

Maschinelle Datenanalyse im Informationszeitalter – Können oder müssen wir ihr vertrauen?

Hans-Jörg Schulz, Thomas Nocke

Kurzfassung:

Die allgemeine Zunahme der zu verarbeitenden Datenmengen ist nicht nur im wissenschaftlichen Umfeld, sondern im gesamtgesellschaftlichen Rahmen zu beobachten. Diese Entwicklung führt zum vermehrten Einsatz rechnergestützter Analysemethoden, deren Komplexität im Zusammenspiel mit den gewaltigen Datenmengen zu Ergebnissen führen, die kaum mehr von Hand überprüft werden können. Diese Arbeit zeigt einige der aus dieser Entwicklung resultierenden Schwierigkeiten und Gefahren an ausgewählten Beispielen auf und diskutiert die Entwicklung intuitiver Mensch-Maschine-Schnittstellen als möglichen Ausweg.

1. Einleitung

Registrierkassen im Supermarkt, Mautbrücken über der Autobahn, elektronisches Rezept mit der Gesundheitskarte – die Durchdringung aller Lebensbereiche unserer heutigen Gesellschaft mit moderner IT hat eine für den Einzelnen unüberschaubare Informationslandschaft geschaffen. Dank fortgeschrittener Mess- und Überwachungstechnologie laufen in wissenschaftlichen, militärischen und kommerziellen Datenzentren riesige Informationsmengen auf, deren Analyse und Archivierung ohne maschinelle Hilfe schon lange nicht mehr zu bewerkstelligen ist. Doch was passiert eigentlich bei solch einer rechnergestützten Datenanalyse? Mathematik und Informatik haben in den vergangenen Jahrzehnten die Basis gelegt, um die Unterstützung des Analyseprozesses (engl. *Data Mining*) durch die Entwicklung immer neuer und immer komplizierterer Verfahren zu gewährleisten.

Ausgehend von einer Erläuterung des wissenschaftlichen Erkenntnisprozesses mit seinen unterschiedlichen charakterisierenden Aspekten, wird die maschinelle Datenanalyse nachfolgend in dessen Begriffsraum eingeordnet. Ferner werden mit dem Data Mining und der Visualisierung zwei prominente Beispiele maschineller Analysemethoden vorgestellt, deren Vor- und Nachteile aufgezeigt und schließlich zum Visuellen Data Mining kombiniert. Anschließend werden die aus dieser Kombination gewonnenen Prinzipien im Umgang mit rechnergestützter Datenanalyse verallgemeinert und der damit in engem Zusammenhang stehende Begriff der Intuition diskutiert. Der Abschluss dieser Arbeit widmet sich darauf aufbauend der Beantwortung der im Titel gestellten Frage nach dem Vertrauen.

2. Der wissenschaftliche Erkenntnisprozess

Der Wissensraum wird nach (Aamodt, Nygård 1995) schematisch durch die sogenannte *Wissenspyramide* dargestellt (s. Abb.1). Dabei verdeutlichen die drei unteren Stufen der Pyramide grob die Semiotik der zu analysierenden Daten (das sogenannte Datenmodell) und die oberen drei Stufen die aus den zugrunde liegenden Daten extrahierten Erkenntnisse. Der wissenschaftliche Erkenntnisprozess verläuft innerhalb der Wissenspyramide, also in der Regel von unten nach oben, wobei aber u.U. bereits gewonnenes Wissen aus den drei oberen Stufen mit einbezogen wird.



Abbildung 1: Wissenspyramide (Aamodt, Nygård 1995)

Der Erkenntnisprozess kann nach unterschiedlichen Aspekten charakterisiert werden. Dazu zählen u.a. dessen Motivation (Erkenntnisinteresse), dessen Methode (Erkenntnispraxis) und die Qualität der gewonnenen Erkenntnisse (Erkenntniskategorie).

