METL: Méthodes Statistiques de Traduction Automatique 3C124

Paola Merlo

année académique 20019-2020

Questions on Attention Neural Machine Translation and MT evaluation

1. To sum up : Advantages and disadavantages of NMT (slides 47-48)

   (a) Choose two of each and explain.

2. Attention

   (a) Motivation for attention : what is the bottleneck problem of the encoded-decoder architecture in NMT ? (slide 58)

   (b) How do you compute the attention scores ?

   (c) It can be claimed that attention produces a kind of soft alignement : why ? (slide 68)

   (d) What are the advantages of attention ? (slide 74)

3. Go back and read slides 49-50 and also the hand-written notes on MT evaluation. We will discuss analyzing machine translation systems and evaluation metrics (modified from Stanford exercice assignment 4).

4. Analyzing NMT Systems.

   Here we present a series of errors we found in the outputs two NMT model. For each example of an Italian source sentence, reference (i.e., gold) English translation, and NMT (i.e., model) English translation, please :

   (a) Identify the error in the NMT translation.

   (b) Provide possible reason(s) why the model may have made the error (either due to a specific linguistic construct or a specific model limitation).

   (c) Describe one possible way we might alter the NMT system to fix the observed error. There are more than one possible fixes for an error. For example, it could be tweaking the size of the hidden layers or changing the attention mechanism.

   Below are the translations that you should analyze as described above. Note that out-of-vocabulary words are underlined. Rest assured that you don't need to know Italian to answer these questions. You just need to know English. The Italian words in these questions are similar enough to English that you can mostly see the alignments. If you are uncertain about some words, please feel free to use resources like Google Translate to look them up.

   (a) Source Sentence : Dunque, un altro dei miei favoriti, "La notte stellata".
   Reference Translation : So another one of my favorites, The Starry Night.
   NMT Translation : Heres another favorite of my favorites, The Starry Night.

(b) Source Sentence : Voi sapete che quello che io faccio é scrivere per i bambini, e sono probabil-
mente l'autore per bambini più letto in America.

Reference Translation : You know, what I do is write for children, and Im probably Americas
most widely read childrens author, in fact.

NMT Translation : You know what I do is write for children, and in fact, Im probably the
author for children, more reading in the U.S.

(c) Source Sentence : E' stato un mio amico a farlo Richard Bolingbroke.

Reference Translation : A friend of mine did that Richard Bolingbroke.

NMT Translation : A friend of mine did that Richard ¡unk¿

(d) Source Sentence : Basta girare intorno all'isolato per vederlo come un'epifania.

Reference Translation : Youve just got to go around the block to see it as an epiphany.

NMT Translation : You just have to go back to the apple to see it as an epiphany.

(e) Source Sentence : Mi ha salvato la vita, permettendomi di andare nel bagno dei professori.

Reference Translation : She saved my life by letting me go to the bathroom in the teachers
lounge.

NMT Translation : She saved my life by letting me go to the bathroom in the womens room.

5. Calculating BLEU.

(a) Please consider this example :

Source Sentence s : l'amore tutto può

Reference Translation $r1$: love can always find a way

Reference Translation $r2$: love makes anything possible

NMT Translation $c1$: the love can always do

NMT Translation $c2$: love can make anything possible

Please compute the BLEU scores for $c1$ and $c2$. Let $\lambda_i = 0.5$ for $i \in 1, 2$ and $\lambda_i = 0$ for $i \in 3, 4$
(this means we ignore 3-grams and 4-grams, i.e., don't compute $p3$ or $p4$). When computing
BLEU scores, show your working (i.e., show your computed values for $p1, p2, len(c), len(r)$ and
BP). Note that the BLEU scores can be expressed between 0 and 1 or between 0 and 100. In
this question we are using the 0 to 1 scale.

Which of the two NMT translations is considered the better translation according to the BLEU
Score ? Do you agree that it is the better translation ?

(b) Our hard drive was corrupted and we lost Reference Translation $r2$. Please recompute BLEU
scores for $c1$ and $c2$, this time with respect to $r1$ only. Which of the two NMT translations now
receives the higher BLEU score ? Do you agree that it is the better translation ?

(c) Due to data availability, NMT systems are often evaluated with respect to only a single reference
translation. Please explain (in a few sentences) why this may be problematic.

(d) List two advantages and two disadvantages of BLEU, compared to human evaluation, as an
evaluation metric for Machine Translation.