

KI-Training nur mit realen Daten? Wie Datensynthese Variabilität sicherstellt

Dennis WITTCHEN

*Hochschule für Technik und Wirtschaft Dresden
Friedrich-List-Platz 1, D-01069 Dresden*

Kurzfassung: Datensätze sind die Grundlage für das Training von Machine Learning (ML) Modellen. Sie nehmen entscheidend Einfluss auf die Güte der Modelle und müssen daher in hinreichender Menge vorhanden sein und dabei eine möglichst große Variabilität der Zielgrößen abbilden. Dies kann jedoch nicht immer mit realen Daten gewährleistet werden. Anhand einer Anwendung zur Analyse menschlicher Bewegungen im Bereich der Mensch-Maschine-Interaktion (MMI) wird ein Lösungsansatz zur Kompensation der Defizite bestehender Datensätze mit Hilfe synthetisch erzeugter Bewegungsdaten vorgestellt. Das ML-Modell zur Bewegungssynthese ist dabei in der Lage, bekannte Bewegungen hinsichtlich definierter Parameter (z.B. Körpermaße oder Alter) zu manipulieren bzw. zu adaptieren, oder gänzlich neue Bewegungen zu erzeugen.

Schlüsselwörter: Bewegungssynthese, synthetische Daten, Machine Learning, Mensch-Maschine-Interaktion

1. Einleitung und Motivation

Große Datensätze bilden die Grundlage für das Training lernender Algorithmen aus dem Bereich des Machine Learnings (ML) und haben entscheidenden Einfluss auf die Güte der trainierten ML-Modelle. In diesem Zusammenhang ergeben sich einige grundlegende Fragestellungen, u.a.: Sind ausreichend große Trainingsdatensätze vorhanden? Decken diese eine möglichst große Variabilität der Einflussparameter ab? Im Fall menschlicher Bewegungen ist die Erhebung von Datensätzen mit erheblichem technischen und finanziellen Aufwand verbunden. Je nach Aufnahmesystem ist die Größe des erfassbaren Raums eingeschränkt. Auch die Genauigkeit variiert bei kamerabasierten Systemen von weniger als einem Millimeter bis hin zu mehreren Zentimetern (Menolotto et al., 2020). Zusätzlich geht die Aufbereitung der Daten (z.B. Labeln) mit einem großen Arbeitsaufwand einher. Werden die Daten dabei manuell verarbeitet, unterliegt die Genauigkeit subjektiven Einflüssen, was wiederum Einfluss auf die Qualität der ML-Modelle hat (Varol et al., 2017). Die Erweiterung des Datenbestandes mit synthetisch erzeugten Daten kann dazu beitragen, hinreichend große und variable Datensätze sicherzustellen. Gegenüber realen Datensätzen können die Labels automatisiert generiert werden, da die Parameter der Erzeugung bekannt bzw. gerade die Basis sind. Ungenauigkeiten durch menschliche oder messtechnische Einflüsse (vgl. kamerabasiertes Motion-Capturing) werden ebenso umgangen. Zu diesem Zweck befasst sich das Promotionsprojekt „MotionSynth“ mit der ML-basierten Synthese von individuellen menschlichen Bewegungen (Wittchen & Wacker, 2020). Es wird das Ziel verfolgt, entsprechende ML-Modelle und Werkzeuge zur Synthese - in Form einer Verarbeitungspipeline für Bewegungsdaten - für ein möglichst breites An-

wendungsspektrum zugänglich zu machen und einfach anwendbar zu gestalten. Anwender,-innen sollen in die Lage versetzt werden, valide Datensätze eigenständig erstellen zu können, um diese für das Training ihrer ML-Modelle zu verwenden.

