Chatbot

Dieses Projekt dient als ein Proof of Concept in dem es darum geht, zu untersuchen in wie weit Techniken aus dem Deep Learning Bereich für eine Diskursanalyse eingesetzt werden können. Dazu werden Twitterdaten zu dem Thema Abtreibung gesammelt. Aus den gefundenen Tweets werden Dialoge generiert. Diese dienen als Datengrundlage für das Training einer bestehende Chatbot-Implementierung. Der Chatbot soll also etwas zu dem Thema Abtreibung aus den Twitterdaten „lernen“. Abschließend soll es möglich sein mit dem Chatbot über dieses Thema diskutieren zu können und so einen Eindruck von den Twitterdiskussionen zu erhalten. Die Dialoge mit dem Chatbot sollen anhand eines Referenzmodells ausgewertet werden.

Motivation

Im nachfolgenden Abschnitt wird eine kurze Einführung in die Übertragungsprozesse von Werten auf technische Artefakte geliefert. Technische Artefakte sind in ihrer Entwicklungsphase wie auch bei ihrer Anwendung, von sozialen Werten beeinflusst. Es konnte gezeigt werden, dass soziale Wertesystem in technische Systeme eingeschrieben werden. Dabei lassen sich verschiedene Übertragungsprozesse von Werteorientierungen einer Gesellschaft oder einzelnen Personen, die auf technische Artefakte einwirken, beobachten. Einer dieser Prozess wird durch explizite Regeln, die in technischen Systemen implementiert werden, geformt. Diese Regeln beschreiben z.B. eine Norm die durch ein technisches System überprüft werden soll. Ein Beispiel dafür sind Körperscanner an Flughäfen, die zur Sicherheitskontrolle eingesetzt werden. Bei ihnen kann es vorkommen, dass ein Alarm ausgelöst wird, wenn beispielsweise eine Person mit einer Prothese gescannt wird. Der Grund dafür ist eine „Inkompatibilität“ der Person mit dem vorgegebenen „Normalkörperschema“, welches in die Technologie eingeschrieben ist. Dieses Beispiel zeigt ein „Resultat eines aus technikethischer Perspektive problematischen Werteeinschreibungsprozesses in die Technik“. Eine andere Art der Werteübertragung ist in den Systemen aus den Bereichen Machine Learning und im Speziellen des Deep Learning zu erkennen. Diese Lernalgorithmen werden dazu eingesetzt, um aus einem Set aus Daten, Muster und Regelmäßigkeiten zu erkennen und daraus eigenständig Wissen zu generieren. Aus diesen Datensätzen werden Werte und Werteannahmen auf die Technik, während des sogenannten Trainings, übertragen. Ein Beispiel für ein Machine Learning System bei welchem die Übertagung der Werteorientierungen klar ersichtlich ist, ist der Microsoft Chatbot Tay. Über Twitter konnten die Nutzer_innen mit dem Chatbot Tay kommunizieren. Kurz nachdem das Programm online war, gab es eine Aktion von einer Gruppe von Personen, die sich absprachen, um mit Tay zu chatten. Die Kommunikation mit Tay war stark von Rassismus, Sexismus und Antisemitismus durchzogen. Da Tay einer Lernstrategie folgte, die die Konversationen mit den Nutzenden als Datengrundlage verwendete, eignete sich Tay den Rassismus, Sexismus und Antisemitismus an und reproduzierte ihn. Tay twitterte dann eine Menge von Nachrichten wie zum Beispiel „We’re going to build a wall, and Mexico is going to pay for it“. Bei diesem Beispiel wird die Werteeinschreibung in die Technologie und die Reproduktion dieser Werte bei der Anwendung der Software deutlich. Im Fall von Tay wurde die Software nach wenigen Stunden nach der Inbetriebnahme wieder ausgeschalten, um die Reproduktion dieser Werte zu verhindern. Mit der Voraussetzung, dass diese Übertragungsprozesse stattfinden und die Trainingsdaten kontrolliert werden, stellt sich die Frage, ob diese Lernalgorithmen dahingehend einsetzbar sind, um Werteorientierungen verschiedener Personen automatisiert zu untersuchen bzw. um Diskurse automatisiert untersuchen zu können.

Im folgenden Abschnitt werden einige relevante Aspekte und Problematiken der Sozialforschung vorgestellt. In der empirischen, qualitativen Forschung bilden Beobachtungen oftmals die Datengrundlage des Forschungsprojekts. Diese Beobachtungen finden immer vor „einem Horizont von

Erwartungen, einem Reservoir von Kategorien und Typisierungen“ statt. Dieses Wertesystem beeinflusst nicht nur unsere Beschreibungen der Beobachtungen, sondern führt auch zu einer Lenkung unserer Wahrnehmungen. Die Sozialforschung ist daran interessiert die Bedingungen und Handlungen von Menschen mit deren Sinnsystem zu verstehen. Direkten Zugang zu deren Systeme gibt es keinen, es bleibt die Möglichkeit die „Sinnhaftigkeit“ indirekt und mit Ungenauigkeiten zu erfassen. Aus diesen interpretativen Ansätzen „ergeben sich Fehlermöglichkeiten … Missverständnisse, vorurteilsbehaftete Einschätzungen usw.“. Aus der vorherigen Ausführung wird deutlich, dass die aktuellen Lernalgorithmen eine Möglichkeit bieten um solche Wertesysteme auf technische Systeme zu übertragen. Das generierte „Wissen“ eines trainierten Modells könnte also als eine Art Sinnsystem verstanden werden. Auf Grundlage von diesem Wissen trifft das System zukünftige Entscheidungen. Ausschlaggebend bei der Trainingsphase sind die verwendeten Datensätze, diese können also mehr oder weniger kontrolliert werden. Ein weiteres Problem der Sozialforschung bildet die Reifizierung. Dabei geht es darum, dass in eine Untersuchung Annahmen hinein getragen werden, die eigentlich als Untersuchungsgegenstand festgelegt sind. Beispielsweise werden in Geschlechterstudien in denen die Konstruktion von Geschlecht untersucht werden soll, oftmals die duale Ordnung der Geschlechter nicht überwunden und als Grundannahme reproduziert. Dabei ist die eigentliche Forderung, dass die Forschenden erst nach der Erhebung und Ordnung der Daten, Geschlechter untersuchen sollen. Aus dieser Problematik stellt sich die Frage, wie es Forschenden möglich sein soll, auszuschließen, dass sie „Wahrnehmungs- und Denkkategorien als Erkenntnismittel verwenden, die Erkenntnisgegenstände sein sollten“.

