

Inhaltsbasierte Musikempfehlung mit Convolutional Neuronalen Netzwerken

WEIDHAS PHILIPP

Matr.nr: 123456

philipp.weidhas@st.oth-regensburg.de

WILDGRUBER MARKUS

Matr.nr: 123456

markus.wildgruber@stud.oth-regensburg.de

Zusammenfassung

Hier kommt die Zusammenfassung...

1. EINLEITUNG

Im ersten Halbjahr des Jahres 2017 wurden 62% der Einnahmen der amerikanischen Musikindustrie durch Streaming Plattformen¹ erzielt. Im Vergleich zu Vorjahr erhöhten sich dadurch die Einnahmen um 48% auf 2.5\$ Milliarde [1]. Dieser Erfolg basiert nicht nur auf einer guten Verfügbarkeit der Lieder und einem günstigen Preis sondern auch auf automatischen Musikempfehlungsdiensten, welche dem Nutzer ein angenehmeres Konsumverhalten ermöglichen. Obwohl Empfehlungsdienste in den letzten Jahren viel erforscht wurden, ist das Problem der Musikempfehlung sehr komplex. Neben einer große Anzahl an verschiedenen Stile und Genres, beeinflussen sowohl soziales- und geographisches Umfeld, sowie der aktuelle Gemütszustand die Vorliebe eines Hörers. [2]
TO DO

Bislang werden Informationen über den Hörer durch ein Benutzerprofil repräsentiert. Das Profil enthält nur wenig Hintergrund Informationen des Hörers und beschränkt sich auf Lieder, die ein Benutzer angehört und bewertet hat [2]. Das Nutzen dieser Daten, um Musikvorschläge abzugeben wird als kollaboratives Filtern (KF) bezeichnet. In der Studie

¹wie Spotify, Apple Music, Pandora etc.

von Vigliensoni und Fujinaga [5] zeigt sich ein deutlicher Unterschied zwischen herkömmlichen Benutzerprofilen und das Einfügen von Zusatzinformationen. Durch das Hinzufügen der Features demographischen Hintergrund und Entdeckergeist des Hörers konnte im Vergleich zu einem herkömmlichen Profil eine 12% besser Genauigkeit erreicht werden.

- neues Lied - bessere ergebnis raten

Der weitere Verlauf der wissenschaftlichen Arbeit ist wie folgt organisiert. Im 2. Abschnitt werden verschiedenen Ansätze in den jeweiligen Methodenbereichen vorgestellt. Im 3. Kapitel werden die erfolgreichsten Ansätze miteinander verglichen. Teil 4 zeigt ein eigenes Experiment zu dem Thema. Abschnitt 5 schließt diese Arbeit ab und diskutiert zukünftige Forschungsrichtungen. //TO DO

2. GRUNDLAGEN DER MUSIKEMPFEHLUNG

In dem Gebiet des Musik Information Retrieval (MIR) gibt es vier Kategorien [3] die einen Einfluß auf die Wahrnehmung von ähnlicher Musik haben.

Musikmerkmale sind Eigenschaften, welche aus dem Audiosignal eines Liedes extrahiert werden. Dazu zählen Aspekte wie der Rhythmus, die Melodie, die Harmonie oder die Stimmung eines Stücks.

Als *Musikkontext* versteht man alle Aspekte, die

nicht aus dem Audiosignal abgeleitet werden, sondern Informationen die über ein Musikstück bekannt sind. Beispielsweise Metadaten wie der Titel eines Lieds, das Genre, Name des Künstlers oder das Erscheinungsjahr.

Die *Benutzereigenschaften* beziehen sie auf Persönlichkeitsmerkmale, wie Geschmack, musikalisches Wissen und Erfahrung oder den demographischen Hintergrund.

