Using Neural Networks to Predict Violent Crime Based on Movie Releases

Skylar Furey, Jimmy Kruse, Sivani Pillutla

Introduction

There has long been speculation regarding the effects of media viewership on human behavior. Since the introduction of visual art, researchers and the general public have been questioning the influence it has on society. In particular, great concern has been raised over whether viewing simulated violence on screen could instigate violent behavior among its viewers. Multiple studies have been done on the long term effects of violent video games on aggression in players of all ages (1) (2), to mixed and often inconclusive results. Studies also show there is sometimes a decrease in nearby violent crime immediately following a showing of a violent movie at a theater, but noting that there is no data to show that reduction continues beyond that specific timeframe. (3) Knowing this, can we accurately predict instances of crime based on violent movie releases?

Related Works

With the advent of new technologies, tools for crime detection and prevention have become more innovative and sophisticated. Machine learning as a field has made advances in predicting crimes before they occur, fundamentally changing the way we view policing and crime prevention. A model built at Rutgers University used neural networks to predict criminal activity based on previous arrest and booking records. The model also used features from the environment, such as broken glass, disorder, and antisocial behavior to identify potential hotspots for crime. It even considered rehabilitation initiatives aimed at steering offenders away from further criminal behavior. The model reached an accuracy rate of 83.95%, showing how machine learning can be used to assist in predictive policing. (5)

Another study compared different approaches for crime forecasting, such as neural networks, machine learning, and computer vision techniques for crime forecasting. The study examined a few different approaches, such as the KNN algorithm, which had accuracy rates ranging from 67% to 87%, decision trees with accuracy rates of 60% to 87%, and Naive Bayes models, which performed similarly with accuracy rates between 66% and 87%. However, the study also noted a major issue: as datasets grow more complex and encompass diverse features, the accuracy of predictions tends to decline. This study shows how when machine learning is applied in real life scenarios, models need to be constantly refined and reconfigured to ensure effectiveness and accuracy. (6)

Machine learning is not only changing the field of policing, but is also used in monitoring online behaviors. An article from the University of Chinese Academy of Sciences used a Multilevel Attention Residual Neural Network (MARN) for detecting online rumors and tracking the spread of misinformation on social media platforms like Weibo. The MARN, combining both text and visual data in the model, greatly enhanced its ability to classify and detect rumors and misinformation with high accuracy. This study showcases how machine learning can be used dynamically in digital spaces, a place with unique challenges such as rapid spread of information, and massie societal repercussions. (7)

The movie industry also has widely accepted the use of machine learning, predicting ratings, revenue, reception, and even recommendations with great accuracy. Many models and studies have been done exploring data from movie releases. Notably, a study explored utilizing a Support Vector Machine (SVM) to predict the success of a movie on a scale of 1-5 based on historical data with high accuracy rates between 83.44% to 89.27%. The model identified budget, IMDb rating, and number of screens showing the film as the strongest predictors of box office success. This finding shows how machine learning can be used in the entertainment industry to increase profits, as well as viewer satisfaction. (8)

Machine learning shows promising advancements across various fields, including criminology, social monitoring, and entertainment. While many studies show there may be a significant correlation between viewership of simulated violence on screen to behaviors in real life, significant gaps remain in research predicting if viewing violent movies influence violent crime. By bridging these gaps, we aim to develop a predictive model that can accurately examine the relationship between violent movie releases and subsequent crime trends, offering insights into this understudied area.

Methodology

1. Data Aggregation and Feature Extraction

Our dataset comprises:

1. Movie Metadata: We extract attributes such as runtime, revenue, and release date. Let $x_{metadata}$ represent this feature vector.

2. Temporal Feature Extraction: Using the release date, we construct a feature x_{Month} that represents the month of release. This monthly alignment allows us to match each movie with the relevant crime data in each state.

3. Genre Selection: Each movie is associated with multiple genres. We select only the primary genre g_1, arguing that the first genre in the list most accurately represents the movie's dominant theme. Let $x_{genre}= g_1$, a categorical variable that we later encode using one-hot for model input.

4. Movie Ratings: We extract IMDb, Rotten Tomatoes, and Metacritic ratings, denoted as $r_{imdbRating}$, r_{RT_Rating}, $r_{Metacritic}$, from the dictionary attribute Ratings. Ratings offer insight into public sentiment, with lower ratings potentially correlating with societal discontent during the release period.

2. Image Processing and Feature Extraction from Posters

We hypothesize that visual elements of movie posters (color scheme, brightness, dominant colors like red for intensity) may correlate with the movie's tone and, indirectly, societal effects.

1. Poster Data: For each movie, we retrieve the poster image P and preprocess it using a pre-trained residual neural network (ResNet) based on the ImageNet model. The ResNet extracts high-level features f_{poster}=ResNet(P), where f_{poster} is a feature vector summarizing the poster content.

2. Feature Integration: These extracted features are appended to the dataset as a new attribute vector, $x_{Poster\_Features} = f_{poster}$.

