

B 26 Datenvisualisierung und Data Mining

Daniel A. Keim

B 26.1 Einleitung

Die rasante technologische Entwicklung der letzten zwei Jahrzehnte ermöglicht heute die persistente Speicherung riesiger Datenmengen durch den Computer. Forscher an der Universität Berkeley haben berechnet, dass jedes Jahr ca. 1 Exabyte (= 1 Million Terabyte) Daten generiert werden – ein großer Teil davon in digitaler Form. Das bedeutet aber, dass in den nächsten drei Jahren mehr Daten generiert werden als in der gesamten menschlichen Entwicklung zuvor. Die Daten werden oft automatisch mit Hilfe von Sensoren und Überwachungssystemen aufgezeichnet. So werden beispielsweise alltägliche Vorgänge des menschlichen Lebens, wie das Bezahlen mit Kreditkarte oder die Benutzung des Telefons, durch Computer aufgezeichnet. Dabei werden gewöhnlich alle verfügbaren Parameter abgespeichert, wodurch hochdimensionale Datensätze entstehen. Die Daten werden gesammelt, da sie wertvolle Informationen enthalten, die einen Wettbewerbsvorteil bieten können. Das Finden der wertvollen Informationen in den großen Datenmengen ist aber keine leichte Aufgabe. Heutige Datenbankmanagementsysteme können nur kleine Teilmengen dieser riesigen Datenmengen darstellen. Werden die Daten zum Beispiel in textueller Form ausgegeben, können höchstens ein paar hundert Zeilen auf dem Bildschirm dargestellt werden. Bei Millionen von Datensätzen ist dies aber nur ein Tropfen auf den heißen Stein.

B 26.1.1 Vorteile der visuellen Datenexploration

Für ein effektives Data Mining ist es wichtig, den Menschen in den Datenexplorationsprozess mit einzubinden, um die Fähigkeiten des Menschen – Flexibilität, Kreativität und das Allgemeinverständnis – mit den enormen Speicherkapazitäten und Rechenleistungen moderner Computersysteme zu kombinieren. Die Grundidee der visuellen Datenexploration ist die geeignete Darstellung der Daten in visueller Form, die es dem Menschen erlauben, einen Einblick in die Struktur der Daten zu bekommen, Schlussfolgerungen aus den Daten zu ziehen sowie direkt mit den Daten zu interagieren.

Visuelle Data-Mining-Verfahren haben in den letzten Jahren einen hohen Stellenwert innerhalb des Forschungsbereichs Data Mining erhalten. Ihr Einsatz ist immer dann sinnvoll, wenn wenig über die Daten bekannt ist und die Explorationsziele nicht genau spezifiziert sind. Dadurch dass der Mensch direkt am Explorationsprozess beteiligt ist, können die Explorationsziele bei Bedarf verändert und angepasst werden.

Die visuelle Datenexploration kann als ein Prozess zur Generierung von Hypothesen aufgefasst werden. Sie ermöglicht dem Menschen ein tieferes Verständnis für die Daten, wodurch er neue Hypothesen über die Daten aufstellen kann. Die Hypothesen können dann wiederum mit Hilfe visueller Datenexplorationsverfahren untersucht und verifiziert werden. Die Verifikation kann jedoch auch mit Hilfe von Techniken aus dem Bereich der Statistik und der künstlichen Intelligenz durchgeführt werden. Die Hauptvorteile der Einbindung des Menschen in den Prozess der Datenexploration im Vergleich zu vollautomatischen Verfahren aus der Statistik bzw. Künstlichen Intelligenz sind:

- Der visuelle Datenexplorationsprozess kann stark inhomogene und verrauschte Daten verarbeiten,
- der Benutzer benötigt keine Kenntnisse von komplexen mathematischen oder statistischen Algorithmen und Parametern, und deshalb kann die Datenexploration auch durch Nicht-Spezialisten durchgeführt werden.

Zusammenfassend kann man feststellen, dass die visuelle Datenexploration in vielen Fällen eine einfachere Exploration der Daten erlaubt und oft auch bessere Ergebnisse erzielt, insbesondere wenn die herkömmlichen automatischen Algorithmen nur unzureichende Ergebnisse liefern. Die visuelle Datenexploration bietet darüber hinaus ein besseres Verständnis des Datenexplorationsprozesses sowie der erzielten Ergebnisse. Visuelle Datenexplorations-Techniken werden deshalb in vielen Anwendungsbereichen eingesetzt und in Verbindung mit automatischen Algorithmen sind sie ein unentbehrliches Verfahren zur Exploration wichtiger Informationen aus großen Datenbanken.

