EE 541 – Computational Introduction to Deep Learning

Model Card for CineLit Recommender: Elevating Entertainment Discovery

Nissanth Neelakandan Abirami
University of Southern California
nissanth@usc.edu

Sam Devavaram Jebaraj
University of Southern California
devavara@usc.edu

Model Details:

- **Person or Organization Developing Model:** The CineLit recommendation system was developed collaboratively by a team of Graduate Students from University of Southern California .
- **Model Date:** December 7, 2023.
- **Model Type:** CineLit is a hybrid recommendation system that combines collaborative filtering, content-based filtering, and deep learning approach.
- **Features:** The recommendation system employs collaborative filtering, matrix factorization perspectives, sequential recommendation systems. Data from different CSVs are merged based on the 'id' column. Some columns are dropped, and missing values are handled.
- **Paper or Other Resource for More Information:** The literature review references previous research in university library book recommendation systems, emphasizing the effectiveness of combining content-based filtering with course syllabus information. Evaluation measures such as RSME and K-Fold Cross Validation are highlighted. The code also addresses the improvement of movie recommendations by combining content-based and collaborative-based filtering algorithms. Additionally, it references a unique unsupervised learning-based hybrid recommender system that incorporates collaborative filtering, a content-based approach.

- **Model Summary:**

```
________________________________________________________________
Layer (type)                     Output Shape         Param #
=================================================================
sequential_movie_model (Seque (None, 128)          0
_________________________________________________________________
sequential_user_model (Sequen (None, 128)          0
_________________________________________________________________
dense_1 (Dense)                  (None, 512)          66048
_________________________________________________________________
dropout_1 (Dropout)              (None, 512)          0
_________________________________________________________________
dense_2 (Dense)                  (None, 256)          131328
_________________________________________________________________
dropout_2 (Dropout)              (None, 256)          0
_________________________________________________________________
dense_3 (Dense)                  (None, 128)          32896
_________________________________________________________________
dropout_3 (Dropout)              (None, 128)          0
_________________________________________________________________
dense_4 (Dense)                  (None, 1)            129
=================================================================
Total params: 230,401
```

Intended Use:

The main objective of the CineLit recommendation system's development was to offer a comprehensive and customized solution for content discovery, particularly addressing the difficulties users encounter when browsing the vast array of literary and cinematic content.

Primary Intended Uses:

- **Content Discovery:** CineLit's goal is to help people find books and movies that closely align with their individual preferences. To provide suggestions for interesting and relevant information, the system makes use of sophisticated recommendation systems.
- **Personalized Recommendations:** Giving customers individualized advice based on their past interactions, preferences, and usage patterns is the main use case. CineLit customizes suggestions to each user profile by utilizing reinforcement learning, content-based filtering, and collaborative filtering.
- **Cross-Domain Integration:** CineLit is the only service that combines the film and literature domains in a unique way, giving customers recommendations that flow naturally between the two. It is intended for the system to serve individuals who are interested in pursuing content from various domains.

Primary Intended Users:

- **General Users:** A wide range of general users seeking tailored recommendations for both books and movies are the target audience for CineLit.
- **Subscribers to online streaming services:** People who use services like Amazon Prime Video to explore content and want a more enriching and personalized experience.
- **Book Enthusiasts:** Individuals interested in literature, seeking recommendations that bridge the gap between books and movies.

Out-of-Scope Use Cases:

- **Non-Entertainment Recommendations:** CineLit is not designed for recommending content outside the realms of literature and film. Suggestions pertaining to subjects or goods unrelated to entertainment are deemed outside the purview of this guide.
- **Real-Time Content Updates:** Recommendations are based on past interactions and data that is available at the time of recommendation; the system may not be able to handle real-time content updates. Updates that reflect releases or occurrences in real time are deemed outside of the scope.
- **Customized User Data Sharing:** Use cases that include external data sharing outside of this scope are deemed out of scope, as the main goal is to improve user experience within the recommendation system.

Factors:

Relevant Factors:

- **User Preferences:** The recommendation engine may create a personalized and fulfilling experience for users by using this deep understanding to identify and recommend literary and cinematic works that resonate with them.
- **Historical Interactions:** The system is able to recognize patterns and correlations by gaining important insights into users' interests, behavior, and habits through an analysis of past encounters. These observations play a key role in collaborative filtering, in which users with similar viewing and reading habits are matched by the algorithm, which results in recommendations based on shared preferences.
- **Content Genre:**The recommendation system's core function is to identify and classify information according to genres, which includes both literary and film genres. Consequently, this enhances the relevancy and appeal of the recommended material overall by enabling the recommendation engine to provide users with tailored suggestions that align with their preferred genres.
- **Recommendation System Metrics and Features:** By matching recommendations to the interests of users who have similar tastes, these collaborative filtering metrics act as crucial indications. Information-based filtering also makes use of attributes like keywords, tags, and descriptions, all of

which play a major role in presenting information that is similar to what users have previously liked. Furthermore, characteristics controlling the system's learning and adaptation over time govern its reinforcement learning component. The system's ability to react to changing user preferences depends heavily on these parameters.

