Technical Report

LegalEase: Summarization of legal documents and query-answering chatbot

30th March, 2024

Proposed Method

Issue with the old approach

1. **Legal-Pegasus** : We were initially using 8000+Indian judgemental data for fine tuning the model. We have the corresponding summaries dataset as well .But the issue was the tokenization limit of 1024. So, to increase the input capacity of the existing model, we started giving the whole document into chunks with the token limit of 1000. Meaning that the model will give a summary of every chunk. But this resulted in a huge summary output, not very useful.As well as,the accuracy score which was calculated using ROUGE score is not effective.

Updated Baseline Result : The F1 score of ROUGE-1 and ROUGE-L is around 0.2043.In new result the scores are updated to 0.3574 (for ROUGE-1) and 0.3294 (for ROUGE-L)

```
ROUGE-1 Precision: 0.1313
ROUGE-1 Recall: 0.7600
ROUGE-1 F1 Score: 0.2240

ROUGE-2 Precision: 0.0568
ROUGE-2 Recall: 0.4252
ROUGE-2 F1 Score: 0.1002

ROUGE-L Precision: 0.1198
ROUGE-L Recall: 0.6933
ROUGE-L F1 Score: 0.2043
```

```
ROUGE-1 Precision: 0.2526
ROUGE-1 Recall: 0.6618
ROUGE-1 F1 Score: 0.3574

ROUGE-2 Precision: 0.1236
ROUGE-2 Recall: 0.4049
ROUGE-2 F1 Score: 0.1836

ROUGE-L Precision: 0.2328
ROUGE-L Recall: 0.6112
ROUGE-L F1 Score: 0.3294
```

Old Result   Improved Result

In the above improved result ,We have followed the steps:

Breaking the judgment files into small chunks **->** then using legal_pegasus and using mean_cosine similarity with the corresponding summary files **->** generated summaries for each chunk **->** training our model on these chunks and their corresponding summaries **->** evaluating the results.

Methodology

To improve our model, we first need to train the model on an updated dataset.

1.Generation of our own dataset from existing data

We have 'Judgment'(d) and 'summary'(s) datasets. But as the input token is limited we cannot use the 'Judgment' dataset as a whole to train our model. As we are taking inputs in chunks, we cannot assume that 's' is the summary of all chunks. We need summaries for each chunk of input.

Approach :

1. Suppose we have **'d'** as a complete document and **'s'** is its summary. Using Mean cosine similarity, we first map every sentence in 's' to the most similar sentences in 'd'.

2. Now, initially we have input in chunks format, let say **{d1,d2,...dn}** making a total **n** chunks.So, for every chunk **'di'**, we do Mean cosine similarity with **'s'**,so to get important/relevant sentences in **'di'**, which will contribute to the summary. Combining all these important sentences, we get the summary **'si'** for the chunk **'di'**.

3. In this process some chunks have a lot of important sentences, while some don't have any.

4. Now, for each document in the training data, we have chunks of datasets and their corresponding summaries generated by Mean cosine similarity.So, our training dataset is now suitable ,which will help in training the model better. As our model takes input in chunks and generates summaries.

The output we generated :

Here we see the summaries for each chunk. We have stored all the chunks and corresponding summaries in specified files.

