Circularity in neural computation and its application to musical composition

Georg Hajdu
Center of New Music and Audio Technologies (CNMAT)
University of California
1750 Arch Street, Berkeley 94709
hajdu@cnmat.berkeley.edu

Abstract

Backpropagation neural networks and their simulations can be useful for composers and theorists. Yet, their usefulness and general applicability depends mainly on the nature of the training data presented to the network. Certain musical parameters are easier to formalize than others, and in the case of parameters that exhibit a circular characteristic like pitch or key space--which have been represented as a winding helix (Shepard) or as a torus (Krumhansl&Kessler)--trigonometric functions can provide the means to prepare the training sets.

Introduction

In this paper, I shall describe two environments in which the MAX multi-layer perceptron object, developed by Michael Lee at CNMAT, was used to control circular musical parameters:

Firstly, an interactive model of key space in which a particular location on the torus is transformed into tonal hierarchies. These hierarchies, in turn, serve as the basis for a stochastic process generating melodies in a given key. Movement on the surface of the torus leads to a change of tonal hierarchies perceived as modulation.

Secondly, a compositional environment I used in my opera in progress "Der Sprung." Motivated by the text, a set of networks were trained to different melodies and used for the subsequent melodic and metric interpolation process.

The perceptron training file format requires the configuration of the net to be specified as well as the number of training patterns. A typical pattern includes data for both the input and the desired output associated with it.

Key Space

The most convenient way to establish a graphical model of the toroidal key space is to project it onto the plane of the computer screen:
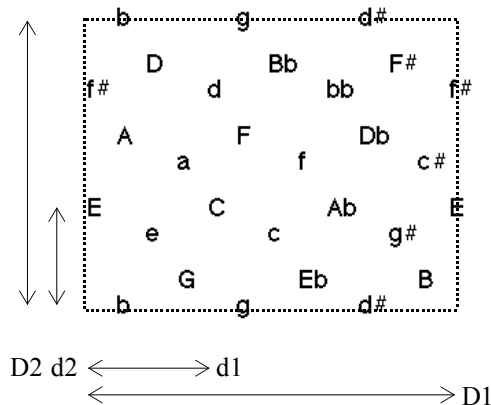
A unique position on the surface of the torus usually represented in terms of two angles A1 and A2 with respect to a reference point (lower left-hand corner) can be expressed as:

$$A1 = d1/D1 * 2\pi \quad A2 = d2/D2 * 2\pi$$

Translating A1 and A2 into training data can be achieved by taking their sine and cosine

$$x1= \sin(A1) \qquad x2= \cos(A1) \qquad y1= \sin(A2) \qquad y2= \cos(A2),$$

before dividing the variables by 2 and adding 0.5 to them to match the input range of the MAX perceptron object (0.-1.):

$$x'1= x1/2+0.5, \quad x'2= x2/2+0.5, \quad y'1= y1/2+0.5, \quad y'2= y2/2+0.5$$

According to Krumhansl&Kessler, a key area is characterized by a unique location on the torus and a particular pitch profile representing tonal hierarchies. For the key space model programmed in MAX, the perceptron (4 input units, 12 hidden units in 1 hidden layer, 12 output units) was trained to 24 patterns for the 24 major and minor key areas. Each pattern consists of 4 input values for the location of a particular key area and 12 output values for its respective profile (pitches in chromatic order). For example, the following line shows the pattern for C major:

x'1	x'2	y'1	y'2	c	c#	d	d#	e	f	f#	g	g#	a	a#	b
0.2500	0.0669	0.2500	0.0669	1.0000	0.0000	0.3033	0.0242	0.5218	0.4514	0.0703	0.7184	0.0387	0.3471	0.0144	0.1576.

After the training period, the network performs the task of interpolating between the profiles (perceived as modulation).

Melodic interpolation

Melodic interpolation is far from being a trivial problem (Polansky). The intervallic space in which interpolation is performed is characterized by at least two distinctive properties: pitch distance (magnitude) and harmonic distance. While pitch distance is usually a function of the frequency ratio of the two fundamentals, harmonic distance is more complicated to formalize and depends a great deal on intuition and experience.
I found that the circle of fifths represents a viable compromise between the two principles, in which the circularity of the chroma circle is preserved and harmonic relationships are allowed to come into play.

