Paper ID: 5014

DON'T LOOK BACK: AN ONLINE BEAT TRACKING METHOD USING RNN AND ENHANCED PARTICLE FILTERING

Mojtaba Heydari and Zhiyao Duan
Audio Information Research Lab, University of Rochester

The 46th IEEE International Conference on Acoustics, Speech & Signal Processing. (ICASSP 2021)

Motivations / Contributions

Motivations:

- **Online beat tracking (OBT)** is one of the most essential MIR tasks with many applications
- **Causality conditions** and **real-time inference**, make performance of the current OBT models significantly lower than the offline methods.
- The most recent researches cover offline beat tracking so that the state-of-the-art online method backs to 2010! without necessarily promising performance.
- Most OBT models apply a **moving window trick** to existing offline models and are not fundamentally online.
- Moving window approach causes some problems such as the initial window requirement and discontinuity.
- Previous filtering-based models use traditional onset strength signal which is not as promising as DNN based activation function.

Contributions:

- We proposed the first DNN based **particle filtering** model for online beat tracking.
- Our model significantly improves online beat tracking accuracy over the state-of-the-art OBT methods.
- **Don't Look Back (DLB)** provides an immediate beat tracking response, which is critical for many OBT applications.
- The model is heavily optimized, leading to promising results for greedy filtering setups (with a few number of particles).

Model Structure

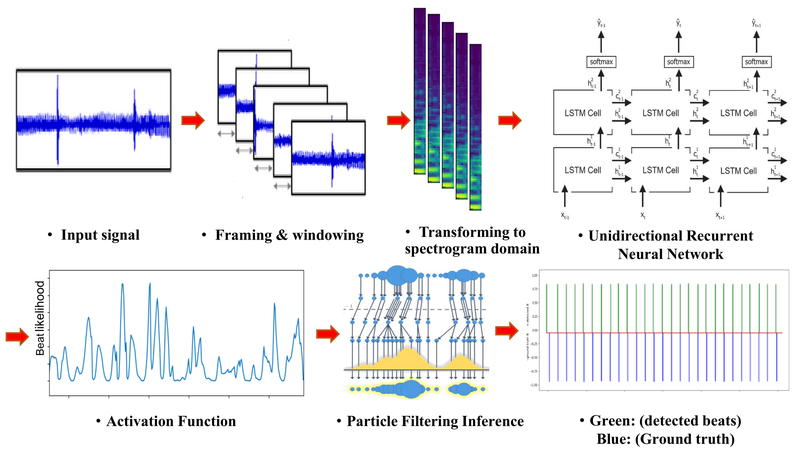

Fig.1: The structure of the proposed model containing preprocessing, beat activation extraction and particle filtering inference

The State Space Model

❖State Space :

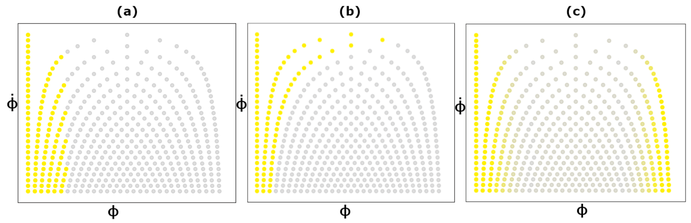

Fig.2: State space and three different beat/non-beat boundaries setup. Φ reperesensts the frame's phase and $\dot{\Phi}$ is the frame's pace

❖Observation Model:

$$p(y_k|x_k) = \begin{cases} b_k & x_k \in B \\ \gamma & x_k \notin B \end{cases} \qquad (1)$$

Where B is beat states, b_k is the beat activation (the output of the softmax) and γ is a small constant number (i.e. $\gamma = 0.03$)

❖Transition Model:

$$p(\dot{\Phi}_k|\dot{\Phi}_{k-1}) = \begin{cases} exp\left(-\left|\lambda\right| \times \frac{\dot{\Phi}_k}{\dot{\Phi}_{k-1}} - 1\right) & for\ x_k \in B \\ \dot{\Phi}_k = \dot{\Phi}_{k-1} & for\ x_k \notin B \end{cases}, \qquad (2)$$

Where λ is a parameter to control the width of resampling area for the tempo fluctuations

Particle Filtering Inference Formulation

According to the importance sampling principle, a high dimensional probability distribution $p(x) = p'(x)/C$ with an unknown model and parameters, and normalization constant C, can be represented by a large number (N) of independent samples $x^{(i)}$ from a known arbitrary proposal distribution $\pi(x)$.