2.1. Erkenntnisinteresse

Die forschungsleitende Fragestellung wird nach (Eberhard 1999:16) durch den Begriff des *Erkenntnisinteresses* beschrieben. Dabei wird (auch schon vor Eberhard) zwischen den folgenden drei Kategorien des Erkenntnisinteresses unterschieden:

- *Phänomenales Erkenntnisinteresse* fragt nach Beobachtungen und Fakten, die sich direkt aus den zu untersuchenden Daten ergeben. Derartige Erkenntnisse sind daher auf der Stufe der „Information“ in die Wissenspyramide einzuordnen.
- *Kausales Erkenntnisinteresse* fragt nach den Gründen für bestimmte Beobachtungen. Dies resultiert in ein Verständnis für das Beziehungsgefüge innerhalb des beobachteten Systems. Da kausale auf phänomenale Erkenntnisse aufbauen (es braucht eine Beobachtung, um nach deren Grund zu fragen), wird es auch darüber in die Stufe des „Wissens“ der Wissenspyramide eingeordnet.
- *Aktionales Erkenntnisinteresse* sucht nun nach konkreten Handlungsanweisungen, um innerhalb des o.g. Beziehungsgeflechts einen bestimmten Effekt zu induzieren. Da hierfür ein Verständnis der kausalen Zusammenhänge innerhalb des Systems nötig ist, werden aktionale über den kausalen Erkenntnissen in die Stufe der „Weisheit“ der Wissenspyramide eingeordnet.

2.2. Erkenntnispraxis

Unabhängig vom Erkenntnisinteresse können Erkenntnisse auf verschiedene Art und Weise gewonnen werden. In der Theorie unterscheidet man zwischen einer *konfirmativen*, hypothesenbeweisenden Erkenntnispraxis, und einer *explorativen*, ungerichteten Erkenntnispraxis (z.B. „Trial&Error-Prinzip“). Die jeweiligen Vor- und Nachteile dieser beiden Verfahren liegen auf der Hand: während die konfirmative Vorgehensweise die Anzahl der zu untersuchenden Variablen durch eine wohlüberlegte Hypothese verringert, birgt sie im

schlimmsten Fall auch die Gefahr übermäßiger Reduktion. Dahingegen hat die explorative Methode den Vorteil, Zusammenhänge aufzudecken, an die im Vorfeld gar nicht gedacht wurde, wobei man allerdings mit deren Verallgemeinerung vorsichtig sein muss, da jede noch so große Datenmenge die Realität i.d.R. immer nur auszugsweise widerspiegelt. Daher werden diese beiden Formen in der Praxis selten isoliert voneinander angewandt, sondern meist wechselseitig miteinander verwoben (Shneiderman 2002).

2.3. Erkenntniskategorien

Erkenntnisse können außerdem in zwei Erkenntniskategorien eingeordnet werden. Während *qualitative Erkenntnisse* einen entdeckten Sachverhalt durch Einordnung desselben in semantische Kategorien beschreiben, charakterisieren *quantitative Erkenntnisse* einen Sachverhalt durch konkrete numerische Werte. Eine Stufung oder Hierarchisierung dieser Erkenntniskategorien ist dabei nicht möglich, da der Erkenntnisweg in beide Richtungen beschriftet werden kann: zum einen ist eine quantitative Untermauerung erster qualitativer Beobachtungen denkbar, zum anderen aber auch die Übersetzung quantitativer (Mess-) Ergebnisse in allgemeinverständliche qualitative Aussagen.

3. Die rechnergestützte Datenanalyse

Die heutige Rechentechnik ist in der Lage, den wissenschaftlichen Erkenntnisprozess unabhängig von Erkenntnisinteresse und -praxis zu unterstützen. So wird der Computer häufig als Analysewerkzeug genutzt, um auch bei großen Datenmengen phänomenalem und kausalem Erkenntnisinteresse nachgehen zu können. Aktionales Erkenntnisinteresse wird dahingegen in der Regel kaum durch analytische Verfahren, sondern eher durch Computersimulationen rechnerisch unterstützt. Obwohl es prinzipiell denkbar wäre, auch Verfahren der Datenanalyse zur Gewinnung aktionaler Erkenntnisse zu nutzen, ist dieser Ansatz in der Praxis kaum vertreten und soll daher auch nicht im Fokus dieser Arbeit liegen.

