Deep Learning Based Collaborative Filtering

Bhavya Patel (bsp75), Paul Kanyuch (pwk29),
Abhigya Sinha (as3883)

Our Model

- Deep Learning-Based collaborative filtering neural network with embedding layers
- Attention mechanism for explainability
- Interactive web interface
- Comprehensive metrics suite
- Visualization tools

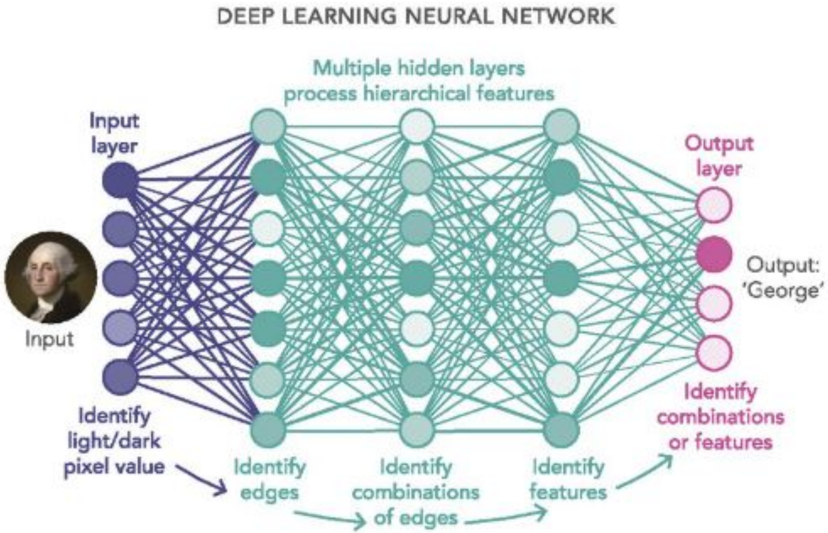
Our Dataset

- Dataset Name: Movielens-Small
  - 100,000 ratings provided by 610 users across 9,724 movies.
  - Explicit feedback in the form of user ratings on a scale from 1 to 5.
- Characteristics:
  - Sparsity: Most users rate only a small fraction of available movies, creating a sparse user-item interaction matrix.
  - Class Imbalance: Due to sparsity, ratings are often biased towards positive feedback, as users tend to rate more movies they like rather than dislike.
- Preprocessing Steps:
  - User and movie IDs are encoded as integers using label encoding.
  - The dataset is split into 80% training and 20% validation sets, maintaining temporal order.

	userId	movieId	rating	timestamp
0	1	1	4.0	964982703
1	1	3	4.0	964981247
2	1	6	4.0	964982224
3	1	47	5.0	964983815
4	1	50	5.0	964982931

	movieId	title	genres
0	1	Toy Story (1995)	Adventure|Animation|Children|Comedy|Fantasy
1	2	Jumanji (1995)	Adventure|Children|Fantasy
2	3	Grumpier Old Men (1995)	Comedy|Romance
3	4	Waiting to Exhale (1995)	Comedy|Drama|Romance
4	5	Father of the Bride Part II (1995)	Comedy

