

AUCOMA - Adaptive Nutzerzentrierte Organisation von Musikarchiven

Sebastian Stober
Andreas Nürnberger

*Institut für Technische und Betriebliche Informationssysteme, DKE Arbeitsgruppe
Otto-von-Guericke-Universität Magdeburg*

Zusammenfassung

Automatische Strukturierung kann den Zugriff auf Musikarchive, speziell die Exploration und Organisation, wesentlich erleichtern. Noch hilfreicher wäre eine Darstellung, die sich an die Art und Weise, wie der Nutzer Musiksammlungen strukturiert, anpasst und somit für ihn intuitiv nachvollziehbar ist. Wir stellen hier ein prototypisches System vor, welches ein personalisiertes Ähnlichkeitsmaß anhand der Nutzerinteraktion mit einer Musiksammlung lernt. Zunächst wird dazu eine wachsende Selbstorganisierende Karte (SOM) trainiert, die nutzerunabhängig ähnliche Musikstücke gruppiert. Der Nutzer kann anschließend die Position von Musikstücken in der Karte durch einfache Drag&Drop-Aktionen verändern. Jede Bewegung verursacht eine automatische Anpassung des der Karte zugrunde liegenden Ähnlichkeitsmaßes, wodurch auch andere Stücke ihre Position ändern können.

Motivation

Aspekten der Personalisierung wurde bisher in der Forschung im Bereich des Music Information Retrieval und der digitalen Musikbibliotheken wenig Beachtung geschenkt. Oftmals wird angenommen, dass alle Nutzer eines Systems Musikstücke auf die gleiche (objektive) Art und Weise vergleichen, was jedoch im Allgemeinen nicht der Fall ist: So mag ein Musiker dazu neigen, stärker auf Strukturen, Instrumentierung oder Harmonien zu achten und dabei vielleicht seinem Instrument eine besondere Gewichtung geben. Nicht-Musiker werden sich beim Vergleichen möglicherweise eher auf die Klangfarbe oder allgemeine Stimmung eines Stückes stützen. Andere wiederum achten, sofern sie der jeweiligen Sprache mächtig sind, auf den Liedtext. Ein System, welches diese Subjektivität berücksichtigt, wird sich besser auf die individuellen Bedürfnisse seiner Nutzer einstellen können. Insbesondere die Exploration und Organisation von Musiksammlungen sind Szenarien, für die dies besonders hilfreich wäre. Hier könnte die Sammlung individuell auf eine Art und Weise strukturiert und dargestellt werden, die für den jeweiligen Nutzer intuitiv verständlich ist.

Derzeitige Systeme bieten jedoch oft bestenfalls die Möglichkeit, vom Nutzer angepasst zu werden. Ihnen fehlt jedoch die Fähigkeit, sich selbständig an den Nutzer anzupassen. Wir stellen im Folgenden einen Ansatz vor, welcher auf der grundlegenden Idee aufbaut, ein Ähnlichkeitsmaß bezüglich (impliziter) Nutzerpräferenzen anzupassen. Dieses personalisierte Ähnlichkeitsmaß kann dann dazu verwendet werden, Musikstücke einer Sammlung zu gruppieren.

Die Idee eines adaptiven Ähnlichkeitsmaßes wurde bereits bei MPeer [2] verwendet. Hier können Nutzer mit Hilfe eines intuitiven Joystick-Interfaces die Gewichtung dreier Facetten (Audioinhalt, Text und "Cultural Metadata") beeinflussen. Auf ähnliche Weise erlaubt die E-Mu Jukebox [13] eine Änderung der Ähnlichkeit. Hier werden 5 Facetten (Klang, Tempo, Stimmung, Genre, Jahr) visuell durch "Adapter" repräsentiert, die auf eine Zielscheibe gezogen werden können, wobei der Nähe zum Zentrum dem jeweiligen Gewicht entspricht. PATS (Personalized Automatic Track Selection) [9], ein adaptives System zur Playlist-Generierung für individuelle Kontexte, kommt ohne explizite Anpassung aus sondern lernt aus dem Nutzerfeedback. Die Idee ähnelt der hier vorgestellten, jedoch unterscheiden sich Nutzungsszenario und Adaptionalgorithmus grundlegend. Zur Ähnlichkeitsbasierten Gruppierung und Visualisierung verwenden wir eine Selbstorganisierende Karte (SOM), ein Ansatz, der bereits mehrfach erfolgreich im Bereich des Music Information Retrieval Anwendung fand, beispielsweise in [8, 5, 6]. Jedoch wird bei keinem dieser Ansätze eine individuelle Anpassung an den Nutzer vorgenommen.