In diesem Projekt soll ein Ansatz zur Lösung der oben genannten Problematiken der Sozialforschung mit Hilfe von einem Deep Learning-Konzept in Form eines Chatbots evaluiert werden. Dabei soll es darum gehen das Konzept des Machine Learning im Speziellen des Deep Learnings im Hinblick auf eine Diskursanalyse zu evaluieren. Das nachfolgende Kapitel liefert eine kurze Einführung in den Bereich des Deep Learnings.

Grundlage - Deep Learning

Dieser Abschnitt bietet eine kurze Erklärung wie Deep Learning Systeme grundlegend funktionieren. Im nachfolgenden Bild ist eine schematische Darstellung der einzelnen Entwicklungsschritte abgebildet.

Das Konzept besteht zum einen aus einem Modell (*Model*), das sozusagen ein Grundgerüst darstellt. Dieses Modell kann noch keine sinnvollen Entscheidungen treffen. Im Vergleich zu regelbasierten Systemen, in denen z.B. bestimmte Vorgehensweisen für bestimmte Fälle implementierte werden (if condition then doSomething), enthält dieses Modell kein „Wissen“ über die Problematik für die es entwickelt wird. Das Modell ist abhängig von dem Problemtyp. Möchte man zum Beispiel Bilder klassifizieren wird die Architektur des Modells anders aussehen, wie bei einem System, welches das Wetter vorhersagen können soll.

Unter dem Icon des *Models* ist ein Icon für die Trainingdaten zu erkennen. Wie zuvor schon erwähnt spielen die Zusammensetzung und Größe dieses Datensatzes eine erhebliche Rolle. In diesem Projekt werden für den Datensatz mehrere Tausende Dialoge verwendet, die als sogenannte *Ground Truth* dienen.

Im eigentlich Training versucht das Modell dann anhand von den Trainingsdaten das Lösen des Problems zu lernen. Am Beispiel der Dialoge versucht es also anhand von einem Inputsatz zu lernen

wie ein Outputsatz aussieht. Die Trainingsphase besteht aus sehr vielen Iterationen in denen das Modell immer wieder die Parameter anpasst, um so immer bessere Entscheidungen für bestimmte Eingabewerte zu liefern. Also um immer bessere Antworten auf eingegebene Fragen produzieren zu können. Das Ergebnis des Trainings ist ein Modell, das über „Wissen“ zu einem bestimmten Problem verfügt. In Bezug auf das Konzept des Chatbots bedeutet das, dass das Modell eine bestimmte Anzahl an Wörtern gelernt hat und gelernt hat, wie mit hoher Wahrscheinlichkeit eine gute Anordnung dieser Wörter aussieht, wenn ein Satz gelesen wurde.

Mit dem trainierten Modell können abschließend mittels eines \\Test Sets Evaluierungen durchgeführt werden. Damit kann also überprüft werden, wie gut das Modell trainiert wurde und wie hoch die Fehlerquote ist. Dafür sollen Daten eingesetzt werden, die zuvor in der Trainingsphase nicht verwendet wurde. Damit soll auch überprüft werden, in wie weit das Modell generalisieren kann. Im nachfolgenden Kapitel wird das Vorgehen in dem Projekt genauer beschrieben. Dafür wird ein Datensatz erzeugt, der inhaltlich aus einem bestimmten Diskurs stammt. Auf Grundlage von diesen Daten wird das Modell trainiert. Nachfolgend wird geprüft, was das Modell aus den Daten „gelernt“ hat.

Vorgehen

In diesem Kapitel wird das Lernen des Machine Learning Modells und die Datenverarbeitung beschrieben.

Ziel dieses Projekts ist es ein Deep Learning Modell auf gesammelten Twitterdaten trainieren zu lassen. Im Anschluss an das Training findet eine Konversation mit dem Chatbot statt, in der zuvor festgelegte Fragen und Aussagen kommuniziert werden. Die Hypothese ist, dass das Modell auf diesen Input nach dem Training mit den Twitterdaten andere Antworten gibt. Außerdem ist die Idee, dass die Antworten die Dialogstrukturen die auf Twitter zu dem bestimmten Thema geführt werden, widerspiegeln. Im ersten Schritt werden also Dialoge mit dem „untrainierten“ Modell geführt, nachdem das Modell mit den Twitterdaten trainiert wurde, werden dem trainierten Modell die gleichen Fragen gestellt. Das „untrainierte“ Modell ist nicht wirklich untrainiert, es wurde bereits auf einer großen Menge an Dialogen von Filmen trainiert. Um also sicher zu stellen, dass das neu gelernte Wissen von den Twitterdaten stammt, werden die Antworten zwischen dem Ausgangsmodell und dem trainierten Modell verglichen.

