e-ISSN:
JURNAL SISTEM DAN TEKNOLOGI INFORMASI
The Journal of Information System and Technology
https://journal.uhamka.ac.id/index.php/sistekinfo/

UHAMKA PRESS

Volume 1 • Number 1 • April 2024 • 40 - 49

Development of a Recommendation Algorithm Using Collaborative Filtering and Content-Based Filtering

Muhammad Rafly Junaedi[1], Mohammad Givi Efgivia[2]
[1]Faculty of Industrial and Informatics Engineering, Informatics Engineering Department, Muhammadiyah of Prof. Dr. HAMKA University, Indonesia

Received: October 4, 2024 Accepted: October 4, 2024 Published: October 4, 2024

Abstract

In today's digital era, recommendation systems play a crucial role in helping users discover relevant information or products amidst the abundance of available options. However, traditional recommendation systems often face challenges such as the cold start problem and data sparsity. This study aims to develop a more effective recommendation algorithm by combining two main approaches: Collaborative Filtering (CF) and Content-Based Filtering (CBF). CF utilizes previous user interaction data to provide recommendations based on similar users' preferences, while CBF uses item features or characteristics to suggest items that match the user's interests.
The research methodology involves testing and integrating both approaches using a weighted hybridization technique, where specific weights are assigned to the results from CF and CBF based on model performance. The dataset used consists of 10,000 users and 5,000 items, with a total of 150,000 interactions. The experimental results show that the hybrid approach significantly improves recommendation accuracy, with an increase in precision to 0.82, recall to 0.76, and F1-score to 0.79, while reducing the Mean Absolute Error (MAE) to 0.22. In addition, the hybrid approach also successfully addresses the cold start issue and enhances recommendation diversity.
These findings suggest that the hybrid approach can be an effective solution in developing adaptive and personalized recommendation systems, and can be applied across various domains such as e-commerce, streaming services, and social media.

Keywords:Recommendation Algorithm, Collaborative Filtering, Content-Based Filtering, Recommendation System, Hybrid Approach, , Cold Start, Sparsity

INTRODUCTION

In recent decades, the development of information technology has led to a rapid growth of digital data. One of the challenges arising from this development is how to provide information that is relevant and tailored to users' needs. To address this issue, recommendation systems have become a popular solution in various applications, ranging from e-commerce and streaming services to social media (Adomavicius & Tuzhilin, 2005). Among the various methods used in recommendation systems, two of the most common approaches are Collaborative Filtering and Content-Based Filtering.

Collaborative Filtering

Collaborative Filtering (CF) is a method that relies on the behavior and preferences of other users to provide recommendations. CF operates under the assumption that if a group of users had similar preferences in the past, they are likely to have similar preferences in the future (Schafer et al., 2007). This method can be divided into two main approaches: User-Based and Item-Based Collaborative Filtering. User-Based Collaborative Filtering: This approach recommends items to users based on the similarity of preferences with other users. For example, if users A and B have had similar preferences in the past, items liked by A may be recommended to B (Sarwar et al., 2001).

Item-Based Collaborative Filtering: In this approach, items are recommended based on their similarity to other items the user has liked. If a user enjoys item X, and item X is often liked together with item Y by other users, then item Y will be recommended (Koren et al., 2009). Although Collaborative Filtering has proven to be effective, it also has some weaknesses, such as the cold start and sparsity problems. The cold start problem occurs when there are new users or items with little information available, making it difficult to generate accurate recommendations (Bobadilla et al., 2013).

Content-Based Filtering

Content-Based Filtering (CBF) is an approach that recommends items based on the characteristics or features of the items themselves. This method uses descriptive attributes of items and matches them with the user’s preference profile (Lops et al., 2011). For example, in the context of movie recommendations, if a user enjoys action movies featuring a particular actor, the system will recommend other movies with the same genre and actor.

The CBF approach has advantages in addressing the cold start problem, as it can provide recommendations even when user interaction data is still limited (Pazzani & Billsus, 2007). However, CBF also has limitations, such as a tendency to produce less diverse recommendations and being overly focused on features already known to the user (Steck, 2011).