Evaluation Factors:

- **Recommendation Accuracy:**The system's precision and suggestion accuracy are crucial metrics that establish the efficacy and dependability of the recommended material. Precision is a measure of the system's capacity to provide pertinent and customized recommendations, which has a direct impact on user happiness. Recommendation engines that are accurate also help to encourage a better user experience by increasing confidence in them.
- **Cross Domain Integration Success:** This feature emphasizes how easily the algorithm may go beyond traditional movie-centric recommendations by adding literary options to its arsenal. The significance is in presenting a comprehensive and all-encompassing method of content discovery, providing users with an extensive array of recommendations that span the literary and cinematic worlds.
- **Cold Start Problem Mitigation:**This measure assesses how well the system adjusts and produces pertinent recommendations in the face of the intrinsic constraint of insufficient user or item data. It is critical to properly handle the cold start issue because it guarantees that customers with little to no past contact history would receive personalized recommendations, improving their overall user experience and recommendation system satisfaction.
- **Algorithm Robustness:**Algorithmic robustness and stability, which include content-based filtering, cooperative filtering, and reinforcement learning algorithms, are essential for guaranteeing reliable performance in a range of scenarios and conditions. This important statistic assesses how well the recommendation engine continues to produce accurate and dependable predictions in the face of shifting user preferences, changing content landscapes.

Metrics:

Recommendation Accuracy Metrics:

- **Accuracy:** The trained model is evaluated on the test dataset, and metrics like top-100 accuracy and RMSE are printed.
- **Precision:** Indicates how relevant the recommended films and books are to the user by calculating the accuracy of positive recommendations relative to the overall number of recommendations.
- **Recall:** Determines how well the system captures all relevant items by dividing the total number of relevant items by the proportion of correctly recommended items.

Metrics for User Engagement:

- **Click-Through Rate (CTR):** This metric gauges how many users click on things that are recommended relative to all users, offering valuable information about how well recommendations encourage user interaction. This was done by predicting the rating that would be given by the user.

Rates of Adaptive Learning:

- **Dynamic Learning Rates:** Based on the unpredictability and diversity of user interactions, learning rates for training are dynamically modified. Here the model’s optimizer was already set with an optimized value for it’s learning rate

Future Scope Techniques:

- **Combining Models:** To reduce variability, the system combines predictions from content-based filtering, collaborative filtering using ensemble approaches.
- **Feedback Loop:** User Feedback Integration: The recommendation process incorporates ongoing user feedback, which enables the system to adjust and reduce uncertainty.

Evaluation of Data:

- **Datasets:** CineLit makes use of a large dataset that was obtained from Kaggle. It consists of three main files: credits.csv, keywords.csv, and movies_metadata.csv. The movies_metadata.csv file, which forms the basis of the content-based filtering strategy, offers critical information about 45,000 films. Movie plot keywords are contained in the keywords.csv file, which enhances the dataset for comprehensive content analysis. In the meantime, the credits.csv file examines movie credits and offers insightful information via JSON-structured objects. Together, these datasets provide the framework for content-based and collaborative filtering systems.
- **Motivation:**The goal of the dataset selection is to build a comprehensive and varied collection that includes works from the literary and film industries. CineLit attempts to provide a comprehensive recommendation experience that smoothly combines literature and film, catering to the various tastes of users by combining movie metadata with keywords and credits.
- **Preprocessing:**To improve the quality of the data for a later study, the textual data is cleaned, especially by removing extraneous characters and punctuation.
- **Training Data:**The preprocessed and combined dataset, which includes textual and numerical elements, makes up the training data. Whereas the content-based filtering strategy makes use of characteristics like genres, keywords, and descriptions, the collaborative filtering method depends on user-movie interactions obtained via user ratings.

Quantitative Analysis:

