tensorflow / **tensor2tensor**

how to reproduce "one model to learn them all" results? #9

New issue

Closed **zzkszzks** opened this issue on 21 Jun 2017 · 3 comments

zzkszzks commented on 21 Jun 2017 • edited

"one model to learn them all" relates to many tasks, but T2T problems are for single task.
How to run multimodel training by T2T?
Thanks a lot!

3

lukaszkaiser commented on 22 Jun 2017 **Member**

Reproducing the MultiModel is not as simple as translation for 2 reasons: (1) you need to get the data (and some are not free and need pre-processing), and (2) it takes over a week to train even in a large distributed setup. I can help guide you through both if you're interested though :).

As for data, 5 out of the 8 problems are hooked up in the generator (t2t-datagen), so it should do the downloading and pre-processing for you. There are: `image_mscoco_tokens_8k_tune`, `wmt_ende_tokens_8k`, `wmt_ende_tokens_8k_rev`, `wmt_enfr_tokens_8k`, `wmt_enfr_tokens_8k_rev`. Then, for `wmt_parsing_tokens_8k`, you can also use the generator, but you'll need to get the Penn Treebank and maybe hack a little bit, similar to wsj_parsing. For `image_imagenet` you can follow the instructions here: https://github.com/tensorflow/models/tree/master/inception -- it's the same file we use at the end. The speech data might be the hardest, we got some help with this from friends who used Kaldi to transform the WSJ corpus into the frequency domain.

So yeah, getting this data together (and remember to use the same tokenizer and vocabulary!) is a bit of work. And then comes the training, which is in principle simple: just set $MODEL=multimodel, $HPARAMS=multimodel_1p8 and $PROBLEM=audio_wsj_tokens_8k_test-image_mscoco_tokens_8k_tune-wmt_parsing_tokens_8k-wmt_ende_tokens_8k_rev-wmt_enfr_tokens_8k-wmt_enfr_tokens_8k_rev-image_imagenet -- except that even on 8 GPUs that'd take a very long time. I'm actively working on making it faster, will be checking in updates to the model, but it's a lot of data to go through.

So the above tells how to replicate the full 8-problem setting. If you want to try smaller-scale, I think it's interesting to try the Transformer on multi-lingual translation. For example, just use the README instructions with `$PROBLEM=wmt_ende_tokens_32k-wmt_ende_tokens_32k_rev-wmt_enfr_tokens_32k-wmt_enfr_tokens_32k_rev` -- that 4-way translation model could be a good warmup before going all-in on the 8-problem one.

5

lukaszkaiser closed this on 22 Jun 2017

zzkszzks commented on 22 Jun 2017

Thank you for the detailed answer!

Assignees
No one assigned

Labels
None yet

Projects
None yet

Milestone
No milestone

Notifications

3 participants

colmantse commented on 11 Aug 2017

Hi,

Just wondering if the transformer four way translation operate like the multi-modal in sense that it applies a prelayer before the transformer for each language? Or it works by using side-constraints, like adding the extra language destination token at the input?

Best,
Colman

martinpopel referenced this issue on 22 Aug 2017

Multilingual translation with shared vocabulary, embeddings and softmax Closed
#242

fstahlberg pushed a commit to ucam-smt/tensor2tensor that referenced this issue on 22 Nov 2017

`Merge pull request #9 from wb14123/fix-readme` ... `abeafd2`