Technische Universität Berlin
Fakultät I - Geisteswissenschaften
Fachgebiet Audiokommunikation
Betreuer: Dr. Steffen Lepa, Prof. Dr. Stefan Weinzierl

Audio-Feature-Analyse als Ergänzung kollaborativen Filterns im Kontext von Musikempfehlungsdiensten

Exposé zur Masterarbeit

Martin Woit
Friedenstraße 92c
10249 Berlin
martin.woit@gmail.com
Matrikelnr.: 32 87 67
18. Oktober 2013

Zusammenfassung

Die Wichtigkeit der Musikempfehlungsdienste steht in der heutigen Zeit außer Frage. Daher soll in der Masterarbeit untersucht werden, inwieweit sich die Empfehlungsgenauigkeit verbessert, wenn man kollaborative Filter mit inhaltsbasierenden Filtern erweitert. Dies soll mit Hilfe eines naiven Bayes-Klassifikators bewerkstelligt werden. Ziel der Arbeit ist es, ein Benchmark zu erarbeiten, welches vorhandene Feature-Kombinationen und Algorithmen in ihrer Eignung zur Verbesserung der Empfehlungsgenauigkeit vergleicht.

1 Einleitung und Fragestellung

Personalisierte Musikempfehlungsdienste spielen in Zeiten von Audio-Streaming-Plattformen wie Spotify[1], Deezer[2] oder Rdio[3] eine immer wichtiger werdende Rolle für den täglichen Musikkonsum. Basierend auf Interessen und zurückliegendem Verhalten der Anwender, nutzen diese Dienste häufig sogenannte kollaborative Filtertechniken, um weitere Musiktitel vorzuschlagen. Eine andere Möglichkeit der Empfehlungsgenerierung stellen inhaltsbasierte Filter da. Hierzu kann eine Audio-Feature-Analyse der dem Dienst zugrundeliegenden Musiken hinzugeordnet werden. Da die Klassifikationen der Feature-Analyse der auf Nutzerverhalten basierenden Methode häufig in Genauigkeit unterlegen ist, soll in der Masterarbeit untersucht werden, ob die Kombination beider die Fehlerrate des kollaborativen Filterns verbessern kann. Als Grundlage soll hierzu ein vorhandener Datensatz mit Nutzer-Ratings eines B2B-Empfehlungsdienstes für Produktionsmusik dienen. Darin enthalten sind 3800 Nutzerbeurteilungen von 40 Produktionsmusiken auf 27 Ausdrucksdimensionen, welche während eines Hörversuchs zu einem Webradio evaluiert worden sind.

Bei der Bearbeitung des Themengebiets ist es interessant herauszufinden, ob sich durch die Extraktion der Features mittels Musikinformationsgewinnung (Music Information Retrieval - MIR) eine oder mehrere Ausdrucksdimensionen der Nutzer-Ratings entbehren ließen.

2 Stand der Forschung

Zu den Themengebieten der hybriden Musikempfehlungsgenerierung existieren zahlreiche Forschungsberichte. Wichtig dabei ist oft die Verknüpfung der kollaborativen und inhaltsbasierten Filtertechniken zur Verbesserung der Empfehlungsgenauigkeit. Yoshii et al. (2006) entwickelten hierzu eine Musikempfehlungsmethode, die Ratings von Nutzern und Audioinformationen verbinden konnte. Als sogenanntes 'aspect model' wurde ein Bayes'sches Netz gewählt, welches nicht erkennbare Nutzerpräferenzen durch verborgene, statistisch ermittelte Variablen ergänzte. Sie konnten zeigen, dass die eingesetzte Methode die beiden

[1] `https://www.spotify.com/`
[2] `http://www.deezer.com/`
[3] `http://www.rdio.com/`

konventionellen Methoden, im Zusammenhang zur Empfehlungsgenauigkeit, übertrafen. Auch wurde der 'kalte Start' überwunden, denn Musikstücke konnten empfohlen werden, auch wenn sie keine Ratings durch User erhalten hatten.

Li et al. (2007) stellten ein kollaboratives Musikempfehlungssystem für Online-Telefonklingeltöne vor. Es basiert auf einem Wahrscheinlichkeitssystem, welches Elemente in Gruppen oder Gemeinschaften klassifiziert. Dabei wurde die Gaußverteilung genutzt, um Vorhersagen für die Nutzer anhand der vorangegangenen Ratings zu erstellen. Ziel war es, Audio-Eigenschaft und kollaborativen Filter mit Hilfe der K-Medoids-Clustermethode zu verknüpfen. Es zeigte sich, dass die vorgestellte Methode die beiden Standardmethoden, kollaboratives und inhaltliches Filtern im Einzelnen, übertraf.

