ERFASSUNG VON AGGLOMERATIONS- UND VERDICHTUNGSPROZESSEN MIT TECHNIKEN DES URBAN DATA MINING

Martin Behnisch, Karlsruhe und Alfred Ultsch, Marburg

Kurzfassung
Durch den Fortschritt in der Informationstechnologie und das immer rapidere Anwachsen der Datenmengen steigen die Anforderungen an Systeme, die Wissen aus Daten extrahieren und darstellen. Urbanes Data Mining wird als Methodik zur Problemlösung verstanden, um logische oder mathematische, zum Teil komplexe Beschreibungen von Mustern und Regelmäßigkeiten in Datensätzen mit Geobezug zu entdecken und Wissen aus Daten zu erzeugen. Die Techniken des Urban Data Mining werden an einem gewählten Untersuchungsszenario in ihrer Anwendung vorgestellt. Das Szenario bezieht sich auf die Erfassung und Strukturierung von bereits erfolgten Agglomerations- und Verdichtungsprozessen in Deutschland. Ausgewählte raumstrukturelle Kenngrößen werden hierzu einer Einzeluntersuchung unterzogen und auf ihre Eignung zur Klassenbildung geprüft. Angestrebt wird eine sachlich-räumliche Differenzierung des deutschen Gemeindesystems.

Gliederung

1. EINFÜHRUNG

Durch den Fortschritt in der Informationstechnologie und das immer rapidere Anwachsen von Datenmengen sind im letzten Jahrzehnt die Anforderungen an Systeme gestiegen, die Wissen aus Daten extrahieren und abbilden. In der Zukunft werden Daten und Informationen zwar erwartungsgemäß im Überfluss verfügbar sein, jedoch wird die Einbindung in Erfahrungszusammenhänge, durch welche erst Wissen geschaffen wird, deutlich schwieriger.

Data Mining trägt unterstützend zum automatischen Erzeugen und Prüfen von Hypothesen bei. Zahlreiche Methoden des Data Mining basieren auf statistischen Verfahren[16]. Der Unterschied zur explorativen Statistik besteht darin, dass die Entdeckung von neuen bzw. verborgenen Zusammenhängen in Daten und die anschließende Wissensgenerierung besondere Berücksichtigung findet[17]. Eingesetzt werden hierzu Verfahren der Wissensentdeckung (Knowledge Discovery), die (natürlich-) sprachliche Darstellungen von Wissen aus Datensammlungen ermöglichen.

Die Raum- und Stadtstruktur bietet als Forschungsgegen-stand für verschiedene wissenschaftliche Disziplinen interessante Arbeitsfelder, wobei gerade in jüngster Zeit der Erkenntnisgewinn durch transdisziplinäre Ansätze unterstützt wird. Es handelt sich um problemorientierte und von den Disziplinen unabhängige, sowohl praxis- als auch theoriebasierte Arbeitsweisen, die auf einer freien Wahl der Methodenanwendung und -entwicklung beruhen.[18]

2. PROBLEMSTELLUNG

Im Spannungsfeld einer geeigneten raumstrukturellen Abgrenzung zur Erfassung von Agglomerations- und Verdichtungsprozesse sei auf die sogenannte BOUSTEDT[19]-Systematik verwiesen, die ausgehend von den in den USA definierten Standard Metropolitan Areas (SMA) Anfang der fünfziger Jahre erarbeitet wurde. Das Modell ist für Planungszwecke und als **Instrument zur Beobachtung des Agglomerationsprozesses** bereits 1953 entwi-

[16] FAYYAD et al. (1996), siehe auch HAND / MANILA (2001)
[17] ULTSCH (2000)
[18] JAEGER et al. (1998)
[19] Vgl. BOUSTEDT (1953), BOUSTEDT (1975 a), BOUSTEDT (1975 b)

ckelt und 1971 nochmals an Lebens- und Arbeitsverhältnisse angepasst worden.

In diesem Ansatz wurde die Berufspendlerquote dazu verwendet, die Einzugsbereiche der Kernstädte zu bestimmen. Im Vergleich zu dem in den USA entwickelten Modell mit einem Kernstadtkreis und einem Umlandkreis, wird das Umland der Kernstädte weiter differenziert, um auf diese Weise ländlich geprägte Bereiche von Vorstädten und Subzentren trennen zu können. Die Gliederung erfolgt in drei Zonen: Ergänzungsgebiet, verstädterte Zone und Randzone.

Die Differenzierung basierte ursprünglich auf der Bevölkerungsdichte, jedoch wurde diese im weiteren Verlauf durch die Einwohner- und Arbeitsplatzdichte (EAD) als exaktere Messgröße ersetzt. Die engere Randzone wurde durch die Pendlerquote in das Kernstadtgebiet definiert. Die weitere Randzone wurde anhand des Anteilswertes der landwirtschaftlichen Berufstätigkeit vom Umland abgegrenzt, wobei der Schwellenwert von 50 % nicht überschritten werden durfte. Als Mindestgröße zur Definition einer Agglomeration wurde eine Einwohnerzahl von 80.000 festgelegt.

Im Jahr 2000 wurden sogenannte BIK-Regionen[20] für das wiedervereinigte Deutschland im Anschluss an eine bereits erneut erfolgte BOUSTEDT-Revision aus dem Jahr 1987 aufgebaut. Die Gemeinden werden hier über ein zielgerichtetes Pendlerverhalten analytisch an die Zentren angebunden, so dass eine fast flächendeckende Struktur von Verflechtungsgebieten unterschiedlicher Größe definiert werden kann. Berechnet wird die Pendlerquote, indem die Zielpendlerquote auf eine gemeinsame Kernstadt gemessen wird (Entscheidungskriterium: Mindestens 7 % der Wohnbevölkerung pendeln als sozialversicherungspflichtig Beschäftigte in diese Kernstadt). Die Gemeinden in einer BIK-Region werden mit einer Vier-Klassen-Systematik anhand der Einwohner- und Arbeitsplatzdichte zusätzlich gegliedert: Kernbereich, Verdichtungsbereich, Übergangsbereich, Peripherer Bereich. Die BIK-

[20] Der Begriff BIK-Regionen (BIK Aschpurwis+Behrens GmbH) ist aus den Stadtregionen nach BOUSTEDT entstanden. Auf der Basis der Volkszählung von 1987 wurde für die alten Bundesländer eine Aktualisierung des Ansatzes aus dem Jahr 1970 erarbeitet. Durch die Wiedervereinigung wurde eine erste methodische Anpassung erforderlich, und im Jahr 2000 sind die BIK-Stadtregionen für ganz Deutschland gemeindescharf nochmals überarbeitet worden. Vgl. BEHRENS / MARHENKE [1997, S. 165-186]

Regionen sind beispielsweise in der Umfragenforschung ein gebräuchliches Instrument, um auf dieser Grundlage weitere Auswertungen durchzuführen. Die Anwendung der Techniken des Urban Data Mining erfolgt vor dem Hintergrund der zuvor geschilderten Untersuchungsansätze. Den Ausgangspunkt bilden die im Jahr 2003 ausgewiesenen Oberzentren (Zentrale-Orte-Konzept). Die Oberzentren dienen zur Festlegung eines Bezugspunktes in einem noch zu definierenden Verflechtungsgebiet. Dazu werden raumstrukturelle Untersuchungsvariablen verwendet und auf ihre Eignung zur Klassenbildung geprüft. Die Klassenbildung und Wissenskonversion werden beschrieben.

Tabelle 1: Übersicht zu 6 raumstrukturellen Kenngrößen

Messgröße	Messvorschrift	Einheit
(1) Verstädterung[21]	Anteil Siedlungs- und Verkehrsfläche an Katasterfläche	[%]
(2) Nutzungsproportion[22]	Anteil Gebäude- und Freifläche an Siedlungs- und Verkehrsfläche	[%]
(3) Konzentration[23]	Einwohner und Arbeitsplätze je km² Gebäude- und Freifläche	[Personen je km²]
(4) Entdichtung[24]	Anteil Ein- / Zweifamilienhäuser am Wohnbaubestand	[%]
(5) Beschäftigungsdisparität[25]	Quotient von sozialversicherungspflichtig Beschäftigten am Arbeitsort und sozialversicherungspflichtig Beschäftigten am Wohnort*100	[dimensionslos]
(6) Erreichbarkeit[26]	Fahrzeit zum nächsten Oberzentrum (PKW)	[Minuten]

Quelle: Bundesamt für Bauwesen und Raumordnung (BBR)

3. DATEN

Um Aussagen über die Raumstruktur zu treffen, ist eine räumlich-differenzierte Betrachtung anzustreben. Die Problematik besteht darin, dass mit zunehmender räumlicher Auflösung das Datenangebot eigenständig im-

[21] Vgl. ARLT et al. (2001, S.5)
[22] Vgl. Laufende Raumbeobachtungen, BBR (2006): Flächenerhebung nach Art der tatsächlichen Nutzung
[23] Vgl. STAACK (1995, S. 128), SIEDENTOP et al. (2005, S. 77), BEHRENS / MARHENKE (1997)
[24] Vgl. SIEDENTOP et al. (2005, S. 76)
[25] Vgl. SIEDENTOP et al. (2003, S. 102) und (2005, S. 79)
[26] Es ist darauf hinzuweisen, dass es sich um verbandsgemeindebezogene Daten handelt, so dass eine gewisse Unschärfe bei diesen Daten vorliegt (siehe BECHER (1995, S.117), SIEDENTOP et al. (2005, S. 81)).

mer schwieriger in großem Umfang beschaffbar wird. Unter dem Aspekt der Erfassung und Strukturierung von Agglomerations- und Verdichtungseigenschaften, werden sechs noch leicht zugängliche raumstrukturelle Kenngrößen herangezogen. Diese sind in Tabelle 1 durch die Messvorschrift charakterisiert und beziehen sich auf die Gesamtmenge der 12430 Gemeinden in Deutschland im Gebietsstand 2004.

4. METHODIK

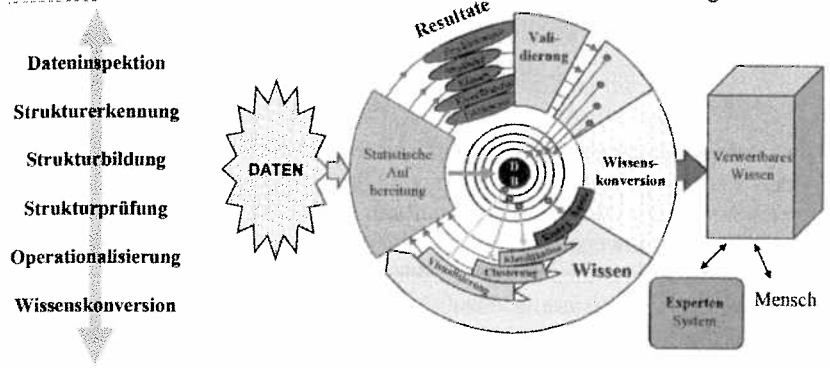
Der Begriff des ‚Urban Data Mining'[27] charakterisiert die Erarbeitung einer für den urbanen Kontext entwickelten Methodik, die dazu dient, logische oder mathematische und zum Teil komplexe Beschreibungen von Mustern und Regelmäßigkeiten in Datensätzen zu entdecken sowie daraus Erkenntnisse abzuleiten und zu bewerten.[28] Wesentliche Verfahrensschritte sind die Dateninspektion, die Strukturerkennung, die Strukturbildung, die Strukturprüfung, die Operationalisierung und die Wissenskonversion. Beschrieben wird ein zyklischer Prozess (Abbildung 1), so dass schrittweise gewonnene Erkenntnisse validiert und als Eingangsstufe der Folgeschritte verwendet werden.