Im Unterschied dazu steht der *Benutzerkontext*, der sich auf die aktuelle Situation des Hörers bezieht. Dabei wird er durch seine Umgebung, seiner Stimmung oder der aktuellen Aktivität beeinflusst. [4]

Es gibt verschiedene Methoden, die in Musikempfehlungssystemen verwendet werden: kollaboratives -, merkmalsbasiertes -, kontextbasiertes Filtern und die hybride Methode. Diese werden genutzt, um Informationen aus den genannten Eigenschaften zu gewinnen und diese für Empfehlungen an den Nutzer zu verarbeiten. [6]

2.1 Kollaborativer Filter

Kollaboratives Filtern prognostiziert Vorlieben eines Hörers, indem es aus unterschiedlichen Benutzer-Lied Verhältnissen lernt. Es basiert auf der Annahme, dass Verhalten und Bewertungen andere Nutzer auf eine vernünftige Vorhersage für den aktiven Benutzer schließen lassen [7]. Durch explizite² und implizite³ Rückmeldung eines Hörers an das Empfehlungssystem empfiehlt dieses neue Lieder, indem es Gemeinsamkeiten auf Basis seiner Bewertungen mit dem Nutzungsverhalten anderer Anwender der gleichen Plattform vergleicht [8].

In der praktischen Umsetzung bedeutet dies: hört ein Anwender ein bestimmtes Musikstück. Dann werden ihm, von der Empfehlungsplattform, Lieder vorgeschlagen welche andere Nutzer, die ebenfalls dieses Lied hörten, hören. Dieses Verfahren geht davon aus, dass durch die

² Bewertungen eines Nutzers

Verbindung der Lieder durch vorhergehende Aufrufe eine gute Aussage darüber getroffen werden kann wie gut diese Stücke zusammen passen. Werden Lieder häufig nacheinander gehört (todo), wird diese Verbindung höher bewertet und die Empfehlung häufiger ausgesprochen. Auch wird das Verhalten und der Musikgeschmack des Kunden selbst durch ein System analysiert, um so über Ähnlichkeiten der Kundenpräferenzen mit derer anderer, diesen wiederum bessere Empfehlungen aussprechen zu können. So werden Lieder einem Musikstil zugeordnet und so zielgerichtet dem Nutzer nahegelegt.

Verschiedene Studien ([8][9]) zeigen, dass KF alternative Methoden in der Genauigkeit übertrifft, weshalb es nicht nur im Bereich der Musikempfehlung als die erfolgreichste gilt.

todo(NUP und NSP in einleitung)

Trotz der Popularität des KF gibt es Probleme, die bei der Verwendung dieser Methode beachtet werden müssen. Das Cold-Start Problem besteht darin, dass noch keine Bewertungen für ein Lied vorliegen, wodurch es auch nicht vorgeschlagen werden kann. Dasselbe Problem gibt es bei einem neuen Benutzer: diesem kann kein guter Vorschlag gemacht werden, da es an Information mangelt welche Art von Musik ihm gefällt. Neben dem Cold-Start Problem gibt es noch weitere Probleme. [7]

2.2 Merkmalsbasierter Filter

2.3 Kontextbasierter Filter

2.4 Hybride Methode

Bei hybriden Methoden werden kollaborative, merkmalsbasierte und kontextbasierter Filter miteinander verknüpft, wodurch ein besseres Empfehlungsergebnis mit weniger Nachteilen der einzelnen Methode zu erzielen. Meistens wird ein kollaborativer Filter mit einem der beiden anderem kombiniert.

Als *gewichtet* wird eine hybride Methode be-

³Beobachten des Konsumverhalten

zeichnet, bei der Empfehlungsrate der einzelnen Methoden durch eine Linearkombination zusammengerechnet wird. Das Ergebnis der Linearkombination stellt den Empfehlungswert eines Liedes dar. Durch unterschiedliche Gewichtung der Methoden kann das Empfehlungsergebnis optimiert werden. Der *wechselnde* Ansatz benutzt ein bestimmtes Kriterium anhand dessen es die Methode zur Vorschlagsbestimmung wechselt. Dies kann beispielsweise dann der Fall sein, wenn der erste Filter kein zuverlässiges Ergebnis⁴ liefert. Dann wechselt das System den Filter und kann ein besseres Empfehlungsergebnis bekommen. Bei *gemischten* hybriden Empfehlungen werden unterschiedliche Techniken direkt miteinander vermischt. Dadurch kann für ein System mit inhaltsbasierten Filter das Cold-Start Problem vermieden werden.

