Vermeidung von Overfitting bei der Vorhersage von subjektiven Tonalitätsbewertungen auf Basis künstlicher neuronaler Netze mit kleinen Trainingsmengen

Tim Holzhäuser, Robert Rosenkranz und M. Ercan Altinsoy
Lehrstuhl für Akustik und Haptik, TU Dresden, 01062 Dresden, Deutschland
Email: Tim.Holzhaeuser@tu-dresden.de

Einleitung

Im Produktdesign wird der akustischen Emission besondere Aufmerksamkeit geschenkt, da sie das Nutzungserlebnis prägt. Tonale Anteile im Geräuschspektrum können als lästig wahrgenommen werden. Vorhandene Ansätze zur Messung von Tonalität, basierend auf Hörmodellen ([1] [2]), zeigen gute Leistungen für synthetische Geräusche, sind jedoch nur bedingt für komplexe Alltagsgeräusche einsetzbar. Die Bestimmung der Tonalität von Alltagsgeräuschen erfolgt daher oft durch Hörversuche. Da Hörversuche mit deutlichem Zeit- und Kostenaufwand verbunden sein können, wären auch in diesem Fall modellbasierte Vorhersagen wünschenswert. Für die Modellierung komplexer Zusammenhänge eignen sich künstliche neuronale Netze (KNN), wobei die Modellierungsqualität in Allgemeinen mit der Anzahl der Trainingsbeispiele steigt. Bedingt durch den Aufwand von Hörversuchen, stehen meist nur kleine Mengen an Trainingsdaten zur Verfügung. Kleine Trainingsdatenmengen enthalten oft mehr statistisches Rauschen als große Trainingsdatenmengen. Modelliert ein KNN das statischen Rauschen einer Trainingsdatenmenge anstatt deren unterliegendes Prinzip, so spricht man von Overfitting. Dieser Beitrag schlägt einen grundlegenden Ansatz zur quadratischen Vergrößerung der Trainingsmenge durch die Transformation von ratioskalierten Daten in ordinalskalierte Daten vor. Die zusätzlichen Trainingsdaten erhöhen die Anzahl der Randbedingungen die ein Modell erfüllen muss, und verringert so die Chance für Overfitting.

Methode

Ziel dieser Arbeit war es, basierend auf subjektiven Tonalitätsbewertungen für eine Auswahl von Stimuli, künstliche neuronale Netze zu trainieren. Trainierte Modelle können dazu verwendet werden, die subjektiv empfundene Tonalität von Stimuli mit unbekannter Tonalitätsbewertung zu schätzen. Im einfachsten Fall schätzt ein KNN direkt die Tonalität eines Stimulus (Abbildung 1). Dieser Ansatz wird als Direkte Regression (DR) bezeichnet. Die Anzahl der Stimuli und Datensätzen sind bei diesem Ansatz identisch, was bei der üblicherweise geringen Anzahl von Stimuli bei Hörversuchen die Gefahr von Overfitting birgt. Eine Möglichkeit diesem Problem zu begegnen, ist die Erzeugung künstlicher Datensätze aus vorhandenen Datensätzen durch die Nutzung anwendungsspezifischer Invarianten (*dataset augmentation*). Typische Techniken für die Erzeugung künstlicher

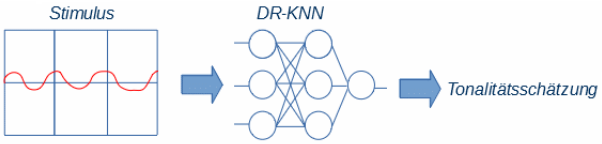

Abbildung 1: Datenfluss bei der Schätzung der subjektiven Tonalitätsbewertung eines Stimulus mit Direkter Regression.

