

MENSEN IN BEWEGING

Het gebruik van “tracking” gegevens binnen de ruimtelijke planvorming

Arend Ligtenberg, Centrum Geo-Informatie Wageningen Universiteit

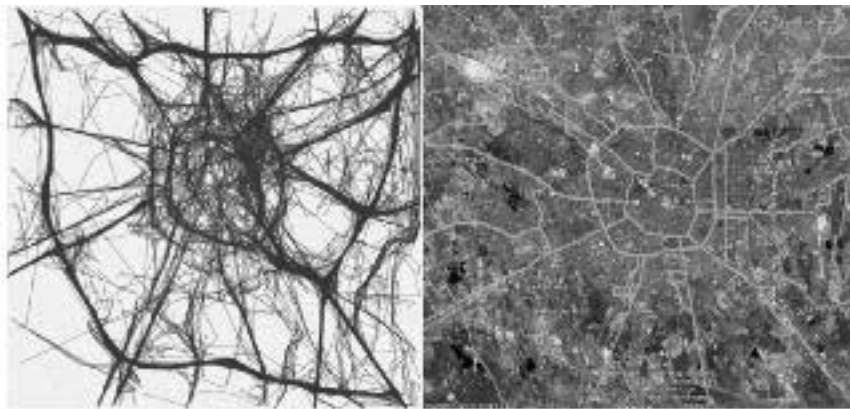
Inleiding

Bijna iedereen maakt gebruik van een mobiele telefoon en veel mensen hebben een navigatie-apparaat in hun auto. Een kenmerk van deze apparaten is, dat ze moeten “weten” waar ze zijn, om van nut te kunnen zijn voor hun gebruiker. Een mobiele telefoon is verbonden met een zendmast, die op een bekende locatie staat. Een navigatiecomputer gebruikt GPS om nauwkeurig te kunnen vaststellen waar het zich bevindt.

Een ander kenmerk is dat deze apparaten ook hun digitale sporen achterlaten, veelal in de vorm van locatie- en tijdsregistraties. Een telefoon doet dit in de logbestanden van het telecombedrijf waar het mobieltje geregistreerd is (de zgn. “hand-over” data) en veel navigatieapparaten bieden de mogelijkheid om routegegevens te uploaden en zo de netwerkinformatie te verbeteren. Het gevolg is dat er uitgebreide gegevensbestanden worden bijgehouden, waarin, afhankelijk van het apparaat waar de registratie vandaan komt, veelal een redelijk gedetailleerd beeld ontstaat van hoe mensen zich bewegen door de tijd en ruimte (zie figuur 1).

Deze informatie is potentieel erg interessant voor verschillende toepassingen. Een aantal van deze toepassingen ligt voor de hand, zoals in de domeinen van de verkeersplanning, openbaar vervoer en (geo)marketing. TomTom bijvoorbeeld gebruikt gegevens van Vodafone gsm gebruikers om betere voorspellingen te doen van de filesituatie. Als iemand met zijn gsm telefoon belt wordt bijna continu de locatie en snelheid doorgegeven aan TomTom. Daarnaast heeft het bedrijf bijvoorbeeld gedetailleerd in beeld gebracht wat per wegsegment de gemiddelde gemeten snelheid is op een bepaald tijdstip. Hiervoor gebruikt het de miljoenen gereden routes die verzameld zijn door TomTom de afgelopen jaren (deze database groeit gemiddeld met een half miljard metingen per dag). Dit is natuurlijk een voor de hand liggende toepassing, maar ook voor ruimtelijke planning zijn er een aantal toepassingsgebieden waarvoor “moving object” data aanvullende analytische mogelijkheden. Het huidige probleem is wellicht dat deze gegevens dan gebruikt gaan worden in toepassingen die niets te maken hebben met het doel waarvoor deze data oorspronkelijk zijn verzameld. Dit levert een aantal interessante uitdagingen op rondom het verzamelen, opslaan, verwerken en de privacy van deze data.

In nationaal en Europees verband worden er een aantal projecten uitgevoerd die de verschillende aspecten van de toe-



Figuur 1: Bewegingen rond en in Milaan tussen 1 en 7 april 2007 (2.075 mlj. punten, 17.000 voertuigen en 200.000 trajectories). Bron Geopkdd project

passing van deze data verkennen. In dit artikel wordt een kort overzicht gegeven van de verschillende activiteiten die binnen deze projecten plaatsvinden. Begonnen wordt met een korte uitleg van een aantal basisconcepten en vervolgens zal een casestudie rondom een toepassing voor het analyseren van recreatie in natuurgebieden worden gepresenteerd.

