

ACCOUNTABILITY IN ALGORITHMIC DECISION MAKING

Vural Mert, Larcher Daniel

1. Juni 2016

Zusammenfassung

Diese Seminararbeit gibt einen kurzen Überblick über die "Algorithmische Entscheidungsfindung", deren Nutzen, über den "Transparency Standard" und die Zukunftsaussichten in "Accountability in Algorithmic Decision Making".

1 Einleitung

Im Zuge der Lehrveranstaltung "Einführung in das wissenschaftliche Arbeiten", an der Universität Innsbruck, haben wir uns mit dem Thema "Accountability in Algorithmic Decision Making" beschäftigt. Am Anfang gibt es eine Einführung in das Thema "Accountability in Algorithmic Decision Making", im nächsten Punkt wird erläutert wozu es überhaupt dient und warum es angewendet wird. Im Anschluss wird der Punkt "Transparency Standard" mit in die Arbeit einbezogen, wobei die 5 Punkte des Standards erklärt werden. Im letzten Abschnitt gibt einen Einblick in die Zukunft.

2 Was ist "Accountability in Algorithmic Decision Making"?

Wie es der deutsche Name schon sagt, wird die "Algorithmische Entscheidungsfindung" im Bereich Algorithmen angewendet, um Entscheidungen zu treffen. Dabei handelt es sich um Entscheidungen, die ohne menschlichen Einfluss, zur Gänze von Algorithmen getroffen werden [1].

3 Wozu "Accountability in Algorithmic Decision Making"?

In der Welt der Informationstechnologie sind Algorithmen nicht wegzudenken. Da heutzutage in allen Bereichen des Alltags elektronische Geräte zur Verwendung kommen, trifft man im Alltag oft auf den Einsatz von Algorithmen.

Die wichtigsten Beispiele hierfür sind Fernseher, Mobiltelefone, Radiosender, Navigationsgeräte und viele andere elektronische Geräte. Sogar in unseren Küchengeräten wie Kaffeemaschine oder Geschirrspüler kommen mittlerweile Algorithmen zum Einsatz. Dies sind alles einfache Beispiele zur Verwendung von Algorithmen, die uns unser Alltagsleben erleichtern.[2]

Auf der anderen Seite gibt es das Gebiet, in dem Algorithmen zum Einsatz kommen, um Entscheidungen von Menschen im Alltag und Arbeitsleben bewusst zu beeinflussen.

Das beste Beispiel dafür ist Facebook. Auf Facebook werden Nachrichten und Postings so sortiert, organisiert und dem User dargestellt, sodass sie die Gedanken und Ideen der Menschen sehr stark beeinflussen. Damit könnte sogar erreicht werden, dass die politische Einstellung der Menschen bewusst in eine Richtung gelenkt werden kann. Systeme, die das Prinzip der Algorithmischen Entscheidungsfindung verwenden, nennt man Nachrichten-Verteilungssysteme [1].

4 Transparency Standard

Der Transparency Standard befasst sich damit, welche Informationen man von einem Algorithmus veröffentlichen kann und sollte. Zu diesem Thema hat Assistenzprofessor Nicholas Diakopoulos einen Workshop abgehalten, in dem er mit 50 Leuten Fallstudien zu 'Automatically Generated News Content', 'Simulation, Prediction and Modelling in Storytelling' und 'Algorithmically Enhanced Curation' diskutierte und anhand dessen 5 Punkte aufstellte, welche Informationen veröffentlicht werden sollten.

4.1 Die 5 Punkte des Transparency Standards

4.1.1 Human Involvement

Als erster Punkt auf der Liste steht das 'Human Involvement' mit dem man eine größere Verantwortlichkeit gegenüber der Öffentlichkeit bezwecken möchte.

Human Involvement befasst sich auf der höheren Ebene der Transparency mit der Motivation und dem Zweck des Algorithmus, aber auch mit seinen Entwicklern. Wer den Algorithmus entwickelt und wer ihn absegnet sollte offengelegt werden. Diakopoulos gesteht aber ein, dass es sich im Alltag als sehr schwierig erweist, diese Zuteilung aufzustellen.