Audiodaten sind: Veränderung der Tonhöhe, Aufaddieren geringer Rauschanteile, Variation von Schalldruckpegel und Wiedergabegeschwindigkeit, harmonische Verzerrung, MP3-Kompression und Frequenzmaskierung [3]. Für die vorliegende Arbeit sind diese Techniken ungeeignet, da sie die wahrgenommene Tonalität von Stimuli verändern können. Diese Arbeit schlägt einen alternativen Ansatz zur Erhöhung der Anzahl von Datensätze vor, der als Paarvergleich (PV) bezeichnet wird. Ein mit diesem Ansatz trainiertes künstliches neuronales Netz (PV-KNN) erhält zwei Stimuli s_i und s_j als Eingabe und schätzt ob die Tonalität von s_i größer oder kleiner als die Tonalität von s_j ist (Abbildung 2). Formal berechnet ein PV-KNN die Klassifikationsfunktion (1)

$$PC(s_i, s_j) = \begin{cases} 0 & \hat{T}(s_i) < \hat{T}(s_j) \\ 1 & \hat{T}(s_i) > \hat{T}(s_j) \end{cases} \quad (1)$$

wobei $\hat{T}(s)$ die geschätzte Tonalität von Stimulus s ist. Der Sonderfall $\hat{T}(s_i) = \hat{T}(s_j)$ ist sehr unwahrscheinlich, solange $s_i \neq s_j$ gilt und wurde deshalb für diese Arbeit nicht betrachtet. Um mithilfe des PV-Ansatzes einen Schätzwert für die Tonalität eines Stimulus zu erhalten, wird eine Referenzmenge $R = \{s_1, s_2, \ldots, s_n\}$ von Stimuli mit bekannter Tonalitätsbewertung T benötigt. Die Bewertungen der Stimuli aus R sollten den zulässigen Tonalitätswertebereich der jeweiligen Anwendung gleichmäßig abdecken. Zudem soll für alle Paare von Stimuli $\forall s_i, s_j \in R$, $T(s_i) \neq T(s_j) | i \neq j$ gelten. Die Stimuli der Referenzmenge sind aufsteigend nach ihrer Tonalitätsbewertung sortiert. Mithilfe eines PV-KNN wird ein Stimulus s_u mit unbekannter Tonalitätsbwertung mit allen Stimuli der Referenzmenge R paarweise verglichen. Sei $E = [PC(s_1, s_u), \ldots, PC(s_n, s_u)]$ die Liste der Resultate dieser paarweisen Vergleiche. Im nächsten Schritt werden zwei Indizes t_{low} und t_{high} so bestimmt, dass für alle

Elemente aus E (2) gilt.

$$PC(s_i, s_u) = \begin{cases} 0 & i < t_{low} \\ 1 & i > t_{high} \end{cases} \tag{2}$$

Für den Fall $t_{high} - t_{low} = 1$ errechnet sich die geschätzte Tonalität des Stimulus s_u aus (3).

$$\hat{T}(s_u) = \frac{1}{2}(T(s_{t_{high}}) + T(s_{t_{low}})) \tag{3}$$

Typischerweise erzeugt ein PV-KNN mit einer gewissen Wahrscheinlichkeit falsche Klassifikationsergebnisse, wodurch $t_{high} - t_{low} > 1$ gilt. In diesem Fall erfolgt eine Auftrennung der Elemente $PC(s_i, s_u)|t_{low} \leq i \leq t_{high}$ nach ihrem Klassifikationsergebnis in zwei Listen L_1 (4) und L_2 (5).

$$L_1 = [T(s_i)|PC(s_i, s_u) = 0] \tag{4}$$
$$L_2 = [T(s_i)|PC(s_i, s_u) = 1] \tag{5}$$

Die geschätzte Tonalität errechnet sich aus dem Mittelwert der Medianwerte von L_1 und L_2 nach (6).

$$\hat{T}(s_u) = \frac{1}{2}(\tilde{L}_1 + \tilde{L}_2) \tag{6}$$

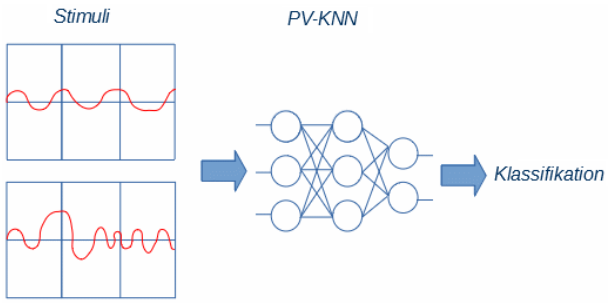

Abbildung 2: Veranschaulichung der Funktionsweise des Paarvergleichs. Ein PV-KNN schätzt welcher der beiden Stimuli die höhere subjektive Tonalitätsbewertung besitzt.