“Moving objects” data

Karakteristieken

De basis voor het huidige denken over ruimte en tijd is gelegd door Torsten Hägerstrand in 1970 die beweging via concepten als “space-time paths” en “space-time prisms” definieerde. Een belangrijk uitgangspunt bij Hägerstrand was, dat tijd en ruimte niet te scheiden zijn, ook niet vanuit analytisch oogpunt. Dit markeert het begin van het denken over het gebruik van ruimtelijk-temporele gegevens in de geografie en aanverwante vakgebieden. Tot die tijd (en in de meeste gevallen tegenwoordig ook nog) werd verplaatsingsgedrag van mensen gezien als geaggregeerd probabilistisch gedrag. Dat betekent dat processen en interacties op lagere schaalniveaus niet expliciet kunnen worden meegenomen in analyses. Dit beperkt het inzicht, niet alleen in de oorzaken van bepaald bewegingsgedrag in relatie tot een sociaal-ruimtelijk omgeving, maar ook, vanuit een toepassingsoptiek, in eventuele oplossingen voor verplaatsingsgerelateerde vraagstukken, bijvoorbeeld de fileproblematiek, bereikbaarheid van stadscentra of recreatiedruk in gevoelige gebieden enz.

In essentie bestaan de ruimtelijk-temporele gegevens, waar we het hier over hebben, uit een aantal 3-dimensionale punten (x,y,t) die (meestal) in chronologische volgorde aan elkaar

| Siemens_ Cella | Cella Adiacente | Intervallo | Handover | Easting Cella1 | Northing Cella1 | Easting Cella2 | Northing Cella2 |
|----------------|-----------------|--------------------|----------|----------------|-----------------|----------------|-----------------|
| MI001D1 | MI034G2 | 25/10 05:00..06:00 | 1 | 9.18944 | 45.46695 | 9.1825 | 45.47306 |
| MI001D1 | MI034G2 | 25/10 16:00..17:00 | 1 | 9.18944 | 45.46695 | 9.1825 | 45.47306 |
| MI001D1 | MI034G2 | 25/10 15:00..16:00 | 3 | 9.18944 | 45.46695 | 9.1825 | 45.47306 |
| MI001D1 | MI034G2 | 25/10 14:00..15:00 | 3 | 9.18944 | 45.46695 | 9.1825 | 45.47306 |
| MI001D1 | MI034G2 | 25/10 13:00..14:00 | 1 | 9.18944 | 45.46695 | 9.1825 | 45.47306 |
| MI001D1 | MI034G2 | 25/10 12:00..13:00 | 2 | 9.18944 | 45.46695 | 9.1825 | 45.47306 |

Figuur 2: deel van een datalog met cell-handover recordings van een mobile operator

zijn gerelateerd (zie figuur 2). Zo'n chronologische constructie van punten wordt vaak aangeduid met de term trajectorie of "geo-spatial lifeline". Zo'n trajectorie is dus een representatie van de verplaatsing van een persoon (voetganger, fietser, automobilist). Een trajectorie heeft een aantal kenmerken

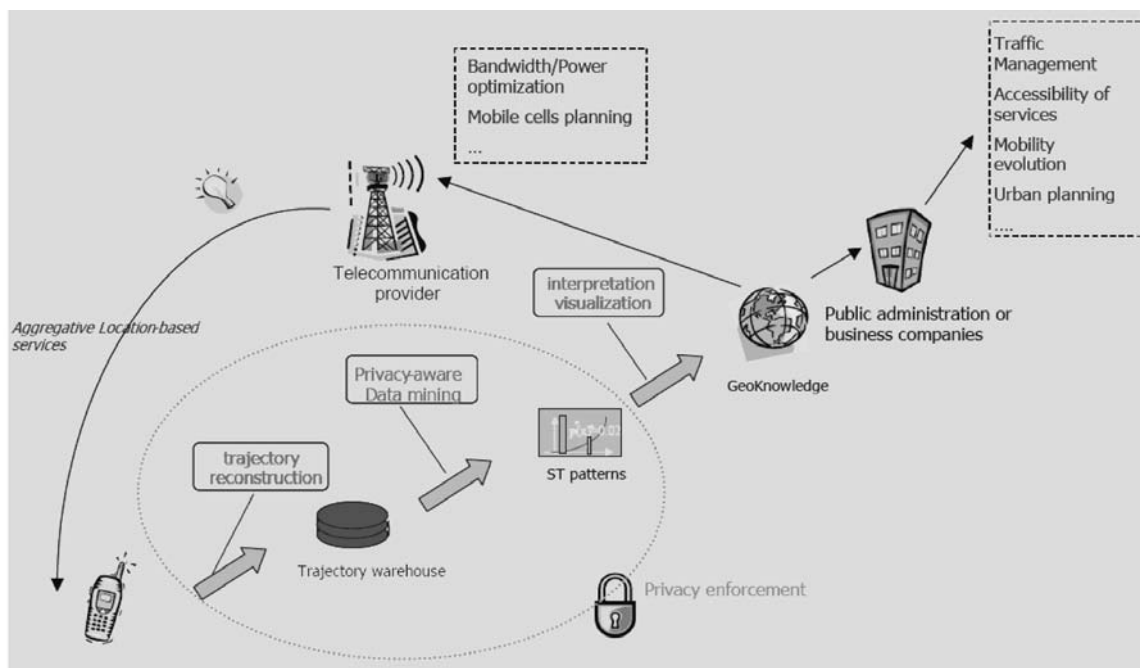
- Een begin en eindpunt met;
- daartussen periodes met continue verplaatsingen en;
- periodes zonder verplaatsingen (stops en starts).

