

ディフュージョンモデルを活用した作曲支援システムの検討

トランスコスモス株式会社

鎌田大輝

Kamada.Hiroki@trans-cosmos.co.jp

作曲における問題点

昨今生成AIでテキストから簡単に作曲ができるようになってきているが、作曲者をサポートするようなサービスは少なく、以前として作曲は難しい領域である。
作曲＝メロディとコード進行を作成することと定義づけたときに、メロディ作成は難易度が高い。メロディ作成には明確な理論がなく、さらには著作権があるため、真似できるものが少ない。
そのためアイデア不足に陥りやすい傾向がある。

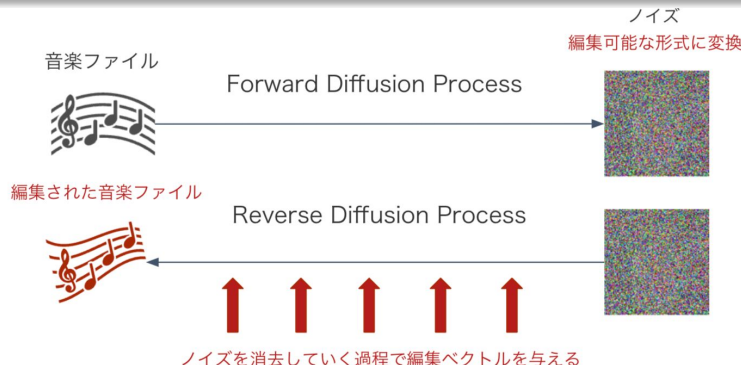
手法・ツールの適用による解決

メロディを入力したら、条件を満たしたメロディ案を複数提供してくれるシステムを作成することで、メロディアイデアが不足する問題の解決を目指す。
今回の修了制作では、ディフュージョンモデルを活用した既存の音楽生成技術を使って以下の条件を満たすようなメロディがどのくらい精度高く作れるのか確かめた。
①元のメロディから変化していること
②元のメロディの雰囲気を保っていること

今回目指すシステムのイメージ



使用技術の概要



今回の音楽生成で使用するディフュージョンモデルは上記のような形である。
入力の音楽ファイルに徐々にノイズを乗せて、編集可能な形式に変換する。その後、今度は逆にノイズを取り除いていく。その過程で、編集ベクトルを与えることでアレンジされた音楽を生成することができる。
今回与える編集ベクトルは主成分を使ったものである。入力メロディの主成分分析を行うことで音声データを次元圧縮し、その入力メロディの特徴を抽出する。その抽出した主成分を元に編集強度を変えることによって、元のメロディの特徴を保ったままの修正が行える。

入力値の整理

打ち込み (パソコンで作成)

A.メロディのみ

B.メロディ+ギターのコード進行

C.メロディ+その他演奏 (ベース、ドラム)

録音 (実際に歌った声)

D.メロディのみ

上記A~Dの4つの入力メロディとその他調整可能な入力値を整理し、どのようにしたらより①②の条件を満たすようなメロディが生成できるのかを確かめた

評価結果

メロディ	主成分	編集強度	元のメロディから変化しているか			元のメロディの雰囲気を保っているか		参考になったか?
			音の数の増減	リズム	音階	外している音※1	元のメロディと同じ箇所があったか	
A メロディのみ (打ち込み)	1	40	0	x	○	0	○	○
		100	0	x	○	1	○	○
		40	+1	○	○	1	○	○
	2	100	0	x	○	1	○	x
		40	0	x	○	0	○	x
		100	-2	○	○	0	○	○
C メロディ+ベース、ドラム	1	40	-3	○	○	0	○	○
		100	+2	○	○	0	○	○
		40	-3	○	○	0	○	○
	2	100	-1	○	○	0	○	○
		40	-1	○	x	0	○	○
		100	+2	○	○	0	○	○
D メロディのみ (録音)	1	40	0	x	x	0	-	x
		100	+2	○	x	0	○	x
		40	0	○	○	0	○	○
	2	100	-	-	-	-	-	x
		40	-1	○	x	0	○	x
		100	-1	○	○	1	○	x

結果

- 生成されたメロディのうち、**①②の条件を満たす出力が88%**となった
- 生成された打ち込みのメロディのうち、**83%が参考となった**
- 生成された録音のメロディはほとんど参考にならなかった

仮説

- 入力メロディと異なるリズムパターンを入力することでメロディのリズムを変化させられるのではないかな?**
- 正確な音階のメロディを入力することでメロディの音階を変化させられるのではないかな?**

➡ 既存技術でも条件を満たすメロディ生成が可能
さらにより質の高いメロディを生成するための仮説が得られた

サービス化にあたっての議論

ニーズはあるのか?

今回の実験結果をPoCとして、初心者作曲者2名に簡易的なアンケートを行った結果、ニッチな領域だがニーズはある可能性が高いことがわかった。

使用感に問題ないシステムが作れるのか?

作れない。現状10秒程度の音源であっても、主成分分析12時間かかってしまうという課題がある。
今後サービス化していくには処理時間を数分程度までに短縮する必要がある

今後の展望

今回生まれた仮説の検証をし、より質の高いメロディ生成が行えるのかを検証していきたい。
また処理時間の課題についても、解決策の調査など、引き続き音楽生成に関する技術的知見を深めていきたい。