

Inhalt

Leistungsnachweis Prüfungsleistung: Klausur

Spätestens der Spiegel Nr. 1 Bestseller von Mai Thi Nguyen-Kim: Die kleinste gemeinsame Wirklichkeit“ zeigt eines ganz deutlich: Wir brauchen mehr und mehr kritisches Methodenwissen

Statements von Dr. Mai Thi Nguyen-Kim:

„Außerdem erklärt jedes Kapitel das, was an wissenschaftlicher Methodik wichtig ist, um über das betreffende Thema sinnvoll zu diskutieren: Wie beurteilt man den Grad der Beweislage? Was ist problematisch daran, unterschiedliche Aspekte zu einem gemeinsamen “Score” zusammenzufassen? Welche statistischen Kriterien gibt es dafür, dass sich Gruppen bezüglich eines Merkmals unterscheiden? Diese Art von Methodenwissen hat breite Anwendungsgebiete; wer es an einem Beispiel verstanden hat, kann es auch auf andere Beispiele übertragen.

Wir brauchen “wissenschaftlichen Spirit – wissenschaftliches Denken, wissenschaftliche Methoden, wissenschaftliche Fehler- und Diskussionskultur“ (<https://scilogs.spektrum.de/relativ-einfach/mai-thi-nguyen-kim-und-die-kleinste-gemeinsame-wirklichkeit-buchbesprechung/>)

Lerninhalte

„Wissenschaftliche Ergebnisse sagen wenig aus, solange ihr nicht die Methoden kennt, mit denen diese Ergebnisse erstellt wurden.“

Noch dazu herrsche in der Wissenschaft eine gewisse Fehlerkultur, ja, sie lebe davon, dass Fehler verbessert würden: „Jeder Irrtum ist eine Erkenntnis, die einen weiterbringt. Man irrt sich vorwärts.“ Wenn das mal keine Prämisse für ein gelingendes Leben ist!

(<https://www.diepta.de/news/die-kleinste-gemeinsame-wirklichkeit>)

„Wissenschaftliche Ergebnisse sagen wenig aus, solange man nicht weiß, auf welchem Weg sie entstanden sind“

<https://www.zdf.de/show/mai-think-x-die-show/maithink-x-folge-01-100.html>

„Wir sollten uns alles hinter Wissenschaft und wissenschaftlicher Evidenz versammeln können“

<https://www.together-for-fact-news.de/interview-nguyen-kim.html>

„Je mehr man googeln kann, desto mehr muss man wissen“

<https://www.zeit.de/gesellschaft/zeitgeschehen/2021-05/mai-thi-nguyen-kim-hass-internet-wissenschaftsjournalismus-pressefreiheit/seite-2>

Die neuesten Entwicklungen im Data Science-Umfeld, wie die Innovationen von Alpha GO Zero, IBM Project Debater, Optimierung

von Computerchips, Autonome Drohnen, die alle mehr und mehr in eine „Black-Box-Situation“ münden, d.h. man weiß nicht, wie der Computer zu den Ergebnissen kommt, und deshalb den Wunsch nach einer „Explainable AI“ hervorrufen, zeigen nur eins: Man muss sich unbedingt mit den methodischen Hintergründen von Algorithmen beschäftigen, um nicht zum nicht-wissenden blindgläubigen Anwender zu verkommen.

Relevanz für Studierende der Onlinekommunikation:

Die berufliche Zukunft und Qualifikationserwartung von AbsolventInnen der Onlinekommunikation wird stark geprägt sein von der Frage, ob diese ein ausgesprochen datengetriebenes Verständnis für die Online & Social Media Marketing- und PR-Welt mitbringen. Sonst werden sie in Zukunft nicht ernstgenommen oder von Konkurrenzqualifikationen verdrängt.