Uitgangspunt is dat elke trajectorie ontstaat vanuit de doelstelling van een persoon om van A naar B (waarbij natuurlijk niet uitgesloten is dat B gelijk is aan A) te gaan. Dit kan zijn: woon-werkverkeer, supermarktbezoek, een wandeling maken of wat dan ook. De moeilijkheid ligt hier in de definitie of iets een "stop" of "start" is, dus onderdeel van een trajectorie, of een begin of eindpunt; dus een trajectorie zelf. Als iemand bijvoorbeeld staat te wachten voor een kruising om over te steken zal dat in de meeste gevallen een stop/start zijn binnen een trajectorie, maar de vraag of de supermarkt een eindpunt is van een trajectorie of een stop in een trajectorie met als begin- en eindpunt iemands huis is minder duidelijk. In z'n algemeenheid hangt deze definitie af van het onderzoeksdoel en de gewenste ruimtelijke en temporele schaal of granulariteit.

Vanuit trajectories kunnen verschillende gegevens worden afgeleid. Onder andere de snelheid, lengte en vorm van trajectories op bepaalde locaties en tijdstippen maar ook bijvoorbeeld clusters, dichtheden (aantal trajectories per oppervlakte eenheid per tijdseenheid), overlappingsen en andere topologische relaties.

Privacy

Een belangrijk aspect van het gebruik van dit soort data is natuurlijk de privacy. Het originele doel van de meeste van de datasets, die interessant zijn om te gebruiken, is anders dan voor ruimtelijke analyses. Telecom operators gebruiken de data om te kunnen factureren en wellicht om de belasting van bepaalde basisstations te kunnen meten. Voor dit gebruik hebben hun klanten getekend. Iets dergelijks geldt ook voor het gebruik van data van GPS en navigatie apparaten. Dit betekent dat, indien deze data gebruikt gaan worden voor ruimtelijke analyse, deze op z'n minst zodanig moeten worden bewerkt, dat met een bepaalde betrouwbaarheid kan worden gesteld dat data niet kunnen worden herleid tot individuen. Het mag duidelijk zijn dat simpel anonimiseren veelal onvoldoende is. Via analyses van beweging en koppeling met andere databestanden, zijn persoonsgegevens meestal eenvoudig te achterhalen. Een (niet ruimtelijk) voorbeeld van een privacyprobleem, waarbij anonimiseren absoluut onvoldoende bleek te zijn, is aangetoond door



Figuur 3: Vraagstukken waar aan gewerkt wordt vanuit het GeoPkdd en People in Motion project.

Amerikaanse onderzoekers, die een geanonimiseerde dataset met gegevens over ziekenhuisopnames, behandelingen enz. ter beschikking kregen van een overkoepelende verzekeringsorganisatie (de zgn. GIC dataset) (Sweeney 2002). Deze organisatie zag hierin geen probleem aangezien de data was geanonimiseerd. Versies ervan werden ook verkocht aan de industrie. Naast ongeveer 100 medische velden bevatte de dataset ook een 5 cijferige zip-code, geboortedatum en geslacht. Door nu deze dataset eenvoudig te combineren met de een voor \$20 aangekochte kiezerslijst van Cambridge Massachussets was het kinderlijk

