



(10) **DE 11 2018 004 687 T5** 2020.06.25

(12)

Veröffentlichung

der internationalen Anmeldung mit der
(87) Veröffentlichungs-Nr.: **WO 2019/084187**
in der deutschen Übersetzung (Art. III § 8 Abs. 2
IntPatÜG)
(21) Deutsches Aktenzeichen: **11 2018 004 687.7**
(86) PCT-Aktenzeichen: **PCT/US2018/057380**
(86) PCT-Anmeldetag: **24.10.2018**
(87) PCT-Veröffentlichungstag: **02.05.2019**
(43) Veröffentlichungstag der PCT Anmeldung
in deutscher Übersetzung: **25.06.2020**

(51) Int Cl.: **G06E 1/00 (2006.01)**

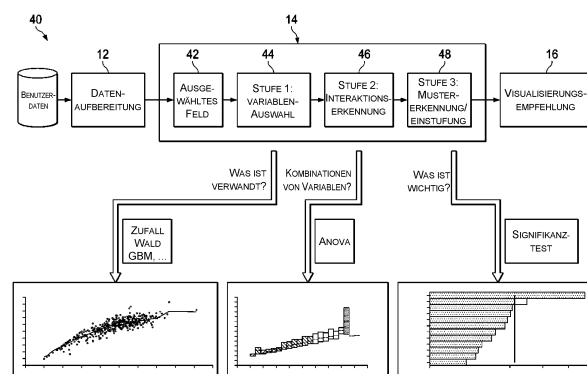
(30) Unionspriorität:
62/576,187 **24.10.2017** **US**
16/168,661 **23.10.2018** **US**
(71) Anmelder:
TIBCO SOFTWARE INC., Palo Alto, CA, US
(74) Vertreter:
**Grünecker Patent- und Rechtsanwälte PartG
mbB, 80802 München, DE**

(72) Erfinder:
**Rope, Daniel J., Palo Alto, CA, US; Berridge,
Andrew J., Palo Alto, CA, US; O'Connell, Michael,
Palo Alto, CA, US; Paolini, Gaia Valeria, Palo Alto,
CA, US; Rajdev, DivyaJyoti Pitamberlal, Palo Alto,
CA, US**

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen.

(54) Bezeichnung: **EINE VORHERSAGEVERARBEITUNGSEINHEIT FÜR MEHRSTUFIGE MUSTERERKENNUNG
UND EMPFEHLUNGEN FÜR VISUELLE ANALYTIK**

(57) Zusammenfassung: Eine Vorhersageverarbeitungseinheit zum Interpretieren von Datenstrukturen beinhaltet eine Interpretationseinheit und einen Visualisierungsgenerator. Die Interpretationseinheit erkennt ein relationales Muster zwischen Zieleigenschaftsvariablen und anderen Eigenschaftsvariablen auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen den Zieleigenschaftsdaten und den anderen Eigenschaftsdaten und erzeugt mindestens eine Metadaten-Eigenschaftsgruppe und zugehörige Ergebnismetriken. Der Visualisierungsgenerator kann mindestens eine Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und den zugehörigen Ergebnismetriken empfehlen. Die Interpretationseinheit beinhaltet mehrere Stufen, die eine Variablenauswahl, Interaktionserkennung und Mustererkennung und -einstufung ausführen. Die Vorhersageverarbeitungseinheit umfasst ferner eine Datenaufbereitungseinheit, die ausgebildet ist, die Datenstrukturen gemäß des Datentyps und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten zu sortieren, zu kategorisieren und zu filtern.



Beschreibung**QUERVERWEIS AUF VERWANDTE ANMELDUNG**

[0001] Diese Anmeldung beansprucht die Priorität der vorläufigen US-Patentanmeldung mit der Nr. 62/576,187, die am 24. Oktober 2017 eingereicht wurde mit dem Titel „Multistage Pattern Discovery for Visual Analytics Recommendations“, und deren gesamter Inhalt hiermit durch Bezugnahme für alle Zwecke mit eingeschlossen ist.

TECHNISCHES GEBIET DER OFFENBARUNG

[0002] Die vorliegende Offenbarung betrifft im Allgemeinen Algorithmen für künstliche Intelligenz und prädikative Funktionseinheiten bzw. Vorhersageverarbeitungseinheiten und insbesondere eine Vorhersageverarbeitungseinheit für mehrstufige Mustererkennung und Empfehlungen für visuelle Analytik.

STAND DER TECHNIK

[0003] Vorhersagende Analytik und Visualisierungsanalytik sind Werkzeuge, die in vielen Bereichen eingesetzt werden. Regierungen, Institutionen und Firmen verwenden diese Werkzeuge, um große Datenmengen zu verwalten und auszuwerten. Die Werkzeuge können von großem Vorteil sein, wenn große Datenmengen ausgewertet werden und Information über die Daten bereitgestellt wird, die verwendet werden kann, um Benutzer Führungsentscheidungen und Verwaltungsentscheidungen treffen zu lassen. Jedoch gibt es viele Nachteile im aktuellen Stand der Technik dieser Werkzeuge. Beispielsweise sind sie nicht skalierbar, sie sind bereichsspezifisch oder sie liefern wenige oder keine Einsichten. Damit gibt es einen Bedarf für eine Verbesserung im aktuellen Stand der Technik für Vorhersagewerkzeuge und Visualisierungsanalytikwerkzeuge.

ÜBERBLICK

[0004] Die vorliegende Offenbarung, die hierin dargelegt ist, umfasst eine Recheneinrichtung mit einem Mechanismus, der ausgebildet ist, Daten aus einer Datenstruktur zu erzeugen, Relationsmuster zwischen Zieleigenschaftsvariablen und anderen Eigenschaftsvariablen zu erkennen und Visualisierungen auf der Grundlage von Relationsmustern zu empfehlen.

[0005] In einem Aspekt richtet sich die vorliegende Offenbarung an eine prädikative Funktionseinheit bzw. Vorhersageverarbeitungseinheit zur Auswertung bzw. Interpretation von Datenstrukturen, die eine Interpretationseinheit und einen Visualisierungsgenerator beinhaltet. Die Interpretationseinheit ist ausgebildet, ein relationales Muster bzw. Relationsmuster zwischen Zieleigenschaftsvariablen

und anderen Eigenschaftsvariablen auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen Zieleigenschaftsdaten und den anderen Eigenschaftsdaten zu ermitteln, und mindestens eine Metadaten-Eigenschaftsgruppe und zugehörige Ergebnismetriken bzw. Ergebnismaßzahlen zu erzeugen. Der Visualisierungsgenerator ist ausgebildet, mindestens eine Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und der zugehörigen Ergebnismetriken zu empfehlen.

[0006] In einigen Ausführungsformen beinhaltet die Interpretationseinheit mehrere Stufen zur Ausführung einer Variablenauswahl, einer Interaktionserkennung und einer Mustererkennung und -einstufung. Die Variablenabhängigkeit ist eine lineare oder eine nicht-lineare Beziehung oder ein nicht-zufälliges Muster. In einigen Ausführungsformen beinhaltet die Vorhersageverarbeitungseinheit eine Datenvorbereitungseinheit, die ausgebildet ist, die Datenstrukturen zu sortieren, zu kategorisieren und zu filtern gemäß dem Datentyp und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten. In einigen Ausführungsformen ist die Interpretationseinheit ferner ausgebildet, einen statistischen Test auszuführen, um zu ermitteln, ob eine Interaktionswirkung signifikant ist. In einigen Ausführungsformen erzeugt der Visualisierungsgenerator unter anderem ein Multivarianten-Diagramm und/oder ein Bivarianten-Diagramm. In einigen Ausführungsformen ist der Visualisierungsgenerator ferner ausgebildet, auf Heuristik basierte Regeln anzuwenden, um die mindestens eine Visualisierung zu empfehlen.

[0007] In einem weiteren Aspekt richtet sich die vorliegende Offenbarung an ein Verfahren zum Betreiben einer prädiktiven Verarbeitungseinheit bzw. Vorhersageverarbeitungseinheit zur Auswertung bzw. Interpretation von Datenstrukturen. Das Verfahren beinhaltet das Ermitteln eines Relationsmusters zwischen Zieleigenschaftsdaten und anderen Eigenschaftsdaten auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen den Zieleigenschaftsdaten und den anderen Eigenschaftsdaten; Erzeugen mindestens einer Metadaten-Eigenschaftsgruppe und zugehöriger Ergebnismetriken; und Empfehlen mindestens einer Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und den zugehörigen Ergebnismetriken.

