

Bestimmung persönlicher Eigenschaften

modifizierter Auszug von
Material zum Fernstudienkurs „Informatik und Gesellschaft“
im Master-Fernstudiengang Informatik (Aufbaustudium)
[https://www.hochschule-trier.de/informatik/fernstudium/module/i-
z/informatik-und-gesellschaft-iug/](https://www.hochschule-trier.de/informatik/fernstudium/module/i-z/informatik-und-gesellschaft-iug/)

Download-Link: www.blaesius.net/IUG4b-bpe-20.pdf

Karl Hans Bläsius

blaesius@hochschule-trier.de

26.10.2020

Bestimmung persönlicher Eigenschaften

Menschen hinterlassen im Internet eine riesige Menge an Daten in unterschiedlichen Formen. Dazu gehören auch verfasste Texte, Fotos, Ton- und Videoaufnahmen. Diese Daten werden automatisch analysiert, um hieraus weitere Informationen abzuleiten. Hierbei können auch spezielle Verfahren zum Einsatz kommen, mit denen persönliche Eigenschaften der betroffenen Personen bestimmt werden.

1. Bestimmung persönlicher Eigenschaften aus Fotos

Bezüglich der Bestimmung persönlicher Eigenschaften aus Texten und Bildern ist besonders Michel Kosinski, Psychologe und Professor in Stanford bekannt geworden. In [WK17] wird ein Verfahren beschrieben, wie aus Fotos von Personen persönliche Eigenschaften wie die sexuelle Orientierung bestimmt werden können. Bei Männern wird eine Trefferquote von 81%, bei Frauen von 71% angegeben. Auch andere Eigenschaften will Kosinski allein aus Fotos bestimmen können, wie z.B. Intelligenz, politische Orientierung, kriminelle Ambitionen.¹

Chinesische KI-Forscher haben in [WZ16] Ergebnisse eines Projektes veröffentlicht, bei dem es darum ging, alleine auf der Basis von Passfotos Kriminelle von Nichtkriminellen zu unterscheiden. Die Trefferrate soll bei 89,5% liegen. Kriminelle stammen öfter aus ärmeren Verhältnissen, haben eher schlecht gepflegte Zähne oder Narben. Solche Merkmale können zu einer entsprechenden Zuordnung führen, aber natürlich auch viele fehlerhafte Zuordnungen erzeugen. Motive für solche Forschungsvorhaben und verwendete Methoden werden kritisch bewertet².

Bereits im 19. Jahrhundert hat der italienische Arzt und Gerichtsmediziner Cesare Lombroso Experimente durchgeführt, um Kriminelle anhand äußerer Körpermerkmale zu erkennen. Den Nazis dienten diese Theorien als Grundlage für rassenbiologische Theorien.³

¹ Süddeutsche Zeitung vom 3.3.2018, Seite 51

² [WZ17] und Süddeutsche Zeitung vom 7.4.2018, Seite 37

³ https://de.wikipedia.org/wiki/Cesare_Lombroso

Mit geeigneten Trainingsdaten könnten solche Verfahren auch angewendet werden, um bestimmte Krankheiten oder Abhängigkeiten (Drogen, Alkohol) mit einer gewissen Wahrscheinlichkeit zu erkennen.

In [RD17] stellen Reece und Danforth Methoden vor, um auf der Basis von Instagram-Fotos Merkmale für Depression abzuleiten. Neben den Fotos selbst werden auch Metadaten verwendet und das Kommunikationsverhalten der Nutzer wird untersucht. Solche Krankheiten können auch auf der Basis von Tonaufnahmen bestimmt werden, also z.B. aus den Daten, die Systeme wie Alexa liefern. Inzwischen wird auch untersucht, ob anhand von Veränderungen der Sprache die Wirksamkeit von Medikamenten festgestellt werden kann.⁴