B 26.1.2 Das Paradigma der visuellen Datenexploration

Nach Shneidermann (Lit. 22) kann die visuelle Datenexploration in drei Schritte untergliedert werden:

- overview,
- zoom and filter,
- details-on-demand.

Diese Gliederung wird auch als Information Seeking Mantra (Lit. 22) bezeichnet. Der Benutzer braucht beim visuellen Datenexplorationsprozess zuerst einen Überblick über die Daten (overview). In dieser ersten visuellen Darstellung kann der Benutzer interessante Muster in den Daten erkennen, die dann im folgenden mit Hilfe von Zoom- und Selektionstechniken (zoom and filter) genauer untersucht werden. Für eine genaue Analyse der Muster benötigt der Benutzer eine Möglichkeit, auf Details der Daten zuzugreifen (details-on-demand). In allen drei Schritten der visuellen Datenexploration können Visualisierungstechniken eingesetzt werden: Visualisierungstechniken können einfach einen Überblick über die Daten erzeugen und erlauben es einem Benutzer, interessante Teilmengen innerhalb der Visualisierung schnell zu erkennen. Während des Fokussierens auf interessante Teilmengen ist es wichtig, einen Überblick

über die Daten beizubehalten, was zum Beispiel durch eine interaktive Verzerrung der visuellen Überblicksdarstellung bezüglich der Foki erfolgen kann. Für die weitere Exploration interessanter Teilmengen benötigt der Datenanalyst eine Möglichkeit, die Daten genauer zu betrachten, um Details zu verstehen. Es ist in diesem Zusammenhang wichtig, dass Visualisierungstechniken nicht nur grundlegende Verfahren für alle drei Schritte bereitstellen, sondern auch die Schwierigkeiten bei den Übergängen zwischen diesen Schritten überbrücken helfen.

B 26.2 Klassifizierung Visueller Data-Mining-Techniken

Visuelles Data Mining bzw. Informations-Visualisierung konzentriert sich auf Daten, die keine 2D- oder 3D-Semantik besitzen und damit keine Standardabbildung auf die zweidimensionale Darstellung des Bildschirms besitzen. Für solche Datenmengen gibt es mittlerweile eine Reihe guter Visualisierungstechniken wie zum Beispiel x-y-Plots, Liniendiagramme und Histogramme. Diese Techniken können für die visuelle Datenexploration hilfreich sein, jedoch sind sie im allgemeinen beschränkt auf relativ kleine und niedrigdimensionale Datenmengen. In den vergangenen Jahren wurde eine Vielzahl neuartiger Techniken für hoch-

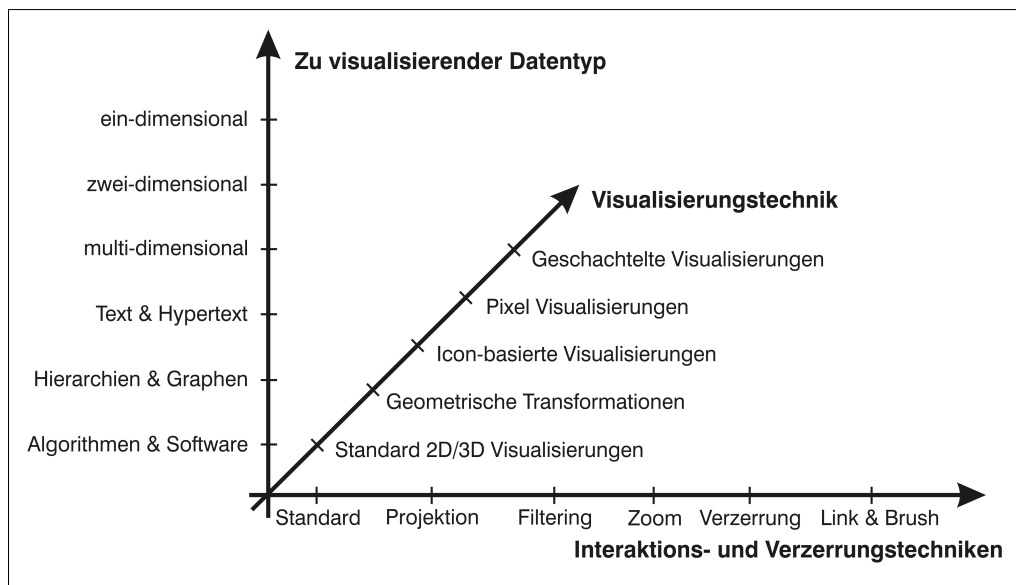


Abb. 1: Klassifikation visueller Data-Mining-Techniken (vgl. Lit. 12)

dimensionale Datenmengen ohne interne 2D- oder 3D-Semantik entwickelt. Ein Überblick über diese Verfahren wird in den aktuellen Büchern gegeben (Lit. 06, Lit. 28, Lit. 25, Lit. 24). Die Techniken können anhand folgender drei Kriterien klassifiziert werden (Lit. 12; siehe Abb. 1):

- (1) der zu visualisierende Datentyp,
- (2) die verwendete Visualisierungstechnik und
- (3) die verwendeten Techniken für Interaktion und Verzerrung.

