Text Classification: Salient and Non-Salient Sentences

Uday Bindal
LTRC, IIIT Hyderabad

***Abstract*—The project deploys linguistically motivated features to classify sentences into salient and non-salient categories using a logistic regression model and infers lexical and syntactic properties that help distinguish the two genres. A movie script typically comprises a large number of scenes; however, only a fraction of these scenes are salient, i.e., important for understanding the overall narrative. This project also proposes a model that uses transformers to classify sentences and then use the salient sentences to generate an abstractive summary of the movie.**
The motivation for this research stems from two key papers in the field. The first paper [1] focuses on classifying text into fiction and non-fiction categories using various linguistic features. The second paper [4] proposes a method for generating summaries of movies, which inspired the use of transformers for sentence classification and subsequent summary generation in this study.

I. Introduction

Abstractive summarization is the process of reducing an information source to its most important content by generating a coherent summary. Several methods have previously relied on content selection for summarization to reduce the input size by performing content selection implicitly using neural network attention but this method uses the most salient scenes of the movie script to generate the summary. In general, salient scenes are longer and have more lexical diversity as compared to the non-salient scenes.
One of the main objectives of this project is to identify the most relevant features that provide deeper insights about the properties that can be used to distinguish these two categories. Three categories of features that have been deployed in logistic Regression are Raw Features, Lexical Features and POS Features.
For the movie summarization part, MENSA dataset was used that includes human annotation of salient scenes in movie scripts. A supervised saliency classification model classifies sentences given a movie script which are then used to generate an abstractive summary. Apart from the transformer model, the classification model also contains a scene-sentence graph to further improve the sentence classification accuracy.

II. Linguistic Features

For saliency classification of sentences, we deployed the following four distinct categories of features in the project: 1. Raw text 2. Lexical 3. POS Tag 4.Sentiment.

A. Raw Text Features

These are the most basic features. We incorporated **Average sentence length** as one of the features for logistic regression.
We found that salient scenes tend to be significantly longer than non-salient scenes, with an average length of 302.05 tokens compared to 120.50 tokens for non-salient scenes. This suggests that salient scenes contain more detailed information.

B. Lexical Features

The following lexical features have been deployed:

- **Lexical Diversity:** establishes the statistical relationship between the type and tokens of the text. The most common approach for measuring the diversity of characters or words is to use the ratio of unique tokens divided by the total number of tokens in a sentence (Type-Token Ratio).
  Non-salient scenes show higher lexical diversity (0.72) compared to salient scenes (0.60). It could be explained by the fact that salient scenes, being longer, may need to repeat key terms and concepts for emphasis or clarity

	precision	recall	f1-score	support
0.0	0.75	0.92	0.82	1091
1.0	0.69	0.37	0.48	541
accuracy			0.74	1632
macro avg	0.72	0.64	0.65	1632
weighted avg	0.73	0.74	0.71	1632

Figure 1. Classification Report

- **Lexical Density:** calculated by taking the ratio of content words (words that are tagged as noun, verb, adjective adverb) to function words (all part of speech tagged words except those of content words). Salient scenes often contain more important or relevant information. This information is usually conveyed through content words (nouns, verbs, adjectives, and adverbs), which increases lexical density.

C. POS Tag Features

- **POS Tag Average:** was calculated to understand their distribution and significance in salient and non-salient scenes. Salient scenes use slightly more personal pronouns (0.07 vs 0.06), indicating more focus on character actions or dialogue.
- **Average Named Entities:** A named entity is a real-world object, such as a person, location, organization, product, etc., that can be denoted with a proper name.

  Salient scenes contain more than twice as many named entities on average (22.29) compared to non-salient scenes (10.36). This indicates that salient scenes are more likely to mention specific people, places, or organizations, which adds to their informational content and potential importance.

D. Sentiment

Sentiment analysis is the process of analyzing text to determine if the emotional tone of the message is positive, negative, or neutral. Average sentiment was computed for both the categories.

Salient scenes have a slightly more positive sentiment (0.07) compared to non-salient scenes (0.02). While both scores are close to neutral, this small difference suggests that salient scenes might contain more emotionally engaging or positively framed content.

