

1.1 Recap last Lecture

- Assignment
 - positiv: gut gelöst, für manche Leute sehr viel einfacher, für manche schwieriger
 - Tradeoff: generalisierung vs. spezifizität
 - * Kunst ist: so generell wie möglich, so spezifisch wie nötig
- interessante Quellen für Sozialwissenschaften
 - Unmenge interessanter Daten, wenig Datensätze
 - Konversion, da Dokument nicht in maschinenlesbarer Form
- gerüstet um mit (fast) allen Textdaten zu arbeiten

1.2 Outline

- heute ein Sprung, der über praktischen Teil von Seminar hinausgeht
 - Mix aus Ethik und Entwicklung NLP
 - moderne NLP leistungsfähiger als je zuvor, aber mit Problemen
- AI ist Werkzeug, erstmal weder gut noch schlecht
 - Vergleich: Motor für Krankenwagen oder Panzer (Verteidigung/Angriff).
 - Internet für Wikipedia oder für Kinderpornographie
- Die Frage ist: Wer profitiert genau von was? Wer verliert dabei?

2.2 Ethics** is more than philosophy. It is **everywhere**.

3.2 An Example

3.3 You are applying for a job at a big company.**

- Ethik ist nicht nur abstrakt und gehört nicht nur in Philosophie
 - nicht Begriff ist wichtig, sondern Denkart
 - nachdenken über Ausgangslage + Konsequenzen
- Anekdoten aus eigenem Bewerbungsprozess
 - als Bewerber
 - für Jobs, die Tools zur automatischen CV-Verarbeitung machen

3.4 Does your CV pass the automatic pre-filtering?

- automatische Vorselektion Bewerbungen
- bestenfalls: naiv, schlechtfalls: anti-liberal/diskriminierend

3.5 Your interview is recorded. :sunglasses: :hot_face: What personal traits are inferred from that?

3.6 Don't worry about the future ...

- je mehr die Systeme können, desto mehr werden sie eingesetzt, desto unsichtbarer wird, was sie nicht können
- Moderne AI lernt Muster aus Daten. Gilt auch für NLP.
 - generalisiert blind oder eben genau so wie gelernt
- Ungleichheit wird reproduziert, gar verstärkt durch Systematik

- Geschlecht, Ethnie, sozioökonomisch
- Problematik ist Reichweite und Intransparenz
 - Menschen auch fallibel, aber mit grösser Varietät
- Anwendungen für tabellarische Daten, Text, Bild, Video
 - self-driving cars (walking pedestrians vs wheel chair pedestrians)
- Entwicklung NLP und Ethik hängen zusammen
 - quasi: je leistungsfähiger NLP, desto mehr Bias wird mitgelernt im aktuellen Paradigma
 - besseres Verständnis = bessere Data Science

3.7 Algorithmic Management of Labour Force

- Akkordarbeit und bei ungenügender Leistung automatische Entlassung?

4.7 An (R)evolution of NLP

4.8 From Bag of Words to Embeddings

- Ein Sprung, wie diese Technologie gewachsen ist
- Probleme
 - Bank kann Kreditinstitut bedeuten, in anderem Satz Parkbank
 - Haus und Gebäude sehr ähnlich, aber nicht reflektiert in Oberflächenform
- Syntax und Semantik um Wörter zu gruppieren

4.9 Bag of Words

- Vektorrepräsentation für ganzes Vokubular
- jedes Wort ist anderes als jedes andere, in unvergleichbarer Weise
- BoW lange Zeit Standard. Ergänzt durch zusätzliche Information wie POS

4.10 Representing a Corpus

- für den Computer müssen Daten tabularisiert werden für weitere Verarbeitung

4.11 “I eat a hot ___ for lunch.”

- Wörter können aber auch anders definiert werden und das will ich hier illustrieren
- Frage an Studis

4.12

- kontextuelle Bedeutung statt intrinsische Definition
- Saussure: Zeichen nur definiert durch andere Zeichen
- relationale Bedeutung: Objekt ist definiert durch Kontext
- lange ohne technische Implementation

4.13 Word Embeddings

- Seit 2013 hat sich alles verändert

- vector = list of numbers -> point in Euclidean space
- Idee: wenn Wort genau gleich gebraucht wird, dann selbe Stelle
- Synonyme, Analogien finden
- alles noch globale Information. Ein Wort hat genau ein Vektor
 - Was passiert mit mehrdeutigen Wörter (z.B. Bank)?
- Frage wie diese Repräsentationen genau gelernt wir nach Pause

4.14 Contextualized Word Embeddings

- alles lässt sich embedden (Wörter, Sätze, Paragraphen, Dokumente)

5.14 Modern NLP is propelled by data

5.15 Learning Associations from Data

- BERT wird in Google Search gebraucht

5.16 Cultural Associations in Training Data

- Analyse umkehren: Nicht nach Pronomen fragen, sondern nach Tätigkeiten
- Model trained on Wikipedia and Books (not Reddit)

5.17 Word Embeddings are biased ...

- Timnit Gebru (Google Ethics Lead) gefeuert für dieses Paper Ende 2020
- Daten sind nicht besser als wir und Gesellschaft trägt extreme Diskriminierungen mit sich
- Pause

5.18 In-class: Exercises I

6.18 Modern AI = DL

6.19 How does Deep Learning work?

6.20 Limitations of data-driven Deep Learning

- ohne Trainingsdaten aktuell nicht zu lösen

6.21 Current State of Deep Learning

- out of domain (schwarze vs weisse Menschen, anderes Textgenre)

7.21 Biased Data and beyond

7.22 Data = Digital Traces = Social Artifacts

- Was sind Daten?
 - Daten sind kein Abbild der Welt, nichts natürliches.
- Analog zu Romanos: Massenmedien sind nicht die Welt
 - Daten sind auch nicht die Welt
- Daten liegen nicht einfach herum, sondern gemacht (siehe Schritte)

7.23 Data vs. Capta

- viel Konflikt geht auf diese Realismus/Konstruktivismus Perspektive zurück

- Konstruktivismus heisst nur, Fragen zu stellen, wieso die Dinge sind, wie sie ausschauen

7.24

7.25 Two Sides of the AI Coin

7.26 Still doubts about practical implications?

- Ungarische Pronomen sind nicht gendered

7.27 And it goes on ...

7.28 Fair is a Fad

- Fair kann ziemlich vieles bedeuten, solange man es selbst definieren kann
 - demokratische Legitimität fehlt für all diese Systeme
- looking beyond data
 - invading privacy
 - economic monopolies
 - (unpaid) AI-trainers and click-workers
 - environmental costs

7.29

7.30 Data represents real life.

8.30 Questions?

8.31 References