

Verborgene Muster

10.8.2026 - | Leibniz-Gemeinschaft

KI-Methoden zur Textanalyse eignen sich auch für die Analyse von Umweltproben. Das zeigt eine Studie am Beispiel Ostsee.

Mikrobielle Gemeinschaften reagieren sensibel auf Umweltveränderungen, doch ihre enorme Vielfalt macht ökologische Muster schwer erkennbar. Ein Forschungsteam des Leibniz-Instituts für Ostseeforschung Warnemünde (IOW) zeigt nun, dass sich eine ursprünglich für Textanalysen entwickelte KI-Methode erfolgreich zur Untersuchung komplexer Umweltproben einsetzen lässt. Das Verfahren identifizierte in der Warnow-Mündung fünf jahreszeitlich aufeinanderfolgende mikrobielle Teilgemeinschaften, bewahrte dabei ökologische und funktionelle Informationen ähnlich gut wie klassische Verfahren und war diesen teilweise sogar überlegen. Die Studie erschien jetzt im Fachjournal „Environmental Microbiome“.

Mikroorganismen sind nahezu überall zu finden und übernehmen zentrale Aufgaben in Ökosystemen: Sie setzen organisches Material um, beeinflussen Stoffkreisläufe und reagieren häufig sehr schnell auf veränderte Umweltbedingungen. Insbesondere in der aquatischen Ökologie kann die Zusammensetzung mikrobieller Gemeinschaften daher wertvolle Hinweise auf den Zustand eines Gewässers geben. Ihre Untersuchung erzeugt jedoch äußerst umfangreiche Datensätze. In einer einzigen Wasserprobe lassen sich oft bis zu Hunderttausende unterschiedlicher genetische Signaturen mikrobieller Lebewesen nachweisen, deren Vorkommen zudem durch Temperatur, Salzgehalt, Nährstoffe, Jahreszeit und zahlreiche weitere Faktoren beeinflusst wird.

Ein Forschungsteam um Anna Kujat, IOW-Doktorandin und Erstautorin der jetzt publizierten Studie, und den wissenschaftlichen Korrespondenzautor Theodor Sperlea untersuchte deshalb, ob sich sogenanntes Topic Modeling für die Analyse solch komplexer Daten eignet. Diese Gruppe von Verfahren des unüberwachten maschinellen Lernens wurde ursprünglich entwickelt, um in großen Textsammlungen Muster zu erkennen. Dabei werden Wörter, die häufig gemeinsam auftreten, zu Themen – auf Englisch „Topics“ – zusammengefasst, die vorher nicht erkannt werden konnten. Übertragen auf mikrobielle Lebensgemeinschaften – sogenannte Mikrobiome – werden anstelle gemeinsam auftretender Wörter Mikroorganismen gruppiert, die in den untersuchten Proben ähnliche Muster bezüglich ihres Vorkommens zeigen. „Wir haben so die Grundidee der Textanalyse auf ein ökologisches Problem übertragen, um aus einer kaum überschaubaren Biodiversität eine kleine Zahl biologisch interpretierbarer Muster zu extrahieren“, erläutert Kujat den Ansatz.

Von mehr als 1.200 Erbgutvarianten zu interpretierbaren Mustern

Als besonders anspruchsvollen Testfall wählten die Forschenden die Warnow-Mündung und angrenzende Küstengewässer der südwestlichen Ostsee. Flusseintrag, Wind und der wechselnde Einfluss von Süß- und Ostseewasser sorgen dort mitunter täglich für Veränderungen. Gleichzeitig wirken städtische Abwässer, Industrie, Tourismus und weitere menschliche Nutzungen auf das System ein. Die Datengrundlage bildeten 1.551 Wasserproben, die zwischen April 2022 und April 2023 an 14 festen Stationen im Rahmen des Projektes OTC Genomics 2* genommen worden waren und ökologisch den Gradienten von reinen Süßwasserhabitaten bis hin zu reinen Ostseehabitaten abdeckten. Neben den mikrobiellen Gemeinschaften erfasste das Team unter anderem Salzgehalt, Wassertemperatur, Chlorophyll-a- und Nährstoffkonzentrationen und nutzte Informationen zum Süßwasserzufluss der Warnow der Staatlichen Ämter für Landwirtschaft und Umwelt Mecklenburg-Vorpommern. Nach Aufarbeitung enthielt der gesamte Datensatz 1.236 unterschiedliche bakterielle

Erbgutvarianten.