Der zu **visualisierende Datentyp** kann wie folgt untergliedert werden (Lit. 22):

- **Ein-dimensionale Daten**, wie zum Beispiel zeitabhängige Daten (vgl. ThemeRiver Visualisierung in Abb. 2)
- **Zwei-dimensionale Daten**, wie zum Beispiel geographische Karten (vgl. Gridfit Visualisierung, Lit. 13)
- **Multi-dimensionale Daten**, wie zum Beispiel tabellarische Daten aus relationalen Datenbanken (vgl. Parallele-Koordinaten-Visualisierung, Lit. 10 und Abb. 3)
- **Text und Hypertext**, wie zum Beispiel Nachrichten oder Web-Dokumente (vgl. ThemeView Visualisierung, Lit. 30, Lit. 29)
- **Hierarchien und Graphen**, wie zum Beispiel Telefon- oder Internetverbindungen (vgl. Skitter Visualisierung, Lit. 08 und Abb. 4)
- **Algorithmen und Software**, wie zum Beispiel Debugging-Operationen (vgl. Tarantula Software Visualisierung, Lit. 07).

Die **Visualisierungstechniken** können wie folgt untergliedert werden:

- **Standard 2D/3D-Visualisierungen**, wie zum Beispiel Balkendiagramme oder X-Y-Diagramme
- **Geometrische Transformationen**, wie zum Beispiel künstliche Landschaften (Lit. 29) und Parallele Koordinaten (Lit. 10, Abb. 3)
- **Icon-basierte Visualisierungen**, wie zum Beispiel die „Strichmännchen“-Visualisierung (Lit. 19)
- **Pixel-Visualisierungen**, wie die Recursive Pattern oder Circle Segments Techniken (Lit. 11)

- **Geschachtelte Visualisierungen**, wie zum Beispiel Treemaps (Lit. 21) oder Dimensional Stacking (Lit. 27).

Die dritte Kategorie der Klassifikation sind die Interaktions- und Verzerrungstechniken. Sie erlauben es dem Benutzer, direkt mit den Visualisierungen zu interagieren. Interaktions- und Verzerrungstechniken können wie folgt untergliedert werden:

- Interaktive Projektion wie im GrandTour System (Lit. 03)
- Interaktive Selektion wie im Polaris System (Lit. 26)
- Interaktives Zooming wie im Spotfire System (Lit. 23)
- Interaktive Verzerrung wie im Hyperbolic Tree (Lit. 17, Lit. 18)
- Interaktives Linking and Brushing wie im XGobi System (Lit. 20, Lit. 04).

Man beachte, dass die drei Dimensionen unserer Klassifikation als orthogonal betrachtet werden können. Orthogonal bedeutet in diesem Zusammenhang, dass für eine zu visualisierende Datenmenge eine beliebige Visualisierungstechnik in Verbindung mit einer beliebigen Interaktions- und Verzerrungstechnik verwendet werden kann. Man beachte ferner, dass ein System mehrere unterschiedliche Datentypen und eine Kombination mehrerer Visualisierungs- und Interaktionstechniken unterstützen kann.

B 26.3 Zu visualisierende Datentypen

Die im Bereich Informations-Visualisierung vorkommenden Daten besitzen in der Regel eine große Anzahl von Datensätzen. Jeder Datensatz entspricht dabei einer Beobachtung, wie zum Beispiel einer Messung bei einem physikalischen Experiment oder einer Transaktion in einem E-Commerce-System, und besitzt eine feste Anzahl an Attributen. Die Anzahl der Attribute kann dabei stark variieren – von einigen wenigen Attributen bis hin zu Tausenden von Attributen. Da bestimmte Attribute wie zum Beispiel die Zeit oder die geographischen Koordinaten eine besondere Bedeutung haben, werden diese in der Regel auch besonders behandelt.

Eindimensionale Daten besitzen in der Regel ein kontinuierliches Attribut, das eine vollständige Ordnung auf den Daten definiert. Ein typisches Beispiel für eindimensionale Daten sind zeitabhängige Daten. Jedem Zeitpunkt können dabei mehrere Datenwerte zugeordnet sein. Beispiele für eindimensionale Daten sind Aktienkurs-Verläufe oder Zeitreihen von Zeitungsmeldungen (vgl. ThemeRiver Visualisierung, Lit. 09 und Abb. 2).

