DEEP RANKING: TRIPLET MATCHNET FOR MUSIC METRIC LEARNING

Rui Lu[1], *Kailun Wu*[1], *Zhiyao Duan*[2], *Changshui Zhang*[1]

[1] *Department of Automation*, *Tsinghua University*

[2] *Department of Electrical and Computer Engineering*, *University of Rochester*

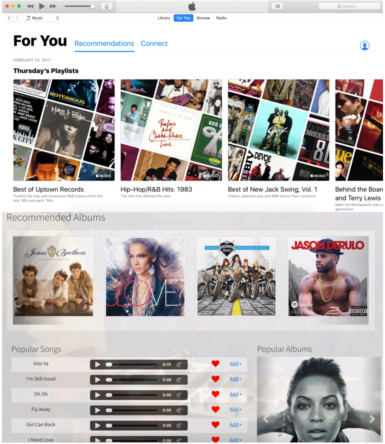
March 8, 2017

Presentation at IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)

Applications of music metric learning

Classification

Recommendation

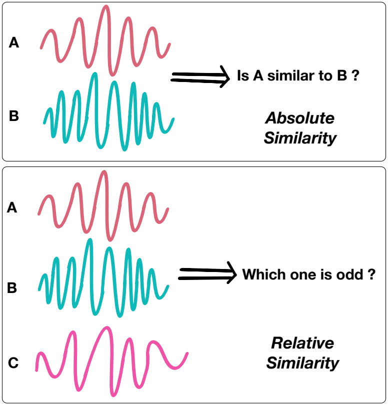


Music metric learning

Basic methods

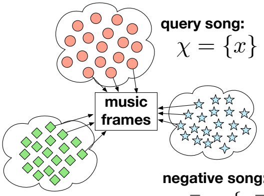
- Supervised methods
  - RITML: learns mahalanobis distance
  - MLR: learn to rank
  - SVM-based
  - ...
- Unsupervised methods
  - Mahalanobis distance
  - PCA
  - ...

Similarity



From frame-level to song-level

Training process

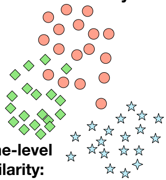


query song: $\chi = \{x\}$

negative song: $\chi^- = \{x^-\}$

positive song: $\chi^+ = \{x^+\}$

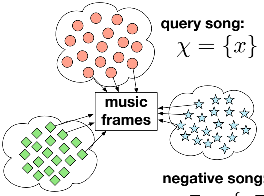
Train with relative similarity



Frame-level similarity:

$$d(x, y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

Testing process



query song: $\chi = \{x\}$

negative song: $\chi^- = \{x^-\}$

positive song: $\chi^+ = \{x^+\}$

$$d(x, y) : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$$

Frame-level similarity:

$$\{d(x, x^+) | \forall x \in \chi, \forall x^+ \in \chi^+\}$$
$$\{d(x, x^-) | \forall x \in \chi, \forall x^- \in \chi^-\}$$

Majority-vote

Song-level similarity

Traditional and deep approaches

Traditional methods

- Handcrafted song-level features
- Linear projections

Deep learning approaches

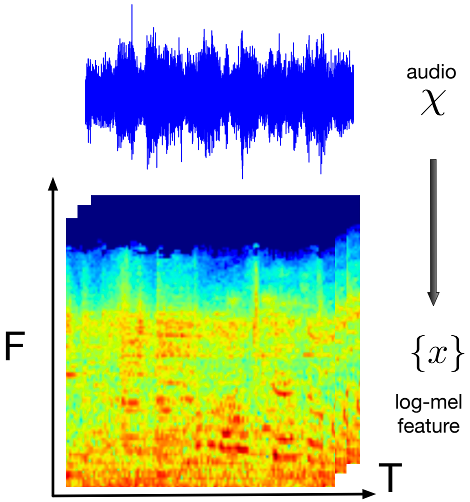
- Learn features automatically
- Highly nonlinear transformations
- Success in various domains
- None for music metric learning

Our approach

Use deep neural networks to learn frame-level relative similarities of music

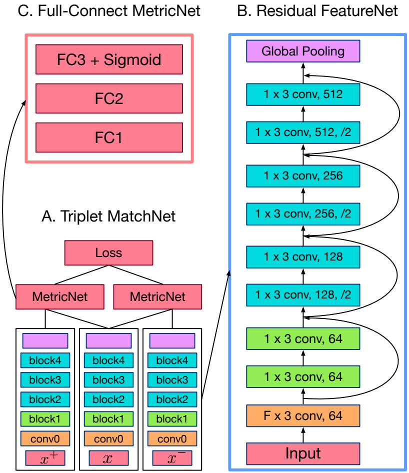
Method

Data preprocessing



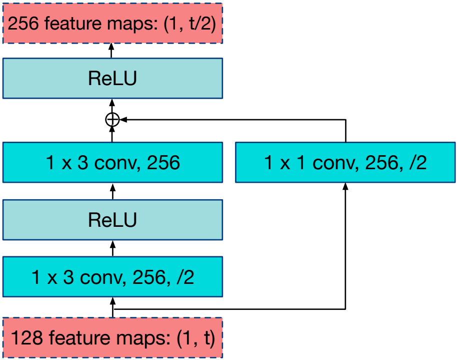
Triplet MatchNet

Network Structure



Residual block

Residual Structure



Advantages

- Easier to optimize
- Accuracy gain from deeper model[1]
- Behave like ensembles of shallow networks[2]

