Comparison of approaches for area segmentation and application to geomarketing analysis

Comparison of area segmentation approaches for the application to geomarketing analysis

Development of an area segmentation algorithm for geomarketing analysis

TOC

# Introduction

## Definition and aim of Geomarketing

## Use cases from the field of Geomarketing

## Motivation and Research question

## Methods

## Outline

# Related Work

## KIT – Institute of Operations Research: discrete optimization and logistic

## RegioGraph

## SimTool

## Easy Map District Manager

## (DISTRICT from CartoGIS)

# Application of area segmentation in the field of Geomarketing

## Optimization of Areas

## Greenfieldanalysis

## Whitespotanalysis

# Approaches of area segmentation and their implementation

## Approaches and Implementation

### Segmentations considering only homogenous distribution

#### segmentation considering just the criteria

#### Segmentation considering sum of criteria / number of locations

### Segmentations considering only distance/compactness

#### segmentation considering just the distance

### segmentations considering criteria and distance

#### criteria + Distance: from inside to outside

##### smallestCritGetsNearest

##### smallestCritGetsTrueNearest

#### criteria + Distance: from outside to inside + inside to outside

#### sum of criteria/number of locations + Distance

#### Distance + criteria - rearranging

## Problems of approachs

### Performance

### Requirements from the field of Geomarketing

#### Formation of holes

#### Inhomogeneous distribution

#### Rearrangement/Infinite loops during rearrangment

#### Need of threshold values

# Comparison of approaches

## Performance

## Problems

## Requirements

## Conclusion

# Application of Algorithm XYZ to Geomarketing analysis

## Optimization of Areas

### Conditions and Aim

### Approach of Optimization algorithm

## Greenfieldanalysis

### Conditions and Aim

### Approach of Greenfield algorithm

## Whitespotanalysis

### Conditions and Aim

### Approach of Whitespotalgorithm

# Realworld scenario: integrated Algorithm to the mapChart Manager

# Discussion and Perspective

## Summary

## Limitations

## Comparison to related work

## Perspective

# Introduction

## Definition and Aim of Geomarketing

* Connection of information in a spatial way to support marketing decisions of companies
* Connects information with spatial reference system 🡪 statistical analysis can be done 🡪 support decisions
* Geomarketing ist eine Teildisziplin des Marketings. Es verknüpft alle in einem Unternehmen zur Verfügung stehenden Informationen mit einem geografischen Koordinatensystem. Durch die statistische Auswertung des daraus resultierenden Datenpools sind so Rückschlüsse auf verschiedene marketingrelevante Aspekte möglich. Mit verbesserten technischen Möglichkeiten gewinnt Geomarketing in jüngster Zeit zunehmend an Bedeutung. Quelle: http://www.omkt.de/geomarketing/

Role of microm

* Connect analysis with compenies: data necessary 🡪 given by microm/creditreform
* Provide webtool for doing geomarketing analysis --> mapchart manager

## Use cases from the field of Geomarketing

### Gebietsplanung, Kundengewinnung, Wettbewerbsbeobachtung, Fahrtzeitzonen,etc

## Motivation and Research question (including aim of the work)

### Um was geht es, was ist das Ziel, warum, was wird untersucht

## Methods

### Wie wird die Untersuchung angestellt, was/welche Tools werden dafür verwendet

## Outline

### Wie ist die arbeit aufgebaut; auf was wird eingegangen in welcher Reihenfolge, was ist in den Kapiteln zu finden

# Related Work

## KIT – Institute of Operations Research: discrete optimization and logistic

Beispiele gegeben, zb basierend auf PLZ Gebieten in DTL, dabei gibt es eine Anzahl von Punkten, die gewisse Aktivität (=Potential) aufweisen.

