

PROJECT

MULTI-DOCUMENT EXYRACTIVE TEXT SUMMARIZATION USING META-HEURISTIC OPTIMIZATION ALGORITHMS

PROJECT REPORT- CODING – PPT REQUIREMENT

I am working on a project involving multi-document extractive text summarization utilizing meta-heuristic optimization algorithms. Let me give you more details about the project. It's about multi-document extractive text summarization using meta-heuristic optimization algorithms. Actually the project is about multi document text summarization using Firefly Algorithm, which is based on a research paper , I will share that paper with u which i have taken as the base paper , that paper is based on single objective optimization and its fitness function uses 3 factors which are Topic Relation Factor, Cohesion factor, and Readability factor.

Now as this paper is using FA and it uses rouge scores (1,2,L,SU4) and datasets used for the task are DUC 2002, 2003, 2004 , now what i have to do is making something unique by using the base paper either i can use another meta heuristic optimisation algorithm like any new algorithm or making changes or adding few more steps in preprocessing, or document representation like here in the paper that is the base paper there are steps like preprocessing, then document representation then fitness function calculation, then it uses Firefly Algorithm, as i have seen in some papers they have added steps like feature extraction or calculating sentence informative score so based on all the above information u need to generate something unique by adding it with the base paper.Either U can follow all the steps in exactly similar manner just by changing the algorithm OR by using the same Firefly Algorithm but then u need to create a new fitness function by combining it with the fitness function of some other paper.Below i have shared the base paper.And i will share some more papers from which you needs to take either the algorithm or the preprocessing and fitness function part.

Coding do

1. Use datasets which (DUC 2002, 2002,2004) AS according to the BASE PAPER.
 2. Additional step include Karna which will be done from any other paper which will fits with the method.
 3. 3 factors are used in base paper (TRF,CF,RF) all these are used in the code.
 4. Fitness function calculation step will include 3 factors plus your additional factor.
 5. Base paper have step you have to do like this Preprocessing step uss meh 4 sub steps hai like case conversion, tokenization, stop word removal, stemming all this will be calculated one by one as mentioned in the base paper.
 6. Additional step include any modern research paper from year 2023 or 2022. After preprocessing and before fitness function calculation.
- And finally at the time of fitness function calculation, combine those 3 base paper factors and additional factors and design a new fitness function code.
7. Use ROUGE scores, such as ROUGE 1, 2, L, SU4, all these are used in base paper and also use some more evaluation metrics.

Take additional steps from another modern research paper published around 2022 or 2023. Take those steps which looks relevant and it fits perfectly with our work. As at last we need to compare it with the base paper that how our method is better than the base paper

Multi document extractive text summarization using nature inspired swarm algorithm.

Now the base paper uses Firefly Algorithm and as a fitness function they used 3 factors (TRF,CF,RF) , datasets used are DUC 2002,2003,2004.

Rouge scores used are ROUGE1,2,L,SU4 , in which they have a good score in rouge 1 and 2 best results by comparing it with other algorithms like PSO, GA. Their method has outperformed these above 2 algorithms.

But their method us single objective.

My work is also on single objective.

The steps they have used are ;

Preprocessing , document representation, fitness function calculation, algorithm apply.

In preprocessing steps used are tokenization, case conversion, stop word ,stemming. For document representation they used TF-IDF CALCULATION. So now, here comes my task to create something unique by following the Base Paper.

I need to take steps from the base paper and from another paper so that it becomes something unique.

As my task is to create something unique so i will follow the base paper as it has used 3 factors in fitness function so i will also use that but i will join one more factor to it. So for joining it i need to add that thing in the starting means after preprocessing step, which will be join with the fitness function.

lease fulfil my requirements. Okay I have to give everything written on details. You will follow every instruction okay once done the coding. You also guide me how to run on my laptop. Remember I need to go run on spyder

PLEASE NOTE

PLEASE READ

Firstly in the report, the problems are the format must be in latex you should need to share me the latex file so that's why you need to do in overleaf platform so that I can also check the latex file.

Secondly, the report only contains information about different meta heuristic optimisation algorithms but where is the discussion about my work, you need to show the graph (that you can see from the report ANPR that's my btech report) take a reference from that graph and even from that report and then you are discussing about the datasets that even I haven't used in my project mention only the things which are there in my project the things that I am including

3rd, literature review, write a literature review properly a descriptive look at my report of mtech 3rd sem that I have shared earlier before the starting of the work I have made literature review of 34 research papers so now you need to include 50 research paper reviews and have to keep their references in the report at last.

Show some results in the report and how the setup has been done.

Add survey findings, in survey findings show that why we have chosen this method.

Add, Future scope.

Just look at the reports that I have shared with you, you will get all the idea how to make it.

Now about ppt, in the ppt again I need latex file, secondly add some diagrams like differences between abstractive and extractive in that you can include diagrams to make it look good and creative

Add flowcharts written by you not copied from google.

Literature review table and references at last don't write anything from the shared ppts those are of different topics not similar to mine.

Again saying look at the ppts i have shared from that u can get an idea what i am trying to say. Just take the idea don't copy it. In ppt, explain about the algorithm that we are using add mathematical formulas make it look realistic .

Like the first step of firefly algorithm is population initialization so, show that how it has been done don't just write description add diagrams, maths to that so that it will be easier for me to explain that how its working. Lastly , just one suggestion take help from those ppts and reports shared by me to get an idea of what i am saying then it will be easier for u to understand.

The university format i have shared with u a long ago , please check it was rar file , so that you can make it in that format

Yeh hai format in latex and need to do in overleaf platform.