Um unser Adaptionverfahren zu demonstrieren, haben wir einen ersten Prototyp entwickelt. Im folgenden stellen wir einzelne Aspekte dieses Systems vor.

Datengrundlage

Für eine erste Analyse verwenden wir den Beatles-Corpus. Dieser umfasst 282 Stücke der Beatles. Für 200 Stücke liegen Audiodaten vor, davon wurden 180 (die Stücke der 12 offiziellen Beatles-Alben) an der Queen Mary University, London von Hand annotiert, so dass hier Informationen über Akkorde und Tonarten vorliegen.¹

Für jedes Musikstück der Sammlung wird eine Reihe von Merkmalen extrahiert, die verschiedene Facetten beschreiben. Zur Merkmalsextraktion aus den Audiodaten verwenden wir das CoMIRVA Framework [10] und JAudio [4]. CoMIRVA unterstützt derzeit 2 komplexe Merkmale, die den allgemeinen Klang eines Musikstückes als Mischung von Gaußverteilungen der Mel Frequency Cepstral Coefficients (MFCCs) nach [1] und [3] beschreiben, sowie ein Merkmal, das aussagt, wie stark und schnell Beats in spezifischen Frequenzbändern gespielt werden [8]. JAudio wird genutzt, um eine globale Klangbeschreibung wie in [12] vorgeschlagen zu extrahieren.

¹Diese manuelle Annotation ist sehr zeitaufwändig und für größere Datensammlungen nicht praktikabel. Daher wird stattdessen eine automatische Akkord- und Tonartklassifikation durchgeführt, wenn keine manuellen Annotationen vorliegen.



Abbildung 1: Das Nutzerinterface des Prototypen. Links die Karte, rechts 2 Zell-Detailfenster.

ren. Aus den Akkordannotationen wird ein Histogramm berechnet und normiert. Weiterhin werden Informationen über den Produktionsprozess (Produzent, Toningenieur, Studio, Aufnahmezeitraum, Musiker und Instrumente) automatisch aus Wikipedia-Artikeln extrahiert.² Die Texte der Musikstücke werden automatisch über den Webservice von LyricWiki³ bezogen und mit Standardverfahren des Text Information Retrieval (Stemming, Stopwortfilterung, TFIDF Termwichtung) verarbeitet und repräsentiert. Daneben werden herkömmliche ID3-Metainformation (Titel, Album, Jahr) als Merkmale verwendet. Insgesamt werden derzeit 21 Merkmale zur Ähnlichkeitsberechnung herangezogen. Dazu wird die gewichtete Summe der elementaren Ähnlichkeiten auf den einzelnen Merkmalen berechnet.

Adaption

Basierend auf einer Gleichgewichtung aller Merkmale wird zunächst eine (wachsende) Selbstorganisierende Karte bestehend aus hexagonalen Zellen gelernt. Der dazu verwendete Lernalgorithmus wird in [7] detailliert beschrieben. Eine gelernte Karte ist in Abb. 1 dargestellt. Jede Zelle entspricht dabei einer Gruppe ähnlicher Stücke und wird visuell durch das Albumcover repräsentiert, welches am häufigsten mit den enthaltenen Stücken verknüpft ist, und bezüglich der Gruppengröße eingefärbt. Mit einem Klick auf eine Zelle wird ihr Inhalt in einem Detailfenster angezeigt. Der Nutzer kann nun beliebige Stücke aus einem solchen Fenster per Drag&Drop in andere Zellen oder entsprechende Detailfenster verschieben. Eine solche Aktion führt zur Aktualisierung der Merkmalsgewichtung und somit zu einer Neuordnung aller Stücke zu den Zellen der Karte. Das manuelle Verschieben eines einzelnen Stückes kann so zu einer automatischen Verschiebung weiterer Stücke führen. In einem iterativen Prozess aus Nutzerauktion und resultierende Adaption wird das Ähnlichkeitsmaß schrittweise an die (möglicherweise unterbewußten) Vorstellungen des Nutzers angepasst.