2. Bestimmung persönlicher Eigenschaften aus Texten

Es gibt verschiedene Projekte zur Bestimmung persönlicher Eigenschaften aus Texten. Auch in diesem Punkt haben Arbeiten von Michal Kosinski für Furore gesorgt (z.B. [KSG13]). Während seiner Tätigkeit an der Universität Cambridge hat Kosinski ein Verfahren entwickelt, das aus Internet-Spuren, wie z.B. Facebook-Likes, persönliche Eigenschaften von Menschen bestimmt, u.a. bezüglich politischer Einstellungen. Dieses Verfahren hat Alexander Nix, ehemals Chef von Cambridge Analytica verwendet, um Wählerprofile zu erstellen und für Donald Trump die Wahl zu beeinflussen. Anfang 2018 wurde bekannt, dass Cambridge Analytica hierbei auch in großem Umfang Daten von Facebook verwendet hatte.

Kosinski kann mit seinen Algorithmen aus den Spuren im Internet mit 93% Sicherheit bestimmen, ob diese von einem Mann oder einer Frau stammen, zu 82%, ob Christ oder Muslim, zu 85%, ob Demokrat oder Republikaner, zu 65%, ob Single oder in Partnerschaft. Viele weitere persönliche Eigenschaften versucht er aus Texten zu bestimmen, die immer nur mit einer gewissen Wahrscheinlichkeit gelten. Kosinski sagt, mit hinreichend vielen „Likes“, die im Internet leicht gesammelt werden können, können die Algorithmen Menschen besser einschätzen, als Kollegen, Freunde, Partner.⁵

Auch Arbeiten von Dirk Hovy, Margaret Mitchell und anderen befassen sich mit der Bestimmung persönlicher Eigenschaften aus Texten. In [BMH17] wird als Zielsetzung beschrieben, psychische Erkrankungen früh zu erkennen und

⁴ [Vol18], Seite 109 - 110

⁵ Süddeutsche Zeitung vom 3.3.2018, Seite 51

zu behandeln, um so Depressionen und Selbstmorde zu verhindern. Die Autoren geben an, dass sie Neigungen zu Depressionen an den im Internet hinterlassenen Textspuren (z.B. likes) erkennen können.

3. Bestimmung persönlicher Eigenschaften aus Tonaufnahmen

In verschiedenen Projekten werden Verfahren entwickelt, um aus dem akustischen Sprachsignal von Tonaufnahmen automatisch Gefühle, Gesundheitsmerkmale und Charaktereigenschaften zu bestimmen (siehe z.B. [Sch07], [SB13]). Hierbei geht es nicht darum, den Inhalt des Gesprochenen zu analysieren oder zu verstehen. Stattdessen werden aus Kriterien wie z.B. Höhen, Tiefen und Tempo des Gesprochenen Merkmale bestimmt, aus denen weitreichende Schlüsse über den emotionalen Zustand sowie Gesundheits- und Charaktereigenschaften gezogen werden.

Inzwischen sind große Datenmengen verfügbar auf deren Basis „deep-learning“-Algorithmen sehr viele Merkmale bestimmen können, um daraus automatisch Gefühlszustände zu berechnen. Kausale Zusammenhänge sind in diesen Merkmalen häufig nicht erkennbar.

Für die Bestimmung solcher Merkmale kann es vielfältige Anwendungen geben. Zum Beispiel lassen sich aus einer Emotionsanalyse der Tonaufnahmen eines Films Rückschlüsse auf die Erfolgsaussichten ziehen. Psychische Erkrankungen lassen sich einfacher und vielleicht sogar zuverlässiger diagnostizieren, als dies durch Psychotherapeuten auf der Basis von Gesprächen oder Fragebogen möglich ist. Die Erkennung eines Gefühlszustands kann von digitalen Medien auch für spezielle Werbemaßnahmen oder zur Manipulation verwendet werden. Amazon hat ein Patent angemeldet für die Erkennung von Emotionen und Krankheiten aus der Stimme. Auch in Zusammenhang mit Stellenbewerbungen werden bereits Stimmanalysesysteme eingesetzt, um Persönlichkeitsmerkmale eines Bewerbers zu erkennen.⁶

