An Integrated Visio-Textual Approach to Movie Genre Classification

Sartaj Baveja
Karthik Srinivasan

Our Data

- Web-scraped IMDB plot summaries and posters
    - dataset contains 78,008 records with 27 unique genres

- Restrict predictions to top 10 genres (sparsity of data for the rest of the labels)

- 60% of the movies are labeled with Drama, Comedy or Romance

- Plot summaries (in english) from over 154 unique language movies released between 1906-2020

Modeling Methodology

Naive Labelling

Assign most common labels

Posters (CNN)

Pretrained ResNet50 model*

Oversampling

*Resnet_50_v2

BERT

Pretrained BERT model*

BERT Finetuning

Oversampling

*bert_en_uncased_L-12_H-768_A-12

Integrated Viso-textual

Pretrained ResNet50 model

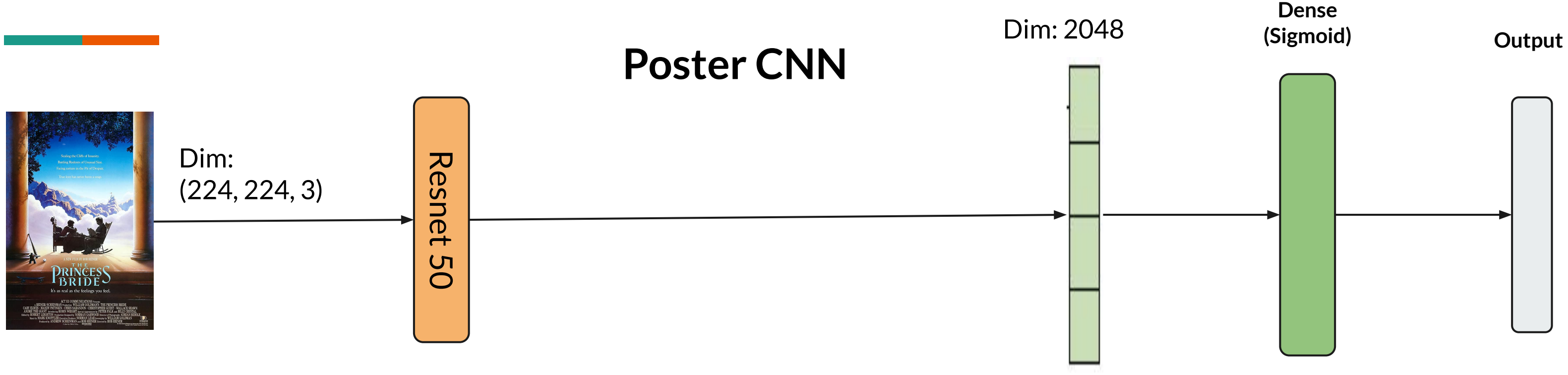
Pretrained BERT model

BERT Finetuning

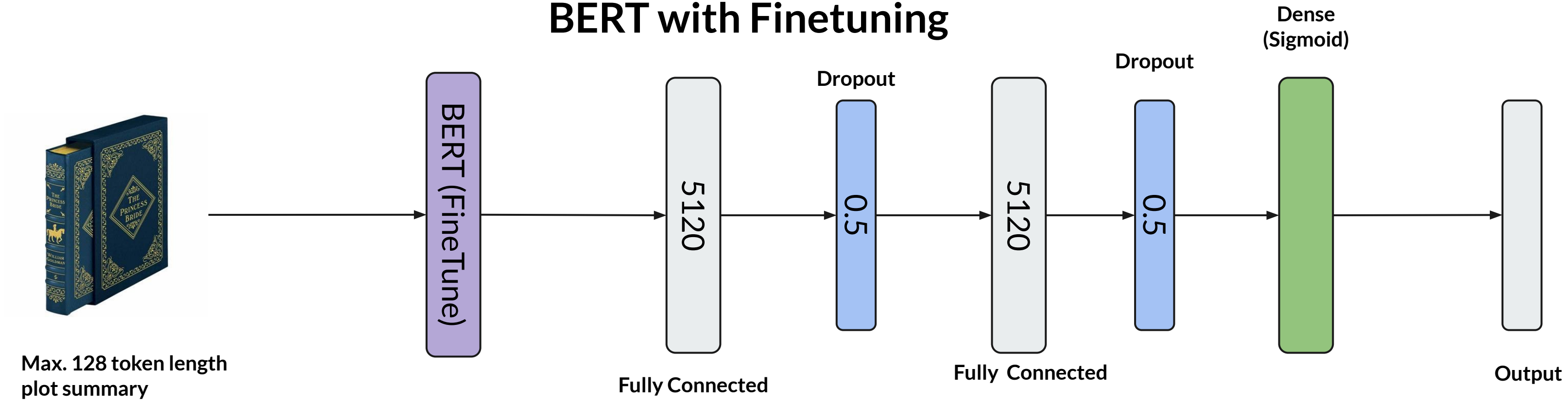
Oversampling

Model Architectures

