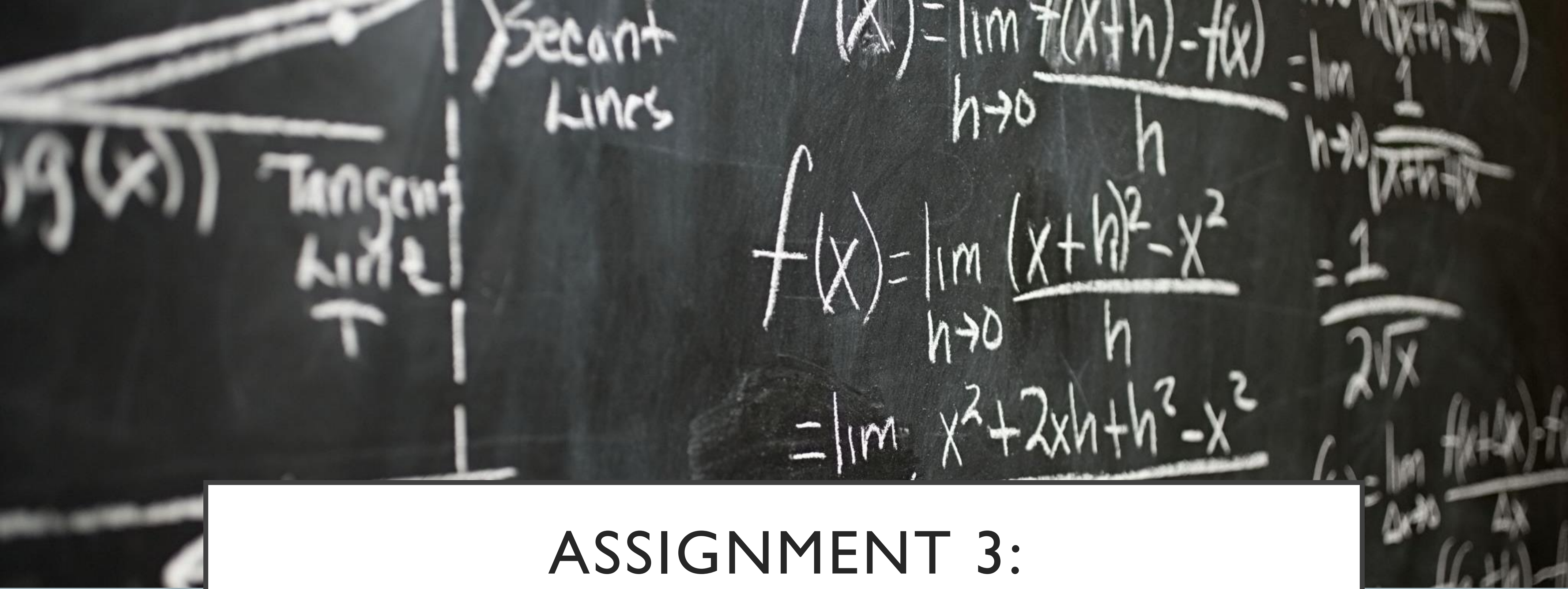


ASSIGNMENT 3: RECOMMENDER SYSTEM

Presenter: Linh Thuy Do – s3927777

TABLE OF CONTENT

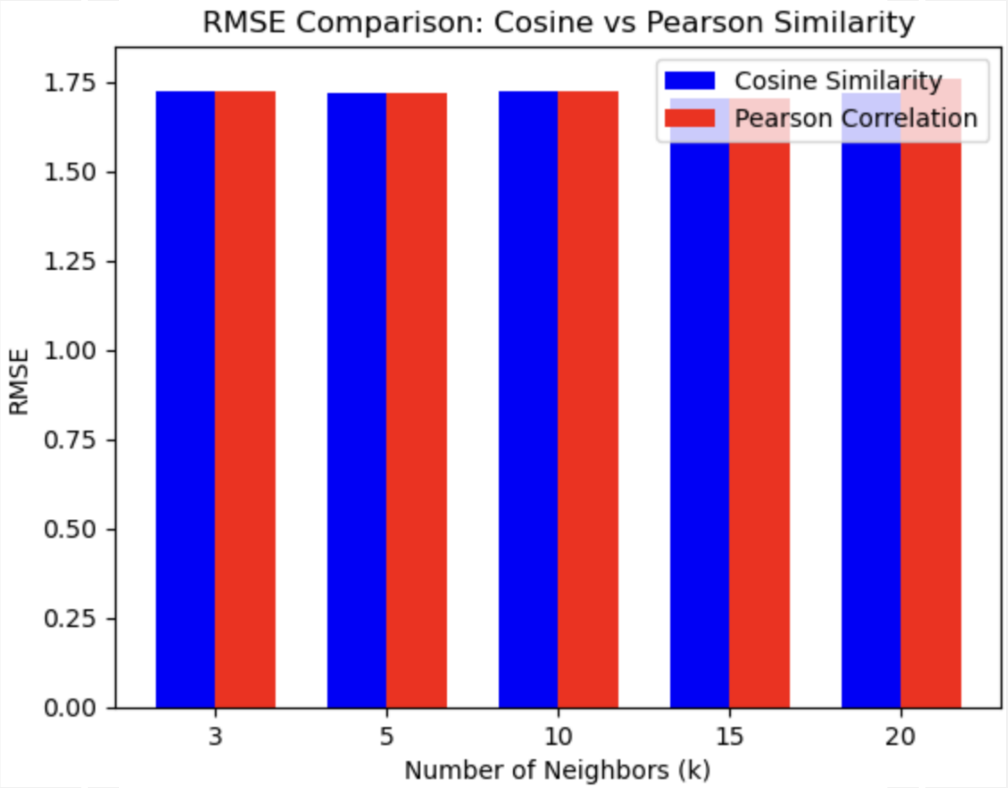
TASK 1: KNN-BASED COLLABORATIVE FILTERING RESULTS

Task Overview

- **Objective:** Implement k-Nearest Neighbors (kNN)-based Collaborative Filtering to predict movie ratings for a selected user.
- **Evaluation Metric:** RMSE (Root Mean Square Error)
- **Comparison:** Cosine Similarity vs. Pearson Correlation

Results

- **Cosine Similarity:**
  - Lowest RMSE at k=15: 1.7056
  - RMSE decreased as k increased, with optimal accuracy at k=15 before a slight increase.
- **Pearson Correlation:**
  - Similar trend with lowest RMSE at k=15: 1.7056
  - RMSE increased after k=15, showing diminishing returns.



TASK 2: MATRIX FACTORIZATION-BASED RECOMMENDATION

1. Matrix Factorization Technique and Improvement Approach

a. Matrix Factorization Technique:

