

Bilingual Word Embeddings

TOPIC: **Implementation of BWE for comparable data**

- Read the paper given below.
- Download the training and test data presented in Section 3.
 - https://sites.google.com/site/ivanvulic/Vulic1000_test.tar.bz2
- Reimplement the model described in Section 2.
 - Train the model with the downloaded training data.
 - Evaluate on the test data. Try to modify the parameters to get as good results as reported in the paper.
- What was particularly easy/difficult in this practical project?
 - Document your code and indicate what was easy/difficult

Sources:

- Vulic and Moens, *Bilingual Word Embeddings from Non-Parallel Document-Aligned Data Applied to Bilingual Lexicon Induction*, ACL 2015.

Bilingual Word Embeddings

TOPIC: Bilingual Lexicon induction with BWE

- Present the task of Bilingual Lexicon Induction.
- Download the following test data, which is (also) described in the paper cited below.
 - https://sites.google.com/site/ivanvulic/Vulic1000_test.tar.bz2
- Train Bilingual Word Embedding models on this data.
 - Four models are described in Sections 2.1 (theory) and Section 3 (baseline models). Footnotes indicate where to find the source code.
- Evaluate the models on BLI and comment the results
 - Provide us with the files of the trained models and show us how you performed the evaluation.

Sources:

- Vulic and Moens, *On the Role of Seed Lexicons in Learning Bilingual Word Embeddings*, ACL 2016.

Neural Machine Translation

TOPIC: En-De translation on different domains

- Download an NMT model trained on En-De
 - http://data.statmt.org/rsennrich/wmt16_systems/de-en/
- Download a news, IT and biomedical test set from the WMT 2016 shared task
 - <http://www.statmt.org/wmt16/>
- Translate each data set using the downloaded model
- Make a detailed error analysis for each domain
 - For each domain, give a list of systematic errors made by the system. Explain why you consider these errors „systematic“ (how often and in which context did they occur?)

Sources:

- Sennrich, Haddow and Birch, *Edinburgh Neural Machine Translation Systems for WMT 16, WMT 2016, system description.*

Machine Translation Evaluation

TOPIC: Metrics for Automatic Machine Translation Evaluation

- What is the basic idea behind automatic MT evaluation?
- Which metrics have been proposed for MT evaluation?
 - Find examples in the scientific literature, such as BLEU, TER, METEOR, BEER, ChrF, etc.
 - Describe one (or more) metric(s) of your choice in more detail.
- Look for freely available tools which implement some of those metrics, download them, and try to run them.
 - How well do scores of different metrics correlate? Are the results similar on different languages? What else do you observe?
- What makes a "good" metric for automatic MT evaluation?

Sources:

- Bojar, Graham, Kamran, and Stanojević, *Results of the WMT16 Metrics Shared Task*, WMT 2016.
- Translations of a standard test set (newstest2016 from WMT):
<http://data.statmt.org/wmt16/translation-task/wmt16-submitted-data-v2.tgz>

Word Order in MT

TOPIC: Pre-reordering for Machine Translation

- Languages have different syntactic structure. MT systems must be able to put translated words into different order.
- Pre-reordering methods take a sentence in a source language $L1$ and (without translating) change the word order in a way that makes it more similar to sentences in target language $L2$.
- Tools like *Otedama* can be used to learn pre-reordering rules from data and then apply the learned rules.
 - How does *Otedama* work?
 - Try to apply it with $L1=English$ and $L2=German$. Try to find settings for *Otedama*'s parameters that work on your task and with your data.
 - (Optional: What are similarities & differences of *Otedama* and *Lader*?)

Sources:

- <https://github.com/StatNLP/otedama>
- <https://github.com/neubig/lader>
- Hitschler et al., *Otedama: Fast Rule-Based Pre-Ordering for Machine Translation*. The Prague Bulletin of Mathematical Linguistics, vol. 106, pp. 159-168. Charles University, Prague, Czech Republic, 2016.

Automatic Post-editing (APE)

TOPIC: Automatic Post-editing for Machine Translation

- Post-editing is the correction of errors in machine-translated content, and typically done by humans with the purpose of bringing error-prone MT output to publishable quality.
- Can *automatic* post-editing (APE) fix errors in MT output?
- Build an APE system for the WMT16 APE shared task. Use any method you like, and any publicly available software and tools. Run your APE system and analyze the results.

Sources:

- <http://www.statmt.org/wmt16/ape-task.html>
- Section 7 of: Bojar et al., *Findings of the 2016 Conference on Machine Translation*. WMT 2016.
- Three *Automatic Post-editing Shared Task* papers (system descriptions) from WMT 2016:
see <http://www.statmt.org/wmt16/papers.html>

System Combination

TOPIC: Machine Translation System Combination

- Based on a publicly available toolkit of your choice, build a machine translation system combination engine.
 - Use data from the WMT 2011 System Combination Shared Task: <http://www.statmt.org/wmt11/system-combination-task.html>
- Can you also combine five of the English-to-Czech MT systems from the WMT 2016 News Translation Shared Task?
 - Cf. <http://www.statmt.org/wmt16/translation-task.html> and http://matrix.statmt.org/matrix/systems_list/1844 .
 - Hint: Download <http://data.statmt.org/wmt16/translation-task/wmt16-submitted-data-v2.tgz> . Czech output of different MT systems is in [wmt16-submitted-data/txt/system-outputs/newstest2016/en-cs/](http://data.statmt.org/wmt16/translation-task/wmt16-submitted-data/txt/system-outputs/newstest2016/en-cs/) . You can split the files into halves, using the first half for tuning and the second half for testing. Try to combine "uedin-nmt", "NYU-UMontreal-NMT", "jhu-pbmt", "uedin-cu-syntax", "cu-tamchyna".

Sources:

- MANY: <http://www-lium.univ-lemans.fr/~barrault/MANY/>
- MEMT: <http://kheafield.com/code/memt/>
- Jane: <http://www.hltpr.rwth-aachen.de/jane/>