[0008] Das Verfahren kann ferner beinhalten: Ausführen einer ersten, einer zweiten und einer dritten Stufe, wobei Variablenauswahl, Interaktionserkennung und Mustererkennung und -einstufung im Schritt des Ermitteln und Erzeugens erfolgen. Die Variablenabhängigkeit ist eine lineare oder ei-

ne nicht-lineare Beziehung oder ein nicht-zufälliges Muster. Das Verfahren kann ferner das Sortieren, Kategorisieren und Filtern der Datenstrukturen gemäß dem Datentyp und/oder hierarchischen Datenstrukturen und/oder eindeutigen Werten und/oder fehlenden Werten und/oder Datums-/Zeitdaten beinhalten. Das Verfahren kann ferner beinhalten: Ausführen eines statistischen Tests zur Ermittlung, ob ein Interaktionswirkung signifikant ist. Das Verfahren umfasst ferner das Erzeugen eines multivariaten Diagramms und/oder eines bivariaten Diagramms.

[0009] In einem weiteren Aspekt richtet sich die vorliegende Offenbarung an ein nicht-flüchtiges computerlesbares Speichermedium, das eine Gruppe aus Computerbefehlen umfasst, die von einem Prozessor ausführbar sind, der in einer Vorhersageverarbeitungseinheit zur Interpretation bzw. Auswertung von Datenstrukturen arbeitet. Die Computerbefehle sind dazu ausgestaltet, ein Relationsmuster zwischen Zieleigenschaftsdaten und anderen Eigenschaftsdaten auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen den Zieleigenschaftsdaten und den anderen Eigenschaftsdaten zu ermitteln; mindestens eine Metadaten-Eigenschaftsgruppe und zugehörige Ergebnismetriken zu erzeugen; und mindestens eine Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und den zugehörigen Ergebnismetriken zu empfehlen.

[0010] Weitere Computerbefehle können dazu gestaltet sein, das Relationsmuster und die mindestens eine Metadaten-Eigenschaftsgruppe und die zugehörigen Ergebnismetriken in mehreren Stufen zu erzeugen, wobei Variablenauswahl, Interaktionserkennung und Mustererkennung und -einstufung ausgeführt werden; und/oder Datenstrukturen gemäß der Datenarten und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten zu sortieren, zu kategorisieren und zu filtern; und/oder mindestens eine oder mehrere aus einem multivariaten Diagramm und einem bivariaten Diagramm zu erzeugen; und/oder auf einer Heuristik beruhende Regeln anzuwenden, um die mindestens eine Visualisierung zu empfehlen. Die Variablenabhängigkeit ist eine lineare oder eine nicht-lineare Beziehung oder ein nicht-zufälliges Muster.

[0011] Weitere Ausführungsformen, Vorteile und neuartige Merkmale sind in der detaillierten Beschreibung angegeben.

Figurenliste

[0012] Für ein vollständigeres Verständnis der Eigenschaften und Vorteile der vorliegenden Offenbarung sei nun auf die detaillierte Beschreibung und die

begleitenden Zeichnungen verwiesen, in denen entsprechende Bezugszeichen in den Figuren zugeordnete Teile bezeichnen und in denen:

Fig. 1 eine Darstellung eines Flussdiagramms ist, das eine Übersicht über Dateninterpretations- und Visualisierungsfunktionen ist, die zu einem Algorithmus für eine Mehrstufen-, Maschinenlern-Vorhersageverarbeitungseinheit gemäß gewissen beispielhaften Ausführungsformen gehören.

Fig. 2 eine Darstellung eines Algorithmus für eine Mehrstufen-, Maschinenlern-Vorhersageverarbeitungseinheit gemäß gewissen anschaulichen Ausführungsformen ist;

Fig. 3-7, Fig. 8A-8B und Fig. 9A-9B Darstellungen von Visualisierungen sind, die von der Vorhersageverarbeitungseinheit erzeugt sind; und

Fig. 10 ein Blockdiagramm ist, das eine Recheneinrichtung und Systemanwendungen gemäß gewissen anschaulichen Ausführungsformen zeigt.

DETAILLIERTE BESCHREIBUNG

[0013] Obwohl die Herstellung und die Verwendung diverser Ausführungsformen der vorliegenden Offenbarung nachfolgend detailliert erläutert sind, sollte dennoch beachtet werden, dass die vorliegende Offenbarung viele anwendbare erfindungsgemäße Konzepte bietet, die in einer großen Bandbreite von speziellen Zusammenhängen umgesetzt werden können. Die speziellen Ausführungsformen, die hierin erläutert sind, sind lediglich anschaulich und sollen den Schutzbereich der vorliegenden Offenbarung nicht abgrenzen. Im Interesse der Klarheit sind gegebenenfalls nicht alle Merkmale einer tatsächlichen Implementierung in der vorliegenden Offenbarung beschrieben. Es selbstverständlich zu beachten, dass bei der Entwicklung einer derartigen tatsächlichen Ausführungsform zahlreiche implementierungsspezifische Entscheidungen getroffen werden müssen, um die speziellen Ziele des Entwicklers zu erreichen, etwa die Verträglichkeit mit systembezogenen und geschäftsbezogenen Rahmenbedingungen, die von einer Implementierung zur anderen unterschiedlich sein können. Seite ist jedoch zu beachten, dass ein derartiger Entwicklungsaufwand komplex und zeitaufwändig sein kann, aber dennoch eine Routinemaßnahme für den Fachmann darstellt, wenn er im Besitze dieser Offenbarung ist.

[0014] Datenvisualisierungsempfehlungssysteme können auf unterschiedliche Weise erzeugt werden. Beispielsweise setzen vorinstallierte Visualisierungen nicht-fachkundige Benutzer in die Lage, ein Bild ihrer Daten zu erhalten, sind aber nicht in der Lage, algorithmische Beziehungen zwischen Datenfeldern zu erkennen und anzuzeigen. Eine weitere Art ist die statistische Analyse. Statistische Analysen und Vi-

sualisierungen können spezielle mathematische Zusammenhänge darstellen und können diese auf Arten anzeigen, die für Datenwissenschaftler bedeutsam sind, die aber nicht dazu gestaltet sind, eine allgemeine Einsicht für geschäftliche Nutzer zu bieten. Anders ausgedrückt, diesen Werkzeugen mangelt es an den allgemeinen Fähigkeiten, für einen geschäftlichen Nutzer Visualisierungen zu bieten, die ausreichend flexibel sind, um einen beliebigen Geschäftsbereich abzudecken, um ausreichend informiert zu sein, sodass interessierende Eigenschaften und Zusammenhänge von Beginn an aufgezeigt werden. Ohne dies kann die wertvolle Einsicht auf geschäftliche Operationen verloren gehen. Eine weitere Art sind vordefinierte analytische Routinen. Deren Ergebnisse werden in speziellen Visualisierungen (oder Erzählungen) angezeigt. Diese sind nur wirksam, wenn sie bereichsspezifisch sind. Des Weiteren werden in konventionellen Visualisierungsempfehlungssystemen typischerweise nur die variablen Metadaten eingesetzt. Sie untersuchen keine Beziehungen zwischen den Daten.

[0015] In diversen Ausführungsformen werden Beziehungen zwischen den Daten durch einen Algorithmus für eine prädikative Verarbeitungseinheit bzw. eine Vorhersageverarbeitungseinheit, wie dies hierin offenbart ist, untersucht. In diversen Ausführungsformen werden mehrere Stufen des Maschinlernens angewendet, um zweckdienliche Variablengruppen bzw. Variablensätze und Metriken zu ermitteln, die ein heuristisches Visualisierungssystem beeinflussen können. In diversen Ausführungsformen werden Ergebnisse von Maschinlern-Algorithmen angewendet, um Hinweise für eine ansprechendere Darstellung der Visualisierungsverzerrung bereitzustellen und Muster innerhalb der Visualisierung zu kennzeichnen. Der mehrstufige Ansatz wird angewendet, um Muster zur Verwendung in Visualisierungsempfehlungen zu erkennen. Ein mehrstufiges Maschinlernen und heuristisch ausgewählte voraufgebaute Visualisierungen können in einer Vorgehensweise kombiniert werden, um analytische Einsichten für Geschäftsnutzer in Form standardmäßiger Geschäftsdiagramme bereitzustellen. Hierin offenbarte Maschinlern-Algorithmen erkennen Muster innerhalb ausgewählter Variablen, die die Auswahl der Rolle der Variablen, die durch ein heuristisches Visualisierungsempfehlungssystem gemacht werden, beeinflussen können. Maschinlern-Algorithmen schlagen ferner Visualisierungsverbesserungen vor, die dabei helfen können, spezielle Muster darzustellen oder Werte für einen Benutzer hervorzuheben.