Der zyklische Prozess des ‚Urban Data Mining' berücksichtigt die Anwendung von geeigneten Methoden auf einen Datenbestand mit dem Ziel der Wissensentdeckung. Der Unterschied zur Statistik besteht darin, dass zu Beginn nur Daten gegeben sind, es ist kein Modell vorhanden, und die Hypothesen werden noch gesucht. Aufgabe des Data Mining ist die Entdeckung von neuen und / oder verborgenen Zusammenhängen in Daten. Ausgehend von einer Menge von Zahlen wird eine Darstellung von bislang unbekannten Sachverhalten in möglichst natürlichsprachiger Form angestrebt. Wissen wird aus Datensammlungen extrahiert. Die Darstellung des gewonnenen Wissens soll zugleich auch maschinell verarbeitbar sein. In der Regel geschieht die maschinelle Verarbeitung in Form von wissensbasierten Systemen[29].

[27] Der Begriff des Urban Data Mining wurde vom Autor definiert. Das Ziel besteht darin, auf der Basis von Beobachtungen bzw. von Messergebnissen, den Übergang von Daten zu Wissen zu entwickeln und dadurch für den Menschen verwertbare Wissenszusammenhänge zu erzeugen. Es wird auf bereits allgemein anerkannte Methoden des Data Mining zurückgegriffen und zusätzlich der Einsatz von üblichen GI-Werkzeugen zur Verarbeitung von Geoinformationen (Spatial Mining, Geocomputation) berücksichtigt. Ein wesentliches Kennzeichen empirisch-analytischer Theorien ist die empirische Testbarkeit.

[28] BEHNISCH (2007 a)

Abbildung 1: Zyklischer Prozess im ‚Urban Data Mining'



Quelle: In Anlehnung an Ultsch (2006)

STREICH[30] definiert: „Wissen ist die intellektuelle Vernetzung von Informations-‚atomen' bzw. Einzeltatsachen zu komplexen Kenntnisstrukturen auf der Grundlage von Erfahrungstatbeständen und / oder Lernvorgängen von Einzelsubjekten oder Gruppen. Informationen bestehen aus sinnvoll strukturierten Daten, Daten wiederum sind die ‚atomaren' Bausteine für Informationen."

WILKE[31] setzt bei der Wissensarbeit voraus,: „[...] dass das relevante Wissen (1) kontinuierlich revidiert, (2) permanent als verbesserungsfähig angesehen, (3) prinzipiell nicht als Wahrheit, sondern als Ressource betrachtet wird und (4) untrennbar mit Nichtwissen gekoppelt ist, so dass mit Wissensarbeit spezifische Risiken verbunden sind."

[29] Vgl. ALTENKRÜGER / BÜTTNER (1992)

[30] Vgl. STREICH (2005, S. 17 ff.)

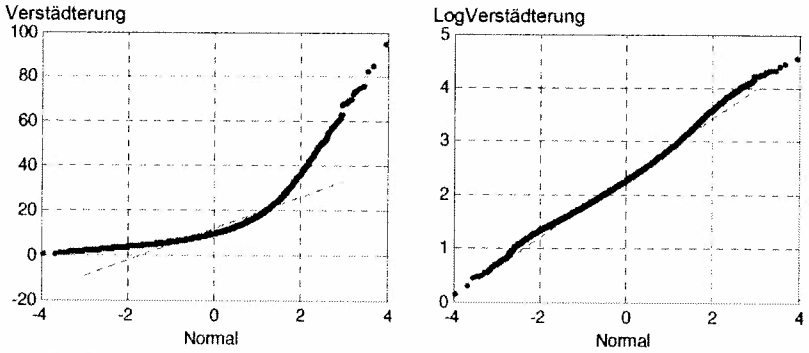
[31] Vgl. WILKE (1998, S. 161-177), siehe auch SCHLEGEL, F.: „Je mehr man schon weiß, je mehr hat man noch zu lernen. Mit dem Wissen nimmt das Nichtwissen in gleichem Grade zu, oder vielmehr das Wissen des Nichtwissens."

Um den Grundvoraussetzungen üblicher Data Mining Methoden zu entsprechen, ist zu Beginn eine umfassende **Dateninspektion** einschließlich Vorverarbeitung erforderlich.

ULTSCH[32] verweist darauf, dass oftmals Verfahren im Data Mining die behandelten Variablen als normal verteilt bzw. einen ähnlichen Verteilungsverlauf voraussetzen, so dass die Variablenvorbehandlung eine wichtige Rolle einnimmt.

Die Dateninspektion beginnt mit der Durchsicht der Objektdaten jeder einzelnen Variablen, indem man sich einen Überblick von Anzahl, Art, Wertebereichen und insbesondere der Verteilung verschafft. Da die Verteilung der Variablen üblicherweise nicht vorab bekannt ist, besteht die Aufgabe darin, eine Hypothese über eine empirisch beobachtete Variable zu gewinnen. Geeignet ist die Visualisierung dazugehöriger Sachverhalte, die neben Lage- und Streuungsmaßen[33] Variablenbeschreibungen ergänzen (z.B. Histogramme, Box-Plots, Quantil/Quantil-Plots (QQ-Plots), Pareto Density Estimation[34], Modellierung mit Gauß-Mixturen).

Abbildung 2: QQ-Plots zur Messgröße ‚Verstädterung'



Quelle: Eigene Bearbeitung

[32] Vgl. ULTSCH (2006), siehe zusätzlich auch bei ERB (1990, S. 57 ff.)

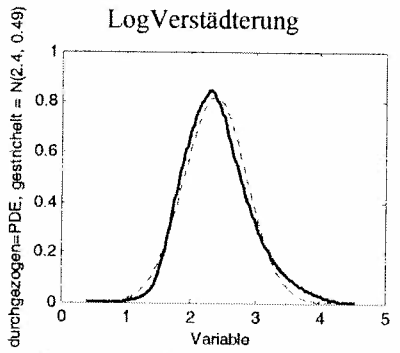
[33] Vgl. HARTUNG (2005, S. 31 ff. und S. 40 ff.)

[34] Vgl. ULTSCH (2001, 2003): Der PDE-Plot berücksichtigt das Pareto Gesetz (80/20-Regel) und sogenannte Pareto-Kugeln. Es handelt sich um empirische Schätzung der Dichte von Daten anhand der informationsoptimalen Menge.

QQ-Plots dienen dazu, eine vorgelegte Verteilung grafisch mit einer standardisierten Verteilung, z.B. Normalverteilung oder Gleichverteilung, zu vergleichen. Bilden die so entstandenen Punkte annähernd eine Gerade, so kann davon ausgegangen werden, dass die beiden Verteilungen gleich sind. Werden Abweichungen von Standardnormalverteilungen festgestellt, so gilt es möglicherweise entsprechende Umformungen (Transformationen) festzulegen, mit denen die Daten in eine bekannte Verteilung transformiert werden können. Die sogenannte ‚ladder of power'[35] ist eine Auflistung für die Größe (power) des Exponenten p zu $y = x^p$. Im Umkehrschluss kann aus dieser Transformation auf die empirische Verteilung geschlossen werden.

Im Folgenden wird für die Messgröße ‚Verstädterung' eine Verteilungsuntersuchung beispielhaft umgesetzt. Abbildung 2 zeigt die QQ-Plots für diese Messgröße. Es werden die Quantile der Variable auf der Y-Achse aufgetragen. Damit ist an dieser Achse das Ablesen des Wertebereichs möglich.

Abbildung 3: PDE-Plot der Messgröße ‚LogVerstädterung'



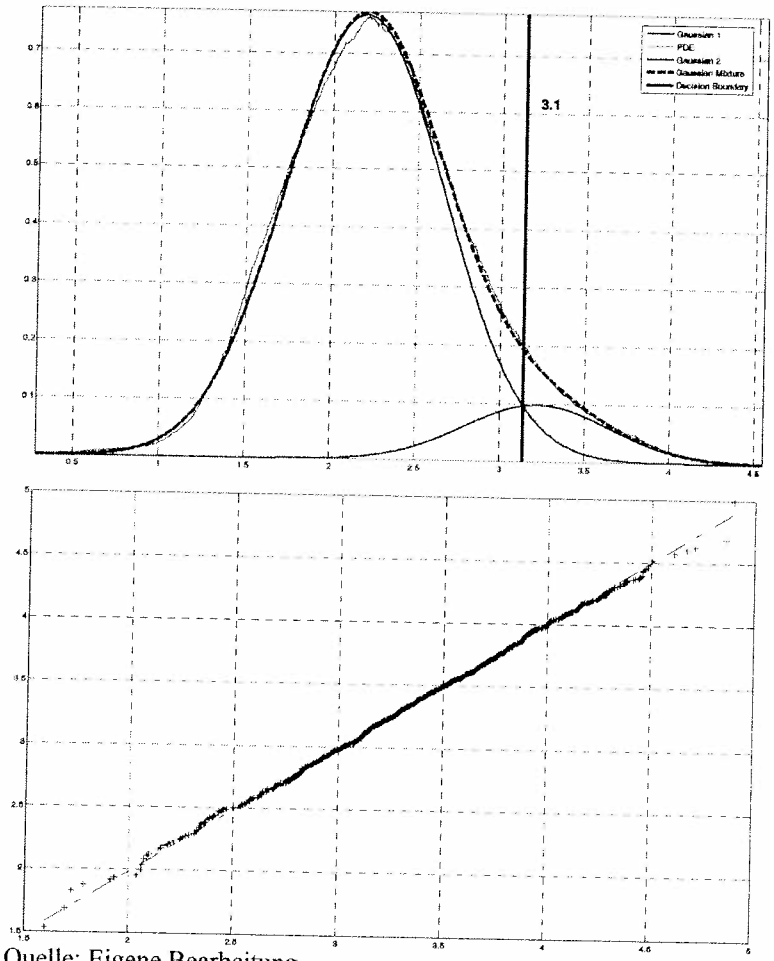
Quelle: Eigene Bearbeitung

Der QQ-Plot der Ausgangsgröße zeigt einen konkaven Bogen, wobei dies auf eine nichtnormale, „schiefe" Verteilung hindeutet. Um eine schiefe Verteilung dennoch charakterisieren zu können, kann eine nichtlineare Transformation angewendet werden. Eine Transformation auf annähernde Normalverteilung kann bei dieser Variablen durch Logarithmieren erreicht wer-

[35] Vgl. HARTUNG (2005, S. 833)

den. In Teilbereichen folgen die logarithmierten Daten gemäß der rechten Abbildung einer Geraden, doch sind zusätzliche Unebenheiten vorhanden.

Abbildung 4: GMM und dazugehöriger QQ-Plot (‚LogVerstädterung')



Quelle: Eigene Bearbeitung

Die Verteilungsuntersuchung der Messgröße ‚Verstädterung' wird ergänzt durch den PDE-Plot. Es handelt sich um einen Kerndichteschätzer, der die Wahrscheinlichkeitsdichte schätzt. Abbildung 3 zeigt, dass die PDE der logarithmierten empirischen Messwerte (durchgezogene Kurve) und der Normalverteilung (gestrichelte Kurve) sich annähern, jedoch keine deutliche Überdeckung vorhanden ist.

Die Prüfung auf Log-Normalverteilung hat ergeben, dass für die Variable ‚Verstädterung' eine komplexere Modellierung sinnvoll wäre. Eine genauere Approximation ist mit einer Gaußschen Mischverteilung (GMM) möglich. Die Verteilungsdichte eines Variablenvektors x lässt sich meistens nur ungenau mit Hilfe einer einzelnen Normal- bzw. Gaußverteilung beschreiben. Zur Schätzung der Parameter eines GMMs ist der Expectation-Maximization (EM) -Algorithmus einsetzbar[36]. So gefundene Lösungen sind besonders von den Initialisierungsparametern abhängig, so dass die Ergebnisse mehrfach zu berechnen sind. Als Gütekriterium eignet sich die Pareto Dichte-Schätzung[37].