Zweidimensionale Daten besitzen zwei spezielle Dimensionen, die jeden Punkt eindeutig charakterisieren. Ein klassisches Beispiel für zweidimensionale Daten sind geographische Koordinaten. Für die Darstellung von zweidimensionalen Daten eignen sich zum Beispiel Standard X-Y-Plots.

Multi-dimensionale Daten: Viele Daten besitzen mehr als drei Attribute und können von daher

nicht mittels einfacher 2D- oder 3D-Darstellungen visualisiert werden. Typische Beispiele für mehrdimensionale Daten sind Tabellen in relationalen Datenbanken, die oft mehrere hundert oder sogar tausend Attribute besitzen. In den meisten Fällen existiert keine einfache Abbildung dieser Attribute in die zweidimensionale Ebene, so dass neuartige Techniken für ihre Visualisierung benötigt werden. Die Parallele-Koordinaten-Technik (Lit. 10) ist ein Beispiel für eine Visualisierungstechnik, die auf einfache Art und Weise die Visualisierung von multidimensionalen Daten erlaubt (siehe Abb. 3).

Text und Hypertext: Im Zeitalter des World Wide Web sind Text und Hypertext zwei wichtige Datentypen. Text und Hypertext unterscheiden sich von den bisher vorgestellten Datentypen, da sie nur

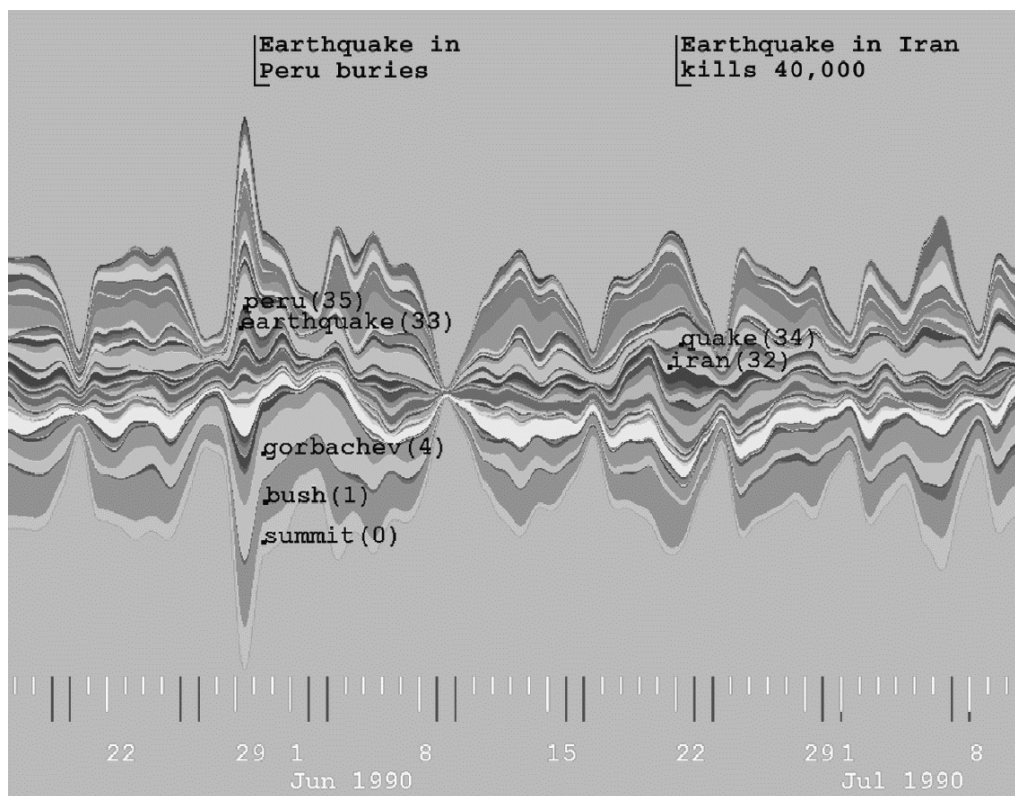


Abb. 2: Die ThemeRiver Visualisierungstechnik (Lit. 09) stellt die thematischen Veränderungen in großen Mengen von Textdokumenten über die Zeit dar. Der Fluss fließt von links nach rechts. Die sich verändernde Breite der einzelnen Segmente des Flusses visualisiert die thematischen Veränderungen. In unserem Beispiel wird ein Archiv von Associate Press Nachrichtenmeldungen von Juni-Juli 1990 visualisiert. Die wichtigsten Ereignisse dieses Zeitraums, wie das Erdbeben in Peru und das Gipfeltreffen zwischen Bush und Gorbatschow, können dabei gut identifiziert werden (im Original farbig; verwendet mit Erlaubnis von Pacific Northwest National Laboratory (PNNL); PNNL wird vom Battelle Memorial Institut für das US Department of Energy betrieben. © PNNL).

