Automatic Classification of Digital Music by Genre

Kamelia Aryafar Ali Shokoufandeh
Department of Computer Science, Drexel University, Philadelphia, PA, USA

Introduction

- This work proposes a method for automatic categorization of music into different **genre labels** in a large music data set.
- The main challenges in creating an automatic genre classification system are:
  - The robust representation of audio signals in terms of low-level features or high-level audio keywords.
  - The construction of an automatic learning schema to classify these feature vectors into music genres.
- This work proposes the use of an automated method based on **explicit semantic analysis** to identify the most representative genre patterns in a large data set.

Proposed Method

- **Feature selection and pre-processing:**
  - Mel frequency cepstral coefficients (MFCCs) are adopted to represent short-term power spectrum of sound and are known to be very effective for music classification systems [2].
  - For a large data set, k-means clustering of MFCCs creates the audio code-book using the cosine similarity distance measure to reduce the complexity of the feature space.
- **Concept-based representation:**
  - The explicit semantic analysis vectors are used to represent each audio in the concept space rather than the feature space.
  - In contrast to term frequency-inverse document frequency (tf-idf) modeling of textual documents, the ESA utilizes a concept-based representation of MFCCs.

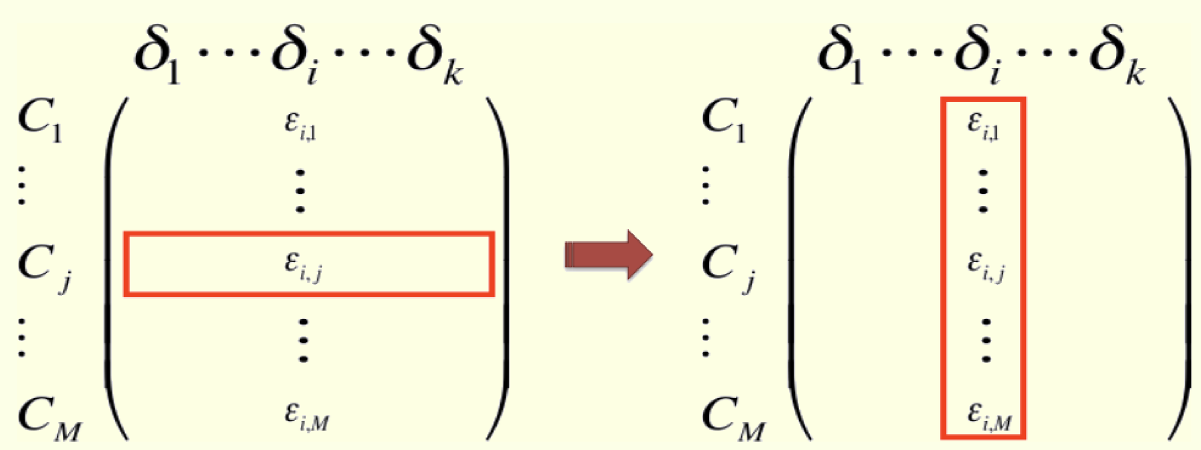
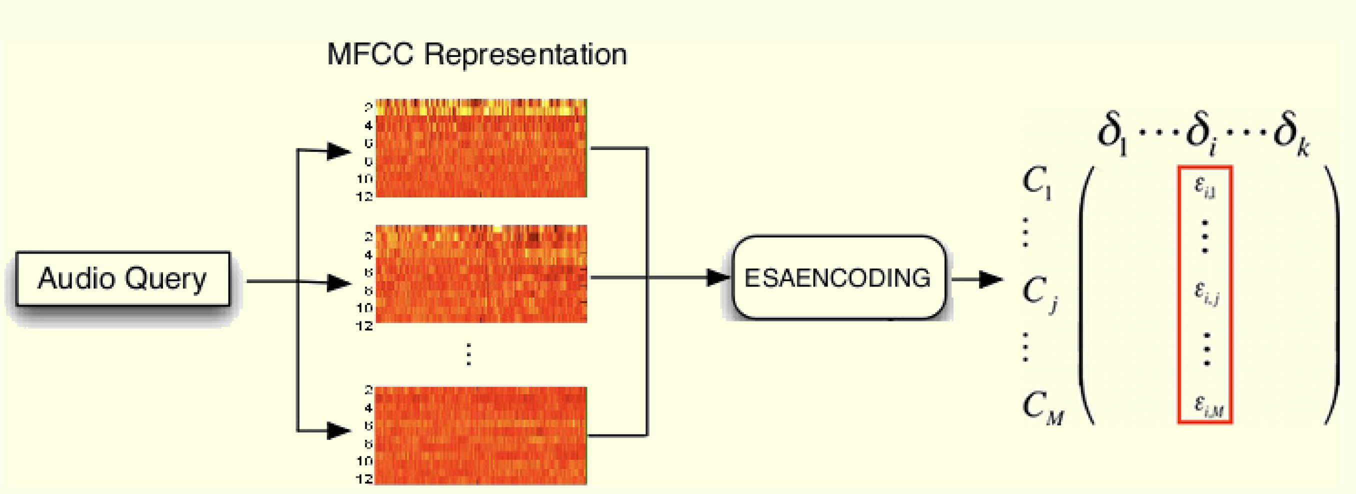