[0016] Der hierin verwendete Begriff Zielvariable bedeutet ein spezielles Attribut, auch als Eigenschaft bzw. Merkmal bezeichnet, das in einer Datentabelle von Interesse ist, deren Änderung durch andere Variablen in den Daten beschreibbar ist. Daten, die die-

ser Zielvariablen zugeordnet sind, werden mit Daten in anderen Variablen innerhalb der Aufzeichnungen der Datentabelle verglichen.

[0017] Mit Verweis auf **Fig. 1** ist ein Flussdiagramm gezeigt, das Dateninterpretations- und Visualisierungsfunktionen zeigt, die mit einem mehrstufigen Maschinlern-Algorithmus einer Vorhersageverarbeitungseinheit gemäß gewissen anschaulichen Ausführungsformen einhergehen, wobei das Diagramm allgemein als 10 bezeichnet ist. Das Flussdiagramm **10** kennzeichnet Merkmale bzw. Eigenschaften, die mit einer Mehrstufen-Vorhersageverarbeitungseinheit mit einer heuristischen Visualisierungsempfehlungseinheit, die durch Maschinlernen erweitert ist, einhergehen. Das Flussdiagramm **10** beinhaltet die Abschnitte: Datenaufbereitung **12**; Erkennung bzw. Entdeckung **14**; und heuristische Visualisierungsempfehlung **16**.

[0018] Die Datenaufbereitung **12** beschreibt Datenaufbereitungseigenschaften, wonach Daten vorverarbeitet werden, um Einstellungen vorzunehmen, die die Qualität der Daten und daher die Vorhersagefähigkeiten des Algorithmus verbessern. Insbesondere werden die Daten durch Sortieren, Kategorisieren und Filtern der Datenstrukturen gemäß der Datenart und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten aufbereitet. Die Rohdaten können durch Datenarten (beispielsweise Datum oder Zeit) und Hierarchie (beispielsweise Jahr, Monat, Stunde, Minute etc....) gekennzeichnet werden und können weiter so gekennzeichnet werden, dass sie die Eigenschaften Eindeutigkeit und/oder Fehlen und/oder Zeit besitzen. Einstellungen können auf Variablen mit fehlenden Daten (beispielsweise: Entfernen oder Zuteilung) auf kategorische Variablen, die einen Schwellenwert unterscheidbarer Werte (beispielsweise Entfernen oder markiert für Neueinstufung) übersteigen, angewendet werden, und Variablen mit nur einem einzigen Wert können ignoriert werden. Die Denkweise besteht darin, Variablen auszuschließen, die nicht genug Information oder Kategorien enthalten, die mit höherer Wahrscheinlichkeit entsprechende Bezeichnungen anstatt Vorhersageeinheiten für das Ziel sind. Wenn ein Benutzer ein Ziel auswählt, das aus einem der zuvor genannten Gründe ausgeschlossen wurde, wird keine Einsicht erzeugt, das heißt, der Benutzer wird ein standardmäßiges Histogramm oder ein Balkendiagramm der Zielvariablen sehen.

[0019] Variablen eines Datums- und Zeitdatentyps können in das wahrscheinlichste oberste Element bzw. oberste Elemente ihrer eigenen Datumshierarchie (beispielsweise Jahr, Monat, ...) umgewandelt werden. In diversen Ausführungsformen können mehrere Ebenen der Hierarchie erzeugt werden. Die ursprüngliche Datumsvariable kann ver-

worfen werden. Das bzw. die oberste/n Hierarchieelement/e werden zu der Datumsvariablen. Numerische Variablen können unter Anwendung mehrerer Techniken unterteilt werden und die Ergebnisse können zusammengefasst werden, wodurch die Robustheit der Ergebnisse verbessert wird. Diese können als Variablentransformationen bezeichnet werden. Des Weiteren kann der Algorithmus automatisch Variablen transformieren, so dass sie normiert, unterteilt werden oder andere Berechnungen auf der Grundlage statistischer Metadaten angewendet werden, das heißt, es werden Minimum/Maximum, Momente, Prozentanteile, Häufigkeitszahlen und dergleichen ermittelt. Kategorische Variablen mit zu vielen Ebenen können eine unnatürlich große Wirkung auf die Eigenschaftsbedeutung haben. Somit können sie neu eingestuft werden, um die Anzahl an Ebenen zu reduzieren. Es können diverse Verfahren eingesetzt werden, um das erneute Einstufen zu bestimmen, wozu die Verwendung spezieller Schwellenwerte oder Untersuchungen der Häufigkeitsverteilung gehören. Eindeutige Werte einer kategorischen Variablen können gezählt werden. Dies hilft dabei zu ermitteln, wie eine Variable im Datenaufbereitungsschritt behandelt wird. Es kann eine neue Variable, die Zufallsdaten enthält, während der Datenaufbereitung in eine Datentabelle eingeführt werden, um eine Basislinie für Signal gegenüber Rauschen zu erhalten. Diese Technik bietet einen Mechanismus, um einen Signifikanzschwellenwert für Beziehungen für analytische Routinen zu ermitteln, die keinen expliziten Test bereitstellen.

[0020] Sobald die Daten aufbereitet sind, wird die Datenerkennung **14** für das ausgewählte Ziel durch den Algorithmus der Vorhersageverarbeitungseinheit ausgeführt. Dabei können Maschinenlernalgorithmen, etwa Zufallswald, gradientenverstärkte Bäume oder statistische Verfahren, etwa Pearson-Korrelation, Cramer V, ANOVA, verwendet werden. Beziehungen zwischen dem Ziel und anderen Variablen werden berechnet und eingestuft. Nichtsignifikante Beziehungen werden nicht verwendet. Die Einstufung von Variablen kann Erkenntnisse, die über die Beziehungen hinausgehen, nutzen, etwa die Anzahl und die Relevanz spezieller Anmerkungen. Die Einstufung von Variablen ist eine einzelne Maßnahme für Einstufung über 2-Variablen- und 3-Variablen-Beziehungen hinweg. Der Algorithmus für die Variablenbeziehung kann Beziehungen zwischen beliebigen Gruppen aus Spalten ermitteln. Die erzeugte Variableneinstufung wird als Eingabe für **16** bereitgestellt.

[0021] Die durch **14** erzeugte Informationen kann dann auf Visualisierungen mit bestem Praxisverhalten mittels heuristischer Regeln, die eine gute Visualisierung auswählen, angewendet werden. Mehrere potentielle Visualisierungen können erzeugt werden und schlechte Wahlmöglichkeiten können auf der

Grundlage der Regeln, die in **14** bereitgestellt werden, in Verbindung mit heuristischen Visualisierungsaspekten gefiltert werden. Dies führt zusammen zu einer globalen Bewertung oder Einstufung. Diese Regeln werden verwendet, um Visualisierungsarten, Achsen und Anmerkungen zu ermitteln. Die globale Bewertung, das heißt, Einstufung, kann auf die erzeugten Graphen angewendet werden, und es kann eine umfassende Liste an Visualisierungen angezeigt werden.

[0022] Eine Stärke des Algorithmus für Vorhersageverarbeitungseinheiten besteht darin, dass es unerheblich ist, ob die Beziehung linear, nicht-linear, als Cluster vorliegend und dergleichen, ist. Der Algorithmus ist in der Lage, jegliche interessierenden Beziehungen aufzufinden, in denen die Werte in den Vorhersagespalten die Werte in den Zielspalten auf eine gewisse nicht-zufällige Weise bedingen. Die Anwendung von Stufen in dem Vorhersagevorgang unterscheidet die Ergebnisse des Algorithmus für Vorhersageverarbeitungseinheiten, da es dadurch möglich ist, Beziehungen, Interaktionen und Muster in kombinierter Weise aufzufinden. Der Algorithmus für Vorhersageverarbeitungseinheiten ist in der Lage, lineare/ nicht-lineare Beziehungen aufzufinden und auch die Muster und Ausreißer darzustellen.