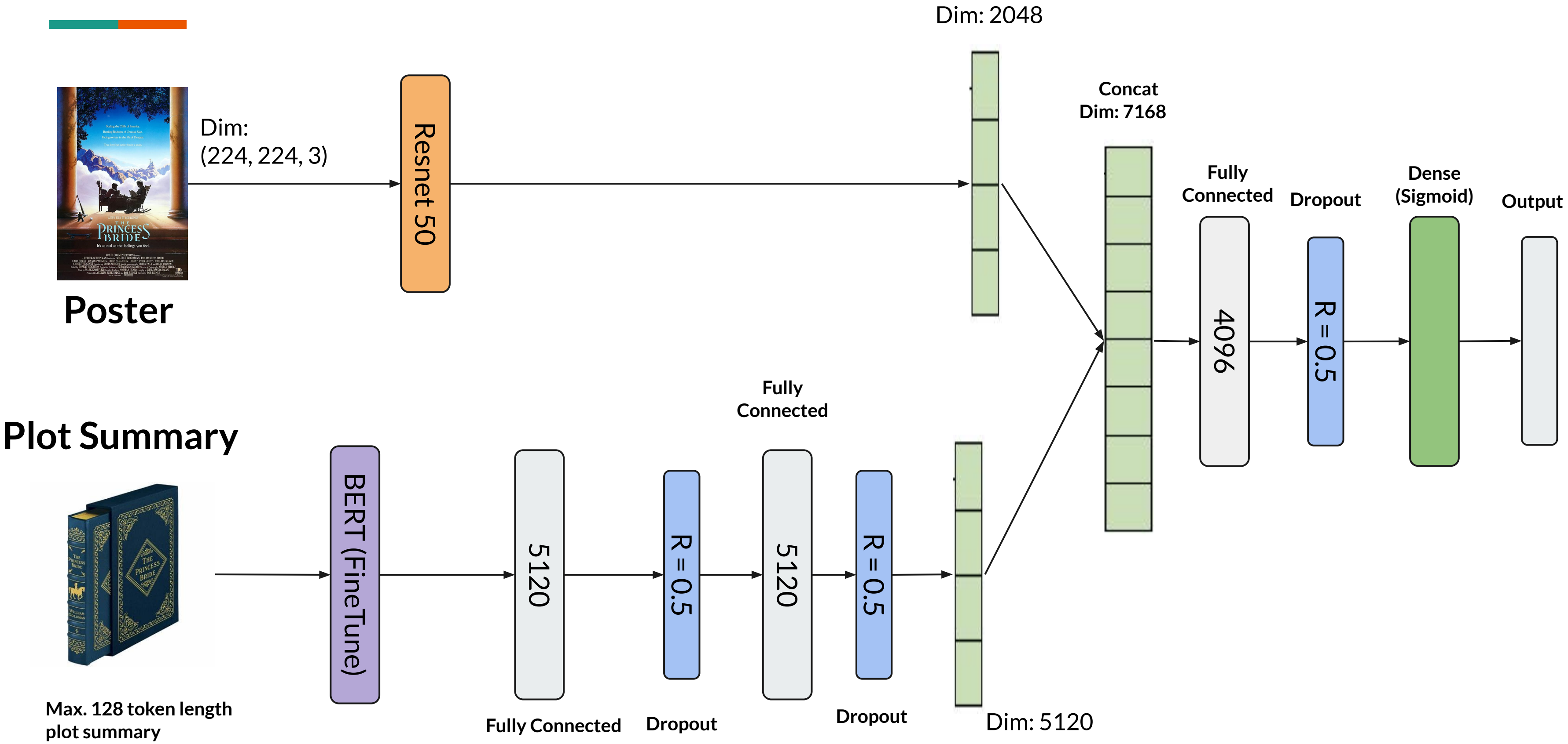
Poster CNN



BERT with Finetuning



Model Architectures



Accuracy Metric and Loss Function

Accuracy Metric

Weighted F1 Score accounts for the class imbalance in the hold-out validation dataset

$$Macro\ F_{1,i} = \frac{2 \cdot TP_i}{2 \cdot TP_i + FN_i + FP_i + \epsilon}$$

$$Wtd\ F_1 = \frac{\sum_{i=1}^{N} Freq_i \cdot Macro\ F_1, i}{\sum_{i=1}^{N} Freq_i}$$

Loss Function

Macro Double Soft F1 directly maximizes the F1 scores (unlike binary cross entropy)

$$\begin{aligned} Loss_i = 1 & \\ & - 0.5 \left(\frac{2 \cdot TP_i}{2 \cdot TP_i + FN_i + FP_i + \epsilon} \right) \\ & + 0.5 \left(\frac{2 \cdot TN_i}{2 \cdot TN_i + FN_i + FP_i + \epsilon} \right) \end{aligned}$$

$$Total\ Loss = \frac{1}{N} \sum_{i=1,N} Loss_i$$

Model Performance

- Oversampling used to decrease label imbalance
  - Train sample size increased from 46K to 76K
- 2% improvement over baseline BERT + FineTuning
- Model only predicts 14 unique combinations of labels while >400 exists, likely due to data sparsity
- Sources of error may include labeling errors, non-standardized plot summaries
- Model finds it hard to distinguish Comedy from Drama
- Movie release year has a small impact on prediction accuracy (across all languages)

Label	**Training** Oversampling	**Training** No Oversampling	**Validation**
Drama	0.49	0.55	0.56
Comedy	0.14	0.34	0.34
Romance	0.27	0.20	0.20
Thriller	0.19	0.20	0.19
Action	0.16	0.15	0.15
Crime	0.15	0.13	0.13
Horror	0.30	0.12	0.12
Adventure	0.24	0.09	0.09
Mystery	0.16	0.07	0.08
Sci-Fi	0.20	0.06	0.06

Table 1: Frequency of labels in train and validation datasets.

Model	**No Oversampling**	**Oversampling**
Naive Classifier	0.209	0.209
Poster ANN	0.232	0.210
BERT	0.192	0.214
BERT + FineTuning	0.433	0.431
Integrated model	0.429	**0.453**

Table 2: Weighted F_1 scores.