Enharmonic equivalence

The following method for melodic interpolation is pulse oriented and suggests the use of two separate networks for the melodic and rhythmic domains. Theoretically, the number of possible "source melodies" can be as large as desired. Nevertheless, for the MAX perceptron object it should be kept between 2 and 12. The training patterns for the "melodic" network consists of m values serving as indices for the m numbers of melodies involved, e.g. 0. 1. 0. for the second of three melodies. The output values for n pulses indicate the position of a particular pitch on the circle of fifths. For a closed circle of fifths with enharmonic equivalence, two values per pulse are necessary, as opposed to an open circle where one value is sufficient. In the first case, one value is taken from the circle based on C, the other from the circle a tritone below (the distances are determined in clockwise direction):

$$a_{1,n}=(D_c(T_1(n)))/12, \qquad b_{1,n}=(D_{Gb}(T_1(n)))/12$$
$$a_{2,n}=(D_c(T_2(n)))/12, \qquad b_{2,n}=(D_{Gb}(T_2(n)))/12$$
$$\ldots\ldots \qquad\qquad\qquad \ldots\ldots$$

$D_k(l)$ = Distance of pitch class l on the clockwise circle of fifths with respect to pitch class k,
e.g. $1 = D_c(G)$.
$T_m(n)$ = Pitch class of melody m for pulse n.
An additional $c_{m,n}$ value would be required if also taking octave register into consideration.

After interpolation, the resulting values $(a_{1.2...,n}, b_{1.2...,n})$ are subject to further interpretation. The smaller of the two values is sampled and an offset is added to it (offset$_a$=0, offset$_b$=0.5):

$$v= a_{1.2...,n}, \text{ if } a_{1.2...,n} < b_{1.2...,n}; \quad v= b_{1.2...,n}+0.5, \text{ if } a_{1.2...,n} \geq b_{1.2...,n}$$

The pitch class is determined by $T_{1.2}(n) = I_c(\text{Int}(V*12+.5)) \bmod 12$.
I_k=Inverse function of $D_k(l)$, e.g. $G=I_c(1)$.

In my opera *Der Sprung*, enharmonic equivalence was not a desired feature. Therefore, the $b_{m,n}$ values could be disposed of. Instead, I used an open progression of fifths spanning from eb to g#. The following measure shows an example for how a short motif was translated into training data (3. scene, measure 117, saxophone):

The coding for pitch on a 16th-note quantization level (the network consisted of 3 input units, 3 hidden units in 1 hidden layer, 24 output units) is as follows:

1. 0. 0. 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.00 0.42 0.42 0.42 0.75 0.5 0.42 0.17 0.17 0.17 0.17 0.17 0.17 0.75 0.75

Rhythm was dealt with in an analogous way:

1. 0. 0. 0.00 0.00 0.00 0.00 1.00 0.50 0.50 0.50 0.50 0.50 1.00 0.50 1.00 1.00 1.00 1.00 1.00 0.50 0.50 0.50 0.50 0.50 1.00 0.50

Here, the values starting after the third number indicate the *sound status* $s_{m,n}$ of a particular pulse.

$s_{m,n}$ =1 if pulse is attacked,
=.5 if pulse is sustained,
=0 if pulse is silent.

After interpolation an intermediate value (.33 - .67) can still lead to an attack if preceded by a value smaller than .33.

The actual interpolation involved three melodies (H, S and R) located at equal distance in the corners of a triangular melodic space and was performed over a total period of three minutes. The following graph shows the interpolation path at a resolution of 1000 sample points. For the graphical plot, another network was trained to translate the input data into spatial coordinates.
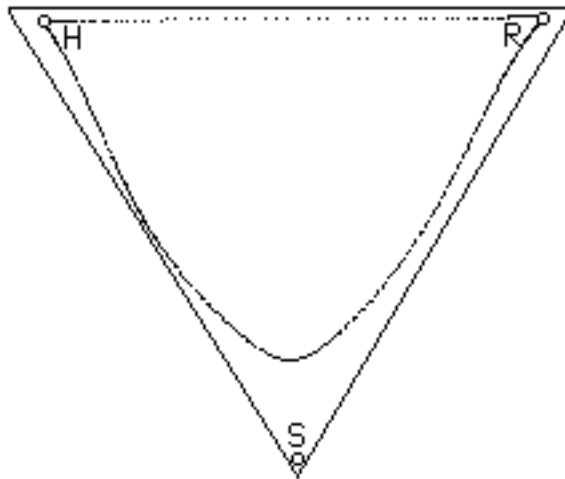
References

Hajdu, G. (1994). *Der Sprung - First Act*. Ph.D. Thesis, UC Berkeley.

Krumhansl, C.L. & Kessler, E.J. (1982). Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical key. *Psychological Review*, *89*, 334-68.

Krumhansl, C. (1990). *Cognitive foundations of musical pitch*. Oxford University Press.

Lee, M., Freed, A. & Wessel, D. (1991). Real-time neural network processing of gestural and acoustic signals. *Proceedings of the ICMC* (Montreal).

Polansky, L. (1992). More on morphological mutation functions. Recent techniques and developments. *Proceedings of the ICMC* (San Jose), 57-60.

Shepard, R.N. (1964). Circularity in judgments of relative pitch. *Journal of the Acoustical Society of America*, *36*, 2346-53.