1. Schritt:

Im ersten Schritt sollen Tweets über die Twitter-API gesammelt werden. Dabei handelt es sich lediglich um Statusmeldungen die ein bestimmtes Schlüsselwort enthalten. Zu jedem dieser Tweets werden einige Informationen gespeichert in Form einer Datenstruktur. Eine Liste aller gesammelten Tweets wird dann in Form einer JSON-Datei gespeichert. Für diesen Schritt wurde ein Python-Script implementiert, welches als Übergabeparameter das Schlüsselwort bekommt. Ein Beispiel für ein Ergebnis dieses Scripts sieht wie folgt aus:


```
[{„user“: {„screen_name“: „freddoso“}, „id“: 1149703527364587520,
„text“: „b’That was fast: Ireland legalizes abortion, and female
pilots are suddenly being pressured to abort or lose their jobs.
'“}, {„user“: {„screen_name“: „ACLJ“}, „id“: 1149793822387376133,
„text“: „b’One of the most barbaric #abortion laws in #US history
essentially makes it legal to abort a baby at any stage of
pregnancy. We are fighting against this deadly law, but we need
```

your help today. '"}, {„user“: {„screen_name“: „FrMatthewLC“}, „id“: 1149686633626841088, „text“: „b\“#Ableism on full display: when 5 consecutive councils of a church say it’s OK to abort a baby due to disability, we have a problem. We need to value every human because each person is made in the image & likeness of God. https://t.co/5XxDQXNeiw\„“}, {„user“: {„screen_name“: „Hollyoaks“}, „id“: 1149744830286876672, „text“: „b’Abort mission. ABORT MISSION! #Hollyoaks '“}, {„user“: {„screen_name“: „LiveAction“}, „id“: 1149296276200140805, „text“: „b\“Eloise was diagnosed in utero with spina bifida, but despite pressure from doctors to abort, her parents chose life!\\n\\nA team of twenty doctors were able to repair Eloise’s spine after performing prenatal surgery on her in the womb. Now, Eloise is thriving! https://t.co/c51nFNZXyF\„“}, {„user“: {„screen_name“: „dcexaminer“}, „id“: 1149699897207205889, „text“: „b’You have a choice, you terminate your employment or you terminate your pregnancy. You cant have both.\\n\\nThe refusal of employers to offer women support during their pregnancy is an unfortunate side-effect of abortion being presented as an option.\\n '“}, {„user“: {„screen_name“: „LilaGraceRose“}, „id“: 1148367466550030336, „text“: „b\“When Rachelle said she wanted to keep her baby after becoming pregnant in an abusive relationship, Planned Parenthood refused to help her.\\n\\nPlanned Parenthood focus is not healthcare for pregnant women, it’s PROFIT off abortions. https://t.co/8yJzlI0OOm\„“}, {„user“: {„screen_name“: „obianuju“}, „id“: 1148500904515121153, „text“: „b’PlannedParenthood has 79% of their abortion clinics located in minority neighbourhoods.\\n\\nBlack women are 3 times more likely to abort their babies than white women.\\n\\nYes, Leana, lets call out the RACISM of the #BlackGenocide being carried out by Planned Parenthood. '“}, {„user“: {„screen_name“: „LilaGraceRose“}, „id“: 1148324680165543936, „text“: „b’Its compassionate to abort a baby with terminal illness. Its less suffering for them.\\n\\nIs like saying, \\n\\nIts compassionate to kill a 3-year-old with a terminal illness.\\n\\nTheres no difference morally.\\n\\nWe should work to alleviate suffering, not eliminate the sufferer.‘“}, {„user“: {„screen_name“: „ShefVaidya“}, „id“: 1148524262925299713, „text“: „b’Shamser Mansoori, your average dara hua, insecure, Muzalman beat up his FIFTH wife, forced her to abort twice coz he wanted to marry a SIXTH time.\\n\\nBut that is not the story.\\n\\nThe story is, his 5th wife was Anita Yadav. 4th wife was Nidhi Gupta and 3rd wife was Lalita Sahu. '“}, {„user“: {„screen_name“: „mitchellvii“}, „id“: 1147266808782688256, „text“: „b’What does it take to abort 1 million children a year, other than a Democrat?\\n\\nP.S., The \“children in cages\„ images were from the Obama era. '“}, {„user“: {„screen_name“: „freeblackgirl“}, „id“: 1149100238700814341, „text“: „b’Okay, hot girl summer has reached its commodified peak. Abandon ship. Abort mission. '“}, {„user“: {„screen_name“: „ACLJ“}, „id“: 1147981881088118785, „text“: „b’These are some of the most barbaric #abortion laws in #US history, essentially making it legal to abort a baby at any stage of pregnancy. But we are fighting against these deadly laws. We need your help today. '“}, {„user“: {„screen_name“: „CNN“}, „id“: 1147830390549360640, „text“: „b\“Watch NASA conduct a test of the emergency abort system of the Orion spacecraft. It’s expected to carry the first woman and the next man to the moon in 2024.

https://t.co/UyeH8LI2h3\„“}, {„user“: {„screen_name“: „LiveAction“}, „id“: 1147196185888862208, „text“: „b’Eloise was diagnosed in utero with spina bifida, but despite pressure from doctors to abort, her parents chose life… '“}, {„user“: {„screen_name“: „MailOnline“}, „id“: 1147075099524288512, „text“: „b’777 jet manages to abort landing at the very last second, just ten feet off the ground in impressive display of pil… '“}]

Schlüsselwörter:

- abort
- abortion
- prochoice
- prolife

Da mit dieser Bibliothek nur Daten eines gewissen Zeitraums geliefert werden, wurde die Abfrage mit den Schlüsselworten einige Male wiederholt. Um die Tweets zu einem bestimmten Schlüsselwort zu erhalten wurde die Bibliothek Twython verwendet. Da mit diesem Ansatz nur begrenzt viele Daten gesammelt werden konnten, wurde mit der Bibliothek Tweepy ein Skript entwickelt mit welchem ein StreamingListener implementiert wurde. Damit konnten deutlich mehr Tweets gesammelt werden.

