

Proposal for Semester Project

Patterns & Trends in Environmental Data / Computational Movement Analysis Geo 880

Semester:	FS24
Data:	Bewegungsdaten von menschlicher Aktivität
Title:	Erfassung und Visualisierung von Bewegungsdaten: Einblicke in Aktivitätsmuster und geografische Bewegungen
Student 1:	Pfister Nadja
Student 2:	Martig Serafin

Abstract

Im Rahmen dieser Projektarbeit erfolgt eine umfassende Analyse von Bewegungsdaten zweier Personen. Dabei werden Muster in der Bewegungsaktivität, die geografische und zeitliche Verteilung von Bewegungen sowie mögliche Korrelationen zwischen der Wahl des Verkehrsmittels, der Tagestemperatur und der Niederschlagsmenge untersucht. Die Daten werden mithilfe von Daten aus Google Maps und Garmin GPX erhoben. Temperatur- und Niederschlagsdaten werden von Dritten erworben. Die Ergebnisse dieser Arbeit werden visualisiert, diskutiert und in einem Bericht festgehalten.

Research Questions

Die Arbeit wird sich folgenden Forschungsfragen widmen: 1. Zeit- und Datumsanalyse • Wann finden die meisten Bewegungen statt (z.B. morgens, abends, werktags, am Wochenende)? • Wie lange dauern verschiedene Aktivitäten (Bewegung vs. Pause)? 2. Geografische Analyse • Wo hält sich die Person häufig auf? • Welche Stadtteile oder Regionen werden am häufigsten besucht? 3. Vergleichende Analysen • Wie unterscheiden sich Bewegungsmuster zwischen verschiedenen Zeiträumen? • Wie unterscheiden sich Bewegungsdaten verschiedener Personen? 4. Clustering und Hotspot-Analyse • Welche Bereiche zeigen eine hohe Bewegungsdichte? • Wo sind die Hotspots der Aktivitäten?

In dieser Arbeit wird die zeitliche und räumliche Verteilung von Aktivitäten, differenziert nach Tageszeiten und Wochentagen untersucht. In der geografischen Analyse werden häufig frequentierte Aufenthaltsorte und besuchte Stadtteile identifiziert. In einer weiteren Analyse werden Bewegungsmuster zwischen verschiedenen Zeiträumen und Personen verglichen. Clustering- und Hotspot-Analysen identifizieren und visualisieren Bereiche mit hoher Bewegungsdichte und Aktivitätsschwerpunkten. Bei hinreichender Kapazität erfolgt eine Korrelationsanalyse zwischen den Bewegungsmustern und –arten und in Abhängigkeit von der Tagestemperatur und der Niederschlagsmenge.

Results / products

Die Ergebnisse dieser Arbeit werden auf Basis einer Bewegungsanalyse erstellt. Die Möglichkeit einer Korrelationsanalyse zwischen der Bewegungsart und der Tagestemperatur sowie der Niederschlagsmenge zum jeweiligen Zeitpunkt wird eruiert. Die Analysen werden durch Visualisierungen veranschaulicht. Es werden einerseits Karten der erhobenen Bewegungsdaten erstellt und Polygonflächen der Daten visualisiert. Andererseits erfolgt eine grafische Erläuterung der Daten mittels Boxplots, Punkt- und Balkendiagrammen, welche Korrelationen aufzeigen.

Data

• Google Maps Zeitachse • GPX-Dateien von Garmin • Openstreetmap • Meteodaten z.B. von Meteo Schweiz > noch nicht eingeholt

Analytical concepts

Die Datenvorverarbeitung und -bereinigung umfasst die Integration von Bewegungsdaten aus Google Maps und GPX-Dateien, im Zeitraum vom 07. April bis zum 28. April 2024. Die Zeitstempel werden auf 15 Minuten-Intervalle gerundet. Diese Daten werden mithilfe von R automatisch konvertiert, was eine konsistente und fehlerfreie Verarbeitung ermöglicht. Im Rahmen der Zeit- und Datumsanalyse werden Aktivitätsmuster nach Tageszeit und Wochentag untersucht. Dabei wird auch die Dauer der verschiedenen Aktivitäten analysiert, um zwischen Bewegungs- und Pausenphasen zu unterscheiden. Die geografische Analyse zielt darauf ab, häufige Aufenthaltsorte der analysierten Personen zu identifizieren. Diese Daten werden visualisiert, indem die geografischen Bewegungen mittels Polygonflächen dargestellt werden. In einer weiteren Analyse, werden verschiedene Vergleiche angestellt. Einerseits werden Bewegungsmuster an verschiedenen Zeitpunkten miteinander verglichen, um Veränderungen im Bewegungsverhalten vor und nach bestimmten Ereignissen und/oder nach einer bestimmten Uhrzeit zu untersuchen. Andererseits werden Bewegungsdaten von zwei verschiedenen Personen miteinander verglichen, um individuelle Bewegungsmuster und Gemeinsamkeiten im Bewegungsverhalten zu identifizieren. In einer Clustering- und Hotspot-Analyse werden Clustering-Methoden angewendet, um Bereiche mit hoher Bewegungsdichte zu ermitteln. Eine Hotspot-Analyse wird durchgeführt, um häufig frequentierte Orte zu identifizieren und somit zentrale Orte im Bewegungsverhalten einer Person zu erkennen. Weiter wird untersucht, ob ein Aufeinandertreffen von Person A und Person B am selben Ort zum selben Zeitpunkt stattgefunden hat. Sofern die Distanz zwischen den beiden Personen weniger als 100 Meter beträgt, wird dies als Aufeinandertreffen gewertet.