Magno und Sable (2008) führten eine Feature-Analyse durch, indem sie die Mel-Frequenz-Cepstrum-Koeffizienten (Mel-Frequency Ceptral Coefficients - MFCCs) der zu untersuchenden Songs bestimmten. Die Koeffizienten wurden in drei verschiedenen Algorithmen verwendet, um daraus Empfehlungen für Nutzer zu generieren. Es konnte gezeigt werden, dass ein signalbasierter Empfehlungsalgorithmus, hier im speziellen ein Expectation-Maximization-Algorithmus verfeinert mit einem Monte-Carlo-Sampling, mit den kommerziellen Algorithmen bekannter Firmen mithalten konnte. Auch Yoshii et al. (2008) machten sich die MFCCs zunutze und stellten einen hybriden Empfehlungsalgorithmus vor, bei welchem ein generatives Wahrscheinlichkeitsmodell herangezogen wurde, das die kollaborativen und inhaltsbasierten Daten verbinden konnte. Mit diesem sogenannten "three-way aspect model" war es möglich, steigende Userzahlen und Ratings (Ratingmodell von Amazon) einfach in das System zu integrieren. Im Vergleich zu den Einzelfiltermethoden in vier verschiedenen Versionen (speicher- oder modellbasiert verknüpft mit kollaborativ oder inhaltsbasiert), war die vorgeschlagene Methode hinsichtlich der Empfehlungs*ge*nauigkeit ebenbürtig bzw. überlegen. Auch bei der Empfehlungs*un*genauigkeit lag das "three-way aspect model" vorn, da dieses im Mittel am niedrigsten war.

Yoshii und Goto (2009) präsentierten ein kontinuierliches pLSI-basiertes (Probabilistic Latent Semantic Indexing) Modell für hybride Musikempfehlung. Da bei dieser Methode die Anzahl der Parameter sehr hoch und dies ein Indikator für eine unangemessene Empfehlung von bestimmten Items an alle User ist ("hubs"), untersuchten Yoshii und Goto drei Glättungstechniken zur Behebung des Problems. Gaußsche Parameterverknüpfung und künstlerbasierte Item-Gruppierung konnten die Modellkomplexität reduzieren und somit die Genauigkeit des hybriden Musikempfehlungsdienstes verbessern. Die Parameterverknüpfung reduzierte die Anzahl der sogenannten "hubs" zusätzlich. Die dritte Glättungstechnik (multinomiale Glättung) konnte nicht überzeugen, da sie die Empfehlungsgenauigkeit sogar noch verschlechterte.

Lu und Tseng (2009) zeigten mit ihrem personalisierten hybriden Musikempfehlungssystem, dass die Empfehlungsgenauigkeit bei 90% liegen kann, wenn neben kollaborativen und inhaltsbasierten, auch emotionenbasierte Filtertechniken herangezogen werden.

Bu et al. (2010) stellten eine Methode vor, die Social-Media-Informationen (z.B. Freundschaftsbeziehungen, Mitgliedschaftsverhältnisse) mit akustikbasierten Musikinformationen verband. Da Social-Media-Informationen viele verschiedene Objekte und komplexe Beziehungen beinhalten können, wurde ein Hypergraph als Modell eingesetzt. Dadurch war

es möglich, die vielen einzelnen Gebilde in geeignete Objektmengen zu verknüpfen. Im Vergleich übertraf das postulierte Schema fünf weitere Einzelempfehlungsalgorithmen signifikant (u.a. kollaborativ, akustikbasiert oder auch inhaltsbasiert, Einzelmethoden des Modells). In einer weiteren Methode kombinierten auch Su et al. (2010) musikalischen Inhalt und Kontextinformationen. In Verbindung eines Clusterverfahrens und der Hinzunahme einer heuristischen Vorgehensweise zur Nutzerpräferenzzusammenstellung war das vorgestellte Modell im Vergleich zu anderen Methoden überlegen (u.a. nutzerbasiert, gegenstandsbasiert).

Das System von Liu et al. (2010) konnte Musik-Wiedergabelisten erstellen, indem es die Zeit analysierte, die ein Nutzer damit verbrachte bestimmte Musiktitel zu hören. Zur individuellen Empfehlung wurden auch Features aus den Wellendaten und Noten der Musikstücke extrahiert. Als Systemkern kam ein modifiziertes neuronales Netzwerk zum Tragen. Um einen möglichen "kalten Start" zu umgehen, zogen Liu et al. zusätzlich noch eine kollaborative Filtertechnik hinzu. Dieses Problem konnte auftreten, wenn ein neuer Nutzer noch keine Musik mit dem vorgeschlagenen System gehört hatte.