[0023] Es sei nun auf **Fig. 2** verwiesen. Hier ist ein Algorithmus einer mehrstufigen Maschinenlern-Vorhersageverarbeitungseinheit gemäß gewissen anschaulichen Ausführungsformen gezeigt, der allgemein als **40** bezeichnet ist. Der Algorithmus **40** kann in mehreren Stufen eingesetzt werden, um Einsichten zum Kuratieren und zur Visualisierung zu erzeugen, die zur Nutzung durch den Benutzer geeignet sind. Der Algorithmus **40** beinhaltet die Funktionen für die Datenaufbereitung **12**, die Erkennung **14** und heuristische Visualisierungsempfehlung **16**. Die Erkennung bzw. Auffindung **14** beinhaltet ein ausgewähltes Feldelement **42**, ein Stufe 1-Variablenauswahlelement **44**, ein Stufe 2-Interaktionsdetektionselement **46** und Stufe 3- Mustererkennungselement.

[0024] In diesen Stufen können Maschinenlernwerkzeuge, etwa Zufälliger Wald, GBM (Gradienten Boost Maschine), ANOVA (Analyse der Varianz) und statistische Signifikanzprüfung eingesetzt werden. Die Ausgabe dieser Stufen kann verwendet werden, um die Visualisierungsempfehlung **16** zu beeinflussen. Die Verwendung eines speziellen Algorithmus gegenüber einem weiteren kann parametrisiert werden, wodurch eine nutzerspezifische Anpassung möglich ist. Beispielsweise können einige Verfahren bei speziellen Datensätzen effizienter sein als bei anderen Verfahren. Es kann eine andere Technik, die für diese Stufe anzuwenden ist, ausgewählt werden, wenn diese Ergebnisse für das betrachtete Geschäftsproblem ungeeignet sind. Variablen können gemäß der

Stärke ihrer Beziehung mit der Zielvariable eingestuft werden.

[0025] Im Allgemeinen werden in dem Algorithmus **40** Benutzerdaten abgetastet und es wird die Datenaufbereitung **12** ausgeführt, die es den nachfolgenden Analysestufen ermöglicht, dass diese in einer effizienten und besser wirksamen Weise arbeiten. Zu den Präparationstechniken bzw. Aufbereitungstechniken können eine oder mehrere der folgenden gehören:

Variablentyperkennung - Bestimmung von kategorischen/kontinuierlichen Typen, während Problemfälle berücksichtigt werden, etwa kategorische Variablen, die als Ganzzahlen codiert sind;

Verarbeitung von fehlenden Daten - Zuteilung für kontinuierliche Variablen, etwa Addieren einer fehlenden Kategorie für kategorische Daten; und

Variablentransformationen - automatische Variablentransformationen zur Normierung, Unterteilung und Ausführen anderer Berechnungen auf der Grundlage statistischer Metadaten, das heißt, Bestimmung von Min/Max, Momenten, Prozentanteilen, Häufigkeitszählwerten und dergleichen.

[0026] Der Benutzer kann dann eine spezielle interessierende Zielvariable, das heißt, ein ausgewähltes Feld **42**, auswählen. In einigen Ausführungsformen ist dies gegebenenfalls die einzige Eingabe, die der Algorithmus **40** von einem Benutzer fordert. Die Auswahl der Variablen nach der Datenaufbereitung ermöglichen es dem Algorithmus **40** jegliche Variablen zu entfernen oder zu markieren, die nie zu einer zweckdienlichen Einsicht führen würden (zum Beispiel: Variablen mit einem konstanten Wert, oder Variablen mit zu vielen fehlenden Werten).

[0027] In diversen Ausführungsformen beinhaltet der Algorithmus **40** Maschinenlernfunktionen, die verwendet werden, um die Daten aufzubereiten, um zu ermitteln, welche Variablen am besten die Variabilität in der benutzerausgewählten Variablen erläutern, Stufe 1-Variablenauswahl **44**. Die Stufe **1** findet die Variablen auf, die unabhängig einer benutzergewählten Zielvariable zugeordnet sind. Diese sind für Bivarianten- (2-Variablen) Diagramme zwischen jeweils den unabhängig zugeordneten Variablen gegenüber der ausgewählten Variablen zweckdienlich. Wie in **Fig. 2** gezeigt ist, kann eine Variablenauswahlfunktion verwendet werden, um diese Zuordnung zu ermitteln. In der Stufe 2-Interaktionserkennung **46** werden Kombinationen dieser Variablen erkannt. Im Verbund können sie einen größeren Teil der Änderung der benutzergewählten Variablen erklären, als sie einzeln dazu in der Lage wären. Diese Variablengruppen können für Multivarianten-Visualisierungen verwendet werden. Beispielsweise können alle Paa-

re aus Variablen, die in Stufe **1** aufgefunden werden, untersucht werden. Ferner kann, wie in **Fig. 2** dargestellt ist, eine vorhersagende Modellierung oder eine statistische Technik, etwa ANOVA, in der Stufe **2** eingesetzt werden. In der Stufe 3-Mustererkennung/Einstufung **48** kann ein statistischer Signifikanztest ausgeführt werden, um zu ermitteln, ob eine Interaktionswirkung signifikant ist. Wenn die Interaktionswirkung als signifikant erkannt wird, wird die Gruppe aus drei Variablen zur Benutzung innerhalb einer Multivarianten-Visualisierung beibehalten. In der Stufe **3** findet der Algorithmus **40** signifikante wichtige Beziehungen zwischen Variablen auf. Die verwendeten Techniken können Variablen-Bedeutungstechniken, statistische Hypothesenprüfung und einfache Pearson-Korrelation mit einschließen. In ähnlicher Weise kann eine statistische Prozedur der besten praktischen Vorgehensweise eingesetzt werden, um die Signifikanz der Interaktionswirkung zwischen zwei (oder mehr) Variablen zu ermitteln.

[0028] Das Ergebnis des Mehrstufenprozesses ist eine Liste aus Variablengruppen und Ergebnismetriken, die durch das Visualisierungsempfehlungssystem **16** verwendet werden können, um geeignete Visualisierungen zu definieren bzw. festzulegen. Die Ergebnismetriken können eingesetzt werden, um eine heuristische Visualisierungsempfehlungseinheit zu beeinflussen bzw. zu steuern, um damit die Beziehung zwischen den Variablen besser darzustellen. Beispielsweise können Regeln einer heuristischen Visualisierungsempfehlung in einer beliebigen Entscheidung resultieren, um eine einzelne Variable der x-Achse zuzuordnen, während eine weitere Variable als eine Farbvariable mit einer Legende angewendet wird. Die Maschinenlern-Metriken können eine stärkere Beziehung zu der y-Achsenvariable für eine dieser Variablen anzeigen, wodurch es dem Empfehlungssystem möglich ist, eine Diagrammkonfiguration auszuwählen, die die geschäftliche Einsicht besser darstellt.

[0029] Ferner kann die heuristische Visualisierungsempfehlung **16** die Metriken benutzen, um Ausreißer, das heißt, ungewöhnliche Werte, zu erkennen und Erkenntnisse zu bezeichnen, die in der endgültigen Visualisierung hinzugefügt werden können. Beispielsweise bei einer gegebenen kategorischen und kontinuierlichen Variablengruppe kann eine Routine ermitteln, dass der Mittelwert der kontinuierlichen Variablen für eine gegebene Kategorie der kategorischen Variablen im Vergleich zu den anderen Kategorien ungewöhnlich groß ist. Die heuristische Visualisierungsempfehlung **16** kann diese Informationen nutzen, um diese Kategorie in einem Balkendiagramm hervorzuheben oder um den Punkt in einem Punktdiagramm hervorzuheben. Ein weiteres Beispiel besteht darin, dass eine Merkmalsextraktion (beispielsweise Jahr und Monat) aus Metadaten für Datums-/Zeitbereiche verwendet wird, um die beste

Zusammenführung für das Erstellen einer Wärmekarten-Visualisierung für eine einzige Datums-/Zeitvariable aufzufinden. Ein weiteres Beispiel besteht darin, Metadaten für die Anzahl separater Ebenen und die Existenz oder Nicht-Existenz von Ausreißern zu verwenden, um zwischen einem Boxen-Diagramm, einem Balkendiagramm oder einer Wärmekarten-Visualisierung als die geeignete Visualisierung für die kontinuierlichen und kategorischen Variablenpaare zu unterscheiden. Es gibt eine Reihe von Verfahren für das Erkennen von Ausreißern, abhängig von der Art der Variablen. Diese werden für eine verbesserte Erkennung kombiniert. Beispielsweise können für kontinuierliche Variablen und kontinuierliche Variablenpaare eine Kaskade aus gitterbasierten und regressionsbasierten Verfahren eingesetzt werden. Für kategorische Variablen und kategorische Variablenpaare können gegenseitige Häufigkeitsverteilungen und der Informationsinhalte eingesetzt werden, um seltene Ebenen hervorzuheben.