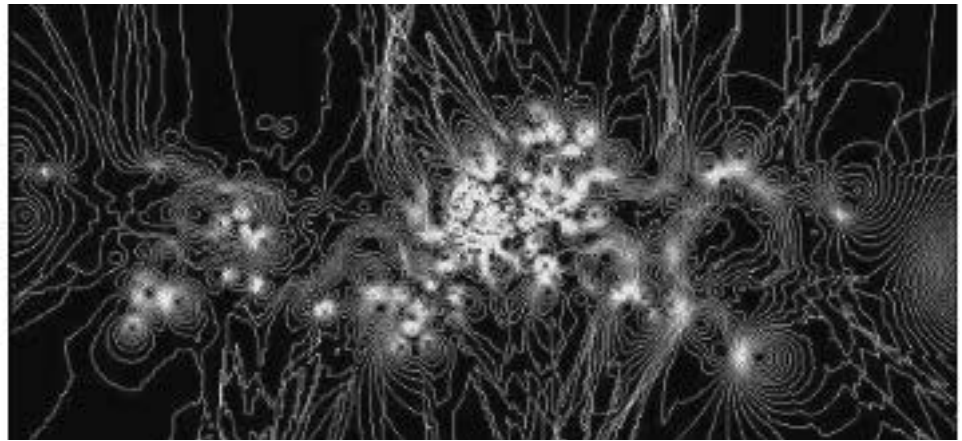
eenvoudig de persoon opnieuw te koppelen aan zijn medische gegevens. Voor de gouverneur van Massachussets (destijds William Weld) bleek bijvoorbeeld dat 6 anderen zijn geboortedatum hadden, slechts drie daarvan waren man en hij was de enige binnen zijn 5 cijferige Zip code gebied. Dit is slechts een simpel voorbeeld maar het geeft wel aan wat het belang is van het ontwikkelen van goede concepten en bijbehorende technieken om de privacy te waarborgen.

Een van de concepten die geschikt lijken te zijn voor “moving object” databases, is dat van *K*-anonymity, eventueel in combinatie met quasi identifiers. Dit betekent dat onder een regime van *K*-anonymity ieder deel van de beschikbare data gelijk is aan ten minste *k* andere delen van informatie betreffende het gedeelte van de data wat privacy gevoelig is.

“Work in progress”

Bovenstaande aspecten worden op dit moment verder uitgewerkt in een aantal projecten waar het Centrum Geo-informatie van Wageningen Universiteit bij betrokken is. Het eerste project betreft het project “Geographic Privacy-aware Knowledge Discovery and Delivery” oftewel GeoPkdd waar in EU verband samengewerkt wordt met een aantal partners aan met name de informatie theoretische aspecten van “moving object” data. Figuur 3 geeft een overzicht van de verschillende onderwerpen waaraan gewerkt wordt binnen dit project.

Op het laagste niveau speelt de opslag van de ruimtelijk-temporele data in specifieke datastructuren en “datawarehouses” een belangrijke rol. Typisch gaat het om grote hoeveelheden dynamische gegevens afkomstig vanuit verschillende sensoren. Om hiermee wat te kunnen, zijn onder andere technieken, waarmee de trajectorie kan worden gereconstrueerd, “Spatial-temporal datamining” en special query talen nodig. Te denken valt aan specifieke clustering algoritmes die om kunnen gaan met zowel de temporele en geografische dimensies als ook de geometrische aspecten (bijvoorbeeld similariteit in vorm door de tijd) waarmee patronen in de data kunnen worden gevonden en geanalyseerd met in acht neming van benodigde privacy (ST patterns in figuur



Figuur 4: Patronen van het gebruik van mobiele telefoons in Graz – <http://senseable.mit.edu/graz/>

3). Op het hoogste niveau gaat het met name om de verschillende toepassingen die mogelijk zijn. Het onderzoek op dit niveau richt zich onder andere op het proces van de vertaling van de relatief betekenisloze patronen (zie bijvoorbeeld figuur 3), het resultaat van de datamining, naar betekenisvolle modellen in de context van een toepassing (Wachowicz et al 2008). Binnen het RGI-160 project “People in Motion” richten we ons op de verkenning van manieren waarop dit kan. Door het combineren van “moving object” gebaseerde patronen met ander geodata kunnen relaties tussen verplaatsingen van mensen of (grote) groepen van mensen en de omgeving worden geïnventariseerd. Een concrete case waaraan gewerkt wordt, is het gebruik van GPS- (en eventuele mobiele telefoon-) data van wandelaars en fietsers in het Dwingelerveld en de Grebbeberg om zo een beter inzicht te krijgen in de gevolgde routes, de “points of interest” en de interacties met het landschap en de overige wandelaars en fietsers. De veronderstelling is dat als gedetailleerder kan worden geanalyseerd hoe mensen zich gedragen in een gebied, dit aanvullende inzichten geeft, vergeleken met de meer traditionele wijzen van recreatie onderzoek, wat overwegend via interviews plaatsvindt. Eind 2008 lopen de bovenstaande projecten ten einde en zal hierover meer duidelijkheid zijn.

Literatuur

- Hägerstrand T.; 1970; What about people in regional science?; *Papers of the Regional Science Association*, 24:7-21
- Sweeney L.; 2002; *k*-anonymity: a model for protecting privacy. *International Journal on Uncertainty, Fuzziness and Knowledge-based Systems*, 10 (5), 557-570
- Wachowicz M., Ligtenberg A., Renso C., Gürses S.; 2008; Characterising the next generation of mobile applications through a privacy-aware geographic knowledge discovery process; in F, Giannotti and D. Pedreschi (eds) *Mobility, Data Mining and Privacy*, 39-72; Springer-Verlag.