Research Plan for the Next Conference

The goal is to do a complete study on method-level comment generation for the methods that have dependent methods in them. Intuition is that these kinds of methods are prevalent in codebase and when doing summarization of code, ignoring the dependent methods will result in poor summarization. Our technique aims to utilize the information of dependent methods while doing code summarization and generate meaningful method level comment generation for code understanding. We plan to formulate this research into three parts:

Need to prove the importance of the research

Before proposing anything, we need to show that this research is important. For this we will try to do some qualitative analysis, primarily:

1. We will show a statistic focusing on **how many methods are dependent and how many methods are independent** in real world scenarios, by mining some popular GitHub repositories.
2. Another statistic we can show is that **how often dependent methods change over time and how often independent methods change over time**. We aim to use Codeshovel tool for that.

By showing these statistics we can build a solid ground for our research.

Prepare Dataset and analyze with the baseline models

After proving the importance of the research, we will move forward to make a benchmark dataset by mining GitHub repositories and run the baseline models available for code summarization on the prepared dataset. Intuition is that these models would not perform on satisfactory level. More specifically, we will try to:

1. Prepare a dataset from GitHub open-source projects (**Research Contribution**). This dataset will be specially built for methods having dependencies in them.
2. Then generate method summaries with different baseline models, on our dataset.
3. Then we will do manual evaluation, showing how acceptable the method summaries are. We can utilize our **summer students** in this part.
4. Intuition is generated summaries will not be useful for code understanding, which will take us in developing our Code Summarization technique.

Development of new technique

After figuring out the gaps in the existing study, we will move forward to developing our own technique for method-level comment generation for methods with decencies. In this stage, we will:

1. Propose and develop a novel technique for code summarization
2. Generate summaries on our benchmark dataset with the developed technique
3. Check the qualities of the generated summaries **(Survey / Controlled experiment / Manual Evaluation)**
4. Compare with GILT tool, if possible.
5. Development of a tool based on our technique. (if necessary or may be that will be future work)

This will conclude the study.

Targeted Conferences

I am planning to target good conferences, as I am feeling optimistic about the work. Currently in mind are:

1. **FSE:** September 5
2. **MSR:** Mid-November