MultiTask Pre-Training for E-Commerce Product Search

Xuange Cui, Wei Xiong, Songlin Wang

Heading1, Multi-Task Pre-training

In pre-training stage, we adopt mlm task, classification task and contrastive learning task to achieve considerably performance

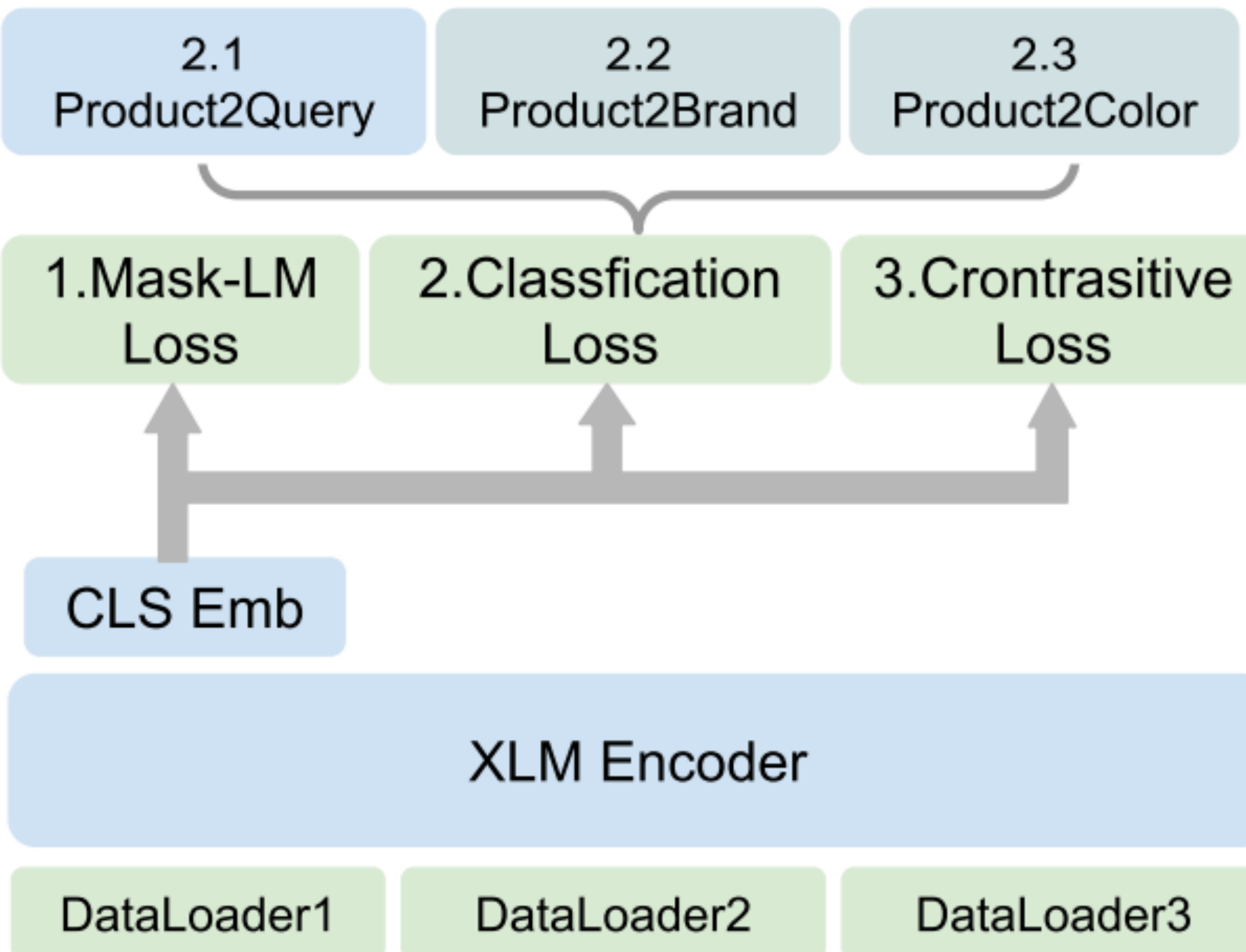

Figure 1: A schematic overview of our novel pre-training tasks. These tasks encourage the encoded representations to be more general.

Algorithm 1: Training a MultiTask model.

Input: DataSet $\mathcal{D} = \{(x, y, z)_i\}_{i=1}^{|\mathcal{D}|}$
Initialize model parameters Θ randomly ;
Model trainer T that takes batches of training data as input to optimize the model parameters Θ ;
Set the max number of epoch: $epoch_{\max}$;
for *epoch in* $1, 2, ..., epoch_{\max}$ **do**
  Shuffle $\mathcal{D}$ by mixing data from different tasks ;
  for $\mathcal{B}$ *in* $\mathcal{D}$ **do**
    // $\mathcal{B}$ is a mini-batch of pre-training task ;
    Compute loss : $L(\Theta)$;
    1. $L(\Theta)$ = Mask LM Loss ;
    2. $L(\Theta)$ += Classification Loss ;
    3. $L(\Theta)$ += Contrastive Learning Loss ;
    Optimize the model using $L(\Theta)$;
  end
end
Output: Pre-trained Model Θ

Heading2, Fine-tuning Methods

In fine-tuning stage, we use confident learning, exponential moving average method （EMA）, adversarial training（FGM）, regularized dropout strategy（R-Drop）and embedding mixup.