Je nach Bedarf unterstützt die heutige Rechentechnik die Datenanalyse auf zwei verschiedene Arten: entweder übernimmt die Technik die komplette Analyse oder sie bereitet die Daten lediglich in einer Weise auf, dass der Nutzer die Daten darauf basierend selbst analysieren kann. Für beide Vorgehensweisen wird nachfolgend je ein Beispiel diskutiert.

3.1. Data Mining

Der Begriff *Data Mining* bezeichnet die rechnergestützte Analyse großer Datenmengen um darin vermutete Zusammenhänge zu bestätigen oder unvermutete Beziehungen zu entdecken. Diese Zusammenhänge oder Beziehungen werden als *Muster* (engl. *patterns*) bezeichnet und häufig durch Methoden der Statistik und der künstlichen Intelligenz erschlossen. Das Data Mining wird i.d.R. auf Daten angewandt, die eigentlich für ganz andere Zwecke erhoben wurden. Ein aktuelles Beispiel ist die Verwendung von zu Abrechnungszwecken gespeicherten Telefonverbindungsdaten für das Aufspüren von Terrorzellen in den Vereinigten Staaten (Mena 2004). Aufgrund dieser Entkopplung zwischen dem primären Zweck der Daten und der Intention des Data Mining wird das Data Mining auch als *sekundäre Datenanalyse* bezeichnet (Hand, Mannila, Smyth 2001).

Das Data Mining birgt allerdings auch Gefahren. So ist es durch die zunehmende Verarbeitungsgeschwindigkeit moderner Computer möglich, immer kompliziertere und ausgefeiltere Analyseverfahren in akzeptabler Rechenzeit auf immer größere Datenmengen anzuwenden. Fehlendes Verständnis für die Funktionsweise dieser Verfahren und den

Bedeutungszusammenhang ihrer Ergebnisse kann somit auf Seite des Anwenders zu deren Einsatz als Blackbox führen. In der Praxis bedeutet dies, dass die Verfahren zwar lediglich über einige Parameter an die eigenen Bedürfnisse und die Charakteristika der zu analysierenden Daten angepasst werden müssen, diese Anpassung ohne detailliertes Wissen um die intern verwendeten Verfahren den Laien jedoch rasch überfordert. Doch die unbedachte Parametrisierung resultiert u.U. in Fehlinterpretationen, die auf Grund der schiereren Datenmenge von Hand nicht falsifizierbar sind. Ferner führt die Fülle an existierenden Analyseverfahren zuweilen zu dem paradoxen Problem, dass unterschiedliche, aber jeweils durchaus plausible Analysemethoden zu voneinander verschiedenen Ergebnissen gelangen und somit Raum für projektive Interpretationen bieten.

Ein Beispiel, an dem diese Effekte deutlich werden ist die Diskussion um die sogenannte *Hockeyschläger-Kurve*, die weithin als Nachweis für die globale Klimaerwärmung gilt (Mann, Bradley, Hughes 1998). Doch eine zweite Analyse (McKittrick, McIntyre 2005) derselben Ausgangsdaten kommt zu einem etwas anderen Ergebnis, welches zwar ebenfalls einen rapiden Temperaturanstieg in den letzten Dekaden aufzeigt, jedoch zusätzlich einen ähnlichen Anstieg der Temperatur im Mittelalter aufweist (vgl. Abb. 2). Bis heute konnten sich beide Forschergruppen nicht einigen, welche der komplizierten Analysemethoden denn nun die geeignetere ist und damit, welches Diagramm die Realität am besten wiedergibt.