[0030] Es sei nun auf **Fig. 3-9** verweisen. Dort sind Visualisierungen dargestellt, die durch die Vorhersageverarbeitungseinheit **40** gemäß gewissen anschaulichen Ausführungsformen erzeugt sind. Die Darstellungen zeigen, wie eine Tabelle aus Aufzeichnungen verarbeitet und interpretiert werden kann, wobei eine Zielvariable verwendet wird, das heißt, ein Eigenschaftsattribut der Tabelle, um lineare oder nicht-lineare Beziehungen oder jegliche nichtzufälligen Muster zwischen als Ziel ausgewählte Attributvariablen oder anderen Attributvariablen innerhalb der Tabelle zu ermitteln.

[0031] Es sei nun auf **Fig. 10** verwiesen. Dort sind eine Rechenanlage **100** und ein Systemanwendungsmodul **200** gemäß anschaulichen Ausführungsformen gezeigt. Die Rechenanlage **100** kann einem der diversen Computer, Mobilgeräte, Laptop-Computern, Servern, eingebetteten Systemen oder Rechensystemen entsprechen, die hierin dargestellt sind. Das Modul **200** kann ein oder mehrere Hardware- oder Softwareelemente aufweisen, beispielsweise eine andere OS- bzw. BS-Anwendung und Benutzer- und Kernel-Raum-Anwendungen, die dazu gestaltet sind, die Rechenanlage **100** in der Lage zu versetzen, die diversen Verfahren und Verarbeitungsfunktionen, wie sie hierin dargestellt sind, auszuführen. Die Rechenanlage **100** kann diverse interne oder externe Komponenten, etwa einen Prozessor **100**, einen Systembus **120**, einen Systemspeicher **130**, Speichermedien **140**, eine Eingabe-/Ausgabeschnittstelle **150** und eine Netzwerkschnittstelle **160** zur Kommunikation mit einem Netzwerk **170**, beispielsweise Funk/GPS, Bluetooth oder WIFI, aufweisen.

[0032] Die Rechenanlagen können als ein konventionelles Computersystem, eine eingebettete Steuerung, ein Laptop-Computer, ein Server, ein Mobilgerät, ein intelligentes Telefon, ein am Körper tragbarer

Computer, eine anwendungsspezifische Einrichtung oder durch eine beliebige andere Hardwareplattform oder eine Kombination oder eine Vielfalt davon implementiert werden. Die Rechenanlagen können ein verteiltes System sein, das ausgebildet ist, unter Anwendung mehrerer Rechenanlagen, die über ein Datennetzwerk oder ein Bussystem miteinander verbunden sind, zu arbeiten.

[0033] Der Prozessor **110** kann so gestaltet sein, dass er Codebefehle ausführt, um die Operationen und die Funktionen, die hierin beschrieben sind, durchzuführen, den Strom aus Anforderungen und Adressenzuordnungen zu verwalten und Berechnungen auszuführen und Befehle zu erzeugen. Der Prozessor **110** kann ausgebildet sein, die Funktionsweise der Komponenten in den Rechenanlagen zu überwachen und zu steuern. Der Prozessor **110** kann ein Prozessor für Allgemeinzwecke, ein Prozessorkern, ein Mehrfachprozessor, ein umkonfigurierbarer Prozessor, eine Mikrosteuerung, ein digitaler Signalprozessor („DSP“), eine anwendungsspezifische integrierte Schaltung („ASIC“), eine Steuerung, eine Zustandsmaschine, eine getorte Logik, diskrete Hardwarekomponenten, eine beliebige andere Verarbeitungseinheit oder eine Kombination oder Vielzahl davon sein. Der Prozessor **110** kann eine einzige Verarbeitungseinheit, er kann mehrere Verarbeitungseinheiten, ein einziger Verarbeitungskern, mehrere Verarbeitungskerne, Verarbeitungskerne für spezielle Zwecke, er kann Co-Prozessoren oder eine beliebige Kombination davon sein. Gemäß gewissen Ausführungsformen kann der Prozessor **110** zusammen mit anderen Komponenten der Rechenanlage **100** eine softwarebasierte oder hardwarebasierte virtualisierte Rechenanlage sein, die in einer oder mehreren anderen Rechenanlagen ausgeführt wird.

[0034] Der Systemspeicher **130** kann nicht-flüchtige Speicher enthalten, etwa einen Nur-Lese-Speicher („ROM“), einen programmierbaren Nur-Lese-Speicher („PROM“), einen löschbaren programmierbaren Nur-Lese-Speicher („EPROM“), einen Flash-Speicher oder eine andere Einrichtung, die in der Lage ist, Programmbefehle oder Daten mit oder ohne anliegende Leistung zu speichern. Der Systemspeicher **130** kann ferner flüchtige Speicher beinhalten, etwa einen Direktzugriffsspeicher („RAM“), einen statischen Direktzugriffsspeicher („SRAM“), einen dynamischen Direktzugriffsspeicher („DRAM“) und einen synchronen dynamischen Direktzugriffsspeicher („SDRAM“). Andere Arten von RAM können ebenfalls eingesetzt werden, um den Systemspeicher **130** zu implementieren. Der Systemspeicher **130** kann unter Anwendung eines einzigen Speichermoduls oder unter Anwendung mehrerer Speichermodule implementiert werden. Obwohl der Systemspeicher **130** als Teil der Rechenanlage dargestellt ist, erkennt der Fachmann, dass der Systemspeicher **130** separat zu der Rechenanlage **100** vorgesehen sein kann, ohne

von dem Bereich der vorliegenden Technik abzuweichen. Es sollte auch beachtet werden, dass der Systemspeicher **130** eine nicht-flüchtige Speichervorrichtung, etwa das Speichermedium **140**, enthalten oder in Verbindung damit arbeiten kann.

[0035] Das Speichermedium **140** kann eine Festplatte, ein Diskettenlaufwerk, einen Kompaktdisketten-Nur-Lese-Speicher („CD-ROM“), eine digitale Vielseitigkeitsdiskette („DVD“), eine Blu-Ray-Diskette, ein Magnetband, einen Flash-Speicher, eine andere nicht-flüchtige Speichervorrichtung, einen Halbleiter-Speicher („SSD“), eine Magnetspeichereinrichtung, eine optische Speichereinrichtung, eine elektrische Speichereinrichtung, eine Halbleiterspeichereinrichtung, eine physikalisch basierte Speichereinrichtung oder eine andere Datenspeichereinrichtung oder irgendeine Kombination oder eine Vielzahl davon beinhalten. Das Speichermedium **140** kann ein oder mehrere Betriebssysteme, Anwendungsprogramme und Programmmodule, Daten oder andere Information speichern. Das Speichermedium **140** kann Teil der Rechenanlage oder mit dieser verbunden sein. Das Speichermedium **140** kann Teil einer oder mehrerer anderer Rechenanlagen sein, die mit der Rechenanlage in Verbindung stehen, etwa Server, Datenbankserver, Cloud-Speichereinrichtungen, durch Netzwerk angebundene Speicher und dergleichen.

[0036] Die Anwendungsmodule **200** und andere OS-Anwendungsmodule können ein oder mehrere Hardware- oder Softwareelemente aufweisen, die dazu ausgebildet sind, diese Rechenanlage zu unterstützen, um die diversen Verfahren und Verarbeitungsfunktionen, die hierin dargestellt sind, auszuführen. Das Anwendungsmodul **200** und andere OS-Anwendungsmodule können einen oder mehrere Algorithmen oder Sequenzen aus Befehlen enthalten, die als Software oder Firmware in Verbindung mit dem Systemspeicher **130**, dem Speichermedium **140** oder mit beiden gespeichert sind. Die Speichermedien **140** können daher Beispiele eines maschinen- oder computerlesbaren Mediums repräsentieren, in welchen Befehle oder Code zur Ausführung durch den Prozessor **110** speicherbar ist. Zu Maschinen- oder computerlesbaren Medien gehören vorzugsweise jedes Medium oder jede Medium, die verwendet werden, um Befehl für den Prozessor **110** bereitzustellen. Derartige maschinen- oder computerlesbare Medien, die zu dem Anwendungsmodul **200** und zu anderen OS-Anwendungsmodulen gehören, können ein Computersoftwareprodukt umfassen. Es sollte beachten werden, dass ein Computersoftwareprodukt, das das Anwendungsmodul **200** und andere OS-Anwendungsmodule umfasst, auch einem oder mehreren Prozessen oder Verfahren zur Zuführung des Anwendungsmoduls **200** und anderer OS-Anwendungsmodule zu der Rechenanlage über ein Netzwerk durch ein beliebiges signaltragendes Medium oder durch

andere Kommunikations- oder Bereitstellungstechnik zugeordnet sein können. Das Anwendungsmodul **200** und andere OS-Anwendungsmodule können ferner Hardwareschaltungen oder Information zur Konfigurierung von Hardwareschaltungen umfassen, etwa Mikrocode oder Konfigurationsinformation für ein FPGA oder eine andere PLD. In einer anschaulichen Ausführungsform können das Anwendungsmodul **200** und andere OS-Anwendungsmodule Algorithmen enthalten, die in der Lage sind, die durch die hierin präsentierten Flussdiagramme und Computersysteme beschriebenen Funktionen auszuführen.

