Critiquing Task- versus Goal-oriented Approaches: A Case for Makam Recognition

Kaustuv Kanti Ganguli[1,2], Sertan Şentürk[2,3], Carlos Guedes[2]

[1] College of Interdisciplinary Studies, Zayed University, UAE

[2] Music and Sound Cultures Research Group, New York University Abu Dhabi, UAE

[3] Independent Researcher, UK

Motivation

Task vs Goal Oriented

Goal: Optimization of evaluation metrics, e.g. accuracy

Task: What a model is intended to learn, e.g., modal features such as intervals, note sequence, relative salience, etc.

> Computational Musicology and Music Information Retrieval (MIR) combining top-down vs. bottom-up approaches

> Relevant in non-Eurogenetic music repertoires where the grammatical rules are rather prescriptive

> Combining domain knowledge and data-driven optimisations

Supervised vs Unsupervised

> Acoustic feature mapping with label?

> Complexities of the makams pertaining to the same tonal material

Mode Recognition

Scale — Mode — Tune

> E.g. Turkish/Arabic makam/maqam, Indian raags/ragas, Gregorian chants

> Classify each excerpt to one of the 20 makams based on audio features

> Goal: state-of-the-art classification accuracy via feature engineering and machine learning techniques

> Task: enable the model to understand the feature space and explain the similarity and variances of allied makams

Makam Characterisation

> Melodic framework defined by intervals, phrases, modulations, stylistic ornamentations etc.

> All three makams use the same tonal material but different melodic progressions

> The latters use transposition (lit: şed) of Çargah makam

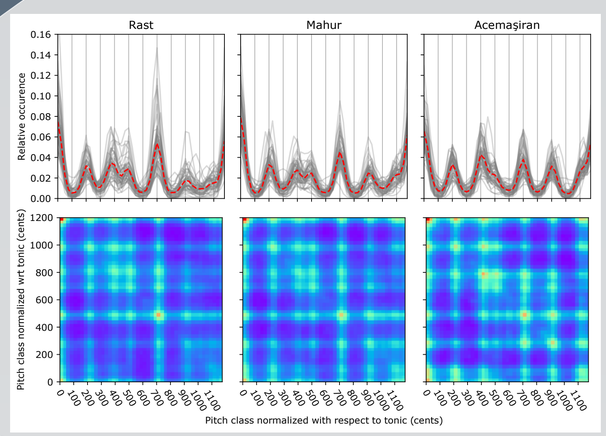
> The intervals are the same except 1-comma difference on the 3rd and 7th scale degrees

> The complexity of the musical characteristics should not be attributed to as shortcomings of computational models

> We identify makam-pairs that are highly discriminable and on the other hand, highly confusable in practice

Experiments

Pitch-Class Distributions (PCD)



Time-Delayed Melody Surfaces (TDMS)

(Supervised) Classification

> Goal-oriented approach, i.e., the best combination of features and classifiers in order to achieve the optimal accuracy

> TDMS with 1-second time-delay index, 12.5 cents of smoothing kernel width and a compression exponent of 0.5

(Unsupervised) Clustering

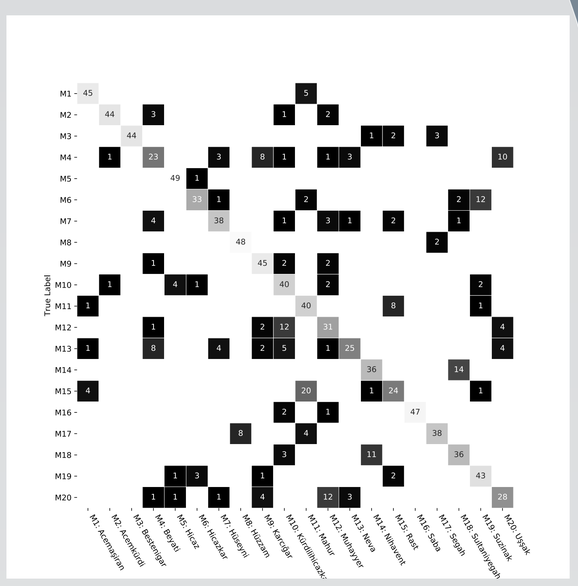
> Task-oriented approach to bridge the gap the best classification achieves

Evaluation

Classification Accuracy

Model	PCD	TDMS
Support Vector Machine	71.0∓ 3.2%	**77.2∓ 3.5%**
Multilayer Perceptron	70.9∓ 3.8%	74.6∓ 5.0%
k-nearest Neighbors	68.2∓ 3.4%	70.2∓ 3.1%
Logistic Regression	66.8∓ 3.9%	75.5∓ 4.1%

