

MATCH – A MUSIC ALIGNMENT TOOL CHEST

Paper by: Simon Dixon and Gerhard Widmer

Kamlesh Lakshminarayanan

USC EE 675 / ISE 575 • Feb 2, 2010

Prof Elaine Chew

What is MATCH ?

- Tool kit for accurate automatic aligning of audio recordings of different renditions of the same piece of music.
- Analyse structure.
- Study relationship between tempo and phrase.

How does it work

- Dynamic time warping (DTW) – Measure similarity of sequences that may vary in time or speed.
- DTW is applied in video, audio, graphics and speech recognition.

Basic Optimization

- Setup cost function.
- Minimize it . Get the "optimal" solution.
- Ideal zero cost "match" situation.
- Penalized for a mismatch.

Basic definitions

- Cost function - function of input prices and output quantity. Its value is the cost of making that output given those input prices.
- A common form:
- $c(w_1, w_2, y)$ is the cost of making output quantity y using inputs that cost w_1 and w_2 per unit.

- Minimum cost path – $W = W_1, W_2, \dots$ where each W is an ordered pair (i, j) such that (i, j) is an element of W means that points U_i and V_j are aligned.
- Alignment done with respect to a local cost function $d(i, j)$ which is in matrix form. It assigns a match cost for aligning each pair (u, v)

contd..

- PATH COST $D(W)$ – Sum of local match costs along the path.
- Minimum cost path – Dynamic programming using recursion.

$$D(i,j)=d(i,j)+\min(\text{neighbours})$$

Path constraints

- Local and global path constraints. Prevent pathological solutions and reduce search space
- Sakoe-chiba bound- path lies between fixed distance from the diagonal
- Weight of 2 given for diagonal steps to avoid biases.

Algorithm

- Forward path estimation algorithm – center of the cost matrix.
- Use of adaptive diagonal to lessen the risk of missing the optimal solution and greater system efficiency.
- Active area of matrix – top row , right column.

- Minimum cost path of each cell is evaluated and the one with the lowest minimum cost path indicates the direction of the optimal path.
- Cell in top right implies algorithm is on target.
- Matrix expanded in appropriate direction if target is missed.

Implementation specs for audio frames

- Spectral representation – FFT
- Hamming window- 2048 point
- Linear log frequency- priority for lower frequency components.

Audio frame

- Audio frame representation uses a half wave rectified first order difference.
- Only increase in energy in each frequency bin is taken into account
- Positive spectral difference vectors are compared using euclidean distance.

Testing and Results

- Bosendorfer computer monitored piano
- BeatRoot data
- From other types of music

Piano Data

- 22 piano artists, 2 excerpts from chopin
- Etude – 70 to 95 sec
- Ballade – 112 to 150 sec
- Alignment performed on all pairs of performances.
- Time alignment difficulty – difference in notes and asynchronies in chords.

Score event ?

- Set of simultaneous notes according to score.
- Calculate $t(i,e)$ - the average onset time.
- I-interpretation. E-score event.
- Alignment defined in terms accuracy of mapping "e" from one "i" to another.
- For each "e" path should cross through $(t(i_1,e), t(i_2,e))$.
- Error-manhattan distance from here to point on alignment path.

MATCH - Software

- Media player like GUI.
- Location of second file and the corresponding location in the first file.
- Display forward and backward cost matrix

contd...

- Java 1.5
- Linux 3 Ghz
- Time taken – 4% of sum of durations
- Time resolution-20 ms

Other Data

- Pieces where beat is annotated using the beat tracking software BeatRoot.
- Classical guitar solos , orchestra performances,piano concerts.