Intro (5 datasets)

Learn continually from dynamic streams.

* A tweeter classifier needs to deal with trending topics which are constantly emerging.
* Suffer from catastrophic forgetting.

Some of researchers:

* Purely replay (store and re-trained previous tasks) => increases memory costs and training time.
* Regularization (add constraints on model parameters L1, L2 regularization) => utilizes previous examples efficiently via the constraints => preserve knowledge stored in the model to prevent them from changing too much while learning new tasks.

Problems: when learning tasks (e.g., new tasks; task generic information and task specific information) should be treated differently => these generic representation might function consistently while task specific representations might need to be changed significantly.

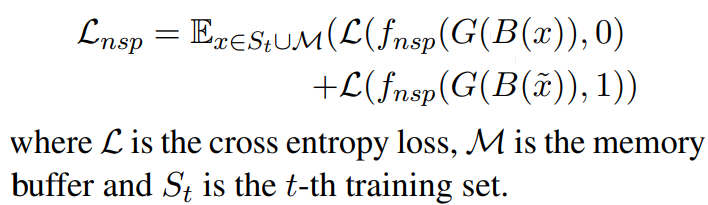
* An information disentanglement (Gỡ rối thông tin) based regularization method.

Detail

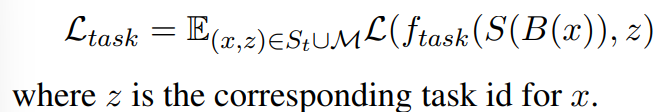
* **Task generic:** Disentangle the text hidden representation space (e.g., output representation of BERT)
* **Task specific:** (2 auxiliary tasks)
  + **Next sentence prediction (generic information) ~ (e.g., content; syntax)**
  + **Task identifier prediction (specific representations) ~ (e.g., style; polarity)**
* **When learn new task:** 
  + Constrain the task generic representation (stable)
  + Representations of task specific (flexible)
* **To not increase of memory and training time => exemplar (1% from K-Means).**

**Methodology:**

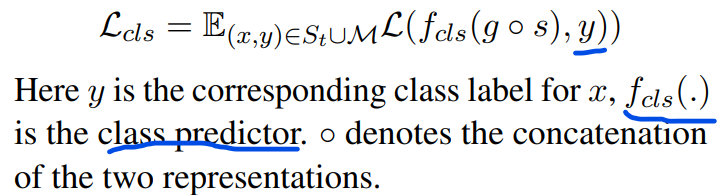
* **Information Disentanglement (ID):** 
  + To disentangle sentence representations into task generic space and specific space.
  + Separate regularizations are imposed on them for continual text classification.
* **Task Generic Space:** The generic information should roughly remain the same (syntactic knowledge). G(.) ~ **next sentence prediction => extract task generic g from r.**
  + Insert a [SEP] token into sentence: John đi đến nhà sách [SEP]. Sau đó, anh ấy quay trở lại nhà.
  + IsNext: John đi đến nhà sách *[SEP]*. Anh ta mua một quyển sách.
  + NotNext: John đi đến nhà sách *[SEP]*. Sau đó, anh ấy quay trở lại nhà.

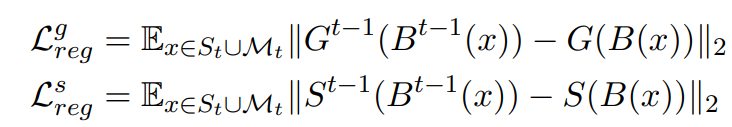


* **Task Specific Space:** For example, on sentiment classification words: “good” or “bad” could be very informative (i.e., sentiment classification), but they might not generalize well for tasks (i.e., topic classification). **S(.) ~ task-identifier prediction task => extract specific s from r.**

****

* **Text classification:** combine the g (from G(B(x))) and s (from S(B(x)))

****

* **Information Disentanglement (ID) based Regularization.**
  + **X from (S\_t) of t task is trained by B^{t-1}(.) G^{t-1}(.) and S^{t-1}(.) from t-1 task.**
  + **Two regularization losses separately:**
  + ****

**Zero-shot text classification with Reuse and Disentanglement information**