BACH CORAL HARMONY PREDICTION

Abstract:

This research paper presents a comprehensive comparative analysis of two prominent machine learning algorithms, Random Forest and Support Vector Classification (SVC), applied to the task of predicting chorale harmony. Chorale harmony prediction involves associating a chord label with each time point in a musical flow, revealing the underlying harmony and aiding in musical analysis. The study investigates the performance of Random Forest and SVC algorithms on a dataset comprising 60 chorales by J.S. Bach, each labeled with one of 101 chord labels. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to assess the predictive capabilities of the algorithms. Experimental results indicate that both Random Forest and SVC demonstrate promising performance, with Random Forest exhibiting slightly higher accuracy on average. However, SVC shows competitive performance in terms of precision and recall. Furthermore, the paper discusses the implications of these findings for applications in music information retrieval and computational musicology, highlighting the strengths and limitations of each algorithm in the context of chorale harmony prediction.

Introduction:

Chorale harmonization is a fundamental aspect of Western classical music, particularly in the context of Baroque music. J.S. Bach's chorales, in particular, serve as canonical examples of harmonically rich compositions. The task of chorale harmony prediction involves assigning chord labels to individual events within a chorale, providing valuable insights into the harmonic structure and aiding in musical analysis. With the advent of machine learning techniques, automated methods for chorale harmony prediction have gained traction, offering efficient and scalable solutions for music analysis tasks.

Dataset Description:

The dataset used in this study comprises 60 chorales by J.S. Bach, spanning a range of musical styles and harmonic progressions. Each chorale consists of a sequence of events, with each event labeled with one of 101 chord labels representing the underlying harmony. The dataset includes information such as pitch classes, bass notes, meter, and chord labels for each event, extracted from MIDI sources and manually annotated by experts.

Methodology:

The comparative analysis focuses on two machine learning algorithms: Random Forest and Support Vector Classification (SVC). Random Forest is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. SVC, on the other

hand, is a discriminative classifier that separates classes by finding the hyperplane that maximizes the margin between them in the feature space.

Experimental Setup:

The experiments are conducted using a cross-validated approach, where the dataset is divided into training and testing sets multiple times to ensure robust evaluation. Evaluation metrics such as accuracy, precision, recall, and F1-score are computed for each algorithm to assess its predictive performance. Hyperparameters for each algorithm, such as the number of trees in Random Forest and the choice of kernel in SVC, are tuned using grid search to optimize performance.

Results and Discussion:

The experimental results demonstrate the effectiveness of both Random Forest and SVC algorithms in predicting chorale harmony. Random Forest achieves slightly higher accuracy on average compared to SVC, indicating its robustness in handling complex feature interactions and noisy data. However, SVC exhibits competitive performance in terms of precision and recall, suggesting its efficacy in capturing subtle patterns in the dataset. The discussion highlights the trade-offs between the two algorithms in terms of computational complexity, interpretability, and generalization performance.

Conclusion:

In conclusion, this research paper presents a comparative analysis of Random Forest and Support Vector Classification algorithms for chorale harmony prediction. While Random Forest demonstrates superior accuracy, SVC offers competitive performance in terms of precision and recall. The findings contribute to the understanding of machine learning approaches for music analysis tasks and provide insights into the strengths and limitations of each algorithm. Future research directions may explore ensemble methods combining Random Forest and SVC for improved predictive performance.

References:

- Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine learning, 20(3), 273-297.