

Die Daten verstehen

Mit visueller Datenanalyse interpretiert Prof. Dr. Daniel Keim große Datenmengen und gibt ihnen ein Gesicht



Von den Galapagos-Inseln aus (links oben) fliegen Albatrosse zu den Fischgründen an der Küste Perus (rechts im Bild, mittig). Auf ihrem Rückweg fliegen sie jedoch zunächst an der Küste entlang mehrere hundert Kilometer nach Süden, um von dort aus den Rückflug über das Meer anzutreten. Die Färbung der Pfeile zeigt jeweils die Stärke des Rückenwinds an.

Nie zuvor konnten Tiere so gut beobachtet werden wie durch die heutigen Technologien zur Positionsbestimmung – zum Beispiel durch GPS. Nie zuvor wurden aber auch so immense Datenmengen erhoben, dass sie für den Tierbeobachter nicht mehr zu bewältigen sind. Es kommen fürwahr viele Daten zusammen, wenn ein Vogelflug per Sender aufgezeichnet wird. Noch mehr Daten werden es, wenn die Bewegung ganzer Schwärme an Zugvögeln nachverfolgt wird. Völlig unüberblickbar werden die Daten spätestens dann, wenn nicht nur Positionsdaten der Tiere im 3D-Raum vorliegen, sondern zugleich auch noch Umgebungsdaten hinzukommen: Windrichtung und -geschwindigkeit, örtliche Vegetation, Niederschlag, nicht zuletzt auch Körperdaten wie der Herzschlag der Tiere.

„Mit den Tierbewegungsdaten haben wir multidimensionale, hochaufgelöste räumlich-zeitliche Daten vorliegen, für die es keine einfache, klassische Visualisierungs- oder Analysemöglichkeit gibt.“

Prof. Dr. Daniel Keim

Das Ergebnis sind viele Gigabyte an Daten. Wenn schon ein einzelner dieser Datensätze für den Menschen nicht mehr überschaubar ist, wie sollen dann viele tausende dieser Datensätze sinnvoll miteinander in Zusammenhang gebracht werden? Die Lösung bietet die Informatik mit neuen Methoden der visuellen Analyse großer Datenmengen.

„Mit den Tierbewegungsdaten haben wir multidimensionale, hochaufgelöste räumlich-zeitliche Daten vorliegen, für die es keine einfache, klassische Visualisierungs- oder Analysemöglichkeit gibt“, berichtet Prof. Dr. Daniel Keim, Professor für Datenanalyse und Visualisierung an der Universität Konstanz. In enger Zusammenarbeit mit den Konstanzer Biologen entwickelt er neue Analysewerkzeuge, die automatische Algorithmen mit interaktiven visuellen Verfahren kombinieren, um komplexe Daten möglichst effektiv und auf einen Blick verstehbar zu machen. Wo zuvor ein Wirrwarr an Zahlen war, ist nun eine Landkarte mit eingezeichneten Flugbahnen zu sehen. Weitere Datensätze wie Windbedingungen, Herzfrequenz der Tiere oder Angaben zur örtlichen Vegetation lassen sich interaktiv mit der Kartenvisualisierung verknüpfen und stufenweise einblenden.

Das Rätsel der Galapagos-Albatrosse

Die Stärke von Keims Verfahren ist genau diese Verknüpfung multipler Datensätze in ein und derselben Visualisierung, wodurch die Daten erst interpretierbar werden. Aus Positionsdaten allein lässt sich das Verhalten der Tiere noch nicht verstehen. Warum beispielsweise nehmen Galapagos-Albatrosse einen deutlich längeren Rückweg in Kauf, wenn sie von ihren Nistplätzen auf den Galapagos-Inseln zu den Fischgründen an der südamerikanischen Küste fliegen? Auf dem Hinweg fliegen sie schnurstracks nach Osten an die peruanische Küste. Auf dem Rückweg fliegen sie hingegen zunächst eine Schleife von mehreren hundert Kilometern nach Süden, bevor sie zu ihren Nistplätzen zurückkehren.

Aus den Positionsdaten heraus ist dieser Umweg nicht nachzuvollziehen. Auch geografisches Wissen über den Humboldtstrom hilft nur bedingt weiter, dieser fließt nämlich gegenläufig von Süden nach Norden. Erst die Verknüpfung mit weiteren Umgebungsdaten kann das Rätsel lösen: Unter Einblendung der lokalen Windbedingungen zeigt sich auf der visualisierten Landkarte, dass die Vögel auf diese Weise den Rückenwind maximieren. Trotz eines Umwegs von vielen hundert Kilometern können die Albatrosse somit die Strecke mit einem insgesamt besseren Energiehaushalt bewältigen. Die Vögel fliegen sogar nachts, um die Windbedingungen optimal auszunutzen.

Wichtig ist Daniel Keim das Zusammenspiel mit den Biologen. „Wir ermöglichen ihnen, anhand einer geeigneten visuellen Präsentation und Analyse der Daten ihre Hypothesen zum Tierverhalten zu überprüfen und neue Hypothesen zu generieren. Im Gegenzug gibt uns ihre Arbeit neue Anstöße für unsere Algorithmen und Visualisierungen. Diese können dann optimiert und auf größere Datenmengen angewendet werden“, erklärt Keim.

Zudem ermöglicht nur die Zusammenarbeit mit den Biologen, Fehler zu vermeiden und das Tierverhalten richtig einzuschätzen. „Schließlich liegen uns unvollständige Daten vor, da die Vögel zum Beispiel auch mit unbesenderten Tieren interagieren“, verdeutlicht Keim: „Wenn zum Beispiel ein Raubvogel den Weg kreuzt, ändert das die Flugroute. Den Zusammenhang kann der Computer anhand der vorliegenden Daten nicht verstehen. Das kann nur der Experte deuten, der Wissen über die entsprechende Spezies hat.“

| gra.



Prof. Dr. Daniel Keim ist Professor für Datenanalyse und Visualisierung an der Universität Konstanz. Zu seinen Arbeitsschwerpunkten zählen Datenbanken, Data Mining sowie die visuelle Analyse großer Datenmengen.