Auch in Zusammenhang mit „Predictive Policing“ wird versucht, Eigenschaften aus der Stimme eines Menschen zu bestimmen, um diesen so als potentiellen Gefährder zu charakterisieren. Zum Beispiel wirbt ein amerikanisches Softwareunternehmen damit, dass auf Basis eines

⁶ Die Zeit Nr. 7, 2019, 7.2.2019, Seite 27 bis 28

zehnminütigen Telefoninterviews festgestellt werden kann, ob ein Bewerber ein Risiko für das Unternehmen darstellt.⁷

4. Kritik

Die Bestimmung persönlicher Eigenschaften aus Fotos oder Texten ist eine Klassifikationsaufgabe, wobei es in der Regel viele Kriterien geben wird, die zu einer Entscheidung beitragen. Diese Kriterien sind rein statistischer Natur und kaum inhaltlich begründet, stellen also keine kausalen Zusammenhänge dar.

Die Bestimmung persönlicher Eigenschaften ist unsicher und vage. Ähnlich wie bei den Begriffen „schwer“ und „viel“ ist eine „Neigung zur Depression“ nicht nur wahr oder falsch, sondern eine solche Eigenschaft gilt in bestimmtem Maße. Ähnliches gilt für andere persönliche Eigenschaften, wie politische Einstellung, Neigung zur Kriminalität, usw.

Die Algorithmen von Kosinski zur Bestimmung persönlicher Eigenschaften aus Fotos werden kritisiert, weil kausale Zusammenhang suggeriert werden, die es nicht gibt.⁸ Dies spielt aber keine Rolle, solange die „neuen“ zu untersuchenden Fälle zur Lerngrundlage passen. Statistische Verfahren funktionieren, auch ohne dass kausale Zusammenhänge vorliegen oder bestimmt werden. Diese Verfahren können trotzdem zu guten Ergebnissen führen. Es gibt viele Anwendungen, bei denen sehr gute Ergebnisse auf der Basis einer statistischen Auswertung von Symptomen erzielt werden, ohne dass kausale Zusammenhänge vorliegen. Ähnliches gilt auch für die Arbeiten von Wu, Zhang zur Bestimmung von Kriminellen aus Fotos. Sobald sich die Bedingungen ändern, z.B. bei einer Anwendung auf andere ethnische Gruppen oder eine andere Altersstruktur, werden die Ergebnisse vermutlich deutlich schlechter. Die Arbeiten von Wu, Zhang wurden auch kritisiert und als wertlos dargestellt, da sie auf einer falschen Datengrundlage basieren. Kriminelle tragen T-Shirt, andere Hemd mit Jackett. Auch wenn dies stimmt, können die Verfahren mit anderen Datengrundlagen wiederholt werden und zu Klassifikationsergebnissen mit eventuell schlechteren Erkennungsraten führen.

Ähnlich argumentiert auch Joran Lanier, der folgendes Beispiel verwendet: Wenn viele andere Menschen, die gerne das Gleiche essen wie man selbst, vom Bild einer Person mit rosa Umrahmung eher abgestoßen werden als bei blauer

⁷ Süddeutsche Zeitung vom 2.3.2019, Seite 33

⁸ Süddeutsche Zeitung vom 7.4.2018, Seite 37

Umrahmung, dann wird dies bei einem selbst wahrscheinlich genauso sein und es ist völlig egal, warum das so ist. Statistiken sind auch dann zuverlässig, wenn keine inhaltlichen Zusammenhänge vorliegen.⁹

Da eine Klassifikation nach rein statistischen Kriterien, ohne inhaltliche Zusammenhänge, sinnvoll sein kann, werden solche Verfahren trotz der Kritik vermutlich auch in Zukunft angewendet. Die Anwendung von Verfahren für die Bestimmung persönlicher Eigenschaften ist aber aus folgenden Gründen sehr problematisch:

