



Hochschule für Angewandte Wissenschaften Hamburg
Hamburg University of Applied Sciences

Hausarbeit

Eduard Weigandt

Visual Analytics

Eduard Weigandt

Visual Analytics

Hausarbeit eingereicht im Rahmen der Vorlesung AW1

im Studiengang Master of Science Informatik
am Department Informatik
der Fakultät Technik und Informatik
der Hochschule für Angewandte Wissenschaften Hamburg

Betreuende Prüfer: Prof. Dr. Kai von Luck & Prof. Dr. Bettina Buth

Eingereicht am: 9. März 2014

Eduard Weigandt

Thema der Arbeit

Visual Analytics

Stichworte

Visualisierung, Visual Analytics, Anwendungsfall, Konferenzen

Kurzzusammenfassung

In der hier präsentierten Recherche werden die Ergebnisse und der Verlauf aus der Suche nach einem möglichen Anwendungsfall für ein zukünftiges Projekt vorgestellt. Dafür wurde das Forschungsfeld der Visual Analytics mit den relevanten Konferenzen sowie den derzeitigen Schlüsselpersonen angeschaut. Für einen besseren Einstieg ins Thema werden die gefundenen Techniken und Verfahren kurz vorgestellt. Von diesen werden interessante Anwendungsfelder für das nächste Studiensemester ausgewählt. Dazu zählt auch ein Ausblick den man für weitere Recherchen nutzen kann, um ein tieferes Verständnis über das Thema zu erlangen. Zum Schluss wird ein möglicher Anwendungsfall aus der Bioinformatik vorgestellt. Zudem werden anschließend Vorschläge zum weiteren experimentellen Vorgehen gemacht. Dazu gehören mögliche Fragestellungen die man während des Projekts untersuchen kann.

Inhaltsverzeichnis

1	Einführung	1
1.1	Motivation	1
1.2	Definition	1
1.3	Forschung und Entwicklung	2
2	Konzepte	3
2.1	Visualisierung	4
2.2	Data Mining	6
2.2.1	Methoden	7
3	Anwendungsfall	8
4	Umfeld & Fazit	10

1 Einführung

In den letzten Jahren wurde das Speichern von großen Datenmengen bezahlbarer, zudem wurden immer mehr Systeme geschaffen die zunehmend mehr Daten erzeugen und verarbeiten können. In dieser Situation kommt man an einen Punkt, wo man von einem Problem der Überladung durch Information spricht (Keim u. a., 2010). Es besteht die Gefahr das man sich in den Daten verliert und dadurch seine Aufgaben nicht erfüllen kann. Wegen dieses Problems kann es sein, dass man auf irrelevante Informationen schließt, wegen der falschen Verarbeitung oder Präsentation der Daten. Dadurch werden Forschungsgebiete interessant, die eine alternative Methoden zur *Daten Exploration* und *Erkenntnisgewinnung* anbieten.

1.1 Motivation

Es steckt ein großes Potential in der effektiven Auswertung von großen Datenmengen die zu einem Erkenntnisgewinn führen, denn die bisherigen Systeme kommen mit der Vielfalt der Daten oder der gestellten Aufgaben an ihre Grenzen. Die verfügbaren Analyse-Verfahren sind für sich alleine gesehen mächtig, jedoch reichen diese oft nicht aus, um eine komplexe Aufgabe automatisch zu lösen. Eine Datenquelle könnte eine Fernsehnachricht, eine E-Mail oder ein Schriftstück sein. Daraus kann man schon erkennen, dass die Daten auch in unstrukturierter Form vorliegen können. Nach Keim u. a. (2010) kann man vollautomatische Analysen nur bei wohl-definierten und gut-verstandenen Problemen einsetzen, weil man vor dem Problem steht dass die erlangten Ergebnisse dieser Verfahren nur schwer nachzuvollziehen sind.

Aus den oben genannten Gründen wird Visual Analytics in den letzten Jahren immer beliebter in der Erforschung, weil es ein Bindeglied zwischen der Analyse und dem Experten, der eine Aufgabe bzw. Fragestellung nachgeht, darstellt. Durch Visual Analytics soll es den Menschen ermöglicht werden einen besseren Einblick in die Daten zu bekommen.

1.2 Definition

Unter Visual Analytics versteht man die unterschiedlichsten Forschungsfelder. Die fachübergreifende Natur des Gebiets ist zum einen die große Schwierigkeit und zum anderen eine gewollte Entscheidung. Zu den Forschungsfeldern zählen z.B. *Visualisierung*, *Data Mining*, *Data Management* und *Stochastik* sowie viele weitere. Durch die Kombination dieser Bereiche soll es einem Experten ermöglicht werden, die gesammelten Daten aus einer bestimmten Domäne besser zu Verstehen und zu Erforschen. Ein grundlegender Moment zur weiteren Entwicklung in der Forschung zum Thema wurde in der Arbeit von Thomas u. Cook (2005) gelegt. Darin wurde eine detaillierte Übersicht des ganzen Themengebiets erstellt, aus der

eine Agenda für weitere Forschung entstanden ist, sowie viele neue weiterführende Projekte und Vorschläge zur Verbesserung der Forschung. Aus dieser Arbeit stammt auch die folgende Definition:

“Visual analytics is the science of analytical reasoning facilitated by interactive visual interfaces.”