> LLM Model Training using Transformers > New_summary > ≡ 1.txt

```
    It raises the question whether munici pal property tax and urban immoveable property tax payable under the relevant Bombay Acts are allowable deductions under section 9 (1) (iv) of the
    The court held that the income from the property was not subject to the tax provisions of Section 9 (1) of the Income Tax Act and therefore the tax officer had no power to determine the
    The Tribunal, however, agreed to refer two questions of law to the High Court of Judicature at Bombay, namely, (1) Whether the municipal taxes paid by .<n>the applicant company are an a
    Chunk 3: Summary - The question for our determination is whether the munic ipal property tax and urban immoveable property tax can be deducted as an allowance under clause (iv) of sub se
    The court held that the tax shall be payable by an assessee under the head ' income from property' in respect of the bona fide annual value of property consisting of any buildings or lar
    Chunk 4: Summary - The Taxman's Act of 1939 amended the tax treatment of interest paid on the security of a property acquired, constructed, repaired, renewed or reconstructed with borrow
    Chunk 6: Summary - The Court of Appeal has ruled that the principle that the outgoings were not part of the assessee's income at all should be extended to cases where obligatory payments
    The Court also held that, in view of the opening words of the new sub clause in the Act, the expression "capital charge" also cannot have reference to a charge on the property, and we th
    Chunk 7: Summary - The sub clause provides for a deduction of annual sums so charged, such sums not being capital sums, the limiting words being intended to exclude cases where capital r
    Kania J., however, took a different view and observed that he was not prepared to accept the sugges tion that a document which provides for a certain payment to be made monthly or annual
    Chunk 9: Summary - Section 143 of the City of Bombay Municipal Act, 1888, authorises the levy of a general tax on all buildings and lands in the city.
    The primary responsibility to pay this property tax is on the lessor (vide section 146 of the Act).<n>In order to assess the tax provision has been made for the determination of the annu
    Section 24 (2) (b) of the Bombay Finance Act, 1932, provides: "Property taxes due under this Act in respect of any building or land shall, subject to the prior payment of the land revenu
    Chunk 12: Summary - Section 9 of the Indian Income tax Act provides for the tax to be collected on the annual value of the land or the building and are annual taxes, although it may be
    The learned Attorney General, on the other hand, argued that although the taxes are assessed for the year the taxes in question fell clearly within the language of section 9 (1) (iv) of
    Chunk 13: Summary - The Bombay High Court has ruled that the tax liability on a property is an annual one and the property having been subjected to it is immediately attracted by the pr
    It is apparent from the whole tenor of the two Bombay Acts that the Taxes are in the nature of an annual levy on the property ' and are assessed on the annual value of the Property each
    Chunk 14: Summary - The words "property taxes due under this Act" mean property taxes for which a person is liable under the Act. Taxes payable during the year have been made a charge o
    The provisions of the Act facilities for the discharge of the liability do not in any way affect their true nature and character.<n>If the annual liability is not discharged in the manne
    Chunk 15: Summary - The Income Tax Court of Bombay has ruled that it is not necessary to consider what the exact meaning of the words 'land revenue' is.
    Chunk 16: Summary - The law as to them varies from province to province and they may not be necessarily a charge on property in all cases.<n>It is elementary that the primary duty of a
    The Bombay Municipal Act does not cover municipal taxes which are made to charge on the property under section 212 of the Act. Chunk 17: Summary - In the case of Moss Empires Ltd. vs In
    Reliance was placed on a decision of the High Court of Madras in Mamad Keyi vs Commissioner of Income tax, Madras (3) in which moneys paid as urban immoveable property tax under the Bom
    Chunk 19: Summary - The Allahabad High Court has ruled that the Income Tax Tribunal's decision in the case of Gappumal Kanhaiya Lal vs Commissioner of Incometax (the connected appeal) wa
```
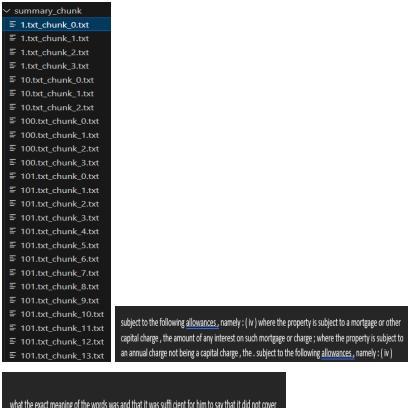
The stored Chunk_Judgement:

∨ judgment_chunk
≡ 1.txt_chunk_0.txt
≡ 1.txt_chunk_1.txt
≡ 1.txt_chunk_2.txt
≡ 1.txt_chunk_3.txt
≡ 10.txt_chunk_0.txt
≡ 10.txt_chunk_1.txt
≡ 10.txt_chunk_2.txt
≡ 100.txt_chunk_0.txt
≡ 100.txt_chunk_1.txt
≡ 100.txt_chunk_2.txt
≡ 100.txt_chunk_3.txt
≡ 101.txt_chunk_0.txt
≡ 101.txt_chunk_1.txt
≡ 101.txt_chunk_2.txt
≡ 101.txt_chunk_3.txt
≡ 101.txt_chunk_4.txt
≡ 101.txt_chunk_5.txt
≡ 101.txt_chunk_6.txt
≡ 101.txt_chunk_7.txt
≡ 101.txt_chunk_8.txt
≡ 101.txt_chunk_9.txt
≡ 101.txt_chunk_10.txt
≡ 101.txt_chunk_11.txt
≡ 101.txt_chunk_12.txt
≡ 101.txt_chunk_13.txt
≡ 101.txt_chunk_14.txt
≡ 101.txt_chunk_15.txt
≡ 101.txt_chunk_16.txt

