Hochschule Düsseldorf
University of Applied Sciences

Fachbereich Medien
Faculty of Media

WISSENSCHAFTLICHE VERTIEFUNG

Autmoatic Music Transcription

01.10.2024 – 28.02.2025

Autor:
Benedikt Kolodziej
878007
benedikt.kolodziej@study.hs-duesseldorf.de
Medieninformatik (B. Sc.)

Betreuender Professor:
Prof. Dr. Dennis Müller
dennis.mueller@hs-duesseldorf.de

Zeitraum
28.05.2025 - xy

Inhaltsverzeichnis

Abbildungsverzeichnis

1 Einleitung

1.1 Automatic Music Transcription

Musik ist seit Jahrtausenden ein zentraler Bestandteil unserer Gesellschaft. Während etwa 2000 bis 0 v.Chr. musikalische Werke meist mündlich überliefert wurden, entwickelte sich in diesem Zeitraum auch eine Notenschrift. Diese Notenschrift ermöglichte es, Musikstücke einfacher zu erlernen und einem breiteren Publikum zugänglich zu machen. Durch die Digitalisierung erhielten Digital Audio Workstations zunehmend Einzug in die Musikproduktion, wodurch Notenblätter oft nicht mehr notwendig waren und es weniger Bedarf gab diese Lieder zu übersetzen in Notenschrift.

An dieser Stelle setzt Automatic Music Transcription (AMT) an. AMT ist ein Prozess, bei dem eine Audiospur als Input gegeben wird und diese durch Computerprogramme Notenblätter oder, was weiter verbreitet ist, MIDI-Dateien als Output wiedergeben. Dabei werden durch mehrere Prozesse die Eigenschaften der Noten, zum Beispiel Frequenz oder Lautstärke, analysiert und im Kontext des Musikstückes analysiert.

1.2 Herausforderungen und Hindernisse

Anstatt das man selber diese Lieder, alleine durchs Gehör, in Notenschrift überträgt würde diese Aufgabe eine Software für einen erledigen. Dieses Ziel ist jedoch schwer zu erreichen, da Musik mehrdimensional ist durch zum Beispiel Zeit, Tonhöhe und Polyphonie. Vor allem bei polyphonen Musikstücken haben herkömmliche Algorithmen viele Schwierigkeiten. In diesen Fällen müssen sie nämlich viele verschiedene Stimmen gleichzeitig analysieren und im späteren auch die jeweiligen Töne voneinander differenzieren und eindeutig einem Instrument zuordnen. Ein weiteres Problem ist die Individualität jedes Musikstückes. In realen Aufnahmen können leichtes Rauschen, kleine Spielfehler oder stilistische Mittel wie Vibrato auftreten, die je nach Interpreten unterschiedlich klingen. Zudem sind die meisten AMT-Modelle auf westliche Tonleiter trainiert. Dies kann zu Problemen führen, wenn man zum Beispiel arabische oder indische Musikstücke transkribieren möchte.

1.3 AMT und Künstliche Intelligenz,

Um diese Vielfalt zu bewältigen, ist ein neuer, oft genutzter Ansatz, die Nutzung von künstlicher Intelligenz und Machine Learning.

Im Gegensatz zu Algorithmen ist KI flexibler und kann sich besser einstellen auf kleine Abweichungen in Musikstücken. In vorherigen Modellen wurden meist direkt elektronische Audioaufnahmen oder MIDI-Dateien verwendet, da realitätsnahe Audioaufnahmen meist zu viele Störfaktoren haben. Durch KI kann man nun mehr auf reale Audioaufnahmen zurückgreifen und das Modell somit besser anpassen für einen realistischen Gebrauch.

Auch die Mehrdimensionalität von Musik kann KI deutlich besser bewältigen als Algorithmen. Das neuronale Netz einer KI ist mehrdimensional aufgebaut, wodurch dieses verschiedene Patterns, Stimmen oder andere Eigenschaften besser zuordnen und erlernen kann. Auf der anderen Seite müssen klassische Algorithmen diese verschiedenen Dimensionen explizit modellieren und sind nicht in der Lage, Muster selbstständig zu erkennen. Sie folgen nur dem, was zuvor vom Menschen fest einprogrammiert wurde.

Um ein AMT-Modell mit KI zu kreieren, muss man sich auch für ein KI-Modell entscheiden. Hier werden meistens Recurrent Neural Networks (RNN) oder Convolutional Neural Networks (CNN) benutzt. Da Musik keine definierbaren richtigen oder falschen Musikstücke, Notenabfolgen oder anderes hat, kann man keine Reinforcement Learning KI-Modelle nutzen. Dementsprechend braucht man auch ein zuverlässiges Datenset aus Audiodateien und deren zugehörigen MIDI-Dateien.