Confusion Matrix

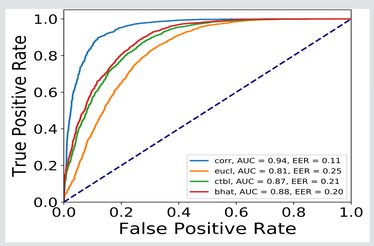


Dataset

> OTMM recognition dataset, largest for makam recognition

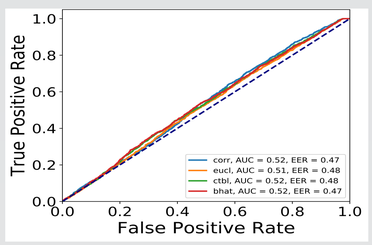
> 1000 recordings (50 rec * 20 makams) with rich metadata, e.g. tonic frequency

Retrieval Scenario



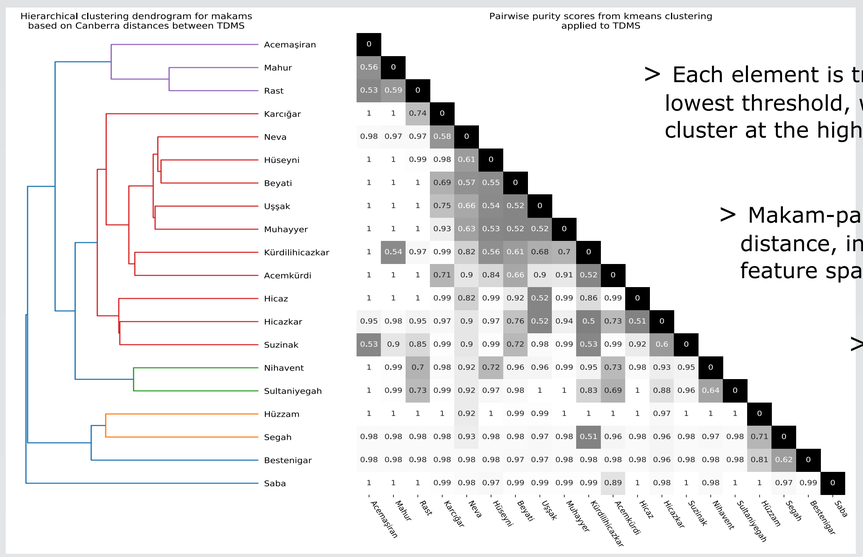
Query Retrieval Score

(Hüzzam, Hicaz)



(Rast, Mahur)

Cluster Purity Matrix



Hierarchical Clustering (Dendrogram)

> Each element is treated as distinct clusters at the lowest threshold, whereas there is a single giant cluster at the highest threshold

> Makam-pair (Rast, Mahur) show a very low distance, indicating high similarity in the feature space

> Cluster purity values confirm no random clustering

Visit the companion page!

Literature Review

> Pitch distribution-based template comparison techniques, with a recent interest in new features such as TDMS, chroma, etc.

> For the OTMM recognition, the state-of-the-art accuracy is approx. 77% under different settings by Demirel et al. and Yeşiler et al.

> Very few works on comparative study on allied makam-pairs; notable work by Bozkurt et al.

> We borrow a raga recognition method from Gulati et al. and allied mode-pair analogy from Ganguli et al.

Array of alternate configurations:
https://sertansenturk.com/work-research/ismir-2022-makam/

Major insights

> We employed methodologies that combine domain knowledgeand data-driven optimisations with a view to understandingthe makam recognition 'task' in depth

> We report comparable accuracy (77.2%) with the state-of-the-art using the newly adapted TDMS feature with SVM classifier

> We advocate that good supervised learning performance is a necessary but insufficient condition for a computational representation-cum-distance-measure to be considered informative for all purposes

References

> A. Karakurt, S. Şentürk, and X. Serra, "MORTY: A toolbox for mode recognition and tonic identification," DLfM 2016

> S. Gulati, J. Serrà Julià, K. K. Ganguli, S. Şentürk, and X. Serra, "Time-delayed melody surfaces for raga recognition," ISMIR 2016

> K. K. Ganguli and P. Rao, "On the distributional representation of ragas: experiments with allied raga pairs," TISMIR, vol. 1, no. 1, 2018

> B. Bozkurt, R. Ayangil, and A. Holzapfel, "Computational analysis of Turkish makam music: Review of state-of-the-art and challenges," JNMR, vol. 43, no. 1, 2014