Gemäß Abbildung 4 lässt sich die gegebene Verteilung mit zwei Gauß-Mixturen gut modellieren. Durch den Schnittpunkt der Kurvenverläufe der einzelnen Gaußverteilungen (Modus) wird die sogenannte Bayes'sche Entscheidungsgrenze gebildet. Diese liegt hier bei 20% (siehe Log(3,1)) und dient dazu, die Objekte aufgrund der Variablenausprägungen in zwei Klassen einzuteilen. Der QQ-Plot des aufgestellten GMMs folgt deutlich einer Geraden.

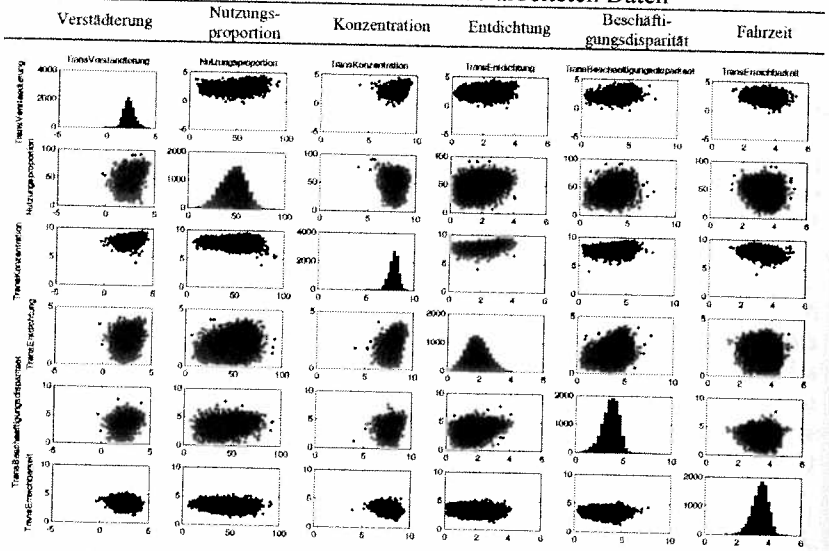
Ist die Beschreibung von einzelnen Variablen erfolgt, beginnt nachfolgend die Untersuchung der Datensätze auf Zusammenhänge bzw. Abhängigkeiten zwischen zwei oder mehreren Variablen. Es wird geprüft, ob redundante Information in den Datensätzen existiert und Hinweise auf die Struktur des Datensatzes bzw. des Grundproblems zu erkennen sind. Hierzu sind einerseits visuelle Methoden wie Streu-Diagramme (Scatter-Plots) und andererseits statistische Maßzahlen wie Korrelationsmaße[38] einsetzbar, die den Abhängigkeitsgrad der Variablen messen. Tabelle 1 vermittelt am gegebenen

[36] Vgl. HAND et al. (2001, S. 281)
[37] SCOTT, D.W. (1992), ULTSCH (2003)
[38] Vgl. HARTUNG (2005, S. 72 ff.), Pearsonscher Korrelationskoeffizient, Spearmans Rangkorrelationskoeffizient, Kendalls Rangkorrelationskoeffizient.

Untersuchungsfall eine praktische Vorstellung von Variablenzusammenhängen (Scatter-Plot). Die Histogramme verdeutlichen den einheitlichen Verteilungsverlauf infolge zuvor gewählter Transformationen.

VOGEL[39] verweist darauf, dass Korrelationen innerhalb der Merkmalsstruktur unvermeidbar sind, doch kann verschieden darauf reagiert werden (z.B. durch Variableneliminierung, Gewichtungsschema[40] oder Faktorenanalyse[41]).

Tabelle 1: Scatter-Plot zu vorverarbeiteten Daten



Quelle: Eigene Bearbeitung

Um Objekte mehrdimensional miteinander vergleichen zu können, ist es notwendig, ein quantifizierbares Maß zu verwenden, welches Prinzipien zur Bestimmung der Gleichheit, Ähnlichkeit bzw. Verschiedenheit berücksichtigt. Bei der Festlegung einer geeigneten Metrik sind die Erkenntnisse der Datensichtung und damit die Grundeigenschaften der Daten mit einzubeziehen. Im Sinne der Vergleichbarkeit von Daten ist eine Entscheidung zum Umgang mit Ausreißern, d.h. Objekte mit extremer Werteausprägung als auch der Fehlstellenbehandlung zu treffen. Da in den meisten Fällen die Variablen nicht in gleicher Dimension vorliegen, sind die Variablen geeignet zu skalieren, z.B. durch Normierung, Standardisierung oder Lineare Transformation. Die Wahl des Ähnlichkeits- bzw. Distanzmaßes ist von entscheidender Bedeutung, da über Ähnlichkeit[42] bzw. Unähnlichkeit von Objekten entschieden wird. Zum Zweck der Validierung sollte die Angabe von besonders ähnlichen / unähnlichen Objekten erfolgen.

[39] Vgl. VOGEL (1975, S.52 ff.)

[40] Vgl. FISCHER, M. (1982, S.54): „Hat man eine gewisse Anzahl relevanter und sinnvoller Attribute ausgewählt, deren Entdeckung und Formulierung sicherlich wissenschaftliche Kreativität erfordert, so muss man sich entscheiden, ob und gegebenenfalls wie man die einzelnen Attribute gewichtet (externes Gewichtungsproblem)."

[41] Vgl. ÜBERLA (1971, S. 155)

Der nächste wichtige Schritt im Data Mining ist die **Strukturerkennung**, d.h. hochdimensionale Daten sind für einen menschlichen Betrachter geeignet darzustellen, wobei dies in der Regel in graphischer Form geschieht. Einsetzbar sind neben Leiterdiagrammen insbesondere die Projektionsverfahren, welche die wesentlichen Eigenschaften eines Datenbestandes aus dem zunächst hochdimensionalen und uneinsehbaren Ursprungsraum in einen darstellbaren zwei- oder dreidimensionalen Repräsentationsraum übertragen.

Es ist zwischen linearen[43] und nichtlinearen Projektionsmethoden zu trennen, wobei gerade die nichtlinearen Projektionen sich dazu eignen, strukturelle Eigenschaften (räumliche Beziehungen, Nachbarschaftsverhältnisse) der Daten abzubilden. Genannt seien die Multidimensionale Skalierung, Sammons-Abbildungen und insbesondere die Merkmalskarten, die den neuronalen Netzen zuzuordnen sind.[44]

Selbstorganisierende Merkmalskarten nach KOHONEN[45] (SOM) sind mit ihrem unüberwachten Lernverfahren geeignet, die inhärenten Strukturen des meist hochdimensionalen Eingaberaums auf einen 2-dimensionalen Raum zu projizieren. Übliche SOM sind durch eine geringe Anzahl Neuronen charakterisiert. Verwendet man eine sehr große Anzahl Neuronen, so ist es mög-

[42] Vgl. EVERITT (1980, S. 19) oder BOCK (1974, S. 44 ff. und S. 77)

[43] Lineare Projektionen dienen eher der Dimensionsreduktion, um z.B. im weiteren Vorgehen eine geringere Menge von Merkmalen betrachten zu müssen.

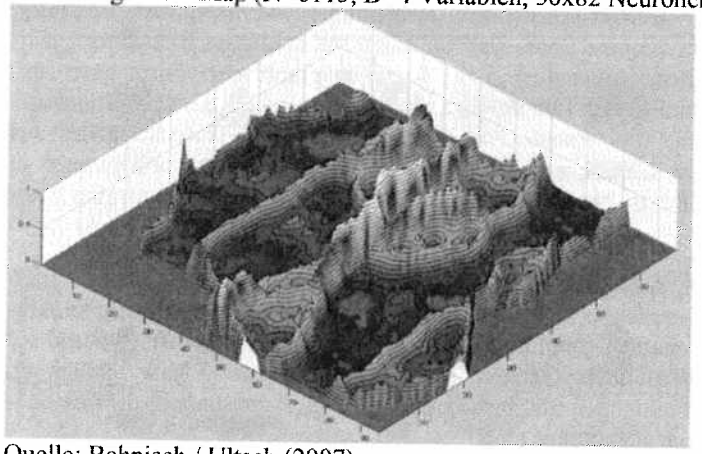
[44] Verwiesen wird auf die Untersuchung eines Datensatzes zu zwei ineinander geschachtelten Toroedern, der eine nichtlineare Entflechtung erfordert, vgl. ULTSCH (2005).

[45] Vgl. Kohonen (1982)

lich, Strukturen in der Merkmalskarte durch Emergenz abzubilden. Die emergenten Selbstorganisierenden Merkmalskarten[46] (ESOM) wurden entwickelt, um eine Strukturerkennung in den Daten durch eine 3-dimensionale Landschaftsdarstellung zu ermöglichen. Zusammengehörende Daten liegen in Tälern, während unterschiedliche Bereiche (Klassen) durch Mauern oder Gebirgszüge getrennt werden.

Die Abbildung 5 zeigt eine Strukturerkennung in Form einer U*-Map (Inseldarstellung). Es handelt sich um Daten aus einer anderen Bearbeitung, die eine hochdimensionale Struktur repräsentieren und zu klar unterscheidbaren Clustern führen (siehe im Gegensatz dazu Abbildung 8).

Abbildung 5: U*-Map (N=8113, D=4 Variablen, 50x82 Neuronen)



Quelle: Behnisch / Ultsch (2007)

Die Klassifikation bildet ein wesentliches Instrument im ‚Urban Data Mining'. Es sind drei Aufgabenstellungen zu nennen, die als Klassifizierungsprobleme[47] existieren.

I. Aufgliederungsproblem (Klassifikation, im engeren Sinn)
Es liegen keine Informationen über Gruppen in der Gesamtheit vor. Die Aufgabe besteht darin, die Gesamtheit oder Stichproben aus der Gesamtheit

[46] Vgl. ULTSCH (1999): Methoden sind die U-Matrix, P-Matrix oder U*-Matrix.
[47] Vgl. VOGEL (1975, S. 3-5), DEICHSEL / TRAMPISCH (1985, S. VII)

in eine zunächst unbekannte Anzahl möglichst homogener und einander möglichst ungleichartiger Gruppen zu zerlegen. Die zu lösenden Hauptprobleme bestehen in der Bestimmung der Anzahl der Gruppen und in der Zuordnung der Einheiten zu diesen Gruppen. Die Datenmatrix bildet die Informationsquelle.

II. Schichtungsproblem (Mischform)
Zu bestimmen ist eine Untergliederung der Gesamtheit in Teilgesamtheiten (Schichten) mit a priori vorbestimmter Anzahl. Die Datenmatrix ermöglicht eine Bearbeitung und wird zusätzlich durch das Wissen über Anzahl und Eigenschaften der Schichten ergänzt.

III. Zuordnungsproblem (Diskriminanzproblem)
Es liegen Informationen über Anzahl und Eigenschaften von Teilgesamtheiten a priori vor. Die Aufgabe besteht darin, den schon bekannten Teilgesamtheiten die aus einer Grundgesamtheit entnommenen Einheiten anhand ihrer Variablenwerte mit möglichst großer Sicherheit zuzuordnen. Die bereits definierten Teilgesamtheiten und die Datenmatrix unterstützen die Vorgehensweise.

Die Ergebnisse eines Klassifikationsvorganges eignen sich zur Entwicklung von Maßstäben und Bewertungsskalen. Bei der Klassifikation im engeren Sinne handelt es sich um induktive Verallgemeinerungen über die Objekte, indem ein gemeinsamer Begriff (Semantik) gefunden wird.

