

SIGNALBASIERTE MUSIKALISCHE ÄHNLICHKEITSSUCHE IM PRAXISEINSATZ

Matthias Eichner und Dirk Schönfuß

*mufin GmbH
dschoenfuss@mufin.com*

Kurzfassung: Musikempfehlungssysteme sind zu einem wertvollen Hilfsmittel geworden, um große Musiksammlungen zu verwalten und neue Musik zu entdecken. Die meisten Anbieter setzen dafür auf die Mitwirkung ihrer Nutzer. Im Folgenden möchten wir unsere Erfahrungen bei Entwicklung und Betrieb eines Musikempfehlungssystems vorstellen, das auf inhaltsbasierten Daten basiert, welche mit Hilfe von Methoden der Signalverarbeitung gewonnen werden. Für die Ähnlichkeitssuche setzen wir dabei sowohl signalnahe Merkmale als auch semantische musikalische Eigenschaften ein, die mit Hilfe von Verfahren des maschinenbasierten Lernens ermittelt werden. Das resultierende System kann mit weiteren Datenquellen – z.B. Expertenannotationen oder nutzergenerierte Daten – ergänzt werden, liefert jedoch auch autark bereits sehr gute Ergebnisse. Um eine solche Technologie ausgehend vom Laborstatus so fortzuentwickeln, dass sie den Anforderungen einer Web-Anwendung mit einem Katalogbestand von mehreren Millionen Titeln genügt, waren jedoch große Herausforderungen zu bewältigen. Ein weiteres Thema sind Bedienkonzepte, die das Potential der inhaltsbasierten Technologie optimal nutzen und es dem Nutzer erlauben, seine eigene Sammlung oder große Katalogbestände auf intuitive Weise zu durchstöbern.

1 Automatische Verfahren zur Musikempfehlung

Die Art und Weise wie Nutzer Musik hören hat sich durch die Verbreitung mobiler Abspielgeräte und das Internet im letzten Jahrzehnt grundlegend geändert. An die Stelle von Musikgeschäften sind Online-Portale getreten und die Musiksammlung liegt zunehmend mehr auf elektronischen Speichermedien vor als in Form von CDs im Regal. Es ist bequemer und leichter geworden, Musik eines bestimmten Interpreten zu finden und zu erwerben. Gleichzeitig ist es schwieriger, sich in dem enormen Angebot an Musik zurechtzufinden, neue Musik nach seinem Geschmack zu finden oder nur die eigene Musiksammlung zu verwalten. Untersuchungen zeigen, dass die meisten Nutzer dazu tendieren, immer dieselben Titel und Interpreten zu hören, weil Ihnen das manuelle Zusammenstellen von Abspielisten zu umständlich ist oder weil sie schlicht vergessen haben, welche Musik sie überhaupt besitzen.

Automatische Verfahren zur Musikempfehlung setzen hier an: Mit ihnen lässt sich die Ähnlichkeit zwischen zwei Songs auf der Basis ihrer Eigenschaften berechnen und so für jeden Titel eine Rangliste der ähnlichsten Songs in der Datenbasis erstellen. Als Informationsquelle für solche Systeme kommen prinzipiell redaktionell gepflegte Attribute, von Nutzern generierte Daten sowie inhaltsbasierte, aus dem Audiosignal ermittelte Merkmale in Frage.

Die redaktionelle Pflege von musikalischen Eigenschaften, die für eine Ähnlichkeitssuche ausreichend sind, ist aus Zeit- und Kostengründen für große Kataloge oder Independent-Musik nicht durchführbar – für privat produzierte Musik können außerdem gar keine Empfehlungen abgegeben werden.

Bei Online-Anbietern wie Amazon hat sich die Empfehlung nach dem Prinzip „Wer dies gekauft hat, hat auch jenes gekauft“ durchgesetzt. Dieser auf Verkaufsstatistiken beruhende

Ansatz hat die nachteilige Eigenschaft, dass populäre Alben noch populärer werden und wenig bekannte Interpreten nie empfohlen werden. Das Dilemma wird besonders an den folgenden Zahlen deutlich: 1% der verkauften Musiktitel generieren 80% des Umsatzes [1][2]. Die Verkaufszahlen folgen dem Potenzgesetz: Wenige Alben werden überdurchschnittlich häufig verkauft, wogegen die Mehrheit kaum oder gar nicht nachgefragt wird.