Der Autor bringt zwei wichtige Punkte in der Definition von Visual Analytics vor. Zum ersten den analytischen Prozess und zum zweiten die Besonderheit dass dieser mit Hilfe von interaktiven visuellen Schnittstellen durchzuführen ist. Zusammengefasst besteht Visual Analytics aus den drei Komponenten: Der Benutzerinteraktion, Visualisierung und der Daten Analyse. In Abbildung 1.1 aus der Veröffentlichung

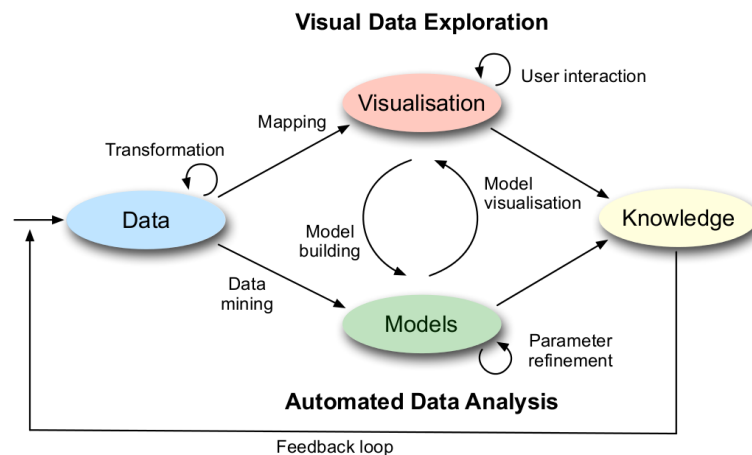


Abbildung 1.1: Verlauf der Erkenntnisgewinnung in einem Visual Analytics Prozess

von Keim u. a. (2010) sieht man zu diesem Zweck eine gute Übersicht des dahinter liegenden Prozesses. Die Kombination aus automatischen und visuellen Analysemethoden in Verbindung mit der Interaktion eines Menschen bilden den entscheidenden Teil in der Erlangung neuen Wissens. Vor dem eigentlich analytischen Prozess muss eine Vorverarbeitung bzw. Transformation der Daten vorangehen, um eine einheitliche Repräsentation zu bekommen mit der man Arbeiten kann. Diese vorverarbeiteten Daten kann man mittels Methoden aus dem Bereich des *Data Mining* analysieren, um ein Modell der vorhandenen Daten zu bekommen, welches der Analyst dann bewerten und ggf. anpassen kann. Dieser Schritt geschieht mit Hilfe von unterschiedlichen Visualisierungen. Die einzelnen Schritte führen dazu, dass man während der Erkundung der Daten neues Wissen erlangt, welches man dazu nutzen kann, um bessere Datenquellen auszuwählen.

1.3 Forschung und Entwicklung

Die Agenda von Thomas u. Cook (2005) zur Forschung und Entwicklung von Visual Analytics wurde im *National Visualization and Analytics Center (NVAC)* aus dem *Pacific Northwest National Laboratory*

(PNNL) ausgearbeitet. Das Interesse dieses Thema zu dieser Zeit voran zu treiben wurde durch das Bedürfnis nach mehr Sicherheit getrieben. Der Fokus lag darin Techniken und Werkzeuge zu entwickeln, mit deren Hilfe es möglich ist eine bessere Terrorabwehr zu ermöglichen und Verbesserungen im Katastrophenschutz zu realisieren. Begünstigt wurde dieses Vorhaben durch die Terroranschläge in den vorherigen Jahren. Aus diesem Grund wurde auch das Projekt vom *Department of Homeland Security* finanziert. Die führende Meinung besagte, dass die gesammelten Daten die benötigten Informationen zur Bekämpfung von Gefahren beinhalten und ein Experte diese durch eine visuell unterstützte Analyse besser finden kann. Das finale Resultat bestand aus einem vollständigen Überblick über das Themengebiet mit dem aktuellen Stand der Forschung und den zukünftigen Zielen und die dazu passenden Vorschläge zur Umsetzung. Einige Jahre später wurde auf europäischer Seite ein ähnliches Projekt ins Leben gerufen, um das Potential in der technologischen Weiterentwicklung und den wirtschaftlichen Gewinnen durch Visual Analytics zu fördern. Das *VisMaster* (Keim u. a., 2010) Projekt setzte auf den Erkenntnissen von Thomas u. Cook (2005) auf und präsentierte die neuen Entwicklungen und Ergebnisse seit der ersten Agenda. Die Finanzierung¹ dieses Projekts wurde durch die Europäische Union ermöglicht. Die wissenschaftliche Koordination wurde dabei von Prof. Dr. Daniel A. Keim² aus der Universität von Konstanz (UKON) übernommen und die allgemeine Koordination von Dr. Jörn Kohlhammer von dem Fraunhofer IGD³. Die beiden sind zwei wichtige wissenschaftliche Ressourcen in der Visual Analytics Community und gut bekannt.