Punkte können auch als Kreuzungen für Straßenabschnitte vorliegen

Keine Zentren vorhanden! Sondern lediglich Gebietsaufteilung ohne Beachtung von Zentren

Einstellungsmöglichkeiten:

* Anzahl der Gebiete
* Anzahl der Suchrichtungen, notwendig, wenn Gebiete in linkes und rechtes Problem zerteilt werden
* Balance tolerance, wie weit die Werte voneinander abweichen dürfen
* Wichtung der Balance: 1 gleichverteilung ist am wichtigsten, 0 Kompaktheit ist am wichtigsten
* Maß für die Messung der Kompaktheit
* Art der Teilung der Gebiete mittels Linie oder Zone

Vorgehen:

* Gebiet wird halbiert, anschließend wird ein Gebiet halbiert, danach das andere, usw
* Teilung erfolgt so, dass das Gebiet geteilt wird
  + Was am größten
  + Was ziel durch teilung am nächsten kommt
* Muss vorher bereits interpoliert oder ähnliches sein, da teilgebiete am ende so entstehen, dass Balance erreicht
* Nutzt Algorithmus "Recursive Partitioning Algorithm" ist in folgendem Paper dargestellt:  
  J. Kalcsics, S. Nickel und M. Schröder: Towards a Unified Territorial Design Approach - Applications, Algorithms and GIS Integration, TOP 13(1), 1-74 (2005)

## (RegioGraph) stattdessen Map&Market

## SimTool

## Easy Map District Manager, Demo verfügbar

<http://www.easymap24.de/Video_Manuelle_Gebietsplanung.php>

http://www.easymap24.de/Video\_Zusatzdaten\_und\_Standorte.php

## (DISTRICT from CartoGIS)

# Application of area segmentation in the field of Geomarketing

## Optimization of Areas

### Was ist das, was wird gemacht, warum

## Greenfieldanalysis

### Was ist das, was wird gemacht, warum

## Whitespotanalysis

### Was ist das, was wird gemacht, warum

# Approaches of area segmentation and their implementation (alle mit Flow-Diagramm)

## Zu betrachtende Kriterien: Kompaktheit, keine Löcher, homogene Verteilung, etc. siehe KIT Seite Planungskriterien; Tests ausgeführt auf PLZ5 und PLZ8 Gebiete mit existierenden Sparkassenstandorten

## Approaches and Implementation

### Segmentations considering only homogenous distribution

#### segmentation considering just the criteria

##### Zuordnung des Polys zu Standort, wo weniger Criteriasum, Zuordnung erfolgt nach Reihenfolge der Polys

##### **Kriterien:** homogene Verteilung der Variable

##### **Erkenntnis:** Variable ist homogen verteilt, aber keine Kompaktheit und Zusammenhängigkeit gegeben; in dieser Form nicht zu empfehlen, sondern Kombination mit Distanz notwendig; Ergebnis nicht reproduzierbar, da Anordnung des Flickenteppichs abhängig von Reihenfolge der Gebiete in Datenbank, da einfach Reihenfolge in Datenbank abgearbeitet wird; dennoch entsteht immer ein Flickenteppich; homogene Gleichverteilung ist nicht an Schwellwert gekoppelt, somit besteht Gefahr einer Diskrepanz, da zb das letzte zugewiesene Gebiet ein zu hohes Kriterium hat und so die Summe dieses Gebietes nach oben schnellt

#### Segmentation considering sum of criteria / number of locations

##### X=sum of criteria/number of locations; jeder Standort bekommt so viele Polys damit x erfüllt **Kriterien:** homogene Verteilung der Variable

###### **Erkenntnis:** Kompaktheit ist bedingt durch Reihenfolge der Polys in der DB --> stehen benachbarte Polys hintereinander, dann kompakt, ansonsten kann Flickenteppich entstehen; homogene Verteilung ist nur bedingt möglich: wenn ersten Standorte zu viel über y drüber sind, weil die letzte Zuordnung so groß vom Wert her war, weißt der letzte Standort eine große Diskrepanz auf (siehe 4 locations); Zuordnung der Gebiete zu location nicht an die Position des Standortes gebunden, sondern das zugehörige Gebiet kann ganz woanders liegen (siehe 4 locations, PLZ5 und PLZ8); keine Beachtung der Distanz zum Standort während der Zuordnung --> somit unbedingt Beachtung der Distanz notwendig, damit sinnvolleres Ergebnis