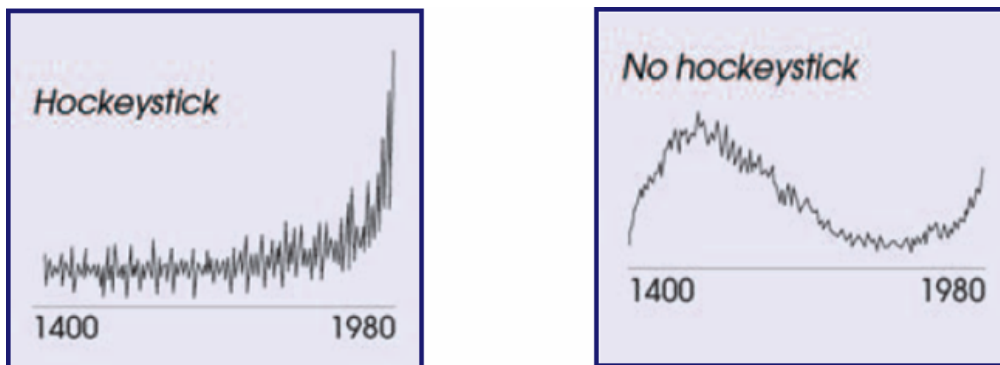


Abbildung 2: Die „Hockeyschläger-Kurve“ von M.Mann (1998) auf der linken Seite und die alternative Darstellung von McKittrick&McIntyre (2005) auf der rechten Seite

3.2. Visualisierung

Die *Visualisierung* bezeichnet das Erzeugen geeigneter visueller Repräsentationen gegebener Daten, um deren effektive Auswertung durch den Anwender zu ermöglichen (Schumann, Müller 2000:4). Damit ist die Visualisierung ein Beispiel für die Aufbereitung der Daten in eine Form, die dem Anwender den Zugang erleichtert und es ihm erlaubt, seine Analysen selbst durchzuführen. Dies geschieht i.d.R. durch Interaktion mit der Visualisierung, also durch eine Manipulation der grafischen Ausgabe etwa durch Zoom- oder Rotationsfunktionen oder z.B. auch durch die manuelle Umsortierung von Koordinatenachsen.

Abb. 3 zeigt zwei visuelle Analyseschritte bei der Auswertung „heißer Sommer“ in den Jahren 1893-1997 um Potsdam. Während auf der linken Seite die einzelnen Jahre in Dekaden abgetragen wurden, hat man auf der rechten Seite dieses Intervall interaktiv von zehn auf sechs Jahre reduziert. Diese geringfügige Alteration der Visualisierung durch den Anwender liefert bereits erste Resultate. Denn während in der linken Darstellung keinerlei Regelmäßigkeit zu entdecken ist, wird aus der rechten Darstellung sofort deutlich, dass in jedem fünften und sechsten Jahr anscheinend ein vergleichsweise kühler Sommer auftritt.

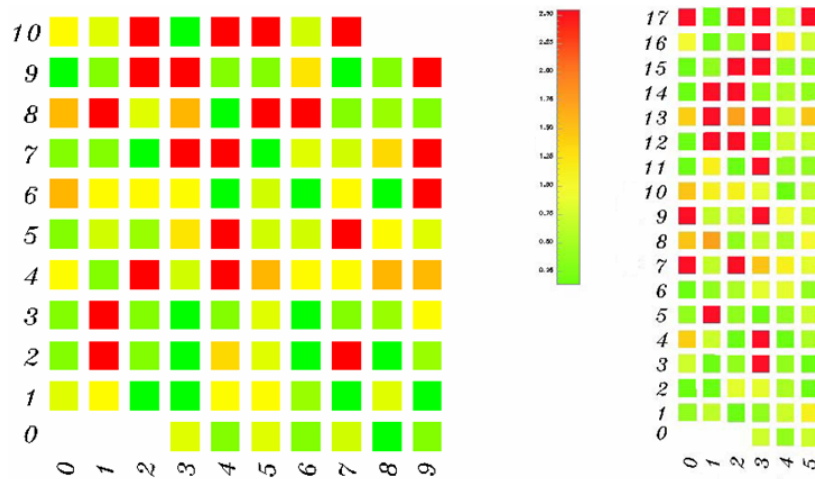


Abbildung 3: Heiße Sommer von 1893-1997 (Messstation Potsdam)

Das obige Beispiel zeigt aber auch, wie sensibel der Analyseprozess auf die Wahl geeigneter Darstellungen reagiert. So kann eine ungeeignete visuelle Repräsentation im besten Fall die Analyse verkomplizieren – im schlimmsten Fall jedoch falsche, in den Daten eigentlich nicht vorhandene Zusammenhänge suggerieren.

