Semantic Control of Generative Musical Attributes

Stewart Greenhill Majid Abdolshah Vuong Le
Sunil Gupta Svetha Venkatesh
Applied Artificial Intelligence Institute, Deakin University, Australia

1. Introduction

Deep generative neural networks can perform many musical tasks, such as composing melodies and accompaniments, rendering expressive performances and synthesizing singing voices, but to be a useful tool for musicians **controllability** is essential.

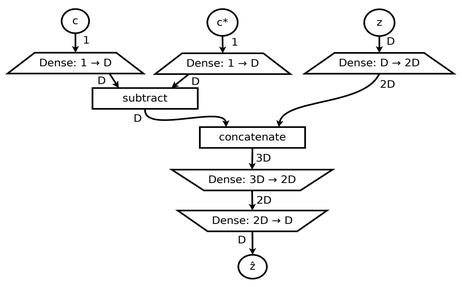
This means that it is important to be able to navigate the **latent space** of the model from which novel output is generated, and where dimensions represent high level semantic attributes such as note density, syncopation, genre, and arousal.

Previous approaches have focused on **regularisation** of the latent space, **disentanglement** of the semantic features, and use of **attribute vectors** for navigation, but this is not a complete solution because of potential non-linear relationships and "holes" in the latent space where generated output is invalid [1].

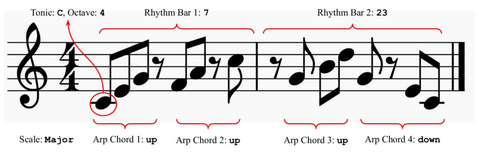
Our approach called **Semantic Neural Latent Traversal** (SeNT) uses a secondary neural network to model the relationship between latent codes and semantic attributes, allowing precise changes to musical attributes, supporting non-linear relationships and accounting for context.

2. Method

We train a neural network by showing it examples of successful changes to semantic attributes of a melody, so that it learns how to make such changes in the latent space of the generative network.



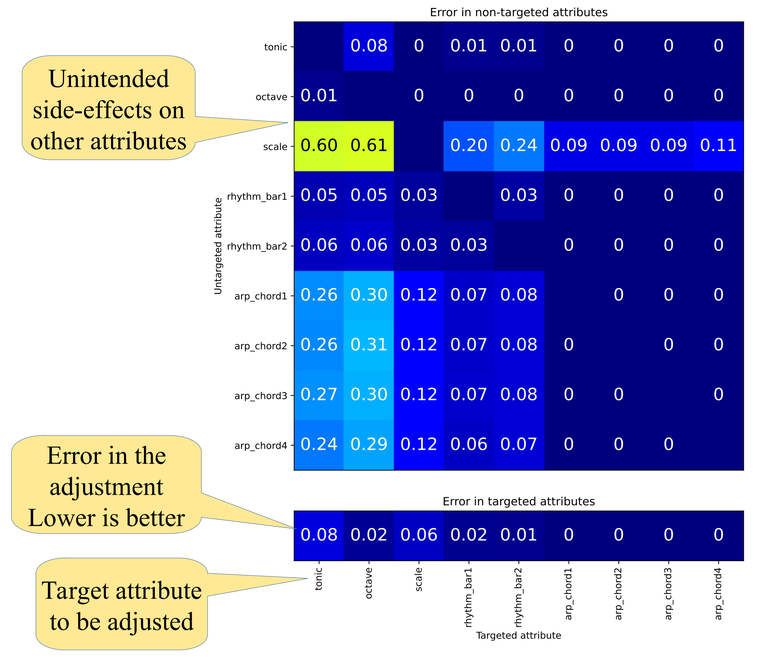
We evaluate the method using the **dMelodies** data set [2] which includes VAE generative models, and a large database of two-bar melodies with 9 attributes related to pitch and rhythm.



For a set of generated melodies, we test the ability of the method to adjust each of the attributes, by measuring how accurately it achieves the desired change, and any unwanted "side-effect" changes to other attributes of the melody.

