NLP2 Report

May 26, 2017

Empirical evaluation (max 3 points): Is the experimental setup sound/convincing? Are experimental findings presented in an organised and effective manner? Here we expected to learn about:

1. the data, experimental setup, practical constraints
2. training conditions (the various choices and hyperparameters discussed and justified)
3. ablation studies if applicable
4. test results
5. a critical discussion of findings

1 the data, experimental setup, practical constraints

1. You have hyper-parameters: SGD learning rate schedule, regularisation strength. Investigate them. Plot likelihood and validation BLEU.
2. You have features of 3 sorts: segmentation, lexical translation, word order. Conduct ablation experiments.
3. Report BLEU1 (it gives an ideal of how well your model performs lexical selection), report BLEU4 (it gives an idea of lexical selection and word ord).

1.1 experimental setup

To evaluate the model we use a Chinese-English corpus consisting of 44k sentence pairs [**?**]. This corpus has two advantages: the sentences are relatively short (at most 30 words, and most of them shorter than 10), and the test-set contain multiple reference-translations [**?**]

We use an adaptive scheme for the learning rate of SGD following [**?**], where we let the learning rate decay as an inverse exponential of the number of parameter-updates. That is, we define the learning δ_t at parameter-update t to be

$$\delta_t = \delta_0(1 + \gamma\delta_0 t)^{-1}$$

where δ_0 is the initial learning rate at $t = 0$ and γ and aditional hyperparameter to determined. See the plot **??** for the effect of different values of γ.

C: **R:** I'd like to reserve two twin rooms. **V:** My like reserve two one pair people room.
C: ? **R:** Do you do alterations? **V:** You may alterations you?
C: **R:** I'd like some crayfish. **V:** My like like some smaller lobster.
C: **R:** I have a sore pain here. **V:** My here very hurts. **G:** I'm so hurt here.
C: Chinese here! **R:** My's happy, green sir. **V:** Mary is not so old as henr .
C: ? **R:** How much is the breakfast? **V:** Breakfast much much?

Table 1: Example Viterbi-translations (V) against a reference translation (R)

1.2 Determining hyperparameter

In order to determine the settings for the hyperparameters involved in the Stochastic Gradient Descent (SGD) we performed a series of experiments on a small selection of training-data (200 sentence-pairs). These concern the choice of δ_0 and γ, the strength of the L_2 regularizer λ, and the size of the minibatch. Plots **??** report of the experiments to determine the choice of δ_0 and γ by the average likelihood of training-sentences over iterations, and plots **??** by the evolution of the BLEU score of the predictions the training-set. The plots in **??** report of the experiments performed to determine the right batch-size by tracking the evolution of the BLEU-score.

In these plots you can clearly see that...

The final settings chosen based on the experiments are $\delta_0 = 10$, $\gamma = 1.0$, $\lambda = 1$ and a minibatch-size of 1 sentence.

1.3 results

In tabel **??** we list the results of the

Table **??** lists the final choices of training

C: **R:** Can you change yen into dollars? **V:** Please my Japanese change into dollar. **I:** Please to yen exchange into dollars. **G:** Please convert the yen into US dollars.
C: , **R:** Well, it's not so serious. **V:** Well, nothing big problem. **I:** Well, nothing big problem.
C: **R:** I'm on my way to Chicago from Minneapolis. **V:** In from 'm 'm go to a I. **I:** The from minneapolis Minneapolis to Chicago the 'm.

Table 2: Example Viterbi (V) and IBM1 heuristic translations (I) against a reference translation (R)

a alterations you ? you
i the i alter ? do
your ? you a the a
? you alter your a
i a i alterations ? you
the the a do ? can
alterations ? you can you i i the
alterations you ? i the i
may ? alter you a the a
may do ? alter i a a
i ? you may
a can alter ? you your
i the i may alterations ? you
alter do ? can your the the the
alter ? you may your a the i
alterations you ? may a
alter do ? you i i i
the the i your may do ? alterations
i alter you ? can you
a the a may you ? alterations
a do ? alterations may you
the a i alterations ? can
? do may your the i i
your ? you can the a the
? you alterations can your the the a
can ? do alterations your the
alterations you ? your a a i
? do alterations can
i a the you ? alterations you
can ? do alterations your a a i
your can alterations do ?
the the i your may ? alter
a a a ? you alter may your
your ? you alterations can i a i
alter you ? can your the the the
? you alterations can i
may ? alterations your a a
alterations you ? may a
you ? may you i the a
i the i your alterations ? may
your ? do alter can i the the
your ? you can i a i
i the the ? you alter
? you your i
alter you ? your i the i
? do alter your the a a
i alterations ? you may
? you may the the the
may do ? i a

Table 3: An impression of the massive search-space of possible translations for the sentence '?'. Shown is a selection of 50 out of a total of 26.514 sentences in the target forest. Note for example that only 1.458 sentence (< 6%) have the question mark at the end of the sentence—the most modest requirement for a proper translation of the sentence. Viewed this way, it is quite an achievement that the trained model managed to selecte 'you may alterations you ?' as the viterbi translation.

feature k	ϕ_k
span:rhs:src-rc:1-2	8.754e-14
span:rhs:src-rc:2	4.006e-20
span:rhs:src-lc:0-2	3.390e-20
span:rhs:src-lc:1-1	-0.514
span:rhs:tgt-lc:0-0	-0.531
type:translation	-1.387
type:deletion	-1.387
trans:/one	-4.916e-39
trans:/a	-4.916e-39
skip-bigram:*	-4.916e-39
trans:/me	-4.916e-39
type:terminal	-2.085
type:insertion	-2.085
ins:i	-2.085
span:rhs:src:0	-2.085
span:lhs:tgt:2-3	-2.085
ins:the	-2.085
ins:a	-2.085
span:lhs:tgt:1-1	-2.128
span:lhs:tgt:1-2	-2.129
span:lhs:tgt:3	-21.709
span:lhs:tgt:0-3	-21.709

Table 4: A selection of high-scoring features from the trained weights vector