S. Burkhart führt dazu aus :

„Viele HR-Abteilungen und Geschäftsführer brüsten sich damit, dass sie auf einfache digitale Anwender Jobs 100te von Bewerbungen erhalten, um Stakeholder und den Markt in die Irre zu führen. Besonders kleinere Digitalbuden und Pseudo-Startups schaffen einen Graumarkt, in welchem junge Menschen durch Karrierechancen geblendet werden, die es tatsächlich aber nicht gibt. In der ersten KI-Welle, die 2019/2020 beginnen wird, werden über 90 Prozent dieser “Pseudo-Digital- Jobs” wegfallen. Für den Aufbau eines Sales-Funnels werden keine 20 Duale Studenten benötigt, sondern EIN qualifizierter KI-Experte. Diesen einen KI-Experten zu finden, DAS ist die Leistung des Unternehmens und von HR.“

(S. Burkhart: Millennials als Treiber für den Kulturwandel in Organisationen, 2019, S. 35)

Es reicht schon ein Blick auf die LinkedIn-Liste der „Hottest Skills That Got People Hired“, um zu sehen, dass Qualifikationsnachweise im Bereich der Data Science, Künstlichen Intelligenz bzw. des Maschinellen Lernens zu den Top Skills gehören. Von den Einstiegsgehältern ganz zu schweigen.

Die effiziente Befassung mit Daten und deren Ergebnisinterpretationen sowie eine kritisch-rationale Sicht auf die Datenergebnisse und ein Schutz vor einer kritiklosen Übernahme der per Software ermittelten Datenauswertungen erfordert eines ganz bestimmt: METHODEN-KENNTNISSE.

Emilio Galli-Zugaro (Leiter Unternehmenskommunikation ALLIANZ AG bis 2016):

„Kaufen und nutzen Sie keine Algorithmen, ohne sie zu verstehen und überprüfen zu können. Der Einsatz von Big Data ist ein großes Risiko, wenn man sich nur auf künstliche Intelligenz und Fremdgehirne verlässt. Gerade in der Kommunikation“

Um die bestehenden Fähigkeiten der Studierenden zu erweitern und zusätzliche, praxistaugliche und berufs- und arbeitsmarktrelevante Alleinstellungsmerkmale zu vermitteln, soll in diesem Fachmodul ein methodisch-operatives Rüstzeug im Umgang mit Data- und Web Mining-Verfahrenstechniken bzw. Maschinellen Lernverfahrensansätzen mit auf den Weg gegeben werden. AbsolventInnen des Bachelorstudiengangs sollen später schnell, kostengünstig, effizient und zielführend für verschiedene Kund/innen, Nutzer/innen und Entscheider Online & Social Media Marketing sowie PR-Dienstleistungen auf hohem qualitativem und wissenschaftlichem Niveau vollbringen. Auf einem der wichtigsten beruflichen Arbeitsfelder der Zukunft, der Aufbereitung von strukturierten und unstrukturierten (Massen-)Daten, nicht zuletzt durch die aufgekommene „Big Data“-Diskussion angestoßen, sind zur Erlangung von arbeitsmarktrelevanten, wettbewerbsfähigen Qualifikationsalleinstellungsmerkmalen u.a. gute methodische skills zur Analyse derartiger strukturierter und unstrukturierter Datenmengen dringend notwendig. Dazu muss man u.a. auf die in der „scientific and practice community“ bekannten und akzeptierten quantitativ-qualitativen, heuristisch- statistischen Verfahren zurückgreifen. Aber dies nicht kritiklos und „blind“. Das moderne Management benötigt Mitarbeiter, die fundierte (empirische) Analyse-, Klassifikations- und Prognosemethoden kennen und beherrschen, aber auch deren Aussagekraft und Grenzen bei der Datenerhebung, -aufbereitung, -analyse und -aufbereitung richtig einschätzen können; gerade im Zeitalter der (webbasierten) Massendatenproduktion („Big Data“) ist hier ein kritisch-wacher Sachverstand notwendig, denn die Ankündigungen sind beeindruckend:

Was sind das für Analyse-, Klassifikations- und Vorhersagemethoden, was können sie und was können sie nicht?

Warum wird über eine vereinfachte Darstellung nicht die eigentliche (begrenzte) Substanz dieser oft mathematisch formelhaft komplex dargestellten Methoden offengelegt, wie in diesem Fachmodul vorgesehen?