### Segmentations considering only distance/compactness

#### segmentation considering just the distance

##### Zuordnung des Polys zu dem Standort, der am nähsten

##### **Kriterien:** geringe Fahrtdistanz = Kompaktheit

##### **Erkenntnis:** Kompaktheit wird gewährleistet, aber es können Löcher entstehen, wenn Standorte ungünstig liegen, große Diskrepanz der Untersuchungsvariable --> keine homogene Verteilung

##### **Frage:** um das Gesamtergebnis kompakter zu bekommen, erst nach Distanz und dann umsortierung damit homogen möglich?

### segmentations combining criteria and distance

#### criteria + Distance: from inside to outside

###### smallestCritGetsNearest

Zuordnung des Polys zu Standort, der am nähsten; Standort mit wenigsten Criteriasum bekommt nächstes Poly  
**Kriterien:** Kompaktheit + Gleichverteilung  
**Erkenntnis:** Kompaktheit ist nur bedingt gewährleistet; bei Zuordnung der Polys entstehen Löcher, da zwar Distanz geprüft wird, aber nur von den Polys, die bisher noch nicht verteilt sind, dazwischen kann aber bereits ein Gebiet einem anderen Standort zugeteilt worden sein; somit Prüfung auf Zusammenhängigkeit notwendig; Es können „Nasen“ entstehen, da andere näher liegende Polys bereits anderen Standort zugeordnet sind; Die Werte sind relativ gut gleichverteilt; homogene Verteilung nicht gebunden an Schwellwert, somit Gefahr von Diskrepanz  
**Fazit:** Unbedingt notwendig ist die Prüfung von zusammenhängenden Gebieten; Kompaktheit ist nur bedingt gewährleistet

###### smallestCritGetsTrueNearest

Zuordnung des Polys zu Standort, der am nähsten, aber nur, wenn das zu keinem anderen Standort gehören würde, weil dieser näher; falls anderer Standort näher, schaue, ob es von diesem Gebiet ein Poly gibt, was ich nehmen kann, weil näher zu mir --> nehme das Poly weg; nur wenn das Gebiet benachbart; falls nicht, nehme von den benachbarten gebieten ein Poly von dem Gebiet, was höchste CritSumme; Standort mit wenigsten Criteriasum bekommt nächstes Poly  
**Kriterien**: Kompaktheit + Gleichverteilung  
**Erkenntnis**: Die Werte sind sehr gut gleichverteilt, vermutlich besteht jedoch die Gefahr, dass Diskrepanz entstehen kann, je nach Größe des Kriteriums je Gebiet, da lediglich Verteilung aller Gebiete, aber keine nachträgliche Prüfung, ob gleichverteilt; Kompaktheit wird besser gewährleistet als bei smallestCritGetsNearest, da immer das zugeordnet wird, was am nahsten gelegen, Gefahr von Löchern nur, wenn durch Form des Gebietes bedingt (zb wenn Gesamtgebiet eine Nase hat, die mehrere Gebiete beinhaltet, die Standort A zugeordnet sind, aber das Gebiet in der äußersten Ecke Standort B zugeordnet wird, da am nahesten); Gefahr von Nasen besteht, siehe 4 Locations; Standort liegt zT außerhalb des Gebietes siehe PLZ5  
**Fazit**: Das Ergebnis ist besser als bei smallestCritGetsNearest, aber nachträgliche Prüfung auf Gleichverteilung notwendig, um diese zu gewährleisten

##### criteria + Distance*:* from outside to inside + inside to outside

###### Einschränken der Anzahl an Polys, indem äußere Polys, die eindeutig einem Standort zugeordnet werden können, zugeordnet werden: alle Polys werden 1x durchgegangen; anschließend Zuordnung wie bei smallestCritGetsNearest