[0037] Die Eingabe-/Ausgabe- („I/O“) Schnittstelle **150** kann dazu ausgebildet sein, ein oder mehrere externe Geräte anzubinden, Daten aus dem einen oder den mehreren externen Geräten zu empfangen, und Daten an das eine oder die mehreren externen Geräte zu senden. Derartige externe Geräte zusammen mit den diversen internen Geräten können auch als Periphergeräte bezeichnet werden. Die I/O-Schnittstelle **150** kann sowohl elektrische als auch mechanische Verbindungen zur Ankopplung der diversen peripheren Geräte an die Rechenanlage oder den Prozessor **110** beinhalten. Die I/O-Schnittstelle **150** kann dazu ausgebildet sein, Daten, Adressen und Steuersignale zwischen den peripheren Geräten, der Rechenanlage oder dem Prozessor **110** auszutauschen. Die I/O-Schnittstelle **150** kann ausgebildet sein, eine Standardschnittstelle zu implementieren, etwa kleine Computersystemschnittstelle („SCSI“), eine seriell angeschlossene SCSI („SAS“), einen Faserkanal eine periphere Komponentenverbindung („PCI“), einen PCI Express (PCIe), einen seriellen Bus, einen parallelen Bus, eine angebundene erweiterte Technik („ATA“), eine serielle ATA („SATA“), einen universellen seriellen Bus („USB“), Thunderbolt, FireWire, diverse Videobusse und dergleichen. Die I/O-Schnittstelle **150** kann dazu ausgelegt sein, nur eine einzelne Schnittstelle oder Bus-Technik zu implementieren. Alternativ kann die I/O-Schnittstelle **150** dazu ausgelegt sein, mehrere Schnittstellen oder Bustechniken zu implementieren. Die I/O-Schnittstelle **150** kann so gestaltet sein, dass Teil des Systembusses **120** ist oder den Systembus repräsentiert oder in Verbindung damit arbeitet. Die I/O-Schnittstelle **150** kann einen oder mehrere Puffer zur Zwischenspeicherung von Übertragungen zwischen einem oder mehreren externen Geräten, internen Geräten, der Rechenanlage oder dem Prozessor **120** aufweisen.

[0038] Die I/O-Schnittstelle **120** kann die Rechenanlage mit diversen Eingabegeräten einschließlich von Mäusen, berührungsempfindlichen Bildschirmen, Tastgeräten, elektronischen Digitalisiergeräten, Sensoren, Empfängern, Berührflächen, Eingabekugeln, Kameras, Mikrofonen, Tastaturen, jeglichen Zeigergeräten oder Kombinationen davon, verbinden. Die I/O-Schnittstelle **120** kann die Rechenanlage mit diversen Ausgabegeräten verbinden, wozu

Videoanzeigen, Lautsprecher, Drucker, Projektoren, Geräte mit taktiler Rückmeldung, Automatisierungssteuerungen, robotische Komponenten, Aktoren, Motoren, Gebläse, Elektromagnete, Ventile, Pumpen, Sender, Signal-Emitter, Leuchten und dergleichen gehören.

[0039] Die Rechenanlage **100** kann in einer vernetzten Umgebung unter Anwendung logischer Verbindungen durch die NIC **160** zu einem oder mehreren anderen Systemen oder Rechenanlagen im Netzwerk arbeiten. Das Netzwerk kann Weitbereichsnetzwerke (WAN), lokale Netzwerke (LAN), Intra-Netzwerke, das Internet, Netzwerke mit drahtlosem Zugang, verdrahtete Netzwerke, Mobilnetzwerke, Telefonnetzwerke, optische Netzwerke oder Kombinationen davon beinhalten. Das Netzwerk kann paketbasiert sein, kann durch Schaltung vermittelt sein, kann eine beliebige Topologie aufweisen und kann jegliches Kommunikationsprotokoll verwenden. Kommunikationsverbindungen innerhalb des Netzwerks können diverse digitale oder analoge Kommunikationsmedien, etwa faseroptische Kabel, Optik für den freien Raum, Wellenleiter, elektrische Leiter, drahtlose Verbindungen, Antennen, Hochfrequenzkommunikationseinheiten und dergleichen, beinhalten.

[0040] Der Prozessor **110** mit den anderen Elementen der Rechenanlage oder den diversen Peripheriegeräten, die hierin erläutert sind, über den Systembus **120** verbunden sein. Es sollte beachtet werden, dass der Systembus **120** innerhalb des Prozessors **110**, außerhalb des Prozessors **110** oder sowohl als auch liegen kann. Gemäß einigen Ausführungsformen können die Prozessoren **110**, die anderen Elemente der Rechenanlage oder die diversen Peripheriegeräte, die hierin erläutert sind, in einer einzigen Einrichtung, etwa einem System auf einem Chip („SOC“), einem System auf einem Gehäuseträger („SOP“) oder einem ASIC-Bauteil integriert sein.

[0041] Ausführungsformen können ein Computerprogramm umfassen, das die hierin beschriebenen und dargestellten Funktionen umsetzt, wobei das Computerprogramm in einem Computersystem implementiert ist, das Befehle umfasst, die in einem maschinenlesbaren Medium gespeichert sind, und einen Prozessor umfasst, der die Befehle ausführt. Es sollte jedoch beachtet werden, dass es viele unterschiedliche Möglichkeiten zum Implementieren von Ausführungsformen in der Computerprogrammierung gibt, und die Ausführungsformen sollten nicht als eine Einschränkung auf eine Gruppe von Computerprogrammbefehlen erachtet werden, sofern dies nicht anderweitig für eine anschauliche Ausführungsform offenbart ist. Ferner ist ein geschulter Programmierer in der Lage, ein derartiges Computerprogramm zu erstellen, um eine Ausführungsform der offenbarten Ausführungsformen zu implementieren, wobei dies auf der Grundlage der angefügten Flussdiagramme,

Algorithmen und zugehörigen Beschreibung in dem Anmeldungstext erfolgt. Daher ist die Offenbarung einer speziellen Gruppe aus Programmcodebefehlen nicht als notwendig zu verstehen, um ein adäquates Verständnis darüber zu erhalten, wie Ausführungsformen zu erstellen und anzuwenden sind. Ferner erkennt der Fachmann, dass ein oder mehrere Aspekte von hierin beschriebenen Ausführungsformen durch Hardware, Software oder eine Kombination davon umgesetzt werden können, wie sie in einem oder mehreren Rechensystemen vorgesehen sind. Ferner sollte der Verweis auf einen Vorgang, der von einem Computer ausgeführt wird, nicht so verstanden werden, dass dieser durch einen einzigen Computer auszuführen ist, da mehr als ein einzelner Computer den Vorgang ausführen können.

[0042] Die hierin beschriebenen anschaulichen Ausführungsformen können in Verbindung mit Computerhardware und -software eingesetzt werden, die die Verfahren und die Verarbeitungsfunktionen, die zuvor beschrieben sind, ausführen. Die Systeme, Verfahren und Prozeduren, die hierin beschrieben sind, können in einem programmierbaren Computer, in einer computerausführbaren Software oder in einer Digitalschaltung umgesetzt werden. Die Software kann auf computerlesbaren Medien gespeichert sein. Beispielsweise können computerlesbare Medien eine Diskette, einen RAM, einen ROM, eine Festplatte, abnehmbare Medien, einen Flash-Speicher, einen Speicherstift, optische Medien, magnet-optische Medien, eine CD-ROM und dergleichen, umfassen. Eine Digitalschaltung kann integrierte Schaltungen, Gatter-Arrays, Logikanordnungen mit Funktionsblöcken, feldprogrammierbare Gatter-Arrays (FPGA) und dergleichen, mit einschließen.