Mit Beginn der 70er-Jahre begann sich die Clusteranalyse durch die Möglichkeit des vermehrten Computereinsatzes als eine eigenständige Analyseform zu etablieren. Ihre Entwicklung verdankt die Clusteranalyse dem Wunsch, den Klassifikationsprozess systematisch und quantitativ erfassen zu wollen und durch Berücksichtigung numerischer Kriterien die Güte von Gruppierungen „objektiv" zu vergleichen. Es handelt sich um Verfahren, die sich auf Objekte stützen und als Technik geeignet sind, diese Objektmenge, von der meist zunächst keine Gruppenstruktur bekannt ist, in homogene Teilmengen zu zerlegen, d.h. im Ergebnis eine Konfiguration von Clustern (**Strukturbildung**).

Die Verfahren der Clusteranalyse[48] sind außerordentlich zahlreich und können nach verschiedenen Kriterien systematisiert werden. Es erfolgt die Durchmusterung und Auswertung von entweder Ähnlichkeits- oder Distanzmatrizen. Die deterministischen Clusteranalyseverfahren berechnen Cluster, so dass Klassifikationsobjekte mit einem Grad der Zuordnung von 0 oder 1 einem Cluster zugehören. LOFTI ZADEH[49] veröffentlicht 1965 das Konzept der unscharfen Mengen und schafft einen Ansatz zum Umgang mit Vagheit und legt den Grundstein für unscharfe Verfahren.

Die Idee der dichtebasierten Clusteranalyse[50] besteht darin, Regionen im Merkmalsraum zu bestimmen, die eine hohe Anzahl von Objekten (also eine hohe Dichte) aufweisen. Weiterhin sollen diese Regionen durch Bereiche mit einer kleinen Anzahl von Objekten (geringe Dichte) abgegrenzt sein. Das Auffinden von dichtebasierten Clustern erfolgt durch Überprüfung aller Objekte hinsichtlich ihrer Eigenschaft als sogenanntes ‚Kernobjekt' und der Beurteilung der ‚Dichte-Erreichbarkeit'. Ein Cluster wächst, solange die Dichte von Objekten in ihrer Nachbarschaft einen Schwellwert überschreitet. Bei der Erkennung der Cluster ist die Form der Cluster unerheblich.

Im Kontext der Klassifizierung sei zusätzlich auf Gaußsche Mixtur-Modelle (Gaussian Mixture Models, GMM)[51] verwiesen. GMMs sind eng mit dem Bayes'schen Klassifizierer verwandt und gelten als ein probabilistisches Modell für multivariate Wahrscheinlichkeitsdichten[52].

Die **Strukturprüfung** ermöglicht die Validierung von gefundenen Clusterungen. Es sei bemerkt, dass die recht ungenau beschriebene Zielsetzung der Clusteranalyse, Cluster möglichst ähnlicher Objekte zu bilden, unterschiedliche Interpretationen zulässt. Denn es ist nicht zu unterscheiden zwischen richtigen oder falschen Gruppierungen, sondern vielmehr im Sinne der jeweiligen Anwendung nach brauchbaren bzw. unbrauchbaren Lösungen.

Eine Möglichkeit der Strukturprüfung basiert auf einem Vergleich der Ergebnisdaten einer Clusterung mit einer bereits vorab a priori bekannten

[48] Vgl. SOKAL / SNEATH (1963), BOCK (1974), SPÄTH (1975), HÖPPNER et al. (1999) und DEIMER (1986)
[49] Vgl. ZADEH (1965)
[50] Vgl. SANDER (1999) und ULTSCH (2005)
[51] THINH, BEHNISCH und ULTSCH (2006) – Anwendung (räumlich orientiert).
[52] Vgl. REYNOLDS et al. (2000)

Klassifizierung. Als Maßzahlen für die Qualitätsprüfung im Sinne einer Übereinstimmung zwischen einer bekannten Klassifikation und zusätzlich ermittelten Clusterung eignen sich die Sensitivität, Spezifität, Akkuratheit und die sogenannte ROC-Kurve (Receiver-Operation-Characteristic).

Eine weitere Möglichkeit der Strukturprüfung besteht darin, unabhängig von einer bereits a priori bekannten Klassifizierung die Homogenität bzw. Heterogenität innerhalb des Clusters und über die Clustergrenzen hinweg zu bestimmen.

Mit Blick auf andere multivariate Verfahren eignet sich die Diskriminanzanalyse[53] dazu, anhand einer Clusterstruktur z.B. die Unterschiede zwischen Clustern hinsichtlich vorgegebener Variablen zu analysieren oder die Trennkraft der Variablen zu ermitteln. Die Regressionsanalyse[54] verfolgt das Ziel, ein funktionales Modell zwischen einer abhängigen (erklärte Variable) und einer oder mehreren unabhängigen Variablen (erklärende Variablen) zu finden.

Die **Operationalisierung** bildet die Grundlage für die nachfolgende Wissenskonversion und ermöglicht insbesondere die Wissensgewinnung aus bestehenden Daten unter Einbeziehung bereits entdeckter Strukturen. Es werden Zuordnungsvorschriften gesucht, die die gewonnenen Klassifikationen charakterisieren und darüber hinaus eine nachträgliche Zuordnung von nicht klassifizierten Daten realisieren.

Im Kontext des Data Mining bilden sogenannte Klassifikatoren eine Quelle zur Wissensdarstellung. Es ist zwischen subsymbolischen[55] und symbolischen[56] Klassifikatoren zu unterscheiden. Während ein subsymbolischer Klassifikator die Aufgabe ohne ein genaues Verständnis der Klassen erle-

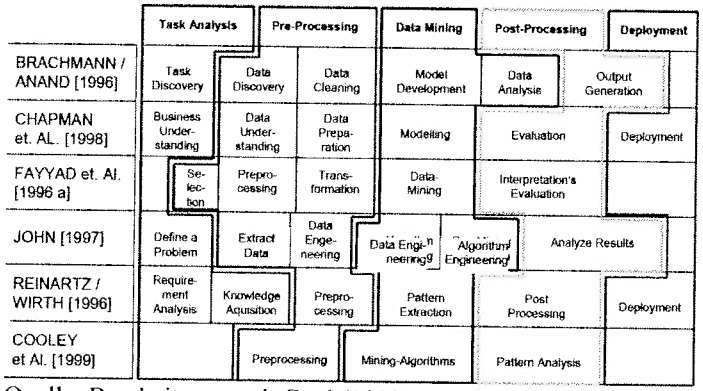
[53] Vgl. BAHRENBERG (2003, S. 318) und ERB (1990, S.9/10)
[54] Vgl. BACKHAUS (2006, S. 49)
[55] Zu subsymbolischen Klassifikatoren zählen Nearest-Neighbour-Klassifikatoren, aber auch künstliche neuronale Netze. Das Konzept der Fuzzy Pattern Klassifikation (vgl. BOCKLISCH (1987)) wird hier ebenso eingestuft.
[56] Zu symbolischen Klassifikatoren zählen Bayes'sche Klassifikatoren und Klassifikatoren, die mit Entscheidungsbäumen oder Entscheidungsregeln arbeiten. Genannt seien die folgenden Algorithmen: Classification and Regression Trees (CART, vgl. BREIMAN (1984)), Interactive Dichtomizer 3 (ID3, vgl. QUINLAN (1986)) und SIG* (Signifikanz der merkmalsbasierten Klassenbeschreibung, vgl. ULTSCH (1991)). Die Generierung der Entscheidungsbäume und auch von Entscheidungsregeln wird in der Disziplin der künstlichen Intelligenz unter dem Stichwort des sog. Maschinellen Lernens erforscht.

digt, stellt ein symbolischer Klassifikator die Anforderung einer nahezu natürlichsprachigen Beschreibungsform an die einzusetzenden Algorithmen. Gerade symbolische Klassifikatoren tragen dazu bei, dass der Mensch ein Verständnis für Klassen gewinnt und eine Abstraktion von Klassen möglich wird.

Der Vorteil der genannten Algorithmen CART und ID3 besteht in der Möglichkeit, die Entscheidungen für eine Klasse graphisch als Entscheidungsbaum darzustellen. Der Algorithmus sig* zielt auf das Verstehen eines Clusters anhand ausgewiesener Regeln ab (Diagnose) und legt dabei weniger Wert auf die Hierarchie der Entscheidungen.

Bei der Bestimmung der Güte von Klassifikatoren ist nicht nur die Klassifikationsleistung im Vergleich zu gegebenen Klassifikationen zu untersuchen, sondern auch die Fähigkeit der Klassifikatoren, neue Datensätze einzuordnen. In der Regel werden drei Datenmengen gebildet: Lern- und Testdatensatz sowie ein Validierungsdatensatz.

Abbildung 6: Hauptschritte in Prozessmodellen des KDD



Quelle: Bearbeitung nach Gaul (1998)

Die genannten Techniken aus dem Gebiet der künstlichen Intelligenz fördern die Wissensdarstellung und unterstützen bei Bedarf zusätzlich die Nutzung von Wissen in maschinellen Systemen und eignen sich z.B. für Monitoring-

bzw. Diagnosesysteme[57]. Diese müssen in der Lage sein, die von ihnen getroffenen Diagnosen nicht nur zu treffen, sondern sie auch zu begründen. Es werden Schätzkalküle verwendet, die zur Abbildung von unvollständigem, widersprüchlichem oder annäherndem Wissen verwendet werden.

Tabelle 2: Konversionsarten und Methoden

Sozialisation	**Observieren** (Beobachten z.B. eines Experten), **Imitieren** (Nachahmung der Handlung z.B. eines Experten), **Praktizieren** (Überführung der theoretischen Grundlagen in praktische Erfahrung) und **Kommunizieren** (direkte verbale Vermittlung von Wissen)
Externalisierung	**Reflektieren** (individuelle Konzentration auf Ideen, die Arbeit und die damit verbundene Ex-plizierung des Wissens), **Metapherbildung** (lebendige, anschauliche Versprachlichung von Zusammenhängen), **Analogiebildung** (Aufzeigen funktionaler Gemeinsamkeiten zwischen getrennten Wissensgebieten und direkter Transfer) und **Modellbildung** (komplexe Zusammenhänge problemspezifisch vereinfachen und strukturieren)
Kombination	**Sortieren** (Neuanordnung), **Hinzufügen** (Entstehung), **Vereinigen** (Zusammenfügen), **Aggregieren** (Ansammlung), **Selektieren** (Auswahl), **Kategorisieren / Klassifizieren** (Generierung) und **Rekombinieren** (Erzeugung aus dem Bestand)
Internalisierung	Prüfendes und vergleichendes Nachdenken über **Lesen** (Texten), **Sehen** (Bilder bzw. Grafiken) und **Hören** sowie vereinzelt **Tasten** und **Riechen**

Quelle: Eigene Bearbeitung unter Verwendung von Nonaka / Takeuchi (1997), Preece et al. (1994), Schreiber et al. (2000), Hyttinen (2004, S.14 ff.)

Abbildung 6 zeigt die Einordnung von Arbeitsschritten des Data Mining in bekannte Prozessmodelle des ‚Knowledge Discovery in Databases' (KDD).

[57] Relevanz: „Aktion Demographischer Wandel", http://www.aktion2050.de

GAUL[58] stellt in Zusammenhang mit Data Mining fest, dass viele Werkzeuge davon nicht über eine Möglichkeit der **Wissenskonversion** verfügen. In der Tabelle 2 sind Arten der Wissenskonversion im Allgemeinen und dazugehörige Methoden zusammenfassend aufgeführt.

