In this work, we address the problem of multivariate time series classification. In particular, we focus on classification of music sequences and propose general, and simple discrete multivariate representations (direct feature quantization, DFQ) and define a family of efficient discrete multivariate similarity kernels (DMV-SK) with manifold embedding.

Experiments using new DFQ representation and manifold kernel embedding show excellent classification performance on music time series classification tasks with significant 25%-40% improvements over existing state-of-the-art time series classification methods.

**Background.** Kernel-based approaches for sequence classification show some of the most accurate results for a variety of problems, such as image classification [12], text categorization [7, 17], speech analysis [2], biological sequence analysis [15, 7, 5], time series and music classification [16, 7].

Classification of multivariate time series can be addressed using codebook learning approach. In a codebook learning framework, multivariate time series \( X \), a sequence of \( n = |X| \) identically sized (continuous-valued) vectors, \( x_i \in \mathbb{R}^d \), is transformed into a univariate discrete sequence of codebook IDs \( c(x_i) = (c_1, c_2, \ldots, c_D) \), \( c_i \in \{1, \ldots, D\} \) for each of the vectors \( x_i \) to the nearest codeword vector in the codebook \( C = \{C_1, C_2, \ldots, C_D\} \). The codeword sequence \( c(x) \) is essentially a discrete sequence over finite alphabet \( \Sigma = \{1, \ldots, D\} \).

A number of state-of-the-art approaches to sequence classification over finite alphabet \( \Sigma \) rely on measuring sequence similarity using fixed-length representations \( \Phi(X) \) of sequences as the spectra (\( |\Sigma|^k \)-dimensional histogram) of counts of short substrings (\( k \)-mers), contained, possibly with up to \( m \) mismatches, in a sequence, c.f., spectrum/mismatch methods [9, 10].

In contrast, in this work we aim at methods that directly exploit these multivariate sequence representations to improve accuracy and propose an efficient discrete multivariate similarity kernels that as we show empirically provide effective improvements in practice over traditional 1D sequence kernels as well as other state-of-the-art time series methods for a number of challenging classification problems.

**Direct feature quantization (DFQ).** In a typically used codebook learning framework, input sequences of identically sized vectors are typically first encoded using codebook IDs, then standard 1D string kernel methods can be applied (e.g., [7, 12]). In this work we consider alternative multivariate discrete representations of original multivariate (continuous-valued) sequences.

A discrete multivariate representation of the original continuous-valued multivariate sequence \( X = (x_1, x_2, \ldots, x_n), x_i \in \mathbb{R}^d \) is obtained by direct quantization of each of the \( d \) feature dimensions. In this approach, each feature \( f_j, j = 1 \ldots d \) is quantized by dividing its range \( [f_{\text{min}}, f_{\text{max}}] \) into finite number of intervals. In the simplest case, the intervals can be defined, for instance, using uniform quantization, where the entire feature data range is divided into \( B \) equal intervals of length \( \delta = (f_{\text{max}} - f_{\text{min}})/B \) and the index of quantized feature value \( Q(f) = \lfloor (f - f_{\text{min}})/\delta \rfloor \) is used to represent the feature value \( f \).

Figure 1 shows an example of DFQ representation for 3-dimensional time series \( X \) with multivariate DFQ representation obtained using uniform binning (\( B=64 \)) along each of the three dimensions. Partitioning of the feature data range could also be obtained by using 1D clustering, e.g. \( k \)-means, to adaptively choose discretization levels.

Table: Original multivariate time series \( X \) vs discrete multivariate representation (DFQ(X))

<table>
<thead>
<tr>
<th>dim</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>43</td>
<td>37</td>
<td>31</td>
</tr>
<tr>
<td>2</td>
<td>37</td>
<td>38</td>
<td>32</td>
</tr>
<tr>
<td>3</td>
<td>39</td>
<td>39</td>
<td>33</td>
</tr>
</tbody>
</table>

Univariate representation (VQ(X), codebook size=2048)

<table>
<thead>
<tr>
<th>dim</th>
<th>1</th>
<th>2</th>
<th>3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>146</td>
<td>124</td>
<td>108</td>
</tr>
<tr>
<td>2</td>
<td>123</td>
<td>123</td>
<td>123</td>
</tr>
<tr>
<td>3</td>
<td>128</td>
<td>128</td>
<td>128</td>
</tr>
</tbody>
</table>

Figure 1: Proposed discrete multivariate representation (DFQ). Discrete representation DFQ(X) is obtained from the original continuous-valued multivariate sequence \( X \) by discretizing each feature dimension. In contrast, typically used vector quantization approach (codebook) represents multivariate sequence as one-dimensional discrete sequence of codeword indices.

**Manifold discrete similarity kernels.** Similar to [6], we define similarity kernel between two multivariate sequences \( X \) and \( Y \) as

\[
K_{MV}(X, Y) = \sum_{\alpha \in X} \sum_{\beta \in Y} K(\alpha^{2D}_{X}, \beta^{2D}_{Y})
\]

where \( \alpha^{2D}_{X} \) and \( \beta^{2D}_{Y} \) are \( d \times k \) submatrices of \( X \) and \( Y \) and \( K(\alpha^{2D}_{X}, \beta^{2D}_{Y}) \) is a kernel function defined for measuring similarity between two submatrices. One possible definition for the submatrix kernel \( K(\cdot, \cdot) \) is cumulative row-based comparison

\[
K(\alpha^{2D}_{X}, \beta^{2D}_{Y}) = \sum_{i=1}^{d} I(\alpha^{2D}_{X}_{i}, \beta^{2D}_{Y}_{i})
\]

where \( I(\cdot, \cdot) \) is some kernel function, e.g., the linear kernel.

\[pkuksa@nec-labs.com\]
where $I(\cdot, \cdot)$ is a similarity/indicator function for matching 1D rows $\alpha_k^D$ and $\beta_k^D$. While typical string kernel methods assume Euclidean feature space and use Euclidean distance, a probabilistic manifold assumption can be more natural and effective (see e.g. [17, 4]). For example, under a multinomial manifold, the kernel/distance $K_{\text{manifold}}(X, Y)$ of two vectors $\Phi(X)$ and $\Phi(Y)$ could correspond to Bhattacharyya affinity measure between two probability distributions, i.e.

$$K_{\text{manifold}}(X, Y) = \sqrt{\Phi(X)} \cdot \sqrt{\Phi(Y)}$$  \hspace{1cm} (3)

**Results.** We test proposed methods on a number of multi-class time series classification tasks: (1) 10-class music genre classification\(^1\), (2) 6-class music genre recognition (ISMIR contest\(^2\)), and (3) 20-class music artist identification (artist20 dataset\(^3\)). For all tasks input sequences are sequences of 13-dimensional MFCC vectors.

We compare with a number of other state-of-the-art methods for time series including multivariate auto-regressive models [14], multilinear models [13], as well as methods with more problem-specific and sophisticated features (aggregate AdaBoost [1], classifier fusion with rich spectral and cepstral features [8], non-negative matrix factorization-based approaches [3]).

We use state-of-the-art spectrum/mismatch [9] and spatial (SSSK) [7] kernels as our basic 1D similarity kernels $K(\cdot, \cdot)$. We compare the proposed method to the above approaches specifically developed for music classification that also use many other features in addition to MFCC. For example, using DFQ similarity kernels with SSSK [7] achieves significantly higher accuracy of 90.7% compared to only 70% when using 1D kernels. We observe similar overall improvements for DFQ-MVSK similarity kernels on another benchmark dataset (ISMIR genre contest), Table 2.

**Music genre recognition.** As shown in Table 1, on a widely used benchmark dataset [11] (10 genres, each with 100 sequences), DFQ-MVSK kernels improve over traditional 1D kernels and other state-of-the-art methods, including DWC [11], aggregate feature AdaBoost [1], approaches specifically developed for music classification that also use many other features in addition to MFCC. For example, using DFQ similarity kernels with SSSK [7] achieves significantly higher accuracy of 90.7% compared to only 70% when using 1D kernels. We observe similar overall improvements for DFQ-MVSK similarity kernels on another benchmark dataset (ISMIR genre contest), Table 2.

**Artist recognition.** We also illustrate the utility of our DFQ method on multi-class artist identification on the standard artist20 dataset with 20 artists, 6 albums each (1413 tracks total). Table 3 lists results for 6-fold album-wise cross-validation with one album per artist held out for testing. Using 2D similarity kernels with direct uniform quantization of MFCC features yields a much lower 25.7% error compared to 42.9% for 1D kernels.

### References


---

\(^1\)http://opih.cs.uvic.ca/sound/genres

\(^2\)http://ismir2004.ismir.net/genre_contest/index.htm

\(^3\)http://labrosa.ee.columbia.edu/projects/artistid/