[0043] Die beispielhaften Systeme, Verfahren und Prozesse, die in den zuvor dargestellten Ausführungsformen beschrieben sind, sind anschaulicher Natur und in alternativen Ausführungsformen können gewisse Vorgänge in einer anderen Reihenfolge und/oder parallel zueinander ausgeführt, vollständig weggelassen und/oder in unterschiedlichen anschaulichen Ausführungsformen kombiniert werden, und/oder es können gewisse zusätzliche Vorgänge ausgeführt werden, ohne von dem Schutzbereich und dem Grundgedanken der diversen Ausführungsformen abzuweichen. Daher sind derartige alternative Ausführungsformen in der Beschreibung hierin mit eingeschlossen.

[0044] Im hierin verwendeten Sinne, sollen die Singularformen „einer, eine, ein“ und „der, die, das“ die Pluralformen ebenfalls mit einschließen, sofern der Zusammenhang dies nicht anders angibt. Es sollte ferner beachtet werden, dass die Begriffe „umfassen bzw. beinhalten“ und/und oder „umfassend bzw. mit“ bei Verwendung in dieser Anmeldung das Vorhandensein angegebener Merkmale, Zahlen, Schritte,

Operationen, Elemente und/oder Komponenten bezeichnen, aber das Vorhandensein oder das Hinzufügen eines oder mehrerer anderer Merkmale, Zahlen, Schritte, Operationen, Elemente, Komponenten und/oder Gruppen davon nicht ausschließen. Im hierin verwendeten Sinne beinhaltet der Begriff „und/oder“ jede mögliche Kombination aus einem oder mehreren der zugehörigen aufgeführten Elemente. Im hierin verwendeten Sinne, sollten Redewendungen, etwa „zwischen X und Y“ und „zwischen ungefähr X und Y“ so verstanden werden, dass X und Y mit eingeschlossen sind. Im hierin verwendeten Sinne sollen sprachliche Wendungen, etwa „zwischen ungefähr X und Y“, „zwischen ungefähr X und ungefähr Y“ bedeuten. Im hierin verwendeten Sinne sollen sprachliche Wendungen, etwa „von ungefähr X bis Y“ „von ungefähr X bis ungefähr Y“ bedeuten.

[0045] Im hierin verwendeten Sinne kann „Hardware“ eine Kombination aus diskreten Komponenten, eine integrierte Schaltung, eine anwendungsspezifische integrierte Schaltung, ein feldprogrammierbares Gatter-Array oder eine andere geeignete Gerätekomponente bezeichnen. Im hierin verwendeten Sinne kann „Software“ ein oder mehrere Objekte, Mittel, Programmfäden, Codezeilen, Unterroutinen, separate Softwareanwendungen, zwei oder mehr Codezeilen oder andere geeignete Softwarestrukturen sein, die in zwei oder mehr Softwareanwendungen in einem oder mehreren Prozessoren (wenn ein Prozessor ein oder mehrere Mikrocomputer oder andere geeignete Datenverarbeitungseinheiten, Speichereinrichtungen, Eingabe-/Ausgabeeinrichtungen, Anzeigen, Dateneingabeeinrichtungen, etwa eine Tastatur oder eine Maus, Peripheriegeräte, etwa Drucker und Lautsprecher, zugehörige Treiber, Steuerkarten, Energiequellen, Netzwerkgeräte, Verbindungsstationen oder andere geeignete Geräte, die unter der Steuerung von Softwaresystem in Verbindung mit dem Prozessor oder anderen Geräten arbeiten, enthält) aktiv sind, oder andere geeignete Softwarestrukturen beinhalten. In einer anschaulichen Ausführungsform kann Software eine oder mehrere Codezeilen oder andere geeignete Softwarestrukturen, die in einer Softwareanwendung für Allgemeinzwecke aktiv ist, etwa in einem Betriebssystem, und eine oder mehrere Codezeilen oder andere geeignete Softwarestrukturen beinhalten, die in einer Softwareanwendung für spezielle Zwecke aktiv sind. Im hierin verwendeten Sinne bezeichnet der Begriff „Koppeln bzw. Verbinden“ und entsprechende zugehörige Begriffe, wie „Koppeln“ und „gekoppelt“, eine physikalische Verbindung (etwa einen Kupferleiter), eine virtuelle Verbindung (etwa durch frei wählbare, zugewiesene Speicherstellen eines Datenspeichers), eine logische Verbindung (etwa durch Logikgatter eines Halbleiterbauelements) oder andere geeignete Verbindungen, oder eine geeignete Kombination derartiger Verbindungen. Der Begriff „Daten“ kann eine geeignete Struktur zur Verwendung, Übertragung oder

Speicherung von Daten, etwa ein Datenfeld, einen Datenpuffer, eine Datennachricht mit Datenwert und Sender-/Empfänger-Adressdaten, eine Steuerungsnachricht mit dem Datenwert und einen oder mehrere Operatoren, die als Empfangssystem oder die Komponente veranlassen, eine Funktion unter Anwendung der Daten auszuführen, oder andere geeignete Hardware- oder Softwarekomponenten für die elektronische Verarbeitung von Daten bezeichnen.

[0046] Im Allgemeinen ist ein Softwaresystem ein System, das in einem Prozessor aktiv ist, um vorbestimmte Funktionen in Reaktion auf vorbestimmte Datenfelder auszuführen. Beispielsweise kann ein System durch die Funktion, die es ausführt und durch die Datenfelder definiert sein, an denen die Funktion ausgeführt wird. Im hierin verwendeten Sinne ist ein NAME-System, wobei NAME typischerweise der Name der allgemeinen Funktion ist, die von dem System ausgeführt wird, eine Referenz auf ein Softwaresystem, das ausgebildet ist, in einem Prozessor aktiv zu sein, und die offenbarte Funktion an den offenbarten Datenfeldern auszuführen. Sofern nicht ein spezieller Algorithmus offenbart ist, kann dann jeder beliebige geeignete Algorithmus, der dem Fachmann für das Ausführen der Funktion unter Anwährend der zugehörigen Datenfelder bekannt ist, als ein Algorithmus betrachtet werden, der innerhalb des Schutzbereichs der Offenbarung liegt. Beispielsweise kann ein Nachrichtensystem, das eine Nachricht erzeugt, die ein Senderadressenfeld, ein Empfängeradressenfeld und ein Nachrichtenfeld enthält, somit eine Software bezeichnen, die in einem Prozessor ausgeführt wird, der das Sender-Adressenfeld, das Empfänger-Adressenfeld und das Nachrichtenfeld aus einem geeigneten System oder einer Einrichtung des Prozessors erhalten kann, etwa aus einer Puffereinrichtung oder einem Puffersystem, und somit können das Sender-Adressenfeld, das Empfänger-Adressenfeld und das Nachrichtenfeld in ein geeignetes elektronisches Nachrichtenformat umgewandelt werden (etwa eine elektronische Mail-Nachricht, eine TCP/IP-Nachricht oder ein anderes geeignetes Nachrichtenformat, das ein Absender-Adressenfeld, ein Empfänger-Adressenfeld und ein Nachrichtenfeld hat) und somit kann die elektronische Nachricht unter Anwendung elektronischer Nachrichtensysteme und Einrichtungen des Prozessors über ein Kommunikationsmedium, etwa ein Netzwerk gesendet werden. Der Fachmann ist in der Lage, die spezielle Codierung für eine spezielle Anwendung auf der Grundlage auf der vorhergehenden Offenbarung zu erstellen, die dazu beabsichtigt ist, anschauliche Ausführungsformen der vorliegenden Offenbarung anzugeben, und nicht dafür vorgesehen ist, eine Lerneinheit für jemanden bereitzustellen, der kein Fachmann ist, etwa für einen, der nicht mit der Programmierung von Prozessoren in einer geeigneten Programmiersprache vertraut ist. Ein speziellen Algorithmus zum Ausführen einer Funktion kann in Form

eines Flussdiagramms oder in anderen geeigneten Formaten bereitgestellt werden, wobei die Datenfelder und zugehörigen Funktionen in einer beispielhaften Reihenfolge von Operationen angegeben sind, wobei die Reihenfolge nach Bedarf geändert werden kann und nicht als Beschränkung dient, sofern nicht explizit eine Beschränkung diesbezüglich erwünscht ist.

