Semester Project – Weekly report
14/03 to 21/03

Title: Extending Dynamic Structure in Memory Network for Response Generation

Teacher: Faltings Boi

Supervisor : Mi Fei

Student: André Cibils

1. Goal of the passed week:
    - Prepare presentation (17/03)
    - Try YerevaNN's code
    - Play with YerevaNN's code

2. Work done & problems encountered:

    YerevaNN's code works really well, but it takes a lot of time on my old laptop. I trained the network on the 14-bAbi task (which is time reasoning) only, as it took 4 days. I got the same results (accuracy ~75% [1]).

    Howover, the original paper achieved an accuracy of 100% for the task 14. Multiple reasons can be found to this huge difference: first of all, there is hyperparameters to tune. Secondly, YerevaNN uses GloVe, which, according to Therne (the guy who did the tensorflow implementation) gives significantly worst results. Finally, the paper describe the need for multiple episodes for some tasks, and I'm not sure YerevaNN's version allow it (I need to check it).

    YerevaNN only implemented basic DMN and some enhancements (dmn_basic, dmn_batch, dmn_qa_draft and dmn_smooth) but not DMN+. The changement are described here:
    - DMN basic is the DMN described in the original paper: AMA:DMN for QA
    - DMN smooth uses the square of the Euclidean distance instead of abs in the attention mechanism.
    - DMN batch is equivalent to DMN smooth but it has minibatch training support.
    - DMN qa draft is a draft version designed for answering multiple choice questions.

    I look up the code: it is interesting and somehow pretty clear, but it lacks comments, description and I don't agree with some implementation choices. I added a lot of comments and documentation, which helped me understand the code quite a lot.

    YerevaNN's version also have the option "reccurent or feedforward" for the answer module. I didn't look it up yet, but this is interesting.

3. Potential goal for next week:

    I'm planing to continue to change the code, deplace some function, adding more doc, etc. This should be enough to feel that the code is "mine".

    I think we can keep the DMN smooth and batch implementation, adding these as options. Transforming the DMN into DMN+ seems crucial as well in order to achieve the same results than thoses in the paper. We can also work on hyperparameter tuning and removing GloVe, but DMN+ does not use GloVe anyway, so I think this is the priority.

4. References

    [1] http://yerevann.github.io/2016/02/05/implementing-dynamic-memory-networks/#initial-experiments