Die Entwicklung der ML-Modelle und Werkzeuge zur Bewegungssynthese erfolgt zunächst entlang der Anforderungen spezieller Domänen, u.a. aus dem Bereich der Mensch-Maschine-Interaktion (MMI). Beim Anlernen von Industrierobotern basierend auf menschlichen Bewegungen haben anthropometrische Parameter mitunter großen Einfluss auf die Ausführung einer Bewegung. Für eine optimale Interpretation der Bewegungstrajektorien und Übertragung auf eine Roboterbewegung müssen ML-basierte Bewegungsanalysen diese Parameter berücksichtigen. Die parametrisierte Bewegungssynthese stellt in diesem Zusammenhang sicher, dass die Trainingsdaten die entsprechenden Parameter in hinreichender Menge und Variabilität enthalten. Der Entwicklungsprozess von Bewegungsanalysen kann durch den reduzierten Aufwand während der Datenerhebung maßgeblich beschleunigt werden. Im weiteren Verlauf des Projektes sollen die Ergebnisse aus den einzelnen Anwendungskontexten genutzt werden, um eine modulare, domänenübergreifende Pipeline zur Verarbeitung und Synthese von Bewegungsdaten mit generischen Schnittstellen zu etablieren.

2. Zielstellung

Den iterativen Entwicklungsprozess einer Pipeline für die Verarbeitung und Synthese menschlicher Bewegungen sowie deren modularen Aufbau beschreiben Wittchen & Wacker (2020) für den Anwendungsbereich der Rehabilitationsmedizin.



Abbildung 1: Anlernen eines Industrieroboters anhand menschlicher Bewegungen (modifizierte Abbildungen, Bildrechte der Originale: Wandelbots GmbH)

In dieser Arbeit soll mit dem Anlernen von Industrierobotern anhand menschlicher Bewegungen eine Anwendung aus dem Bereich der MMI beschrieben werden. Das Anlernen eines Roboterarms geschieht hier bspw. mittels eines Eingabegerätes (vgl. Abb. 1 links), mit dem eine 3D-Trajektorie im Raum aufgezeichnet wird (vgl. Abb. 1 Mitte). Diese Trajektorie enthält Informationen der zeitlichen Veränderungen in Position und Orientierung des Gerätes. Diese werden analysiert, optimiert (z.B. geglättet) und in Befehle für die Robotersteuerung übersetzt. Die aufbereitete Trajektorie entspricht damit der Trajektorie des Endeffektors eines Roboterarms. Das Analyseverfahren hat jedoch keine Kenntnisse über die Intentionen und Rahmenbedingungen, unter denen die Bewegung entstanden ist, u.a. die Körperhaltung (vgl. Abb. 1 rechts). Die Unterscheidung einer gezielt unruhigen Bewegung von einer, die durch eine überstreckte Haltung resultiert, ist somit nicht möglich. Durch Hinzufügen von Bewegungsdaten des Menschen in den Trainingsprozess der ML-Bewegungsanalyse, können diese Zusammenhänge jedoch erkannt und erlernt werden. Da anhand realer Aufnah-

men nicht immer sichergestellt werden kann, dass ausreichend viele und variable Daten für die jeweiligen Bewegungen vorhanden sind, sollen synthetische Bewegungsdaten diese Defizite kompensieren. Die synthetisch erzeugten Daten enthalten neben der Bewegung des Menschen und der Trajektorie des Eingabegerätes auch Informationen (Labels) über die Bewegungsparameter (z.B. Haltungsbeurteilungen) und anthropometrische Faktoren. Zudem können diese Bewegungen durch Rauschen und weitere Störparameter noch weiter angereichert werden. Diese Daten dienen dem ML-Bewegungsanalyseverfahren als erweiterter Trainingsdatensatz.

3. Stand der Forschung

3.1 Bewegungsdatensätze

In vielen Forschungsbereichen - u.a. der Robotik - wurden zahlreiche (reale) Bewegungsdatensätze veröffentlicht. Einige davon etablierten sich als Referenzdatensätze zur Evaluierung von Algorithmen zur Bewegungssegmentierung oder Aktivitätserkennung, darunter Human3.6m (Ionescu et al., 2014). Datensätze im Kontext der Ergonomie (Maurice et al., 2019) oder Arbeitsprozessen (Niemann et al., 2020) sind hingegen seltener zu finden. Solche Bewegungen werden mit unterschiedlichen Technologien aufgezeichnet, z.B. mit kamerabasiertem Motion-Capture oder mittels Inertial-Measurement-Units (vgl. Menolotto et al., 2020). Je nach Aufnahmesystem und Weiterverarbeitung unterscheiden sich dabei die Art der Daten (z.B. Videos, 3D-Skelettdaten, oder 3D-Oberflächen) und die Datenformate der Datensätze zum Teil deutlich. Für Skelettdaten kommen zudem verschiedene Topologien zum Einsatz, die sich in Anzahl und Hierarchie der einzelnen Gelenkpunkte unterscheiden. Dies ist jedoch ein Hemmnis bei der Integration mehrerer Datensätze zu einem Trainingsdatensatz. Terlemez et al. (2014) entwickelten daher mit der Master-Motor-Map ein System, welches die Vereinheitlichung und Konvertierung zwischen den Topologien ermöglicht.