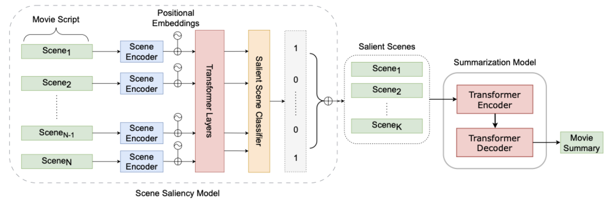

Figure 2. The architecture of the scene saliency detection and summarization models

III. Experiment and Results

All these features were used to train a logistic regression model. We evaluate model performance using classification report (Figure 1). We selected the optimal features by applying recursive feature elimination with cross-validation (RFECV) on the 4 feature sets described. The 2 selected optimal features are *Lexical Diversity and Sentence Length.*

IV. Scene Saliency Classification Model

We train a neural network- based classification model to predict scene saliency. We formulate this task as a sequence labeling task where the model takes a sequence of scenes M = S1, S2, ..., SN as input and predicts a sequence of binary labels Y = y1 , y2 , ..., yN denoting whether a scene is salient.

The model consists of two components, as shown in Figure 2. The first component is a scene encoder which computes scene representations by concatenating the sentences in the scene and encodes them using a pre-trained language model. To preserve the sequence of the scenes, we add positional encodings to scene representations obtained from the first component.

Next, to learn contextual scene representation across the whole movie, we further encode the scene embeddings generated by the scene encoder using a transformer block (L layers stacked), with unmasked self-attention initialized with random weights .

We also used a graph-based algorithm. we constructed a movie script graph such that nodes in the graph correspond to the scenes in the movie M . The edge eij between any two scene nodes Si and Sj represents their similarity, with the edge weight being the similarity score. The centrality of a node Si measures the importance of that node and is computed as follows:

$$\text{centrality}(S_i) = \lambda_1 \sum_j e_{ij} + \lambda_2 \sum_j e_{ij} \qquad (1)$$

where 1 and 2 are weights for forward-looking (edges to following scene nodes) and backward- looking (edges to preceding scene nodes) and sum to one. We then select top-K nodes as the salient scenes based on their centrality score.

V. SUMMARIZATION MDEL

Given a movie with a set of salient scenes M = S1, S2, ..., SK , the goal is to generate a target summary S = s1, s2, ..., sm. We use a Longformer Encoder- Decoder (LED) architecture. To handle long input sequences, LED uses efficient local attention with global attention for the encoder. The decoder then uses the full self-attention to the encoded tokens and to previously decoded loca- tions to generate the summary.

VI. QA-BASED EVALUATION

To evaluate the performance of our model, we used QAEval, a question-answering-based evaluation that generates question-answer pairs using the reference summaries. It then uses the model-generated summaries to answer these questions, thereby measuring information overlap. Metrics like ROUGE (lexically based) and BERTScore fail to compare content-based factual consistency.

REFERENCES

[1] Arman Kazmi, Sidharth Ranjan, Arpit Sharma, and Rajakrishnan Rajkumar. Linguistically motivated features for classifying shorter text into fiction and non-fiction genre. In Nicoletta Calzolari, Chu-Ren Huang, Hansaem Kim, James Pustejovsky, Leo Wanner, Key-Sun Choi, Pum-Mo Ryu, Hsin-Hsi Chen, Lucia Donatelli, Heng Ji, Sadao Kurohashi, Patrizia Paggio, Nianwen Xue, Seokhwan Kim, Younggyun Hahm, Zhong He, Tony Kyungil Lee, Enrico Santus, Francis Bond, and Seung-Hoon Na, editors, *Proceedings of the 29th International Conference on Computational Linguistics*, pages 922–937, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.

[2] Paramita Mirza, Mostafa Abouhamra, and Gerhard Weikum. AligNarr: Aligning narratives on movies. In Chengqing Zong, Fei Xia, Wenjie Li, and Roberto Navigli, editors, *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 427–433, Online, August 2021. Association for Computational Linguistics.

[3] Pinelopi Papalampidi, Frank Keller, and Mirella Lapata. Movie plot analysis via turning point identification. In Kentaro Inui, Jing Jiang, Vincent Ng, and Xiaojun Wan, editors, *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1707–1717, Hong Kong, China, November 2019. Association for Computational Linguistics.

[4] Rohit Saxena and Frank Keller. Select and summarize: Scene saliency for movie script summarization. In Kevin Duh, Helena Gomez, and Steven Bethard, editors, *Findings of the Association for Computational Linguistics: NAACL 2024*, pages 3439–3455, Mexico City, Mexico, June 2024. Association for Computational Linguistics.