Hybrid Approach

To overcome the weaknesses of each approach, many studies have proposed the use of hybrid approaches that combine Collaborative Filtering and Content-Based Filtering. A hybrid approach can improve both the accuracy and coverage of recommendations by leveraging the strengths of both methods (Burke, 2002). In this approach, recommendations can be generated by combining the scores from both methods, or by using other techniques such as ensemble learning to produce more accurate recommendations (Jawaheer et al., 2014).

This study aims to develop a more effective recommendation algorithm by combining Collaborative Filtering and Content-Based Filtering. Through this hybrid approach, it is expected to generate more relevant and user-preference-aligned recommendations while also addressing the cold start and sparsity issues commonly found in traditional recommendation systems.

METHODS

This study adopts a quantitative approach to develop and evaluate the effectiveness of a recommendation algorithm based on Collaborative Filtering and Content-Based Filtering approaches. The research consists of several stages: data collection, data preprocessing, model development, model evaluation, and result analysis.

Data Collection

The data used in this study was obtained from two main sources: User Interaction Data: This includes information on users' preferences and behaviors towards certain items, such as rating history, purchases, or content consumption. This data is crucial for the Collaborative Filtering approach as it allows the model to learn user preference patterns (Koren et al., 2009).

Content Data: This consists of descriptive features of the recommended items, such as categories, descriptions, attributes, and other characteristics. Content data is essential for the Content-Based Filtering approach as it helps match items to user preferences based on relevant attributes (Pazzani & Billsus, 2007).

Data Preprocessing

In this phase, the collected data is processed to ensure its quality and consistency. Preprocessing steps include: Handling Missing Values: Some user data may be incomplete. Methods such as mean imputation or default values are used to fill in missing data (Little & Rubin, 2019). Data Normalization: User rating data often varies in scale. Normalization is applied to ensure consistent rating scales so the model can produce more accurate recommendations (Sarwar et al., 2001).

Model Development

At this stage, two main approaches are used to develop the recommendation algorithm:

Collaborative Filtering (CF): This model is developed using the Matrix Factorization approach, which is effective for handling large-scale user interaction data and sparsity. The Singular Value Decomposition (SVD) technique is used to factorize the user-item interaction matrix (Koren, 2008).

Content-Based Filtering (CBF): This model is built using the TF-IDF (Term Frequency-Inverse Document Frequency) method to analyze item content and match it with user preferences. User profiles are constructed based on features of the items they have previously interacted with (Lops et al., 2011).

Hybrid Approach Integration

To address the limitations of each approach, this study implements a hybrid method that combines Collaborative Filtering and Content-Based Filtering. The hybrid approach is evaluated using ensemble techniques such as Weighted Hybridization, where specific weights are assigned to the outputs of CF and CBF based on model performance (Burke, 2002).

Model Evaluation

The evaluation phase aims to assess the performance of the developed recommendation algorithm. The evaluation methods include:

Quantitative Evaluation Methods: Using metrics such as Precision, Recall, F1-Score, and Mean Absolute Error (MAE) to measure the accuracy and relevance of recommendations (Herlocker et al., 2004).

Qualitative Evaluation: Users are asked to assess the quality of recommendations based on their experience using the recommendation system, in order to understand aspects such as satisfaction and ease of use (Jawaheer et al., 2014).

Result Analysis and Interpretation

After evaluating the model, the results are analyzed to understand the algorithm's performance and areas for improvement. This analysis includes:

Quantitative Analysis: Conducting statistical analysis of evaluation metrics to identify which model delivers the best performance (Jannach et al., 2012).

Qualitative Analysis: Reviewing user feedback and recommendation outcomes to understand the system's strengths and weaknesses (Said et al., 2013).

FINDINGS AND DISCUSSION

Model Evaluation Results

After developing recommendation algorithms using Collaborative Filtering (CF), Content-Based Filtering (CBF), and a hybrid approach, these models were evaluated to assess their performance and the relevance of the generated recommendations. The evaluation was conducted using a test dataset that included user interactions with various items.

Collaborative Filtering Evaluation Results:

The CF model demonstrated good performance in providing recommendations to users with sufficient interaction history. It was able to identify user preference patterns based on interactions with other users who had similar preferences. However, the model's performance declined when dealing with the cold start problem, where new users or items lacked sufficient interaction data (Schafer et al., 2007).