3.3. Die Kombination der Ansätze zum Visuellen Data Mining

Während also das automatische Data Mining die Datenanalyse rechnerisch durchführt und dem Nutzer i.d.R. nach ihrem Abschluss die Interpretation der Ergebnisse überlässt, dient die Visualisierung in erster Linie der grafischen Aufbereitung der Daten, um dem Nutzer selbst die Analyse zu ermöglichen. Der Vorteil der Analyse entweder durch den Computer oder den Nutzer liegt in deren unterschiedlichen Analysefähigkeiten, wie sie in Abb. 4 dargestellt sind. So kann je nach Bedarf eine entsprechende Analyseverfahren ausgewählt werden. Sollen große Datenmengen lediglich durchsucht und numerisch erfasst werden, ist sicherlich die Verwendung des Data Mining sinnvoll. Soll jedoch auch Allgemeinwissen und Kreativität mit in die Analyse einfließen, dann ist der Wahrnehmungsapparat des Menschen meist überlegen und sollte durch geeignete Visualisierungen unterstützt werden.

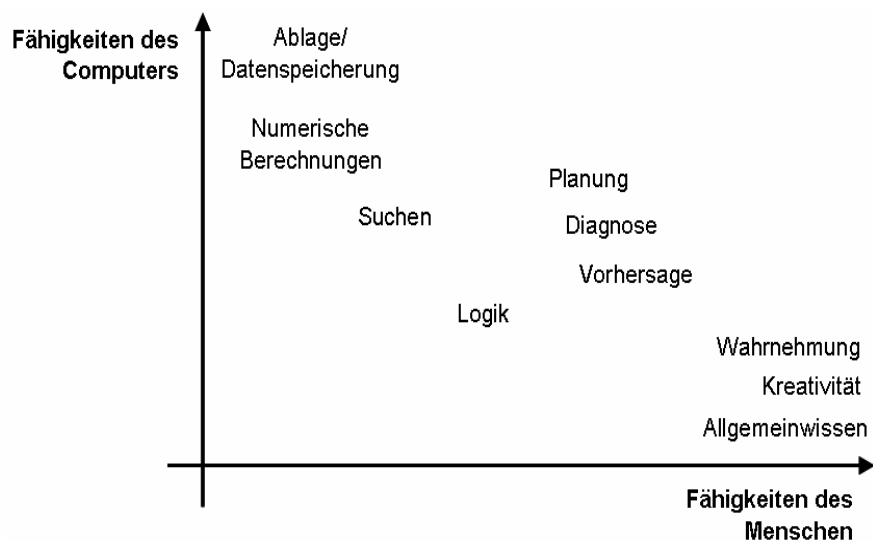


Abbildung 4: Aspekte der Datenanalyse im Kompetenzfeld von Computer und Mensch (nach Daniel Keim, Universität Konstanz)

Auch die technischen Grenzen der Analysemethoden spielen bei deren Auswahl eine Rolle. So benötigen die rechnerischen Verfahren des Data Mining häufig sehr viel Rechenzeit (engl. *runtime bottleneck*) und Speicherplatz (engl. *memory bottleneck*) und eignen sich kaum zum Treffen qualitativer Aussagen. Dahingegen haben die darstellenden Verfahren der Visualisierung Probleme, die großen Datenmengen mit ihren komplexen Zusammenhängen auf dem nur begrenzt verfügbaren Darstellungsplatz unterzubringen (engl. *screen bottleneck*), oder es in einer Weise zu tun, welche die Wahrnehmungsfähigkeiten des Menschen nicht überfordert (engl. *mind bottleneck*). Ferner eignet sie sich kaum zur Gewinnung quantitativer Aussagen.