// TODO Hybride Methoden können einige Nachteile von kollaborativen Filtern entfernen. Allerdings stehen auch sie vor dem NUP. Dennoch sind hybride Methoden sehr beliebt, da Information über einen neuen Benutzer schnell herausgefunden werden oder durch Profilangaben bereits nach der Registrierung vorhanden sind. [10]

3. NEURONALE NETZEN IN DER MIR

CNN sind durch das biologische Sehen inspiriert und konnten den ersten großen Erfolg im Bereich der Bildklassifizierung [11] verzeichnen. Trotzdem werden CNN auch in verschiedenen Audibereich, wie der Spracherkennung [12] sowie in der MIR mehr genutzt und erforscht.

In der MIR nutzen die ersten Forschungen CNNs, um die Aufgabe der Musikgenre-Klassifizierung [13] zu untersuchen. Die Ergebnisse⁵ zeigen, dass eine automatisierte Klassifizierung die herkömmliche Methode MFCC deutlich übertrifft. Das erste CNN für inhaltsbasierte Musikempfehlung [2] benutzt zunächst

⁴semantische Unterschiede

eine Matrix-Faktorisierung um Eigen Vektoren für alle Lieder zu erhalten. Anschließend wird das Neuronale Netz für die Zuordnung der Audio-Inhalte zu den Eigen Vektoren genutzt. [6]

Im nachfolgenden Absatz wird der Aufbau, das Training und die Optimierung eines CNN beschrieben.

3.1 Convolutional Neuronale Netze

todo

Im Unterschied zu regulären DNN verwendet das CNN Neuronen, die drei Dimensionale angeordnet sind. Durch diese Anordnung ist es möglich größere Inputdaten in derselben Geschwindigkeit zu verarbeiten wie zuvor [14]. Um eine CNN Architektur zu erstellen werden drei Haupttypen von Schichten verwendet: Convolutional Layer (CL), Pooling Layer (PL) und ein Fully-Connected Layer (FCL).

3.1.1 Schichten eines Convolutional Neuronale Netzwerks

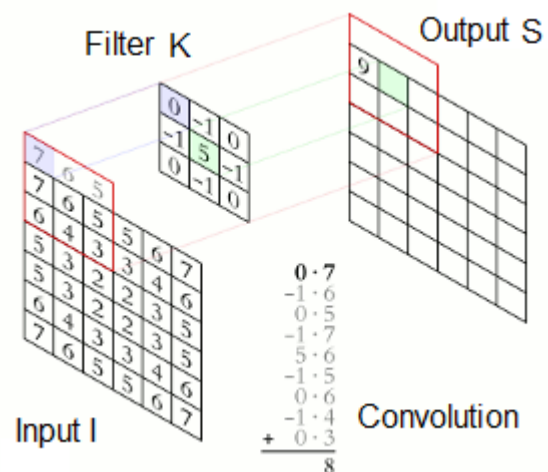


Abbildung 1: Faltung einer 6x6 Matrix mit einem 3x3 Filter [16]

⁵richtigen Klassifizierung

Convolutional Layer

In einem CL findet eine Faltung der Eingangsdaten, in Form einer Matrix, und einem oder mehreren Filtern statt. Ein Filter dient beispielsweise zur Glättung oder zur Verkleinerung der Daten. Eine Verkleinerung der Eingangsmatrix findet statt, wenn ein Filter ohne Zero-padding⁶ verwendet wird. Die Parameter eines Filters werden zufällig initialisiert, können aber mit Hilfe eines Optimierungsverfahrens (3.1.3) angepasst werden. Werden mehrere Filter auf die Eingangsdaten angewendet, ändert sich die Tiefe der gesamten Ausgangsmatrix entsprechend der Anzahl der Filter. [14]