Tabelle 3: Datenaufbereitung der raumstrukturellen Variablen

Messgröße	grob inspiziert	GMM (Grenzen)	Regel	Klassengröße
(1) Verstädterung	Log(Data) folgt Normalverteilung	Bimodal, 2 Gauß-verteilungen, Grenze: 20 %, Log (Data): 3.1	Klasse 1: Data ≤20 Klasse 2: Data >20	Klasse 1, [11029] Anteil: 88,73 Klasse 2, [1401] Anteil: 11,27 %
(2) Nutzungs-proportion	Ohne Transformation	Bimodale Verteilung, 2 Gauß-Verteilungen, Grenze: 40 %	Klasse 1: Data ≤40 Klasse 2: Data >40	Klasse 1, [3938] Anteil: 31,68 Klasse 2, [8492] Anteil: 68,32
(3) Konzentration	Log(Data) folgt Normalverteilung	Multimodal, 3 Gaußverteilungen, Grenzen: 2500, 4000 Log(Data): 7.9, 8.3	Klasse 1: Data ≤2500 Klasse 2: 2500<Data ≤4000 Klasse 3: Data >4000	Klasse 1, [5263] Anteil: 42,34 % Klasse 2, [4802] Anteil: 38,63 % Klasse 3, [2365] Anteil: 19,03 %
(4) Entdichtung	Umkehransatz der Variablendaten: y=log(100-Data) Log((100-Data)+1) folgt Normalverteilung	Multimodal, 3 Gaußverteilungen, Grenze: 2 %, 15 % Log(Data): 1,1 und 2,8 invertierte Grenze: 98 %, 80 %	Klasse 1: Data <85 Klasse 2: 85≤Data<98 Klasse 3: Data ≥98	Klasse 1, [1420] Anteil: 11,42 % Klasse 2, [8869] Anteil: 71,35 % Klasse 3, [2141] Anteil 17,22 % 100 % Einfamilienhäuser [518]
(5) Beschäftigungs-disparität	Log(Data+1) folgt Normalverteilung	Multimodal, Drei Gaußverteilungen, sachlogisch erzwungen: 100	Klasse 1: Data <100 Klasse 2: Data ≥100	Klasse 1, [10808] Anteil: 86,95 % Klasse 2, [1622] Anteil: 13,05 %
(6) Erreichbarkeit	Log(Data+1) folgt Normalverteilung	Bimodal, 2 Gauß-Verteilungen Grenze: 30 Min.. Log (Data): 3.3	Klasse 1: Data=0 Klasse 2: 0<Data ≤30 Klasse 3: Data > 30 Klasse 0: Data ='NAN'	Klasse 1, [133] Anteil: 1,07 % Klasse 2, [5092] Anteil: 40,96 % Klasse 3, [6964] Anteil: 56,02 % NAN: [241]

Quelle: Eigene Bearbeitung

[58] GAUL, W. (1998, S.145 ff.)

5. DATENAUFBEREITUNG

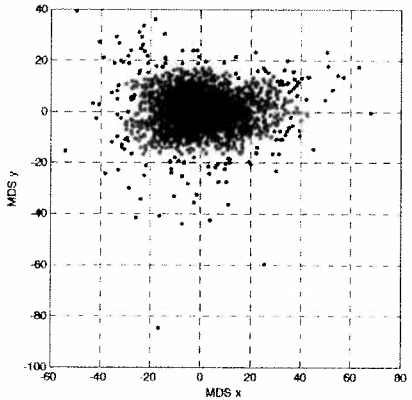
Im Kontext von Agglomerations- und Verdichtungsprozessen werden für das Untersuchungsjahr 2004 sechs raumstrukturelle Kenngrößen der Einzeluntersuchung unterzogen. Tabelle 3 enthält Ergebnisse der Datenaufbereitung.

6 KLASSENBILDUNG

Auf Grundlage der zuvor untersuchten raumstrukturellen Variablen wird geprüft, ob und in welcher Weise eine Klassenbildung überhaupt plausibel und sinnvoll ist. Die Objekte einer Klasse sollen sich dabei möglichst ähnlich sein und zu Objekten anderer Klassen deutliche Unterschiede aufweisen.

Abbildung 7 zeigt die Anwendung der mehrdimensionalen Skalierung und verweist auf die Schwierigkeit, klar unterscheidbare Gruppen mit Hilfe von Cluster-Algorithmen mit diesen Variablen bilden zu können. Es zeichnet sich eine deutliche Punktwolke ab.

Abbildung 7: MDS mit 6 raumstrukturellen Kenngrößen



Quelle: Eigene Bearbeitung

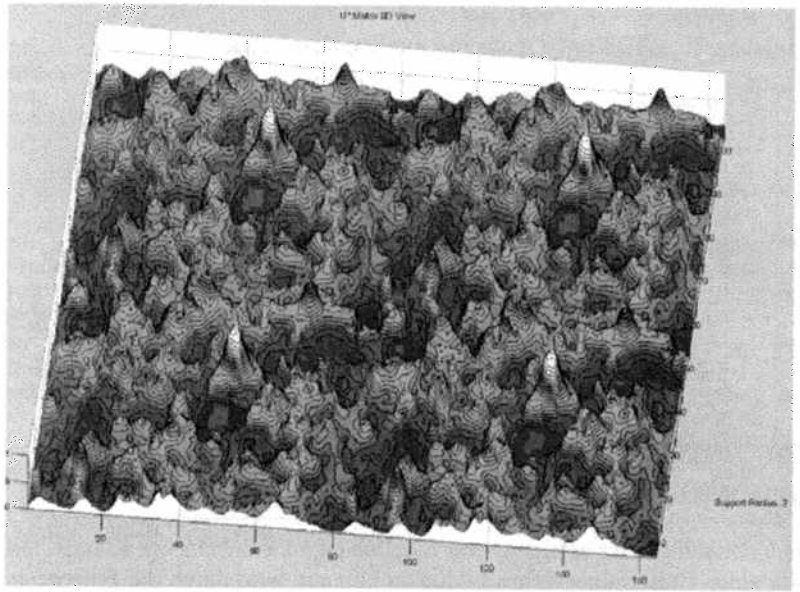
Die Vermutung, dass es nur schwer möglich ist, unterscheidbare Klassen mit Hilfe eines Clusteralgorithmus bilden zu können, wird durch die Emergente SOM aus Abbildung 8 zusätzlich bestätigt. Aufgrund dieser beiden Abbil-

dungen wird die Notwendigkeit einer Strukturerkennung vor einer geplanten Strukturbildung deutlich betont.

Es bilden sich bei den Daten der 12430 Gemeinden keine klaren Grenzen heraus. Möglichkeiten der Gruppierung lassen sich nicht erkennen, wie dies bei anderen Datensätzen bereits gezeigt wurde (siehe Abbildung 5).

Aufgrund erkannter Gruppierungsschwierigkeiten gemäß MDS und U*-Matrix wird entschieden, dass für eine mehrdimensionale Betrachtung die bereits gewonnene Einzelklassifizierung im Rahmen der Verteilungsuntersuchung besser geeignet ist, um auf Basis dieser ermittelten Klassen eine Gesamtbetrachtung durchzuführen. Im Hinblick auf eine sinnvolle Interpretierbarkeit von Klassenergebnissen werden drei der sechs Variablen (‚Verstädterungsgrad', ‚Konzentration', ‚Fahrzeit') gewählt. Ausgesucht werden die Variablen nicht nur aus inhaltlichen Aspekten, sondern auch unter dem Gesichtspunkt einer guten Objekttrennung (3D-Scatter-Plot).

Abbildung 8: U*-Matrix, (N=12430, D=6, 50x82 Neurons)



Quelle: Eigene Bearbeitung

Die erste Klassifikationsvariable bildet die Messgröße ‚Erreichbarkeit'. Diese unterstützt die Klassifikation in der Form, dass eine Aussage über die Fahrzeit im motorisierten Individualverkehr zum nächstgelegenen Oberzentrum in Deutschland getroffen werden kann. Charakterisiert wird dadurch die Verflechtung zwischen Umland und Zentrum. Infolge der Verteilungsuntersuchung wurde eine Entscheidungsgrenze von 30 Minuten gefunden, die im Sinne eines abstrahierten gravitationstheoretischen Verständnisses dazu eingesetzt wird, um die Peripherie und das direkt durch das Oberzentrum beeinflusste Umland zu unterscheiden.

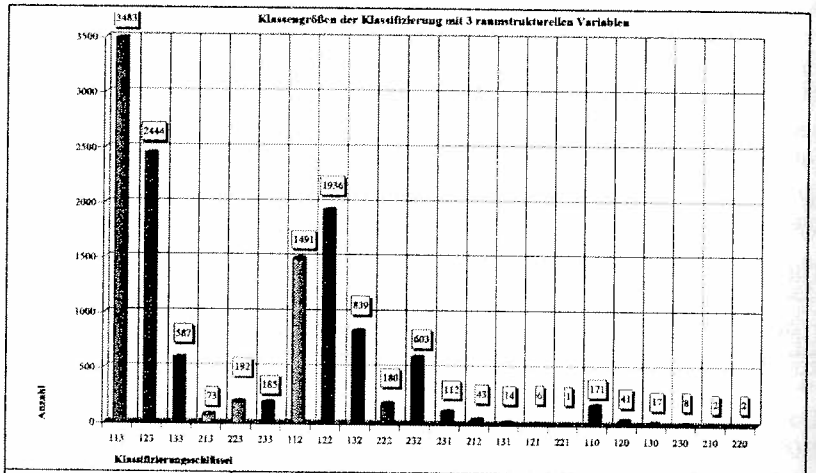
Die zweite Klassifikationsvariable wird durch die Messgröße ‚Konzentration' erzeugt, um eine Differenzierung nach der vorhandenen Auslastung der Gebäude- und Freifläche durch Bewohner und Beschäftigte vornehmen zu können.

Das deutsche Gemeindesystem wird dadurch in Anlehnung an die von BOUSTEDT bereits formulierte Einwohner- und Arbeitsplatzdichte regional ausdifferenziert. Diese in der Vergangenheit häufig eingesetzte Dichtegröße bezieht sich noch auf die Gemeindefläche insgesamt, so dass eine gewisse Unschärfe entsteht, da nicht die tatsächlich bebaute Fläche berücksichtigt wird und damit das Messergebnis erheblich von der Gebietsgröße einer Gemeinde beeinflusst wird. Aus diesem Grunde wird insbesondere mit Bezug auf SIEDENTOP[59] die Gebäude- und Freifläche als genauere räumliche Bezugsgröße herangezogen. Infolge der Verteilungsuntersuchung und der Modellierung der Verteilung mit GMMs wurden zwei Entscheidungsgrenzen erarbeitet (2500 und 4000 Einwohner und Arbeitsplätze je km² Gebäude- und Freifläche).

Die dritte Klassifikationsvariable bezieht sich auf die Messgröße ‚Verstädterung' und wird in die geplante Klassifikation integriert, da vor dem Hintergrund des technologischen Wandels und der sich ändernden wirtschaftlichen Bedingungen nicht nur Umstrukturierungsprozesse innerhalb eines Oberzentrums selbst stattfinden, sondern auch gerade im Umland eine Verstädterungsausdehnung erfolgt bzw. Auswirkungen der Schrumpfung sichtbar werden.

