When Counterpoint Meets Chinese Folk Melodies

Nan Jiang†, Sheng Jin†, Zhiyao Duan‡, Changshui Zhang†

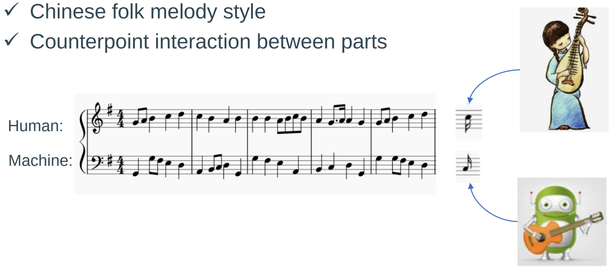
† Institute for Artificial Intelligence, Tsinghua University (THUAI),
State Key Lab of Intelligent Technologies and Systems,
Beijing National Research Center for Information Science and Technology (BNRist),
Department of Automation, Tsinghua University, Beijing, China
‡ Department of Electrical and Computer Engineering, University of Rochester

Acknowledgement:
National Key R&D Program of China (No. 2018AAA0100701),
Beijing Academy of Artificial Intelligence (BAAI),
National Science Foundation grant No. 1846184.

Introduction

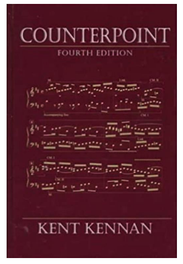
Human-machine collaborative duet improvisation.

- ✓ Chinese folk melody style
- ✓ Counterpoint interaction between parts



Task: Incorporating Western counterpoint interactions into Chinese folk melodies for online human-machine collaborative duet improvisation.

✘



- **Chinese folk melody**: typically presented in a *monophonic* form or with accompaniments that are less melodic.
- **Counterpoint**: mediation of two or more musical voices into a meaningful and pleasing whole.

Challenges

- Out-of-domain data (Chinese folk duets are scarce)

  Monophonic Chinese folk melodies + Bach chorales

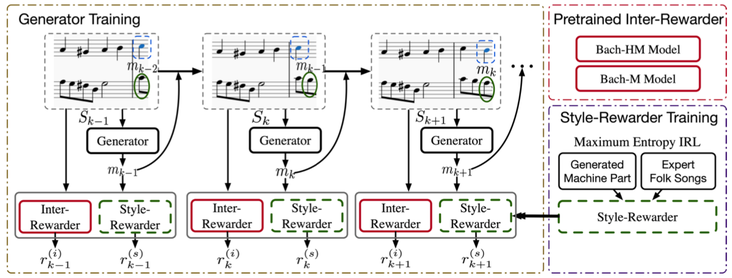
- Counterpoint pattern is coupled with western-music style patterns

  Extract counterpoint interaction pattern & eliminate Western music style

Our Solutions

- Reinforcement Learning → Design task-specific reward functions
- Measure counterpoint interaction using *mutual information*

FolkDuet



Framework of FolkDuet: a Generator and two rewarders

Inter-Rewarder: models the counterpoint interaction in Western music
Style-Rewarder: models the melodic pattern of Chinese folk melodies
The Generator is trained using reinforcement learning with these two rewards.

Rewarders

Inter-Rewarder:
measures the degree of interaction between human and machine parts through a mutual information informed measure.

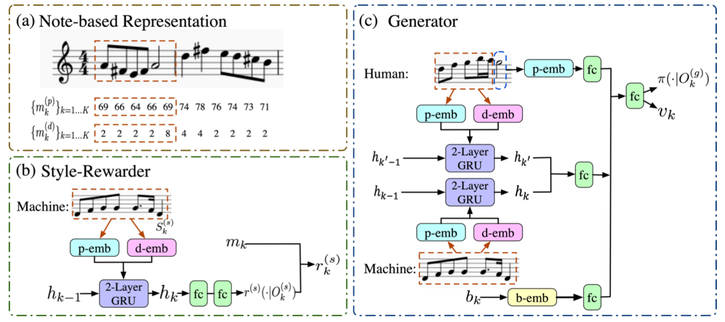
$$I(X,Y) = \sum_{X,Y} P(X,Y)\log\frac{P(X,Y)}{P(X)P(Y)}$$
$$= \sum_{X,Y} P(X,Y)\left[\log P(Y|X) - \log P(Y)\right] \approx \sum_{X_t,Y_t\sim P_{X,Y}} \left[\log P(Y_t|X_t) - \log P(Y_t)\right]$$
$$= \sum_{X_t,Y_t\sim P_{X,Y}} \left[\log \prod_{k=1}^{K^{y_t}} P(y_k^{(t)}|X_t, y_{t<t_k}^{(t)}) \cdot P(y_0^{(t)}|X_t) - \log \prod_{k=1}^{K^{y_t}} P(y_k^{(t)}|y_{t<t_k}^{(t)}) \cdot P(y_0^{(t)})\right]$$
$$= \sum_{X_t,Y_t\sim P_{X,Y}} \sum_{k=1}^{K^{y_t}} \left[\log P(y_k^{(t)}|X_t, y_{t<t_k}^{(t)}) - \log P(y_k^{(t)}|y_{t<t_k}^{(t)})\right] + C(X_t, y_0^{(t)}).$$