- **Keine Unterscheidung Unsicherheit - Vagheit:** Unsicherheit und Vagheit werden häufig nicht unterschieden. Stattdessen wird einfach von Erkennungsraten für die charakteristischen Eigenschaften gesprochen. Der Aspekt der Vagheit wird hierbei einfach weggelassen, d.h. eigentlich vage Werte, also Werte zwischen 0 und 1 im Sinne der Fuzzy-Logic werden „binarisiert“, also zu 0 ab- oder zu 1 aufgerundet. Vermutlich gilt Ähnliches in den meisten Fällen auch für die Datengrundlage für das Lernen: Personen werden einfach nur in 2 Kategorien (gilt - gilt nicht) eingeteilt, ohne zu bestimmen in welchem Maße eine Eigenschaft gilt. Eine genaue Bestimmung vager Werte wird in der Regel auch nicht möglich sein.
- **Schlüsse aus unsicherem Wissen sind unsicher:** Die Ergebnisse der Bestimmung persönlicher Eigenschaften aus Fotos oder Texten sind immer unsicher. Unsicheres Wissen kann mit sicheren oder unsicheren Regeln verknüpft werden. Das Ergebnis ist in jedem Fall wieder unsicher. In einer Kette von Schlussfolgerungen können so weitere unsichere Daten erzeugt werden. Alle so erzeugten Daten können falsch sein. Wenn unsichere Ausgangsdaten falsch sind, sind auch alle daraus gezogenen Schlüsse ungültig. Wenn bei den Schlussfolgerungen keine Konflikte (Widersprüche) auftreten, kann die Ungültigkeit in der Regel nicht festgestellt werden, egal welches Verfahren zur Behandlung von Unsicherheiten verwendet wird.
- **Keine Korrektur von falschen Werten:** Auch beim Datenaustausch zwischen Unternehmen kann es unsichere Daten geben, wie z.B. bei der automatischen Erkennung und Verarbeitung von Rechnungen oder der Klassifikation von Gerichtsurteilen. Hierbei gibt es aber hinreichend viele Möglichkeiten für Plausibilitätsprüfungen. Des Weiteren werden bei solchen Anwendungen unsichere bzw. falsche Werte manuell

⁹ [Lan18], Seite 12

überprüft und gegebenenfalls korrigiert. Bei der automatischen Bestimmung persönlicher Eigenschaften aus Texten oder Bildern wird eine Korrektur von falschen Werten kaum möglich sein. Die betroffenen Personen wissen in der Regel auch nicht, welche Informationen auf eine solche Weise bestimmt werden und können somit auch keine Korrektur veranlassen. Das heißt, diese Daten werden verwendet, ohne Unsicherheit und Vagheit zu berücksichtigen. Das Merkmal wird als gültig angenommen und angewendet. Die Analyse-Ergebnisse werden nicht auf Wahrheit überprüft, sondern werden selbst zur Wahrheit.

- **Kein Lernfortschritt:** Lernende Systeme sind auf Feedback angewiesen. Da falsche Werte in der Regel nicht überprüft und korrigiert werden, gibt es auch keinen Lernfortschritt. Die Erkennungsraten bleiben auf dem Stand, der aus der ursprünglichen Datengrundlage resultierte.
- **Massenanwendung:** Riesige Mengen von Texten, Fotos und Tonaufnahmen haben die großen Internetkonzerne z.B. auf der Basis von Alexa oder Siri gesammelt. Viele dieser Daten, z.B. Fotos sind im Internet auch frei verfügbar und könnten für entsprechende Analysen verwendet werden. Mit Hilfe von inzwischen sehr erfolgreicher Gesichtserkennung sind auch Fotos mit mehreren Personen (Gruppenaufnahmen) für diese Zwecke nutzbar. Das Missbrauchspotential durch Staaten und auch durch Unternehmen (Verkauf der Daten an andere Unternehmen, Diskriminierung bei Stellenbewerbungen, Bankgeschäften, usw.) ist sehr hoch.

Das Bestimmen von persönlichen Eigenschaften aus Texten, Bildern, Tonaufnahmen oder Videosequenzen ist daher äußerst problematisch und abzulehnen. Die Ergebnisse sind immer unsicher. d.h. sie gelten nur mit gewisser Wahrscheinlichkeit, also nur für einen Teil der Personen, denen sie zugeordnet sind. Außerdem sind diese Ergebnisse vage, d.h. selbst wenn eine entsprechende Eigenschaft einer Person zugeordnet werden kann, gilt diese Eigenschaft nur in bestimmtem Grad.

