Machine Translation

16: Wrap Up and Exam Preparation

Rico Sennrich

with credit to Sharon Goldwater

University of Edinburgh

Neural Machine Translation Toolkits

large list at `https://github.com/jonsafari/nmt-list`

selected tools

- Nematus (Edinburgh; Theano/Tensorflow; RNN)
  `https://github.com/EdinburghNLP/nematus`
- Marian (Edinburgh; C++; RNN, Transformer) `https://marian-nmt.github.io/`
- OpenNMT (Harvard; Torch; RNN) `https://github.com/OpenNMT/OpenNMT`
- XNMT (CMU; DyNet; RNN, Transformer) `https://github.com/neulab/xnmt`
- Neural Monkey (Charles University Prague; Tensorflow; RNN)
  `https://github.com/ufal/neuralmonkey`
- tensor2tensor (Google; Tensorflow; Transformer)
  `https://github.com/tensorflow/tensor2tensor`
- fairseq (Facebook; Torch; CNN) `https://github.com/facebookresearch/fairseq`
- Sockeye (Amazon; MXNet; RNN, CNN, Transformer)
  `https://github.com/awslabs/sockeye`

Example Models and Training Scripts

- Edinburgh's WMT 2017 systems and training scripts
  `http://data.statmt.org/wmt17_systems/`
- replication of Edinburgh systems with Marian
  `https://github.com/marian-nmt/marian-examples/tree/master/wmt2017-uedin`

Alternative Textbooks/Tutorials

- Kyunghyun Cho (2015): Natural Language Understanding with Distributed Representation
  `https://arxiv.org/abs/1511.07916`
- Graham Neubig (2017): Neural Machine Translation and Sequence-to-sequence Models: A Tutorial
  `https://arxiv.org/abs/1703.01619`

Reminder: Learning Goals

at the end of this course, you will:

- understand the linguistic and computational challenges of MT
- know how MT is formalized as a machine learning problem
- have experience reading, using and modifying an MT implementation
- understand state-of-the-art approaches to MT and their limitations

Format of Exam

- Three main questions, each with several sub-parts.
- You **must** answer two out of 3 questions.
- If you attempt all three, only Q1 and Q2 will count (not the best 2!)
- The exam lasts two hours.
- You write your answers in a blank exam book.

Topics of Exam

material covered

- all lectures
- required reading: Koehn 13.2-5; 13.7-8

focus

- you will **not** have to memorize/reproduce equations...
- ...but you should understand and be able to work with them
- you will **not** work with computer code...
- ...but be prepared to discuss algorithms and architectures
- some questions will be **bookwork**: things you can memorize
- most questions will involve **application**, **analysis**, or **synthesis**

Example of Application

question

Identify two linguistic phenomena that make the automatic translation of this sentence difficult. Assume that the target language is German:

To solve it, he took the lead in the investigation.

Example of Application

question

Identify two linguistic phenomena that make the automatic translation of this sentence difficult. Assume that the target language is German:

To solve it, he took the lead in the investigation.

answer

- *lead* is ambiguous: leadership or metal
- *it* is an anaphoric pronoun, and we need to know the antecedent to produce the correct gender marking in German

Example of Analysis

question

neural machine translation operates with fixed and (for computational reasons) relatively small network vocabularies. Discuss three methods to increase the effective vocabulary size of a neural MT system, and discuss strengths and weaknesses of each.

Example of Analysis

question

neural machine translation operates with fixed and (for computational reasons) relatively small network vocabularies. Discuss three methods to increase the effective vocabulary size of a neural MT system, and discuss strengths and weaknesses of each.

answer

- we can use a dynamically changing subset of the vocabulary in the network at training and test time, computing an approximative softmax on this subset. This allows the increase of the model's vocabulary size with little computational cost, but word representations for rare words may be poor, and this solution does not allow the translation of unseen words.
- we can represent rare words via a special UNK symbol, and use a backoff dictionary to replace an UNK on the target side with the translation of the source word that it is aligned to. This is relatively effective for copying of names and other 1-to-1 translations, but relies on the attention mechanism, and does not consider 1-to-many/many-to-1 translations.
- we can use subword-level modelling (e.g. characters or BPE units). This allows the network to represent/generate an open vocabulary of words. A weakness is that increasing the sequence length can make training/inference less efficient.

Example of Synthesis

question

assume you want to translate between languages that are closely related, are very similar syntactically, and differ mostly in vocabulary and spelling of some words. You have a small amount of parallel training data. How could you exploit the similarity of the languages to learn a better neural MT system?

During the Exam

- Read all the questions first and decide which two to answer. (Plan on taking 5–10 minutes to do this.)
- You do not need to answer the sub-parts in order, but you **must** be clear which part you are answering when.
- Read the questions carefully
  - Make sure you answer the question that is being asked, and that you have fully answered it.
  - Be concise. Don't write down everything you know about topic X, but answer the question.
  - clearly cross out any work in the exam book that you do not want to count towards your mark.
    → If both correct and incorrect answers are there, you may not get credit for the correct answer.