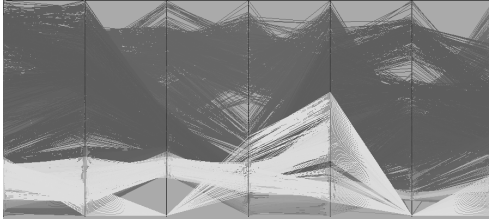


Abb. 3: Parallele Koordinaten-Visualisierung (im Original farbig)

schwierig durch einfache Datentypen mit fest definierter Größe beschrieben werden können. Viele der bekannten Standard-Visualisierungstechniken können deshalb nicht für ihre Visualisierung verwendet werden. Damit aber Text und Hypertext dennoch visualisiert werden können, werden sie in den meisten Fällen in sogenannte Beschreibungsvektoren (Feature-Vektoren) transformiert.

Ein Beispiel für eine einfache Transformation ist das Zählen aller nicht-trivialen Wörter im Text.

Hierarchien und Graphen: Eine Vielzahl von Datensätzen lässt sich nicht durch die bisherigen Datentypen beschreiben, da die Datensätze komplexe Beziehungen untereinander besitzen. Diese Beziehungen können mit Hilfe von Graphen modelliert werden. Ein Graph besteht aus einer Menge von Objekten, den Knoten, und Verbindungen zwischen diesen Objekten, den Kanten. Eine Hierarchie von Objekten kann dabei als eine spezielle Art von Graph aufgefasst werden, bei der die Verbindungen nur in Top-Down-Richtung verlaufen. Beispiele für Hierarchien und Graphen sind die Verbindungen in Telefon- oder Computernetzwerken, das Kaufverhalten von Kunden in E-Commerce-Angeboten, das Filesystem auf Festplatten und die Hyperlinks im World Wide Web.