3.2 Digitale Menschmodelle

Menschliche Bewegungen werden auf Skelettdaten repräsentiert. Die Knoten solcher Skelette (Gelenke) sind über verschiedene Topologien untereinander verknüpft und stellen eine vereinfachte, digitale Repräsentation eines Menschen dar - ein sogenanntes digitales Menschmodell (DMM). Diese werden hinsichtlich ihres Anwendungsbereichs unterschieden. Anthropometrische DMM werden im Bereich der Arbeitswissenschaft eingesetzt, um bspw. Arbeitsplätze oder Produkte bereits während der Entwurfs- und Konstruktionsphase für bestimmte Populationen bedarfsgerecht zu dimensionieren (vgl. Bubb, 2019). Dabei spielt neben den Skelettdaten auch die realistische Darstellung der physikalischen Ausmaße eines digitalen Avatars eine Rolle (vgl. Bubb, 2019). Neben einigen allgemeinen Funktionen besitzen DMM jeweils spezialisierte Funktionen zur Manipulation (z.B. Inverse-Kinematics), Analyse (z.B. Haltungsanalysen) und Ausgabe (z.B. Visualisierung) (vgl. Bullinger-Hoffmann & Mühlstedt, 2016: S. 81-91). Im Unterschied zu anthropometrischen DMM liegt der Fokus bei biomechanischen DMM auf der Abbildung mechanischer Eigenschaften. Diese DMM werden eingesetzt, um mechanische Auswirkungen zu untersuchen, die durch externe Krafterwirkungen verursacht werden (vgl. Bubb, 2019).

3.3 Bewegungssynthese

Gu et al. (2016) unterscheiden vier Ansätze zur Bewegungssynthese: (1) manuelle Methoden (Key-Frame basierte Methoden), (2) Methoden basierend auf physikalischen Modellen, (3) videobasierte Methoden und (4) datengetriebene Ansätze unter Nutzung von Motion-Capture-Daten. Bei der manuellen Methode erstellen Animator,-innen einzelne Posen einer Bewegung (Key-Frames). Die kontinuierliche Bewegung zwischen den Key-Frames wird durch einfache Algorithmen interpoliert. Jedoch ist diese Methode sehr zeitaufwendig und bedarf viel Erfahrung, um natürlich wirkende Bewegungen zu erzeugen. Die Synthese basierend auf physikalischen Gesetzen nutzt u.a. die Massenverteilung einzelner Körpersegmente und biomechanische Prinzipien, wie Drehmomente der Gelenke (Jiang et al., 2019). Eine weitere Spezialisierung ist die Bewegungssynthese anhand von Muskelkontraktionen (Lee et al., 2019). Physikalische Modellierungen erzeugen zwar biomechanisch korrekte Bewegungen, es ist jedoch herausfordernd, den Bewegungsstil einer bestimmten Person abzubilden. Videobasierte Verfahren nutzen Algorithmen der Computer-Vision, um zunächst die menschliche Pose in 2D-Daten und deren zeitliche Änderung zu extrahieren. Im Anschluss werden die 3D-Informationen aus mehreren Kameraaufnahmen aus verschiedenen Positionen rekonstruiert. Werden einzelne Körperteile nicht von allen Kameras erfasst, können die entsprechenden Gelenkpunkte ggf. nicht rekonstruiert werden. Die gewonnenen Bewegungen können wiederum in kleinere Segmente unterteilt werden und anschließend zu neuen Bewegungen zusammengesetzt werden. Motion-Capture basierte Verfahren nutzen bereits vorhandene 3D-Bewegungsdaten. Diese können wiederum auf verschiedene Weise manipuliert werden, z.B. durch das Einschränken der Freiheitsgrade einzelner Gelenke, oder Zerlegung in kleinere Segmente. Bewegungssegmente können anschließend zu neuen, realistisch anmutenden Bewegungen zusammengesetzt werden. Auch die Übertragung eines bestimmten Bewegungsstils ist möglich (Aberman et al., 2020). Einige Verfahren sind auch in der Lage, die Bewegungssynthese in Echtzeit durchzuführen, bspw. Laufanimationen auf diversen Untergründen (Holden et al., 2017) oder die Interaktion mit Objekten (Starke et al., 2019).