Experiment

Als Basismenge für die Experimente fungierten 5 Tonaufnahmen von Elektromotoren. Neben Variationen der Schalldruckpegel wurden Bandsperren, Hoch- und Tiefpassfilter eingesetzt, um aus der Basismenge 180 zusätzliche Stimuli zu erzeugen. In Hörversuchen erfolgte die Bewertung der Tonalität der 185 Stimuli durch Probanden auf einer quasikontinuierlichen Skala von 0 bis 100. Die Probandengruppe bestand aus 13 Personen (6w, 7m) im Altersbereich von 23 bis 63 Jahren. Die in dieser Arbeit verwendeten KNN-Modelle benötigen Eingabedaten konstanter Größe. Um dies zu erreichen, fanden jeweils nur die ersten 2 Sekunden Spielzeit der nicht-transienten Stimuli Verwendung. Anschließend wurde für jeden Stimulus eine Datenmatrix der spezifischen Prominenz im Zeitverlauf mithilfe der Software HEAD ArtemiS [4] ermittelt. Der Effektivschalldruckpegel (A-Bewertung) eines jeden Stimulus wurde der zugehörigen Datenmatrix als zusätzlichen Zeile angefügt. Die abschließende elementweise Berechnung des dekadischen Logarithmus reduziert die betragsmäßige Varianz der Datenmatrixelemente und erleichtert das Training der KNNs.

Erzeugung der Datensätze für den PV-Ansatz

Ein PV-Datensatz besteht aus der Kombination von zwei Datenmatrizen. Bei n Datenmatrizen existieren maximal n^2 Kombinationsmöglichkeiten. Kombinationen einer Datenmatrix mit sich selbst werden nicht betrachtet, wodurch n Kombinationsmöglichkeiten wegfallen. Desweitern können Paare von Datenmatrizen d_i und d_j auf zwei Arten zu Datensätzen kombiniert werden: (d_i, d_j) und die symmetrische Variante (d_j, d_i). Der Einfluss symmetrischer Kombinationen auf die Trainingsergebnisse wurde im Experiment untersucht. Werden symmetrische Kombinationen nicht berücksichtigt, so halbiert sich die Anzahl der verfügbaren Datensätze. Desweitern ist bei der Erzeugung der Datensätze zu berücksichtigen, dass die Tonalitätsbewertungen T der Stimuli Stichprobenmittelwerten darstellen, die von den Mittelwerten der Gesamtpopulation abweichen können. Dies hat zur Folge, dass das Ergebnis des Größenvergleichs zweier Tonalitätsbewertungen vom Ergebnis des Größenvergleichs der zugehörigen Populationsmittelwerte abweichen kann (insbesondere bei einer kleinen betragsmäßigen Differenz der Tonalitätsbewertungen in Kombination mit hoher Varianz in den Stichproben). Zur Reduktion der Fehlentscheidungswahrscheinlichkeit wurde die Stichproben mithilfe der Effektstärke Cohens d [5] verglichen. Dabei wurden nur Stimulikombinationen mit einer Mindesteffektstärke der Trainingsdatenmenge zugefügt. Eine hohe Effektstärke verringert die Fehlentscheidungswahrscheinlichkeit, verringert aber tendenziell auch die Anzahl der verfügbaren Datenmatrizenpaare, was die Gefahr für Overfitting erhöht. Zur Ermittlung einer geeigneten Effektstärke wurden in mehreren Trainingsläufen verschiedenen Werten für Cohens d getestet.

