

Auf die Größe kommt es an

Maßstab und Skalierung in Geosimulationsmodellen

Christoph Fink
 Doctorate College GIScience
 Dept. of Geoinformatics Z_GIS
 University of Salzburg, Austria
 christoph.fink@sbg.ac.at

Kontext

Agenten-basierte Modelle (ABM) eignen sich hervorragend dazu, Gesellschaftstheorien auf den Prüfstand zu stellen. Ein brennendes Thema aktueller sozialgeographischer Forschung stellen die sozialen Dynamiken in den postmodernen Städten dieser Welt dar: Gentrifikation als prominentes Beispiel der vielfältigen Prozesse etwa greift in die soziale Zusammensetzung von Stadtteilen ein, verändert diese, und beeinflusst direkt und indirekt das Umzugsverhalten ihrer BewohnerInnen. Umzugsverhalten als kollektives Verhalten wiederum ist geradezu prädestiniert dazu, in einem Geosimulationsmodell abgebildet zu werden. Um soziales Verhalten valide in ABM simulieren zu können, müssen allerdings noch einige methodologische Fragestellungen beantwortet werden. Eines der vorrangigen Probleme ist der Einfluss von Maßstab und Skalierung – von Eingabedaten wie Analyse – auf die Ergebnisse von Modellen. Im vorliegenden Forschungsprojekt wird ein bestehendes Modell urbaner sozialer Transformationen in allen seiner wichtigen Dimensionen Skalierungsphänomenen ausgesetzt: skaliert werden räumliche, zeitliche und soziale Bezugseinheit, und bedeutende sozioökonomische Variablen und Indizes.



Diskussion

Viele Bereiche geographischer Forschung sind anfällig für Änderungen und Schwankungen im Analysemaßstab. Das *Modifiable Areal Unit Problem* (MAUP), so benannt von OPENSHAW (cf. 1984), beschreibt ein spezifisches Problem, das von der unabdingbar willkürlichen Abgrenzung geostatistischer Raumeinheiten her rührt. KOCH und CARSON (2013) betrachteten Skalierungsprobleme solcher Art, und legten dabei spezielles Augenmerk auf Simulationsmodelle. Dabei kamen sie zum Schluss, dass es sinnvoll wäre, das Konzept des MAUP analog zum Raum auch auf eine zeitliche und soziale Komponente auszudehnen, und schlugen dafür die Bezeichnungen MTUP und MSUP vor. Das vorliegende Forschungsvorhaben zielt darauf ab, genau diese Effekte sichtbar und interpretierbar zu machen, und sie näher zu quantifizieren. Der erste Schritt, die Disaggregation/Synthesierung einer Population im Maßstab 1:1 unter Zuhilfenahme der Originaldaten eines ABM, ist erfolgreich abgeschlossen. Der Quelltext der erstellten Werkzeuge ist unter der im QR-Code im unteren rechten Eck dieses Posters hinterlegten Webadresse einsehbar. Da weder die Anwendung darauf abzielt, noch der Algorithmus dazu im Stande wäre, die „reale“ Bevölkerung abzubilden, und der Anspruch auch mit datenschutzbezogenen Bedenken kollidieren würde, kann kein Ground Truthing betrieben werden; eine intrinsische Validierung durch Re-Aggregation fiel allerdings positiv aus. Nun wird in einem nächsten Schritt eine Experimentierumgebung entwickelt, die die disaggregierten Daten schrittweise nach Modelldimensionen (Raum, Zeit, soziale Bezugseinheit, ...) re-aggregiert, und jeweils eine Serie Modelldurchläufe ausführt. Es wird erwartet, aus den zu erwartenden Daten Rückschlüsse auf den Grad der MAUP-, MTUP- und MSUP-Effekte ziehen zu können, sowie durch die offene und modulare Gestaltung weitere Simulationsmodelle analysieren zu können, und wertvolle Hinweise zur Optimierung der Model-Designs geben zu können.

