Machine Learning Engineering Bootcamp

Capstone Project Proposal

Background:

- 3GPP (3rd Generation Partnership Project) is a global consortium of telecommunications standards organizations that develops protocols and technical specifications for mobile telecommunications.
- 3GPP's work is divided into "Releases," each containing a set of specifications for a specific technology version (e.g., Release 15 introduced 5G). Each release work is conducted by meetings.
- 3GPP RAN1 (Radio Access Network Working Group 1) meetings are specialized gatherings within the 3GPP organization focused on the standardization of the physical layer of radio access technologies. RAN1 is one of several working groups under the Radio Access Network Technical Specification Group (RAN TSG) within 3GPP.
- Participants in 3GPP RAN1 meetings are typically representatives from various companies, organizations, and academic institutions.
- Participants submit technical contributions that propose changes, optimizations, or new features. These contributions are then discussed, revised, and either approved or rejected through consensus from the participants.
- The participants submit their contributions usually a week prior to the meeting date. This allows the meeting participants to study and understand the views of the participants.
- Each meeting consists of many different topics categorized under agenda items. For each agenda item, participants submit their contributions (available on 3GPP ftp server). On average, each meeting receives 1500~2000 contributions from the participants. These documents are technical in nature and written in a domain specific language.

Motivation and Problem Statement:

- Many stakeholders, participants and observers seek to quickly grasp the key ideas of certain technical contributions. For instance, a standard delegate focused on a specific agenda item in a particular meeting wants an overview of the submissions, while an engineering manager leading a team of researchers and engineers needs to understand the broader direction of technological development. Academic professors involved in related research may also be

interested in the general scope of the discussions. Government bodies and policymakers aim to understand the direction of these discussions to assess their potential impact. Similarly, telecommunications operators are concerned with how ongoing talks might affect their operations in the future. Providing clear, concise summaries of these technological discussions allows stakeholders, participants and observers to better prepare, evaluate the impact on their respective fields, and focus on advancing their technology and research efforts.

- A common approach is to generate summaries manually. For instance, to understand the perspectives of other companies and develop competitive technologies, each participating company reviews contributions and creates summaries. These summaries are used by standard delegates, engineers, researchers, engineering managers, and senior leadership to gain insights into the anticipated direction of upcoming discussions and technological advancements. Studying and summarizing such a large number of contributions from many different participants within a very short time (usually a week) is a challenging task along with other responsibilities and meeting preparational work.
- Government bodies and policymakers often outsource this work at high cost due to limited in-house expertise or resources. Academic professors may also spend significant amounts of valuable time studying these topics. Operators and other observers, lacking sufficient resources, may experience delays in their efforts or overlook important information.
- These manual, labor-intensive, domain expertise processes are time-consuming and expensive, with a significant financial, social and economical impact to the respective companies, organizations or governments.

Objective:

- Automatically generate summaries of 3GPP standard contribution documents—a large collection of technical, domain-specific papers written in English—for use by companies (research and development domain companies, telecommunication operators/service providers, equipment manufacturers), government and other organizations.
- These summaries will help such users quickly and effectively understand the direction of discussions in the upcoming 3GPP meeting in a timely manner. Additionally, they enable the companies to develop new technologies, products, better positioning and preparing for the discussions. These summaries enable governments, organizations and operators to prepare for the upcoming technology developments in a timely manner.
- The summary report should be organized in a format easy to follow by many audience/interested parties mentioned. Summary should include the document

ID, company, title, the purpose, the main points (eg. observations, proposals, remarks etc). The summary should be concise, extract the appropriate details as necessary. The summary avoids too many details, fabricated information (eg. information not mentioned in the document).

The dataset:

- The contributions for each meeting is available (public data) in a zip file in 3GPP ftp server. The zip file may contain single or multiple files. These files are in the format of MS word files, especially the long text documents while few documents may be in some other file formats such as powerpoint slides.
- Each contribution contains texts, plots, graphs, diagrams. Summary generation will be done on the text of the document.
- An example of 3GPP ftp folder for meeting #118 is provided here: https://www.3gpp.org/ftp/TSG_RAN/WG1_RL1/TSGR1_118/Docs
- Some example contributions for agenda item 9.1.1 and some non-contribution files for RAN1 meeting #118 is provided in folder "DataSet" and described in the table below.

File name	Description	Content
R1-2405808.zip	Contribution in agenda 9.1.1 from company FUTUREWEI	Contribution
R1-2405950.zip	Contribution in agenda 9.1.1 from company Google	Contribution
R1-2406492.zip	Contribution in agenda 9.1.1 from company NVIDIA	Contribution
R1-2405780.zip	Draft Agenda of RAN1#118 meeting	Not a contribution
R1-2405798.zip	LS in. Contains two word files.	Not a contribution
TDoc_List_Meeting_RAN1 #118.xlsx	TDoc list document	Not a contribution