Training der KNNs

Aus der Basismenge wurden zufällig 122 Stimuli zur Erzeugung von Trainingsdatensätze ausgewählt. Mithilfe der restlichen 63 Stimuli wurde die Modellleistungsfähigkeit getestet. Trainingsdatensätze bestehen aus 3498 (DR-Ansatz) bzw. 6996 (PV-Ansatz) Einzelwerten. Für die Verarbeitung von hochdimensionalen Datensätzen eignen sich *Convolutional Neural Networks* (CNN). Durch die spezielle Architektur von CNNs ist die Anzahl der Modellparameter (Gewichte + Schwellen) nur im geringem Maße von der Dimension der Datensätze abhängig. Ein betragsmäßig kleines Verhältnis von Modellparametern zur Anzahl von Trainingsdatensätzen reduziert die Chance für Overfitting. Als Trainingsalgorithmus fand das Gradientenabstiegsverfahren Adam [6] mit einer Lernrate von $0,01$ Verwendung. Jeder Trainingslauf erfolge über 150 Epochen. Tabelle (1) zeigt die verwendete Netzwerkarchitektur für den PV-Ansatz.

Resultate

Messung der Leistungsfähigkeit

Zur Beurteilung der Leistungsfähigkeit der beiden Ansätze fand der Approximationsfehler $Err(a)$ über alle

Tabelle 1: Architektur des PV-KNN: Die Notation der Netzwerkschichten orientiert sich am *Keras*-Framework [7].

$Convolution2D(5, kernel = (3,3), relu)$
$MaxPool2D(pool\_size = (2,2))$
$BatchNormalization$
$Convolution2D(5, kernel = (3,3), relu)$
$MaxPool2D(pool\_size = (2,2))$
$BatchNormalization$
$Convolution2D(5, kernel = (3,3), relu)$
$MaxPool2D(pool\_size = (2,2))$
$BatchNormalization$
$Flatten$
$Dense(2, softmax)$

n Testbeispiele Verwendung (7),

$$Err(a) = \frac{1}{n} \sum_{i=0}^{n} |\hat{T}_a(s_i) - T(s_i)| \tag{7}$$

wobei $\hat{T}_a(s_i)$ die geschätzte Tonalität für den Stimulus s_i durch den Ansatz a darstellt. Um den Einfluss der zufälligen Initialisierung von Gewichten und Schwellen auf die Leistungsfähigkeit der KNNs zu ermitteln, wurden mehrere Trainingsläufe durchgeführt. Die Schwankung in der Leistungsfähigkeit der KNNs hat Schwankungen bei der Schätzung der Tonalität durch die untersuchten Ansätze zur Folge. Sei $M_a = \{Err_1(a), ..., Err_k(a)\}$ die Multimenge von k Approximationsfehlern durch den Ansatz a. Der Gesamtapproximationsfehler $ErrSet(M)$ errechnet sich aus dem arithmetischen Mittel aller Elementen von M nach (8).

$$ErrSet(M) = \frac{1}{|M|} \sum_{\forall m \in M} m \tag{8}$$

Ermittlung geeigneter Werte für Cohens d
Tabelle (2) zeigt die Resultate der Versuche zur Ermittlung eines geeigneten Wertes für Cohens d. Vier mögliche Ausprägungen für die Effektstärke wurden betrachtet [5]. Pro Ausprägung wurden zwei Sätze von Trainingsdaten erzeugt (mit und ohne symmetrische Kombinationen). Dargestellt sind der Gesamtapproximationsfehler $ErrSet$ sowie die Standardabweichung $SdSet$ einer Menge von je 5 Approximationsfehlern. Der niedrigste Gesamtapproximationsfehler $ErrSet$ ergab sich für Cohens $d = 0,2$. Zudem wird deutlich, dass die Verwendung symmetrischer Kombinationen in den Trainingsdaten die Leistungsfähigkeit der PV-KNNs verbessert (verringerter Gesamtapproximationsfehler sowie verringerte Standardabweichung). Ein Verzicht auf die Verwendung symmetrischer Kombinationen ist nur sinnvoll, wenn kurze Trainingsrechenzeiten angestrebt werden.