Methodik (I) Datendisaggregation

Zunächst müssen dazu die Eingabedaten auf die größtmöglich sinnvolle Auflösung disaggregiert werden, um sie später schrittweise wieder zu aggregieren. Ziel ist, im Vergleich zum getesteten Modell möglichst wenig zusätzliche Datenquellen zu benötigen. Im vorliegenden Falle sind bereits Volkszählungsdaten auf der Basis von Zählsprengeln vorhanden. Enthalten sind unter anderem verschiedene demographische und sozioökonomische Variablen und Indizes. Zur Disaggregation werden weiters ein feinmaschiger Raster der Bevölkerungszahl, sowie Polygone der Gebäude hinzugezogen. Diese Daten sind für die meisten Länder verfügbar; als Testdaten wurden Daten des Institut national de l'information géographique et forestière (IGN 2012) und des Institut national de la statistique et des études économiques (INSEE 2012a,b) verwendet. Zunächst werden die Gebäude einem regionalen Schwellwert ihrer Fläche nach gefiltert (Median des Baublocks ± 30), um Ausreißern wie innerstädtischen Einkaufszentren und Kiosken habhaft zu werden, bevor die Werte des Bevölkerungsrasters aliquot auf die verbleibenden Gebäude aufgeteilt werden. Dann wird ein modifizierter Inverse-Distance-Weighting-Algorithmus

angewandt, der lokale Werte in jedem Gebäude für alle in den Zählsprengeln verfügbaren Variablenspalten berechnet:

$$u(B_x) = \frac{\sum_{i=1}^n u(S_i) \left(\frac{1}{d(B_x, S_i)} \right)^p}{\sum_{j=1}^n \left(\frac{1}{d(B_x, S_j)} \right)^p} \cdot \frac{1}{u(S \ni B_x)} \cdot \frac{s(B_x)}{\sum_{k \in S_x} s(B_k)}$$

$u(A)$Wert u in Polygon A (bzgl. rel. z. Fläche) BGebäudepolygon
 pconst. SZählsprengel-Polygon
 $d(A,B)$euklidische Distanz zwischen den Schwerpunkten der Polygone A und B
 $s(A)$„seed“ Population (Gewichtung) in (Gebäude-)Polygon A

Die erhaltenen Daten erlauben es, Individuen in Haushalte zu gruppieren (Haushaltsgrößen und -zahlen sind verfügbar), die wiederum auf die Gebäude aufgeteilt werden. Schließlich erhalten wir Datensätze von (a) Gebäuden mit ganzzahliger Bevölkerung, (b) Gebäuden mit dezimalen und ganzzahligen Werten in allen Spalten der Eingabedatensätze, und (c) einen Satz interrelationaler Tabellen von Individuen, Haushalten und Gebäuden. Die Daten sind in SpatialLite Datenbanken gespeichert, und mit ihrer jeweiligen räumlichen Position und Ausdehnung verlinkt.

Methodik (II) Iterative skalierte Simulation

Schließlich wird ein Framework geschaffen, das agentenbasierte Modelle in allen Skalierungskombinationen ablaufen lassen kann. Netlogo (WILENSKY, 1999), eine beliebte Entwicklungsumgebung für ABM, bietet mit „BehaviourSpace“ eine ähnliche Experimentierumgebung an, die es allerdings nicht erlaubt, die räumlichen Grundbezugsgrößen zu variieren. Die explizite Berücksichtigung der (euklidischen) Räumlichkeit der modellierten Phänomene zeigt sich auch darin, dass das im Rahmen dieses Projektes entwickelte Framework Geodaten nicht nur als Eingabe-Datenquelle betrachtet, sondern alle Operationen direkt in der als Backend dienenden SpatialLite- bzw. PostGIS/Postgresql-Datenbank ausführt und speichert. Grundsätzlich wird dabei jede Kombination aller Maßstäbe/Skalierungen jeweils aller Dimensionen in Monte-Carlo-Verfahren vielfach durchlaufen. Im Gegensatz zum klassischen Monte-Carlo-Verfahren werden die Resultate jedoch nicht unmittelbar gemittelt, sondern vorerst alle Interaktionen und alle Zustände einzeln aufgezeichnet. Dabei wird auf die Fähigkeit PostGIS' zur Versionskontrolle zurückgegriffen, was später eine uneingeschränkte Visualisierung jedes