Future Work

- Increase dataset specifically to address imbalance
- Evaluate other NLP models such as Universal Sentence Encoder, ELMo etc.
- Fine tune poster image ResNet models, evaluate object segmentation for better performance

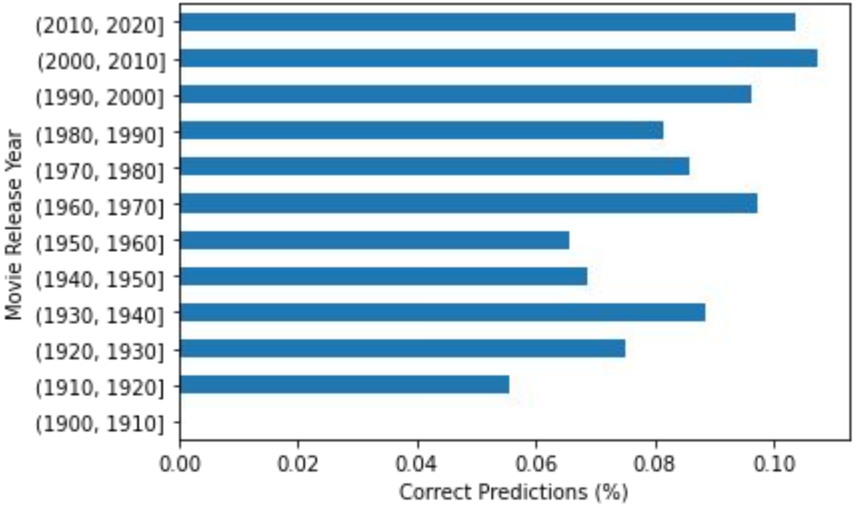
Appendix

Weighted F1 Score Analysis

Label	BERT+FineTuning	Integrated
Drama	0.37	0.39
Comedy	0.17	0.16
Romance	0.08	0.08
Thriller	0.06	0.07
Action	0.04	0.04
Crime	0.02	0.03
Horror	0.04	0.04
Adventure	0.02	0.02
Mystery	0.02	0.01
Sci-Fi	0.02	0.02

Table 2: Weighted F1 scores for each label in the validation dataset.