Der Fokus bei dem europäischen Projekt lag in der Formung der zukünftigen Entwicklungen in den Informations- und Kommunikations-Technologien. Beide Veröffentlichungen liefern einen guten Startpunkt zur weiteren Vertiefung in das Thema, da viele erforschte und noch unbekannte Aspekte mit den entsprechenden Problemen in dem noch jungen Forschungsfeld erläutert werden. So z. B. werden Anwendungen und Verfahren im Bereich der Betrugsentdeckung, Analyse von Verbraucherdaten oder Gesundheitsdienstleistungen genannt. Ein interessanter Abschnitt in beiden Arbeiten ist der über die Bewertung von den eingesetzten Systemen und zur Erstellung von richtigen Metriken, die die Systeme unter einander vergleichbar machen. Nach der Aussage von Keim u. a. (2010) gibt es immer noch keine soliden Funde in dem Gebiet der Evaluierung von Visual Analytics Verfahren. Zu dem war es interessant bei weiteren Recherchen zur Kriminalitätsbekämpfung zu sehen, dass es sehr schwierig oder fast unmöglich war an konkrete Daten zu gelangen, die man zur Daten Analyse nutzen könnte. Das macht eine Kooperation mit einer entsprechenden Institution⁴ unausweichlich, wenn man diesen Anwendungsfall für ein experimentelles Projekt weiter verfolgen möchte.

¹http://cordis.europa.eu/programme/rcn/853_en.html (Zu letzt besucht am 04.03.2014)

²Profil: <http://www.vis.uni-konstanz.de/mitglieder/keim/>

³IGD: <http://www.igd.fraunhofer.de/Institut/Abteilungen/IVA?id=36&lang=ENGLISH>

⁴Bundeskriminalamt

2 Konzepte

2.1 Visualisierung

Nach Ltifi u. a. (2009) ist das Ziel von *Information Visualisierung* die Begabung eines Menschen zu nutzen, dass dieser auf natürliche und scharfsinnige Weise Informationen verstehen kann. Jeder kennt das Sprichwort: "Ein Bild sagt mehr als tausend Worte." Jedoch muss man hinzufügen, dass die Darstellung einer Information die Aussage stark beeinflussen kann. Weiter erklärt Ltifi u. a. (2009) das ohne einen Standard diese Informationen für jeden Menschen etwas anderes bedeuten kann. Damit ist es wichtig für jeden Anwendungsfall die effektivste Darstellung zu wählen. Die möglichen Arten können die unterschiedlichsten Formen annehmen, die wohl bekanntesten sind *Scatter Plots*, Graphen oder Bar Charts (Card u. a., 1999) die ersten beiden sieht man in Abbildung 2.1. In einem Graphen werden die Werte ei-

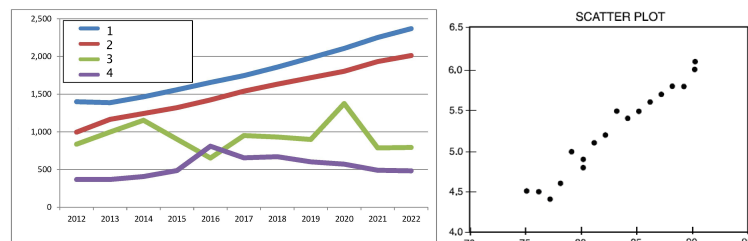


Abbildung 2.1: Beispiel eines Graphen und einem Scatter Plot

ner Funktion in den definierten Dimensionen aufgezeigt. Bei einem Scatter Plot wiederum werden viele Einzelwerte auf einer Zweidimensionalen Fläche aufgezeichnet. Diese Ansicht liefert eine gute Möglichkeit z.B. Ausreißer in den Werten zu identifizieren. Eine Alternative zu diesen Visualisierungen liefern z. B. *Clustering* oder das ikonografische Verfahren der *Chernoff Gesichter* (Chernoff, 1973). Die Cluster Analyse kann man gut mit einem Scatter Plot vergleichen, der sich gut zur visuellen Gruppierung von Werten in einem geografischen Raum eignet. Beim *Clustering* werden wie bei einem *Scatter Plot* die Beziehungen zwischen den Werten sichtbar gemacht.

In dem linken Bild von Abbildung 2.2 sieht man fünf identifizierte Bereiche (Cluster) und deren Verbindungen. In einem Bereich sind jeweils die Begriffe gruppiert die am engsten zu einem Thema in Beziehung stehen. Dabei ist die meist benutzte Form, die einzelnen Gruppierungen aufzuzeigen, eine Baumstruktur. Eine kuriose Variante der ikonografischen Darstellung von Daten mit mehreren Variablen sieht man im rechten Bild. Dabei werden die jeweiligen Variablen entweder in der Breite oder Höhe der einzelnen Bestandteile eines Gesichts repräsentiert. Bei dieser Art der Darstellung nutzt man die Fähigkeit des Menschen einen Ausdruck in einem Gesicht zu deuten. Einen Vergleich zwischen den geometrischen Visualisierungen in Form von *Scatter Plots* und den ikonografischen Darstellungen in Form von z.B. *Chernoff Gesichtern* findet man in Rabelo u. a. (2008). Dabei wurde herausgefunden dass die ikonografischen Arten sich für kontinuierliche und diskrete Arten von Werten eignen, aber schlecht in der Skalierbarkeit sind,