4. Methodik

4.1 Analyse bestehender Bewegungsdatensätze

Mit Hilfe einer systematischen Literaturrecherche werden Bewegungsdatensätze aus unterschiedlichen Domänen zusammengetragen. Das Ziel der anschließenden Analysen besteht darin, Kenntnisse über den Umfang und die Güte der enthaltenen Bewegungsdaten zu erlangen. Dazu werden die Datensätze zunächst einer Metadaten-Analyse unterzogen, um Kenntnisse über den Kontext, die Struktur und den Umfang des jeweiligen Datensatzes zu erhalten (z.B. Domäne, Anzahl der Bewegungen und Bewegungsklassen, Art des Aufnahmesystems, sowie Anzahl und Verteilung anthropometrischer Parameter). Nachgelagert findet die Analyse der eigentlichen Bewegungsdaten statt. Die zu analysierenden Parameter können dabei je nach Anwendungskontext unterschiedlich ausfallen, z.B. Minimum und Maximum der Armstreckung, oder der Beschleunigungsverlauf der Hand während einer Greifbewegung. Anhand dieser beiden Analysen können unterrepräsentierte Bewegungen oder Parameter identifiziert werden. Des Weiteren können die Analysen genutzt werden, um für eine Anwendung geeignete Kombinationen unterschiedlicher Datensätze abzuleiten.

4.2 Bewegungssynthese

Zunächst wird ein ML-Modell entwickelt, das Zeitreihen von 3D-Gelenkdaten analysieren und synthetisieren kann. Hierzu bieten sich Architekturen wie Variational Auto Encoder (VAE) oder Recurrent Neural Networks (RNN) an. Das ML-Modell wird zunächst anhand der in Abschnitt 4.1 beschriebenen Datensätze vortrainiert, sodass es in der Lage ist, vielseitige und biomechanisch plausible Bewegungen zu erzeugen. Um den Einfluss von Körperhaltungen zu erlernen, müssen Referenzdatensätze in realen Szenarien aufgenommen werden und entsprechend (mittels Labels) annotiert werden. Mit diesen Daten kann das ML-Modell weiter trainiert werden. Zur Laufzeit bekommt dieses Modell als Eingabe entweder eine manuell animierte Zielbewegung oder einzelne Key-Frames, die zu einer kontinuierlichen Bewegung vervollständigt werden. Desweiteren müssen die Zielparameter (z.B. anthropometrische Werte) an das Modell übergeben werden, die als Grenzwerte für die Synthese verwendet werden sollen. Anhand dieser Eingaben können anschließend Variationen der Zielbewegung generiert werden. Diese werden zur Visualisierung und Validierung auf ein DMM übertragen.