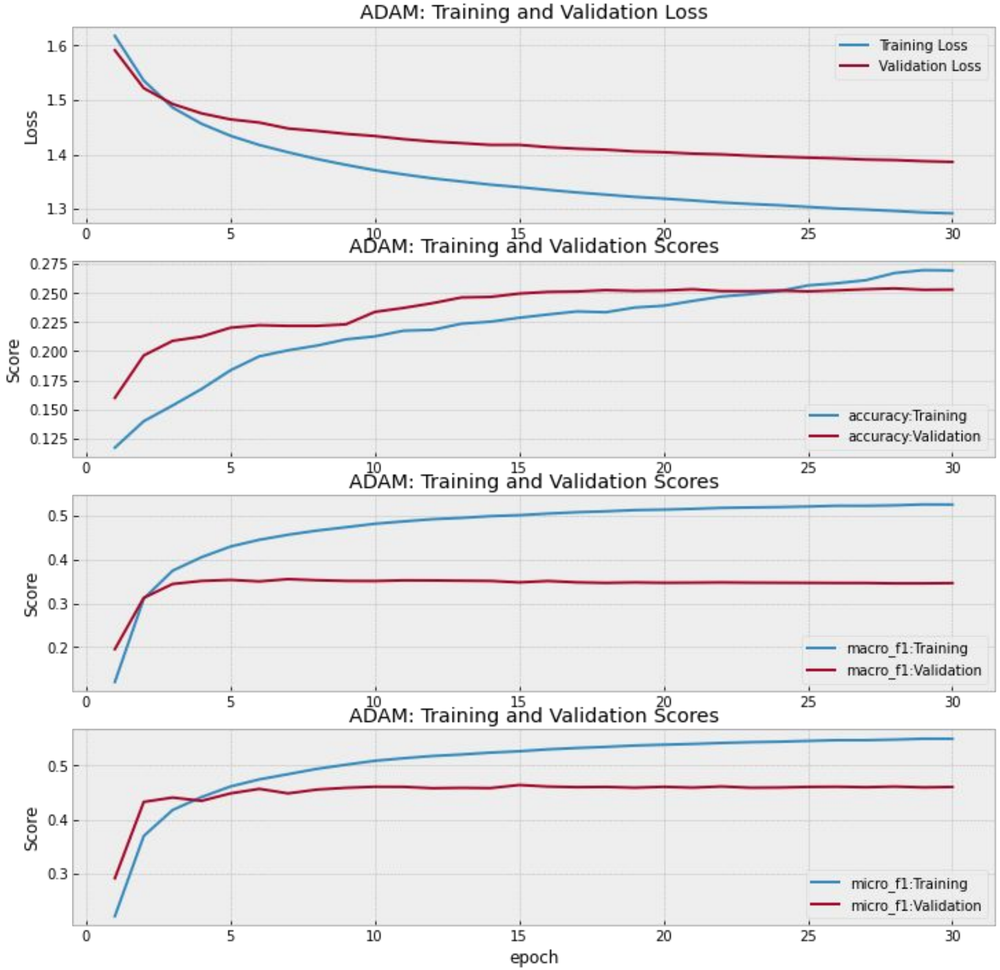
Effect of Release Year on Predictions



Loss/Accuracy Curves for Integrated Model

- Loss curves saturate after first few epochs

- F1-scores saturate in as few as 5 epochs (the trends are same upon repeated runs)



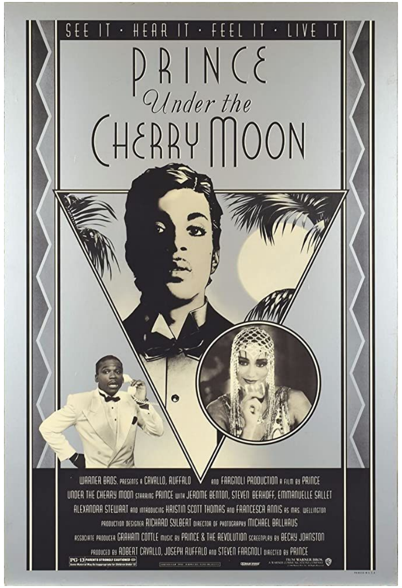
A few examples...

'Jimmy Wayne Collins finds himself adrift in Memphis, Tennessee. Forced to return home to the piney woods of Southeast Texas, Jimmy will face his imprisoned brother, his dying father and the demons he left behind.'

Actual : Drama
Predicted (Integrated) : Drama
Predicted (BERT) : None

"In a small town in California, the mysterious Jeff Bailey owns a small gas station and is in love with the local Ann. When a stranger just arrived in town meets him, Jeff is ordered to travel to meet the powerful criminal Whit Sterling. Before traveling, Jeff calls Ann and tells her the story of his life, when he was a private eyes hired by Whit for US$ 5,000.00 to find his former mistress Kathie that had shot Whit and stolen US$ 40,000.00. The competent Jeff finds Kathie in Acapulco….

Actual : Drama, Thriller. Crime, Romance
Predicted (Integrated) : None
Predicted (BERT) : Thriller, Action, Crime, Horror, Adventure, Mystery, Sci-Fi



'Two cousins from Miami, Florida are in the Mediterranean, enjoying life by scamming money off of rich women. One day, they read about a young woman, Mary Sharon (Dame Kristin Scott Thomas), set to inherit fifty million dollars from her father. At first, Tricky (Jerome Benton) has Christopher Tracy (Prince) talked into romancing her for her money, but as he gets to know her, Christopher falls in love with her. This love comes between the cousins, and Tricky tells all about the plan.'

Actual : Drama, Comedy, Romance
Predicted (Integrated) : Drama, Comedy, Romance
Predicted (BERT) : Drama, Comedy, Romance

An elderly man reads the book ""The Princess Bride"" to his sick and thus currently bedridden adolescent grandson, the reading of the book which has been passed down within the family for generations. The grandson is sure he won't like the story, with a romance at its core, he preferring something with lots of action and ""no kissing"". But the grandson is powerless to stop his grandfather…

Actual : Adventure, Romance
Predicted (Integrated) : Drama
Predicted (BERT) : Drama