3. Data Preprocessing and Scaling

To prepare the data for modeling, we concatenate all feature vectors into a single vector:

$$x = [x_{metadata}, x_{month}, r_{imdbRating}, r_{RT\_Rating}, r_{Metacritic}, x_{Poster\_Features}]$$

Each numeric feature is scaled to zero mean and unit variance using StandardScalar. This scaling ensures that features with large variances do not disproportionately influence the model.

To normalize the outputs due to the high variance from state to state, the model was also ran with the crimes reported in each state divided by the state's population in 2020.

Another strategy we used to combat the high variance was to predict the totals for the entire country instead of the individual states.

4. Model Architecture: Convolutional Neural Network (CNN)

Our primary model is a CNN designed to integrate visual and non-visual features.

1. Feature Input Layers: The model initially processes non-image features $\hat{x}_{metadata}$, $\hat{x}_{month}$, $\hat{r}_{imdbRating}$, $\hat{r}_{RT\_Rating}$, $\hat{r}_{Metacritic}$ through dense layers to capture interactions among numeric, categorical, and rating features.

2. Feature Fusion: After several dense layers, we concatenate the outputs from the dense layers and ResNet output to form a joint feature representation. Let $c_{fusion}=[c_{dense}, x_{Poster\_Features}]$, where c_{dense} is the outputs of the dense layers.

3. Output Layer: The final layer has a linear activation to predict violent crimes reported y_{pred} for each state during the release month. Our loss function is mean absolute error (MAE):

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|$$

where $\hat{y}_i$ is the model's prediction and y_i is the actual violent crime count.

5. Evaluation and Hypothesis Testing

Our model's performance will be evaluated against a baseline created by predicting the mean from the training set for each record in the test set. We also used a grid search with cross validation to determine the best hyperparameters for our model. s.

Results

The optimal hyperparameters were determined using a Grid Search with 5-fold cross-validation. The search evaluated combinations of the following: three hidden layer configurations [(32), (32, 32), (32, 32, 32)], two activation functions (ReLU and tanh), three optimization solvers (adam, lbfgs, and sgd), three regularization strengths ([0.0001, 0.001, 0.01]), three learning rate schedules (constant, invscaling, and adaptive), and four initial learning rates ([0.0001, 0.001, 0.01, 0.1]).

The best-performing configuration utilized the ReLU activation function and the ADAM optimizer. ReLU facilitated efficient gradient-based updates, while ADAM's dynamic learning rate adjustments allowed the model to fine-tune weights effectively throughout training. The adaptive learning rate schedule further optimized weight updates by dynamically adjusting the learning rate as the model's performance improved or worsened over epochs.

Despite identifying the best hyperparameters, the model's performance remained suboptimal. The best model achieved a mean absolute error (MAE) of 3,921, which is only 13.6% better than the baseline of predicting the training set mean for every record. When applied to predicting country-wide crime reports, the model underperformed, yielding results worse than the baseline.

Analysis of the residual distribution (Figure 1) shows that most predicted crime report values deviate by less than 2,500 from their actual values. However, the distribution is heavily skewed to the right, driven by states like Texas and California, which have larger populations and significantly more reported crimes. The model did not sufficiently account for the relationship between state population size and the volume of reported crimes, indicating that the state variable's influence on predictions was inadequately captured.

To address this, crime rates were normalized by each state's population, and the residual distribution was re-examined (Figure 2). While normalization reduced the skew and produced a distribution closer to normal, a slight right skew persisted. Additionally, the normalized model's performance declined, achieving an MAE only 2.6% better than the baseline, further highlighting its limited predictive power.

Conclusion

Our project set out to predict violent crime rates based on movie releases, combining tabular data with features extracted from movie posters using a pre-trained ResNet model. Our original idea to predict movie revenue with this data was not novel enough so we went after something more ambitious. In retrospect, what we settled on was probably a little too ambitious. At least, too ambitious to do with the movie data as the sole dataset for our predictions. As a result of the difficulty of the problem, as well as our limited experience building machine learning models, we did not have much success in the goal of predicting crime rates by state. Despite the lack of success in this area, we did end up learning a lot. The project provided invaluable experience and opportunities as our first foray into machine learning.

One of the key lessons we learned was the importance of flexibility in the modeling process. Our architecture, which combined ResNet-derived image features with tabular data in a custom sequential model, was unnecessarily complicated and poorly suited for the task. It made our data path disjoint and poorly optimizable. A more sophisticated architecture might be to process tabular and image data separately through parallel layers in order to concatenate the two feature sets and pass them through further layers. The original architecture we had was simply our first try at making a working model, and as we progressed further into other aspects of the project we did not take the time to reevaluate and reconsider our previous design choices. This failure to reevaluate and simplify our approach reflected a lack of experience and underscores the value of continuous evaluation when working on machine learning projects. It is easy to get tunnel vision and lose sight of the big picture. This was a great learning experience for us.