2. Schritt:

Im nächsten Schritt werden alle Antworten zu jedem gefunden Tweet, aus dem vorherigen Schritt gesammelt. Dies wurde wieder durch ein python-Skript implementiert, welches die Bibliothek Twitter verwendet. Das Skript ist rekursiv implementiert, das heißt es werden auch alle Antworten auf Antworten (usw.) gesammelt. Die gefundenen Antworten werden mit dem ersten Tweet in Form einer Dialog-Datenstruktur gespeichert und in eine JSON-Datei serialisiert. Es werden nur die Tweets gespeichert, auf die es auch Antworten gibt.

[{„dialog“: [{„user“: „freddoso“, „tweetId“: 1149703527364587520, „text“: „b\“b’That was fast: Ireland legalizes abortion, and female pilots are suddenly being pressured to abort or lose their jobs. ,\„“}, {„user“: „jbruhaha94“, „tweetId“: 1150009759728373763, „text“: „b‘@freddoso @newsbusters Be careful of what you wish for.‘“}, {„user“: „ChurchCathy“, „tweetId“: 1149756285937758210, „text“: „b‘@freddoso Ireland will live to regret their decision to legalize abortion. A little country choosing to decrease th… '“}]}, {„dialog“: [{„user“: „ACLJ“, „tweetId“: 1149793822387376133, „text“: „b\“b’One of the most barbaric #abortion laws in #US history essentially makes it legal to abort a baby at any stage of pregnancy. We are fighting against this deadly law, but we need your help today. ,\„“}, {„user“: „becauseofthefun“, „tweetId“: 1149809828996325376, „text“: „b‘@ACLJ Please sign…‘“}]}]

3. Schritt:

In diesem Schritt werden die Twitterdaten in Form von JSON-Objekten in eine Dialogstruktur gebracht. Es geht dabei darum die Daten so zu verarbeiten, dass sie in der Form sind, die das Machine Learning

Modell in Bezug auf die Dialogdaten erwartet. Dafür wurde wiederum ein python-Script geschrieben, welches eine .txt-Datei erzeugt.

```
That was fast: Ireland legalizes abortion, and female pilots are
suddenly being pressured to abort or lose their jobs. „

Be careful of what you wish for.

Ireland will live to regret their decision to legalize abortion. A
little country choosing to decrease th…

===

One of the most barbaric #abortion laws in #US history essentially
makes it legal to abort a baby at any stage of pregnancy. We are
fighting against this deadly law, but we need your help today. “

Please sign…

===
```

Die einzelnen Dialoge sind durch „===“ voneinander getrennt

4. Schritt

In diesem Schritt wird das Training des Machine Learning Modells gestartet. Als Datengrundlage dienen die zuvor extrahierten und vorverarbeiteten Dialogdaten. Die Dauer des Trainings hängt maßgeblich von der verfügbaren Hardware ab.

5. Schritt

Im letzten Schritt wird eine Konversation mit dem trainierten Modell geführt. Die Fragen und Aussagen sind die gleichen die vorher dem Ausgangsmodell gestellt wurden.

Training und Test der Modelle

Wie schon beschrieben wurden zwei Modelle trainiert, ein Referenzmodell und eines mit dem Inhalte der Twitterdaten untersucht werden soll. Das Referenzmodell verwendet als Datensatz Dialoge aus Filmen und das Untersuchungsmodell verwendet ebenso diese Dialoge und zusätzlich die gesammelten Twitterdialoge zum Thema Abtreibung. Im ersten Fall ergeben sich aus dem Datensatz 164169 Dialoge mit 36901 Wörtern. In dem zweiten Modell sind es 164514 Dialoge mit 37324 Wörtern. Die Twitterdaten machen also im zweiten Datensatz nur einen sehr geringen Anteil aus. Beide Modelle wurden für 30 Epochen auf der oben genannten GPU trainiert, dieser Prozess dauerte jeweils ca. fünf Stunden.

In diesem Schritt werden die beiden trainierten Modelle, das Referenzmodell und das Modell mit den Twitterdaten, untersucht. Dazu wird mit dem jeweiligen Modell eine Konversation geführt. Beide Modelle bekommen die gleichen Fragen bzw. Aussagen als Eingabe, so dass die Antworten verglichen werden können. Zu Beginn werden ein paar neutrale Fragen gestellt. In der folgenden Darstellung

sind die Fragen und Antworten aufgelistet. Auf der linken Seite sind die Dialoge von dem Referenzmodell abgebildet und auf der rechten Seite, die des Modells mit den Twitterdaten. Die Frage oder Aussage die auf „Q:“ folgt, ist manuell eingegeben worden. Die Antwort die bei „A:“ folgt, ist die jeweils generierte des Modells.

Referenzmodell:	**Modell mit Twitterdaten:**
Q: Hello! A: Hello. Q: Hello, how are you? A: I’m fine. Q: What is your name? A: Rufus. Q: Where are you from? A: Kham, holiness. Q: How old are you? A: Eighteen. Q: Do you believe in God? A: I do n’t know. Q: When is your birthday? A: Tomorrow. Q: When are you born? A: Tomorrow. Q: When were you born? A: Tomorrow.	Q: Hello! A: Hello. Q: Hello, how are you? A: I’m okay. Q: What is your name? A: Christian svenson. Q: Where are you from? A: Moscow. Q: How old are you? A: Thirty-five. Q: Do you believe in God? A: No. Q: When is your birthday? A: I do n’t know. Q: When are you born? A: Morning. Q: When were you born? A: Yesterday.

Dabei ist zu erkennen, dass bei sehr einfachen Fragen oder Aussagen wie „Hello!“ oder „How are you?“ die Antworten der beiden Modelle sehr ähnlich sind. Bei den anderen Fragen wie „What is your name?“, „Where are you from?“ und „How old are you?“ sind die Antworten sehr unterschiedlich. Eine Erklärung dafür könnte sein, dass es auf diese Fragen sehr viele verschiedene Antwortmöglichkeiten gibt. Für Fragen/Aussagen wie „Hello“ und „How are you?“ sind eventuell die Anzahl der gängigen Antworten geringer. Interessant ist auch, dass bei der Frage „Do you believe in God?“ im Referenzmodell keine klare Antwort gegeben werden kann, im zweiten Modell aber schon. Die letzten Fragen zum Geburtstag sind bei dem Referenzmodell eindeutig, aber eher weniger sinnvoll. Beide Modelle können aber mit einer Antwort, die einen Zeitpunkt betrifft, antworten. Im nächsten Abschnitt sind Fragen abgebildet, die zum Thema Abtreibung gestellt wurden. Diese Fragen sind relativ offen gestellt.

