Share your Model instead of your Data: Privacy Preserving Mimic Learning for Ranking

Mostafa Dehghani, Hosein Azarbonyad, Jaap Kamps, and Maarten de Rijke

University of Amsterdam

Motivation

With the recent success of deep learning in many fields, IR is also moving from traditional statistical approaches to neural network based approaches. Supervised neural networks are data hungry and training an effective model requires a huge amount of labeled samples.

Problem:

- For many IR tasks, academia does not have access to a large amount of data.
- Privacy and confidentiality concerns prevent many data owners from sharing the data.
- Thus today the research community can only benefit from research on large-scale datasets in a limited manner.

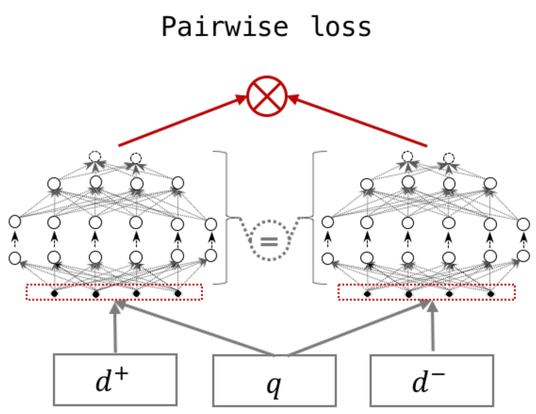
Solution:

- **Privacy Preserving Mimic Learning**
  - Use *aggregated noisy* versions of your models trained on *different partitions* of the trained data to *annotate large-scale unlabeled* datasets.
  - Use the created dataset for training new models.

Neural Ranker with Mimic Learning

- Train a very deep and wide teacher network on the original training data which leads to a big model that is able to express the structure from the data very well.
- Use the teacher model to annotate a large unlabeled dataset.
- Employ the annotated set to train a neural network which is called a student network.
- The student model makes predictions similar to the teacher model with nearly the same or even better performance.

In our experiments, as the ranker, we have used **Rank Model** by Dehghani et al. 2017* to construct teacher and student models:



- The input query and documents are passed through a representation learning layer that learns the representation of the input data instances, i.e. (q, d^+, d^-), and consists of three components:
  - an embedding function $\varepsilon : \mathcal{V} \rightarrow \mathbb{R}^m$ (where $\mathcal{V}$ denotes the vocabulary and m is the number of embedding dimensions)
  - a weighting function $\omega : \mathcal{V} \rightarrow \mathbb{R}$, and
  - A compositionality function $\odot : (\mathbb{R}^m, \mathbb{R})^n \rightarrow \mathbb{R}^m$.
- The model is optimized using the hinge loss (max-margin loss function) on batches of training instances and it is defined as follows:

$$\mathcal{L}(b;\theta) = \frac{1}{|b|}\sum_{i=1}^{|b|} \max\{0, 1 - \text{sign}(s_{\{q,d_1\}_i} - s_{\{q,d_2\}_i}) \left(\mathcal{S}(\{q,d_1\}_i;\theta) - \mathcal{S}(\{q,d_2\}_i;\theta)\right)\}$$

Teacher and student neural networks configurations.

Parameter	Teacher	Student
Number of hidden layers	3	3
Size of hidden layers	512	128
Initial learning rate	1E-3	1E-3
Dropout	0.2	0.1
Embedding size	500	300
Batch size	512	512

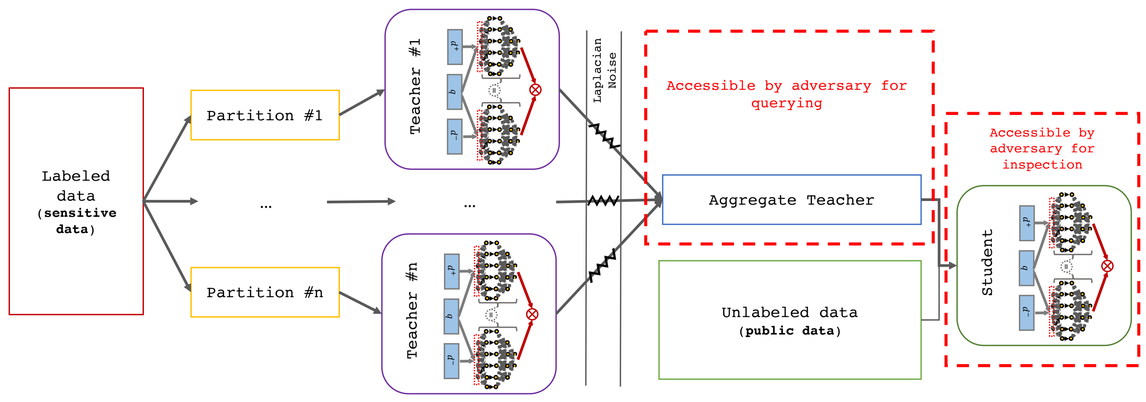
* Mostafa Dehghani, H. Zamani, Al. Severyn, J. Kamps, and W. B. Croft. "Neural Ranking Models with Weak Supervision", In Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR17).

Neural Ranker with Privacy Preserving Mimic Learning

There is a risk of privacy problems, both where the adversary is just able to query the model, and where the model parameters are exposed to the adversaries inspection.

Privacy preserving annotator/model sharing

- The sensitive training data is divided into n partitions.
- On each partition, an independent neural network model is trained as a teacher.
- Once the teachers are trained, an aggregation step is done using majority voting to generate a single global prediction.
- Laplacian noise is injected into the output of the prediction of each teacher before aggregation.



Experiments

Datasets

- Robust04 with a set of 250 queries (TREC topics 301-450 and 601-700) with judgments.

Exp 1: Effectiveness of Neural Ranker with Mimic Learning

- Research Question 1: Can we use mimic learning to train a neural ranker?

Performance of teacher and student models with different training strategies.

Training strategy	model	MAP	P@20	nDCG@20
Full supervision	Teacher	0.1814	0.2888	0.3419
	Student	0.2256	0.3111	0.3891
Weak supervision	Teacher	0.2716	0.3664	0.4109
	Student	0.2701	0.3562	0.4145

- Outcome 1: We can train a neural ranker using mimic learning. Using weak supervision to train the teacher model, the student model performs as good as the teacher model.

Exp 2: Effectiveness of Privacy Preserving Mimic Learning

- Research Question 2: Are privacy preserving mimic learning methods effective for training a neural ranker?

Performance of teachers (average) and student models with noisy and non-noisy aggregation.

Model	MAP	P@20	nDCG@20
Teachers (avg)	0.2566	0.3300	0.3836
Non-noisy aggregated teacher	0.2380	0.3055	0.3702
Student (non-noisy aggregation)	0.2337	0.3192	0.3717
Noisy aggregated teacher	0.2110	0.2868	0.3407
Student (noisy aggregation)	0.2255	0.2984	0.3559

- Outcome 2: Using the noisy aggregation of multiple teachers as the supervision signal, we can train a neural ranker with an acceptable performance.

Conclusions

- Sharing data raises many concerns such as violating the privacy of users.
- Instead of the data, we can train a model on the data and share the model.
- A student ranker model trained on a dataset labeled based on predictions of a teacher model, can perform almost as well as the teacher model.
- Privacy preserving mimic learning can be further used to guarantee the privacy of users.

NWO Netherlands Organisation for Scientific Research

University of Amsterdam

SIGIR 2017 Workshop on Neural Information Retrieval (Neu-IR'17)