Eine weitere Möglichkeit, auf Nutzerdaten zurückzugreifen, setzen Online-Plattformen wie lastfm.com ein. Hier vergeben Nutzer den Musiktiteln Schlüsselwörter (Tags), welche die Songs nach ihrer Meinung gut beschreiben. Je mehr Nutzer dies tun, desto zuverlässiger sind die Informationen und umso besser funktioniert die Suche nach Titeln mit ähnlichen Eigenschaften. Werden die Tags sehr sorgfältig vergeben, dann können beispielsweise auch subjektiv empfundene Eigenschaften, Informationen aus dem sozialen Kontext, der Epoche oder regionale Informationen in die Suche mit einfließen, die sich rein aus dem Audiosignal nicht ermitteln ließen. Allerdings leidet auch dieser Ansatz darunter, dass für Titel von populären Künstlern deutlich mehr Tags vergeben werden als für unbekannte Künstler.

Signalbasierte Verfahren sind davon unbeeinflusst: Mit ihnen werden Eigenschaften wie z. B. Perkussivität, Tempo und Instrumentierung direkt aus dem Musiksignal extrahiert. Die Empfehlung ähnlicher Titel ist daher unbeeinflusst von Dingen wie Verkaufsrang, Popularität oder Fan-Gemeinde in einer Online-Community – außerdem kann die Technologie auf jedes Musikstück angewendet werden und funktioniert auch rein lokal ohne die Notwendigkeit, Daten von einem zentralen Server abzurufen. Die Verarbeitung eines Titels erfolgt in der Regel in Sekundenbruchteilen und die Ergebnisse lassen sich kompakt in Datenbanken speichern.

Mufin bietet eine inhaltsbasierte Empfehlungstechnologie an, die die genannten Vorteile aufweist, jedoch mit nutzerbasierten Datenquellen kombinierbar ist. Eine solche Kombination verschiedener Wissensquellen verspricht die besten Ergebnisse und die höchste Akzeptanz durch den Nutzer.

2 Herausforderungen bei der signalbasierten Musikempfehlung

Um ein signalbasiertes Musikempfehlungssystem praxistauglich zu machen, das den Performance-Anforderungen eines Echtzeit-Websystems entspricht und mit einer Datenbankgröße von mehreren Millionen Musikstücken arbeitet, mussten wir uns verschiedenen Herausforderungen stellen. Zunächst einmal stellt eine solche Klassifikationsaufgabe enorme Ansprüche bezüglich Speichergröße und Rechenleistung, da die dort verarbeiteten, aus Audio-Lowlevel-Feature erstellten Feature-Vektoren eine deutlich höhere Komplexität besitzen als die vergleichsweise einfachen Datentypen, die in einem System verwendet werden, das mit Nutzer- oder Expertendaten arbeitet.

Das System besteht aus drei wesentlichen Bestandteilen: In Schritt (1) wird aus dem Audiosignal ein Fingerprint erstellt, in Schritt (2) wird aus dem Fingerprint ein kompaktes Modell des Songs gebildet und in Schritt (3) erfolgt eine Berechnung der Ähnlichkeit auf der Basis dieser Song-Modelle ([3], Abbildung 1).

Zunächst wurde untersucht, welche Low-Level-Features geeignet sind, um als Basis für die Analyse musikalischer Eigenschaften zu dienen. Bei der Erstellung der Song-Modelle musste ein Kompromiss zwischen geringem Speicherbedarf und einer möglichst hohen Aussagekraft sowie einer geeigneten Granularität der Eigenschaften gefunden werden. Einerseits möchte man so viele Features wie möglich berücksichtigen, um ein optimales Resultat zu erhalten, andererseits muss man die Größe der Song-Modelle so weit wie möglich reduzieren, um die Berechnung der Empfehlungen innerhalb einer realistischen Zeitspanne zu ermöglichen.