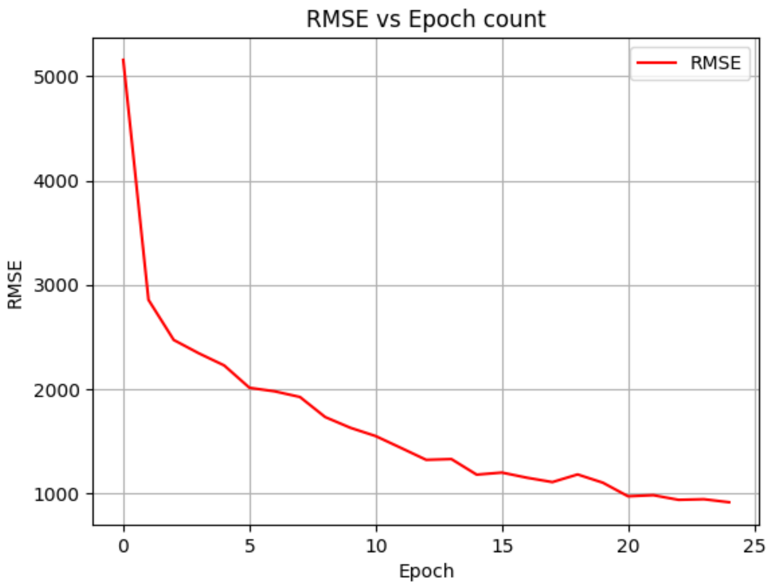
- **Unitary results:** CineLit's effectiveness in unitary analysis is demonstrated by discrete criteria like accuracy, recall, and precision. Precision gauges how accurate product recommendations are, making sure that suggested films and novels suit users' tastes. Recall measures how well the system can retrieve pertinent items, highlighting a wide range of user interests. Accuracy gives a comprehensive picture of CineLit's performance by measuring the total correctness of predictions. These unified results demonstrate the system's capacity to provide consumers with precise, thorough, and accurate recommendations, improving their experience finding content.
- **Intersectional results:** In intersectional analyses, the combined effectiveness of content-based and collaborative filtering in CineLit is assessed. The system attempts to offer recommendations that take advantage of both inherent content attributes and collaborative insights from user activities by evaluating the junction of these two techniques. Intersectional results show how collaborative and content-based methods work effectively together, proving that the system can provide a wide range of thoughtful recommendations that satisfy users' multifaceted preferences.

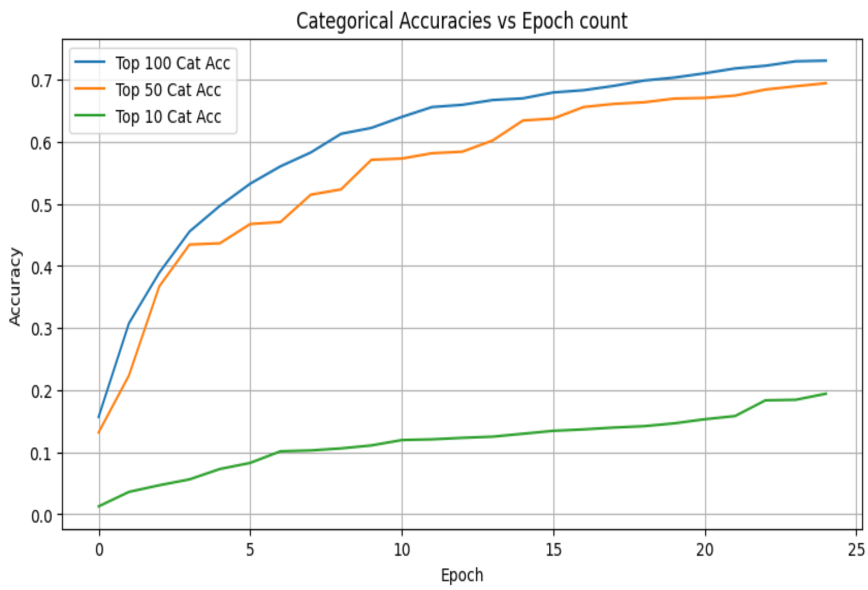
Ethical Considerations:

- **Data:** CineLit is dedicated to protecting the security and privacy of user information and recognizes the sensitivity of that data. To prevent unwanted access, stringent data protection procedures are put in place, guaranteeing that user choices and interactions are handled with the highest discretion.
- **Human Life:** The user's preferences and well-being are given top priority by the recommender system. Recommendations that can encourage harmful material or negatively impact people should be avoided for ethical reasons. CineLit promotes a wide variety of content that complies with moral guidelines with the goal of pleasantly and responsibly enhancing user experiences.
- **Mitigations:** CineLit checks its recommendation algorithms on a regular basis and includes fairness limits to address potential biases and unforeseen outcomes. The system's ethical foundation is based in part on proactive modifications and continuous monitoring, which guarantee that recommendations are fair, varied, and considerate of users' views.
- **Risks and Harms:** CineLit strives hard to reduce any potential hazards while acknowledging the dangers that may be connected with content suggestions. Avoiding anything that might be damaging, offensive, or discriminating is part of this. The user feedback tools of the system are essential for quickly detecting and fixing any problems. A typical example can be that if a user's content is well

registered by the recommender system to be mostly that of kids. The model must not pop adult or even semi-adult content to the user.

Model Evaluation:





The deep learning model, the model that was trained with a word embedding of 128 and a Keras framework, achieved 70% accuracy and saturated as the number of epochs increased. The accuracy was calculated with respect to Top 100, Top50 and Top10 and Root Mean Square Error. A noticeable observation is that there are few junk recommendations visible in both approaches with a lower similarity score in the below the recommended table for both the deep learning model as well the previous prediction through cosine similarity.