3. Results

When adjusting an attribute of a melody, we want the value of the target attribute to be close to the target value, and the other non-target attributes to be close to their original values. This is measured through the **target error** and **non-target error** which will both be low when an attribute is successfully adjusted. This example shows some side-effects on the scale when adjusting the pitch by changing the tonic or octave.

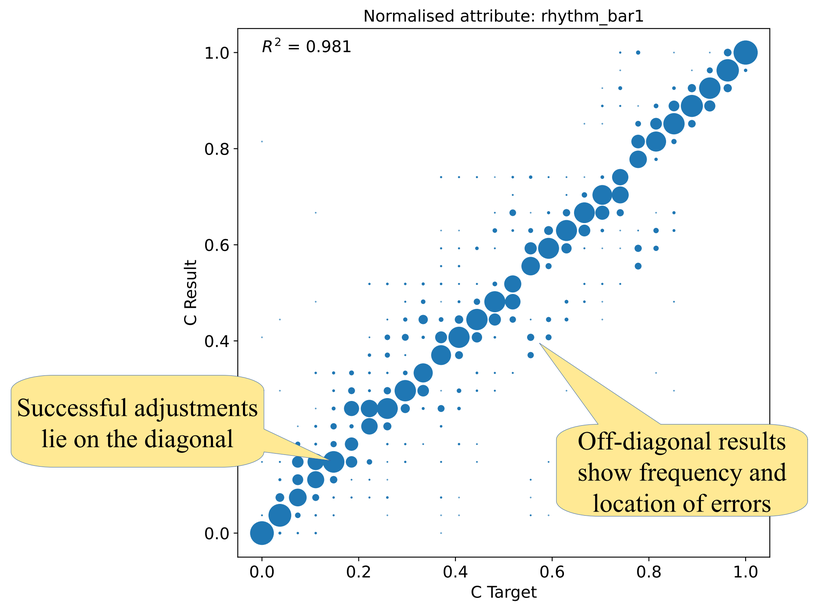


We measure the correlation R^2 between the intended target value and the achieved result. We would like this to be high, but it varies depending on the generative model and target attribute

VAE	To	Oc	Sc	R1	R2	A1	A2	A3	A4
β	.09	.78	.19	.96	.97	.11	.45	.56	.32
AR	.07	0	-1.3	.95	.97	.98	.99	1	1
I	.26	.34	.98	.99	.97	1	1	1	1
S2	.77	.93	.78	.98	.99	1	1	1	1

Correlations between target and achieved results for four different VAE models. Attributes are Tonic (To), Octave (Oc), Scale (Sc), Rhythm Bar (R) and Arp Chord (A).

Best performing models were the regularised I-VAE and S2-VAE models which are trained with supervision and have better disentanglement



4. Conclusions

The SeNT method provides way to navigate latent spaces of generative models by avoiding holes and handling nonlinear relationships. It does not replace disentanglement, and performs best when the latent space is well structured.

Attributes of the dMelodies data set which depend on pitch (Scale and Arp Chord) were found to be less stable when adjusting overall pitch via Tonic or Octave, and the definition of these attributes may be fragile to small errors in the note pitch.

Although each SeNT network controls just one attribute, more complex operations could be performed using a sequence of separate attribute changes. While we demonstrate it on VAE (Variational Auto-Encoders) generative models, it can also be used in other latent space models such as GAN (Generative Adversarial Networks)

References

[1] A. Pati and A. Lerch, "Is disentanglement enough? On latent representations for controllable music generation," in *Proceedings 22nd International Society for Music Information Retrieval Conference*, 2021.

[2] A. Pati, S. Gururani, and A. Lerch, "dMelodies: A music dataset for disentanglement learning," in *Proceedings 21st International Society for Music Information Retrieval Conference*, 2020.

Further information

For questions or comments please contact:
s.greenhill@deakin.edu.au

Data and code is available online at:
https://github.com/stewartgreenhill/sentgen

Information on our other work at: https://a2i2.deakin.edu.au