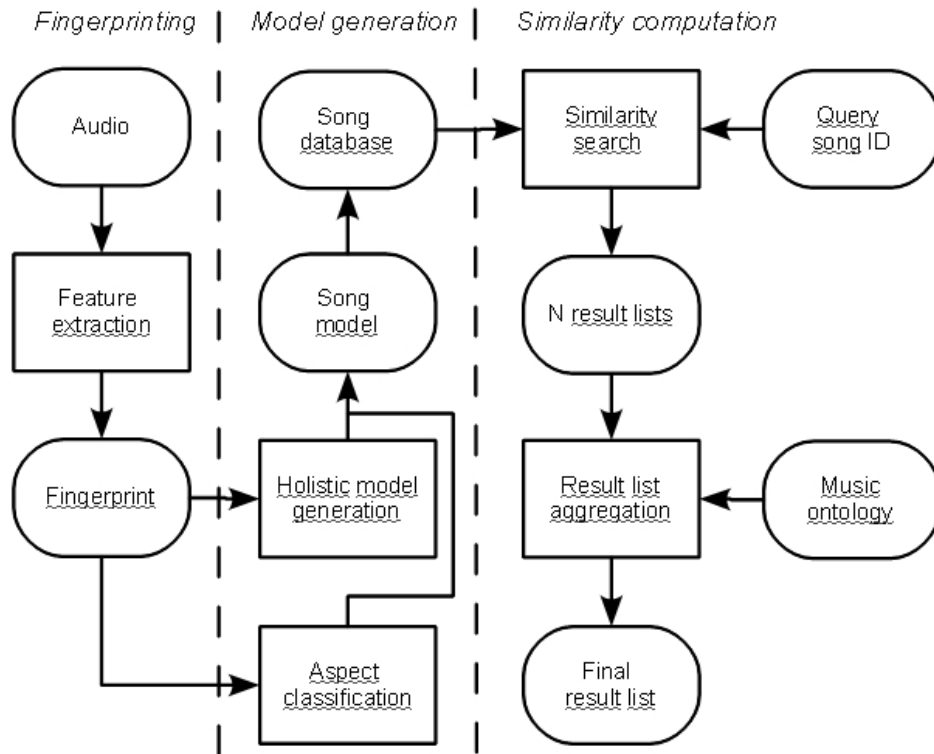


Abbildung 1: Blockschaltbild des Musikempfehlungssystems

iTunes, der weltgrößte Musikshop, hat im Januar 2009 bekannt gegeben, dass nun mehr über 10 Millionen Musikstücke zum Kauf verfügbar sind. Als wir unser Forschungsprojekt im Jahr 2006 begannen, betrug die Größe eines Fingerprints im Speicher ca. 15 KB – daher war es selbst mit 32 GB Hauptspeicher nur möglich, etwa 2 Millionen Songs in die Datenbank aufzunehmen. Zu damaligen Zeit dauerte ein einziger Anfrage nach ähnlichen Stücken auf einer IntelTMXeonTM CPU ca. 211 Sekunden, d.h. es hätte ca. 13,4 Jahre gedauert, eine komplette Berechnung des Kataloges durchzuführen.

Zunächst einmal optimierten wir bei nahezu gleich bleibender Qualität die Größe der Song-Modelle mit Hilfe eines genetischen Algorithmus'. Dadurch ist es uns heute möglich, selbst Modelle von größten Musikkatalogen mit weniger als 16 GB Hauptspeicher komplett zu laden.

Eine weitere Herausforderung stellte die Komplexität des Suchalgorithmus' dar. Unser Ziel war es, den Suchraum für jede Anfrage durch Vorhersage der Teile des Katalogs einzuschränken, in dem sich die ähnlichen Titel mit hoher Wahrscheinlichkeit befinden werden. Es gelang uns, durch die Analyse der Korrelation der Ergebnisse mit den von uns verwendeten signalnahen Features und musikalischen Eigenschaften mit den ermittelten Ähnlichkeiten, geeignete Merkmale für einen solchen Mechanismus zu finden [4]. Eine solche Optimierung stellt wiederum einen Kompromiss zwischen Ergebnisqualität und Geschwindigkeitsgewinn dar.