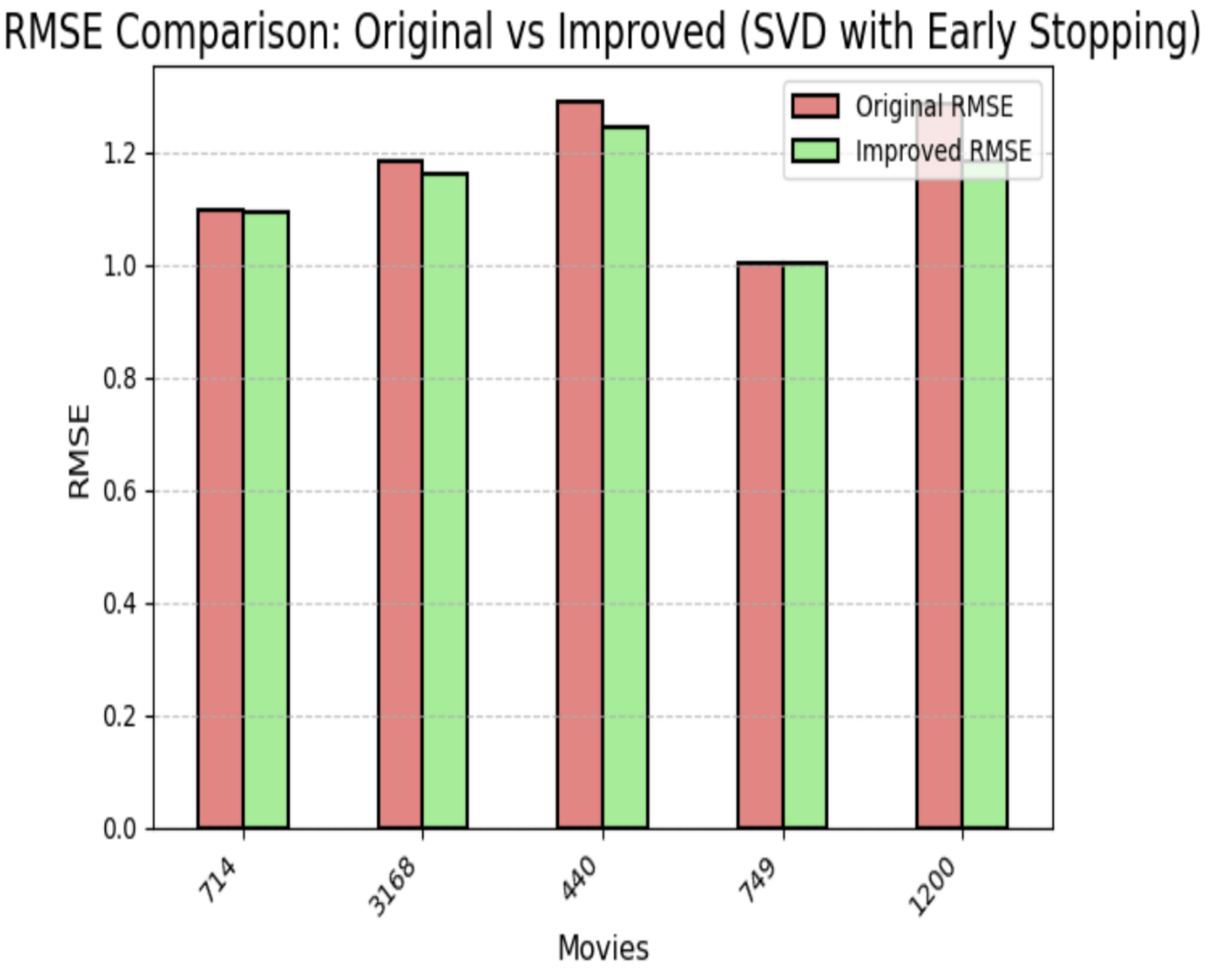
- Overview: SVD decomposes the user-item matrix into three smaller matrices, capturing relationships between users and items.
- Components:
  - U: User factors representing preferences.
  - Σ: Singular values indicating feature importance.
  - V^T: Item factors capturing item characteristics.
- Goal: Reconstruct the matrix to predict missing ratings based on user-item interactions.

b. Improvement Approach: Early Stopping in SVD

- Challenge: Standard SVD may overfit, especially if it continues to iterate without meaningful improvement.
- Early Stopping Solution:
  - min_delta: Defines minimum RMSE improvement threshold.
  - patience: Specifies maximum iterations with no improvement before stopping.
- Implementation: Tested different numbers of latent factors (20, 50, 100) and applied early stopping.
- Objective: Reduce unnecessary iterations, avoid overfitting, and improve prediction accuracy.