Bis auf das Speicherplatzproblem werden alle genannten problematischen Aspekte beider Analysemethoden durch deren Kombination zum *Visuellen Data Mining* adressiert. So ist es beispielsweise möglich, große Datenmengen durch einen gezielten Data Mining Schritt auf darstell- und wahrnehmbare Maße herunterzubrechen und somit der Visualisierung überhaupt erst zugänglich zu machen. Umgekehrt ist es durch einen interaktiven visuellen Analyseschritt nun denkbar, zeitraubende Data Mining Schritte der Analysekompetenz des Nutzers zu überlassen. Insgesamt ermöglicht das Visuelle Data Mining eine enge Kopplung qualitativer und quantitativer Analyse, so dass Erkenntnisse beider Erkenntniskategorien in einem Analysedurchgang gewonnen werden können. Auf der anderen Seite erhöht sich durch diese Verzahnung die Komplexität des Analyseprozesses und führt dadurch eventuell zusätzliche Fehlerquellen ein.

4. Intuition im Analyseprozess

Das im vorangegangenen Abschnitt beschriebene Visuelle Data Mining eliminiert nicht nur eine Anzahl der aufgezeigten Nachteile der einzelnen Verfahren, sondern es ermöglicht dem Anwender in erster Linie seine Intuition über die visuelle Datenrepräsentation in den abstrakten, algorithmischen Analyseprozess des Data Mining einzubringen. Das bedeutet, dass der Anwender im Kontext seines Wissens eine Einsicht für die charakteristische Struktur der Daten entwickelt und dadurch die Auswahl und Parametrisierung der analytischen Methoden präziser und sicherer vornehmen kann. Zusätzlich dient dieselbe Visualisierung, über die der Analyst die zugrunde liegenden abstrakten Daten manipuliert und analysiert, auch dem verständlichen Feedback der rein rechnerischen Verfahren des Data Mining an den Analysten und bildet daher eine brauchbare Mensch-Maschine-Schnittstelle, die es in der einen Richtung ermöglicht, die komplexen Analyseergebnisse des Data Mining zu vermitteln, aber dieses umgekehrt darüber auch nachvollziehbar zu steuern und zu parametrisieren. Dieser Zusammenhang ist noch mal in der nachfolgenden schematischen Abb. 5 zusammengefasst.

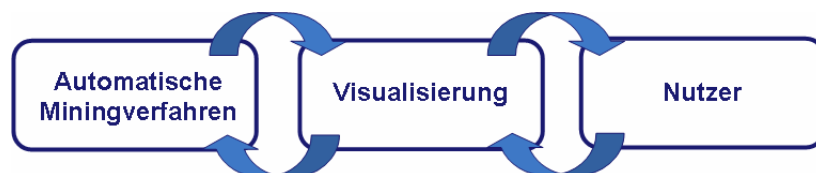


Abbildung 5: Die Visualisierung als Schnittstelle zwischen Mensch und Maschine

Die Notwendigkeit der Integration menschlicher Intuition in den Analyseprozess wird nachfolgend begründet und die intuitive Visualisierung an zwei Beispielen exemplarisch dargestellt.

4.1. Die Informationsexplosion

Nach einer Schätzung der Berkeley School of Information Management and Systems wurden allein im Jahr 2002 insgesamt 5 Exabytes an Daten gespeichert (Lyman, Varian 2003). Das entspricht (wiederum einer Schätzung zufolge) in etwa dem Informationsgehalt aller je von Menschen gesprochenen Wörter. Dieses immense Datenaufkommen steigt zudem jährlich um ca. 30% – eine Entwicklung, die allgemein als *Informationsexplosion* bezeichnet wird. Der amerikanische Zukunftsforscher John Naisbitt hat sich bereits 1982 in seinem Bestseller „MegaTrends“ mit dieser Tendenz beschäftigt und ist zu dem Schluss gekommen, dass die menschliche Intuition in Zukunft eine noch viel stärkere Rolle spielen wird als heute, da allein mit ihr die anfallenden Datenmengen noch innerhalb eines realistischen Zeitrahmens zu bewältigen sind¹.