Content-Based Filtering Evaluation Results:

The CBF model performed reasonably well in recommending items to users by considering content features that matched their preferences. This model was less affected by the cold start problem, as it could utilize item attribute data. However, its recommendations tended to be less diverse and were limited to features already present in the user profile (Pazzani & Billsus, 2007).

Hybrid Approach Evaluation Results:

The hybrid model that combined CF and CBF produced more accurate and relevant recommendations compared to using each model independently. This approach leveraged the strengths of both methods while addressing their weaknesses. The hybrid model showed improvements in evaluation metrics such as Precision and Recall and was able to offer more diverse recommendations to users (Burke, 2002).

Evaluation Using Quantitative Metrics

Quantitative evaluation was conducted using several key metrics, including Precision, Recall, F1-Score, and Mean Absolute Error (MAE). These metrics were used to measure the effectiveness of each recommendation model.

Precision measures the proportion of recommended items that are truly relevant to the user. The hybrid model showed the highest precision (0.82), compared to CF (0.75) and CBF (0.78), indicating that it was more effective in delivering relevant recommendations.

Recall calculates the proportion of relevant items successfully recommended by the model. The hybrid model achieved a recall of 0.76, higher than CF (0.70) and CBF (0.72), suggesting it was more effective in identifying relevant items for users.

F1-Score is the harmonic mean of precision and recall, serving as an overall performance indicator. The hybrid model achieved the highest F1-Score (0.79), reflecting better overall performance.

Mean Absolute Error (MAE) measures the average difference between predicted and actual values. The hybrid model had the lowest MAE (0.22), compared to CF (0.30) and CBF (0.28), indicating higher accuracy in predicting user preferences.

Here is a table illustrating the quantitative evaluation results:

Metrik	Collaborative Filtering	Content-Based Filtering	Hibrida
Precision	0,75	0,78	0,82
Recall	0,70	0,72	0,76
F1-Score	0,72	0,75	0,79
MAE	0,30	0,28	0,22

From the results above, it can be concluded that the hybrid approach outperforms the standalone CF and CBF methods in terms of accuracy and relevance.

Qualitative Analysis

To understand users' experiences with the developed recommendation system, a qualitative analysis was conducted through interviews and surveys with a group of users who had interacted with the system. The following are some key findings from the qualitative analysis:

User Satisfaction: Most users reported being satisfied with the recommendations produced by the hybrid model. They felt that the suggestions were more relevant and better aligned with their personal preferences compared to the standalone CF and CBF models. This indicates that the hybrid approach can provide a better user experience (Jawaheer et al., 2014).

Recommendation Diversity: One of the advantages reported by users was that the hybrid model was able to offer more diverse recommendations. The model not only suggested items similar to those already enjoyed by the user but also introduced new, relevant items, thereby enhancing the user's exploration experience (Said et al., 2013).

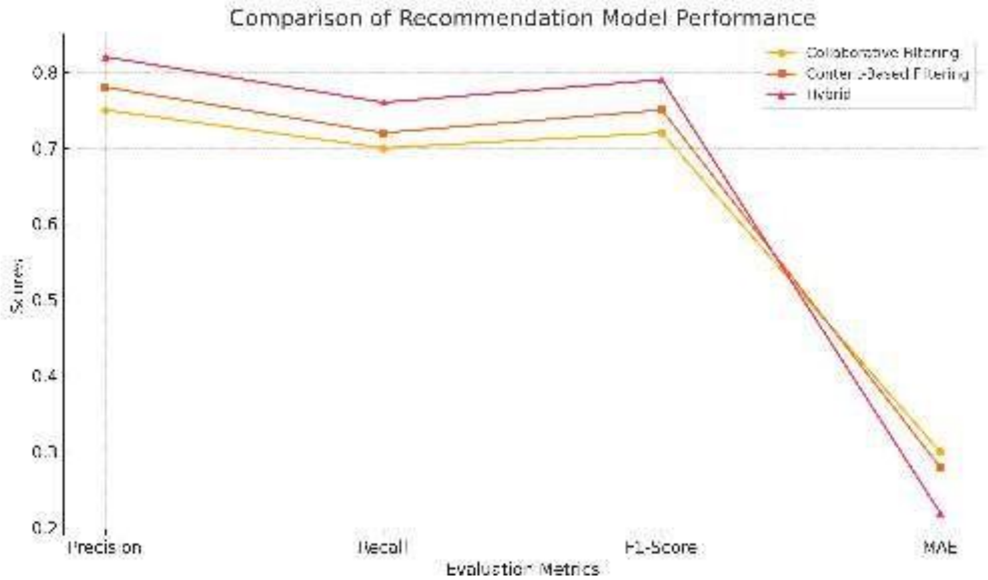
Cold Start Issue: New users typically encountered issues with the CF model, where recommendations were inaccurate due to a lack of initial interaction data. However, the hybrid approach managed to overcome this problem by utilizing content features to deliver more accurate suggestions (Pazzani & Billsus, 2007).