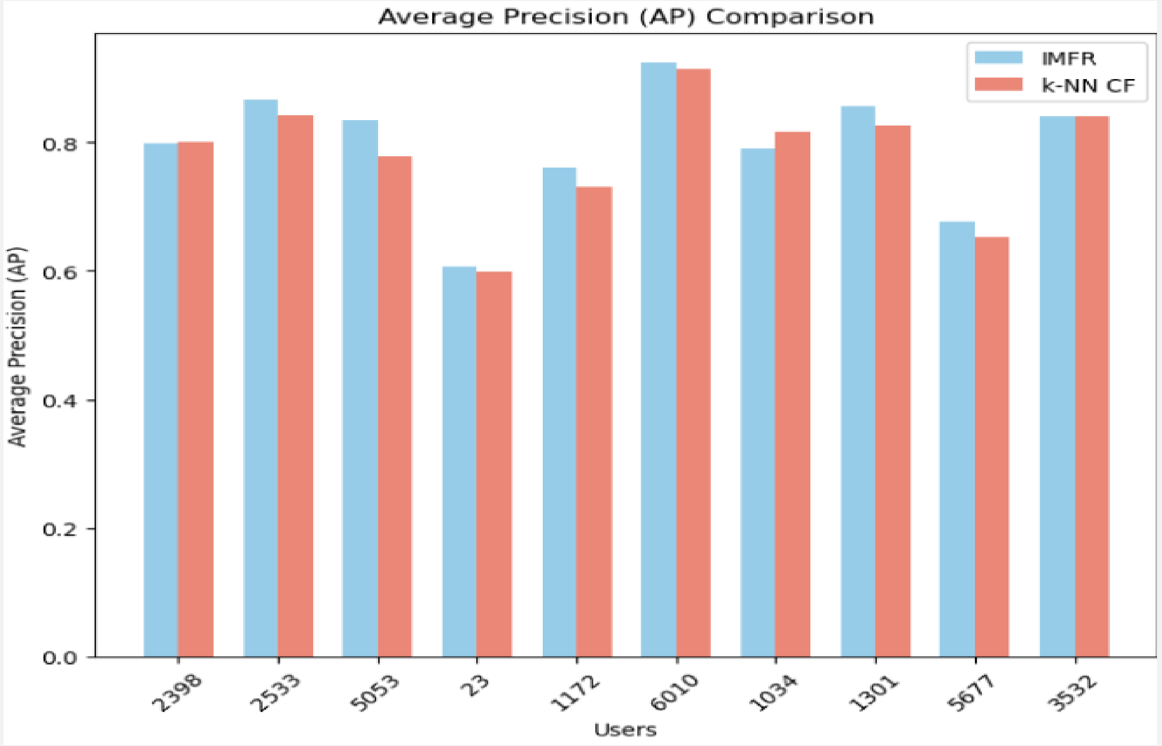
2. RESULTS AND CONCLUSION

1. **RMSE Comparison (Original vs Improved)**:
   - Original SVD:
     - RMSE values: Generally higher for the 5 selected movies, indicating less accurate predictions.
     - Overfitting Issue: No stopping criteria, leading to potential overfitting on noise in data.
   - Improved SVD with Early Stopping:
     - Reduced RMSE: RMSE decreased across all selected movies, showing enhanced accuracy.
     - Optimal Latent Factors: 100 factors with early stopping achieved the best balance of accuracy and complexity.
2. **Why the Improved Approach Works Better:**
   - Prevents Overfitting: Early stopping avoids extra iterations that adapt to noise, focusing only on meaningful data patterns.
   - Efficient Model: Early stopping allows the model to achieve low RMSE with fewer, more impactful iterations, improving generalization to new data.