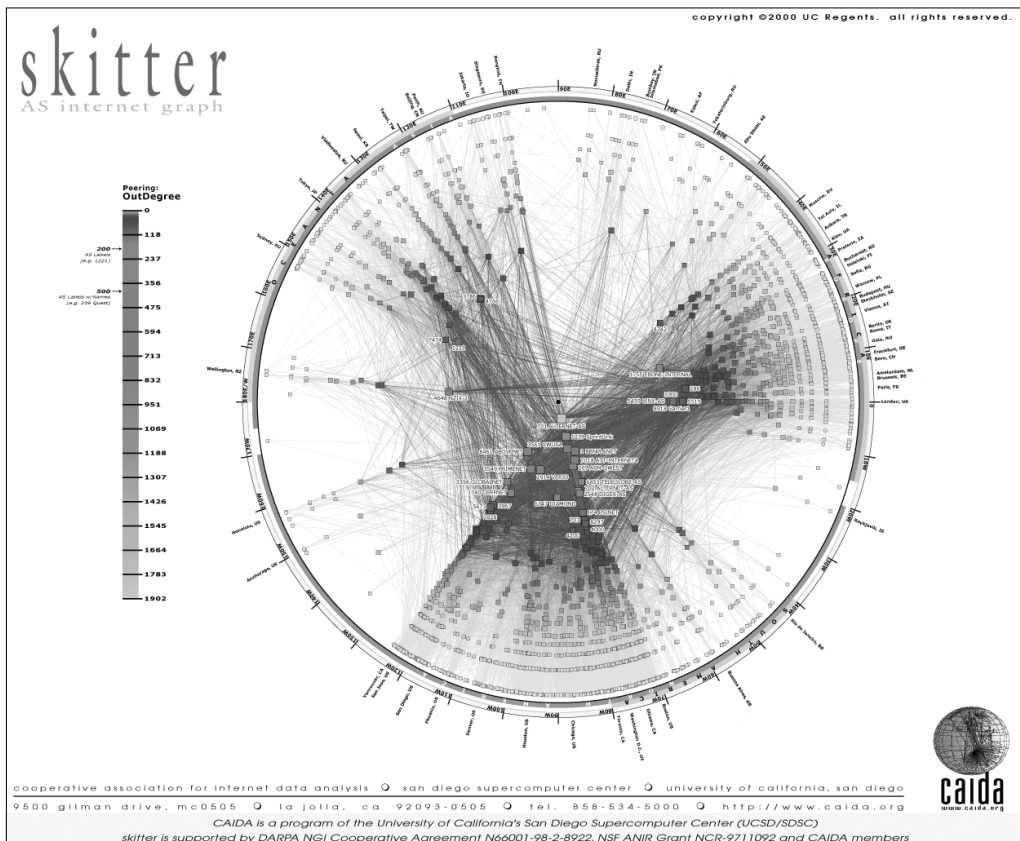


Abb. 4: Die Skitter-Visualisierungstechnik zeigt einen globalen Internet-Graphen. Die Knoten sind in Polarkoordinaten angeordnet, wobei die wichtigen Knoten mit einer hohen Anzahl von Verbindungen weiter innen liegen als die Knoten mit einer geringen Anzahl von Verbindungen (im Original farbig; verwendet mit Erlaubnis von CAIDA – Cooperative Association for Internet Data Analysis, k claffy „Skitter Graph“ Internet Map © 2000 Regents of the University of California).

Algorithmen und Software: Eine weitere interessante Klasse von Datentypen sind Algorithmen und Software. Die Erstellung großer Softwareprojekte ist nach wie vor eine Herausforderung. Das Ziel der Visualisierung ist die Vereinfachung der Softwareentwicklung durch ein besseres Verständnis der Algorithmen und des Quellcodes. Das kann zum Beispiel durch eine geeignete Visualisierung der Struktur der Software (z.B. des Funktions-Aufrufgraphen) geschehen oder durch eine Visualisierung aller von Fehlern betroffenen Teile des Programms erreicht werden

B 26.4 Visualisierungstechniken

Für die Visualisierung von Daten existieren eine Vielzahl von Visualisierungstechniken. Neben den weitverbreiteten Standard-2D/3D-Techniken, wie zum Beispiel x-y-(bzw. x-y-z-)Diagrammen, Balkendiagrammen, Liniendiagrammen usw., stehen heute eine Reihe weiterentwickelter Techniken zur Verfügung:

Geometrische Transformationen versuchen interessante Projektionen der multidimensionalen Datenmenge zu finden, um sie dann visuell darzustellen. Die Klasse der Visualisierungstechniken, die auf geometrischen Transformationen basieren, umfassen Verfahren der explorativen Statistik wie zum Beispiel Scatterplot-Matrizen (Lit. 02, Lit. 05) und Techniken, die unter dem Oberbegriff Projection Pursuit zusammengefasst werden.

Icon-basierte Visualisierungen: Eine visuelle Darstellung wird dabei durch die Abbildung der Attribute eines Datensatzes auf die Eigenschaften eines Icons erzeugt.

Pixel-Visualisierungen mit Abbildung jedes Datenwerts auf ein farbiges Pixel. Dabei werden die Pixel gemäß der Dimensionen gruppiert (siehe Abb. 5), bei der Recursive Pattern Technik (Lit. 16) in rechteckigen Teilbereichen und bei der Circle Segments Technik (Lit. 01) in Kreissegmenten. Die Pixel, die zu einem Datensatz gehören, sind dementsprechend über die Teilbereiche verstreut und stehen nur über die relative Position innerhalb der Teilbereiche in Beziehung. Über die visuelle Beziehung zwischen den Teilbereichen ist es möglich, lokale Beziehungen zwischen den Attributen, Korrelationen und Ausnahmen zu finden.

Geschachtelte Visualisierungen (Stacked Display Techniques) partitionieren die Daten gemäß eines oder mehrerer Attribute und visualisieren die Daten dann in hierarchischer Form, wobei die Wertebereiche der Attribute ineinander geschachtelt werden.

B 26.5 Interaktions- und Verzerrungstechniken

Für eine effektive Datenexploration sind Interaktions- und Verzerrungstechniken unverzichtbar. Durch Verwendung von Interaktionstechniken kann der Datenanalyst die Visualisierungen gezielt bezüglich der Explorationsziele verändern. Interaktionstechniken erlauben zudem eine Kombination verschiedener Visualisierungstechniken. Verzerrungstechniken helfen bei der Fokussierung auf Details, ohne dabei den Überblick über die Daten zu verlieren. Die Idee von Verzerrungstechniken (distortion techniques) ist das Hervorheben von Ausschnitten der Visualisierung mit vielen Details