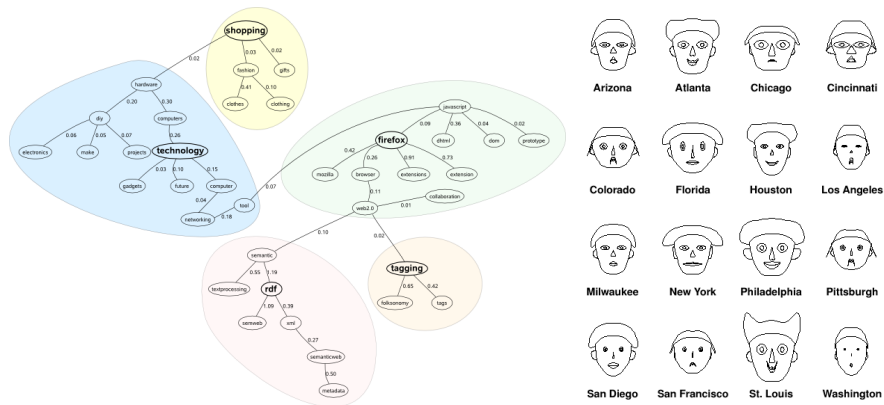


Abbildung 2.2: Beispiel von Clustering und Glyphen in Form von Chernoff Gesichtern

weil die abzubildende Variablenanzahl limitiert ist. Zu dem sind die vorhandenen Beziehungen zwischen den einzelnen Variablen schwer untereinander zu bewerten. Also muss man festhalten, dass es wichtig ist vor der Wahl einer Visualisierung sich damit kritisch auseinander zu setzen. Am Ende muss man also entscheiden welche Formen sich für die vorliegenden Daten und Fragestellungen am besten eignen.

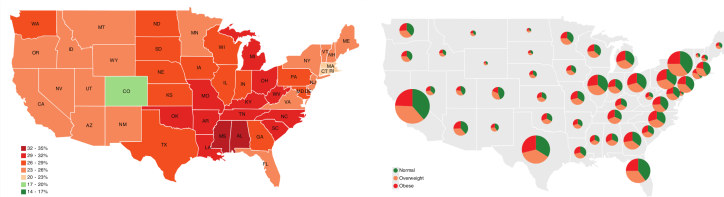


Abbildung 2.3: Visualisierungen zur Fettleibigkeit in Amerika (Teil1)

Am Anfang des Abschnitt 2.1 ist die Behauptung gefallen, dass die Darstellung einer Information die Aussage beeinflussen kann, dafür liefern die drei Bilder in Abbildung 2.3 und 2.4 ein gutes Beispiel (Heer u. a., 2010). In den drei unterschiedlichen Bildern werden jeweils Daten zum prozentualen Anteil von Fettleibigkeit in den U.S.A. dargestellt. Die unterschiedlichen Variationen der aufgeführten Werte sind jeweils in der Form und Farbe kodiert. In dem Artikel von Heer u. a. (2010) wird zudem darauf hingewiesen, dass die ursprüngliche Aussage in den Farbschattierungen durch die darunterliegende Fläche beeinflusst werden kann.

Weitere Forschung. Das Forschungsgebiet in der Visualisierung ist noch lange nicht ausgeschöpft, so wurde auf der kürzlich stattgefundenen *IEEE Visualization Conference 2013* (IEEE VIS) in der Präsentation eines Gastredners, dass Thema "*Information Visualization: Challenges and Opportunities*" vorgestellt, bei dem neue Herausforderungen identifiziert wurden. IEEE VIS ist eine der ältesten und größten Konferenzen auf dem Gebiet der *Information Visualisierung*. Somit bleibt die Entwicklung von neuen Formen der Repräsentation von Daten immer noch ein spannendes Thema, da die Entwicklung noch viel

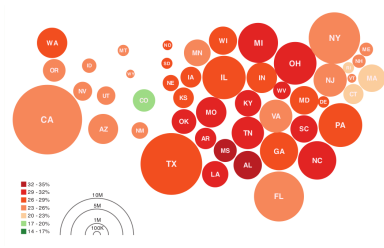


Abbildung 2.4: Visualisierungen zur Fettleibigkeit in Amerika (Teil2)

Potential besitzt und die Bemühungen aus Thomas u. Cook (2005) und Keim u. a. (2010) noch nicht abgeschlossen sind.

Fazit Um eine effektive Visualisierung zu entwickeln, muss eine intensivere Recherche in dem Bereich der Evaluierung erfolgen, sowie den vorhandenen Techniken, um die beste Variante der Darstellung für die vorliegende Fragestellung zu wählen. Die Arten der Daten geben einen weiteren kritischen Punkt zur Auswahl der richtigen Visualisierung, je nach Wahl kann die Skalierung des Verfahrens leichter erfolgen oder unnötig erschwert werden. Dabei darf man nicht vergessen, dass man die Domäne der Daten verstanden haben muss, jedoch zu welchem Grad es nötig ist, die Domäne verstanden zu haben, müssen weitere Recherchen am Anwendungsfall liefern. Zu dem kann die Aussage einer Information durch Formen und Farben in ihrer Kombination verfälscht werden. Es sind also noch reichlich Fragen offen, die man in Verbindung mit einem Anwendungsfall untersuchen muss.