Die Forschenden testeten zwei Varianten des Topic Modeling und verglichen sie mit bewährten Verfahren, die große und unübersichtliche Datensätze auf wenige zentrale Muster verdichten. Für die quantitative Bewertung prüfte das Team, wie gut die verkleinerten Datendarstellungen noch ökologische Variablen wie Salzgehalt, Chlorophyll-a-Konzentration oder Probenahmeort abbildeten. Zusätzlich wurde untersucht, inwieweit funktionelle Ähnlichkeiten zwischen den Mikroorganismen erhalten blieben. Dabei zeigte sich: Insgesamt war Topic Modeling den konventionellen Verfahren von der Analyseleistung mindestens ebenbürtig und schnitt in einzelnen Analysen sogar besser ab.

Fünf Teilgemeinschaften folgen dem Rhythmus der Jahreszeiten

Für eine vertiefende ökologische Auswertung wählte das Team ein Topic Modeling-Verfahren aus, das in den quantitativen Tests sowohl ökologische als auch potenziell funktionelle Informationen besonders gut herausfiltern konnte. Von insgesamt elf ermittelten mikrobiellen Teilgemeinschaften waren fünf für den Mündungsbereich der Warnow charakteristisch. Sie traten nicht gleichzeitig auf, sondern folgten in einer deutlichen saisonalen Abfolge aufeinander mit Schwerpunkten des Auftretens im Frühjahr, Sommer, Frühherbst und während kalter Bedingungen bis Ende November. Anders als in den Süßwasser-Abschnitten der Warnow und an der Küste fand das Team für die Mündung keine Teilgemeinschaft, die durchgängig das ganze Jahr dominierte.

Die fünf Gruppen unterschieden sich sowohl in ihrer taxonomischen Zusammensetzung als auch in den Umweltbedingungen, unter denen sie bevorzugt auftraten. Zwei sommerliche Teilgemeinschaften waren beispielsweise vor allem mit Temperaturen oberhalb von 20 °C und Salzgehalten zwischen 15 und 20 PSU verbunden. Die Frühjahrsgruppe trat dagegen bei Salzgehalten von 8 bis 15 PSU und Temperaturen zwischen 8 und 17 °C auf. Im Herbst nahm der Süßwasserzufluss zu und der Salzgehalt längerfristig ab. Die für die Mündung typischen Teilgemeinschaften wurden daraufhin schrittweise durch Gemeinschaften ersetzt, die eher für Süßwasser oder Küstengewässer charakteristisch waren. „Die Methode hat nicht nur gezeigt, dass sich die mikrobielle Gemeinschaft verändert, sondern macht auch sichtbar, welche Gruppen von Organismen gemeinsam auf bestimmte Umweltbedingungen reagieren. Gerade in einem so dynamischen Lebensraum wie der Warnow-Mündung eröffnet das eine zusätzliche ökologische Perspektive“, hebt Theodor Sperlea hervor.

Neue Chancen für das Umweltmonitoring

Das Neue an der Studie ist die Anwendung eines ursprünglich aus der Linguistik stammenden Verfahrens im Kontext einer ökologischen Anwendung in einem dynamischen natürlichen Lebensraum, das zudem einer systematischen methodischen Bewertung unterzogen wurde. Dabei konnte die Forschenden zeigen, dass Topic Modeling komplexe Mikrobiomdaten stark verdichten und dennoch wichtige ökologische Zusammenhänge erhalten kann.

„Damit könnte Topic Modeling künftig helfen, Umweltveränderungen früher zu erkennen und Gewässer gezielter zu überwachen. Zunächst soll es bewährte Analyseverfahren jedoch nicht ersetzen, sondern eher ergänzen“, kommentiert Theodor Sperlea. „Unsere Studie macht jedoch die Tür auf für den Import weiterer Methoden aus der Computerlinguistik in die Analyse hochkomplexer Datensätze wie die aus der Umweltmikrobiologie – einschließlich der Algorithmen, die hinter den aktuell so viel diskutierten Large Language Models stecken“, so der IOW-Forscher abschließend.

* Das Projekt OTC Genomics 2 wird im Rahmen des Ocean Technology Campus Rostock vom Bundesministerium für Forschung, Technologie und Raumfahrt gefördert.

Originalpublikation

Anna S. Kujat, Christiane Hassenrück, Stefan Lüdtke, Matthias Labrenz, Theodor Sperlea (2026): Enhancing the understanding of environmental microbiomes through topic modeling: a quantitative and qualitative analysis. Environ. Microbiome, 21: 90. doi.org/10.1186/s40793-026-00927-2

Weitere Informationen und Kontakt

Pressemitteilung des Leibniz-Instituts für Ostseeforschung Warnemünde (IOW)

<https://www.leibniz-gemeinschaft.de/ueber-uns/neues/forschungsnachrichten/forschungsnachrichten-single/newsdetails/verborgene-muster>