[59] Vgl. SIEDENTOP et al. (2005, S. 77)

Abbildung 9: Klasseneigenschaften der mehrdimensionalen Betrachtung



Semantik	Verstädterungsgrad			Konzentration			Erreichbarkeit			Anteil
	Median	Ø	StAbw	Median	Ø	StAbw	Median	Ø	StAbw	[in %]
113 = Peripherie, geringe Dichte	7,30	7,82	3,1	1736,2	1709	495,8	46,20	49,3	15,4	28,02
123 = Peripherie, mittlere Dichte	9,70	10,1	3,8	3064,1	3120	411,3	41,17	44,1	11,8	19,66
133 = Peripherie, hohe Dichte	10,97	11,2	4,2	4565,3	4781	730,5	39,44	43,5	11,4	4,72
213 = Regionalzentrum, geringe Dichte	25,17	27,1	7,4	2033,8	1843	618,4	41,50	47,3	16,1	0,59
223 = Regionalzentrum, mittlere Dichte	24,19	27,6	9,6	3198,5	3244	430,1	40,13	43,8	13,6	1,54
233 = Regionalzentrum, hohe Dichte	25,71	28,2	8,5	4942,0	5229	997,8	36,73	41,7	12,9	1,49
112 = Inter-Agglomerativ, geringe Dichte	7,60	8,2	3,1	1948,0	1863	457,5	22,46	21,6	5,8	12,00
122 = Inter-Agglomerativ, mittlere Dichte	10,59	10,9	3,7	3147,4	3183	416,6	21,95	21,3	5,9	15,58
132 = Inter-Agglomerativ, hohe Dichte	14,11	13,6	4,1	4664,9	4891	797,3	20,72	20,4	6,0	6,75
222 = Agglomerationseinheit, mittlere Dichte	23,47	26,3	7,9	3431,9	3366	403,0	20,49	19,5	6,1	1,45
232 = Agglomerationseinheit, hohe Dichte	27,68	30,9	10,2	5622,9	5880	1444,8	18,54	18,3	6,0	4,85
231 = Agglomerationszentrum	37,57	40,0	12,6	7557,6	7874	2073,8	0	0	0	0,90
212 = Sonderfälle: Agglomerationseinheit	27,65	30,0	8,8	1710,6	1624	660,1	19,24	19,8	5,97	0,35
131 = Sonderfälle: Agglomerationszentrum	17,48	15,9	3,6	4859,4	5057	846,1	0	0	0	0,11
121 = Sonderfälle: Agglomerationszentrum	9,64	12,2	4,7	3439,0	3504	375,4	0	0	0	0,05
221 = Sonderfälle: Agglomerationszentrum	22,07	22,0		3899,1	3899		0	0		0,01

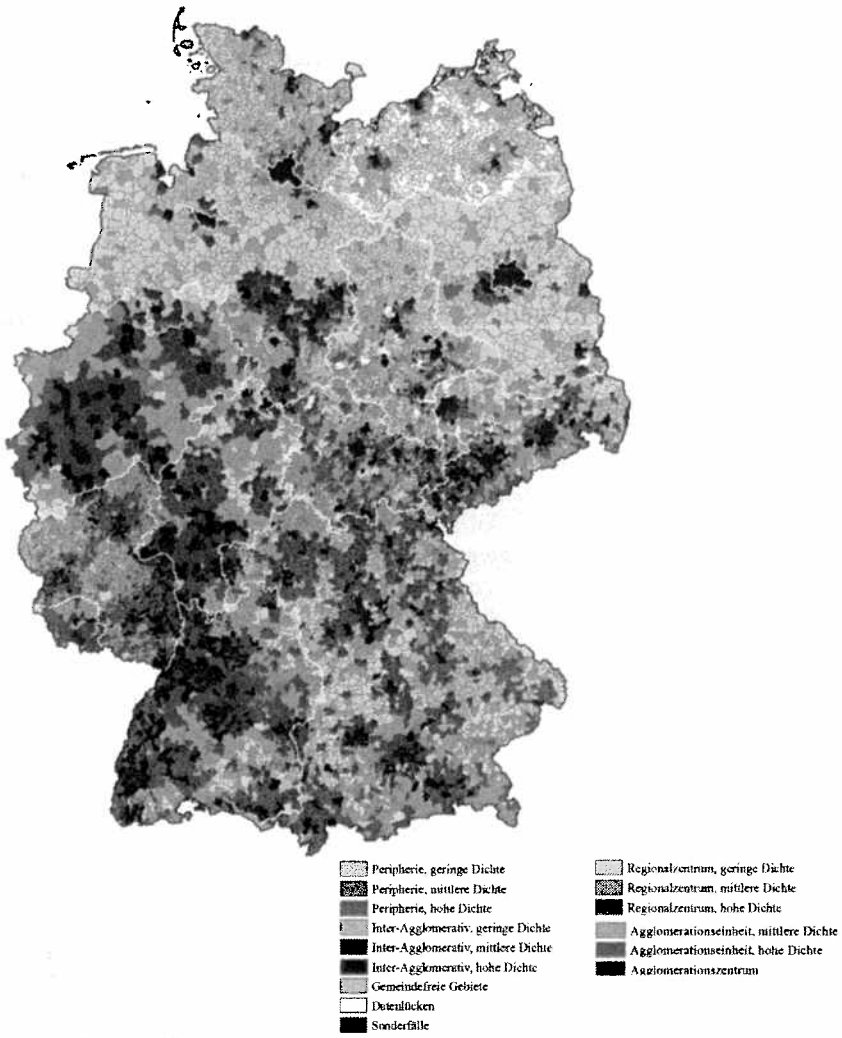
Ziffer 1
1 = hoch verstädtert (≤20), [%]
2 = geringer verstädtert (>20), [%]

Ziffer 2
1 = Niedrige Konzentration (≤ 2500), [P./km²]
2 = Mittlere Konzentration(>2500 < 4000), [P./km²]
3 = Hohe Konzentration > 4000, [P./km²]

Ziffer 3
0 = keine Daten, 'Nan'
1 = Oberzentrum, Fahrzeit = „0"
2 = Fahrzeit ≤ 30 Minuten
3 = Fahrzeit > 30 Minuten

Quelle: Eigene Bearbeitung

Abbildung 10: Verortung des Klassifizierungsergebnisses



Quelle: Eigene Bearbeitung

Monozentrische Regionen lösen sich teilweise auf, und es entstehen polyzentrale Regionen (agglomerieren). Neue Zentren des demographischen und wirtschaftlichen Wachstums entstehen sowohl durch Prozesse der Suburbanisierung als auch durch standörtlich differenzierte Bedingungen. Zukünftig werden angesichts vermehrter Schrumpfungsprozesse auch Phasen einer Disurbanisierung auftreten.

Die Verteilungsuntersuchung ermöglichte die Unterscheidung der Gemeinden nach dem Grad der Verstädterung in geringer und hoch verstädterte Gemeinden (Grenze: 20%).

Mit Hilfe von GMMs wurde die zugrunde liegende klassenbedingte Wahrscheinlichkeitsdichte berechnet, auf Basis derer ein Likelihood-Ratio-Klassifizierer dann ein gegebenes Muster einer Kategorie zuweist. Die Wahrscheinlichkeit der Zugehörigkeit eines Datensatzes zu einer Klasse wird als a posteriori Wahrscheinlichkeit bezeichnet. Anhand von Entscheidungsgrenzen, die durch den Schnittpunkt der Kurvenverläufe der Gaußverteilungen (Modus) gebildet wurden, lassen sich die Objekte klassifizieren.[60]

Es ist anzumerken, dass bei einer Klassifizierung anhand von Entscheidungsgrenzen mit steigender Anzahl der Untersuchungsvariablen sehr große Klassenzahlen entstehen können. Insgesamt wurden bei diesem Untersuchungsansatz mit drei Variablen 22 Klassen aus den Gemeindedaten ermittelt. Dies kommt der theoretisch zu erwartenden Klassenzahl von 24 nahe. Abbildung 9 beschreibt die Klasseneigenschaften.

Gemäß der Verortung der Klassen aus Abbildung 10, zeichnet sich in Deutschland eine regional sehr unterschiedliche Charakteristik von Gemeindestrukturen ab, die im Rahmen des Arbeitsschrittes der Wissenskonversion begründet wird.

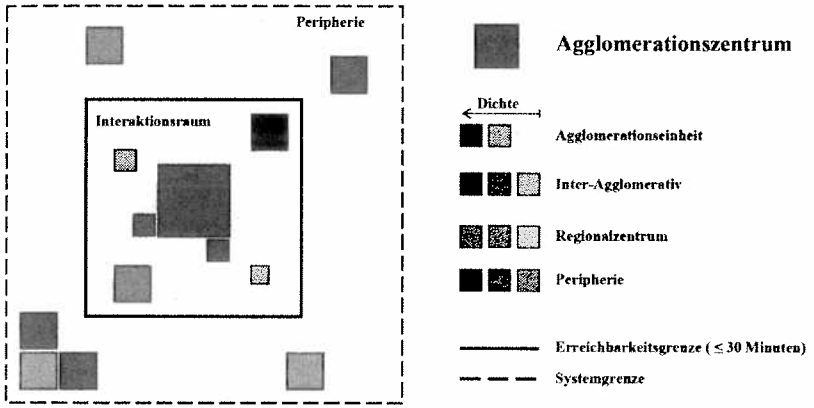
7 WISSENSKONVERSION

Das Ergebnis der Klassenbildung ermöglicht eine Raumbeobachtung, die der Grundidee folgt, die heutige Situation der Verdichtung bzw. Agglomerationen abzubilden. Abbildung 11 zeigt schematisch die Klassengrundstruktur.

[60] Vgl. LAURITZEN (1996) und BILMES (1997)

In Bezug auf ein gewähltes Agglomerationszentrum wird ein Verflechtungsgebiet untersucht, wobei als Verflechtungsmaß die Fahrzeit im motorisierten Individualverkehr zu den bestehenden Oberzentren verwendet wird. Die Oberzentren werden bei dieser Klassifizierung unter dem Begriff des Agglomerationszentrums beschrieben, und ein sogenannter Interaktionsraum entsteht zwischen den Gemeinden, die eine Fahrzeit von höchstens 30 Minuten zum Agglomerationszentrum aufweisen. Außerhalb dieses sogenannten Interaktionsraumes befinden sich die Gemeinden, welche der Peripherie zugeordnet werden. Die Messgröße ‚Konzentration' als modifizierte Siedlungsdichte (Auslastung der Gebäude- und Freifläche durch Einwohner und Beschäftigte) dient zum Aufbau von drei Dichteklassen.

Abbildung 11: Grundstruktur der Klassenbildung



Quelle: Eigene Bearbeitung

Der Grad der Verstädterung als Anteil der Siedlungs- und Verkehrsfläche an der Gemeindefläche insgesamt ermöglicht die Identifizierung von besonders verstädterten Gemeinden. In diesem Ansatz werden die Gemeinden, welche innerhalb des Interaktionsraums liegen, als Agglomerationseinheiten bezeichnet. Diese tragen bei ähnlichen Verstädterungseigenschaften der Nachbargemeinden zum Wachstum der Agglomeration insgesamt bei. Gemeinden, die in der Peripherie liegen und einen besonders großen Verstädterungsgrad aufweisen, werden als Regionalzentrum bezeichnet. Eine Gemeinde gilt als inter-agglomerativ, wenn diese sich im Interaktionsraum

befindet, jedoch nicht über einen Verstädterungsgrad oberhalb von 20 % verfügt.

Die einzelnen Agglomerationszentren werden untereinander mit ihrem dazugehörigen Interaktionsraum vergleichbar. Hinsichtlich der Einwohner und Arbeitsplätze je km² Gebäude- und Freifläche und dem Verstädterungsgrad wurden verschiedene Verflechtungsmuster erfasst. Die Regionen Rhein-Main, Rhein-Ruhr sowie die Region Stuttgart repräsentieren hochverdichtete und deutlich verstädterte Agglomerationen. Es ist zu vermuten, dass in Zukunft bei weiter fortschreitendem regionalen Wachstum von Bevölkerung und Beschäftigten eine noch stärkere Agglomeration entstehen wird und damit die Übergänge zwischen Verflechtungsgebieten noch schwieriger abgrenzbar sein werden. In Bezug auf einige ostdeutsche Agglomerationszentren (z.B. Chemnitz, Neubrandenburg oder Schwerin) lassen sich deutlich geringere Dichte- und Verstädterungseigenschaften bei den Gemeinden im Verflechtungsgebiet erkennen. Hier kann nicht von einer homogen dichten städtischen Agglomeration gesprochen werden, sondern eher von einem Kerngebiet, welches als Oberzentrum auf das Umland einen Einfluss ausübt.