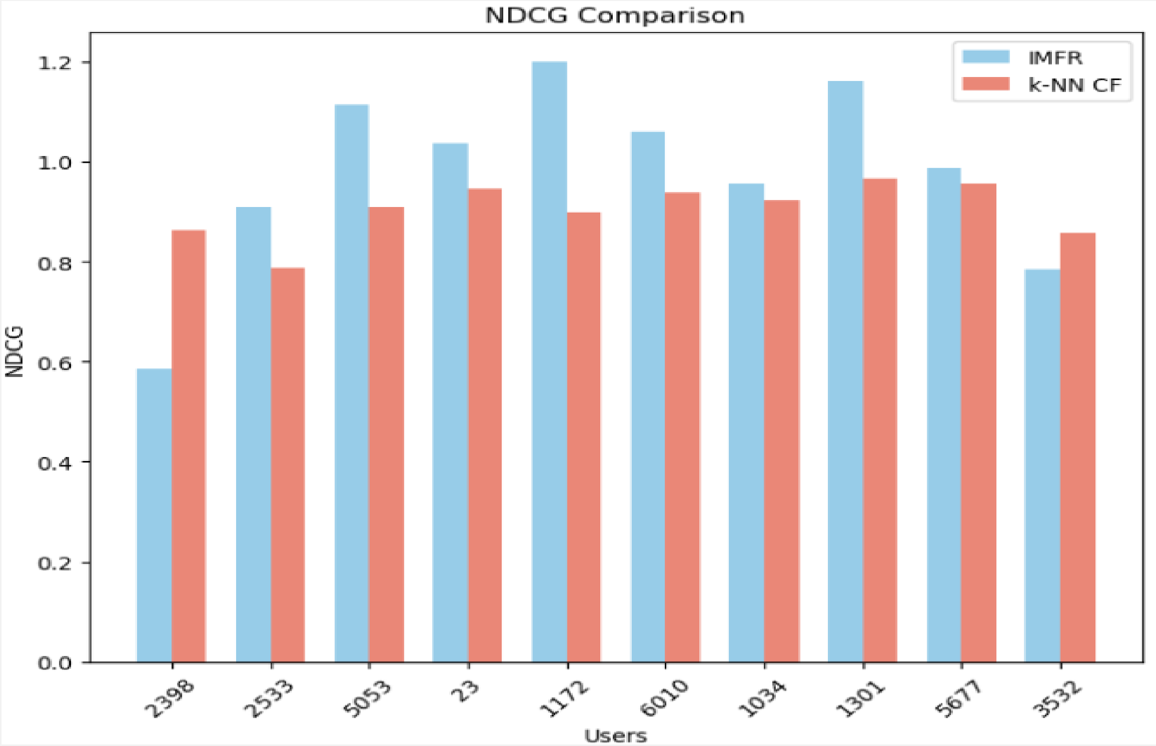


TASK 3: RANKING-BASED EVALUATION AND COMPARISON

AP COMPARISON FOR KNNCF AND IMFR



NDCG COMPARISON FOR KNNCF AND IMFR



LIMITATIONS OF KNNCF AND IMPROVEMENT OPPORTUNITIES

- **Limitations of KNNCF:**
  - Dependency on Similarity: Relies heavily on user similarity, which is challenging for sparse data or unique user preferences.
  - Cold Start Problem: Struggles with new users or items with limited ratings.
  - Ranking Quality: KNNCF does not capture deeper user-item interactions, leading to lower ranking precision.
- **Improvement Opportunities:**
  - Hybrid Approach: Integrate content-based or matrix factorization methods to supplement recommendations.
  - Adaptive Weighting: Adjust similarity scores based on user activity or preferences.
  - Dimensionality Reduction: Apply techniques like SVD to improve ranking quality by focusing on influential patterns.

WHY IMFR DELIVERS BETTER PERFORMANCE

Strengths of IMFR:

- Latent Factor Modeling: Captures complex user-item interactions beyond surface similarity.
- Ranking Precision: Matrix factorization enhances ranking, positioning items based on user preference predictions.
- Adaptive Learning: Early stopping in SVD prevents overfitting, ensuring better generalization.

Impact on Evaluation Metrics:

- Higher AP: Indicates more relevant top recommendations.
- Higher NDCG: Shows improved ranking, with relevant items effectively prioritized.

THANK YOU FOR WATCHING

REFERENCES

1. Y. Koren, R. Bell, and C. Volinsky, "Matrix factorization techniques for recommender systems," *IEEE Computer*, vol. 42, no. 8, pp. 30–37, Aug. 2009, doi:10.1109/MC.2009.263.
2. F. Ricci, L. Rokach, and B. Shapira, *Recommender Systems Handbook*, 2nd ed. New York, NY, USA: Springer, 2015, doi:10.1007/978-1-4899-7637-6.
3. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.
4. C. C. Aggarwal, *Recommender Systems: The Textbook*. Cham, Switzerland: Springer, 2016, doi:10.1007/978-3-319-29659-3.
5. A. Paterek, "Improving regularized singular value decomposition for collaborative filtering," in *Proc. KDD Cup and Workshop*, 2007, pp. 39–42.
6. X. Su and T. M. Khoshgoftaar, "A Survey of Collaborative Filtering Techniques," *Advances in Artificial Intelligence*, vol. 2009, pp. 1–19, 2009.
7. R. Burke, "Hybrid Recommender Systems: Survey and Experiments," *User Modeling and User-Adapted Interaction*, vol. 12, no. 4, pp. 331–370, Nov. 2002, doi:10.1023/A:1021240730564.