Vergleich von DR-Ansatz und PV-Ansatz
Tabelle (3) zeigt die wichtigsten Resultate des Vergleichs von DR-Ansatz und PV-Ansatz. Die dargestellten Leistungsheuristiken ($ErrSet$ und $SdSet$) fassen die Ergebnisse von 21 Trainingsläufen zusammen. Durch die erhöhte Anzahl an Trainingsdatensätzen (generiert mit

Tabelle 2: Abweichung geschätzter Tonalität von wahrgenommener Tonalität zur Ermittlung eines optimalen Wertes für die Effektstärke Cohens d

Cohens d	Keine Symmetrie		Symmetrie	
	$ErrSet$	$SdSet$	$ErrSet$	$SdSet$
0	10,33	7,60	8,99	7,54
0,2	8,66	6,62	8,59	6,53
0,5	9,93	8,02	9,04	6,80
0,8	10,77	7,45	9,20	6,64

Cohens $d = 0.2$) konnte der Gesamtapproximationsfehler von $ErrSet = 13,40$ auf $ErrSet = 8,94$ verbessert werden. Durch den Vergleich der Konfidenzintervalle beider Ansätze wird ersichtlich, dass die Verbesserung in der Vorhersagequalität von subjektiven Tonalitätsbewertungen durch den PV-Ansatz statistisch relevant ist. Ein Vergleich der Ansätze mit einem t-Test ([8]) zeigt höchst signifikante Ergebnisse (p-Wert$< 0,001$).

Tabelle 3: Vergleich der beiden untersuchten Ansätze zur Schätzung subjektiver Tonalitätsbewertungen.

Methode	DR	PV
Trainings-datensätze	122	12428
Testdatensätze	63	3304
Trainingsläufe	21	21
ErrSet	13,40	8,94
SdSet	4,30	2,64
Konfidenz-intervall $\gamma = 99\%$	[10,99; 15,82]	[7,45; 10,42]

Zusammenfassung und Ausblick

Durch die Transformation von ratioskalierten Daten in ordinalskalierte Daten kann der Einfluss von Overfitting bei der Vorhersage von subjektiven Tonalitätsbewertungen verringert werden. Die erhöhte Anzahl an verfügbaren Trainingsbeispielen senkt den durchschnittlichen Approximationsfehler und reduziert die Schwankungen in der Leistungsfähigkeit trainierter Modelle. Der Paarvergleichsansatz ist prinzipiell für alle Lernprobleme mit rationskalierten Daten anwendbar. Deshalb erscheint die Untersuchung der Anwendbarkeit des PV-Ansatz für die Schätzung von subjektiven Bewertungen anderer psychoakustischer Größen aussichtsreich.

Literatur

[1] ECMA-74: Measurement of Airborne Noise emitted by Information Technology and Telecommunications Equipment (16th ed) Ecma International (2019-06)

[2] DIN 45681: Akustik-Bestimmung der Tonhaltigkeit von Geräuschen und Ermittlung eines Tonzuschlages für die Beurteilung von Geräuschimmissionen. Deutsches Institut für Normung e.V. (2005-03)

[3] Mauch, M. und Ewert, S.: The Audio Degradation Toolbox and its Application to Robustness Evaluation. ISMIR (2013)

[4] HEAD ArtemiS Suite, URL: https://www.head-acoustics.com/de/nvh_artemis_suite.htm

[5] Cohen, J.: Statistical power analysis for the behavioral sciences (2nd ed) Hillsdale, New York, Erlbaum Associates, 1988

[6] Kingma, D. und Ba, J.: Adam: A Method for Stochastic Optimization. arXiv:1412.6980 (2014)

[7] Keras-Framework, URL: https://keras.io/

[8] Welch, B.: The Generalization of 'Student's' Problem When Several Different Population Variances Are Involved. Biometrika 34 (1947), (1/2):28–35,