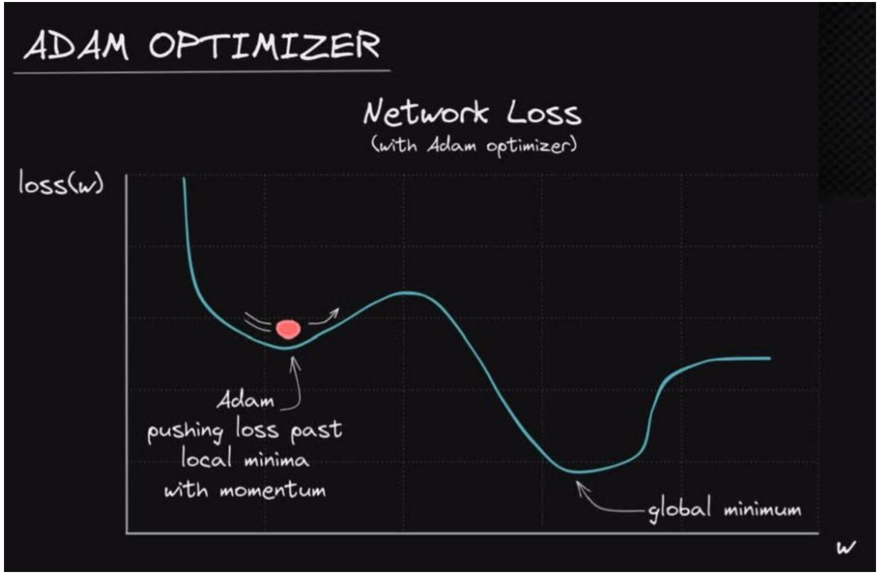
Model Architecture



- Input Representation:
  - User IDs and Movie IDs are mapped to dense 128-dimensional embeddings.
  - Embedding layers learn latent features for users and movies.
- Embedding Concatenation - Combines user and movie embeddings into a single feature vector.
- Hierarchical Neural Network:
  - Fully connected layers with progressive dimensionality reduction:
    - 512 → 256 → 128 → 64 neurons.
  - ReLU activation captures non-linear relationships.
- Regularization:
  - Batch normalization stabilizes training by mitigating internal covariate shifts.
  - Graduated dropout prevents overfitting (0.3 → 0.2 → 0.1).
- Output Layer: Single neuron with Sigmoid activation, scaled to a 1-5 system to match database ratings.

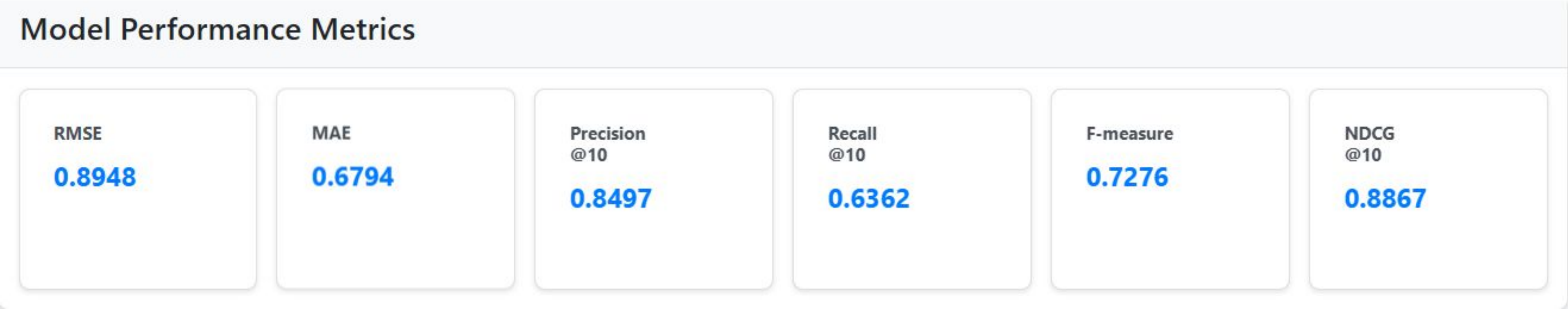
Optimization Strategies

- Xavier Initialization:
  - Ensures balanced weight initialization across layers and prevents scales from diminishing/amplifying.
  - Prevents vanishing or exploding gradients, facilitating stable training.
  - Designed to handle deep networks by sustaining even gradient propagation.
- Adam optimizer
  - Combines the benefits of momentum and adaptive learning rates for efficient convergence.
  - Automatically adjusts learning rates for individual parameters based on their gradients.
  - Utilizes 1st and 2nd moment estimates of gradients for updates.
  - Integrated with a learning rate scheduler to reduce the rate when validation loss plateaus, preventing overfitting to the training data.



Evaluation Metrics for Sample Run

- RMSE: 0.8968 (indicating strong predictive accuracy).
- Mean Absolute Error (MAE): 0.6870
- Precision@10: 0.8497 (measuring relevance of top recommendations).
- Recall@10: 0.6362 (capturing the proportion of relevant items retrieved).
- F-measure: 0.7276 (balancing precision and recall).
- NDCG@10: 0.8851 (demonstrating effective ranking performance for top recommendations).



Top 5 Recommendations for UserID 1

Recommendations for User 1

Two Days, One Night (Deux jours, une nuit) (2014)
Predicted Rating: 4.61

Jules and Jim (Jules et Jim) (1961)
Predicted Rating: 4.57

Gaga: Five Foot Two (2017)
Predicted Rating: 4.57

From the Earth to the Moon (1998)
Predicted Rating: 4.55

Casablanca (1942)
Predicted Rating: 4.54

Bonus: Explainability

- Uses an attention-based mechanism to provide transparency in how recommendations are generated, with a recommendation strength scored between 0 and 1.
- Two key factors: user preferences and movie characteristics
    - Attention weights always total 1
- User preferences weight shows how much the recommendation relies on the user's rating history, while the movie characteristics weight indicates the contribution of the movie's inherent features.
- A high-confidence recommendation (0.9) might have different weightings, such as 0.65 for user preferences and 0.35 for movie characteristics.
- This approach enhances system trust by making the recommendation process transparent, allowing users to understand the rationale behind specific movie suggestions.