Figure: The classic tf-idf model is extended to represent MFCCs in the concept space rather than the feature space.

  - A higher value $\mathcal{E}_{i,j}$ indicates a codeword δ_i, that is frequent in concept $\mathcal{C}_j$ but rare in other concepts.



- **Genre classification:**
  - A supervised learning schema is trained on the training set to learn genre labels associated with each music sample in the concept space.
  - The trained classifier is used to assign genre labels to the music samples in the testing set [1].

Technical Terms

- Set of features $\{f_1, ..., f_\ell\}$
- Code-book of features $\mathcal{D} = \{\delta_1, ..., \delta_k\}$
- Set of audio signals in the data set $C = \langle (f_1, w_1), ..., (f_\ell, w_\ell) \rangle$ where $\mathcal{C} = \{C_1, ..., C_M\}$
- $M \times k$ ESA matrix $\mathcal{E}_{\mathcal{C},\mathcal{D}}$, where:
  - $tf(C, x) = \frac{\sum_{i=1}^{\ell} w_i \times d(f_i, x)}{\sum_{i=1}^{\ell} w_i}$ is the term frequency.
  - $idf_\delta = \log \frac{M}{\sum_{i=1}^{M} \chi(\delta, C_i)}$ is the inverse document frequency.
  - $tfidf(C, \delta) = tf(C, \delta) \times idf_\delta$.
  - $\mathcal{E}_{\mathcal{C},\mathcal{D}}[i, j] = tfidf(C_i, \delta_j)$.
- Set of t ordered pairs $\mathcal{T} = \{(A_1, L_1), ..., (A_t, L_t)\}$ of audio sequences, A_i, and their corresponding genre labels, L_i

Concept-based Representation

- For a given audio sequence A, Algorithm 1 computes the ESA vector for A.
- This is accomplished by computing the MFCC features of A and aggregating the ESA vectors corresponding to the best matching codewords in D.

Algorithm: ESAEncoding$(A, \mathcal{D}, \mathcal{E})$
Input: A: input audio, $\mathcal{D}$: code-book, $\mathcal{E}$: ESA matrix
Result: $\mathcal{E}(A)$: ESA-representation of A
$\{f_1, ..., f_\ell\} \leftarrow \mathbf{MFCC}(A)$;
$\mathcal{E}(A) \leftarrow \bar{\mathbf{0}}$;
foreach $f \in \{f_1, ..., f_\ell\}$ **do**
  $\delta^* = \max_{\delta \in \mathcal{D}} d(f, \delta)$;
  $\mathcal{E}(A) = \mathcal{E}(A) + \mathcal{E}(\delta^*)$;
end
return $\mathcal{E}(A)$

Algorithm 1: Construction of the ESA vector of an audio sequence.

- $\mathcal{E}(A)$ is the concept-based representation of audio sequence A.