2.2 Data Mining

Bei Visual Analytics kommen viele fachübergreifende Themenbereiche zusammen, die sich gegenseitig in der endgültigen Analyse unterstützen sollen. Wie in der Definition in dem Abschnitt 1.2 gesehen, bestehen diese Bereiche aus der Visualisierung, Benutzerinteraktion und der Datenanalyse. Der nun folgende Abschnitt liefert einen kleinen Ausschnitt aus dem Bereich zur Extraktion von Informationen aus großen Datenmengen. In der Arbeit von Järvinen u. a. (2009) werden die einzelnen Methoden zum Thema des Data Mining in die folgenden (frei übersetzten) Kategorien eingeteilt:

- **Vorhersage:** Methoden, die etwas Unbekanntes vorherzusagen unter Verwendung der vorliegenden Variablen. Davor werden jedoch vordefinierte oder ausgesuchte Ziele erstellt, an denen man sich orientiert.
- **Beschreibung:** Suche nach Mustern, die die Daten im Allgemeinen beschreiben.

Data Mining ist schon ein lang erforschtes Feld, welches man aus dem Prozess des *Knowledge discovery and data mining* (KDD) her kennt. In KDD geht es um die semi- oder vollautomatische Analyse von großen Datenmengen unabhängig von der Domäne oder dem Wissen eines Experten. Im Gegensatz dazu nutzen Methoden zur Visualisierung bestimmtes Hintergrundwissen, Kreativität und Intuition, um ein demgegenüber kleines Problem zu lösen (Keim u. a., 2010). Die meisten Bemühungen im Einsatz von

KDD Techniken in Verbindung mit Visual Analytics findet man in den letzten Jahren in der Bioinformatik. Das sieht man z.B. am Keynote Thema "*Visualization vs Expertise: Case Studies in Lexicography and Genomics*", welches auf der *IEEE VIS 2013* von Erez Lieberman Aiden präsentiert wurde. Darin wurden Fallstudien und Erfahrungen vorgestellt, die beim Erforschen von Genom Faltungsstrukturen in Verbindung mit der Menschheitsgeschichte unter zur Hilfenahme von Visualisierungen erzielt wurden.

2.2.1 Methoden

In der Kategorie der Vorhersage gibt es viele unterschiedliche Methodenarten. Die entscheidenden sind jedoch *classification* und *regression*. Bei der *classification* baut man sich einen Trainingssatz an Daten auf, anhand deren man die neuen Daten dann in Klassen einordnen kann. Die exemplarischen Techniken dafür bilden die *Entscheidungsbäume*, *Regelsysteme* oder *Neuronalen Netze*¹. Die Verfahren zur *regression* nutzen lineare und nicht-lineare Modelle zur Vorhersage von einem Wert basierend auf den bekannten Variablenwerten. Ein Beispiel dafür sind Statistiken, die einen Trend vorhersagen wie man in Abbildung 2.6c sehen kann. Eine Anwendung finden diese Techniken z.B. in Berechnung der möglichen Verkäufe durch die gemachten Werbeausgaben oder die Vorhersage des Aktienmarktes. Das Themenfeld der Vorhersage von Ereignissen ist aktuell in der *Special Interest Group on Knowledge Discovery and Data Mining 2013* (SIGKDD) Konferenz von einem Hauptredner behandelt worden (Scott u. Varian, 2013). In dieser Präsentation wurden die Ergebnisse einer Untersuchung vorgestellt, bei der das Verbraucherverhalten und die getätigten Waffenkäufe durch Suchmaschinenanfragen vorhergesagt wurden. Dem gegenüber stehen bei der Kategorie der Beschreibung einmal das *clustering*, *association rule discovery*, *sequential pattern discovery* und *anomaly detection*.

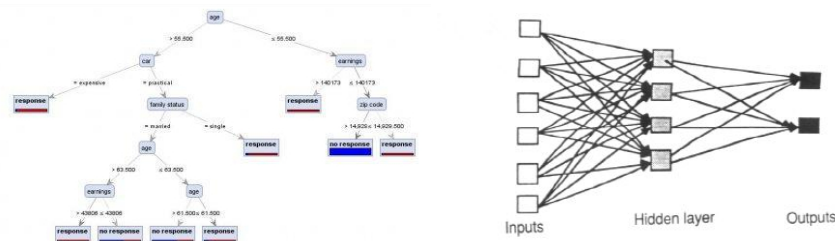


Abbildung 2.5: Links ein Entscheidungsbaum und rechts ein Neuronales Netz.

Das *clustering* basiert auf einem Algorithmus, der eine Einteilung von Werten in Gruppen macht, wie bei der gleichnamigen visuellen Methode in der Visualisierung im Abschnitt 2.1. Ein beliebtes Anwendungsfeld dieser Technik ist z. B. die Erkennung von verschiedenen Kundentypen zur Entwicklung einer besseren Strategie in der Vermarktung.

