

15-400 Report 1

Lichen Zhang

January 29, 2020

1 Major Changes

Previously, I proposed to improve the PinSAGE architecture using some local clustering algorithm, but then I found a problem: what I really what to show is using local clustering algorithm instead of some very naive random walks will in fact improve the performance of the SAGE architecture. So there is a better way to test this idea: an earlier paper called GraphSAGE, their main goal was to test the performance of this architecture, so I will adapt their strategy, instead of implementing an entire system. There's in fact no very big change, since nonetheless I need to write the code for local clustering before touching other parts.

2 Accomplishments

I have implemented many utility and data structures for the first local clustering algorithm, namely, the capacity releasing diffusion algorithm. Also, I have met with my mentor to discuss more deeply and try to better understand the paper for this algorithm.

3 Meet Milestones

Originally I planned to use one month to finish the test version of capacity releasing diffusion, it seems I am a little bit ahead of the plan, since I can finish the test version next week.

4 Surprises

By reading and better understanding the paper, we might be able to modify the algorithm and simplify it.

5 Look Ahead

Within next two weeks, I will finish the test version of capacity releasing diffusion, and hopefully, integrating it into the GraphSAGE architecture. Also, next week I will discuss the proof of the paper with my mentor and see whether we can modify that to make it simpler and better.

6 Revisions

Currently, the only revision is instead of implementing an entire recommender system that can be exploited in industry, I will integrate into a smaller architecture called GraphSAGE. After reading the code, I found that it is much easier to do the integration, so I might have time to try several different local algorithms.

7 Resources

Currently, no.