Social Tags (Genre, Stil, Stimmung, Meinung des Nutzers, genutzte Instrumente) wurden in der Methode von Nanopoulos et al. (2010) verwendet. Da diese Daten häufig sehr spärlich vom Nutzer angegeben werden, wurden zusätzlich Audio-Features zur Ähnlichkeitsbestimmung herangezogen. Dadurch konnten fehlende Social Tags durch vorhandene, ähnlicher Musiktitel ersetzt werden. Die verwendeten Methoden waren ein "Gaussian Mixture Model" basierend auf MFCCs, angewendet auf eine Kullbach-Leibler-Divergenz, welche die Ähnlichkeit zwischen zwei Wahrscheinlichkeitsverteilungsfunktionen berechnet. Im Vergleich zu realen Daten von Last.fm[4] war das verwendete Model bezüglich der Empfehlungsqualität überlegen.

Wang et al. (2012) präsentierten ein adaptives, kontextbezogenes und inhaltliches Filtermodell (Adaptive Context-Aware Content Filtering Model - ACACF), welches mit Hilfe eines Bayes'schen Systems auf dem Mobiltelefon kontextbezogene Aktivitätsklassifikationen (u.a. lernen, laufen, einkaufen) und Musikinhaltsanalyse verbinden konnte. Die Evaluation des Systems ergab eine hohe Zufriedenheit der Teilnehmer bezüglich der Empfehlungsgenauigkeit geeigneter Musikstücke.

3 Methode und Quellen

3.1 Algorithmen

In der Masterarbeit soll ein naiver Bayes-Klassifikator erarbeitet werden, der Vektoren signalbasierter Audio-Features und vorhandene Nutzerratings innerhalb eines Algorithmus verbindet. Dieses hybride Musikempfehlungssystem soll zeigen, dass die Empfehlungsgenauigkeit eines kollaborativen Filtersystems durch die Hinzunahme einer inhaltsbasierten Methode gesteigert werden kann. Die statistische Unabhängigkeit bzw. funktionale Abhängigkeit (Rish, 2001) der vorhandenen Merkmale wird bei einer Klassifikation mittels

[4] http://www.lastfm.com

Bayes vorausgesetzt. Da nicht davon ausgegangen werden kann, dass die in der Arbeit verwendeten Werte statistisch unabhängig sind, wird nach Decker (2005) zunächst eine Hauptkomponentenanalyse (PCA - Principal Component Analysis) durchgeführt werden, welche eine Dekorrelation und eventuelle Dimensionsreduzierung der Daten zur Folge hat. Die gewonnenen Werte werden im Anschluss durch eine unabhängige Komponentenanalyse (ICA - Independent Component Analysis) entmischt, und es kann dann davon ausgegangen werden, dass die statistische Abhängigkeit der Komponenten minimal ist (S. 30 ff).

Eine Hauptherausforderung wird dabei die Dimensionierung eines gemeinsamen Merkmalsraums per PCA und ICA sein, da die Audiofeatures im Vergleich zu den Ratings (n pro Musiktitel) eine Dimension Variabilität weniger besitzen.

Gefolgt wird die vorangegangene Berechnung von einer Klassifikation der kollaborativen und inhaltsbasierten Daten anhand des zu entwickelnden naiven Bayes-Algorithmus und die Untersuchung, inwieweit eine Verbesserung der Empfehlungsgenauigkeit entstanden ist. Hierzu bietet sich als Benchmark die Methode des Mean-Absolute-Error (MAE) an (Breese et al., 1998; Li et al., 2007, S. 481). Je kleiner dieser Wert ausfällt, desto besser ist die Empfehlungsgenauigkeit. Auch wird eine Orientierung an der Masterarbeit von Böhringer (2013) angestrebt, da die dort verwendeten Daten als Trainungsstichprobengröße ohne Audiofeatures betrachtet werden können. Das zugehörige Empfehlungsgenauigkeitsmaß beinhaltet den Prozentsatz korrekt klassifizierter Titel.

Zu guter Letzt soll die Eignung der Feature-Kombinationen und verwendeten Algorithmen aufgezeigt werden. Interessant dabei wäre herauszufinden, ob einige der Studie zugrundeliegenden Nutzerratings durch Audio-Features ersetzt werden können.