Nach Järvinen u. a., 2009 konzentriert sich *sequential pattern discovery* auf das Auffinden von Informationen, die in dem Ablauf der zeitlichen Ereignisse auftauchen. Dabei werden Regeln gefunden, die die Abhängigkeiten in den sequentiellen Abfolge beschreiben (siehe Abbildung 2.6b). Ein großer Anwendungsbereich dafür bildet der Telekommunikationssektor, in dem durch diese Analysetechniken die

¹http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html

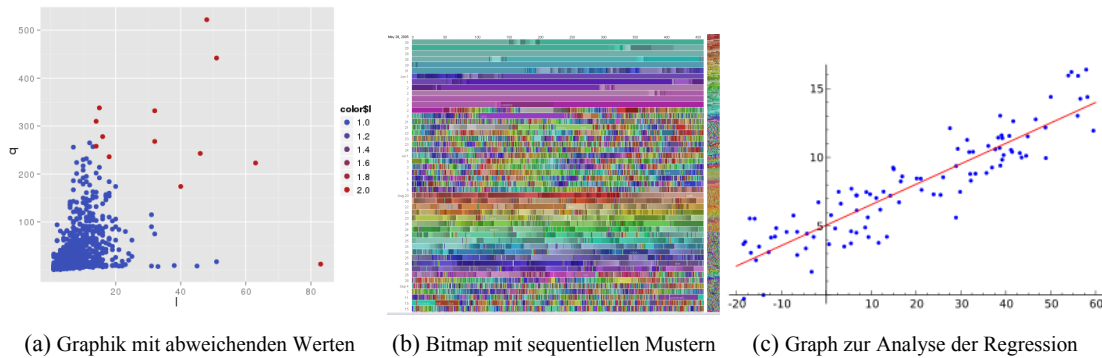


Abbildung 2.6: Beispiel Visualisierungen einzelner Techniken.

aufgezeichneten Daten von Sensoren ausgewertet werden, um einen möglichen zukünftigen Ausfall dieser zu entdecken.

Bei den Methoden im Bereich der *anomaly detection* wird nach Ausnahmen bzw. Anomalien in den Werten gesucht, welche vom normalen Verhalten abweichen. Ein wichtiger Anwendungsbereich für solche Algorithmen ist die Betrugsentdeckung in Finanzinstituten. Eine grafische Darstellung sieht man in Abbildung 2.6a, wo man einen *Scatter Plot* mit blau sowie rot markierten Werten sieht. Die roten sind die sichtbaren Abweichungen zu den blauen Punkten. Wieder erkennt man eine Beziehung zwischen *Data Mining* und der *Visualisierung*, die die gleichen Ziele und Fragestellungen verfolgen und unterstützend eingesetzt werden können, um etwas zu verifizieren.

Fazit. Wie man sieht, gibt es viele unterschiedliche Algorithmen zur Extraktion von Informationen. Es ist spannend zu sehen, wie man diese mit den entsprechenden *Visualisierungen* verbinden kann, um eine bessere Gewinnung von neuem Wissen zu unterstützen. Bei einem entsprechenden Anwendungsfall muss dann untersucht werden, wie gut diese dann in der Kombination funktionieren.

3 Anwendungsfall

In Abschnitt 2.1 werden einige Anwendungsfälle zur Bioinformatik vorgestellt. Die Themen zur Untersuchung vom menschlichem Genom (Lieberman-Aiden u. a., 2009) sind wichtig, um den Menschen besser zu verstehen. Darüber hinaus können diese Erkenntnisse helfen bestimmte genetische Krankheiten vielleicht in der Zukunft zu heilen. Diese Aussichten machen die weitere Erforschung im Zusammenhang mit Visual Analytics in diesem Gebiet erstrebenswert. Davon abgesehen ist die Untersuchung des Verlaufs einer Krankheit in Form von einer Darstellung der Entwicklung einer Pandemie oder Epidemie entscheidend, um Menschenleben zu retten (Livnat u. a., 2012). Hierzu sieht man in Abbildung 3.1 ein Visual Analytics System zur Exploration und Simulation, was für solch einen Fall entwickelt wurde. Den Entwicklern war es wichtig den Fokus auf die passenden *Visualisierung* und *Datenexploration* zu legen,

um bei der Entscheidungsfindung eine optimale Hilfe zu sein. Dabei war es entscheidend den ausgeführten Prozess der Experten aus dem Gesundheitswesen zu verstehen, um ein Werkzeug zu bauen was den Benutzern neue Wege aufzeigen sollte.

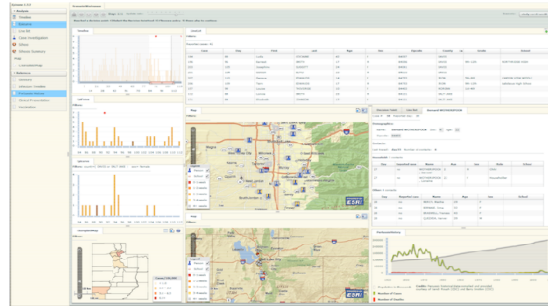


Abbildung 3.1: Screenshot von Epinome

Die schwierige Bewertung der eingesetzten Techniken wurde dadurch vereinfacht, indem ausgewählte Szenarien und Krankheitsverläufe in den Daten eingebettet wurden. Damit die erzielten Erkenntnisse bewertet werden konnten, mussten die möglichen Entwicklungen der Krankheitsausbrüche limitiert werden. So wurde die Größe des Problemraums erfolgreich eingeschränkt.