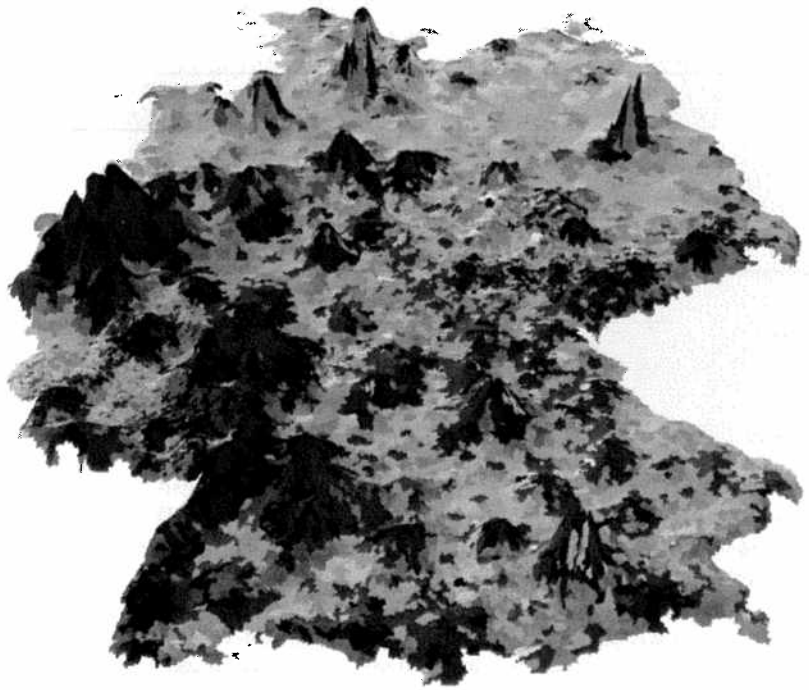
Gerade in den Gebieten mit einer geringeren Verdichtung in den Verflechtungsgebieten fallen entweder Gemeinden mit hoher Dichte oder sogar Gemeinden mit hoher Dichte und großer Verstädterung in einem gewissen Erreichbarkeitsabstand zum Agglomerationszentrum auf. Exemplarisch herausgegriffen sei das Agglomerationszentrum Kempten und die speziellen Gemeinden mit einer hohen Dichte: Marktobersdorf, Pfronten und Immenstadt im Allgäu.

Innerhalb der Peripherie sind zwischen den dazugehörigen Gemeinden ebenfalls Besonderheiten im Hinblick auf die Dichte und den Verstädterungsgrad messbar. Die als Regionalzentren definierten Gemeinden sind durch einen höheren Verstädterungsgrad gekennzeichnet. Es ist zu erwarten, dass diese Gemeinden weitere besondere Eigenschaften haben und auch einen stärkeren Einfluss auf die Nachbargemeinden ausüben. Zu bemerken ist, dass es sich vielfach um Gemeinden mit Stadtrecht handelt. Darüber hinaus werden Gemeinden in der Peripherie mit einer hohen Dichte identifiziert, die möglicherweise weitere regionalspezifische Besonderheiten aufweisen.

Zusätzlich vorstellbar ist zukünftig eine weitere Klassenerklärung im Sinne der Operationalisierung. Für diesen Zweck sei an dieser Stelle nochmals

verwiesen auf regelbasierte Ansätze (z.B. sig*-Algorithmus oder U-Know-Algorithmus) oder hierarchisierte Entscheidungsbäume (z.B. CART). Auf diese Weise lassen sich beispielsweise weitere Erkenntnisse über die von Agglomerations- und Verdichtungsprozessen betroffenen Gemeinden gewinnen.

Abbildung 12: Überlagerung von Verstädterungseigenschaften gemäß Abbildung 10 und berechneter Gebäudedichte.



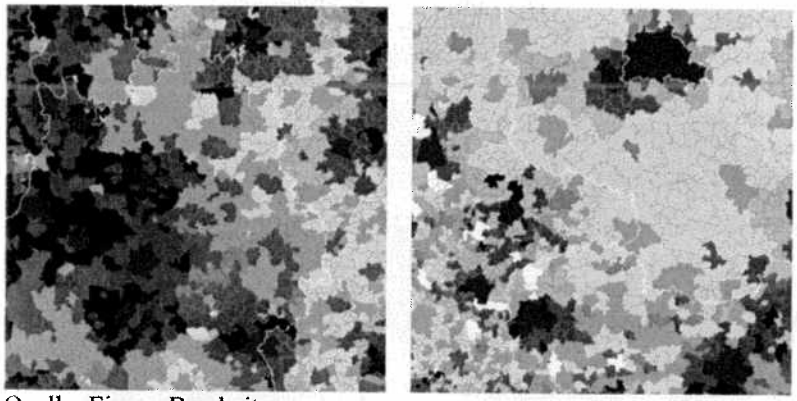
Quelle: Eigene Bearbeitung

Abbildung 12 vermittelt einen visuellen Eindruck von einer eindimensionalen Informationsüberlagerung. In Überlagerung dargestellt werden die durch Klassenbildung erzeugten Verstädterungseigenschaften mit der Gebäudedichte in Deutschland, wobei große Höhenwerte Ausdruck einer größeren Gebäudedichte mit Bezug auf die Katasterfläche der Gemeinde repräsentie-

ren. Eine hohe Gebäudedichte deutet sich gerade in den Gemeinden des zuvor definierten Interaktionsraums an. In BEHNISCH (2007 b) wird ein Schätzansatz ausführlich vorgestellt und gezeigt, dass vermutlich 20% der Gemeinden in Deutschland bereits 80 % des Gebäudebestandes enthalten.

In Abbildung 13 sind Regionalstrukturen in einem größeren Detailausschnitt ersichtlich. Es handelt sich um die Region Stuttgart und den Großraum Berlin.

Abbildung 13: Detailansicht zu Agglomerations- und Verdichtungseigenschaften (Stuttgart, links und Berlin, rechts)



Quelle: Eigene Bearbeitung

In der Region Stuttgart kann bereits von flächenhaften Verstädterungsausprägungen gesprochen werden. Im Interaktionsraum von Stuttgart werden hoch verdichtete und auch hoch verstädterte Gemeinden erfasst. Die ermittelte Erreichbarkeitsgrenze von maximal 30 Minuten Fahrzeit zum nächsten Agglomerationszentrum führt zur Identifizierung der Interaktionsräume. In der Region Stuttgart und den benachbarten Agglomerationszentren Pforzheim, Karlsruhe und Heilbronn sowie Tübingen grenzen diese aneinander und überlagern sich in Teilbereichen. Im Großraum Berlin wird in direkter Nachbarschaft zum Agglomerationszentrum eine größere Anzahl Gemeinden als Regionalzentrum erkannt. Diese zeigen oftmals eine geringe bis mittlere Dichte. Es handelt sich um ein frühes Suburbanisierungsstadium. Das sich in der Industrialisierung heraus gebildete Siedlungssystem mit einem Ring größerer Mittelstädte um das Zentrum ist weitgehend noch erhalten.

8. SCHLUSSFOLGERUNG

Es wurde gezeigt, dass die Klassenbildung auf Basis von drei ausgewählten Messgrößen in Anlehnung an BOUSTEDT[61] zur Erfassung und Strukturierung des Agglomerations- und Verdichtungsprozesses geeignet ist. Das Ergebnis ermöglicht einerseits die Identifizierung von polyzentrischen Regionalstrukturen im Verflechtungsbereich von Oberzentren und andererseits die Aufdeckung von monozentrisch geprägten geringer besiedelten Regionalstrukturen.

Mit Blick auf die Verstädterung zeichnet sich der Rückgang des Freiraums am Rande der Kernstädte im engeren suburbanen Raum deutlich ab. Die Suburbanisierung hat dazu geführt, dass sich in Teilbereichen flächenhaft bebaute Stadtregionen entwickelt haben, die sich weit in das Umland ausdehnen und im Kontext der Zwischenstadtdiskussion eine große Relevanz erlangt haben.

Es sei ergänzend angemerkt, dass im Verflechtungsbereich von Oberzentren die Ausweisungspraxis von zentralen Orten diskutiert und als reformbedürftig betrachtet wird. Es wird von vielen Seiten die zusätzliche Zentralitätsstufe ‚Metropolregion' debattiert.[62] Diese sogenannten ‚zentralörtlichen Kooperationsräume'[63] werden nicht von den klassischen Zentrale-Orte-Kategorien repräsentiert, die innerhalb des Zentrale-Orte-Konzeptes der Raumordnungspolitik entwickelt wurden.

Im methodischen Sinne hat die dargelegte Einzeluntersuchung der sechs raumstrukturellen Variablen eine problemorientierte Sicht auf die Notwendigkeit einer Verteilungsuntersuchung geliefert. Zusätzlich wurde auf den Einsatzvorteil von Verfahren der Strukturerkennung hingewiesen. Gerade mit Hilfe einer Emergenten SOM wurde erkannt, dass keine deutliche Struktur in den Daten der sechs Messgrößen vorhanden ist. Die Anwendung eines Clusteralgorithmus hätte daher nicht zu einer klar unterscheidbaren Gruppenstruktur geführt.

---

61 Vgl. BOUSTEDT (1953), BOUSTEDT (1975 a), BOUSTEDT (1975 b)

62 Vgl. ARL [2002], SIEDENTOP [2003, S. 192 ff.], siehe ‚Metropolregion' bei BLOTEVOGEL [2005]

63 Vgl. ARL [2002]

Weiterhin unterstützt die Modellierung der Verteilung den Aufbau von Entscheidungsgrenzen und fördert den Aufbau von üblichen Schwellwerten im urbanen Forschungsfeld. Die Verortung von einzelnen Messgrößen erfolgt nicht auf subjektiv festgelegten Klassengrenzen oder aus Expertenwissen. Vielmehr bilden Gauß-Mixtur-Modelle eine Charakteristik in den Daten ab und liefern eine zusätzliche Entscheidungsgrundlage.

Unter dem Aspekt einer zunehmend feststellbaren Verfügbarkeit von raumbezogenen digitalen Daten und leistungsstarker GI-Systeme ist damit zu rechnen, dass sich weitere umfassende Lösungsmöglichkeiten für die räumliche Analyse in den nächsten Jahren ergeben werden[64]. Das Datenmaterial wird vielfach nicht auf administrative Einheiten aggregiert, so dass im Sinne der geographischen Abbildung eine genauere Wiedergabe der tatsächlich vorhandenen räumlichen Situation vorstellbar ist.

Literatur

Altenkrüger, Doris und Büttner, Winfried (1992): Wissensbasierte Systeme – Architektur, Entwicklung, Echtzeit-Anwendungen. Vieweg Verlag, Braunschweig.

ARL, Akademie für Raumforschung und Landesplanung (2002): Fortentwicklung des Zentrale-Orte-Systems. In: Blotevogel, H. (Hrsg.): Bericht des Ad-hoc-Arbeits-kreises „Fortentwicklung des Zentrale-Orte-Systems". ARL, Hannover.

Backhaus; Erichson; Plinke; Weiber (2006): Multivariate Analysemethoden – Eine anwendungsorientierte Einführung. 11. Auflage. Springer Verlag, Berlin.