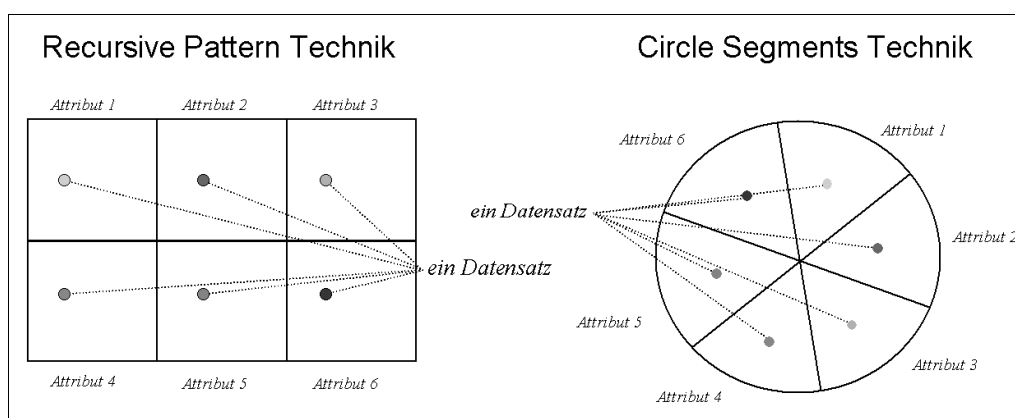


Abb. 5: Pixel-Visualisierung (im Original farbig)

unter Beibehaltung der Darstellung aller übrigen Teile der Visualisierung mit geringerem Detaillierungsgrad. In diesem Zusammenhang unterscheidet man zwischen dynamischen und interaktiven Techniken. Während bei dynamischen Techniken Veränderungen der visuellen Darstellung automatisch vorgenommen werden, erfolgen bei interaktiven Verfahren Veränderungen an der visuellen Darstellung durch Benutzerinteraktion. Man unterscheidet:

- Dynamische Projektion,
- Interaktive Filterung,
- Interaktives Zooming,
- Interaktive Verzerrung,
- Interaktives Linking and Brushing.

B 26.6 Zusammenfassung und Ausblick

Die Exploration großer Datenmengen ist ein sehr wichtiges, aber schwieriges Problem. Informations-Visualisierungs-Techniken können helfen, dieses Problem zu lösen. Die Verwendung von Informations-Visualisierungs-Techniken in zahlreichen Systemen hat gezeigt, dass sie zum Aufspüren interessanter Informationen (wie zum Beispiel Korrelationen, Cluster, funktionale Abhängigkeiten und Ausnahmen) in großen Datenmengen beitragen können. Die Zahl der Anwendungen, die Informations-Visualisierungs-Techniken für eine verbesserte Datenexploration verwenden, steigt stetig an. Beispiele für Bereiche, in denen die visuelle Datenexploration heute schon erfolgreich eingesetzt wird, sind Betrugserkennung, Marketing und Data Mining zum Beispiel in bio-molekularen Datenbanken. Die Aufgabe zukünftiger Forschung ist die Integration von Informations-Visualisierungs-Techniken mit traditionellen Techniken aus den Bereichen Statistik, maschinelles Lernen und Operations-Research. Erste Ansätze in diese Richtung sind zu finden und auch kommerzielle Systeme beginnen sich in diese Richtung zu orientieren. Der Vorteil einer solchen Integration ist eine Steigerung der Qualität und Effizienz des Datenexplorationsprozesses. Zusätzlich müssen die visuellen Data-Mining-Techniken mit großen Datenbank- und Data-Warehouse-Systemen verbunden werden. Das ultimative Ziel ist ein integriertes, leicht bedienbares und verständliches Datenexplorationssystem, das eine schnelle Exploration sehr großer Datenmengen ermöglicht.