Will man durch formelhafte Berechnungskomplexität und komplizierte Herleitung wissenschaftlich beeindrucken, nach der Devise: Je schwieriger und schwer verständlich, desto besser die Analyse-, Klassifikations- und Prognosequalität? Baue ich hier eine eigene (fiktive, realitätsferne) Wissenschaftswelt auf, die lediglich dem armseligen „Beeindrucken“ gilt, die häufig dogmatisch und autoritär erscheint, statt dem eigentlichen Ziel, die ökonomische und soziale Realität zu erklären und zu prognostizieren?

Im Fokus der Lehrveranstaltung steht das Qualifikationsziel der anwendungsorientierten Vermittlung von Verfahrenstechniken des empirisch-experimentellen Data- und Web-Mining, insbesondere mit Bezug auf die Grundlagen des Maschinellen Lernens.

Nicht ausgebildet werden Programmierfähigkeiten, das machen Informatiker.

Ausgebildet sollen TeilnehmerInnen werden, die verstehen wollen, was es heißt, Daten zu analysieren, welche Algorithmen es gibt und wie diese funktionieren bzw. wo deren methodischen Schwächen sind, damit sie später nicht zu blindgläubigen Anwender/- innen degenerieren. Hier dient die angestrebte Qualifikation dazu, aus der Unmündigkeit des Nichtwissens zur Hype um die „Big Data“ herauszukommen.

Die Vermittlung mathematisch/heuristischer-statistischer, insb. „multivariater Verfahren“, löst oftmals ein „ungutes“ Gefühl aus, deshalb werden stellen sich für den Dozenten besondere Herausforderungen.

Dies erfolgt in Form einer „Anti-Hegel“-Lehrveranstaltung:

„Er hat dazu geführt, dass es in Universitäten – in vielen Universitäten, natürlich nicht in allen – eine Tradition gibt, Dinge hegelianisch auszudrücken, und dass die Leute, die das gelernt haben, es nicht nur als ihr Recht ansehen, so zu sprechen, sondern geradezu als ihre Pflicht. Aber diese sprachliche Einstellung, die Dinge schwierig und damit eindrucksvoll auszudrücken, die macht die deutschen Intellektuellen unverantwortlich... Die intellektuelle Verantwortlichkeit besteht darin, eine Sache so deutlich hinzustellen, dass man dem Betreffenden, wenn er etwas Falsches oder Unklares oder Zweideutiges sagt, nachweisen kann, dass es so ist“

Es gibt eine Art Rezept für diese Dinge:... Man sage Dinge, die großartig klingen, aber keinen Inhalt haben, und gebe dann Rosinen hinein – die Rosinen sind Trivialitäten. Und der Leser fühlt sich gebauchpinselt, denn er sagt, das ist ja ein ungeheuer schweres Buch! (Sir Karl Popper 1990)

Die Lehrveranstaltung soll daran gemessen werden, ob sie den kritisch-rationalen Anmerkungen von Karl Popper Folge geleistet haben.

Dumme Fragen gibt es nicht. Dazu nochmal Dr. Mai Thi Nguyen-Kim „Ich glaube, der traditionelle Journalismus hat ein Arroganzproblem – sie setzen für viele ihrer Inhalte und auch für ihre Sprache eine gewisse Bildung voraus, von der sich viele Jugendliche nicht abgeholt fühlen. Ein Beispiel ist Rezo. Auch bei der Diskussion um ihn schwingt immer eine gewisse akademische Arroganz mit. Wenn Rezo "worken" sagt in seinem Video statt arbeiten, dann erreicht er damit seine Zielgruppe – und das zählt. Die beste Sprache ist die, die ankommt.

Ich stelle mir Wissenschaftsvermittlung wie eine Zwiebel vor: Innen, im Kern, sind die wissenschaftlichen Studien, die Originale, die nur für eine kleine Auswahl zugänglich sind.

Das Einzige, was zählt, ist die Stärke der Evidenz. In einem Video können wir dann aber auch genau das thematisieren und erklären, wie Wissenschaft funktioniert und warum Evidenz nicht immer stark genug ist, um ein Fazit zu ziehen.