$$p(x) = \frac{p\prime(x)}{C} = \frac{p\prime(x)}{C\,\pi(x)}\pi(x) = \frac{p\prime(x)}{C\,\pi(x)} \lim_{N\to\infty} \frac{1}{N}\sum_{i=1}^{N} \delta(x - x^{(i)})$$

$$= \lim_{N\to\infty} \sum_{i=1}^{N} \frac{\omega_i}{\sum_{i=1}^{N} \omega_i}\, \delta(x - x^{(i)}) \qquad (3)$$

Where $\omega_i = p'(x^{(i)})/\pi(x^{(i)})$is each particle i's importance weight.

$$\omega_K^{(i)} = \frac{p'(x_{0:K}^{(i)}|y_{0:K})}{\pi(x_{0:K}^{(i)}|y_{0:K})} = \frac{p(y_K|x_K^{(i)})\, p(x_K^{(i)}|x_{K-1}^{(i)})}{p(x_K^{(i)}|x_{K-1}^{(i)})}\, \omega_{K-1}^{(i)} \Rightarrow$$

$$\omega_K^{(i)} = p\left(y_K|x_K^{(i)}\right)\ \omega_{K-1}^{(i)} \qquad (4)$$

Algorithm and Performance

❖ Algorithm:

Initializing *the particle uniformly across all state space and loop through following steps:*

I. *Sample particles from proposal distribution which in our case is* ***transition probability****. (****motion****)*
II. *Compute the new importance weights (4) based on observation probability derived from* ***LSTM Beat activations****.*
III. *Resample based on* ***new importance weights*** *to discard unlikely hypotheses and remain more likely ones. (****correction****)*
IV. *Take the* ***median*** *of the positions of all particles. Classify the frame as beat if it is within the beat boundary and is far from previously detected beat with a threshold.*

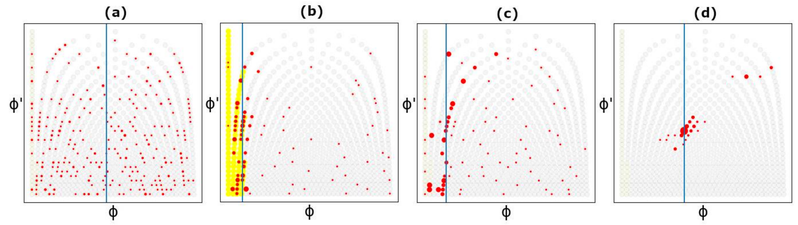

Fig.3: The performance of the system from initializing randomly to convergence to the frame location through the beat interval.

Evaluation

Table.1: F-measure report of the online/offline inferences with and initialization time for online models (GTZAN dataset)

Method	*F-measure (5 Sec. skip)*	*F-measure (without skip)*	*Initialization elapse Time*
Online beat tracking methods			
DLB	73.77	71.44	0 seconds
ACF [1]	64.63	51.74	6 seconds
IBT [2]	68.99	62.75	5 seconds
Aubio [3]	57.09	55.91	3 consistent beat periods
Offline beat tracking methods			
CF [4]	67.97	67.74	---
DBN [5]	77.75	77.36	---

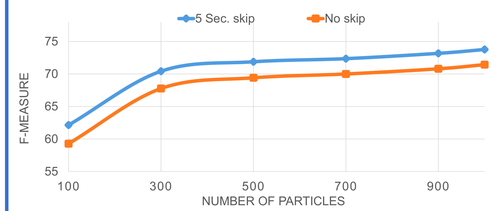

Fig.4: DLB performance for different number of particles.

Advantages :

- ✅ Significantly better performance than previous models
- ✅ No initialization requirement
- ✅ Robust against greedy filtering
- ✅ General approach (No strong assumption)
- ✅ Robust against tempo fluctuation

Drawbacks :

- ❌ lower speed comparing to some none sample-based models (Still faster that real-time)

References

1. S. Böck and M. Schedl, "Enhanced beat tracking with context-aware neural networks," in *Proc. Int. Conf. Digit. Audio Eff (DAFx)*, 2011, pp. 135–140.
2. J. L. Oliveira, F. Gouyon, L. G. Martins, and L. P. Reis, "IBT: A real-time tempo and beat tracking system," in *Proc. Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, 2010, pp. 291–296.
3. P. M. Brossier, "Automatic Annotation of Musical Audio for Interactive Applications," Ph.D. dissertation, Queen Marry University, London, UK., August, 2006.
4. S. Böck, F. Krebs, and G. Widmer, "Accurate tempo estimation based on recurrent neural networks and resonating comb filters," in *Proc. Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, 2015, pp. 625–631.
5. S. Böck, F. Krebs, and G. Widmer, "A multi-model approach to beat tracking considering heterogeneous music styles," in *Proc. Int. Soc. Music Inf. Retr. Conf. (ISMIR)*, 2014, pp. 603–608.