

FORSCHUNG KOMPAKT

FORSCHUNG KOMPAKT

1. Februar 2019 || Seite 1 | 3

Machine Learning: Social Media-Daten auswerten

Software für die automatisierte Erkennung von Fake News

Erfundene Nachrichten, verdrehte Fakten – Fake News verbreiten sich rasant im Netz und werden oft unbedacht geteilt, vor allem in den Sozialen Medien. Fraunhofer-Forscherinnen und -Forscher haben ein System entwickelt, das Social Media-Daten automatisiert auswertet und bewusst gestreute Falschmeldungen und Desinformationen gezielt herausfiltert. Das Tool nutzt dafür sowohl inhaltliche als auch Metadaten, wobei es die Klassifikation mit Machine Learning erworben hat und diese in Interaktion mit dem Nutzer während der Anwendung verbessert.

Falschmeldungen werden zur Stimmungsmache oder Hetze gegen einzelne oder mehrere Personen genutzt. Sie sollen die öffentliche Meinung zu bestimmten aktuellen Themen beeinflussen und manipulieren. Diese Fake News verbreiten sich rasant über das Internet, vor allem über Soziale Medien wie Facebook und Twitter. Sie zu identifizieren ist schwierig. Hier setzt ein Klassifikationstool des Fraunhofer-Instituts für Kommunikation, Informationsverarbeitung und Ergonomie FKIE an. Es wertet Informationen aus Social Media-Beiträgen automatisiert aus. Das System erschließt große Datenmengen. Es bewertet nicht nur Texte, sondern bezieht auch Metadaten in die Analyse ein und bereitet die Ergebnisse grafisch auf. »Mit unserer Software fokussieren wir uns auf Twitter und Webseiten. In den Tweets werden die Links veröffentlicht, unter denen die eigentlichen Fake News zu finden sind. Die sozialen Medien liefern sozusagen den Trigger. Die eigentlichen Falschmeldungen finden sich häufig auf Webseiten, die denen von Nachrichtenagenturen nachempfunden und nur schwer von den Originalen zu unterscheiden sind. Oftmals liegen ihnen DPA-Meldungen zugrunde, die sprachlich verändert wurden«, erläutert Prof. Dr. Ulrich Schade, Wissenschaftler am Fraunhofer FKIE, dessen Forschungsgruppe das Tool entwickelt hat.

Im ersten Schritt bauen Schade und sein Team Bibliotheken mit seriösen Beispielbeiträgen auf sowie mit solchen Texten, die der Nutzer als Fake News klassifiziert. Mithilfe dieser Lernsets wird das System trainiert. Um Falschmeldungen herauszufiltern, wenden die Forscherinnen und Forscher »Machine Learning«-Verfahren an, die automatisiert nach bestimmten Merkmalen in den Texten und den Metadaten suchen. Das können beispielsweise in einem politischen Kontext auf semantischer Ebene Formulierungen und Wortkombinationen sein, die sich weder im alltäglichen Sprachgebrauch noch in der journalistischen Berichterstattung finden wie »die aktuelle Bundeskanzlerin«. Zu

Kontakt

Janis Eitner | Fraunhofer-Gesellschaft, München | Kommunikation | Telefon +49 89 1205-1333 | presse@zv.fraunhofer.de

Silke Wiesemann | Fraunhofer-Institut für Kommunikation, Informationsverarbeitung und Ergonomie FKIE | +49 228 9435-103 | silke.wiesemann@fkie.fraunhofer.de
Fraunhoferstraße 20 | 53343 Wachtberg | www.fkie.fraunhofer.de

den Merkmalen zählen auch sprachliche Fehler. Dies ist insbesondere dann der Fall, wenn der Autor für die Formulierung der Fake News Deutsch und nicht seine eigene Muttersprache verwendet. Dann deuten etwa falsche Gedankenstriche, Orthografie-, Deklinations- oder Satzbaufehler darauf hin, dass eine Meldung eine Fake News sein könnte. Unangemessene Ausdrücke oder umständliche Formulierungen können ein weiteres Indiz sein.

»Füttern wir unser Tool mit vielen Merkmalen, so sucht sich das Verfahren selbstlernend diejenigen heraus, die funktionieren. Entscheidend ist auch, welche »Machine Learning«-Verfahren dann die besten Ergebnisse liefern. Das ist sehr aufwändig, da man die diversen Algorithmen mit unterschiedlichen Kombinationen von Merkmalen durchrechnen lassen muss«, so Schade.

Metadaten liefern entscheidende Hinweise

Zu den Merkmalen zählen auch Metadaten. Diese spielen eine wichtige Rolle, wenn es darum geht, richtige von falschen Meldungen zu unterscheiden: Wie häufig wird gepostet, wann wird ein Tweet abgesetzt und um welche Uhrzeit. Aufschlussreich ist der Zeitpunkt eines Posts. Er kann darauf hinweisen, aus welchem Land und welcher Zeitzone der Sender Meldungen absetzt. Eine hohe Sendefrequenz deutet auf Bots hin, was die Wahrscheinlichkeit einer Fake News erhöht. Die Social Bots senden ihre Links an sehr viele Nutzer, um etwa Unsicherheit in der Bevölkerung zu verbreiten. Auch die Vernetzung der Accounts und Follower kann für Analysten von großer Bedeutung sein.

Die Sendedaten und deren Anzahl aber auch die Netze der Follower lassen sich in Form von Heatmaps und Graphen visualisieren. Aus der Struktur der Netze und ihrer Knoten lässt sich beispielsweise ablesen, welcher Knoten eine Fake News in Umlauf gebracht oder eine Fake News-Kampagne initiiert hat.

Auch Hate Speech lässt sich automatisiert erkennen. Beiträge, die sich als Nachrichten ausgeben, aber in Passagen Hate Speech verwenden, verweisen gern über Links auf Fake News. »Hier ist es dabei wichtig, einen Klassifikator zu entwickeln, der die eindeutigen Fälle identifiziert. Dies sind beispielsweise Ausdrücke wie »Politischer Abschaum« oder »Nigger«, sagt der Linguist und Mathematiker.

Die Wissenschaftlerinnen und Wissenschaftler können ihr System an unterschiedliche Arten von Texten anpassen, um diese zu klassifizieren. Sowohl Behörden als auch Unternehmen nutzen das Tool, um gezielt Desinformation aufzudecken und umfassend zu bekämpfen. »Unsere Software lässt sich für jeden Kunden individuell anpassen und trainieren. Behörden kann es als Frühwarnsystem dienen«, sagt Schade.



Abb. 1 Um Fake News zu erkennen, bewertet das Tool des Fraunhofer FKIE nicht nur Texte, sondern bezieht auch Metadaten in die Analyse ein.

© Fraunhofer FKIE

FORSCHUNG KOMPAKT
1. Februar 2019 || Seite 3 | 3