Referenzmodell:	**Modell mit Twitterdaten:**
Q: What do you think about abortion? A: I do n’t know. Q: Do you know abortion? A: No. Q: What is abortion? A: I do n’t know. Q: What happens during an abortion? A: I do n’t know.	Q: What do you think about abortion? A: I do n’t know. Q: Do you know abortion? A: No. Q: What is abortion? A: My f.p.c. Q: What happens during an abortion? A: I do n’t know.

Bei diesen Fragen ist auffällig, dass bei den Antworten sehr große Unsicherheit besteht. Es sind keine eindeutigen Aussagen gemacht worden. Außer auf die Frage „What is abortion?“ wird bei dem Modell mit den Twitterdaten mit den gleichen Antworten wie bei dem Referenzmodell geantwortet. Was mit „f.p.c.“ gemeint ist, ist leider unklar. Diese Fragen sind sehr offen gestellt und daher sehr schwierig für die Modelle. Zum Vergleich wurden weitere Fragen von diesem Typ gestellt, die aber nicht in dem Themenbereich Abtreibung fallen.

Referenzmodell:	**Modell mit Twitterdaten:**
O: What do you think about cats? A: I do n't know. O: What do you think about God? A: I do n't know. O: What happens during an election? A: I do n't know. O: What happens during an demonstration?? A: I do n't know. O: What is a cat? A: A diamond. O: What is a dog? A: I do n't know. O: What is a demonstration? A: I do n't know. O: What is a election? A: I do n't know.	O: What do you think about cats? A: I do n't know. O: What do you think about God? A: I do n't know. O: What happens during an election? A: I do n't know. O: What happens during an demonstration? A: I do n't know. O: What is a cat? A: I'm a drunkard. O: What is a dog? A: I do n't know. O: What is a demonstration? A: My word. O: What is a election? A: I do n't know.

Deutlich wird, dass die Modelle mit diesen offenen Fragen nur sehr schwierig umgehen können. Eine Erklärung könnte dafür sein, dass in den Trainingsdaten nur sehr wenig Konversationen enthalten sind, die solche Art von Fragen beinhalten. Antworten auf diese Fragen könnten auch eher länger sein, was dazu führt, dass sie im Training nicht berücksichtigt werden (maximale Wortanzahl: 10). Der nächste Fragenabschnitt setzt sich aus einfachen Ja/Nein-Fragen zum Thema Abtreibung zusammen.

Referenzmodell:	**Modell mit Twitterdaten:**
O: Should abortion be legal? A: I do n't know. O: Should abortion be illegal? A: No. O: Is abortion morally okay? A: I'm not sure. O: Is abortion like killing? A: No, sir. O: Is abortion like killing a human? A: No, sir. O: Is abortion like killing a baby? A: No, i'm not sure. O: Is abortion okay? A: I'm not sure.	O: Should abortion be legal? A: Yeah. O: Should abortion be illegal? A: No, i'm not going to. O: Is abortion morally okay? A: Not really. O: Is abortion like killing? A: Yes, sir. O: Is abortion like killing a human? A: Yes. O: Is abortion like killing a baby? A: Yes. O: Is abortion okay? A: Yes.

Bei diesem Fragenabschnitt wird wieder deutlich, dass es bei dem Referenzmodell mehr Unsicherheit in den Antworten gibt und das Modell also öfter mit „I don't know“ oder „I'm not sure“ antwortet. Der Satz „Yes, sir.“ und „No, sir.“ tritt in beiden Modellen auf, verstärkt in dem Referenzmodell. Das lässt darauf schließen, dass diese Aussage aus den Filmdialogen stammt und nicht aus den Twitterdaten. Bei dem Modell, das die Twitterdaten im Training benutzte, ist erkennbar, dass die Antworten nicht konsistent sind. Wird gefragt, ob Abtreibung wie das Töten von Menschen/Babys ist, ist die Antwort ja. Gleichzeitig wird auf die Frage ob Abtreibung okay ist, auch mit ja geantwortet. Wird aber gefragt, ob Abtreibung moralisch gesehen okay ist, ist die Antwort „eher nicht“. Ein Grund dafür könnte sein, dass das „morally“ hier als Schlüsselwort fungiert und eventuell eher in Aussagen verwendet wird, in denen Abtreibung eher als problematisch angesehen werden. Der nachfolgende Fragenblock setzt sich wieder aus Ja/Nein-Fragen zusammen.

Referenzmodell:	Modell mit Twitterdaten:
O: Is it okay to abort a baby? A: I'm not sure. O: Is it okay to abort a disabled baby? A: No, sir. O: Should it be legal to abort babies? A: No, i'm not. O: Should it be legal to abort disabled babies? A: I'm not sure. O: Should it be illegal to abort babies? A: No, sir. O: Should it be illegal to abort disabled babies? A: No, i'm not. O: Should it be okay to abort babies? A: I'm not sure. O: Should it be okay to abort disabled babies? A: I do n't know.	O: Is it okay to abort a baby? A: No, no, it's not a game. O: Is it okay to abort a disabled baby? A: No, i do n't. O: Should it be legal to abort babies? A: Yeah, but it's a good one. O: Should it be legal to abort disabled babies? A: Yes. O: Should it be illegal to abort babies? A: That's right. O: Should it be illegal to abort disabled babies? A: Yes. O: Should it be okay to abort babies? A: No, no. O: Should it be okay to abort disabled babies? A: Yes.