3.2 Musik-Informations-Gewinnung

Zur signalbasierten Audio-Feature-Analyse soll die Open-Source-Matlab-Bibliothek „MIR-toolbox" von Lartillot und Toiviainen (2007) genutzt werden. Sie verfügt über sogenannte „Low-Level"- und „High-Level"-Features. Durch das modulare Framework ist es möglich, blockweise Analysen durchzuführen, welche parametrisiert, wiederverwendet und umsortiert werden können (Lartillot, 2013, S. 8). In der Literatur (2) werden u.a. „MFCCs", „Spectral Centroid" oder „Spectral Flatness" zur Feature-Bestimmung genutzt, aber auch tonale („Pitch Chroma") und temporale Analysen („Onset Detection", „Beat Histogram", „Tempo Detection") kommen zum Einsatz. Die Auswertung geeigneter Features wird dementsprechend im oben genannten vergleichenden Benchmark wiederzufinden sein.

4 Vorarbeiten

Als Vorarbeit wurde die Evaluation einer geeigneten Open-Source-Software zur Musikinformationsgewinnung durchgeführt. Die Entscheidung, eine existierende Software zu nutzen, entstand durch die Überlegung, dass die eigenständige Programmierung vorhandener Audio-Feature-Algorithmen sehr zeitintensiv ist und nicht dem Erkenntnisgewinn dient. Somit wird es möglich sein, das Hauptaugenmerk auf die Entwicklung eines Empfehlungs-

algorithmus auf Basis eines naiven Bayes-Klassifikators und die Auswertung der daraus resultierenden Empfehlungsgenauigkeit zu legen. Mögliche Kandidaten für die Feature-Analyse waren u.a. *jAudio* (Mcennis et al., 2005), *Yaafe* (Mathieu et al., 2010), *MIRtoolbox* (Lartillot und Toiviainen, 2007) und die softwarebasierte Methode von Lerch (2008), die zur Audio-Feature-Extraktion bei Musikaufführungen, im Speziellen von Orchestermusik, erarbeitet wurde.

Die Entscheidung fiel auf die *MIRtoolbox*, da sie hauptsächlich für die Nutzung mit Matlab konzipiert wurde. Matlab wird im Fachgebiet der Audiokommunikation vorrangig zur Berechnung von komplexen Algorithmen und Abläufen genutzt, weshalb dieses Merkmal einen hohen Stellenwert bei der Suche nach einer geeigneten Software einnahm. Weiterhin erwies sich der objektorientierte Aufbau, die gute Dokumentation der Toolbox und die Tatsache, dass die Software noch erweitert und weiterentwickelt wird, als ausschlaggebend.

Die Software zur Feature-Extraktion bei Orchestermusik (Lerch, 2008) wurde nicht ausgewählt, da es sich bei den in der Masterarbeit zu untersuchenden Musiken vorrangig um Studioproduktionen handelt und sie deshalb für eine korrekte Nutzung ungeeignet wäre. Weiterhin ergab die Recherche, dass die Software fehlerbehaftet ist.

jAudio ist eine Feature-Extraktionsbibliothek, welche in Java geschrieben wurde. *Yaafe* hat zwar eine Matlab-Schnittstelle, wurde aber seit 2011 nicht mehr aktualisiert.

Zu einer weiteren Vorarbeit gehört auch eine Literaturrecherche zum Thema *Music Information Retrieval* und *Audio Content Analysis* und die Erstellung einer Mindmap auf Grundlage des Buches von Lerch (2012).

5 Arbeits- und Zeitplan

Zeit in Wochen	**Was wird durchgeführt?**
2	Recherche geeigneter Audio-Features und Extraktion
8	Implementierung eines naiven Bayes-Klassifikators in Matlab (inkl. PCA, ICA) und Verknüpfung der Audio-Features und User-Ratings
2	Auswertung der Empfehlungsgenauigkeit (MAE, Prozentsatz korrekt klassifizierter Titel)
3	Erstellung eines Benchmarks und Auswertung
6	Niederschrift der Arbeit
3	Überarbeitungszeit
2	Puffer
26	Wochen (insgesamt; 5,98 Monate)

Literatur

Böhringer, G. (2013): *Dimensionen der Ausdrucksqualität von Produktionsmusik*. unvollendete Masterarbeit (Stand: 18.10.2013), Technische Universität Berlin, Deutschland, Berlin.

Breese, J. S., Heckerman, D. und Kadie, C. (1998): *Empirical analysis of predictive algorithms for collaborative filtering*. In: *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*, 43–52.

Bu, J., Tan, S., Chen, C., Wang, C., Wu, H., Zhang, L. und He, X. (2010): *Music recommendation by unified hypergraph: combining social media information and music content*. In: *Proceedings of the international conference on Multimedia*, 391–400.