Literatur

- Dieser Aufsatz basiert auf dem detaillierteren Beitrag von Daniel A. Keim: Datenvisualisierung und Data Mining. Datenbank Spektrum 2 (2) S. 30-39, 2002 - <http://www.datenbank-spektrum.de>
- 01 M. Ankerst; D. A. Keim and H.-P. Kriegel. Circle segments: A technique for visually exploring large multidimensional data sets. In Proc. Visualization 96, Hot Topic Session, San Francisco, CA, 1996.
 - 02 D. F. Andrews. Plots of high-dimensional data. *Biometrics*, 29: 125-136, 1972.
 - 03 D. Asimov. The grand tour: A tool for viewing multidimensional data. *SIAM Journal of Science & Stat. Comp.*, 6: 128-143, 1985.
 - 04 A. Buja; D. F. Swayne and D. Cook. Interactive high-dimensional data visualization. *Journal of Computational and Graphical Statistics*, 51: 78-99, 1996.
 - 05 W. S. Cleveland. *Visualizing Data*. AT&T Bell Laboratories, Murray Hill, NJ, Hobart Press, Summit NJ, 1993.
 - 06 S. Card; J. Mackinlay and B. Shneiderman. *Readings in Information Visualization*. Morgan Kaufmann, 1999.
 - 07 J. Eagan; M. J. Harrold; J. A. Jones and J. Stasko. Visually encoding program test information to find faults in software. In Technical Report, Georgia Institute of Technology, GIT-GVU-01-09, 2001.
 - 08 B. Huffaker; A. Broido; K. Claffy; M. Fomenkov; S. McCreary; D. Moore and O. Jakubiec. Visualizing internet topology at a macroscopic scale. In http://www.caida.org/analysis/topology/as_core_network, 2001.
 - 09 S. Havre; B. Hetzler; L. Nowell and P. Whitney. Themeriver: Visualizing thematic changes in large document collections. *Transactions on Visualization and Computer Graphics*, 2002.
 - 10 A. Inselberg and B. Dimsdale. Parallel coordinates: A tool for visualizing multidimensional geometry. In Proc. Visualization 90, San Francisco, CA, pages 361-370, 1990.
 - 11 D. Keim. Designing pixel-oriented visualization techniques: Theory and applications. *Transactions on Visualization and Computer Graphics*, 6 (1): 59-78, Jan-Mar 2000.
 - 12 D. Keim. Visual exploration of large databases. *Communications of the ACM*, 44 (8): 38-44, 2001.
 - 13 D. Keim and A. Herrmann. The gridfit approach: An efficient and effective approach to visualizing large amounts of spatial data. In Proc. Visualization 98, Research Triangle Park, NC, pages 181-189, 1998.
 - 14 D. A. Keim; M. C. Hao; U. Dayal and M. Hsu. Pixel bar charts: A visualization technique for very large multi-attribute data sets. *Information Visualization Journal*, 1 (1): 1-14, Jan. 2002.

- 15 D. A. Keim and H.-P. Kriegel. Visdb: Database exploration using multidimensional visualization. *Computer Graphics & Applications*, 6: 40-49, Sept. 1994.
- 16 D. A. Keim; H.-P. Kriegel and M. Ankerst. Recursive pattern: A technique for visualizing very large amounts of data. In *Proc. Visualization 95*, Atlanta, GA, pages 279-286, 1995.
- 17 J. Lamping; R. Rao and P. Pirolli. A focus + context technique based on hyperbolic geometry for visualizing large hierarchies. In *Proc. Human Factors in Computing Systems CHI 95 Conf.*, pages 401-408, 1995.
- 18 T. Munzner and P. Burchard. Visualizing the structure of the world wide web in 3D hyperbolic space. In *Proc. VRML '95 Symp*, San Diego, CA, pages 33-38, 1995.
- 19 R. M. Pickett and G. G. Grinstein. Iconographic displays for visualizing multidimensional data. In *Proc. IEEE Conf. on Systems, Man and Cybernetics*, IEEE Press, Piscataway, NJ, pages 514-519, 1988.
- 20 D. F. Swayne; D. Cook and A. Buja. *User's Manual for XGobi: A Dynamic Graphics Program for Data Analysis*. Bellcore Technical Memorandum, 1992.
- 21 B. Shneiderman. Tree visualization with treemaps: A 2D space-filling approach. *ACM Transactions on Graphics*, 11 (1): 92-99, 1992.
- 22 B. Shneiderman. The eye have it: A task by data type taxonomy for information visualizations. In *Visual Languages*, 1996.
- 23 B. Shneiderman. Dynamic queries, starfield displays, and the path to spotfire. In <http://www.cs.umd.edu/hcil/spotfire>, 1999.
- 24 H. Schumann and W. Müller. *Visualisierung: Grundlagen und allgemeine Methoden*. Springer, 2000.
- 25 B. Spence. *Information Visualization*. Pearson Education Higher Education publishers, UK, 2000.
- 26 C. Stolte; D. Tang and P. Hanrahan. Polaris: A system for query, analysis and visualization of multi-dimensional relational databases. *Transactions on Visualization and Computer Graphics*, 2002.
- 27 M. O. Ward. Xmdvtool: Integrating multiple methods for visualizing multivariate data. In *Proc. Visualization 94*, Washington, DC, pages 326-336, 1994.
- 28 C. Ware. *Information Visualization: Perception for Design*. Morgan Kaufman, 2000.
- 29 J. A. Wise. The ecological approach to text visualization. *Journal of the American Society for Information Science*, 50 (13): 1224-1233, 1999.
- 30 J. A. Wise; J. J. Thomas; K. Pennock; D. Lantrip; M. Pottier; A. Schur and V. Crow. Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In *Proc. Symp. on Information Visualization*, Atlanta, GA, pages 51-58, 1995.