In addition to what has been described above, we did learn a number of important things to consider and explore when working with neural networks. We saw early on that increasing the number of layers or nodes does not actually always improve the model and can often lead to worse performance. In general a simpler solution is preferable and it can make implementing and subsequent changes or improvements much easier. We gained practical experience in acquiring, cleaning, filtering, and processing data for use in machine learning. This includes using a variety of normalization and preprocessing techniques which are essential skills for producing any future machine learning projects.

Another major oversight was our failure to expand the dataset. Our narrow focus on movie-related data, while aligned with the original project goal, likely limited the model's ability to capture meaningful patterns. This was partly because we were too stuck on our initial plans and were reluctant to make major changes. We most likely would have been able to produce a more successful project overall if we were open to exploring additional input data. Because we were so focused on trying to do something novel and challenging we failed to consider such changes. Other data such as more specific dates as well as socioeconomic, demographic or other data should have been incorporated in order to build a more robust model.

One of the biggest challenges we faced was the difficulty of predicting real-world events with complex causes. Violent crime is influenced by an array of socio-economic, psychological, and situational factors. Attempting to explain its variance using our inputs, numerical movie-related. While the project ultimately did not yield a successful model, it offered insights into integrating diverse data types and the challenges of building machine learning systems for complex problems. The experience reinforced the importance of careful planning, iterative model development, and prioritizing meaningful data over adhering to a rigid framework.

Future expansion of this project would include redesigning the architecture of the model in a more sophisticated manner. This would require more research and help from others as well. If we want to continue to explore the way media and entertainment affect crime, then we would likely want to focus on a smaller area to start. From this we could find more specific data and expand our work from there. We could also include music, art, online media or other related things as additional input data. Alternatively, we could shift the target variable to something more related to movies than crime. While revenue was not creative enough, perhaps we could use movie data along with other data to predict the cultural impact/cultural relevance a film will have some time after. This would likely become a much harder task to find target data for though as well.

Despite the challenges we had, we did learn a variety of valuable things and have greatly expanded our abilities to tackle similar, or not so similar, projects in the future. The limitations we encountered highlight the importance of carefully incorporating diverse, high-quality datasets and more directly relevant features when tackling complex societal issues. Our future projects following this one will greatly benefit from the experience garnered here.

References:

1) https://pmc.ncbi.nlm.nih.gov/articles/PMC6790614/
2) http://gip.uniovi.es/docume/pro_vs/vv_ado.pdf
3) https://econweb.ucsd.edu/~gdahl/papers/movies-and-violence.pdf
4) https://dl.acm.org/doi/pdf/10.1145/3325112.3328221
5) https://dl.acm.org/doi/pdf/10.1145/3325112.3328221
6) https://link.springer.com/content/pdf/10.1186/s42492-021-00075-z.pdf
7) https://www.frontiersin.org/journals/physics/articles/10.3389/fphy.2021.711221/full
8) https://ieeexplore.ieee.org/document/8281839

Figures:

Figure 1: Residual distribution of model based on Crime Rate Report counts.

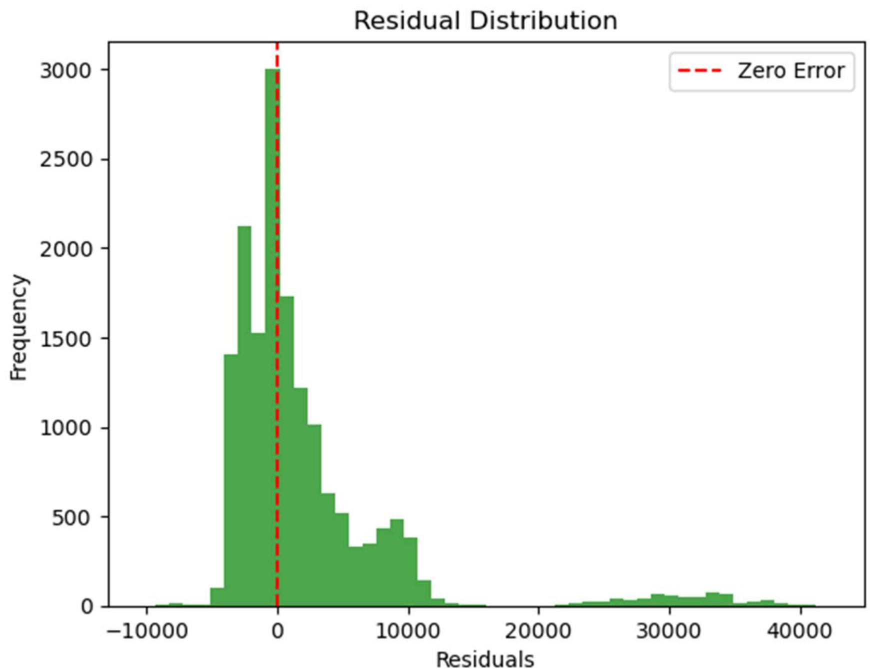

Figure 2: Residual distribution of model based on crime rate report proportions compared to state populations.