In Abbildung 1 ist die Eingabematrix I eine 6x6 Matrix und K ein 3x3 Filter. Die Ausgangsmatrix S wird an den Stellen (i,j) durch die Gleichung (1) berechnet. Eine genauere Herleitung der Gleichung findet der Leser u. a. bei [15](328f).

$$S(i,j) = (I \star K)(i,j) \quad (1)$$

$$(I \star K)(i,j) = \sum_m \sum_n I(i+m, j+n) K(m,n) \quad (2)$$

Pooling Layer

Ein PL wird zwischen zwei CL eingefügt. Ihre Funktion besteht darin, die Größe der Daten zu reduzieren und damit die Anzahl der Parameter für das nächste CL. Durch die Reduzierung wird die Berechnung des gesamten Netzwerkes beschleunigt. [14]

Ein PL wandelt die Ausgabe eines CL, durch eine statistische Zusammenfassung von nebeneinander liegenden Ausgängen um. Verschiedene Methoden für ein PL sind: Max Pooling [17], eine Übergabe der größten Zahl in einem rechteckigen Umfeld; die Durchschnittsberechnung des Umfeldes oder ein gewichteter Durchschnitt basierend auf der Entfernung eines zentralen Punktes [15](355).

⁶Eine Matrix wird am Rand um Nullen erweitert.
Bsp. aus einer 7x7 Matrix wird eine 9x9 Matrix

Abbildung 2 zeigt einen 2x2 Max-Filter, der auf eine 4x4 Datenmatrix angewandt wird. Die Verschiebung oder Stride des Filters ist 2 d.h. der Filter wird zunächst auf der y-Achse verschoben. Erreicht er dort das Ende wird er um eine Stride auf der x-Achse verschoben und beginnt wieder mit der y-Verschiebung.

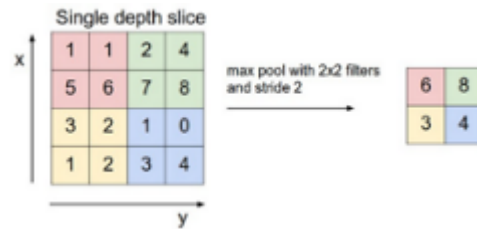


Abbildung 2: Maxpooling mit einem 2x2 Filter[14]

Fully-Connected Layer

toDo

Neuronen in einer FCL haben Verbindungen zu allen Knoten der vorherigen Schicht. Ihre Aktivierung wird durch eine Matrixmultiplikation und einem Bias-Offset berechnet [14]. Die FCL wird als Ausgangsschicht verwendet um aus der Eingangsmatrix einen Vektor zu erzeugen.

3.1.2 Training

Cross entropie

3.1.3 Optimierung

Dropout mini batch verfahren

3.2 Musikempfehlung mit Neuronalen Netzwerken

Anwendung eines CNN zur automatischen Musikempfehlung

Ansatz: Um das TODO Verzweigung TODO vorgestellte Problem der Semantischen Lücke zu lösen, kommt ein solches neuronales Netz zum Einsatz [2]. Dieser Ansatz wird wie folgt umgesetzt: das Netz wird darauf trainiert latente Faktoren aus einzelnen Musikstücken zu generieren, welche für eine Empfehlung verwendet werden. Dieses Verfahren wird anschließend im Vergleich mit einem konventionellen Ansatz, welcher dem Bag-of-Words Prinzip folgt. Daraus folgend wird beurteilt, inwiefern mittel CNN Eigenschaften zur Spezifizierung des Nutzergeschmacks generiert und ausgewertet werden können. In den folgenden Absätzen TODO wirklich Absätze? TODO wird auf diesen Ansatz nun genauer eingegangen.

