

Studie: Maschinelern-Verfahren können Psychosen erkennen

Per Inhaltsanalyse der gesprochenen Worte lassen sich offensichtlich Anzeichen dafür finden, dass eine Person mit erhöhter Wahrscheinlichkeit zu einem späteren Zeitpunkt eine Psychose entwickelt.

Angeblich erzielten Wissenschaftler mit Hilfe von Deep-Learning-Algorithmen eine Trefferwahrscheinlichkeit von 93 Prozent. Ein entsprechendes Paper von Neguine Rezaii, Elaine Walker und Phillip Wolff mit dem Titel „[A machine learning approach to predicting psychosis using semantic density and latent content analysis](#)“ ist offen zugänglich. Wenn Algorithmen Auffälligkeiten finden, die menschlichen Beobachtern verborgen blieben, ist das immer einen Blick wert.

Der Ansatz ist, die verwendeten Worte in der alltäglichen Sprache der Probanden mit automatisierten Inhaltsanalysemethoden zu untersuchen. Dabei zeigte sich ein Zusammenhang zwischen zwei Variablen und der Wahrscheinlichkeit, dass die Person schizophrene Symptome zeigt. Die eine Variable ist die Häufigkeit, mit der Worte verwendet werden, die mit Klängen, Tönen, auditiven Signalen in Zusammenhang stehen. Die andere Variable ist, dass die Aussagen der Person vage sind, die Sätze semantisch wenig dicht sind, wie es heißt.

Dass sich psychotische Tendenzen einer Person in ihrer Sprache niederschlagen, ist seit längerem bekannt. Mentale Zustände wie Ärger und Wut lassen sich in der Stimme erkennen. Allerdings sind offenbar einige linguistische Marker bisher

verborgen geblieben. Auch geübten Diagnostikern, die mit psychotischen Patienten zu tun haben, fiel der Zusammenhang in der verwendeten Sprache bisher nicht auf.

Als Material verwendeten die Forscher allgemeine Kommunikation in einer Social Media – Plattform (Reddit.com). Mit in Anwendung implementierten Word2Vec-Modellen wurden große Mengen an Text vektorisiert, so dass mit diesem Dataming-Prozess ein Korpus an Charakteristiken aus „normaler“ Sprache entstand.

Damit konnte man dann Unterschiede zu verwendeter Sprache finden, wie sie in diagnostischen Interviews von Personen verwendet wurden, die als Schizophrenie-Risikopersonen gelten. Diese transkribierten Diagnosegespräche entstanden im Rahmen einer breiter angelegten Studie in den USA, der *North American Prodrome Longitudinal Study (NAPLS)*. In dieser Studie hat man es mit Probanden zu tun, die auf Schizophrenie hin untersucht worden sind. Diese Texte wurden ebenfalls mit Word2Vec analysiert.

Diese Studie zeigt, dass es latente Auffälligkeiten in Sprachmuster und Sprachgebrauch gibt, die selbst erfahrenen Diagnostikern bislang nicht aufgefallen sind. Angeblich, so das Ergebnis dieser Studie, nutzen Betroffene, die an Schizophrenie leiden oder von Psychosen geplagt werden, mehr als der Durchschnitt Begriffe, die sich um Geräusche oder Laute drehen. Zudem ist ihre Sprache weniger dicht an Fakten. Die Forscher gaben Trefferwahrscheinlichkeiten jenseits von 90 Prozent an, was erstaunlich hoch wirkt. Daher sei es naheliegend, dass man mit der „digital phenotyping“ genannten Methodik mentale Krankheiten erkennen kann, auch wenn diese als solche bei einem Individuum noch nicht diagnostiziert sind.

Allerdings finde ich zwei Dinge beachtenswert beziehungsweise kritisch: Die Studie vergleicht „normale Sprache“ mit „besonderer Sprache“. Zum einen nutzt man geschriebenes Wort

aus einem sozialen Netzwerk, zum anderen, um Kontraste zu finden, verbale Sprache, die in schriftliche Sprache umgesetzt worden ist. Dass in der „besonderen Sprache“ vermehrt Begriffe auftauchen, die mit Lauten und Geräuschen zu tun haben, könnte auch damit zusammenhängen, dass die Diagnosegespräche waren. Es ist gut möglich – zumindest sollte man das untersuchen -, dass es mehr als die „normale Sprache“ aus der sozialen Plattform um ein Artefakt handelt, da die diagnostischen Interviews strukturiert sind, und zwar gemäß der Anleitung „Structured Interview for Prodromal Syndromes“ ([PDF](#)). Das heißt: Der Diagnostiker, der das Interview typischerweise führt, fragt gezielt nach Auffälligkeiten in der Wahrnehmung, was das Hören mit einschließt. Schließlich möchte er ja kennzeichnende Auffälligkeiten finden, um einen Patienten zu diagnostizieren.

Die Autoren thematisieren neben der auffällig häufigen Verwendung von Begriffen rund im Geräusche oder Töne auch eine gewisse Armut im Ausdruck, bezeichnet als vage Aussagen, was typisch für die Gruppe der von Psychosen oder schizophrenen Symptomen Betroffenen sei. Ich denke, hier müsste man diskutieren, inwieweit die Art der Datenerhebung entsprechende Effekte generieren kann: Die strukturierten Interviews werden zu Diagnosezwecken mündlich geführt. Schriftlich drückt sich der Durchschnittsmensch womöglich faktenreicher und gezielter aus als mündlich. Schriftlicher Text ist anders strukturiert als transskribierte Sprache.

Womöglich sind die Probanden zurückhaltender, da ihnen die Testsituation bewusst ist, was bei den Usern auf Reddit nicht der Fall war. Die wussten zwar, dass sie sich gewissermaßen öffentlich äußerten, schrieben ihre Texte aber nicht im Bewusstsein, damit auf Psychosen getestet zu werden.

Der zweite Punkt ist, dass hier eigentlich eine klassische Unterschiedshypothese getestet wird, ohne dass eine Theorie vorgelagert wurde, die erklärt, weswegen von Schizophrenie betroffene Patienten in ihrer Sprache vermehrt

Begrifflichkeiten rund um Hören, Töne, Geräusche verwenden. Die Fallzahl liegt wohl bei 30 bis 40 Probanden, was für Unterschiedshypothesentests wiederum gering erscheint. Zehn Probanden waren als Kontrollgruppe definiert. Der Maschinenlernansatz bezieht sich auf die Anwendung eines neuronalen Netzes zur Klassifizierung von Sprache beziehungsweise Text.