Bahrenberg, G.; Giese, E.; Nipper, J. (2003): Statistische Methoden in der Geographie. Gebrüder Borntraeger, Berlin

Becher, St. (1995): Klassifikation der regionalen Immobilienmärkte der Bundesrepublik Deutschland (Dissertation). Universität Mainz.

Behnisch, Martin (2007 a): Urban Data Mining - Operationalisierung der Strukturerkennung und Strukturbildung von Ähnlichkeitsmustern über

[64] Vgl. THINH [2004 a, S.52 ff.], der zusätzlich einen Überblick von den nationalen und internationalen Programmen zur Entwicklung und Forschung im Bereich Geoinformation anführt.

die gebaute Umwelt. Dissertation, Institut für industrielle Bauproduktion, Universität Karlsruhe (TH).

Behnisch, Martin (2007 b): Estimation of Building Stocks. In: Building Research and Information (Accepted). Taylor and Francis, Colchester.

Behnisch, Martin und Ultsch, Alfred (2007): Urban Data Mining. Proceedings of the 31st Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications. Springer series Studies in Classification, Data Analysis, and Knowledge Organization.

Behrens, K. / Marhenke, W. (1997): Die Abgrenzung von Stadtregionen und Verflechtungsgebieten in der Bundesrepublik Deutschland. In: Statistisches Landesamt Baden-Württemberg (Hrsg.): Jahrbuch für Statistik und Landeskunde Baden-Württemberg, S. 165-186, Statistisches Landesamt Baden-Württemberg, Stuttgart.

Bilmes, J. (1997): A Gentle Tutorial on the EM Algorithm and its Application to Parameter Estimation for Gaussian Mixture and Hidden Markov Models. Technical Report, University of Berkeley, ICSI-TR-97-021. http://crow.ee.washington.edu/people/bulyko/papers/em.pdf

Blotevogel, H. (2005): Neuformulierung des Zentrale-Orte-Konzepts. Kurzfassung eines Vortrages im Rahmen einer Fachtagung des Ministeriums des Innern und für Sport, Oberste Landesplanungsbehörde, am 14.07. 2005 in Budenheim bei Mainz.

Bock, Hans Hermann (1974): Automatische Klassifikation. Studia Mathematica/Mathematische Lehrbücher, Band XXIV. Vandenhoeck & Ruprecht in Göttingen.

Bocklisch, S. F. (1987): Prozeßanalyse mit unscharfen Verfahren, VEB Verlag Technik, Berlin.

Boustedt, O. (1953): Die Stadtregion. Ein Beitrag zur Abgrenzung städtischer Agglomerationen. Allgemeines statistisches Archiv 37, S.13-26.

Boustedt, O. (1975 a): Grundriß der empirischen Raumforschung – Teil 3 (= Taschenbücher zur Raumplanung Band 6): Siedlungsstrukturen. Veröffentlichung Akademie für Raumforschung und Landesplanung, Hannover.

Boustedt, O. (1975 b): Grundriß der empirischen Raumforschung – Teil 4: Regionalstatistik. (= Taschenbücher zur Raumplanung Band 7), Veröffentlichung Akademie für Raumforschung und Landesplanung, Hannover.

Breiman, L.; Friedman, J.H.; Olshen, R.A.; Stone, C.J. (1984) Classification and Regression Trees. Wadsworth, California.

Deichsel und Trampisch (1985): Clusteranalyse und Diskriminanzanalyse. Gustav Fischer Verlag, Stuttgart.

Deimer (1986): Unscharfe Clusteranalysemethoden – Eine problemorientierte Darstellung zur unscharfen Klassifikation gemischter Daten. Dissertation. Johann Wolfgang Goethe-Universität, Frankfurt am Main.

Erb, Wolf-Dieter (1990): Anwendungsmöglichkeiten der linearen Diskriminanzanalyse in Geographie und Regionalwissenschaft. Schriften des Zentrums für regionale Entwicklungsforschung der Justus-Liebig-Universität Gießen, Bd. 39, Weltarchiv, Hamburg.

Everitt, B.S. (1980): Cluster Analysis. Quality and Quantity, New York.

Fayyad, Usama; Piatesky-Shapiro, Gregory; Smyth, Padhraic (1996): The KDD Process for Extracting Useful Knowledge from Volumes of Data. In: Communications of the ACM, Vol. 39, No 11, November 1996, ACM Press, New York.

Fischer, M. M. (1982): Eine Methodologie der Regionaltaxonomie: Probleme und Verfahren der Klassifikation und Regionalisierung in der Geographie und Regionalforschung. Heft 3 der Bremer Beiträge zur Geographie und Raumplanung (Hrsg.: Bahrenberg, Gerhard; Barth, Hans-Karl; Leuze, Eva; Taubmann, Wolfgang), Presse- und Informationsdienst, Universität Bremen.

Gaul, W. G.; Säuberlich, F. (1998): Classification and Positioning of Data Mining Tools. Herausforderungen der Informationsgesellschaft an Datenanalyse und Wissensverarbeitung. 22. Jahrestagung - Gesellschaft für Klassifikation, Proceedings. Springer Verlag, Berlin.

Hartung, Joachim (2005): Statistik – Lehr- und Handbuch der angewandten Statistik. 14. unwesentlich veränderte Auflage, Oldenbourg, München.

Hand, David J.; Mannila, Heikki; Smyth, Padhraic (2001): Principles of Data Mining, Bradford Book.

Höppner, F.; Klawonn, F.; Kruse, R.; Runkler, T. (1999): Fuzzy cluster Analysis. Aktualisierte Version der deutschen Ausgabe: Höppner, F. (1997). John Wiley & Sons, New York.

Hyttinen, Laura (2004): Knowledge conversions in knowledge work - a descriptive case study. Licentiate Thesis. Espoo: Helsinki University of Technology.
http://www.tuta.hut.fi/library/raportit/tps_report/pdf/raporttiLauraHyttinen1.pdf, (Stand: 16.06.2006).

Jaeger, J.; Scheringer, M. (1998): Transdisziplinarität: Problemorientierung ohne Methodenzwang. – GAIA 7 (1): 10–25, oekom verlag, München.

Kohonen, T. (1982): Self-Organized Formation of Topologically Correct Feature Maps. Biological Cybernetics Vol. 43, pp. 59-69.

Lauritzen, S. (1996): Graphical Models. Oxford University Press.

Nonaka, I.; Takeuchi, H. (1997): Die Organisation des Wissens. Wie japanische Unternehmen eine brachliegende Ressource nutzbar machen. Campus Verlag, Frankfurt am Main

Preece, J.; Rogers, Y.; Sharp, H.; Benyon, D.; Holland, S.; Carey, T. (1994): Human-Computer Interaction. Addision Wesley, Boston.

Quinlan, J. R. (1986): Introduction of Decision Trees. Machine Learning Vol. 1, Springer Netherlands, Dordrecht.

Reynolds, D. A.; Quatieri, T. F.; Dunn, R. B. (2000): Speaker verification using adapted gaussian mixture models. Digital Signal Processing, 10(1-3), 19-41.

Sander, Jörg (1999): Generalized Density-Based Clustering for Spatial Data Mining. Utz Verlag, München.

Schreiber, G.; Akkermans, H.; Anjewierden, A.; De Hoog, R.; Shadbolt, N.; Van de Velde, W.; Wielinga, B. (2000): Knowledge Engineering and Management - The CommonKADS Methodology. MIT Press, Cambridge, Massachusetts.

Scott, D.W. (1992): Multivariate Density Estimation. John Wiley & Sons, New York.

Siedentop, St.; Kausch, St.; Einig, K. und Gössel, J. (2003): Siedlungsstrukturelle Veränderungen im Umland der Agglomerationsräume. Bundesamt für Bauwesen und Raumordnung (Hrsg.), Heft 114, Bonn.

Siedentop, St.; Kausch, St.; Guth, D.; Stein, A.; Wolf, U.; Lanzendorf, M.; Harbich, R.; Hesse, M. (2005): Mobilität im suburbanen Raum. Neue verkehrliche und raumordnerische Implikationen des räumlichen Strukturwandels. Abschlussbericht zum Forschungsvorhaben 70.716 im Auftrag des Bundesministeriums für Verkehr, Bau- und Wohnungswesen (BMVBW).

Sokal, R.R.; Sneath, P.H.A (1963): Principles of Numerical Taxonomy. Freeman, W.H., San Francisco.

Späth, H. (1975): Cluster-Analyse-Algorithmen zur Objektklassifikation und Datenreduktion. Oldenbourg Verlag, München, Wien.

Staack, J. (1995): Die Klassifikation deutscher Städte nach ihrer regionalen Zentralität. Dissertation, Universität Hamburg. Europäische Hochschulschriften, Peter Lang Verlag, Frankfurt am Main.

Streich, Bernd (2005): Stadtplanung in der Wissensgesellschaft: ein Handbuch. 1. Auflage, VS Verlag für Sozialwissenschaften, Wiesbaden.

Thinh, Nguyen Xuan (2004): Entwicklung von mathematisch-geoinformatischen Methoden und Modellen zur Analyse, Bewertung, Simulation und Entscheidungsunterstützung in Städtebau und Stadtökologie. Habilitationsschrift der Universität Rostock, Fakultät der Agrar- und Umweltwissenschaften.

Thinh, Nguyen Xuan; Behnisch, Martin; Ultsch, Alfred (2006): Examination of several results of different cluster analysis with a separate view to balancing the economic and ecological performance potential of towns and cities. In: Bock, H.H., Gaul, W., Vichi, M. (Hrsg.): Proceedings of the 31st Annual Conference of the German Classification Society on Data Analysis, Machine Learning, and Applications. Springer series Studies in Classification, Data Analysis, and Knowledge Organization.

Überla, Karl (1971): Faktorenanalyse - eine systematische Einführung für Psychologen, Mediziner, Wirtschafts- und Sozialwissenschaftler. 2. Auflage, Springer Verlag, Berlin.

Ultsch, A. (1991): Konnektionistische Modelle und ihre Integration mit wissensbasierten Systemen. Habilitationsschrift der Universität Dortmund.

Ultsch, A. (1999): Data Mining and Knowledge Discovery with Emergent Self-Organizing Feature Maps for Multivariate Time Series, Kohonen Maps, pp. 33-46, Elsevier, Amsterdam.

Ultsch, A. (2000): The Neuronal Data Mine. In: Proceedings 2nd Int. ICSC Symposium on Neural Computation NC, Berlin.

Ultsch, A. (2001): Eine Begründung der Pareto 80/20 Regel und Grenzwerte für die ABC Analyse. Technical Reports No. 30, Dept. of Mathematics and Computer Science, University of Marburg, Germany.

Ultsch, A. (2003): Pareto Density Estimation: A Density Estimation for Knowledge Discovery. Baier, D. and Wernecke, K.D. (Eds), in: Innovations in Classification, Data Science, and Information Systems - Proceedings 27th Annual Conference of the German Classification Society (GfKL) 2003. Berlin, Heidelberg, Springer, pp. 91-100.

Ultsch, A. (2005): Clustering with SOM U*C. In: Proceedings Workshop on Self-Organizing Maps (WSOM 2005), Paris, France, pp. 75-82.

Ultsch, A. (2006): Vorlesungsunterlagen: Knowledge Discovery, Fachbereich Mathematik und Informatik, Philipps-Universität Marburg.

Vogel, Friedrich (1975): Probleme und Verfahren der numerischen Klassifikation. Vandenhoeck & Ruprecht, Göttingen.

Wilke, Helmut (1998): Organisierte Wissensarbeit. In: Zeitschrift für Soziologie 27, 3, S. 161-177, Lucius & Lucius Verlag, Stuttgart.

Zadeh, L.A. (1965): Fuzzy Sets. In: Information and Control, Vol. 8, S.338-353.