TODO Unterkapitel einfügen TODO

Als Datenbasis zur Umsetzung dieses neuen Ansatzes wurden verschiedene Musik Datensätze in Betracht gezogen. Diese stehen im Zusammenhang mit dem Million Song Dataset TODO Abkürzung TODO, dieser Datensatz verfügt über Metainformationen und bereits analysierter Audio Informationen von einer Millionen Liedern. Ebenso ist dieser Datensatz öffentlich zugänglich und kann kostenlos heruntergeladen werden, des Weiteren stellt dieser derzeit die größte Forschungsdatenbasis im Gebiet der Musikanalyse dar. Zwei Datensätze im Umfeld des MSD sind von besonderem Interesse, der Echo Nest Taste Profile Subset Datensatz und der The Last.fm Datensatz. Die Problematiken mit diesen Datensätzen, waren zum einen ein schlechte Dokumentation der Generierung der Informationen und der Daten selbst, sowie das keine unverarbeiteten Musikquellen mitgeliefert werden. Das Problem der nicht vorhandenen Rohdaten konnte beseitigt werden, in dem für 99% des Datensatzes Musikschnipsel mit der Dauer von 29

Sekunden von 7digital.com bezogen wurden. Auf der Seite 7digital.com können einzelne Musikstücke 30 Sekunden lang Probegehört werden. TODO anders erläutern TODO TODO generell zu gleich mit dem Million Song Dataset Eintrag im Original Dokument, noch anpassen vll erweiter oder umändern TODO. Der ENTPS sticht durch seine Eigenschaft als größte, bereits mittels kollaborativen Filtern ausgewerteten, Informationsbasis hervor und bietet sich dadurch für ein Weiterverarbeitung mittels CNN an. Auch kann durch die Größe der Daten eine realitätsnaher Versuch unternommen werden. Durch die Verwendung dieses Datensatzes, ist zu jedem Lied, pro Nutzer, die genaue Anzahl gespeichert, wie häufig das Stück vom Nutzer angehört wurde. Wie bereits im Kapitel TODO VERWEIS TODO erklärt wird beim Kooperativen Filter angenommen das ein Nutzer ein Lied nur dann oft hört, wenn es ihm auch gefällt. Um diese Daten für ein Training des neuronalen Netztes zu verwenden, muss ein spezieller Algorithmus angewandt werden. Die üblichen Algorithmen, welche für ein Training eines neuronalen Netzes verwendet werden, dessen Ziel eine Errechnung von Bewertungen haben, wie TODO BEISPIEL nennen TODO können nicht auf Basis dieser Daten zum Einsatz kommen, denn sie können nicht verarbeiten wenn für ein Lied keine Wertung vorliegt. TODO Satz auftrennen TODO Der Umstand das ein Lied keine Wertung erhalten hat, kann mehrere Gründe haben, unter anderem das ein Nutzer dieses Lied schlicht und ergreifen nicht kennt. Der Nutzer könnte allerdings das Lied bereits kennen, aus anderen Quellen, es nicht mögen und deshalb diesen Titel nicht anhören. Dies führt zu zwei unterschiedlichen Szenarien auf Grundlage der gleichen Wertung. Um diesen Umstand richtig bewerten zu können, muss der Algorithmus flexibel sein. Deshalb kommt ein weighted matrix factorization Algorithmus zum Einsatz. Er wurde ursprünglich entwickelt um bei der Bewertung von Fernsehshows und deren automatisierter Empfehlung dem Nutzer gegenüber möglichst

gute Ergebnisse zu erzielen. Auch in diesem Fall gab es rein implizite Rückmeldung seitens des Nutzers. Es wurde gespeichert wie oft ein Nutzer ein Fernsehformat angesehen hat. Auch hier tritt der Fall auf das für Formate von einzelnen Nutzern keine Bewertungen vorliegen. Dies konnte mehrere Gründe haben: der Nutzer kennt das Format nicht, eine Sendung die er noch lieber mochte kommt zur gleichen Sendezeit oder er mag die Sendung nicht. Folglich auch hier eine Ausgangswertung und drei Verschiedene Schlussfolgerungen möglich.