4.2. Beispiele aus der Visualisierung

Damit die menschliche Intuition, also das zuvor beschriebene kontextuelle Verständnis der Beschaffenheit der Daten, in den Analyseprozess integriert werden kann, bedarf es intuitiver Mensch-Maschine-Schnittstellen wie der Visualisierung, die in der Lage sind, komplexe Sachverhalte, lange Datenkolonnen und abstrakte Beziehungen verständlich zu präsentieren. Die Abb. 6 zeigt zwei solcher intuitiver Darstellungstechniken, die z.Zt. aktuelle Forschungsgegenstände sind.

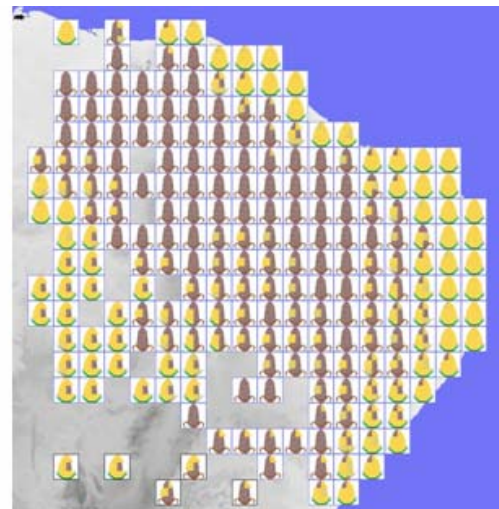
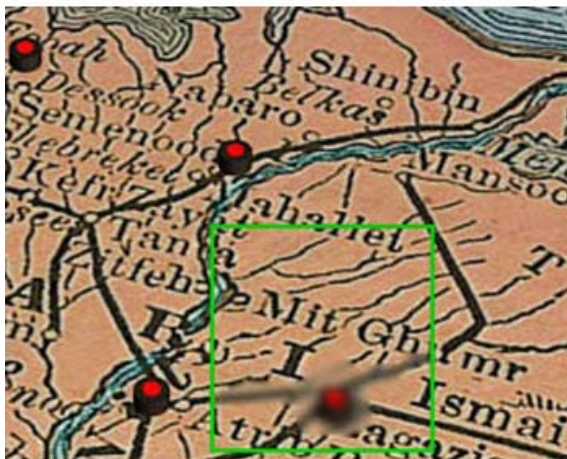


Abbildung 6: Unsicherheitslinse über einer historischen Kartendarstellung (links), Maiskolbenmetapher zur Darstellung der Maisanbaubedingungen in Brasilien (rechts)

Auf der linken Seite der Abb. 6 ist das Konzept der Unsicherheitslinse dargestellt, die interaktiv über einer grafischen Darstellung verschoben werden kann, um in dem darunterliegenden Bereich Informationen, die mit Unsicherheiten behaftet sind, z.B. verschwommen darzustellen. Gerade bei historischen Angaben, wie der abgebildeten Karte, kommen Unsicherheiten häufig vor und können auf diese Weise dem Anwender intuitiv zugänglich gemacht werden.

Auch die metaphorbasierte Visualisierung der Maisanbaubedingungen in Brasilien auf der rechten Seite der Abbildung stellt eine intuitive Visualisierung dar. Hier wird ein ähnlich komplexes Konzept wie das der Unsicherheit, nämlich das der Anbaubedingungen, in eine

¹ "Intuition becomes increasingly valuable in the new information society precisely because there is so much data." (Naisbitt 1982)

einfache, rasch zu erfassende visuelle Repräsentation überführt. Die grafische Metapher des Maiskolbens ist in insgesamt sechs Teile unterteilt, die ganz bestimmte Anbaufaktoren wie die Niederschlagsmenge codiert. Fällt viel Regen, wird dieser Teil des Maiskolbens in saftigem Gelb dargestellt, sonst graduell bis zu einem verdorrten Braun abgeschwächt. Auf diese Weise kann selbst ein Schulkind mit der Karte im Unterricht arbeiten und deren Inhalt auswerten.