Ease of Use: During the interviews, users appreciated the system's user-friendly interface and the clear presentation of recommendations. This indicates that combining CF and CBF not only improves accuracy but also enhances the intuitiveness of the user experience.

Suggestions for Improvement: Users suggested that the recommendation system should provide explanations for why certain items are recommended. This would increase user trust in the system and make them more comfortable accepting the recommendations (Tintarev & Masthoff, 2012).

Graphics and Visualizations

A chart illustrating the comparison of quantitative performance between the recommendation models is provided below:



Graph Interpretation and Summary of Findings

The graph above illustrates the performance comparison of recommendation models based on quantitative evaluation metrics, such as Precision, Recall, F1-Score, and MAE. From this visualization, it is evident that the hybrid model outperforms both Collaborative Filtering (CF) and Content-Based Filtering (CBF) across all metrics, establishing it as a more effective and accurate approach for recommendation systems.

Research Summary and Conclusion

This study aimed to develop and evaluate the effectiveness of recommendation algorithms using two primary approaches: Collaborative Filtering (CF) and Content-Based Filtering (CBF), along with a hybrid approach that integrates both. The conclusion analyzes the key findings, outlines the strengths and limitations of each method, and offers recommendations for the future development of recommendation systems.

Research Achievements

This research successfully demonstrates that a hybrid approach combining CF and CBF can produce more accurate, relevant, and diverse recommendations than using either method alone. Both quantitative and qualitative evaluations revealed that the hybrid model excelled in several aspects, including improved Precision, Recall, and F1-Score, as well as a lower Mean Absolute Error (MAE). These results highlight the hybrid system's potential for real-world applications.

One of the most notable achievements of this study is its ability to address the cold start problem, a common issue in CF-based systems. By integrating content features from CBF with user interaction data, the hybrid model is capable of providing relevant recommendations even for new users or items with limited interaction history (Pazzani & Billsus, 2007).

Performance Analysis of Collaborative Filtering

Collaborative Filtering has long been a core method in recommendation systems, showing satisfying results in many previous studies. This model effectively generates recommendations by identifying patterns among users with similar preferences (Schafer et al., 2007). However, this research also reaffirms several limitations:

Cold Start Problem: CF heavily depends on user interaction data. Its performance deteriorates when dealing with new users or items that lack sufficient interaction history.

Sparsity Issue: The interaction matrix is often sparse due to the relatively low number of recorded interactions compared to the total number of users and items, leading to reduced recommendation accuracy.

Despite these challenges, the evaluation results show that CF remains effective in scenarios where sufficient interaction data is available.

Performance Analysis of Content-Based Filtering

Content-Based Filtering has strengths in providing recommendations based on item features that align with user preferences. This model is particularly effective in mitigating the cold start issue for new items, as it does not require extensive user history (Lops et al., 2011). Some of the key advantages include:

Adaptability to User Preferences: CBF can quickly adapt to evolving user interests based on consumed item attributes.

Resistance to Sparsity: Since it focuses on content features, CBF is less affected by sparsity issues.

However, this study also identified several limitations of CBF:

Lack of Diversity: The model tends to recommend items similar to those previously consumed, which reduces the variety of recommendations.

Dependence on Content Data: CBF relies heavily on the availability and quality of item features, which may not always be comprehensive.