Fragestellungen. Wie man in der Abbildung 3.1 sehen kann, bestehen die verwendeten Visualisierungen aus geographischen Karten, einem Flächenkartogramm¹ und mehreren Balkengraphen. In den Karten sieht man ein wichtiges Konzept aus der Visual Analytics, welches besagt das man sich zuerst eine Übersicht verschaffen muss und dann durch Zoomen und Filtern zusätzliche Details ein- oder ausblenden kann (Shneiderman, 1996). In einem experimentellen Anwendungsfall könnte man weitere Arten zur Darstellung implementieren und mit denen aus *Epinome* in einer kleinen Studie vergleichen. Die Voraussetzung die Domäne gut verstanden zu haben, könnte man umgehen wenn man grundlegende Aspekte aus den Visualisierungen ausarbeitet, die allgemeingültig bei der Darstellung von Informationen sind. Trotzdem müssen sich die stochastischen Verfahren (Shatnawi u. a., 2013) aus der Domäne angeschaut werden, um eine gute Simulation einer Epidemie oder Pandemie nachzustellen. Danach wäre ein möglicher Ausblick bei der Vorhersage von Epidemien in bestimmten Szenarien als eine spannende Erweiterung zu sehen. Bei allen Schritten sollte versucht werden, die Erkenntnisse, die durch ein bestimmtes Verfahren gewonnen wurden für die Bewertung aufzuzeigen.

Die einzige Komponente von Visual Analytics die bisher noch nicht erwähnt wurde besteht in der Benutzerinteraktion mit dem System, da das Ziel des ausgesuchten Anwendungsfalls in der experimentellen Erkundung des Forschungsgebietes liegt. Das zu entwickelnde System wird einen prototypischen Charakter besitzen, der eine gute Evaluierung der Benutzerinteraktion erschwert. Trotzdem kann man sich vorbehalten eine Studie durchzuführen, bei der man unterschiedliche Verfahren vergleicht und in der Interaktion bewertet.

¹Nach Wikipedia: "Eine Choroplethenkarte (auch Flächenkartogramm) ist eine thematische Karte, bei der die Gebiete im Verhältnis zur Verteilungsdichte des thematischen Objektes eingefärbt, schattiert, gepunktet oder schraffiert sind."

4 Umfeld & Fazit

Konferenzen. Im Verlauf der Recherche wurde die eine oder andere relevante Konferenz für Visual Analytics schon genannt. Die *IEEE Visualization Conference*¹, welche einmal jährlich stattfindet, vereint drei große Veranstaltungen zur gleichen Zeit. Darunter einmal die *Conference on Visual Analytics Science and Technology* (VAST) zur direkten Weiterentwicklung des Visual Analytics Gebiets in den unterschiedlichsten Domänen². Darüber hinaus findet bei der VAST jedes mal ein Wettbewerb statt, um die entwickelten Systeme untereinander zu messen. Zu den weiteren Veranstaltungen zählt der *Scientific Visualization Contest* (SCVIS) und die *Information Visualization Conference* (INFOVIS). Darüber hinaus veranstaltet die ACM³ jedes Jahr die *Special Interest Group on Knowledge Discovery and Data Mining* Konferenz seit 1995. Eine weitere interessante Konferenz ist die *Special Interest Group on GRAPHics and Interactive Techniques*⁴ (SIGGRAPH), die sich speziell für die Recherche für weitere Visualisierungen eignet.

Außerhalb der großen Veranstaltungen finden auch kleinere Treffen von Forschern und Anwendern statt, wie z. B. bei der *Visualization and Data Analysis*⁵ (VDA) in die auch einer der Koordinatoren aus dem *Pacific Northwest National Laboratory* kommt. In Europa findet man noch die *Eurographics Conference on Visualization*⁶ (EuroVis) die ähnliche Themen wie die IEEE VIS behandelt.

Weiter Einrichtungen und Projekte. Zusätzlich zu den Konferenzen gibt es viele Institute mit spannenden Projekten. Bei den Fraunhofer ist das IGD⁷ und das IAIS⁸ für Visual Analytics interessant. Beim ersteren werden neue Arten der Visualisierung in Simulationen entwickelt sowie Graphische Informationssysteme und beim letzteren direkt das Thema Visual Analytics⁹. Das Fraunhofer IAIS steht dabei in direkter Kooperation mit der Forschungsgruppe *Knowledge Discovery and Machine Learning*¹⁰ aus der Uni Bonn. Die *Deutsche Forschungsgemeinschaft* (DFG) betreibt zusätzlich das Projekt *Scalable Visual Analytics*¹¹ und das *Bundesministerium für Bildung und Forschung* über die Uni Konstanz das Projekt *Visual Analytics for Security Applications*¹² (VASA).