5. Fazit und Schlussbemerkungen

Es gibt hauptsächlich zwei Probleme, aus denen das Misstrauen und die Unsicherheit im Umgang mit automatischen Analyseverfahren resultieren. Zum einen ist es die Menge der zu analysierenden Daten. Sie ist zu groß, um sie noch im Ganzen zu überblicken und bereits im Vorfeld der eigentlichen Datenanalyse ein grundlegendes, noch durch keine eventuell falsch parametrisierte Berechnung verfälschtes Verständnis für ihre Beschaffenheit zu entwickeln. Dieses Verständnis ist eigentlich umso wichtiger, als dass es häufig für die korrekte Auswahl geeigneter Analyseverfahren benötigt wird. Das zweite Problem ist das mangelnde Verständnis für die hochkomplexen automatischen Analyseverfahren. So ist es kaum noch möglich, deren Ergebnisse nachzuvollziehen oder auch nur mit einem pauschalen Überschlag, wie man ihn aus dem Schulunterricht kennt, zu überprüfen.

Gerade deshalb sind die aktuellen Anstrengungen, den Anwender mit seinem Wissen, seiner Kreativität und vor allem seiner Intuition vermehrt in den Analyseprozess einzubinden so wichtig. Dies geschieht zum einen durch die verstärkte Integration explorativer Analyseschritte (z.B. beim Visuellen Data Mining), wodurch der Nutzer einen viel direkteren Zugang zu den eigentlichen Daten bekommt und damit auch ein höheres Maß an Verständnis für diese erhält. Zum anderen ist man beim Entwurf von Analysetechniken bemüht, deren Funktionsweise durch ein möglichst hohes Maß an Feedback für den Nutzer transparent zu machen. Das bedeutet, dass nicht nur am Ende der Berechnung das Ergebnis ausgegeben wird, sondern laufend aktuelle Zwischenergebnisse abgerufen, ausgewertet und dargestellt werden können.

Es ist also ersichtlich, dass die heutige Gesellschaft aufgrund der riesigen Datenmengen zwar keine andere Wahl hat, als der maschinellen Datenanalyse zu vertrauen – doch auch wenn wir ihr also vertrauen müssen, zeigt die aktuelle Forschung auf dem Gebiet intuitiver Mensch-Maschine-Schnittstellen, dass man intensiv daran arbeitet, den Menschen wieder zunehmend in den Analyseprozess zu integrieren, um das Vertrauen darin zu stärken.

Literaturangaben:

Aamodt, Agnar & Nygård, Mads (1995): Different roles and mutual dependencies of data, information, and knowledge – An AI perspective on their integration, in “Data and Knowledge Engineering”, Bd.16, Nr.3, S. 191ff.

Eberhard, Kurt (1999): Einführung in die Erkenntnis- und Wissenschaftstheorie, 2.Auflage, Kohlhammer-Verlag (EA 1987)

Hand, David & Mannila, Heikki & Smyth, Padhraic (2001): Principles of Data Mining, 1.Auflage, MIT-Press

Lyman, Peter & Varian, Hal R. (2003): How Much Information 2003?, UC Berkeley School of Information Management and Systems

Mann, Michael E. & Bradley, Raymond S. & Hughes, Malcolm K. (1998): Global-Scale Temperature Patterns and Climate Forcing Over the Past Six Centuries, in "Nature", Nr.392, S. 779ff.

McKittrick, Ross & McIntyre, Stephen: Hockey Sticks, Principal Components and Spurious Significance, in "Geophysical Research Letters", Bd.32, Nr.3

Mena, Jesus (2004): Homeland Security. Techniques & Technologies, 1.Auflage, Charles River Media

Naisbitt, John (1982): Megatrends, 1.Auflage, Warner Books Inc.

Schumann, Heidrun & Müller, Wolfgang (2000): Visualisierung. Grundlagen und allgemeine Methoden, 1.Auflage, Springer

Shneiderman, Ben (2002): Inventing discovery tools: combining information visualization with data mining, in "Information Visualization", Bd.1, Nr.1, S. 5ff.

Bildnachweis:

Abb. 2: Kyoto protocol based on flawed statistics, in „Naturwissenschaft & Technik“, Ausgabe Feb. 2005