5. Zu erwartende Ergebnisse

Im Rahmen einer systematischen Literaturrecherche für den Zeitraum von 2000 bis 2020 konnten ca. 120 publizierte Datensätze identifiziert werden. Die Analysen dieser Datensätze werden Aufschluss darüber geben, ob und in welchem Ausmaß Defizite bzgl. der Anzahl und Verteilung von Parametern (z.B. Anthropometrie) vorliegen. Entsprechende Werkzeuge zur Exploration der Datensätze sowie zur Darstellung der Analyseergebnisse werden veröffentlicht. Um die bestehenden Bewegungsdaten für spätere Anwendungen einfacher nutzbar zu machen, wird eine vereinheitlichte Repräsentation der Bewegungsdaten hinsichtlich der Skelett-Topologie unter Zuhilfenahme eines DMM realisiert. Mit Hilfe der Bewegungssynthese werden dann Daten erzeugt, die nicht durch bestehende Datensätze abgedeckt sind. Defizite bzgl. der Datensatzgröße und Verteilung von Einflussparametern werden somit kompensiert. Darüber hinaus werden bisher nicht erfasste Bewegungen synthetisiert und frei zugänglich gemacht. Neben den synthetisierten Daten werden auch die zur Erzeugung eingesetzten Werkzeuge (ML-Modelle und Pipeline) unter OpenSource-Lizenzen veröffentlicht.

6. Diskussion

In dieser Arbeit wurden sowohl die Bedeutung als auch die Herausforderungen dargestellt, die mit Datensätzen im Kontext von ML einhergehen. Hierbei wurde insbesondere auf Anwendungen im Bereich der MMI eingegangen, in denen ML für Bewegungsanalysen zum Einsatz kommt. Als eine Lösungsstrategie für zu kleine oder nicht hinreichend variable Datensätze wurde die Erzeugung und Nutzung synthetischer Daten diskutiert. Aktuelle Forschungsergebnisse aus dem Bereich Computer-Vision zeigen, dass das Training von ML-Modellen anhand synthetisch angereicherte Daten eine Qualitätssteigerung der Modelle mit sich bringen kann (vgl. Varol et al., 2017). Die Übertragbarkeit dieser Erkenntnisse für Anwendungen im Bereich Computer-Animation soll im Rahmen des Projektes erforscht werden. Je Anwendungsfall ist zu evaluieren, wie hoch der Anteil der synthetischen Daten ausfallen darf bzw. muss, um den

Anforderungen entsprechend valide Trainingsdatensätze für die ML-Modelle bereitzustellen. Zur Datensynthese wurde eine Verarbeitungspipeline für menschliche Bewegungen vorgestellt und diskutiert. Mit erfolgreicher Umsetzung dieser Pipeline ist zu erwarten, dass Unternehmen in die Lage versetzt werden, ihre Anwendungen mit Hilfe von synthetischen Daten kosten- und ressourcenschonend weiterentwickeln zu können. Eine Adaption für weitere Anwendungsbereiche wird ebenso angestrebt.