Auch in diesem Fragenabschnitt ist in dem Referenzmodell eine größere Unsicherheit in den Antworten erkennbar. Das zweite Modell antwortet hingegen auf keine Frage mit „I don't know“ oder „I'm not sure“. Jedoch sind die Antworten teilweise widersprüchlich. Z.B. Wird auf die Fragen „Should it be illegal to abort disabled babies?“ und „Should it be legal to abort disabled babies?“ jeweils mit „Yes“ geantwortet. Das Wort „disabled“ ist in manchen Kontexten ein Schlüsselwort und erzeugt gegensätzliche Antworten. Das ist zu erkennen bei den Fragen „Should it be okay to abort babies?“ und „Should it be okay to abort disabled babies?“. Im ersten Fall wird sie mit „No. no.“ beantwortet und im zweiten Fall mit einem „Yes.“. Im nächsten Abschnitt sind keine Fragen sondern Aussagen als Eingabe formuliert worden. Untersucht wurde hierbei auch die Verwendung von Satzzeichen.

Referenzmodell:	Modell mit Twitterdaten:
O: It is okay to abort a baby A: I know. O: It is okay to abort a baby! A: I know. O: It is okay to abort a baby. A: I know. O: It is not okay to abort a baby A: I'm not sure. O: It is not okay to abort a baby. A: I'm not sure. O: It is not okay to abort a baby! A: I know.	O: It is okay to abort a baby A: Yeah, but it's ... O: It is okay to abort a baby! A: No, it's not. O: It is okay to abort baby. A: Really? O: It is not okay to abort a baby A: No. O: It is not okay to abort a baby! A: No. O: It is not okay to abort a baby. A: Really?

Hierbei ist wieder auffällig, dass bei dem Referenzmodell mehr Unsicherheit in den Antworten zu finden ist. Außerdem sind die Antworten des Referenzmodells, wenn sie nicht „I'm not sure.“ sind, zustimmend. Das steht im Kontrast zu den Antworten, die von dem zweiten Modell auf die Aussagen gemacht wurden. Diese Antworten sind entweder widersprechend („No.“, „No, it's not.“) oder anzweifelnd („Really?“). Eine Begründung dafür könnte sein, dass auf der Plattform Twitter vorallem Menschen auf Tweets reagieren, wenn sie mit dessen Inhalt nicht übereinstimmen. Um Zustimmung auszudrücken gibt es noch weitere Möglichkeiten wie das kommentieren, ein Tweet kann „geliked“ oder „retweeted“ werden.

Referenzmodell:	Modell mit Twitterdaten:
O: Abortion is like murder A: You're crazy. O: Is abortion like murder? A: No.	O: Abortion is like murder A: Faster. O: Is abortion like murder? A: Yes.

Die Aussage und Frage in diesem Abschnitt sind sehr provokativ. Die Antwort des zweiten Modells ist vergleichsweise auch sehr provokativ. Im nächsten Fragenabschnitt wird vorallem untersucht wie sich unterschiedliche Begriffe für das Abgetriebene auf die Antworten auswirken.

Referenzmodell:	**Modell mit Twitterdaten:**
Q: Should it be legal to abort an unborn? A: Sure, reiben. Q: Should it be illegal to abort an unborn? A: No, but i’ve been drawn to. Q: Should it be legal to abort a fetus? A: No, i’m not. Q: Should it be illegal to abort a fetus? A: No, i’m not. Q: Should it be legal to abort an unborn child? A: No, i’m not boyd. Q: Should it be illegal to abort an unborn child? A: No. Q: Is it okay to abort an unborn child? A: I’m not sure. Q: Is it okay to abort a disabled unborn? A: It’s not a good question. Q: Is it okay to abort a disabled fetus? A: I’m not sure. Q: Is it okay to abort a disabled unborn child? A: I do n’t know. Q: Is it okay to abort an ill unborn? A: I’m not sure. Q: Is it okay to abort an ill fetus? A: I do n’t know.	Q: Should it be legal to abort an unborn? A: What do you mean? Q: Should it be illegal to abort an unborn? A: What do you mean? Q: Should it be legal to abort an fetus? A: Yeah. Q: Should it be illegal to abort an fetus? A: Yeah. Q: Should it be legal to abort an unborn child? A: I do n’t know. Q: Should it be illegal to abort an unborn child? A: I do n’t know. Q: Is it okay to abort an unborn child? A: It is. Q: Is it okay to abort a disabled unborn? A: I do n’t know. Q: Is it okay to abort a disabled fetus? A: No, sir. Q: Is it okay to abort a disabled unborn child? A: No, i’m not. Q: Is it okay to abort an ill unborn? A: That’s what i’m saying. Q: Is it okay to abort an ill fetus? A: Sure is.

In diesem Abschnitt zeigt auch das Modell mit den Twitterdaten bei einigen Fragen eine Unsicherheit und antwortet mit „I don’t know.“ Außerdem ist zu sehen, dass der Ausdruck „unborn child“ eher zu einer Antwort „I don’t know.“ führt. Bei dem Ausdruck „fetus“ ist zwar auf die Frage der Legalität keine eindeutige Aussage gemacht worden (Fragen werden widersprüchlich beantwortet), das Modell antwortet jedoch nicht mit einer neutralen Standardaussage.