Moreover, we use a multi-granular semantic unit to discover the queries and products textual metadata for enhancing the representation of the model.

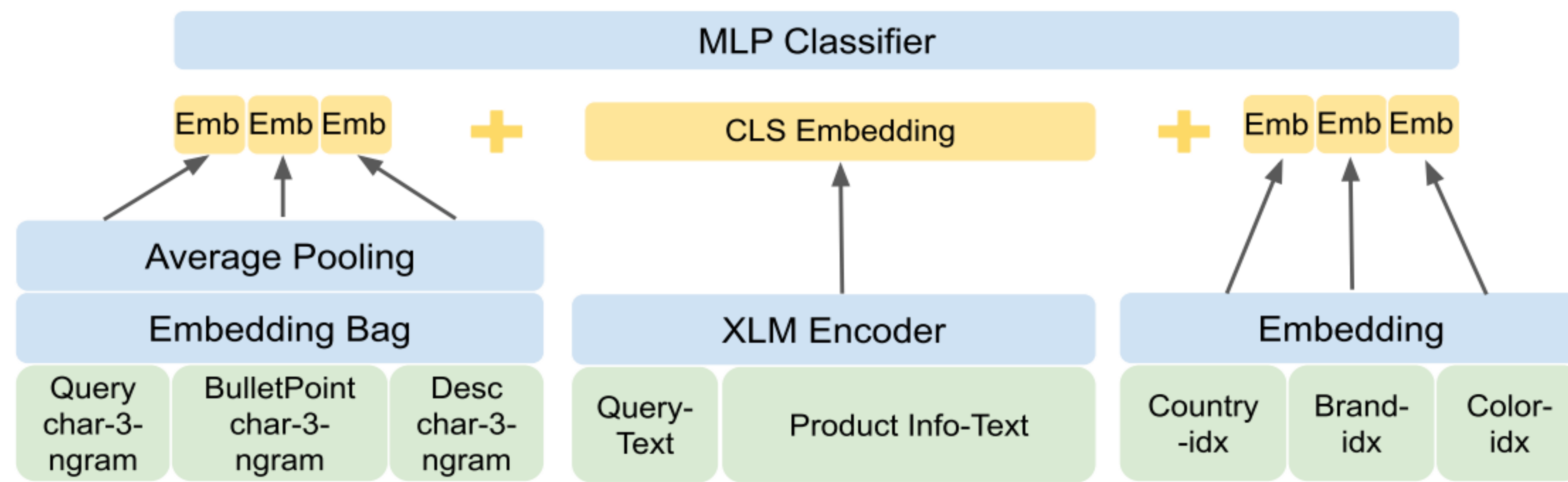

Figure 2: In fine-tuning stage, we concatenate the multi-granular semantic units, the [CLS] embedding from XLM encoder and the IDs' embeddings.

Heading3, Results

SubTask	Model	Metric	Ranking
task1	6 large models	ndcg=0.9025	5th
task2	only 1 large model	micro f1=0.8194	7th
task3	only 1 large model	micro f1=0.8686	8th

Table 2: Performance of our approach on the private leaderboard. In task1, we used six InfoXLM$_{large}$ models that fine-tuned by different datasets or methods. In task2 and task3, we used only one InfoXLM$_{large}$ model with the same network structure, as shown in Figure 2.

Pre-Training Task	CV-MLM Loss	CV-Micro F1
Mask LM	1.966	74.97
+Product2Query	1.969	75.05
++Product2Brand	1.978	75.08
+++Contrastive Learning	2.047	75.08

Table 3: The effect of different pre-training tasks and keep accumulating from top to bottom. We report the cross validation MLM-Loss and Micro-F1 Score × 100 in the task2 setting.

Methods	CV-Micro F1
+EMA	75.19
++FGM	75.30
+++R-Drop	75.43
++++Embedding Mixup	75.43

Table 5: The effect of different strategies and keep accumulating from top to bottom. We report the cross validation Micro-F1 Score × 100 in the task2 setting.

Confident Learning	CV-Metric
with-in-task1	NDCG, +0.005
with-in-task2	Micro-F1, -0.003
with-in-task3	Micro-F1, -0.002

Table 6: The effect of removing 4% noisy labels.

References

https://github.com/cuixuage/KDDCup2022-ESCI