RNNs sind spezialisiert, um zeitliche Abläufe besser im Kontext zu verstehen. In der Musik werden viele Noten hintereinander gespielt, diese müssen harmonisch im Stück übereinstimmen. Das RNN verarbeitet die jeweiligen Sequenzen und merkt sich die Informationen der schon gespielten Noten,

um die darauffolgenden Noten besser einordnen zu können. So können Töne einfacher bestimmten Akkorden zugeordnet werden oder der Rhythmus des Musikstückes erkannt werden.

CNNs hingegen können gut räumliche Strukturen erkennen. Das hifft uns bei der Analyse von Spektrogrammen. Meist werden die verschiedenen Frequenzen der Noten, die gespielt wurden, nach der Verarbeitung des Musikstückes in Spektrogrammen wiedergegeben. Durch die Analyse von dem Spektrogramm können gewisse Frequenzmuster erkannt werden, die dann einem bestimmten Instrument zugeordnet werden können. Das ist vor allem hilfreich dabei verschiedene Stimmen der jeweiligen Instrumente voneinander zu differenzieren.

RNNs und CNNs werden auch häufig kombiniert in AMT-Modellen. Meist in folgender Reihenfolge:

1. **CNN:** Extrahiert folgende Merkmale aus dem Spektrogramm:
    - Frequenzverteilungen und spektrale Muster
    - Tonhöhenlage und damit verbundene Obertöne
    - Klangfarbe einzelner Instrumente
    - Energieverteilung, unter anderem zur Erkennung von Toneinsätzen
    - Harmonische Strukturen wie Akkordfolgen
2. **RNN:** Verarbeitet auf Basis dieser Merkmale die zeitliche Abfolge und erkennt dabei folgende Eigenschaften:
    - Reihenfolge und Übergänge musikalischer Ereignisse
    - Beginn und Ende einzelner Töne zur Bestimmung der Notendauer
    - Rhythmische Muster und zeitliche Gruppierungen
    - Musikalische Phrasen mit zusammenhängender Struktur
    - Wiederholungen, Themen oder längere Abhängigkeiten im Verlauf
3. **Output:** Gibt das transkribierte Musikstück in strukturierter Form aus:
    - Als MIDI-Datei mit exakten Noteninformationen
    - Oder als Transkription in standardisierter Notenschrift

1.4 praktische Anwendungsfelder und Vorteile von AMT

AMT kann auch bei vielen anderen Problemen helfen oder in vielen Bereichen Quality-of-Life-Changes bringen. Zum einen kann der Musikunterricht spannender und interaktiver gestaltet werden. Es gibt eine breitere Auswahl von Musikstücken, die man den Schülern anbieten kann, wodurch diese durch individuell angepasste Musikstücke mehr Spaß und Ehrgeiz beim lernen haben könnten. Zudem kann man die gespielten Musikstücke der Schüler direkt beim Spielen transkribieren und gezielt erkennen, wo der jeweilige Schüler noch Verbesserungsmöglichkeiten hat. Grundsätzlich können deutlich mehr Musikstücke transkribiert werden, wodurch sich große Archive aufbauen lassen. Ein größeres Interesse an Musik wird geweckt, da Musikstücke von beliebten Serien, Filmen oder Spielen leichter für deren Musikbegeisterte Zielgruppe zugänglich sind. Allein dadurch, dass Computerprogramme Musikstücke besser verstehen, können darauf aufbauend weitere Tools für die Musikproduktion entwickelt werden. Auch KI würde davon stark profitieren. KI-generierte Musik würde verbessert werden, da die KI selber ein besseres Verständnis der Musik entwickelt. Audiobasierte Suchmaschinen könnten gewünschte Musikstücke oder bestimmte Videos präziser finden. Musik könnte barrierefreier gestaltet werden, indem gehörlose Menschen sie lesen können und Musiker beim Spielen direktes Feedback erhalten, ob sie die Noten korrekt gespielt haben.

1.5 Motivation und Zielsetzung dieser Arbeit

(Noch nicht angefangen!) Was wird in der Arbeit behandelt, worauf liegt der Fokus (z.B. KI-Methoden für AMT), warum ist das Thema relevant (z.B. für Musiker, KI-Forschung, Musikpädagogik)?

2 Geschichte

Geschichte von ding hier musik und so

3 Fazit

Abschließende Bemerkungen, Reflexion und Ausblick.