**Kriterien:** Kompaktheit + Gleichverteilung  
**Erkenntnis:** Kompaktheit ist besser als bei iii1ai, dennoch entstehen Löcher wodurch Prüfung auf Zusammenhängigkeit notwendig; Kriterium ist annähernd gleich verteilt, wobei das nicht immer gegeben sein muss (wenn ein Gebiet zu viele äußere Polys zugeordnet bekommen hat, dass Abweichung nicht mehr aufgeholt werden kann), somit Umverteilung notwendig, um Gleichverteilung zu wahren  
**Fazit:** Unbedingt notwendig ist die Prüfung von zusammenhängenden Gebieten; Kompaktheit ist nur bedingt gewährleistet, ebenso Gleichverteilung, weniger zerrupfte Gebiete als bei smallestCritGetsNearest und weniger Nasen, somit besser geeignet als smallestCritGetsNearest, da weniger Umverteilung am Ende notwendig, smallestCritGetsTrueNearest scheint dennoch geeigneter

#### sum of criteria/number of locations + Distance

##### jeder Standort bekommt die Polys, die ihm am nahesten liegen, bis X erfüllt **Kriterien:** Kompaktheit + homogene Verteilung **Erkenntnis:** im Gegensatz zu „considering sum of criteria / number of locations“ von der Verteilung her eine Verbesserung zu sehen: die Gebiete weisen eine kürzere Fahrtdistanz auf, zudem sind um den Standort eher gebiet verteilt, die zu diesem gehören, ist jedoch nicht immer gewährleistet, nicht wenn um einen Standort bereits alle Gebiete verteilt sind und einem anderen Standort zugewiesen worden sind, um für diesen X zu erfüllen (siehe PLZ5 und PLZ8) ; homepoly wird nicht beachtet; Diskrepanz zwischen ersten und letzten Standort ist immer noch vorhanden; Bei den letzteren Standort kann es passieren, dass Gebiet nicht mehr zusammenhängend oder weit von Standort entfernt, da nur noch Polygone übrig, die abseits liegen --> somit wäre Umsortierung nötig, jedoch aufgrund der vielen weiteren Mängel ist der Algorithmus nicht zu empfehlen

#### Distance + criteria – rearranging

Zunächst werden die Polys nach Distanz zugeordnet, anschließend wird so lange umverteilt, bis Gleichverteilung erfüllt  
**Kriterien:** Kompaktheit + homogene Verteilung  
**Problem:** bei 4 Standorten wird Endlosscheife erzeugt, Lösung notwendig!  
**Erkenntnis**: Homogene Verteilung der Variable nur bedingt möglich; theoretisch durch Angabe eines Schwellwertes so exakt möglich, wie vom User gefordert; Wenn Diskrepanz der Werte am Anfang jedoch zu groß, endet die Umsortierung in Endlosschleifen; Kompaktheit gut gegeben, aber Gefahr der Bildung von Inseln (siehe 4 locations, PLZ5, PLZ8), somit Prüfung auf Zusammenhängigkeit notwendig; Problem der Endlosschleife wurde gelöst, indem das aktuelle Gebiet kein Teilgebiet des Gebietes nehmen darf, was im Durchlauf zuvor ein Gebiet hinzubekommen hat; Gefahr der Endlosschleife somit nicht zu 100% gebannt, aber bisher bei Tests keine Probleme, zudem wird Threshold hoch gesetzt, falls Endlosschleife  
**Fazit**: Es ist unbedingt auf zusammenhängende Gebiete zu achten

## Problems of approachs

### Performance

#### Needed improvements; warum, welche

### Requirements from the field of Geomarketing

#### Formation of holes

##### Problem + Solution

##### Auf Zusammenhang prüfen; mittel ST\_Intersects oder ST\_Touches; ST\_Touches schneller, da weniger Punkte geprüft werden müssen, in unserem Fall jedoch nicht geeinigt, da nicht immer gewährleistet, dass Gebiete richtig digitalisiert sind und bündig abschließen; Alternative: Speicherung der Nachbargeometrien zu Beginn: vorteilhafter, da so weniger Datenbankzugriffe notwendig --> Performance

