

Nachrichten aus dem Datenraum

Wie hältst Du's mit Technologien?

Die Welt produziert ständig neue Daten, die wiederum etwas über die Welt aussagen. Können wir diese Daten nutzen, um die technologieorientierte Zukunftsforschung besser zu machen? Genau das ist die zentrale Frage und Motivation von Data Driven Foresight. Bislang haben wir uns in dieser Newsletter-Kategorie vor allem mit der Analyse strukturierter Daten, speziell Publikationsdaten, befasst und gezeigt, welche Schlüsse man aus ihnen ziehen kann. In diesem Beitrag nun soll es aber um die Analyse von Textdaten, also von unstrukturierten Daten gehen. Im Rahmen seiner Masterarbeit, die am Lehrstuhl für Technologieanalysen und -vorausschau in der Sicherheitsforschung an der RWTH Aachen University sowie dem Fraunhofer INT entstand, hat sich David Blasko damit befasst, wie man Twitter-Nachrichten, sogenannte Tweets, für die Technologiefrühaufklärung nutzbar machen kann.

Diese Kurznachrichten werden nicht nur für Ankündigungen, sondern oft auch für Kommentare zu einer Vielzahl von Themen genutzt. Dazu zählen neben politischen und gesellschaftlichen Themen durchaus auch Technologiethemata. In diesen Nachrichten kommt in vielen Fällen zum Ausdruck, wie der Absender zu einem Thema eingestellt ist, z. B. ob er ihm positiv oder negativ gegenübersteht. Prinzipiell ist es möglich, die Stimmung eines Textes mit Hilfe der sogenannten Sentimentanalyse automatisiert zu bestimmen. Dafür stehen verschiedene Verfahren aus dem Bereich der Computerlinguistik zur Verfügung, mit denen bestimmt werden kann, ob ein Text negativ, neutral oder positiv gestimmt ist.

In seiner Masterarbeit hat David Blasko ein solches Verfahren angewandt und unter anderem auf Tweets zum Thema Künstliche Intelligenz angewandt. Dazu wurden zunächst jene Tweets herausgesucht, die den Begriff im Text oder als sogenanntes Hashtag enthalten. Nach einer umfangreichen Vorverarbeitung der Daten wurde schließlich die Stimmung der Tweets bestimmt. Auf diese Weise entstand ein dynamisches Stimmungsbild zum Thema Künstliche Intelligenz, welches interessante Einblicke in die Wahrnehmung dieses Themas bietet.

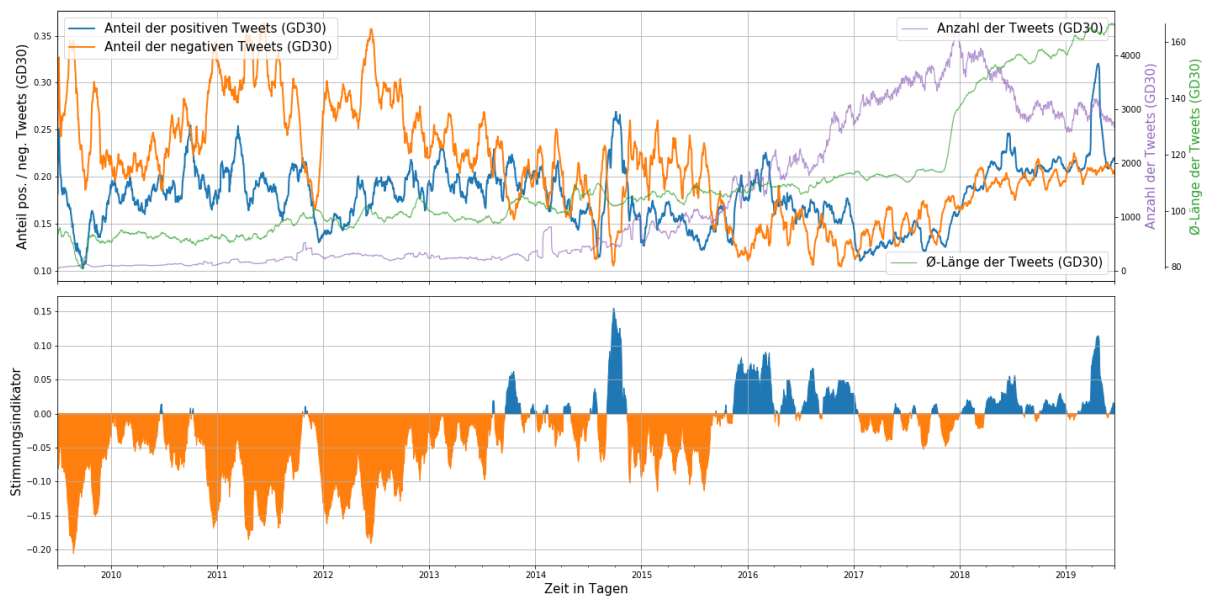


Abbildung 1: Twitter-basiertes Stimmungsbarometer zum Thema Künstliche Intelligenz. Im oberen Teil der Abbildung sind über einen Zeitraum von 30 Tagen gemittelt (GD30) die absolute Zahl der Tweets (violette Linie), deren mittlere Länge (grüne Linie), sowie der Anteil der positiven (blaue Linie) und negativen (orangene Linie) Tweets. In der unteren Abbildung ist ein Stimmungsindikator dargestellt, der anzeigt, wie sehr negative oder positive Stimmungen überwiegen.

Der obere Teil der Abbildung enthält Detailinformationen zu vier Parametern: den Anteil der positiven und negativen Tweets, die absolute Anzahl der Tweets zum Thema Künstliche Intelligenz und ihre durchschnittliche Länge (im November 2017 verdoppelte Twitter die Anzahl der zulässigen Zeichen pro Tweet von 140 auf 280). Dabei wurden die Daten etwas geglättet, um kurzzeitige Schwankungen auszugleichen.

Der untere Teil der Abbildung gibt so etwas wie die Nettostimmung zum Thema wieder und beantwortet damit die Frage, ob die Twittergemeinde der Technologie Künstliche Intelligenz gegenüber eher negativ oder positiv eingestellt ist. Hier kann man die klare Tendenz erkennen, dass diese Nettostimmung am Anfang des Betrachtungszeitraums (2009) eher negativ war. Etwa Mitte 2013 scheint ein Umdenken einzusetzen und positive Einschätzungen beginnen zu überwiegen – immer mal wieder unterbrochen durch eher negativ gestimmte Phasen, welche aber zunehmend weniger werden.

Solche Analysen stellen eine wichtige Ergänzung für die Technologieführaufklärung dar, insbesondere wenn es um die Bewertung einer Technologie geht. In diese fließen nämlich nicht nur technologische, sondern auch ökonomische, soziale sowie politische Aspekte ein. Diese lassen sich wiederum nur begrenzt aus klassischen wissenschaftlichen Publikationen ableiten, sodass die Betrachtung sozialer Medien wie Twitter hier eine sinnvolle Ergänzung darstellt.