Mit Hilfe der durchgeführten Optimierungen ist es uns gelungen, in einem im Januar 2008 durchgeführten Test eine Performance zu erreichen, die die komplette Berechnung von Ähnlichkeitsinformationen für einen Musikkatalog mit ca. 5 Millionen Stücken mit der uns zur Verfügung stehenden Hardware in weniger als einer Woche erlaubt.

3 Neue Möglichkeiten zur Navigation in Musiksammlungen

3.1 Klangbasierte Sortierung

Jede datenbankbasierte Player-Software bietet Funktionen an, seine Sammlung nach Meta-Informationen zu sortieren. Cher und Coldplay haben allerdings zwar den gleichen Anfangsbuchstaben, aber sonst nicht viel gemeinsam. Die Sortierung nach Künstler- oder Trackname zeigt nicht immer unbedingt Titel, die auch musikalisch zueinander passen. In der kostenlosen Software „mufin Player“ kann man die Musiksammlung nach dem Klang der Tracks sortieren und behält so immer die Tracks im Blick, die musikalisch zueinander passen.

3.2 Semantische Suche

In großen Datenbanken wird es immer wichtiger, dass Werkzeuge zur Verfügung stehen, die es dem Nutzer erlauben, gezielt das zu finden, was er wirklich sucht. Durch die Analyse der Musikdateien werden Eigenschaften ermittelt, die es einem Suchsystem ermöglichen, dem Nutzer weitere Suchkriterien, Filter bzw. Möglichkeiten zur Verfeinerung oder Disambiguierung seiner Suche anzubieten. Außerdem wird es möglich, semantische Verknüpfungen zwischen einzelnen Items in einer Datenbank herzustellen – angefangen von Song-zu-Song-Beziehungen bis hin zu Informationen über die Nutzerpräferenzen und Ähnlichkeiten zwischen Interpreten.

3.3 mufin Vision

Die Verwaltung von Musiksammlungen mittels herkömmlicher, textbasierter Informationen wie Künstler oder Musiktitel erlauben es nicht, sich einen schnellen Überblick über eine große Musiksammlung zu verschaffen. Titel, die die gleiche Stimmung haben, können in einer solchen Liste an völlig entgegengesetzten Stellen auftauchen. Die grafische 3D-Darstellung „mufin vision“ wurde mit dem Ziel entwickelt, eine Lösung zur Musikverwaltung und -navigation für heutige digitale Musiksammlungen zu bieten, die selbst Bibliotheken mit tausenden von Musiktiteln einfach zugänglich macht.

Mufin vision stellt jeden Musiktitel in der Musiksammlung in einem Koordinatensystem dar. X,Y und Z-Achse können mit unterschiedlichen Kriterien belegt werden, so dass man die Musiksammlung in verschiedenen Dimensionen erforschen kann. Die Kriterien umfassen zum einen klangbasierte Merkmale wie Tempo, Stimmung, Musikstil, Instrumentierung oder Klangdichte, zum anderen Meta-Informationen wie Veröffentlichungsdatum oder Titellänge. Das ermöglicht einen schnellen und einfachen Zugriff auf ähnliche Musik, da Musikstücke, die dieselben Merkmale teilen, visuell nah beieinander erscheinen. Beziehungen zwischen Titeln werden so direkt sichtbar – es wird deutlich einfacher, Musik zu finden und wiederzuentdecken, die ähnliche Merkmale oder einen ähnlichen Stil hat. Selbst auf kleinen Bildschirmen wie bei tragbaren Geräten oder Handys lässt sich mufin vision einsetzen.

Literatur

- [1] Nielsen Soundscan: State of the industrie 2007.
URL: <http://www.narm.com/2008Conv/StateoftheIndustry.pdf> (22.07.2009)
- [2] Celma, O.: Music Recommendation and Discovery in the Long Tail. PhD-Thesis, Universität Pompeu Fabra, Spain, 2008.
- [3] Bastuck, C.: An extensible and multiperspective approach for music similarity. In 3rd Music Information Retrieval Evaluation eXchange (MIREX 2007), Vienna, Austria, Sept 2007.
- [4] Hübler, S.: Suchraumoptimierung zur Identifizierung ähnlicher Musikstücke. Diplomarbeit, TU-Dresden, Institut für Akustik und Sprachkommunikation, 2008.