WS 2016 Project 2:
Due by: 29 March 2016, 23h55

This project counts towards 30% of your final grade for this course.

1 Project description

The topic of this project is sentiment mining. Given a dataset of LEGO product reviews, the goal is to extract the sentiment (neutral, positive, or negative) from each review[1].

You must use **two** sentiment mining approaches:

1. The first approach can be any automatic sentiment classification method you wish (out of the box, such as http://sentistrength.wlv.ac.uk, https://www.uclassify.com/browse/uclassify/sentiment, http://text-processing.com/demo/sentiment/, or implemented by you).

2. The second approach is the deep learning method described here (you can use the code provided): http://nlp.stanford.edu/sentiment/.

You will be given a dataset of LEGO product reviews, and you are asked to train your two sentiment mining approaches on this data using 3 fold validation. Then, you are asked to measure the accuracy of each of your methods.

In addition, as part of this project, you are asked to assist in producing the ground truth of the dataset that you will be using. The dataset consists of 4902 LEGO product reviews that have been crawled from http://brickset.com/reviews/eurobricks. Each of you is asked to manually read 43 of these reviews and decide (as a human annotator) if each of these reviews as a whole is neutral, positive or negative. The reviews can be found here: https://docs.google.com/spreadsheets/d/1Bfq2fPRiwwCGS6JgpyrVP4OSyPuhuU_sO77xnFKWSoQ/edit?usp=sharing. This is a Google sheet that contains the reviews allocated to each of you. You are expected to add your decision about the sentiment of each of your 43 allocated reviews by Friday 18 March at the latest. **Failing to complete your annotations by Friday 18 March** will automatically result in a -5% reduction of your final grade on this project. The complete 4902 LEGO

[1] You are asked to output the sentiment of each review as a whole, not of each sentence in a review. For example, a review may contain both positive and negative sentences. You are asked to produce a single decision (positive, negative, or neutral) for the whole review.

product reviews with the annotations created by your is the data on which you will run your two sentiment mining approaches.

2 Submission

2.1 What to submit

You must submit **a single tar.gz file** that contains:

1. your report in pdf (not the latex sources), formatted according to the template used in Project 1,
2. the output of your two sentiment classification approaches for each product review, along with their accuracy, and
3. all code that you used.

You should not submit your ground truth annotations. We can see them in the Google sheet.

Everything in your submission must be anonymous (i.e., do not write your name in the report or your code). There are no length restrictions for the report, however it must contain **all the sections in the template**, plus any more sections you wish to add.

2.2 How to submit

- You must upload your submission to Absalon **by 29 March 2016, 23h55, at the latest**.
- If you are unable to submit via Absalon for some reason, you must send your submission by e-mail to c.lioma@di.ku.dk **with cc to** oswin.krause@di.ku.dk and nhansen@di.ku.dk **by 29 March 2016, 23h55 at the latest**.
- Submissions received after the deadline without prior approval for e.g. medical reasons or similar, by C. Lioma, will not take part in the peer-assessment. This results in an immediate -20% reduction of your final portfolio grade (15% for the peer-assessment you will not make + 5% for your missing amendment list - see the *Portfolio Guidelines* for details).

3 Competition

In the first week of April (week 14) we will have a competition of your sentiment mining approaches on unseen data of LEGO product reviews. You will be asked to run the two approaches you submitted in Project 2 on a small set of unseen data. The best performing approach will win the first prize, which includes a package sponsored by LEGO (with actual LEGO products, among other things).

The outcome of the competition does not affect your assessment on this project. Your project will be assessed on the basis of the report you submit, not on the basis of how well or badly you did in the competition. For example, if someone wins the competition but submits a very bad report, he/she will receive a very bad grade, and vice versa.

More details on the exact dates and logistics of the competition will be given in the next weeks.