

Third practical and final report

Ekaterina Shutova

ILLC
University of Amsterdam

9 May 2018

Third practical

- **Evaluate** the word embedding models that you have built:
 - Skip-gram
 - Bayesian skip-gram
 - Embed-align
- **SentEval framework:**
 - Alexis Conneau and Douwe Kiela. 2018. *SentEval: An Evaluation Toolkit for Universal Sentence Representations*.
- Data and evaluation scripts will be provided with the SentEval toolkit
- Detailed instructions will be provided by Samira at tomorrow's lab session.
- **Deadline for the final report:** 31 May, 2018 – firm deadline!

Report

Goal:

- To describe your work building the models and evaluating them
- To explain in detail of your implementation
- To describe your results and their analysis
- To mimic the language and organisation of a research paper

Format:

- four pages in length
- typeset in two column
- use latex if you can – we will provide a template.

Report: Structure

- **Introduction:** summarise the goals, techniques, findings – like in a scientific paper
- **Background:** briefly describe the intuitions behind word embeddings and summarise existing techniques
- **Methods:** describe your models and details of your implementation – we need to be able to reproduce your results
- **Evaluation:** describe the SentEval framework
- **Results and analysis:** present your results and analyse them – what can we learn from them?
- **Conclusion:** the main conclusions of your experiments

General Tips

- Math notation – define each variable (either in running text, or in a pseudo-legenda after or before the equation)
- Define technical terminology you need
- Avoid colloquial language – everything can be said in a scientific-sounding way
- Avoid lengthy sentences, stay to the point!
- Do not spend space on “obvious” things

Results

- Tabulate your results in stages; normally in more than one table
- Each table should have a theme (e.g., comparison in a particular task or a related set of tasks)
- Metric should be clear from table even without having to read the text
- Table caption should be descriptive of the results and the dataset / subset you are using

Analyse the results

- What can we learn from your results?
- Are particular models better suited for particular tasks?
- Why? Does it have to do with the model architecture? With the context definition? With the data used?
- Notion of “interestingness” of a result – can you connect the result to a related observation that might be slightly non-obvious

Null results

- If you find a positive effect of a more advanced model, it obviously worked
- Something else may have worked better, but it does not matter because you brought positive proof
- If you find a negative effect, it may have two reasons: you didn't try hard enough, or the effect is really not there
- You believe the latter, but you have to convince your readers that you tried everything reasonable
- That is why in the “real science world”, null results can be problematic
- Here in the ULL practical, they are not!

An ideal report

- Precise, scientific-sounding, technical, to the point
- Little general “waffle”/chit-chat
- Not boring – because you don’t explain obvious things too much
- Efficient delivery of (only) the facts that we need to know to understand/reimplement
- Results visually well-presented and described with the correct priority of importance of sub-results
- Insightful analysis – speculation should connect to something interesting and not be too much; the reader “learns something new”
- No typos, no colloquialisms – well-considered language
- This normally means several re-draftings (re-orderings of information)

Master project topics (for next year)

- Semantic composition and metaphor
- Multimodal semantics (language & vision)
- Neurocognition of language
- Multilingual NLP and modelling variation