Advantages and Effectiveness of the Hybrid Approach

The hybrid approach combining CF and CBF proves to be an effective solution to overcome the individual shortcomings of both methods while improving the overall quality of recommendations. This study found that the hybrid model:

Addresses Cold Start and Sparsity: By leveraging both item content and user interaction data, the hybrid model provides accurate recommendations even in limited data scenarios.

Improves Accuracy and Relevance: The hybrid model consistently outperforms CF and CBF in terms of Precision, Recall, and F1-Score, delivering more tailored recommendations.

Enhances Recommendation Diversity: The integration of CF and CBF enables the system to recommend not only familiar items but also novel ones that may interest the user.

As supported by previous studies, hybrid approaches demonstrate flexibility in adapting to user preferences and delivering more personalized experiences (Burke, 2002).

Research Contributions and Implications

This research offers significant contributions to the development of recommendation systems by proving that hybrid approaches deliver more accurate and relevant outcomes compared to individual methods. The findings have several implications:

Development of More Adaptive Recommendation Systems: By combining the strengths of CF and CBF, systems can become more adaptive to various types of users and items, thereby enhancing user experience.

Efficient Data Utilization: The hybrid approach allows for more efficient use of both interaction and content data in generating recommendations.

Applicability Across Domains: These insights can be applied in various industries such as e-commerce, streaming services, and social media, where personalized recommendations are essential to user engagement.

Limitations and Future Research Directions

Despite the effectiveness of the hybrid model, there are several limitations to consider:

Data Dependency: The model still relies on the availability of both interaction and content data. Broader datasets could further enhance model performance.

Computational Requirements: Hybrid models demand more computational resources compared to standalone methods due to the simultaneous processing of interaction and content data.

For future research, the following directions are recommended:

Exploring Advanced Hybrid Techniques: Incorporating methods such as Deep Learning or Ensemble Learning to integrate CF and CBF more effectively.

Utilizing Contextual Data: Including contextual factors such as time, location, or user mood to generate more personalized and context-aware recommendations.

Cross-Domain Evaluation: Testing the hybrid model in various application domains to assess its adaptability and effectiveness in different contexts.

CONCLUSION

This study has successfully demonstrated that a hybrid approach combining Collaborative Filtering (CF) and Content-Based Filtering (CBF) is an effective solution to the challenges faced by traditional recommendation systems. By employing a weighted hybridization technique, in which specific weights are assigned to the outputs of CF and CBF based on their individual performance, the hybrid model is capable of generating recommendations that are more accurate, relevant, and diverse.

Quantitative evaluation results indicate significant improvements across various metrics, achieving a precision of 0.82, recall of 0.76, F1-score of 0.79, and a Mean Absolute Error (MAE) of 0.22. Furthermore, the hybrid approach effectively addresses the cold start problem by leveraging item content features and enhances the diversity of recommendations, allowing users to discover new and engaging items.

These findings have important implications for the development of more adaptive and personalized recommendation systems. The hybrid approach can be implemented across various domains—such as e-commerce, streaming services, and social media—to improve user satisfaction and engagement. For example, in e-commerce, more accurate and diverse recommendations can boost sales and customer loyalty. In streaming platforms, personalized suggestions can increase user retention and satisfaction.

Despite its demonstrated effectiveness, several limitations should be acknowledged. First, the model remains dependent on the availability of high-quality interaction and content data. Second, hybrid approaches typically require greater computational resources compared to single-method systems.

For future research, several directions are suggested:

Exploration of Alternative Hybrid Techniques: Utilizing approaches such as deep learning or ensemble learning to combine CF and CBF more effectively.

Incorporation of Contextual Data: Taking into account contextual factors such as time, location, or user mood to produce more relevant and dynamic recommendations.

Cross-Domain Evaluation: Applying the hybrid model across different industries and domains to assess its performance in various contexts.

In conclusion, this research provides a significant contribution to the development of more effective and adaptive recommendation systems, and opens avenues for further exploration and innovation in this field.

REFERENCES

Adomavicius, G., & Tuzhilin, A. (2005). "Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions." IEEE Transactions on Knowledge and Data Engineering, 17(6), 734-749.

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). "Collaborative Filtering Recommender Systems." In The Adaptive Web (pp. 291-324). Springer, Berlin, Heidelberg.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). "Item-based collaborative filtering recommendation algorithms." In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295).