A. Aus (Trainings-)Daten „lernen“: Regeln (Rules) und Muster (Pattern)

1. Informationsökonomische Grundlagen
 - 1.1. Informationstheorie von C. Shannon
 - 1.2. Entropie – der mittlere Informationsgehalt
 - 1.3. Informationsgewinn (Information Gain)
 - 1.4. Klassifikation per Entscheidungsbaum (Decision Tree Learning)
 - 1.5. Anwendungsfälle des ID3-Algorithmus für Decision Trees (+Random Forest)
 - 1.6. Regressionsbäume – ein besondere Variante der Decision Trees
 - 1.7. Naiver Bayes Klassifikator für Data-, Text- und Web-Mining

2. Maschinelles Lernen
 - 2.1. Ziele, Begriffe und Anwendungsfelder des Maschinellen Lernens
 - 2.2. Regeln lernen und Muster erkennen: Supervised / unsupervised learning
 - 2.3. Beispielsalgorithmen: Find-S-Algorithm, Version Space, Find-G-Set-Algorithm
 - 2.4. Optimierungsansätze: Post-Pruning (Candidate Eliminate-Algorithm)
 - 2.5. Regel-Algorithmus: Separate-and-Conquer-Rule (Top-Down-Hill-Climbing)
 - 2.6. Precision & Accuracy
 - 2.7. Bottom-Up-Hill-Climbing (Special-to-General)
 - 2.8. Evaluation der Regeln (Learning outcomes): Sign Test, Covering-Algorithm

B. Daten-Ähnlichkeiten und Daten-Klassen (Cluster)

3. Ähnlichkeitsmessung (similarity measurement) zur Mustererkennung
 - 3.1. Binäre Ähnlichkeitsmaße (Jaccard-, M-Koeffizient etc.)
 - 3.2. Nominale Ähnlichkeitsmaße (Distanzmaße, L-Norm, Euklidische Distanz, Q- Korrelationskoeffizient etc.)
 - 3.3. Cluster-Analytik (agglomerative, divisive Clusteralgorithmen: Single-, Complete-, Average- Linkage; WARD-Algorithmus)
 - 3.4. Maschinelles Lernen und Cluster-Analytik (mit Inverse Distance Weighting, Value Difference Metric VDM, Relief-Algorithm zur Attributgewichtung, RISE, Initiale Accuracy)
 - 3.5. Chi-Quadrat und Korrespondenzanalyse (+ Maschinelles Lernen mit supervised/unsupervised discretization)
 - 3.6. Assoziationsanalyse (Warenkorbanalyse)
 - 3.7. Informationswissenschaftliche Ähnlichkeitsmessung (Vektorraummodell und WDF-IDF-Modell) für Data-, Text- und Web-Mining
 - 3.8. Anwendungsfälle der Ähnlichkeitsmessung (Empfehlungssysteme/Recommendersystems; content-based-filtering, collaborative filtering)

Literatur

1. Alpaydin, Ethem: Maschinelles Lernen, Oldenbourg-Verlag, 2008
2. Oestreich, M.; Romberg, O.: Keine Panik vor Statistik! Erfolg und Spaß im Horrorfach nichttechnischer Studiengänge. 3.Aufl., 2010
3. Foremann, John W.: Big Data smart mit Excel analysieren,

Wiesbaden 2015

4. Zweig, Katharina: Ein Algorithmus hat kein Taktgefühl, München 2019
5. Griffiths, Tom: Algorithmen für den Alltag, München 2019
6. Krämer, W.: So lügt man in der Statistik. 2011
7. Witten, Ian H.; Frank, Eibe; Hall, Mark A.: Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann, 3.Aufl., 2011
8. Ferber, Reginald: Information Retrieval. Suchmodelle und Data Mining für Textsammlungen und das Web, Dpunkt Verlag, 2003
9. Russell, Matthew: Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and other Social Media, O'Reilly Media, 2011
10. Kemper, Hans-Georg; Baars, Henning; Mehanna, Walid: Business Intelligence – Grundlagen und praktische Anwendungen, Vieweg+Teubner, 3.Aufl., 2010

Zusätzliche Unterlagen, Übungsaufgaben und Materialien.