Appeal No. LXVI of 1949. Appeal from the High Court of judicature, Bombay, in a reference under section 66 of the Indian Income tax Act, 1022 . K.M. Munshi (N. P. Nathvani, with him), for the appel lant . ' M.C. Setalvad, Attorney General for India (H. J. Umrigar, with him), for the respondent . 1950 May 26. The judgment of the Court was delivered by MEHR CHAND MAHAJAN J. This is an appeal against a judgment of the High Court of Judicature at Bombay in an income tax matter and it raises the question whether munici pal property tax and urban immoveable property tax payable under the relevant Bombay Acts are allowable deductions under section 9 (1) (iv) of the Indian Income tax Act . The assessee company is an investment company deriving its income from properties in the city of Bombay. For the assessment year 1940 41 the net income of the assessee under the head `` property '' was computed by the Income tax Officer in the sum of Rs. 6,21,764 after deducting from gross rents certain payments. The company had paid during the relevant year Rs. 1,22,675 as municipal property tax and Rs. 32,760 as urban property tax. Deduction of these two sums was claimed under the provisions of section 9 the Act. Out of the first item a deduction in the sum of Rs. 48,572 was allowed on the ground that this item represented tenants ' burdens paid by the assessee, otherwise the claim was disal lowed . The, appeals of the assessee to the Appellate As sistant Commissioner and to the Income tax Appellate Tribu nal were unsuccessful . The Tribunal, however, agreed to refer two questions of law to the High Court of Judicature at Bombay, namely, (1) Whether the municipal taxes paid by the applicant company are an allowable deduction under 555 the provisions of section 9 (1) (iv) of the Indian Income tax Act ; (2) Whether the urban immoveable property taxes paid by the applicant company are an allowable deduction under section 9 (1) (iv) or under section 9 (1) (v) of the Indian Income tax Act . A supplementary reference was made covering a third question which was not raised before us and it is not there fore necessary to refer to it. The High Court answered all the three questions in the negative and hence this appeal. The question for our determination is whether the munic ipal property tax and urban immoveable property tax can be deducted as an allowance under clause (iv) of sub section (1) of section 9 of the Act . The decision of the point depends firstly on the construction of the language employed in sub clause (iv) of sub section (1) of section 9 of the Act, and secondly, on a finding as to the true nature and character of the liability of the owner under the relevant Bombay Acts for the payment of these taxes . Section 9 along with the relevant clause runs thus: (1) The tax shall be payable by an assessee under the head ' income from property ' in respect of the bona fide annual value of property consisting of any buildings or lands appurtenant thereto of Which he is the owner , .. subject to the following allowances, namely : (iv) where the property is subject to a mortgage or other capital charge , the amount of any interest on such mortgage or charge ; where the property is subject to an annual charge not being a capital charge , the . amount of such charge ; where the property is subject to a ground rent , the amount of such ground rent ; and , where the property has been acquired , constructed , repaired , renewed or recon structed with borrowed capital , the amount of any interest payable on such capital ; . . . `` It will be seen that clause (iv) consists of four sub clauses corresponding to the four deductions allowed 556 under the clause . Before the amending Act of 1939, clause (iv) contained only the first , third and fourth sub clauses . Under the first sub clause interest is deductible whether the amount borrowed on the security of the property was spent on the property or not.. There is no question of any capital or other expenditure on the property. The expression `` capital charge '' in the sub clause can not connote a charge on the capital, that is , the property assessed . That would be a redundancy as the opening words themselves clearly indicate that the charge is on the property. We are therefore of opinion that capital charge here could only mean a charge created for a capital sum, i.e. , a charge to secure the discharge of a liability of a capital nature . In 1933 the Privy

and so should not be included in his assessment. Though the decision proceeded on the principle that the outgoings were not part of the assessee 's income at all , the framers of the amending Act of 1939 wanted , apparently , to extend the principle , so far as the assessment of property was concerned , even to cases where obligatory payments had to be made out of the assessee 's income from the property charged with such payments , and the second sub clause , namely , `` where the property is subject to an annual charge not being a capital charge , the amount of such charge '' was added . It is this sub clause which the appellant invokes in support of its claim to deduction of the municipal and urban, property taxes in the present case . In view of the opening words of the newly added sub clause, the expression `` capital charge '' also used therein can not have reference to a charge on the property , and we think it must (1) I.L.R . 60 cal. 557 be understood in the same sense as in sub clause (1) ; that is to say , the first sub clause having provided for deduc tion of interest where a capital sum is charged on the property , this sub clause provides for a deduction of annual sums so charged , such sums not being capital sums , the limiting words being intended to exclude cases where capital raised on the security of the property is made repayable in instalments . In Commissioner of Income tax, Bombay vs Mahomedbhoy Rowji (1), a Bench of the Bombay High Court considered the meaning of these words . As regards `` annual charge, " Beau mont C.J . observed as follows : `` The words , I think , would cover a charge to secure an annual liability . " Kania J., as he then was , said as follows : `` I do not see how a charge can be annual unless it means a charge in respect of a payment to be made annually . " This construction of the words has been followed in the judgment under appeal. In Gappumal Kanhaiya Lal vs Commissioner of Income tax (2) (the connected appeal before us) , the Bench of the Allahabad High Court agreed with the construction placed on these words in the Bombay case , i.e. , the words `` annual charge '' mean a charge to secure an annual liability . It is therefore clear that there is no conflict of judicial deci sions as to the meaning of the phrase `` annual charge '' occur ring in section 3 (1) (iv) and the meaning given is the natural meaning of these words . As to the phrase `` capital charge ", Beaumont C.J . in the case above referred to took the view that the words mean a charge on capital. Kania J., however , took a different view and observed that he was not prepared to accept the sugges tion that a document which provides for a certain payment to be made monthly or annually and charged on immoveable property or the estate of an individual becomes a capital charge . In the Allahabad judgment under appeal these (1) I.L.R . (2) I.L.R . 1944 All. 558 words were considered as not meaning a charge on capital. It was said that if an annual charge means a charge to secure the discharge of an annual liability, then , capital charge means a charge to secure the discharge of a liability of a capital nature . We think this construction is a natu ral construction of the section and is right. The determination of the point whether the taxes in dispute fall within the ambit of the phrase `` annual charge not being a capital charge '' depends on the provisions of the statutes under which they are levied. Section 143 of the City of Bombay Municipal Act, 1888 , authorises the levy of a general tax on all buildings and lands in the city . The primary responsibility to pay this property tax is on the lessor (vide section 146 of the Act) . In order to assess the tax provision has been made for the determination of the annual rateable value of the building in section 154. Section 156 provides for the maintenance of an assessment book in which entries have to be made every official year of all buildings in the city, their rateable value , the names of persons primarily liable for payment of the property tax on such buildings and of the amount for which each building has been assessed . Section 167 lays down that the assess ment book need not be prepared every official year but public notices shall be given in accordance with sections 160 to 162 every year and the provisions o+ the said sec tions and of sections 163 and 167 shall be applicable each year. These sections lay down a procedure for hearing objections and complaints against entries in the assessment book. From these