---

1 Kaiming He et al, Deep residual learning for image recognition, CVPR 2016.

2 Andreas Veit et al, Residual networks behave like ensembles of relatively shallow networks, NIPS 2016.

Final loss

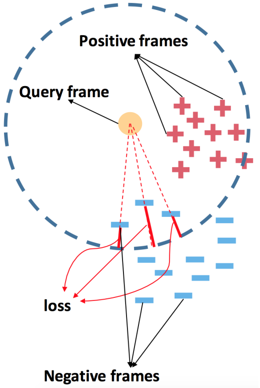
The final loss for training our Triplet MatchNet is:

$$loss(\chi, \chi^+, \chi^-) = \frac{1}{|\{x\}|} \sum_{x \in \{x\}} (\psi(x) + \phi(x)).$$

Where $\psi(x)$ is the rank-based loss; $\phi(x)$ is the contrastive loss.

Rank-based loss

Illustration



Equation

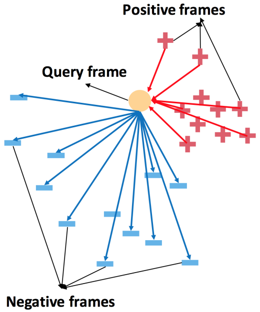
Rank-based loss

$$\psi(x) = \frac{1}{|\{x^-\}|} \sum_{x^- \in \{x^-\}} \max\{0, d^+_{max} - f(x, x^-)\}$$

where $d^+_{max} = \max_{x^+ \in \{x^+\}} f(x, x^+)$;
$f(x, y)$ is the proposed network's final output

Contrastive loss

Illustration



Equation

Contrastive loss

$$\phi(x) = -\frac{\sum_{x^+}\sum_{x^-}[\log(1-d^+)+\log(d^-)]}{|\{x^+\}||\{x^-\}|}$$

where $d^+ = f(x, x^+)$; $d^- = f(x, x^-)$

Experiments

Dataset and Evaluation

MagnaTagATune

- Relative similarity
- 860 triplets like (χ, χ^+, χ^-)
- 993 unique songs
- Each song with 29 seconds

Evaluation

- Constraints Fulfillment Rate
  - Portion of triplets that preserve partial order relationships
- 10-cross validation
- Comparison methods
  - RITML[1]
  - MLR[2], RMLR[3]
  - SVM, Euclidean

---

[1] Daniel Wolff et al, Comparative music similarity modelling using transfer learning across user groups, ISMIR 2015.

[2] Brian McFee et al, Metric learning to rank, ICML 2010.

[3] Daryl KH Lim et al, Robust structural metric learning, ICML 2013.

Constraints Fulfillment Rate comparison

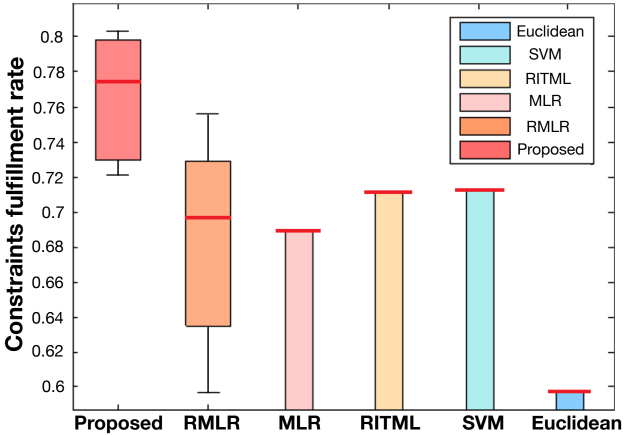


Figure: Constraints Fulfillment Rate by 10-fold cross validation.

Generalization Capability

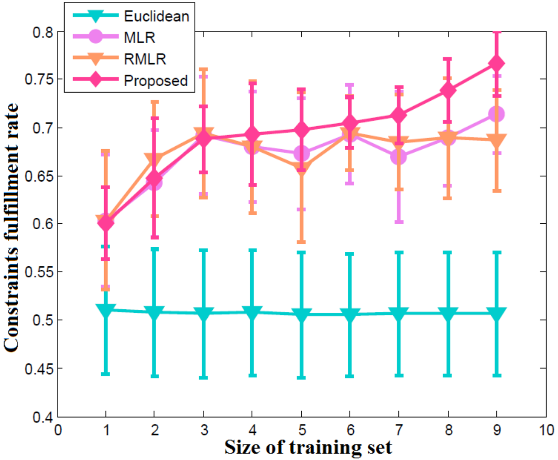


Figure: Generalization capability by 10-fold cross validation.

Better features extracted by Triplet MatchNet

Method	**HandCrafted**	**PCA**	**Proposed**
RMLR	-	65.9 ± 8.3	$\mathbf{71.2 \pm 7.2}$
MLR	68.9	61.7 ± 10.5	$\mathbf{71.7 \pm 6.9}$
Euclidean	59.8	50.7 ± 6.3	$\mathbf{70.6 \pm 3.8}$

Table: Constraints Fulfillment Rate of three baselines working with different features.

Thank you !