02	46.25
ELIAS	79.00	60.37	46.87
ELIAS++	**81.26**	**62.51**	**48.82**

Upto 4% better at R@100 than the next best method on Amazon-670K





bit.ly/elias-paper

NEURAL INFORMATION PROCESSING SYSTEMS