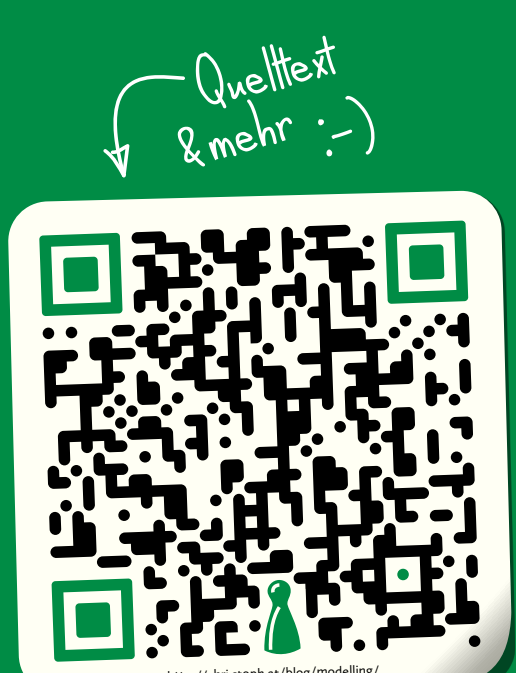
Simulationszustandes erlaubt. Der Visualisierung der Simulationsschritte und der Resultate wird generell besondere Aufmerksamkeit gewidmet. Die Programmierung der Experimentierumgebung wird in Python ausgeführt, da es ein vorzügliches Gleichgewicht zwischen Performanz und Entwicklungsgeschwindigkeit bietet, und bereits Bibliotheken für vielfältige Anwendungen existieren, die beispielsweise die Anbindung an Geodatenbanken und ABM-Umgebungen erleichtert. Zunächst muss auch das zu untersuchende Simulationsmodell in Python übersetzt werden, mittelfristig ist loose-coupling mit ABM-Umgebungen wie Netlogo geplant. Schließlich und endlich ergibt sich eine n -dimensionale Matrix (wobei n die Anzahl der variierten Dimensionen angibt) von aufgezeichneten Simulationen und deren Ergebnissen. Diese werden nach Entwicklung und Berechnung von Indizes einer klassischen Faktoren- und Clusteranalyse unterzogen, aber auch vermehrt Visual Analytics sozusagen in einem Grounded Truth-Verfahren analysiert.

In 20 Sekunden:

Um agentenbasierte Modelle (ABM) kollektiver sozialer Interaktionen hinsichtlich ihrer Anfälligkeit für Skalierung zu evaluieren, wird deren Population zunächst auf die kleinsten sinnvollen Einheiten in den Dimensionen Raum, Zeit und Akteur(e) disaggregiert.

Danach werden die Akteure, ihre Räumlichkeit und Zeitlichkeit wieder schrittweise aggregiert, wobei jeweils eine Reihe von Modelldurchläufen je Skalierungskombination aufgezeichnet wird.

Am Ende steht eine n -dimensionale Matrix von Interaktionen der Agenten und von Resultaten im virtuellen „Raum“, die mittels klassischer Faktoren- und Clusteranalyse, vermehrt Visual Analytics, und in einem Grounded Theory Ansatz analysiert werden.



References:
 IGN (2012): BD PARCELLAIRE*. – available from <http://professionnels.ign.fr/bdparcellaire> (last visited 2013-09-16)
 INSEE (2012a): Bases de données du recensement de la population 2008. – available from <http://www.insee.fr/fr/bases-de-donnees/default.asp?page=recensement/resultats/2008/donnees-detaillees-recensement-2008.htm> (last visited 2013-09-16)
 INSEE (2012b): Données carroyées de la population. – available from http://insee.fr/fr/themes/detail.asp?reg_id=0&ref_id=donnees-carroyees (last visited 2013-09-16)
 KOCH, A. & D. CARSON (2013): Spatial, Temporal and Social Scaling in Sparsely Populated Areas – Geospatial Mapping and Simulation Techniques to Investigate Social Diversity, in: T. Jekel, A. Car, J. Strobl & G. Griesebner (eds.), GI Forum 2013, (pp. 2–5).
 OPENSHAW, S. (1984): The Modifiable Areal Unit Problem. CATMOG, 38.
 WILENSKY, U. (1999): NetLogo. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston. – available from <http://ccl.northwestern.edu/netlogo/> (last visited 2013-09-17)