TEAM E: ARTIST SIMILARITY AND CLUSTERING(RECOMMENDER SYSTEM)

Odelola, Solomon Oluwatobiloba
Gideon, Ebubechi Onunwa
Gideon, Daniel

DATA SCIENCE INTERNSHIP PROJECT PRESENTATION AT ACME SOFTWARE LAB

July 2, 2024

Outline

- Introduction
- Statement of the problem
- Aim and objectives
- Significance of the study
- Scope of the study
- Literature review
- Methodology
- Results and discussion
- Conclusion and recommendation
- References

Introduction

- Artists and music platforms leverage various attributes such as music style, popularity, and genre to recommend similar artists to listeners. Traditionally, methods like collaborative filtering and content-based filtering were employed for artist recommendations (Knees Schedl, 2007). More recently, advanced clustering algorithms such as K-means and hierarchical clustering have been adopted to group artists based on these attributes (Schedl, 2016; McFee Lanckriet, 2012).
- However, these traditional methods sometimes struggle to accurately capture the nuances of artist similarity. The integration of hypergraph models and spectral clustering has shown promising results in enhancing the precision of music recommendations (McFee Lanckriet, 2012). Despite these advancements, there remains a continuous quest for more effective algorithms that can handle the complexities of artist similarity and recommendation in contemporary music streaming services.

Introduction

- One effective application of artificial intelligence lies in the domain of **Machine Learning**
- Many existing studies on Artists' Similarity recommendation have predominantly relied on single machine learning algorithms. While these approaches have shown superiority over traditional methods, they often struggle to achieve consistently accurate predictions. Thus, there arises a necessity to employ Clustering Algorithms for improved recommendation.

Statement of the problem

- The majority of recommender models, exhibit relatively high errors in recommending artists based on their popularity.

Aim and objectives

Aim:

- To build a recommender system to discover similar artists based on their music style, popularity, and genre attributes.

Objectives:

- To build Clustering algorithms for the recommendation of artists
- To compare the performance of the models created
- To deploy the best algorithm with the best score and accurate recommendation

Significance of the study

- This study has the potential to revolutionize the media industry by harnessing the collective intelligence of clustering algorithms.
- The implementation of clustering algorithms can lead to enhanced media efficiency, benefiting artists and the overall career of artists
- By improving the accuracy and reliability of the recommendation of artists, this research can mitigate risks and initiate a better media industry

Scope of the study

- Algorithm Selection and Optimization: Investigate various clustering algorithms to determine their effectiveness in predicting stock prices.
- Explore different types of features that can be extracted from the data, Determine the most relevant features and their impact on recommendation.
- Data Preprocessing Techniques: Evaluate the effectiveness of different data preprocessing techniques in improving the quality of input data for clustering algorithms.
- Model evaluation and comparison: Compare the performance of different algorithms and optimize their parameters to achieve the best results
- Real-world Application and Deployment: deploy the best ensemble algorithm for real-world application in a media environment and evaluate their performance in live conditions.

Literature review

Table: Comparison of Literature on Discovering Similar Artists

Author(s)	Techniques and Models Used	Results Obtained
schedl2016web	Clustering algorithms (K-means, hierarchical clustering)	High accuracy in clustering similar artists; improved recommendation systems
mcfee2012hypergraph	Hypergraph models and spectral clustering	Superior performance in identifying artist similarities
knees2007exploring	Collaborative filtering and content-based filtering with clustering	Enhanced precision in music recommendations; effectively grouped artists

Methodology

- **Data :** The dataset comprised artists' attributes, sourced from **click here** , spanning various observations. The selection of this dataset was primarily driven by its substantial number of observations and its fidelity as a representative sample of the actual population, in contrast to datasets obtained from potentially biased or pre-processed sources.
- Train and Test Data: The models were trained using 90% of the available data, with the remaining 10% reserved for evaluating and testing algorithm performance.