Decker, T. (2005): *Datenklassifikation mittels Bayestechniken: angefertigt am Fraunhofer ITWM in Kaiserslautern.* Diplomarbeit, Technische Fachhochschule, Berlin. URL `http://www.itwm.fraunhofer.de/fileadmin/ITWM-Media/Abteilungen/BV/Pdf/Diplomarbeit_Decker.pdf`, Zuletzt geprüft am 25.07.2013.

Lartillot, O. (2013): *MIRtoolbox 1.5: User's Manual*. URL `https://www.jyu.fi/hum/laitokset/musiikki/en/research/coe/materials/mirtoolbox/MIRtoolbox1.5Guide`, Zuletzt geprüft am 27.09.2013.

Lartillot, O. und Toiviainen, P. (2007): *A matlab toolbox for musical feature extraction from audio*. In: *International Conference on Digital Audio Effects*, 237–244.

Lerch, A. (2008): *Software-Based Extraction of Objective Parameters from Music Performances*. Dissertation, Technische Universität Berlin, Deutschland, Berlin.

Lerch, A. (2012): *Audio Content Analysis: An introduction*. Wiley, Hoboken and N.J.

Li, Q., Myaeng, S. H. und Kim, B. M. (2007): *A probabilistic music recommender considering user opinions and audio features*. In: Information processing & management, **43**, 2: 473–487.

Liu, N.-H., Hsieh, S.-J. und Tsai, C.-F. (2010): *An intelligent music playlist generator based on the time parameter with artificial neural networks*. In: Expert Systems with Applications, **37**, 4: 2815–2825.

Lu, C.-C. und Tseng, V. S. (2009): *A novel method for personalized music recommendation*. In: Expert Systems with Applications, **36**, 6: 10035–10044.

Magno, T. und Sable, C. (2008): *A Comparison of Signal-Based Music Recommendation to Genre Labels, Collaborative Filtering, Musicological Analysis, Human Recommendation, and Random Baseline*. In: *Proceedings of the 9th International Conference on Music Information Retrieval*, 161–166. Philadelphia and USA. URL `http://ismir2008.ismir.net/papers/ISMIR2008_157.pdf`, Zuletzt geprüft am 05.10.2013.

Mathieu, B., Essid, S., Fillon, T., Prado, J. und Richard, G. (2010): *YAAFE, an Easy to Use and Efficient Audio Feature Extraction Software*. In: *ISMIR*, 441–446.

Mcennis, D., Mckay, C. und Fujinaga, I. (2005): *JAudio: A feature extraction library*. In: *International Conference on Music Information Retrieval*, 600–603.

Nanopoulos, A., Rafailidis, D., Symeonidis, P. und Manolopoulos, Y. (2010): *Musicbox: Personalized music recommendation based on cubic analysis of social tags*. In: Audio, Speech, and Language Processing, IEEE Transactions on, **18**, 2: 407–412.

Rish, I. (2001): *An empirical study of the naive Bayes classifier*. In: *IJCAI 2001 workshop on empirical methods in artificial intelligence*, Bd. 3, 41–46.

Su, J.-H., Yeh, H.-H., Yu, P. S. und Tseng, V. S. (2010): *Music Recommendation Using Content and Context Information Mining*. In: Intelligent Systems, IEEE, **25**, 1: 16–26.

Wang, X., Rosenblum, D. und Wang, Y. (2012): *Context-aware mobile music recommendation for daily activities*. In: *Proceedings of the 20th ACM international conference on Multimedia*, MM '12, 99–108. ACM, New York and NY and USA. URL `http://doi.acm.org/10.1145/2393347.2393368`, Zuletzt geprüft am 05.10.2013.

Yoshii, K. und Goto, M. (2009): *Continuous pLSI and Smoothing Techniques for Hybrid Music Recommendation*. In: *Proceedings of the 10th International Society for Music Information Retrieval Conference*, 339–344. Kobe and Japan. URL `http://ismir2009.ismir.net/proceedings/OS4-1.pdf`, Zuletzt geprüft am 05.10.2013.

Yoshii, K., Goto, M., Komatani, K., Ogata, T. und Okuno, H. G. (2006): *Hybrid collaborative and content-based music recommendation using probabilistic model with latent user preferences*. In: Proceeding of the 7th International Conference on Music Information Retrieval (ISMIR 2006), **2006**: 296–301. URL `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.79.3718`, Zuletzt geprüft am 05.10.2013.

Yoshii, K., Goto, M., Komatani, K., Ogata, T. und Okuno, H. G. (2008): *An Efficient Hybrid Music Recommender System Using an Incrementally Trainable Probabilistic Generative Model*. In: IEEE Transactions on Audio, Speech, and Language Processing, **16**, 2: 435–447.