$\log p(Machine|Human) - \log p(Machine)$

Style-Rewarder:
Inverse Reinforcement learning (IRL):
learns to infer a reward function underlying the observed expert behavior.

Style-Rewarder is alternatively updated using the maximum entropy inverse reinforcement learning. Its learning objective is to infer the reward function that underlies the demonstrated expert behavior, *i.e.* the Chinese folk melodies.

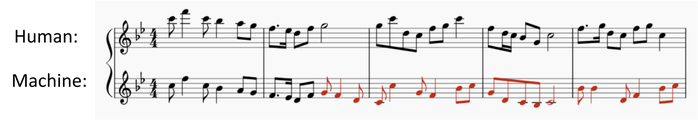
Architectures



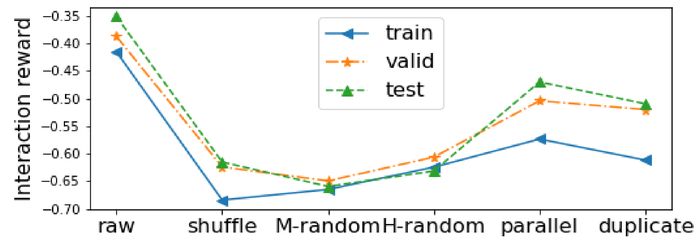
(a) The note-based representation, the network architectures of (b) Style-Rewarder and (c) Generator. p-emb, d-emb and b-emb represent pitch/duration/beat embedding modules, respectively. GRU represents the Gate Recurrent Unit, and fc stands for the fully-connected layer.

Results

Generated Duets



Can interaction reward reflect counterpoint interaction?



original bach's duets	two randomly shuffled parts	random notes in the M part	random notes in the H part	parallel H and M parts	duplicate H and M parts

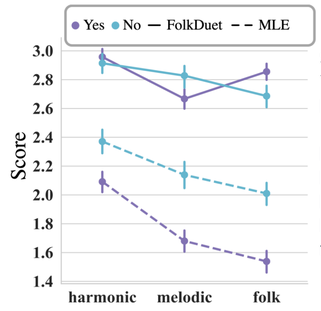
H part and M part are short for human part and machine part respectively.
It shows that this interaction reward achieves the highest score on the original duets.

Objective Evaluation

	PC/bar	PI	IOI	PCH ↓	NLH ↓	key-consist ↑	inter-reward ↑
Dataset	3.90	2.73	2.36	-	-	-	-
MLE	4.21 ± 0.12	$\mathbf{3.02 \pm 0.12}$	2.87 ± 0.10	0.017 ± 0.002	0.036 ± 0.008	0.78 ± 0.01	-0.30 ± 0.02
RL-Duet [27]	3.23 ± 0.01	4.02 ± 0.01	3.64 ± 0.02	0.017 ± 0.001	0.055 ± 0.002	0.71 ± 0.01	-0.50 ± 0.004
FolkDuet	$\mathbf{3.96 \pm 0.12}$	$\mathbf{2.44 \pm 0.14}$	$\mathbf{2.16 \pm 0.10}$	$\mathbf{0.008 \pm 0.001}$	$\mathbf{0.014 \pm 0.004}$	$\mathbf{0.85 \pm 0.01}$	$\mathbf{0.13 \pm 0.03}$

Style: Closer to Chinese folk datasets, in some statistics and distribution distance, e.g. pitch interval (PI), pitch class histogram (PCH).
Counterpoint interaction: Higher key consistency between human and machine parts, higher inter-reward.

Subjective Evaluation



Subjects background:
Yes: have music training/background before
No: have no music training/background before

Evaluation perspectives:
Harmonic: harmonic appealingness of the duets
Melodic: melodic appealingness of the generated part
Folk: prominence of the Chinese folk style of the generated part