Um einzelne Titel bewerten zu können kommt ein Taste Profile Subset (TODO schöne Übersetzung finde) zum Einsatz. In diesem Subset werden für jeden Song die Anzahl der Abspielungen und der Nutzer festgehalten. Basierend auf der Annahme ein Nutzer würde einen Song nur dann mehrmals hören wenn er ihm gefällt. Dies wird als implizite Bewertung verarbeitet, man weiß nun wie oft ein Nutzer den Track gehört hat, nicht allerdings wie gut er im tatsächlich gefällt. Hat ein Musikstück aber noch keine Bewertung auf Basis von Abspielhäufigkeit erhalten, dadurch das der Nutzer dieses Stück noch nicht angehört haben kann dies allerdings auch daran liegen das er das Lied einfach nicht kennt. Ein Rückschluss darauf das es im Gefällt oder nicht gefällt kann dadurch nicht mit Sicherheit getroffen werden. TODO noch umformulieren TODO Auf dieser Basis kann nicht mit Typischen Matrixen gearbeitet werden, deshalb kommt ein weighted matrix factorization (WMF) algorithmus zum Einsatz. TODO genauer erläutern TODO

Um weitere Faktoren für eine Bessere Bewertung mit einfließen lassen zu können wurden folgende Merkmale definiert:

Diese definieren und umschreiben. Alles genauer erläutern und die Quellen einfügen.

- Extract MFCCs from the audio signals. We computed 13 MFCCs from windows of 1024 audio frames, corresponding to 23 ms at a sampling rate of 22050 Hz, and a hop size of 512 samples. We also computed first and second order differences, yielding 39 coef-

ficients in total. • Vector quantize the MFCCs. We learned a dictionary of 4000 elements with the K-means algorithm and assigned all MFCC vectors to the closest mean. • Aggregate them into a bag-of-words representation. For every song, we counted how many times each mean was selected. The resulting vector of counts is a bag-of-words feature representation of the song.

Jetzt die daraus entstehende Umsetzung ins Neuronale Netz einpflegen. Linear Parallel

Vorgehensweise

Erläutern der Objective function

Da wir selbst scheinbar kein Experiment machen sollen auswertung deren Experiment und möglicherweise noch die Erläuterung derer Datensätze des Millions songs Teils. Ohne Eigenes Experiment ist es denke ich Eminent wichtig um den Erfolg mittels CNN darlegen zu können.

3.3 Hybride Musikempfehlung mit einem Neuronalen Netzwerk

Im Unterschied zu der zuvor dargestellten Forschung (3.2) wird in der jetzigen ein Deep Belief Netzwerk (DBN) verwendet, um ein hybrides inhaltsbasiertes Musikempfehlungssystem zu entwickeln. Bisherige inhaltsbasiertes Systeme verfolgen typischerweise einem zweistufigen Ansatz: zunächst extrahieren sie aus Audioinhalte den MFCC Koeffizienten; anschließend prognostizieren sie Musikpräferenzen eines Nutzers. Das nachfolgende Modell führt dieses beiden Schritte simultan und automatisch aus. [6]

Das hybride Modell basiert auf einem hierarchischen linearen Modell mit einem Deep Belief Netzwerk (HLDBN), dass zunächst erläutert wird, um anschließend die Funktionsweise des hybriden Systems darzustellen.