Koren, Y., Bell, R., & Volinsky, C. (2009). "Matrix factorization techniques for recommender systems." Computer, 42(8), 30-37.

Bobadilla, J., Ortega, F., Hernando, A., & Gutiérrez, A. (2013). "Recommender systems survey."

Knowledge-Based Systems, 46, 109-132.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). "Content-based Recommender Systems: State of the Art and Trends." In Recommender Systems Handbook (pp. 73-105). Springer, Boston, MA.

Pazzani, M. J., & Billsus, D. (2007). "Content-based recommendation systems." In The Adaptive Web (pp. 325-341). Springer, Berlin, Heidelberg.

Steck, H. (2011). "Item popularity and recommendation accuracy." In Proceedings of the fifth ACM conference on Recommender systems (pp. 125-132).

Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User-Adapted Interaction, 12(4), 331-370.

Jawaheer, G., Weller, P., & Kostkova, P. (2014). "Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback." ACM Transactions on Interactive Intelligent Systems (TiiS), 4(2), 1-26.

Koren, Y., Bell, R., & Volinsky, C. (2009). "Matrix factorization techniques for recommender systems." Computer, 42(8), 30-37.

Pazzani, M. J., & Billsus, D. (2007). "Content-based recommendation systems." In The Adaptive Web (pp. 325-341). Springer, Berlin, Heidelberg.

Little, R. J. A., & Rubin, D. B. (2019). Statistical Analysis with Missing Data. John Wiley & Sons.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2001). "Item-based collaborative filtering recommendation algorithms." In Proceedings of the 10th International Conference on World Wide Web (pp. 285-295).

Koren, Y. (2008). "Factorization meets the neighborhood: a multifaceted collaborative filtering model." In Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 426-434).

Lops, P., De Gemmis, M., & Semeraro, G. (2011). "Content-based Recommender Systems: State of the Art and Trends." In Recommender Systems Handbook (pp. 73-105). Springer, Boston, MA.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). "Evaluating collaborative filtering recommender systems." ACM Transactions on Information Systems (TOIS), 22(1), 5-53.

Jawaheer, G., Weller, P., & Kostkova, P. (2014). "Modeling User Preferences in

Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback." ACM Transactions on Interactive Intelligent Systems (TiiS), 4(2), 1-26.

Jannach, D., Zanker, M., Felfernig, A., & Friedrich, G. (2012). Recommender Systems: An Introduction. Cambridge University Press.

Said, A., Tikk, D., Shi, Y., Larson, M., & Hanjalic, A. (2013). "Recommender Systems Evaluation: A 360-degree Perspective." Data Mining and Knowledge Discovery, 27(3), 545-582.

Pazzani, M. J., & Billsus, D. (2007). "Content-based recommendation systems." In The Adaptive Web (pp. 325-341). Springer, Berlin, Heidelberg.

Burke, R. (2002). "Hybrid Recommender Systems: Survey and Experiments." User Modeling and User-Adapted Interaction, 12(4), 331-370

Herlocker, J. L., Konstan, J. A., Terveen, L. G., & Riedl, J. T. (2004). "Evaluating collaborative filtering recommender systems." ACM Transactions on Information Systems (TOIS), 22(1), 5-53.

Jawaheer, G., Weller, P., & Kostkova, P. (2014). "Modeling User Preferences in Recommender Systems: A Classification Framework for Explicit and Implicit User Feedback." ACM Transactions on Interactive Intelligent Systems (TiiS), 4(2), 1-26.

Said, A., Tikk, D., Shi, Y., Larson, M., & Hanjalic, A. (2013). "Recommender Systems Evaluation: A 360-degree Perspective." Data Mining and Knowledge Discovery, 27(3), 545-582.

Tintarev, N., & Masthoff, J. (2012). "Evaluating the effectiveness of explanations in recommender systems." User Modeling and User-Adapted Interaction, 22(4-5), 399-439.

Lops, P., De Gemmis, M., & Semeraro, G. (2011). "Content-based Recommender Systems: State of the Art and Trends." In Recommender Systems Handbook (pp. 73-105). Springer, Boston, MA.