INF5820, 2014, Exercises in MT

This is the first time INF5820 has a regular exam. Hence there are no earlier sets of exam exercises. Some of you have wondered on which level the exercises will be – what to expect in terms of mathematical formulas, for example. The following should give a representative sample, and it is good training to actually work with them and write out complete solutions. The list is of course not exhaustive. Don't get shocked if you get questions on the exam which are not included here ☺

MT in general

Exercise 1.1

Classical MT distinguishes between three different strategies. What are the three different strategies? What are their merits and challenges compared to the other two?

Exercise 1.2

The most immediate idea for MT is to use a dictionary and translate word-by-word. Some of the problems for this strategy have to do with typological differences between languages. Describe 3 typological differences between languages which make problems for this simplistic approach and explain how they are problematic.

Exercise 1.3

The most immediate idea for MT is to use a dictionary and translate word-by-word. Some of the problems for this strategy have to do with typological differences between languages, but there are also other problems which do not reside in typological differences. What sort of problems are these? Illustrate with examples and explain why they are problematic.

MT evaluation

Exercise 2.1

How is human evaluation of MT quality performed? Which challenges are there when one designs such a study?

Exercise 2.2

Which other properties of an MT system might be relevant for evaluation besides quality?

Exercise 2.3

a) Some well known automatic evaluation measures in language technology (and information retrieval) are precision and recall. What do they measure? How can they be applied to MT output?

b) The following shows one (human) reference translation and the output of two MT-systems. Evaluate the two outputs with respect to precision and recall.

- Sys.A: You register for the ride at all three cabins the night before
- Sys.B: You enter the trip at all three cabins the night before
- Reference: You may register for the trip at all three lodges the evening before

Exercise 2.4

a) The most well known automatic evaluation system for MT is called BLEU. Explain how BLEU works. You do not have to go into the details of how the brevity penalty is calculated.

b) Strictly speaking, BLEU should be calculated for a whole corpus and not for individual sentences. Why is that?

c) By carefully selecting our examples, we may still be able to apply BLEU to a corpus of one sentence. Calculate the BLEU score of the system output given the 3 reference translations and explain the steps as you go along.

Sys: the valley is wide and great , the trail is good , and on both sides has large spikes .
Ref.A: the trail is good ; the valley is wide and grand , with lofty peaks flanking it on both sides .
Ref.B: the valley is wide and beautiful , the path is good , and there are soaring peaks on both sides .
Ref.C: the valley is broad and splendid , the trail is good , and peaks rise on both sides .

Statistical MT and the IBM models

Exercise 3.1

Explain the noisy channel model and how it splits the translation task into more than one step.

Exercise 3.2

a) Explain what an alignment is in the word-based IBM models.

b) A central formula in describing IBM model 1 is the following

$$P(\mathbf{f},\mathbf{a}\,|\,\mathbf{e}) = \frac{\varepsilon}{(k+1)^m}\prod_{j=1}^{m} t(f_j\,|\,e_{a_j})$$

Explain the different elements of the formula and in particular what e_{a_j} refers to.

c) Which simplifying assumptions are made in IBM model 1 to arrive at this formula?

d) In which respects does IBM model 2 differ from model 1?

Exercise 3.3

Without bringing in mathematical formulas, describe the idea of the generative model underlying IBM model 3 and how it differs from the generative model for IBM models 1 and 2.

Training the system and the EM-algorithm

Exercise 4.1

Which preparations must be done to a training corpus to make it suitable for training an SMT system? We assume that you start out with the electronic version of a book in one language and its translation into another language.

Exercise 4.2

Without going into the mathematical formulas, explain how the EM-algorithm can be used to simultaneously estimate the most probable word alignments of a bilingual corpus and the word translation probabilities.

Exercise 4.3

a) Why can't we train IBM model 3 with the EM algorithm similarly as we trained IBM model 1?

b) What sort of algorithm can be used to train model 3? Explain how it works.

c) IBM model 3 is known to give better results than IBM model 1. Why do we start the training of model 3 by training model 1 first?

Phrase-based translation

Exercise 5.1

What is the mathematical model for basic phrase-based translation?

Exercise 5.2

a) In training a phrase-based model, one starts out with a word-aligned corpus. Describe how phrases are extracted from this. You may illustrate with an example if you prefer.

b) How are the phrase translation probabilities estimated?

Decoding

Exercise 6.1

a) The translation process is a computationally hard problem. Which property of translation explains why it is so hard compared to similar problems, say HMM tagging or speech recognition?

b) Which simplifying assumptions might be made to circumvent this problem? Such assumptions are more undermining for some language pairs than for others. Explain!

Exercise 6.2

a) What is the basic set-up for phrase-based decoding? What are the data structures and in which order are the tasks performed.

b) Phrase-based decoding applies beam search. What does that mean in this context? Will we always get the best solution with this strategy? Why not?

Exercise 6.3

a) It is a particular problem for the intuitive set-up for phrase-based decoding that the system may choose to translate the easiest parts first. How might this arise, and why can it be devastating for the procedure?

b) A way to avoid this problem is to estimate future costs. Explain the basic ideas of this procedure without going into details with respect to how the future costs actually are calculated.

Log-linear models, tuning, reranking

Exercise 7.1

a) Both IBM model 1 and the basic phrase-based model are called log-linear. What does that mean?

b) What does it mean to tune a log-linear model?

c) Explain the main steps in the tuning of a phrase-based model, e.g. in the Moses system.

d) How can more features be added to a log linear model? How can one argue that with tuning, in principle these extra features will not destroy the results?

Exercise 7.2

a) What does it mean to add a reranker to a translation system?

b) How can a reranker for an SMT system be trained?

Exercise 7.3

A reranker can also be added to a rule-based translation system. Explain how a reranker is added to the transfer-based LOGON system.