Exam questions[edit]

January 2021[edit]

Only 3 questions due to Corona.

1. Word embeddings and language models
    1. Compare skip-gram and CBOW
    2. What is a suitable loss function for BERT's masked token prediction step
    3. Skipgram exercise: calculate the probability of X being in the context of Y where you are given the skipgram matrices (like exercise session 1)
    4. Find the word that is the most related to X given the Skipgram word embeddings (same as the weights of the layer between the 1-hot encoded input and the hidden layer => you can use inner product similarity or cosine similarity for this
2. Constraints and ILP
    1. Give examples of and discuss ways to integrate constraints into ML models
    2. Give an example of a constraint which would be useful in the context of finding temporal relations in a text => the ILP examples from the exercise session about temporal relations could be used for example
    3. Exercise where you have to design an ILP constraint. Tip: write all logical binary combinations and mark which are forbidden due to the constrant, then calculate all calculations which offer results which are unique in more than their sign e.g. A+B+C, -A+B+C, A-B+C, and A+B-C are the only relevant possible combination as -A-B-C, +A-B-C, -A+B-C, and -A-B+C just reverse the sign of the previous calculations. Then you find out which (combination of) calculation(s) gives a result which allows you to separate the wrong combinations and make these your constraints.
3. Design a machine translation system with the following constraints
    1. The architecture has to be universal so you can train it for multiple languages and technical domains (e.g. legal, biology, ...)
    2. As a side product your company sells translation memories as well (training data)
    3. Avoid linguistic resources and tools as they might not be available for each language/domain
    4. Make sure the translations are contextually correct

August 2019[edit]

Provided in the PDF below.

January 2018[edit]

1. Discuss RBMT, PFT and NMT. Compare the advantages and the disadvantages of these three methods.
2. Given two syntax trees

        1. Write down the PCFG for these trees.
        2. Parse the sentence "time flies like a banana" using Earley Parsing.
        3. Show the resulting parse tree(s) and their probability.
        4. Discuss the results.
        5. How could we improve this parsing-method?
    3. Give two phrases which were tagged using B-I-O tags combined with argument tags (as seen in the semantic role labeling presentation)
        1. Explain what these tags mean.
        2. If you're using MaxEnt, design and explain 3 feature functions.
        3. If you were to model this using integer linear programming, design 2 constraints and explain.
        4. Assume you're using a conditional random field. Design one extra feature function that ensure the transitions between labels.
    4. Theory questions
        1. I forgot the first one.
        2. Compare classic N-grams with a neural network approach. How do both deal with unseen words?
    5. Design question: you have to design a query answering system that, given a question, will query a relational database and return the results. You had to explain this in sufficient detail, and explain all methods that you would use.

January 2015[edit]

1. Given three phrases.

```
a) Tag all the phrases using the Penn treebank POS tags.
b) Compute the HMM transition probability matrix (do not apply smoothing).
c) Compute the HMM emission probability matrix (no smoothing).
d) Given a new phrase: compute the bigram probabilities for this phrase.
e) Discuss the results of d). What would happen if we aplied smoothing?
```

2. Given two phrases:

Fruit flies like a ripe banana. Time flies like an arrow.

```
a) Write down the PCFG for this corpus.
b) Convert it to CNF.
c) Apply CKY parsing.
d) Show the resulting parse trees and their probabilities.
e) Discuss the results.
```

3. Given a hidden Markov model and a series of observations, compute the most likely joint state of the HMM.

4. Given three text snippets, compute the association of two pairs of words using pointwise mutual information association measure and the Lin association measure.

5. Question 5, January 2015.

6. You want to build a system which finds companies, brands, names, etc. which are mentioned on web pages. How would you do this? Go step by step, in enough detail, using the techniques mentioned in the course. Mention the bottlenecks involved in the process.

January 2015[edit]

1) Given an FSA, give the matching regular expression and give 10 strings that can be generated by the FSA. The FSA was like this:

```
q_0 -> q_1 <-> f <- q_2
       |            ^
       |            |
        - - - - - -
with q_0 -> q_1 = a
   q_1 -> f = b
   q_1 -> q_2 = b
   q_2 -> f = a
   f -> q_1 = a
```

2) You get 2 parse-trees of the following (simple) sentences: "time flies like an arrow" and "time flies like a banana".

```
a) create the corresponding PCFG
b) convert the PCFG to CNF (chomsky normal form)
c) perform CKY algorithm on the sentence "time flies like a banana"
d) draw all possible parse trees and give their probabilty
```

3) Given 5 sentences (sentence 1 is the same as sentence 3 and sentence 2 is the same as sentence 4). The word "anybody" didn't appear in the 5 sentences, the word "nobody" appeared 3 times.

```
a) Making use of bigrams, compute the probability of the word "nobody" and
the word "anybody", using maximum likelihood estimation MLE, Laplace
smoothing and Good Turing smoothing.
b) Compare every method and give their advantages and disadvantages.
```

4) Theoretical question: the question handled "pairwise coreference classifier" (discourse analysis) and had 2 subquestions:

```
a) how are "coreference chains" formed, based on the result of the
classifier? (not completely sure about the exact question but it was
something like this)
```

```
b) give 5 features (seen during the lectures) which can be used for this
classifier
```

5) Theoretical question: given the slide of the last lecture about semantic parsing. On these slides there is an example execution of "stochastic gradient descent", an optimizing algorithm. Explain the different iterations of the examples and explain how the algorithm computes, derives every value of every parameter.

June 2009[edit]

```
1) Same as June 2015 but without the part "give 10 strings that can be
generated by the FSA".
2) Same as June 2015.
3) What is "rhetorical structure theory"? Give relevant examples,
explanation and applications.
4) Given 3 English sentences.
 a) compute the similarity between 'flight' & 'debris' and 'flight' &
'Airbus' using the lin assoc method
 b) in which application can these similartites be used?
5) Theoretical question...cant quite remember
6) Suppose you have an e-mail management prgram that consists of two parts:
 - Task manager: gets all the tasks from the incoming e-mails
 - Time manager: saves the overview of the appointments and reminds you of
deadlines of tasks
How would you do this?
```