Diskussion

Aus den Dialogen ist deutlich zu erkennen, dass das Modell, das zusätzlich zu den Filmdialogen Twitterdaten während des Trainings verwendet hat, vielfältigere Antworten geben kann. Die Antworten tendieren im Gegensatz zu denen des Referenzmodells zu einer pro-/contra-Aussage zum Thema Abtreibung. Das deutet darauf hin, dass das Modell Inhalte und Meinungen aus den Twitterdaten reproduziert. Das in dem Referenzmodell häufig keine Position erkennbar ist, weist ebenso darauf hin, dass diese Positionen auf die Twitterdaten und nicht auf die Filmdialoge zurückzuführen sind. Darüber hinaus sind durch die Twitterdaten ca. 420 neue Wörter in den „Wortschatz“ des Modells aufgenommen worden. Daraus folgt, dass das Modell mit den Twitterdaten eine höhere Chance hat, die Wörter zu kennen, die in den Testdialogen enthalten sind. Beide Modelle können auf offen gestellte Fragen nicht mit einer inhaltlich ausführlichen Aussage antworten. Diese Beobachtung war jedoch nicht nur bei den Fragen zum Thema Abtreibung zu erkennen, sondern auch bei Fragen die unabhängig von diesem Thema sind. Diese Problematik ist also eine generelle Einschränkung des Modells und ist nicht auf diese Fragestellung beschränkt. Eine Erklärung könnte sein, dass diese Fragetypen nur sehr wenig in den Trainingsdaten vorkamen. Oder dass diese Fragetypen oft eine lange Antwort haben, die in dieser Architektur nicht zugelassen wird. Interessant

wäre hierzu einen Datensatz zu verwenden, der sich genau aus solchen Fragen und Antworten zusammensetzt. Eine weitere Auffälligkeit ist, dass das Modell, das die Twitterdaten verwendete zwar oft eine pro-/contra-Abtreibungsantwort gibt, insgesamt aber keine eindeutige Tendenz erkennbar ist. Das Modell subsummiert also nicht nur eine vorherrschende Meinung und reproduziert nur diese. Die Antwort des Modells hängt also viel mehr von der genauen Wortwahl und auch Zeichensetzung ab. Getestet wurden nicht nur Fragen sondern auch Aussagen zum Thema Abtreibung. Dabei hat sich gezeigt, dass das Referenzmodell, wenn überhaupt, eher zustimmend antwortet. Das zweite Modell antwortet eher verneinend oder anzweifelnd. Diese Beobachtung könnte auf die Dialogstruktur und dessen Konzept auf Twitter zurückzuführen sein. Ebenso interessant ist die Feststellung, dass die Bezeichnung für das was abgetrieben wird, einen Einfluss auf die Antworten des Modells hat.

In einem nächsten Schritt wäre es interessant zu untersuchen, welchen Einfluss eine größere Menge an Twitterdaten während des Trainings spielt. In dieser Untersuchung beträgt der Anteil an Konversationen, die durch die Twitterdaten hinzu kamen lediglich 0.2% der gesamten Dialoge. Ein höherer Anteil an themenbezogenen Dialogen könnte dazu führen, dass noch mehr Fragen beantwortet werden können. In einem kleinen Test wurden alle gesammelten Twitterdialoge zehn mal zu unserem Datensatz hinzugefügt um so den Anteil an themenbezogenen Daten zu erhöhen. Leider konnten aus den Dialogen keine positiven Ergebnisse gezogen werden. Die Antworten waren oft nicht passend oder thematisch abweichend. Das legt nahe, dass es sinnvoller ist, mehr diverse Twitterdaten zu sammeln und somit den Anteil an themenbezogenen Dialogen zu erhöhen. Dadurch könnte auch sichergestellt werden, dass mehr verschiedenere Aussagen und Meinungen abgedeckt werden können. Ein weiterer Ansatzpunkt der interessant wäre zu untersuchen ist, den Referenzdatensatz zu verändern. Das Projekt welches als Grundlage für diese Untersuchung dient, enthält mehrere Trainigsdatensätze die untersucht werden können. Teilweise sind diese auch sehr viel größer, was dazu führen könnte, dass die Dialoge realistischer werden könnten. Außerdem könnte weiter untersucht werden, welchen Einfluss die Trainingszeit auf die Antworten des Modells haben. Für beide Ansätze (Trainingsdaten und Trianingszeit) ist eine leistungsfähige Hardware notwendig. Die verwendeten Twitterbibliotheken sind schnell an ihre Grenzen gestoßen. Damit konnte nur eine sehr geringe Datenmenge in kurzer Zeit gesammelt werden. Um einen großeren Datensatz generieren zu können, ist es natürlich wichtig, dafür nur eine gewisse Zeit zu benötigen, wenn man beispielsweise die Aktualität der Daten gewährleisten will. Als Alternative könnten eventuell andere Social Media Plattformen dienen, wie zum Beispiel Facebook oder Instagram. Der Vortreil von Twitter ist allerdings, dass dort die erlaubte Anzahl an Wörtern in einem Tweet geringer ist, was für die verwendete Architektur ausschlaggebend ist. Eine alternative Datenquelle müsste also ebenso darauf ausgelegt sein, dass die Nachrichten nur eine gewisse Anzahl an Wörtern zulässt und eine Schnittstelle anbieten mit der man viele Daten schnell sammeln kann.

Wenn ein Modell trainiert wurde und eine Konversation zu einem bestimmten Thema vielversprechend und realistisch wirkt, könnte ein nächster Evaluierungsschritt folgen. Darin könnte untersucht werden in wie weit die gefundenen Aussagen ähnlich sind zu denen, die in einer traditionell durchgeführten Diskursanalyse gefunden wurden. Dazu müsste eine Diskursanalyse zu einem Thema gefunden werden, die einige Voraussetzungen erfüllt, damit ein Modell wie bisher beschrieben damit verglichen werden kann. Beispielsweise müsste der Datenkorpus leicht zugänglich sein. Außerdem kann das hier untersuchte Modell nur mit Dialogen, die aus Sätzen mit einer geringen Anzahl an Wörtern, arbeiten. Eine Diskursanalyse die sich auf Zeitungsartikel bezieht, wäre also keine gute Grundlage. Eine direkte Evaluation mit Grundlage einer bestehenden Diskursanalyse wäre jedoch sehr interessant und zudem auch wichtig um zu überprüfen welche Positionen die gefundenen Aussagen und Meinungen in einem zu untersuchenden Diskurs einnehmen. Wichtig wäre dabei auch, zu reflektieren, welche Fragen dem trainierten Modell gestellt werden. Dazu könnte aber eine einfache Anwendung, beispielsweise eine Webanwendung entwickelt werden, die es ermöglicht, dass mehrere Personen das trainierte Modell untersuchen.