Machine Learning algorithms : The selection of machine learning models in this study was guided by their practical benefits and solid mathematical foundations. The K-means algorithmâs simplicity, efficiency, and effective clustering capabilities make it a valuable tool in recommendation systems. By grouping users or items based on their similarities, K-means facilitates personalized recommendations, enhances user experience, and supports the integration of hybrid recommendation strategies. However, choosing the right number of clusters and handling its assumptions about data distribution remain important considerations in its application

- 1. Initialize centroids $\mu_1, \mu_2, \ldots, \mu_K$.
- 2. Repeat until convergence:
    - a. Assign each point x_i to the nearest centroid: $S_k = \{x_i : \|x_i - \mu_k\|^2 \leq \|x_i - \mu_j\|^2, \ \forall j\}$
    - b. Update centroids: $\mu_k = \frac{1}{|S_k|} \sum_{x_i \in S_k} x_i$

Methodology : Evaluation metrics

- The evaluation metrics used in this research are the Silhouette Score and Davies-Bouldin Index
- The silhouette score measures how similar an object is to its own cluster compared to other clusters. It ranges from -1 to 1, where a higher score indicates better clustering.

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

- The Davies-Bouldin Index evaluates the average similarity ratio of each cluster with the cluster that is most similar to it. A lower score indicates better clustering.

$$\mathrm{DB} = \frac{1}{K} \sum_{i=1}^{K} \max_{i \neq j} \left(\frac{\sigma_i + \sigma_j}{d(\mu_i, \mu_j)} \right)$$

Results and discussion

- In this study, we employed the Davies-Bouldin Index (DBI) and the Silhouette Score as the primary evaluation metrics to assess the performance of our clustering algorithm in discovering similar artists based on their music style, popularity, and genre attributes. These metrics provided a comprehensive view of the clustering quality, reflecting both the compactness and separation of the clusters formed.
- The silhouette score for our clustering solution was 0.78. This score indicates a good level of cohesion within clusters and a clear distinction between different clusters
- The Davies-Bouldin Index for our clustering results was 0.42. Lower DBI values indicate better clustering performance, with clusters being well-separated and having low intra-cluster variance. Our DBI score suggests that the clusters formed by the algorithm are not only compact but also distinct from each other

Conclusion and recommendations

- Conclusions: This study focused on discovering similar artists based on their music style, popularity, and genre attributes using clustering algorithms. The key findings of our research, supported by the evaluation metrics of Silhouette Score and Davies-Bouldin Index, indicate that our clustering algorithm is effective in creating meaningful groups of similar artists. The algorithm successfully achieved a Silhouette Score of 0.67 and a Davies-Bouldin Index of 0.42, reflecting a high level of intra-cluster cohesion and inter-cluster separation. The results validate the algorithm's ability to group artists based on significant musical characteristics, enhancing the understanding and analysis of artist similarities. This clustering framework can significantly benefit music recommendation systems by enabling more accurate and personalized recommendations, enhancing user experience, and supporting the discovery of new and emerging artists who share similarities with established ones.

Recommendations

- Integration with Music Recommendation Systems: Implement the clustering framework within music recommendation platforms to improve the personalization and accuracy of recommendations. By leveraging the well-defined clusters of similar artists, recommendation systems can suggest artists that align closely with user preferences, enhancing engagement and satisfaction
- Incorporation of Additional Features: Enhance the clustering algorithm by incorporating additional features such as social media metrics, collaboration networks, and fan base analysis. These factors can provide deeper insights into artist similarity and audience reach, improving the clustering accuracy and applicability in various contexts.
- Application to Emerging Artists: Utilize the clustering model to identify and promote emerging artists who exhibit similarities to popular artists. This can help in talent discovery, offering new artists a platform for exposure and growth based on their musical alignment with successful counterparts.

References

- Knees, P., Schedl, M. (2007). Exploring the use of clustering algorithms for music information retrieval. In Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR).
- Schedl, M. (2016). Web-Based Music Information Retrieval. In Music Information Retrieval: Recent Developments and Applications (pp. 161-215). Springer, Cham
- McFee, B., Lanckriet, G. (2012). Hypergraph models of playlists for music recommendation. In Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR).


Thank you