Zusatzaufgabe: Integration von Wetterdaten Bei Vorhandensein entsprechender Kapazitäten erfolgt eine Analyse des Einflusses von Tagestemperatur und Niederschlag auf das Bewegungsmuster der beiden Personen. Zudem wird der Einfluss dieser beiden Faktoren auf die Wahl des Verkehrsmittels untersucht. Dazu werden Wetterdaten von Meteo Schweiz verwendet

(Open Data jedoch erst ab 2026). Die Verfügbarkeit dieser Daten muss jedoch noch abgeklärt werden. Des Weiteren ist zu eruieren, ob die Daten über eine öffentliche API zur Verfügung stehen oder ob der manuelle Download von CSV-Dateien erforderlich ist.

R concepts

Im Rahmen der Datenanalyse wird die Statistiksprache R mit der Software RStudio verwendet. Die vollumfängliche Datenvorerarbeitung, -bereinigung und -analyse erfolgt in diesem Programm. Bewegungsdaten aus Google Maps und GPX-Dateien werden integriert und konvertiert. Neben Basic-R werden folgende Zusatzpakete verwendet: `tidyr`, `readr`, `sf`, `sp`, `dplyr`, `ggplot2`, `lubridate`, `tmap`, `xml2`.

Im Rahmen der Zeit- und Datumsanalyse werden Aktivitätsmuster nach Tageszeit und Wochentag sowie die Dauer der Aktivitäten untersucht. Die geografische Analyse identifiziert häufige Aufenthaltsorte und visualisiert geografische Bewegungen mittels Polygonflächen. Vergleichende Analysen prüfen Bewegungsmuster zwischen verschiedenen Zeiträumen und Personen. Clustering-Methoden identifizieren Bereiche mit hoher Bewegungsdichte und eine Hotspot-Analyse visualisiert Aktivitätsschwerpunkte.

Zusätzlich werden meteorologische Daten von Meteo Schweiz integriert. Bei Bedarf wird auf GIS-Software wie QGIS oder ArcGIS zurückgegriffen, um geografische Analysen durchzuführen und georeferenzierte Daten zu visualisieren. Die Kombination aus R und GIS-Applikationen ermöglicht eine umfassende Analyse der Ergebnisse.

Risk analysis

Die Datenqualität und -verfügbarkeit stellen wesentliche Risiken dar, da die Bewegungsdaten unvollständig oder fehlerhaft sein könnten. Um diese Risiken zu minimieren, werden Datenbereinigungsprozesse durchgeführt und Daten zusätzlich mit einem weiteren GPS-fähigen Gerät aufgezeichnet (Sportuhr). Ein weiteres Risiko ist die Verfügbarkeit von Temperatur- und Niederschlagsdaten. Hierbei sind Vorabklärungen bei Anbietern wie MeteoSchweiz erforderlich.

Des Weiteren könnten technische Herausforderungen bei der Datenanalyse auftreten. Durch bereits erfolgte Schulungen im Umgang mit R und GIS-Tools, sorgfältige Planung und Möglichkeiten von technischem Support sollte diese Herausforderung jedoch zu meistern sein. Da es sich um Bewegungsdaten von realen Personen und privat besuchte Orte handelt, ist der Datenschutz ebenfalls von eminenter Bedeutung. Daher werden alle Bewegungsdaten anonymisiert und die Auflösung der erstellten Karten so gewählt, dass keine Rückverfolgbarkeit auf einzelne Personen möglich ist. Falls Daten verwendet werden, welche von Dritten erhoben wurden, werden Abklärungen bezüglich der Verwendung dieser Daten getroffen. Die Interpretation und Validität der Ergebnisse könnten durch methodische Fehler oder eine unzureichende Dateninterpretation beeinträchtigt werden. Um dies zu verhindern, werden

Validierungstests und Peer-Reviews durchgeführt sowie eine transparente Dokumentation sichergestellt. Zeitliche Einschränkungen könnten zu Verzögerungen im Zeitplan führen. Durch eine detaillierte Projektplanung, ein regelmässiges Monitoring sowie eine flexible Anpassung des Projektumfangs kann diesem Risiko entgegengewirkt werden. Der Ausschluss von Verkehrsdaten birgt das Risiko einer eingeschränkten Analyse, da relevante Einflussfaktoren wie Wetterbedingungen und geografische Daten nicht berücksichtigt werden. Um die Auswirkungen des fehlenden Verkehrs zu bewerten, können Sensitivitätsanalysen durchgeführt werden. Diese Massnahmen helfen, potenzielle Probleme frühzeitig zu identifizieren und zu lösen, um eine erfolgreiche Analyse sicherzustellen.

Questions?

Werden gesammelt für die Coaching-Session.