7. Literatur

- Aberman, K., Weng, Y., Lischinski, D., Cohen-Or, D., & Chen, B. (2020). Unpaired motion style transfer from video to animation. *ACM Transactions on Graphics*, 39(4). <https://doi.org/10.1145/3386569.3392469>
- Bubb, H. (2019). Why do we need digital human models? In *DHM and Posturography* (S. 7–32). Elsevier. <https://doi.org/10.1016/B978-0-12-816713-7.00002-7>
- Bullinger-Hoffmann, A. C., & Mühlstedt, J. (2016). *Homo Sapiens Digitalis—Virtuelle Ergonomie und digitale Menschmodelle*. Springer. <https://doi.org/10.1007/978-3-662-50459-8>
- Gu, P., Zheng, Y., Wang, D., & Wu, Q. (2016). A Review of Human Motion Synthesis and Motion Retrieval Technology. *2016 11th International Conference on Computer Science & Education (ICCSE)*, 976–980.
- Holden, D., Komura, T., & Saito, J. (2017). Phase-functioned neural networks for character control. *ACM Transactions on Graphics*. <https://doi.org/10.1145/3072959.3073663>
- Ionescu, C., Papava, D., Olaru, V., & Sminchisescu, C. (2014). Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7), 1325–1339. <https://doi.org/10.1109/TPAMI.2013.248>
- Jiang, Y., Van Wouwe, T., De Groot, F., & Liu, C. K. (2019). Synthesis of biologically realistic human motion using joint torque actuation. *ACM Transactions on Graphics*, 38(4), 1–12. <https://doi.org/10.1145/3306346.3322966>
- Lee, S., Park, M., Lee, K., & Lee, J. (2019). Scalable muscle-actuated human simulation and control. *ACM Transactions on Graphics*, 38(4), 73:1–73:13.
- Maurice, P., Malaisé, A., Amiot, C., Paris, N., Richard, G.-J., Rochel, O., & Ivaldi, S. (2019). Human movement and ergonomics: An industry-oriented dataset for collaborative robotics. *The International Journal of Robotics Research*, 38(14), 1529–1537. <https://doi.org/10.1177/0278364919882089>
- Menolotto, M., Komaris, D.-S., Tedesco, S., O’Flynn, B., & Walsh, M. (2020). Motion Capture Technology in Industrial Applications: A Systematic Review. *Sensors*, 20(19), 5687. <https://doi.org/10.3390/s20195687>
- Niemann, F., Reining, C., Moya Rueda, F., Nair, N. R., Steffens, J. A., Fink, G. A., & ten Hoppel, M. (2020). LARa: Creating a Dataset for Human Activity Recognition in Logistics Using Semantic Attributes. *Sensors*, 20(15), 4083. <https://doi.org/10.3390/s20154083>
- Starke, S., Zhang, H., Komura, T., & Saito, J. (2019). Neural State Machine for Character-Scene Interactions. *ACM Transactions on Graphics (TOG)*, 38(6), 1–14.
- Terlemez, Ö., Ulbrich, S., Mandery, C., Do, M., Vahrenkamp, N., & Asfour, T. (2014). Master Motor Map (MMM) - Framework and toolkit for capturing, representing, and reproducing human motion on humanoid robots. *2014 IEEE-RAS International Conference on Humanoid Robots*, 894–901. <https://doi.org/10.1109/HUMANOIDS.2014.7041470>
- Varol, G., Romero, J., Martin, X., Mahmood, N., Black, M. J., Laptev, I., & Schmid, C. (2017). Learning from synthetic humans. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*. <https://doi.org/10.1109/CVPR.2017.492>
- Wittchen, D., & Wacker, M. (2020, März). Eine modulare Software-Pipeline zur Synthese funktioneller menschlicher Bewegungen. *66. Frühjahrskongress der Gesellschaft für Arbeitswissenschaft (GfA)*. Digitale Arbeit, digitaler Wandel, digitaler Mensch?, Berlin.

Danksagung: Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes (FKZ 100380879).



Gesellschaft für
Arbeitswissenschaft e.V.

Arbeit HUMAINE gestalten

67. Kongress der
Gesellschaft für Arbeitswissenschaft

Lehrstuhl Wirtschaftspsychologie (WiPs)
Ruhr-Universität Bochum

Institut für Arbeitswissenschaft (IAW)
Ruhr-Universität Bochum

3. - 5. März 2021

GfA-Press

Bericht zum 67. Arbeitswissenschaftlichen Kongress vom 3. - 5. März 2021

**Lehrstuhl Wirtschaftspsychologie, Ruhr-Universität Bochum
Institut für Arbeitswissenschaft, Ruhr-Universität Bochum**

Herausgegeben von der Gesellschaft für Arbeitswissenschaft e.V.
Dortmund: GfA-Press, 2021
ISBN 978-3-936804-29-4

NE: Gesellschaft für Arbeitswissenschaft: Jahresdokumentation

Als Manuskript zusammengestellt. Diese Jahresdokumentation ist nur in der Geschäftsstelle erhältlich.

Alle Rechte vorbehalten.

© **GfA-Press, Dortmund**

Schriftleitung: Matthias Jäger

im Auftrag der Gesellschaft für Arbeitswissenschaft e.V.

Ohne ausdrückliche Genehmigung der Gesellschaft für Arbeitswissenschaft e.V. ist es nicht gestattet:

- den Kongressband oder Teile daraus in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) zu vervielfältigen,
- den Kongressband oder Teile daraus in Print- und/oder Nonprint-Medien (Webseiten, Blog, Social Media) zu verbreiten.

Die Verantwortung für die Inhalte der Beiträge tragen alleine die jeweiligen Verfasser; die GfA haftet nicht für die weitere Verwendung der darin enthaltenen Angaben.

Screen design und Umsetzung

© 2021 fröse multimedia, Frank Fröse

office@internetkundenservice.de · www.internetkundenservice.de