Base Model Description

Improved Attentive Reader

A thorough examination of the cnn/daily mail reading comprehension task

Chen, Danqi, Jason Bolton, and Christopher D. Manning. "A thorough examination of the cnn/daily mail reading comprehension task." arXiv preprint arXiv:1606.02858 (2016).

$$\begin{gathered}R_i = \phi^C_{RNN}(emb(C_i))[:] \\ r = \phi^Q_{RNN}(emb(Q))[-1] \\ a_i = Softmax(R_i^T M r) \\ u = \sum_i a_i R_i \\ o = Softmax(\phi_{Linear}(u))\end{gathered}$$

Generate representations of story tokens and the question, calculate the similarity between tokens and the question, then map the attended word(s) back to the answer.

Modified Model Description

Shared Embedding Reader

$$\begin{gathered}\phi(\cdot) = \phi^{RNN}((\phi^{word\_emb_{L_1,L_2}}(\cdot) + \phi^{entity\_emb}(\cdot)) \\ R_i^{(j)}[:] = \phi^{(C)}(C_i^{(j)})[:] \\ r_i^{(j)} = \phi^{(Q)}(Q_i^{(j)})[-1] \\ \gamma(R, r) = Softmax[Linear(r, \sigma(R, r) \odot R)]\end{gathered}$$

Shared embedding mechanism: WIP

Goal:

$$argmax_{\Theta_\phi, \Theta_\gamma} logP(A|Q, C, \Theta_\phi, \Theta_\gamma)$$

Use embedding vectors mapped to a shared space for L_1 and L_2, use the same embeddings for entity tokens, then use the same network to train on both languages.

Reader + Language Discriminator

$$\begin{gathered}\phi_i(\cdot) = \phi_i^{RNN}((\phi_i^{word\_emb}(\cdot) + \phi^{entity\_emb}(\cdot)) \\ R_i^{(j)}[:] = \phi_i^{(C)}(C_i^{(j)})[:] \\ r_i^{(j)} = \phi_i^{(Q)}(Q_i^{(j)})[-1] \\ \gamma(R, r) = Softmax[Linear(r, \sigma_i(R, r) \odot R)] \\ \delta(\cdot) = \delta^{MLP}(\cdot)\end{gathered}$$

The joint goal can be represented as: $$ argmax*{\Theta*\phi, \Theta*\gamma} [logP(A*|*Q, C, \Theta*\phi, \Theta_\gamma)

- \alpha \cdot log P(\lnot L|Q, C,\Theta*\phi, \Theta*\delta)] $$

Use different (or shared) embeddings for each language, then before the answerer layer, also pass the output through a discriminator network which tries to identify the source language. Through adversarial training, force the representation before the answerer network to be language-independent.

Dataset Collection Method and Description

English QA data:

https://github.com/deepmind/rc-data/ (https://github.com/deepmind/rc-data/) (Hermann et al., NIPS 2015)

Use CNN part of the dataset. Story - Question - Answer tuples, anonymised with coreference resolution.

Spanish QA data:

Collected from www.elmondo.es (via cached links on Wayback Machine) and processed to fit the format of the CNN dataset. Anonymised through named entity recognition.

Dataset	**English**	**Spanish**
Number of articles	0	0
SQA tuples	0	0
Vocabulary size	0	0
Total tokens	0	0

English and Spanish word embeddings:

https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md (https://github.com/facebookresearch/fastText/blob/master/pretrained-vectors.md)

(P. Bojanowski*, E. Grave*, A. Joulin, T. Mikolov, Enriching Word Vectors with Subword Information)

300 dimension word vectors trained on Wikipedia text.

Word alignment dictionary:

http://opus.lingfil.uu.se/download.php?f=EUbookshop%2Fdic%2Fen-es.dic (http://opus.lingfil.uu.se/download.php?f=EUbookshop%2Fdic%2Fen-es.dic)

(Raivis Skadiņš, J�rg Tiedemann, Roberts Rozis and Daiga Deksne (2014): Billions of Parallel Words for Free, In Proceedings of LREC 2014, Reykjavik, Iceland)

EU Bookshop parallel corpus dataset

http://opus.lingfil.uu.se/OpenSubtitles2012.php (http://opus.lingfil.uu.se/OpenSubtitles2012.php) (Jörg Tiedemann, 2012, Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012))

Open Subtitles dataset (more in-corpus aligned words)

Dictionary	**EU Bookshop**	**Open Subtitles**
Total aligned words	0	0
In-corpus aligned words	0	0

Results

Method	English (best)	Spanish (best)	Bilingual (best)	English (avg.)	Spanish (avg.)	Bilingual (avg.)
Individual embeddings	0	0	0	0	0	0
Mapped embeddings	0	0	0	0	0	0
Adversarial training	0	0	0	0	0	0
Mapped embeddings / Adv. Training	0	0	0	0	0	0

Mapping Method	English	Spanish	Bilingual
None	0	0	0
Unconstrained	0	0	0
Normalised + mean shift	0	0	0

Training Curve Comparison

Dictionary Size vs Accuracy

```
In [ ]:
```