Chunk Summary:

summary_chunk
- 1.txt_chunk_0.txt
- 1.txt_chunk_1.txt
- 1.txt_chunk_2.txt
- 1.txt_chunk_3.txt
- 10.txt_chunk_0.txt
- 10.txt_chunk_1.txt
- 10.txt_chunk_2.txt
- 100.txt_chunk_0.txt
- 100.txt_chunk_1.txt
- 100.txt_chunk_2.txt
- 100.txt_chunk_3.txt
- 101.txt_chunk_0.txt
- 101.txt_chunk_1.txt
- 101.txt_chunk_2.txt
- 101.txt_chunk_3.txt
- 101.txt_chunk_4.txt
- 101.txt_chunk_5.txt
- 101.txt_chunk_6.txt
- 101.txt_chunk_7.txt
- 101.txt_chunk_8.txt
- 101.txt_chunk_9.txt
- 101.txt_chunk_10.txt
- 101.txt_chunk_11.txt
- 101.txt_chunk_12.txt
- 101.txt_chunk_13.txt

subject to the following allowances , namely : (iv) where the property is subject to a mortgage or other capital charge , the amount of any interest on such mortgage or charge ; where the property is subject to an annual charge not being a capital charge , the . subject to the following allowances , namely : (iv)

what the exact meaning of the words was and that it was sufficient for him to say that it did not cover municipal taxes which are made a charge on the property under section 212 of the Bombay Municipal

2. Fine-tuning our existing model on this updated dataset.

As we take input in chunks, our model will predict the most essential statements in each chunk and then combine all these outputs to give a final summary. This method will also reduce the summary length compared to the old output. As some chunks may not have any important statements.

Finally, we are doing this:

1. **Preparation**: Importing necessary libraries.
   Reading judgment files.
2. **Chunking and Summarizing**:
   Breaking down judgments into smaller chunks.
   Generating summaries for each chunk using Mean Cosine similarity between the generated chunk and actual summary we have.
3. **Fine-Tuning** LegalPegasus on the new data set we generated.
   Train LegalPegasus with generated summaries, which is chunk wise summarize.
4. **Evaluation**:
   Assessing LegalPegasus performance on test data.
5. **Analysis**:

Analyzing results for further refinement

Evaluation

Our evaluation approach includes
- For evaluation, we have used 3 variants of ROUGE ,that is (ROUGE-1,2 and L) scores as it is best for Natural languages.
- We have used Divide and conquer techniques to deal with large datasets.
- Compared our accuracy with the accuracy given by the same model in the below research paper.
- Fine Tuning our model on a huge dataset, for better performance. Even though we were able to generate a good amount of chunk data and their summaries, we still need to increase the training dataset.

Conclusion

Legal Pegasus was already fine-tuned on legal documents. We improved the input size and used Mean cosine similarity function for generating our own dataset and their summaries. We will train our model using these chunks and summary pairs . We hope to achieve a F1 score of atleast 0.5 and then again proceed further.

The research paper we are following has an F1 score of around 0.864. So, we will try to achieve that level of F1 score.

References :

https://arxiv.org/abs/2004.06190

https://aclanthology.org/2022.aacl-main.77.pdf