Fazit. In der Recherche wurde sichtbar, dass eine kritische Auseinandersetzung mit den passenden Arten der *Visualisierung* nötig ist. Weiterhin ist die richtige Integration von *Daten*, *Visualisierung* und *Analyse*

¹<http://ieevis.org/>

²Anwendungen: Wissenschaft, Medizin, Gesundheitswesen, Medien, Business, Soziale Interaktion, Security und Investigative Analysen

³Association for Computing Machinery

⁴<http://www.siggraph.org/>

⁵<http://vda-conference.org/>

⁶<http://eurovis.swansea.ac.uk/>

⁷<http://www.igd.fraunhofer.de>

⁸<http://www.iais.fraunhofer.de/>

⁹<http://visualanalytics.info/>

¹⁰<http://www-kd.iai.uni-bonn.de/>

¹¹<http://www.visualanalytics.de/>

¹²<http://www.va-sa.net/>

von Anwendungsfall abhängig. Wichtig ist es die Aussage der Daten nicht durch eine falsche Darstellung zu beeinflussen. Es gibt viele *Data Mining* Methoden, die sich in Kombination mit bestimmten Visualisierungen gut oder schlecht eignen. Um die richtigen zu finden müssen diese im Anwendungsfall untersucht werden. Während der Recherche wurde die Bioinformatik als ein spannendes Feld entdeckt, dafür wurde sich die Anwendung *Epinome* angeschaut. Für das finale Projekt wurde aus diesem Grund der Anwendungsfall der Simulation einer Epidemie ausgewählt. Bei der Simulation liegt der Fokus auf dem Verlauf nach dem Ausbruch einer Krankheit. Für einen potentiellen Ausblick eignet sich die Entwicklung eines Systems zur Vorhersage von Krankheiten in ihrer Vermehrung in bestimmten Szenarien.

Literaturverzeichnis

- [Card u. a. 1999] Card, Stuart K. ; Mackinlay, Jock D. ; Shneiderman, Ben: *Readings in information visualization: using vision to think*. Morgan Kaufmann, 1999 4
- [Chernoff 1973] Chernoff, Herman: The use of faces to represent points in k-dimensional space graphically. In: *Journal of the American Statistical Association* 68 (1973), Nr. 342, S. 361–368 4
- [Heer u. a. 2010] Heer, Jeffrey ; Bostock, Michael ; Ogievetsky, Vadim: A Tour Through the Visualization Zoo. In: *Queue* 8 (2010), Mai, Nr. 5, 20:20–20:30. <http://dx.doi.org/10.1145/1794514.1805128>. – DOI 10.1145/1794514.1805128. – ISSN 1542–7730 5
- [Järvinen u. a. 2009] Järvinen, Paula ; Puolamäki, Kai ; Siltanen, Pekka ; Ylikärälä, Markus: Visual Analytics. Final Report / VTT Technical Research Centre of Finland. Espoo, Finland, 2009 (117). – VTT Working Papers 6, 7
- [Keim u. a. 2010] Keim, Daniel A. ; Kohlhammer, Jörn ; Ellis, Geoffrey ; Mansmann, Florian: *Mastering The Information Age-Solving Problems with Visual Analytics*. Florian Mansmann, 2010 1, 2, 3, 6
- [Lieberman-Aiden u. a. 2009] Lieberman-Aiden, Erez ; Berkum, Nynke L. ; Williams, Louise ; Imakaev, Maxim ; Ragozy, Tobias ; Telling, Agnes ; Amit, Ido ; Lajoie, Bryan R. ; Sabo, Peter J. ; Dorschner, Michael O. u. a.: Comprehensive mapping of long-range interactions reveals folding principles of the human genome. In: *science* 326 (2009), Nr. 5950, S. 289–293 8
- [Livnat u. a. 2012] Livnat, Y. ; Rhyne, T. ; Samore, M.: Epinome: A Visual-Analytics Workbench for Epidemiology Data. In: *Computer Graphics and Applications, IEEE* 32 (2012), March, Nr. 2, S. 89–95. <http://dx.doi.org/10.1109/MCG.2012.31>. – DOI 10.1109/MCG.2012.31. – ISSN 0272–1716 8
- [Ltifi u. a. 2009] Ltifi, H. ; Ben Ayed, M. ; Alimi, A.M. ; Lepreux, S.: Survey of information visualization techniques for exploitation in KDD. In: *Computer Systems and Applications, 2009. AICCSA 2009. IEEE/ACS International Conference on*, 2009, S. 218–225 4
- [Rabelo u. a. 2008] Rabelo, E. ; Dias, M. ; Franco, C. ; Pacheco, R.C.S.: Information Visualization: Which Is the Most Appropriate Technique to Represent Data Mining Results? In: *Computational Intelligence for Modelling Control Automation, 2008 International Conference on*, 2008, S. 1228–1233 4
- [Scott u. Varian 2013] Scott, Steven L. ; Varian, Hal R.: Bayesian variable selection for nowcasting economic time series / National Bureau of Economic Research. 2013. – Forschungsbericht 7
- [Shatnawi u. a. 2013] Shatnawi, M. ; Lazarova-Molnar, S. ; Zaki, N.: Modeling and simulation of epidemic spread: Recent advances. In: *Innovations in Information Technology (IIT), 2013 9th International Conference on*, 2013, S. 118–123 9
- [Shneiderman 1996] Shneiderman, Ben: The eyes have it: A task by data type taxonomy for information visualizations. In: *Visual Languages, 1996. Proceedings., IEEE Symposium on IEEE*, 1996, S. 336–343 9
- [Thomas u. Cook 2005] Thomas, James J. ; Cook, Kristin A.: *Illuminating the path: The research and development agenda for visual analytics*. IEEE Computer Society Press, 2005 1, 2, 3, 6