Für den einzelnen Menschen kann aber nicht bestimmt werden, ob die Eigenschaft gilt, oder ob sie nicht gilt. Selbst wenn die Gültigkeit einer Eigenschaft angenommen werden kann, bleibt offen, in welchem Maße dies zutrifft. Ein solcher Grad wird in der Regel nicht bestimmt, sondern der Person wird einfach das Merkmal als gültig zugeordnet.

5. Folgen

Die Verfahren zur Bestimmung persönlicher Eigenschaften könnten massenhaft auf den Daten in sozialen Netzwerken angewendet werden und die Nutzer in Kategorien einteilen. Dies kann zu Vorurteilen und Diskriminierung führen.

Auch bei anderen Anwendungen hat sich in der Vergangenheit gezeigt, wie problematisch das Auswerten von Texten bezüglich persönlicher Eigenschaften sein kann. So wurde in den USA ein Algorithmus angewendet, der auf der Basis von alten Gerichtsentscheidungen Vorschläge macht, welche Gefangenen vorzeitig aus der Haft entlassen werden sollen. Eine Analyse der Ergebnisse hat gezeigt, dass Gefangene schwarzer Hautfarbe hierbei benachteiligt wurden. Die Ursache hierfür lag an den alten Texten. Vorurteile aus der Vergangenheit, die in diesen Urteilen steckten, wurden durch die Lernverfahren verwendet, also mitgelernt.¹⁰

Automatisch bestimmte Eigenschaften wie z.B. sexuelle Orientierung, Neigung zur Kriminalität oder politische Präferenzen können zu erheblichen Nachteilen für die Betroffenen führen und in manchen Staaten sogar die Freiheit und das Leben bedrohen. Zum Beispiel könnten die Bürger, bei denen die Analyse ihres Fotos zu einem entsprechenden Ergebnis führt, diskriminiert oder bestraft werden. Ähnliches könnte bei der Einreise in einen solchen Staat passieren bzw. die Einreise könnte abgelehnt werden. In [WK17] weisen die Autoren auch auf solche Gefahren hin.

Angenommen jemand erhält Kenntnis von Eigenschaften, die von ihm aus Fotos bestimmt wurden und diese Eigenschaften gefallen ihm nicht und sind vielleicht auch falsch. Was kann getan werden? Aufgrund einer eventuell weiten Verteilung solcher Daten, wird es kaum möglich sein, diese Daten an allen relevanten Stellen zu ändern. Bleibt als einzige Möglichkeit eine Gesichtsänderung durch plastische Chirurgie? Als Ergänzung wäre auch das Ändern der Augenfarbe inzwischen möglich. Es ist aber absurd, wenn man wegen Algorithmen solche Maßnahmen ergreifen muss oder glaubt, man müsse solche Maßnahmen ergreifen, damit die Algorithmen möglichst gute Eigenschaften von einem liefern.

¹⁰ Süddeutsche Zeitung vom 25.3.2017, Seite 35

6. Juristische Aspekte

Mögliche juristische Fragestellungen sind:

- Sollte in einem Rechtsstaat geprüft werden, ob die Bestimmung persönlicher Eigenschaften aus Texten, Tonaufnahmen, Bildern oder Videosequenzen sowie deren Anwendung durch eine entsprechende Gesetzgebung verboten wird?
- Widersprechen solche Anwendungen bereits heute geltenden Gesetzen, wie einem Diskriminierungsverbot?
- Auch die europäische Menschenrechtskonvention enthält in Artikel 14 ein Diskriminierungsverbot. Verletzen Forschung und Anwendung der Bestimmung persönlicher Eigenschaften dieses Gesetz?