#### Inhomogeneous distribution

##### Problem + Solution

#### Rearrangement/Infinite loops during rearrangment

##### Problem + Solution

#### Need of threshold values

##### Problem + Solution

# Comparison of approaches

## Performance

## Problems

## Requirements

## Conclusion

## Entscheidungsmatrix nutzen, including evaluation: Welche Planungskriterien werden beachtet und welche gäbe es noch (ausgeglichenheit, Kompaktheit, Zusammenhängend,… 🡪 weitere Beispiele siehe seite des Kit)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Algorithmus | Kompaktheit | Gleichverteilung | Rechenzeit PLZ5 in s | RZ PLZ8 | Gefahren / Probleme | Potential | Fazit |
| I JustCrit | 5 | 3 | 1,338 | 1,972 | * Diskrepanz * Flickenteppich | - | 5 |
| II sumDivNum | 5 | 3 | 1,423 | 2,092 | * Diskrepanz * Flickenteppich | Wenn Distanz mitbetrachtet | 4 |
| III JustDist | 1 | 5 | 113,218 1 min 53s | 797,723  13 min 17s | * Keine homogene Verteilung | Nachher umsortieren | 3 |
| IV CritDistInOut | 3 | 2 | 112,324  1 min 52s | 812,291 13 min 32s | * Diskrepanz * Viel Umsortieren im vgl zu V | Prüfen auf Zusammenhang, umsortierien mit Schwellwert | 3 |
| V CritDistInOutTrueNearest | 3 | 2 | 156,7545  2 min 36s | 997,299  16 min 37s | * Diskrepanz möglich * Standort liegt evtl außerhalb des Gebietes | Prüfen auf Zusammenhang, umsortieren mit Schwellwert | 3 |
| VI CritDistOutInInOut | 3 | 3 | 164,656  2 min 40s | 805,777  13min 25s | * Diskrepanz * Flickenteppich | Prüfen auf Zusammenhang, umsortieren mit Schwellwert | 3 |
| VII sumDivNumDist | 4 | 3 | 110,654  1 min 50s | 809,697  13 min 29s | * Diskrepanz * Flickenteppich |  | 4 |
| VIII DistCrit | 2 | 1 | 138,505  2 min 18s | 952,984  15 min 52s | * Standort liegt evtl außerhalb * Inseln | Schwellwert wird beachtet!; Pürfen auf Zusammenhang | 2 |

Werte von 1-5: 5 ungenügend; 4 ausreichend, 3 befriedigend, 2 gut, 1 sehr gut;

Bewertungs-/Güteformel ausdenken

GüteKompaktheit = Anzahl der Inseln / Anzahl der korrekten Gebiete <- Sollte möglichst klein sein

(GüteGleichverteilung=größterWert-kleinsterWert <- Sollte möglichst klein sein

ODER)

GüteGleichverteilung=SummeDerAbweichungenvomMittelwert/Anzahl an Werten

Gesamtgüte = (GüteKompaktheit + GüteGleichverteilung)\*Zeit

# Application of Algorithm XYZ to Geomarketing analysis

## Optimization of Areas

### Conditions and Aim

### Gegebene Daten, gegebene Standorte, algorithmus der area segmentation

### Approach of Optimization algorithm

### algorithmus der area segmentation unter Beachtung der Bedingungen fürs Geomarketing, Workflowdiagram, was wurde beachtet, wie umgesetzt

## Greenfieldanalysis

### Conditions and Aim

### Gegeben: daten, keine Standorte, Algorithmus zur Area segmentation

### Approach of Greenfield algorithm

### Workflowdiagram

## Whitespotanalysis

### Conditions and Aim

### Gegeben: daten, Standorte + Anzahl wieviel neue Standorte, Algorithmus zur Area segmentation

### Approach of Whitespotalgorithm

### Workflowdiagram

# Realworld scenario: integrated Algorithm to the mapChart Manager

Evaluation

# Discussion and Perspective

## Summary

## Limitations

## Comparison to related work

## Perspective