Zur Berechnung der neuen Gewichtung wird ein quadratisches Optimierungsverfahren verwendet, welches in [11] ausführlich vorgestellt und für Textdaten evaluiert wurde. Die Gewichte werden dabei minimal verändert – gera-

²<http://de.wikipedia.org> – Zu den Stücken der Beatles lagen hier umfangreiche Informationen vor. Das Verfahren muss für weitere Künstler jedoch gegebenenfalls noch erweitert werden.

³<http://lyricwiki.org>

de so stark, dass alle (auch in der Vergangenheit) verschobenen Stücke ihren Zielzellen zugeordnet werden. Dies verhindert eine allzu abrupte Veränderung des Ähnlichkeitsmaßes. Weiterhin wird sichergestellt, dass Gewichte nicht negativ und auch nicht beliebig groß werden.

Erste subjektive Evaluierungen lassen zumindest auf eine Akzeptanz beim Nutzer deuten. Es kann jedoch vorkommen, dass der Nutzer ein Merkmal sehr stark gewichtet, das nicht extrahiert wurde, und somit nicht adaptiert werden kann. Da dieses Problem auch bei einer großen Anzahl von unterstützten Merkmalen nicht auszuschließen ist, muss ein Weg gefunden werden, damit umzugehen. Des Weiteren soll untersucht werden, inwieweit der vorgestellte Ansatz auf andere ähnlichkeitsbasierte Darstellungsmethoden übertragen werden kann.

Literatur

- [1] J.-J. Aucouturier and F. Pachet. Improving timbre similarity: How high is the sky? *Journal of Negative Results in Speech and Audio Sciences*, 1(1), 2004.
- [2] S. Baumann and J. Halloran. An ecological approach to multimodal subjective music similarity perception. In *Proc. of CIM'04*, 2004.
- [3] M. Mandel and D. Ellis. Song-level features and support vector machines for music classification. In *Proc. of ISMIR'05*, 2005.
- [4] D. McEnnis, C. McKay, I. Fujinaga, and P. Depalle. jAudio: An feature extraction library. In *Proc. of ISMIR'05*, 2005.
- [5] F. Mörchen, A. Ultsch, M. Nöcker, and C. Stamm. Databionic visualization of music collections according to perceptual distance. In *Proc. of ISMIR'05*, 2005.
- [6] R. Neumayer, M. Dittenbach, and A. Rauber. PlaySOM and PocketSOMPlayer, alternative interfaces to large music collections. In *Proc. of ISMIR'05*, 2005.
- [7] A. Nürnberger and M. Detyniecki. Externally growing self-organizing maps and its application to e-mail database visualization and exploration. *Applied Soft Computing*, 6(4), 2006.
- [8] E. Pampalk, A. Rauber, and D. Merkl. Content-based organization and visualization of music archives. In *Proc. of ACM MULTIMEDIA'02*, 2002.
- [9] S. Pauws and B. Eggen. PATS: Realization and user evaluation of an automatic playlist generator. In *Proc. of ISMIR'02*, 2002.
- [10] M. Schedl. The CoMIRVA Toolkit for Visualizing Music-Related Data. Technical report, Johannes Kepler University Linz, 2006.
- [11] S. Stober and A. Nürnberger. User modelling for interactive user-adaptive collection structuring. In *Proc. of AMR'07*, 2008.
- [12] G. Tzanetakis. Marsyas submission to MIREX 2007. In *Proc. of ISMIR'07*, 2007.
- [13] F. Vignoli and S. Pauws. A music retrieval system based on user driven similarity and its evaluation. In *Proc. of ISMIR'05*, 2005.