Artikel 9 der EU-DSGVO behandelt die Verarbeitung besonderer Kategorien personenbezogener Daten, Absatz 1 lautet:¹¹

„Die Verarbeitung personenbezogener Daten, aus denen die rassische und ethnische Herkunft, politische Meinungen, religiöse oder weltanschauliche Überzeugungen oder die Gewerkschaftszugehörigkeit hervorgehen, sowie die Verarbeitung von genetischen Daten, biometrischen Daten zur eindeutigen Identifizierung einer natürlichen Person, Gesundheitsdaten oder Daten zum Sexualleben oder der sexuellen Orientierung einer natürlichen Person ist untersagt.“

Absatz 2 dieses Artikels enthält dann aber einige Ausnahmen. Solche Daten dürfen verarbeitet werden, wenn die betroffene Person ausdrücklich eingewilligt oder die Daten offensichtlich öffentlich gemacht hat. Auch das BDSG legt in § 22 weitere Bedingungen für die Verarbeitung besonderer Kategorien personenbezogener Daten fest.¹²

Im Hinblick auf die in Abschnitt 4 beschriebene Argumentation stellt sich die Frage, ob diese gesetzlichen Vorgaben ausreichen. Fotos, die auf Internetseiten, z.B. als Mitarbeiter eines Unternehmens, Ansprechpartner einer Behörde oder Vorstandsmitglieder eines Vereins erscheinen, sind offensichtlich öffentlich gemacht. Auf solche Fotos könnten aber auch die beschriebenen Verfahren angewendet werden, was sicher nicht im Sinne der betroffenen Personen ist. Das Hochladen von Fotos in sozialen Netzen oder anderen Internetportalen erfolgt auf Basis der jeweiligen AGBs, wobei in der Regel einer uneingeschränkten Verarbeitung zugestimmt wird. Die Besitzrechte an den

¹¹ <https://dsgvo-gesetz.de/art-9-dsgvo/>

¹² <https://dsgvo-gesetz.de/bdsg/22-bdsg/>

Fotos werden abgegeben. Der normale Nutzer weiß aber nicht, was alles aus seinen Daten und Fotos abgeleitet werden kann, wenn er die Rechte daran abgibt.

Wer vor 10 Jahren ein Foto bei Facebook hinterlegt hatte, konnte nicht ahnen, dass heute daraus Eigenschaften wie sexuelle Orientierung, Neigung zu Depressionen usw. abgeleitet werden können. Es ist nicht absehbar, welche Daten irgendwann aus den heute vorhandenen Daten abgeleitet und wie diese verwendet werden.

Literatur

- [BMH17] Adrian Benton, Margaret Mitchell, Dirk Hovy: Multitask Learning for Mental Health Conditions with Limited Social Media Data, Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics, 2017
- [KSG13] Michal Kosinski, David Stillwell, Thore Graepel: Private traits and attributes are predictable from digital records of human behavior, PNAS, 2013
- [Lan18] Jaron Lanier: Zehn Gründe, warum du deine Social Media Accounts sofort löschen musst. Hoffman und Campe, Hamburg 2018
- [RD17] Andrew G. Reece, Christopher M. Danforth: Instagram photos reveal predictive markers of depression. <https://arxiv.org/pdf/1608.03282v2.pdf>, 2017
- [SB13] Björn Schuller, Anton Batliner: Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing, Wiley, 2013
- [Sch07] Björn Schuller: Mensch, Maschine, Emotion: Erkennung aus sprachlicher und manueller Interaktion, VDM-Verlag, 2007
- [Vol18] Holger Volland: Die kreative Macht der Maschinen. Beltz Verlag, 2018
- [WK17] Yilun Wang, Michal Kosinski: Deep neural networks are more accurate than humans at detecting sexual orientation from facial images. Journal of Personality and Social Psychology, 2017

- [WZ16] Xiaolin Wu, Xi Zhang: Automated Inference on Criminality Using Face Images, <https://arxiv.org/pdf/1611.04135v1.pdf>, 2016
- [WZ17] Xiaolin Wu, Xi Zhang: Responses to Critiques on Machine Learning of Criminality Perceptions, <https://arxiv.org/pdf/1611.04135v3.pdf>, 2017