Week 14

11-777

Due: Thursday, Nov 30, by 4:30 pm

Instructions

A small yet significant part of the course is answering a few questions based on a paper you just read. By design, some of these questions are open-ended and the purpose is to start a thinking process. Some of those questions may already have been answered in recent literature (or in the given paper itself). We encourage you to, both, come up with your own ideas and survey recent papers before answering these questions.

A few key points to take care of:

- Answer **two out of the three questions** (you are welcome to answer all 3), unless stated otherwise.
- A lot of questions ask you to suggest changes or make alterations. It would be nice to support it with logical / mathematical arguments. Cite all sources used in the process of coming up with your answer. Figures and equations, if they support your arguments will be appreciated.

Questions

Captioning Images with Diverse Objects

1. What happens in typical image-captioning models? Why are auxiliary training objectives important for the mechanism adopted in the paper? How is the described model jointly trained?

2. How does the model learn to describe novel objects using semantic embeddings and inputs from the visual classifier? How is evaluation done for objects absent / infrequent in MSCOCO?

3. In Table 1 and 7, observe that for “couch”, the F1 as well as METEOR scores are significantly higher for DCC than NOC. Why do you think this can be the case? Does this indicate that NOC is biased towards certain kinds of objects? Why / why not?