Project Title:

Objectives:
Our project aims to generate a synopsis from a given report using the principles of Natural Language Processing and Reinforcement Learning. The main aim of such a simplification is to reduce linguistic complexity of the text, while retaining its meaning and all associated original information.

Importance of the Project:
Why synopsis generation?
It is for automatic simplification, so even people without the required skill-set can make sense of the report quickly. Since it is an industrial tie-up, currently the main application of the project will be to analyze corporate reports and decrease the lag between receiving a report and understanding. This can also be extended to various other applications, given datasets specialized enough, like making texts simpler for those with cognitive disabilities.

Why are we using reinforcement learning?
Modes without reinforcement learning often pick-up lines directly from the references. Reinforcement learning can help us in ensuring that the sentences are simplified, so that proposed model doesn't just copy them in toto. Better results presented by algorithms like DRESS and SARI show us the feasibility of the same. Also, reinforcement learning models perform additional tasks like generation and compression tasks.

Background:

Text simplification in an important aspect of modern NLP task. In the last one decade or so different models have been developed to utilize the potential of Reinforcement Learning in different NLP applications including Text simplification. The initial models simply extract the rare and complex words from the report and replace them with more common words. Subsequent approaches are more holistic in the sense that they analyze the entire report as a whole and make simplifications using the given resources and form multiple new sentences. Then all these sentences were compared to find the one that provides the gist of the given document in the simplest way. This simplest summary is considered to be the output of the system.

In recent years an even better model namely DRESS, that uses deep reinforcement learning, has been developed. The major drawback in the previous models was that most of the time their algorithm would just parrot back what was given in the resource or make a few trivial changes at the most.

However, in our proposed approach we proceed in a different way. The proposed algorithm first works out a solution and then on the basis of a few test parameters it is given a reward. The reinforcement method works on the reward values in an iterative way to maximise the overall reward. This ensures that an optimal model is achieved for text simplification. As of now this

model has been implemented for simplifying independent sentences. In our approach we would extend it for simplifying an entire document, or a collection of related documents which may require establishment of relationship between different terms occurring in multiple documents.

Relevant references:

- Sentence simplification with deep reinforcement learning (by Xingxing Zhang and Mirella Lapata)
- Automatic Text Simplification by Horacio Saggion

Work plan:

- Year 1:
- Year 2:
- Year 3:
- Year 4:

Expected outcome:
To generate a deep reinforcement learning model which is fairly accurate <insert accuracy bound +- error> in generating a synopsis.

Budget details & justification: