Efficient time series classification with Multivariate similarity kernels

Pavel P. Kuksa

NEC Laboratories America, Inc
pkuksa@nec-labs.com

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,

$$\mathbf{x} = (x_1, x_2, \ldots, x_n), x_i \in \mathcal{R}^d \ \forall i$$

is transformed into a *univariate discrete* sequence of codebook IDs

$$\mathbf{c}(x) = (c_1, c_2, \ldots, c_n), c_i \in \{1 \ldots D\} \ \forall i$$

by mapping 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 $\mathbf{c}(x)$ 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 Σ 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 a family of 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 $X = (x_1, x_2, \ldots, x_n), x_i \in \mathcal{R}^d$ 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^j_{min}, f^j_{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 $\delta = (f_{max} - f_{min})/B$ and the index of quantized feature value $Q(f) = \lfloor (f - f_{min})/\delta \rfloor$ 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 dicretization levels.

Original multivariate time series X

	t=1	t=2	t=3	t=4	t=5
dim 1	0.43	1.43	3.79	2.53	3.29
dim 2	-0.34	0.91	2.97	1.68	2.12
dim 3	-0.41	0.40	2.22	1.15	1.74

Discrete multivariate representation (DFQ(X), B=64 bins)

	t=1	t=2	t=3	t=4	t=5
dim 1	37	38	40	39	39
dim 2	20	21	22	21	21
dim 3	31	32	33	32	33

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

	t=1	t=2	t=3	t=4	t=5
dim 1	309	173	484	1148	1252

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

$$K_{MV}(X, Y) = \sum_{\alpha_{2D} \in X} \sum_{\beta_{2D} \in Y} \mathcal{K}(\alpha_{2D}, \beta_{2D}) \tag{1}$$

where α_{2D} and β_{2D} are $d \times k$ submatrices of X and Y and $\mathcal{K}(\alpha_{2D}, \beta_{2D})$ is a kernel function defined for measuring similarity between two submatrices. One possible definition for the submatrix kernel $\mathcal{K}(\cdot, \cdot)$ is cumulative *row-based comparison*

$$\mathcal{K}(\alpha_{2D}, \beta_{2D}) = \sum_{i=1}^{d} I(\alpha^i_{2D}, \beta^i_{2D}) \tag{2}$$

Table 1: Music genre recognition (10-class, 13-dim. MFCC features only)

method	Error, %	F1
Baseline 1: MFCC (Mean+Variance)	48.30	51.55
Baseline 2: NMF [3]	26.0	-
Baseline 3: (non-MFCC): DWCH [11]	21.5	-
Baseline 4: MAR (multivariate autoregressive model) [14]	21.7	-
Baseline 5: AdaBoost (MFCC,FFT,LPC,etc) [1]	17.5	-
Baseline 6: Classifier fusion (MFCC,NASE,OSC) [8]	9.4	-
VQ Spectrum (k=1)	34.5	65.61
VQ Mismatch-(k=5,m=2)	32.6	67.51
VQ SSSK-(k=1,t=2,d=5)	31.1	69.08
DFQ Spectrum-(k=1)	23.0	77.23
DFQ SSSK-(k=2,t=3,d=5)	**12.3**	**88.19**
DFQ SSSK-(k=2,t=3,d=5)+FFT64	**9.3**	**91.05**

Table 2: Music genre recognition (ISMIR contest, 6-class, 13-dim. MFCC features only)

method	Error, %	F1
Baseline 1: Multilinear (Cortical) [13]	19.05	-
Baseline 2 (best): NMF [3]	**16.5**	-
VQ Spectrum (k=1)	24.15	68.99
VQ Manifold Spectrum-(k=1)	19.62	76.17
DFQ Manifold Spectrum-(k=3)	**16.74**	**80.57**

Table 3: Music artist identification (20-class, 13-dim. MFCC features only)

method	Error, %	F1
VQ Spectrum (k=1)	42.97	57.26
VQ Manifold Spectrum-(k=6)	34.62	66.22
DFQ Manifold Spectrum-(k=6)	**25.67**	**74.79**

where $I(\cdot,\cdot)$ is a similarity/indicator function for matching 1D rows α_{2D}^i and β_{2D}^i. While typical string kernel methods assume Euclidean feature space and use Euclidean distance, a *probabilistic manifold assumption* on the geometry of the data space could be more natural and effective (see e.g. [17, 4]). For example, under a *multinomial manifold* assumption, the kernel/distance $K_{manifold}(X,Y)$ between $\Phi(X)$ and $\Phi(Y)$ could correspond to Bhattacharyya affinity measure between two probability distributions , i.e.

$$K_{manifold}(X,Y) =< \sqrt{(}\Phi(X)), \sqrt{(}\Phi(Y)) > \tag{3}$$

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 autoregressive 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 approaces [3]).

We use state-of-the-art spectrum/mismatch [9] and spatial (SSSK) [7] kernels as our basic 1D similarity kernels $\mathcal{K}(\cdot,\cdot)$.

Music genre recogntion. 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 DWCH [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] James Bergstra, Norman Casagrande, Dumitru Erhan, Douglas Eck, and Balázs Kégl. Aggregate features and adaboost for music classification. *Mach. Learn.*, 65:473–484, December 2006.

[2] William M. Campbell. Generalized linear discriminant sequence kernels for speaker recognition. In *Acoustics, Speech, and Signal Processing (ICASSP), 2002 IEEE International Conference on*, volume 1, pages I–161 –I–164, may 2002.

[3] A. Holzapfel and Y. Stylianou. Musical genre classification using nonnegative matrix factorization-based features. *Audio, Speech, and Language Processing, IEEE Transactions on*, 16(2):424 –434, feb. 2008.

[4] Tony Jebara, Risi Kondor, and Andrew Howard. Probability product kernels. *J. Mach. Learn. Res.*, 5:819–844, December 2004.

[5] Pavel Kuksa, Pai-Hsi Huang, and Vladimir Pavlovic. Scalable algorithms for string kernels with inexact matching. In *NIPS*, 2008.

[6] Pavel P. Kuksa, Imdadullah Khan, and Vladimir Pavlovic. Generalized similarity kernels for efficient sequence classification. In *SDM*, 2012.

[7] Pavel P. Kuksa and Vladimir Pavlovic. Spatial representation for efficient sequence classification. In *ICPR*, 2010.

[8] Chang-Hsing Lee, Jau-Ling Shih, Kun-Ming Yu, and Hwai-San Lin. Automatic music genre classification based on modulation spectral analysis of spectral and cepstral features. *Multimedia, IEEE Transactions on*, 11(4):670 –682, june 2009.

[9] Christina Leslie and Rui Kuang. Fast string kernels using inexact matching for protein sequences. *J. Mach. Learn. Res.*, 5:1435–1455, 2004.

[10] Christina S. Leslie, Eleazar Eskin, Jason Weston, and William Stafford Noble. Mismatch string kernels for SVM protein classification. In *NIPS*, pages 1417–1424, 2002.

[11] Tao Li, Mitsunori Ogihara, and Qi Li. A comparative study on content-based music genre classification. In *SIGIR '03*, pages 282–289, New York, NY, USA, 2003. ACM.

[12] Zhiwu Lu and H.H.S. Ip. Image categorization with spatial mismatch kernels. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 0:397–404, 2009.

[13] Ioannis Panagakis, Emmanouil Benetos, and Constantine Kotropoulos. Music genre classification: A multilinear approach. In *ISMIR'08*, pages 583–588, 2008.

[14] Halfdan Rump, Shigeki Miyabe, Emiru Tsunoo, Nobutaka Ono, and Shigeki Sagayama. Autoregressive mfcc models for genre classification improved by harmonic-percussion separation. In J. Stephen Downie and Remco C. Veltkamp, editors, *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands, August 9-13, 2010*, pages 87–92. International Society for Music Information Retrieval, 2010.

[15] Sören Sonnenburg, Gunnar Rätsch, and Bernhard Schölkopf. Large scale genomic sequence SVM classifiers. In *ICML '05*, pages 848–855, New York, NY, USA, 2005.

[16] S. V. Vishwanathan, Alexander J. Smola, and René Vidal. Binet-cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes. *Int. J. Comput. Vision*, 73(1):95–119, June 2007.

[17] Dell Zhang, Xi Chen, and Wee Sun Lee. Text classification with kernels on the multinomial manifold. In *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '05, pages 266–273, New York, NY, USA, 2005. ACM.

[1] http://opihi.cs.uvic.ca/sound/genres

[2] http://ismir2004.ismir.net/genre_contest/index.htm

[3] http://labrosa.ee.columbia.edu/projects/artistid/