TOPIC 7, EXERCISE 13

XTREME: Benchmark for Zero-Shot Cross-Lingual Transfer

XTREME is a benchmark for the evaluation of the cross-lingual generalization ability of pre-trained multilingual models. Read the relevant parts of the XTREME paper, alternative and write a short answer (a few sentences) for the following questions:

a) What is zero-shot cross-lingual transfer?

> Zero-shot cross-lingual transfer is scenario where annotated training data is provided in some language, but none is provided in the desired language.

b) What transfer approaches they test in the paper?

> Machine translation and multi-lingual representation-based approaches

c) What kind of datasets/tasks are used in evaluation? What is the general evaluation setting (e.g. train/test languages)?

> - Classification, datasets compiled from corpus': XNLI, PAWS-X. The gen eval setting is natural language inference and paraphrases from miscellanious sources including wikipedia / Quora as accuracy of classification as metric.
> - Structured prediction, datasets from UD and Wikiann and evaluation setting is POS / NER with F1 as metric.
> - QA, the datasets include multilangual datasets and the task is to extract spans with correct answer measured by F1 / EM

d) Summarize their overall findings. How well zero-shot cross-lingual transfer seems to work in their setting, and how do different methods compere with each other?

> Cross-lingual zero-shot transfer models trained on english data seem to outperform mBERT on same tasks in all cases. For structured prediction the improvement is not that significant. Using machine translation to translate training data to target lang improves results.