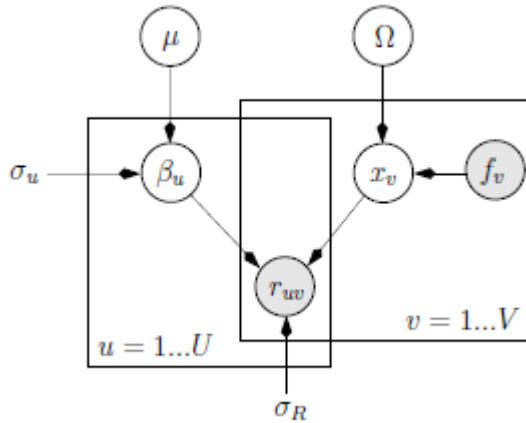


Abbildung 3: Hierarchisches lineares Modell eines Deep Belief Netzwerks [6]

3.3.1 Hierarchisches lineares Modell mit einem Deep Belief Netzwerk

Das in Abbildung 3 gezeigte Modell ist wie folgt definiert: f_v sind Musikmerkmale eines Liedes v , die durch den Eigenvektor x_v automatisch errechnet werden. Die bevorzugte Musik eines Benutzer u wird als Vektor β_u bezeichnet. Ω bezeichnet die Parameter, die das DBNs lernt. Die Bewertung, die u einem Lied v gibt, ist ein Skalarprodukt von r_{xv} und β_u . Durch σ_R wird die Varianz aller Bewertungen des Nutzers betrachtet. μ repräsentiert den allgemeinen Musikgeschmack aller Benutzer, wobei σ_u die Varianz des einzelnen Nutzers definiert. Alle Benutzer und Lieder Paare werden als I bezeichnet. Für eine Regularisierung der Werte wird die Gaußsche Normalverteilung \mathcal{N} verwendet.⁷ [6]

Das Modell wird wie folgt formuliert:

$$\begin{aligned} r_{xv} &\sim \mathcal{N}(\beta'_u x_v, \sigma_R^2) \\ \beta &\sim \mathcal{N}(\mu, \sigma_u^2 I) \\ x_v &= \text{DBN}(f_v; \Omega) \end{aligned}$$

Für das Training des Systems wird die Maximum Likelihood-Funktion oder auch Cross-Entropy verwendet. Als Optimierungsmethode wird das stochastische Mini-Batch Verfahren

⁷ $\mathcal{N}(a,b)$ ist die Normalverteilung mit Mittelwert a und Varianz b . $x \sim p$ zeigt, dass x die Verteilung p erfüllt

genutzt, um ein Overfitting der Parameter zu vermeiden. Nach der Lernphase kann r_{xv} geschätzt werden, wodurch auch neue Lieder empfohlen werden können. [6]

3.3.2 Hybrides Modell mit einem Deep Belief Netzwerk

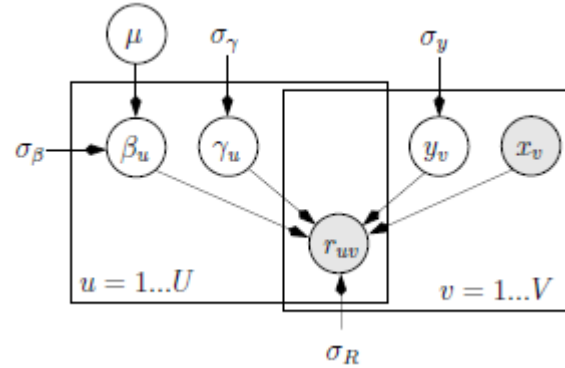


Abbildung 4: Hybrides Empfehlungs Modell [6]

4. VERGLEICH DER VORGESTELLTEN MODELLE

Abschließend wird nun ein Vergleich zwischen dem vorgestellten CNN Ansatz einerseits und dem danach folgendem DBN gezogen. In Versuchen [2] wurde festgestellt, dass das CNN Modell einem BOW System überlegen ist und eine bessere Empfehlungsrate erreicht. Das anschließend erläuterte Versuchsmodell, welches mittels DBN einen hybriden Methode verfolgt, konnte im direkten Vergleich zu CNN eine nochmals verbesserte Empfehlungsgenauigkeit erreichen. Vergleichende Versuchsreihen [6] haben folgendes festgestellt: ein nicht hybrider Ansatz, welcher allein auf Training der beiden Netze ergab, dass das vorgestellte HLDBN Modell genauere Ergebnisse lieferte als das vorgestellte CNN. Die Integration der beiden Modelle in einen hybriden Aufbau ergab wiederum ebenso, dass der Einsatz des HLDBN Netzes zu genaueren Empfehlungsraten führte als die