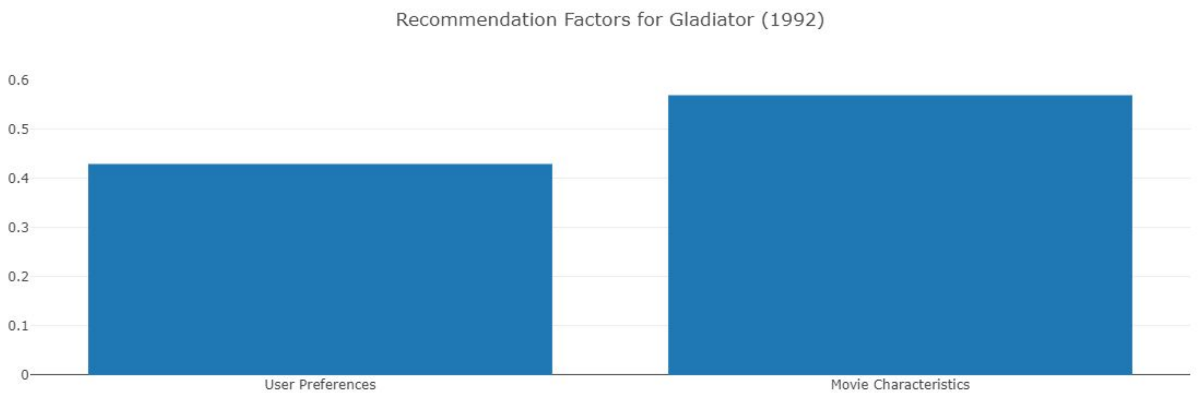
Bonus: Explainability for User 2

Recommendation Explanations

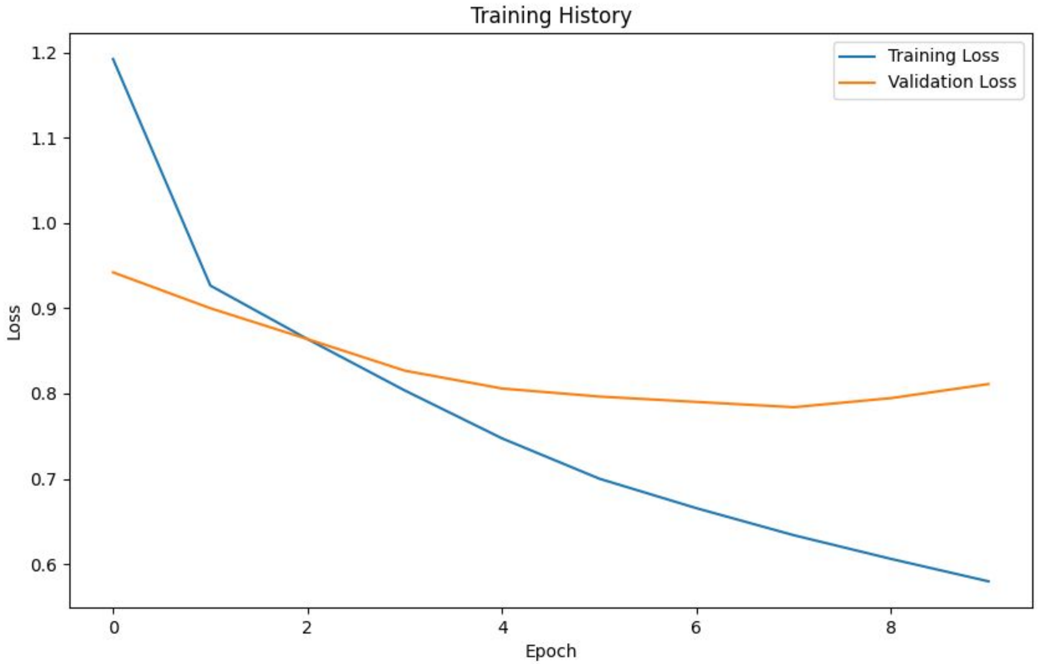
Gladiator (1992)

Predicted Rating: 4.59

```
Recommendation strength: 0.92
Main factors in this recommendation:
- User preferences: 0.43
- Movie characteristics: 0.57
```



Training Process



Conclusion

- Utilized a deep-learning based collaborative filtering system and neural network architectures with dual embedding layers and regularization techniques
- Experimental results on MovieLens dataset showed superior performance:
  - RMSE (0.8968), Precision@10 (0.8497), Recall@10 (0.6362), and NDCG@10(0.8870)
- Hierarchical neural architectures enabled nuanced capture of user-item interactions through progressive dimensionality reduction; graduated dropout and batch normalization provide robust regularization
- 3 promising future research directions: (1) contextual personalization, (2) multimodal learning, and (3) adaptive learning frameworks