Genre Classification

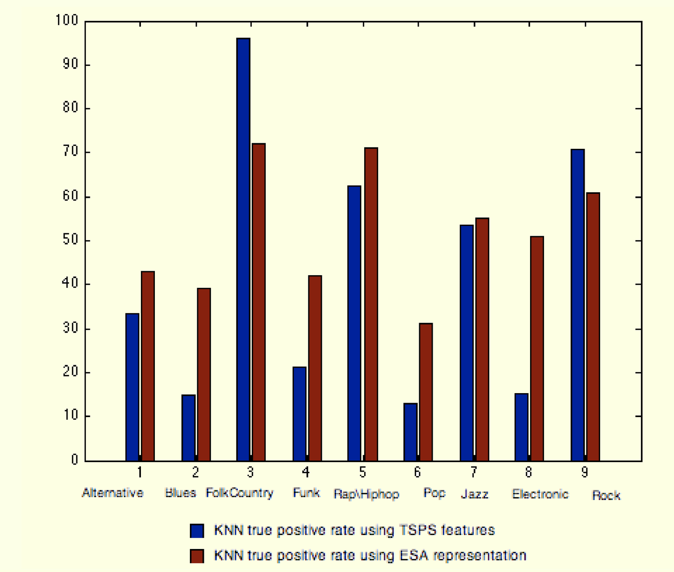
- **Training:**
  - A set of t ordered pairs $\mathcal{T} = \{(A_1, L_1), ..., (A_t, L_t)\}$ of audio sequences, A_i, and their corresponding genre labels, L_i form the training data set.
  - We form the set $\mathcal{E}(\mathcal{T}) = \{(\mathcal{E}(A_1), L_1), ..., (\mathcal{E}(A_t), L_t)\}$, where $\mathcal{E}(A_i)$is the ESA encoding of $(A_i, \mathcal{D}, \mathcal{E}),\ i = 1, ..., t$.
  - The set $\mathcal{E}(\mathcal{T})$ of (ESA-vector, label) pairs will be provided as the training data to a supervised classifier algorithm.
  - The set of hyperplanes that define the gaps between genres, are the outcome of the training on $\mathcal{E}(\mathcal{T})$.
  - Support vector machine (SVM) and k-nearest neighbors (k-NN) are used to build a model that assigns samples to their genre categories.
- **Testing:**
  - For an audio query element q we form $\mathcal{E}(q)$, the ESA representation of q.
  - The classifier, trained on set $\mathcal{E}(\mathcal{T})$, is used to estimate the genre label L_q simply by determining to which side of the genre cell (defined by the set of decision planes and genre gaps) they belong.

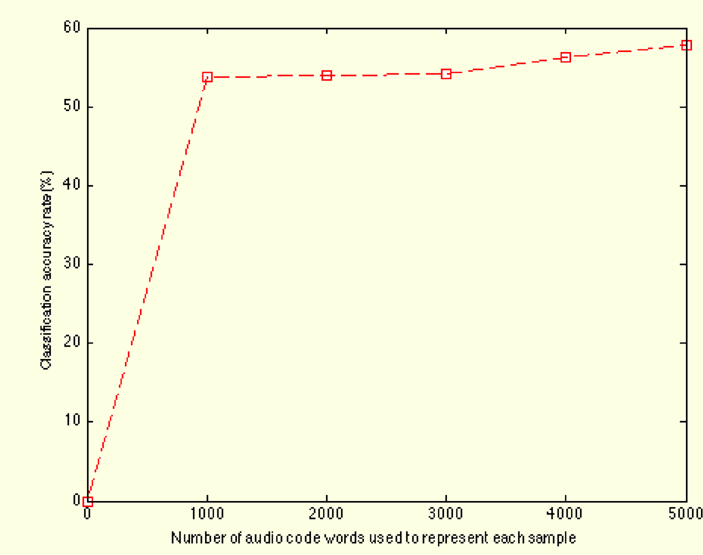
Experimental Results

- **Data set:**
  - We use the publicly available benchmark dataset for audio classification and clustering.
  - The dataset contains samples of 1886 songs obtained from the Garageband site.
  - The data set includes 9 different genre samples of different sizes.

Genre	Samples
alternative	145
blues	120
electronic	113
folk-country	222
funk soul/R&B	47
jazz	319
pop	116
rap/hip-hop	300
rock	504

- **Experimental setup**
  - Validation method: 10-fold cross validation
  - Performance measure: classification accuracy rate
  - Similarity measure: cosine distance





- Aggregation of MFCC features (AM) and temporal, spectral and phase (TSPS) features are compared to the ESA representation of MFCC features.

Method	AM	TSPS	ESA	
			$k:1000$	$k:5000$
Random	22.39	21.68	29.51	25.40
k-NN	35.83	47.40	48.59	51.88
SVM	40.81	51.81	53.76	57.81

Future Work

- Incorporate other audio features
  - Bag of audio keywords
  - Textual metadata
- Music artist identification in specific genre
- Lyrics retrieval using an extended ESA model

Literature

[1] Kamelia Aryafar, Sina Jafarpour, and Ali Shokoufandeh.
Automatic musical genre classification using sparsity-eager support vector machines.
In *Proceedings of the 21st international Conference on Pattern Recognition*, ICPR '12, 2012.

[2] Kamelia Aryafar and Ali Shokoufandeh.
Music genre classification using explicit semantic analysis.
In *Proceedings of the 1st international ACM workshop on Music information retrieval with user-centered and multimodal strategies*, MIRUM '11, pages 33–38, New York, NY, USA, 2011. ACM.