Verknüpfung von CNN und CF. Die Empfehlungsraten für bereits bekannte Lieder, konnten somit im Vergleich zu klassischen Ansätzen verbessert werden. Auch das bereits eingeführte Problem des NSP konnte durch den Einsatz Neuronaler Netze gelöst werden. Sowohl der Einsatz des CNN als auch der des HLDBN Netzes führen hierbei zum Erfolg.

Sowohl der im Punkt 3.2 vorgestellte ansatz mittels eines CNNs sowie der in Punkt 3.3 erläuterte Ansatz mit Einsatz eines hybriden dbn Netzwerkes eine gute Möglichkeit zur automatisierten Musikempfehlung. Beide Verfahren haben in Rahmen von Versuchen bewiesen, dass sie sowohl zuverlässig sind, aber auch das sie einen klassischen Verfahren wie ein BoW-System

LITERATUR

- [1] Joshua P. Friedlander. News and notes on 2017 mid-year riaa revenue statistics. *RIAA*, 2017.
- [2] Aäron van den Oord, Sander Dieleman, and Benjamin Schrauwen. Deep content-based music recommendation. *Advances in Neural Information Processing Systems* 26, 2013.
- [3] Markus Schedl, Arthur Flexer, and Julián Urbano. *The neglected user in music information retrieval research*, volume 36. Springer, 2013.
- [4] Peter Knees and Markus Schedl. *Music Similarity and Retrieval*, volume 41. Springer, 2016.
- [5] Gabriel Vigliensoni and Ichiro Fujinaga. Automatic music recommendation systems: Do demographic, profile, and contextual features improve their performance? *Proceedings of the 17th International Society for Music Information Retrieval Conference*, pages 94–100, 2016.
- [6] Xinixi Wang and Ye Wang. Improving content-based and hybrid music recommendation using deep learning. *Proceedings of the ACM International Conference on Multimedia*, pages 627–636, 2014.
- [7] Òscar Celma. *Music Recommendation and Discovery - The Long Tail, Long Tail, and Long Play in the Digital Music Space*. Springer, 2010.
- [8] B. McFee, T. Bertin-Mahieux, D. P. W. Ellis, and G. R. G. Lanckriet. The million song dataset challenge. *21st International Conference Companion on World Wide Web*, pages 909–916, 2012.
- [9] Luke Barrington, Reid Oda, and Gert Lanckriet. Smarter than genius? human evaluation of music recommender systems. *Proceedings of the 10th International Society for Music Information Retrieval Conference*, pages 357–362, 2009.
- [10] Robin Burke. Hybrid recommender systems: Survey and experiments. *User Model. User-Adapt. Interact.*, pages 331–370, 2002.
- [11] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [12] Alex Graves, Abdel-Tahman Mohamed, and Geoffrey E. Hinton. Speech recognition with deep recurrent neural networks. *Acoustics, Speech and Signal Processing, IEEE International Conference on*, pages 6645 – 6649, 2013.
- [13] Honglak Lee, Yan Largman, Peter Pham, and Andrew Y. Ng. Unsupervised feature learning for audio classification using convolutional deep belief networks. *Advances in Neural Information Processing Systems*, 2009.

- [14] Andrej Karpathy. *Convolutional Neural Networks for Visual Recognition*. Stanford University, 2017. <https://github.com/cs231n/cs231n.github.io/blob/master/convolutional-networks.md#conv>. <http://www.deeplearningbook.org>.
- [15] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.
- [16] wikipedia, 2017. https://de.wikipedia.org/wiki/Convolutional_Neural_Network#/media/File:3D_Convolution_Animation.gif.
- [17] Zhou Y. and Chellappa R. Computation of optical flow using a neural network. *IEEE International Conference*, 71–78, 1988.