4.1.2 Data

Der nächste Punkt auf der Liste betrifft die Daten. Hier gibt es viele verschiedene Zweige der Transparency. Einer davon konzentriert sich auf die Messung der Qualität der benutzten Daten. Hierbei sollten die Daten auf ihre Genauigkeit und Vollständigkeit sowie etwaige Fehler untersucht werden. Auch ihre Zeitlosigkeit sollte in Betracht gezogen werden, da sich die Gültigkeit der Daten im Laufe der Zeit ändern könnte. Ein weiterer, nicht zu unterschätzender Punkt in dieser Untersuchung ist der repräsentative Charakter der Daten, da Algorithmen auf eine bestimmte Personengruppe beschränkt sein können. Die Daten müssen dann auf diese Zielgruppe zugeschnitten sein, damit der Algorithmus korrekt arbeiten kann.

Ein anderer Zweig beschäftigt sich damit, wie die Daten gesammelt, transformiert, bearbeitet (hierbei muss man ebenfalls berücksichtigen, ob dies automatisiert oder durch Menschen geschah) und untersucht wurden.

4.1.3 Modell

Das Modell kann ebenfalls bis zu einem gewissen Punkt veröffentlicht werden. Hierbei sollte man den Input (Features und Variablen des Algorithmus einschließlich der Gewichtung dieser) und die Daten (falls machine-learning verwendet wurde) untersuchen, wie auch die Software-Modeling-Tools, da diese verschiedene Limitierungen oder Voraussetzungen haben.

Bei der Untersuchung des Modells fließt auch der erste Punkt (Human Involvement) wieder mit ein, da sich hier mehrere Fragen auftun, wie, zum Beispiel, die Gewichtung der Features entschieden wurde oder welcher Design-Prozess für alternative Modelle und für die Modell-Vergleichung gewählt wurde.

4.1.4 Inferences

Als vorletzten Punkt auf der Liste betrachtet man die Inferences des Algorithmus, da seine Klassifizierungen und Vorhersagen oft nicht direkt nachvollziehbar sind und Fragen bezüglich seiner Genauigkeit und seines Fehlerpotentials offen lassen.

Bei den Inferences untersucht man, wie viel Spielraum für Fehler übrig bleibt und ob ein Fehler durch menschlichen Einfluss, den Daten-Input oder durch den Algorithmus selbst verursacht wurde.

4.1.5 Algorithmic presence

Der letzte Punkt beschäftigt sich mit der Präsenz des Algorithmus. Hierbei betrachtet man, zu welchem Zeitpunkt ein Algorithmus eingesetzt wird und ob beispielsweise A/B-Testing (Testmethode, bei der das Originalsystem gegen eine modifizierte Variante getestet wird [3]) verwendet wurde. Bei der Betrachtung der Präsenz des Algorithmus muss man beachten, dass technische Systeme einem ständigen Wandel unterliegen und sich die Dynamik des Algorithmus aufgrund ständigem maschinellen Lernens verändern kann. [1]

5 Zukunft

Ein Fortschritt für die Zukunft wäre, den momentanen Stand von maschinellen Lernprozessen in einer Weise zu verbessern und weiterzuentwickeln, so dass sie nicht nur für Entwickler sondern auch für den allgemeinen User verständlich sind. Ein Ansatz hierfür wäre die Verwendung der Bayesian Rule List Technik, bei der verschiedene, für Menschen lesbare, Regeln angewendet werden, welche bei der Verkettung eine leichter verständliche Erklärung der Klassifizierer ergeben.

Andere Ansätze werden in natural language generation (NLG) entwickelt, sodass man einen Output erzeugen kann, welcher erklärt, wie der Algorithmus zu seiner Entscheidung kam. [1]

6 Schlussfolgerung

Diese Arbeit beschäftigt sich mit der Zuständigkeit bei Algorithmischer Entscheidungsfindung und sollte kurz zusammenfassen worum es hierbei geht. Hierzu wurde der Transparency Standard, welcher von Nicholas Diokopoulos entwickelt wurde, erklärt. Abschließend wurden zukünftige Entwicklungen erläutert.

Literatur

- [1] Nicholas Diakopoulos, *Accountability in Algorithmic Decision Making*, Communications of the ACM (February 2016, Vol.59, No.2)

- [2] Algorithmen - welt.de,
<http://www.welt.de/print/wams/vermischtes/article13663091/Die-wunderbare-Welt-der-Algorithmen.html>
[Online; aufgerufen am 25.Mai 2016]

- [3] A/B-Test - Wikipedia,
<https://de.wikipedia.org/w/index.php?title=A/B-Test&oldid=152730476>
[Online; aufgerufen am 26.Mai 2016]