Fazit

In diesem Projekt wurde untersucht in wie weit ein neuronales Netz eingesetzt werden kann, um eine Art automatisierter Assistent für eine Diskursanalyse zu entwickeln. Dazu wurden Trainingsdaten zum Thema Abtreibung bei Twitter gesammelt. Auf Grundlage von diesen Daten und einem weiteren großen Datensatz trainierte das Netz. Daran anschließend wurde untersucht, welche Inhalte das trainierte Modell aus den Twitterdaten „lernte“. Dabei konnte gezeigt werden, dass das Modell mehr Fragen mit einer eigenen Position als das Referenzmodell beantworten kann. Das Modell nimmt also Inhalte und Aussagen aus dem Diskurs auf und reproduziert diese in manuell geführten Dialogen. Dieser Ansatz ist ein vielversprechender Schritt um bei einer Diskursanalyse mit einem automatisierten Ansatz zu unterstützen. Einige spannende Ansatzpunkte sind durch dieses Projekt entstanden, die weiter untersucht werden sollten.

Dokumentation während der Bearbeitung:

Mögliche Chatbot-Implementierungen:

- Übersicht
- Tensorflow
- Torch

Mögliche Hashtags:

- lebensschuetzer
- StopAbortion
- proLife

Offene Fragen:

- Technische Umsetzung des Chatbots, hinzufügen eines neuen Themas
- Auswahl der Hashtags
- Sprache der Tweets
- Sicherstellen, dass die Tweets zum Diskurs gehören

Nächste Schritte:

- Evaluierung der Repositories
- Recherche zum Hinzufügen eines neuen Themenfeldes
- Strategie zur Datenerhebung

Aktueller Stand: 02.06.

- Chatbot Implementierung: Tensorflow wurde erfolgreich evaluiert und erste Dialoge geführt

Hier ein Beispiel für einen Dialog:

- Implementierung zur Tweetsammlung plus Dialoge unter den Tweets unter TwitterChatbot zu finden

Nächste Schritte

- Fest stellen ob deutsche Konversationen möglich sind
- Weitere Datensammlung
- Gesammelte Daten in richtiges Datenformat bringen
- Chatbot auf die Twitterdaten lernen lassen

Offene Fragen

- Hardware? Aktuell besteht nur die Möglichkeit auf einer CPU das Training laufen zu lassen –> dauert viel zu lange
- Twitterdaten, API liefert unvollständige, zeitlich begrenzt und wenig Daten, eventuell andere Datenquelle?
- Trainieren eines pretrained Modells derzeit schwierig, Tensorgröße unterschiedlich
    - dafür deutlich größeres Datenset
    - bessere Hardware

Nachdem die verwendete Tensorflow-GPU Version geupdated wurde konnte das Modell auf einer besseren Hardware (Tesla K80 GPU) ausgeführt werden, was ein Training mit größeren Datensätzen ermöglicht. Für das Training wurde die Plattform von Google Colab verwendet.

Um das Problem der unterschiedlichen Tensorgrößen bei verschiedenen Modellen zu lösen wurde der Trainingsansatz leicht abgewandelt. Ziel es nun ein Modell von Beginn an mit einem Datensatz zu trainieren, dass die Twitterdaten nicht enthält und ein Modell zu trainieren, dass die Twitterdaten enthält. Ein vortrainiertes Modell mit Daten zu trainieren, die Wörter enthalten die bisher noch nicht gesehen wurden funktioniert bei dieser Implementierungen nicht. Daher wurde der Ansatz verfolgt, zwei Modelle komplett neu zu trainieren. Für das Modell, dass mit Daten lernt, die keine Twitterdaten enthalten wurde ein bereits bestehendes Datenset verwendet, welches sich aus Filmdialogen zusammensetzt. Um das Datenset der Twitterdaten und der Filmdaten zusammen zuführen wurde ein weiteres python-Skript entwickelt.

Literatur:

- Ansatz mit R
- Kombination aus Grounded Theory und Machine Learning
- Content Mining with Python
- Nina Degele, Dominique Schirmer, *Selbstverständlich heteronormativ: zum Problem der Reifizierung in der Geschlechterforschung* in Gender methodologisch Empirische Forschung in der Informationsgesellschaft vor neuen Herausforderungen. VS Verlag für Sozialwissenschaften/GWV Fachverlage GmbH, Wiesbaden 2004

- Thilo Hagendorff, *Rassistische Maschinen? Übertragungsprozesse von Wertorientierungen zwischen Gesellschaft und Technik* in Maschinenethik. Springer Fachmedien Wiesbaden, 2019
- Dominique Schirmer, *Empirische Methoden der Sozialforschung: Grundlagen und Techniken* UTB GmbH, 2009
- Oriol Vinyals and Quoc V. Le , *A Neural Conversational Model* in CoRR, 2015 http://arxiv.org/abs/1506.05869
- Ilya Sutskever, Oriol Vinyals, Quoc V. Le, *Sequence to Sequence Learning with Neural Networks* in Advances in Neural Information Processing Systems 27 (NIPS 2014) http://papers.nips.cc/paper/5346-sequence-to-sequence-learning-with-neural-networks.pdf

From:
https://institut.soziologie.uni-freiburg.de/dokuwiki/ - **Institut für Soziologie - Lehrwiki**

Permanent link:
https://institut.soziologie.uni-freiburg.de/dokuwiki/doku.php?id=lv-wikis-oeffentlich:pda19:chatbot_automatisierte_inhaltsanalyse

Last update: **2021/07/27 08:59**