[0047] Die vorhergehende Beschreibung von Ausführungsformen der Offenbarung wurde zum Zwecke der Darstellung und der Beschreibung angegeben. Sie soll nicht umfassend sein oder soll die Offenbarung auf die genaue offenbarte Form beschränken, und es sind Modifizierungen und Varianten im Lichte der vorhergehenden Lehren möglich oder können bei der Umsetzung der Offenbarung erkannt werden. Die Ausführungsformen wurden ausgewählt und beschrieben, um die Prinzipien der Offenbarung und ihre praktische Anwendung dem Fachmann zu erläutern, um die die Offenbarung in diversen Ausführungsformen und mit diversen Modifizierungen Modifikationen einzusetzen, wie sie für die spezielle betrachtete Nutzung geeignet sind. Andere Ergänzungen, Modifizierungen, Änderungen und Weglassungen können an der Gestaltung, den Betriebsbedingungen und der Anordnung der Ausführungsformen vorgenommen werden, ohne von dem Schutzbereich der vorliegenden Offenbarung abzuweichen. Derartige Modifizierungen und Kombinationen der anschaulichen Ausführungsformen sowie anderer Ausführungsformen ergeben sich für den Fachmann beim Studium der Beschreibung. Es ist daher beabsichtigt, dass die angefügten Ansprüche alle derartigen Modifizierungen oder Ausführungsformen mit einschließen.

ZITATE ENTHALTEN IN DER BESCHREIBUNG

Diese Liste der vom Anmelder aufgeführten Dokumente wurde automatisiert erzeugt und ist ausschließlich zur besseren Information des Lesers aufgenommen. Die Liste ist nicht Bestandteil der deutschen Patent- bzw. Gebrauchsmusteranmeldung. Das DPMA übernimmt keinerlei Haftung für etwaige Fehler oder Auslassungen.

Zitierte Patentliteratur

- US 62/576187 [0001]

Patentansprüche

. Beansprucht ist:

1. Eine Vorhersageverarbeitungseinheit zum Interpretieren von Datenstrukturen, wobei die Vorhersageverarbeitungseinheit umfasst:

eine Interpretationseinheit, die ausgebildet ist, ein relationales Muster zwischen Zieleigenschaftsvariablen und anderen Eigenschaftsvariablen auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen den Zieleigenschaftsdaten und den anderen Eigenschaftsdaten zu ermitteln und mindestens eine Metadaten-Eigenschaftsgruppe und zugehörige Ergebnismetriken zu erzeugen;
einen Visualisierungsgenerator, der ausgebildet ist, mindestens eine Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und der zugehörigen Ergebnismetriken zu empfehlen.

2. Die Vorhersageverarbeitungseinheit nach Anspruch 1, wobei die Interpretationseinheit mehrere Stufen zur Ausführung einer Variablenauswahl, einer Interaktionserkennung und einer Mustererkennung und -einstufung enthält.

3. Die Vorhersageverarbeitungseinheit nach Anspruch 1, wobei die Variablenabhängigkeit eine lineare Beziehung oder eine nicht-lineare Beziehung oder ein nicht-zufälliges Muster ist.

4. Die Vorhersageverarbeitungseinheit nach Anspruch 1, die ferner eine Datenaufbereitungseinheit aufweist, die ausgebildet ist, die Datenstrukturen gemäß eines Datentyps und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten zu sortieren, zu kategorisieren und zu filtern.

5. Die Vorhersageverarbeitungseinheit nach Anspruch 1, wobei die Interpretationseinheit ferner ausgebildet ist, einen statistischen Test auszuführen, um zu ermitteln, ob eine Interaktionswirkung signifikant ist.

6. Die Vorhersageverarbeitungseinheit nach Anspruch 1, wobei der Visualisierungsgenerator mindestens ein oder mehrere eines Multivarianten-Diagramms und/oder eines Bivarianten-Diagramms erzeugt.

7. Die Vorhersageverarbeitungseinheit nach Anspruch 1, wobei der Visualisierungsgenerator ferner ausgebildet ist, auf Heuristik beruhende Regeln zum Empfehlen der mindestens einen Visualisierung anzuwenden.

8. Ein Verfahren zum Betreiben einer Vorhersageverarbeitungseinheit zur Interpretation von Datenstrukturen, wobei das Verfahren umfasst:

Erkennen eines relationalen Musters zwischen Zieleigenschaftsdaten und anderen Eigenschaftsdaten auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen den Zieleigenschaftsdaten und den anderen Eigenschaftsdaten;

Erzeugen mindestens einer Metadaten-Eigenschaftsgruppe und zugehöriger Ergebnismetriken; und

Empfehlung mindestens einer Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und der zugehörigen Ergebnismetriken.

9. Das Verfahren nach Anspruch 8, wobei der Schritt des Erkennens und Erzeugens in einer ersten, zweiten und dritten Stufe oder in mehr Stufen ausgeführt wird, in denen die Variablenauswahl, die Interaktionserkennung und die Mustererkennung und Einstufung ausgeführt werden.

10. Das Verfahren nach Anspruch 8, wobei die Variablenabhängigkeit eine lineare oder eine nicht-lineare Beziehung oder ein nicht-zufälliges Muster ist.

11. Das Verfahren nach Anspruch 8, das ferner umfasst: Sortieren, Kategorisieren und Filtern der Datenstrukturen gemäß des Datentyps und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten.

12. Das Verfahren nach Anspruch 8, das ferner Ausführen eines statistischen Tests umfasst, um zu ermitteln, ob eine Interaktionswirkung signifikant ist.

13. Das Verfahren nach Anspruch 1, das ferner Erzeugen mindestens eines Multivarianten-Diagramms und eines Bivarianten-Diagramms umfasst.

14. Ein nicht-flüchtiges computerlesbares Speichermedium, mit einer Gruppe aus Computerbefehlen, die von einem Prozessor zum Betreiben einer Vorhersageverarbeitungseinheit ausführbar sind, um Datenstrukturen zu interpretieren, wobei die Computerbefehle ausgebildet sind, zum:

Erkennen eines relationalen Musters zwischen Zieleigenschaftsdaten und anderen Eigenschaftsdaten auf der Grundlage des Erkennens einer Variablenabhängigkeit zwischen den Zieleigenschaftsdaten und den anderen Eigenschaftsdaten;

Erzeugen mindestens einer Metadaten-Eigenschaftsgruppe und zugehöriger Ergebnismetriken; und

Empfehlen mindestens einer Visualisierung auf der Grundlage der mindestens einen Metadaten-Eigenschaftsgruppe und den zugehörigen Ergebnismetriken.

15. Das nicht-flüchtige computerlesbare Speichermedium nach Anspruch 14, das ferner Computerbefehle enthält, die ausgebildet sind, um das relationale Muster und mindestens eine Metadaten-Eigenschaftsgruppe und zugehörige Ergebnismetriken in einer ersten, einer zweiten und einer dritten Stufe oder in mehr Stufen zu erkennen und zu erzeugen, in denen Variablenauswahl, Interaktionserkennung und Mustererkennung und -einstufung ausgeführt werden.

16. Das nicht-flüchtige computerlesbare Speichermedium nach Anspruch 14, wobei die Variablenabhängigkeit eine lineare oder eine nicht-lineare Abhängigkeit ist.

17. Das nicht-flüchtige computerlesbare Speichermedium nach Anspruch 14, das ferner Computerbefehle enthält, die ausgebildet sind, Datenstrukturen gemäß eines Datentyps und/oder gemäß hierarchischen Datenstrukturen und/oder gemäß eindeutigen Werten und/oder gemäß fehlenden Werten und/oder gemäß Datums-/Zeitdaten zu sortieren, zu kategorisieren und zu filtern.

18. Das nicht-flüchtige computerlesbare Speichermedium nach Anspruch 14, das ferner Computerbefehle enthält, die ausgebildet sind, einen statistischen Test auszuführen, um zu ermitteln, ob eine Interaktionswirkung signifikant ist.

19. Das nicht-flüchtige computerlesbare Speichermedium nach Anspruch 14, das ferner Computerbefehle enthält, die ausgebildet sind, ein Multivarianten-Diagramm und ein Bivarianten-Diagramm zu erzeugen.

20. Das nicht-flüchtige computerlesbare Speichermedium nach Anspruch 14, das ferner Computerbefehle enthält, die ausgebildet sind, auf Heuristik beruhende Regeln zum Empfehlen der mindestens einen Visualisierung anzuwenden.

Es folgen 12 Seiten Zeichnungen

Anhängende Zeichnungen

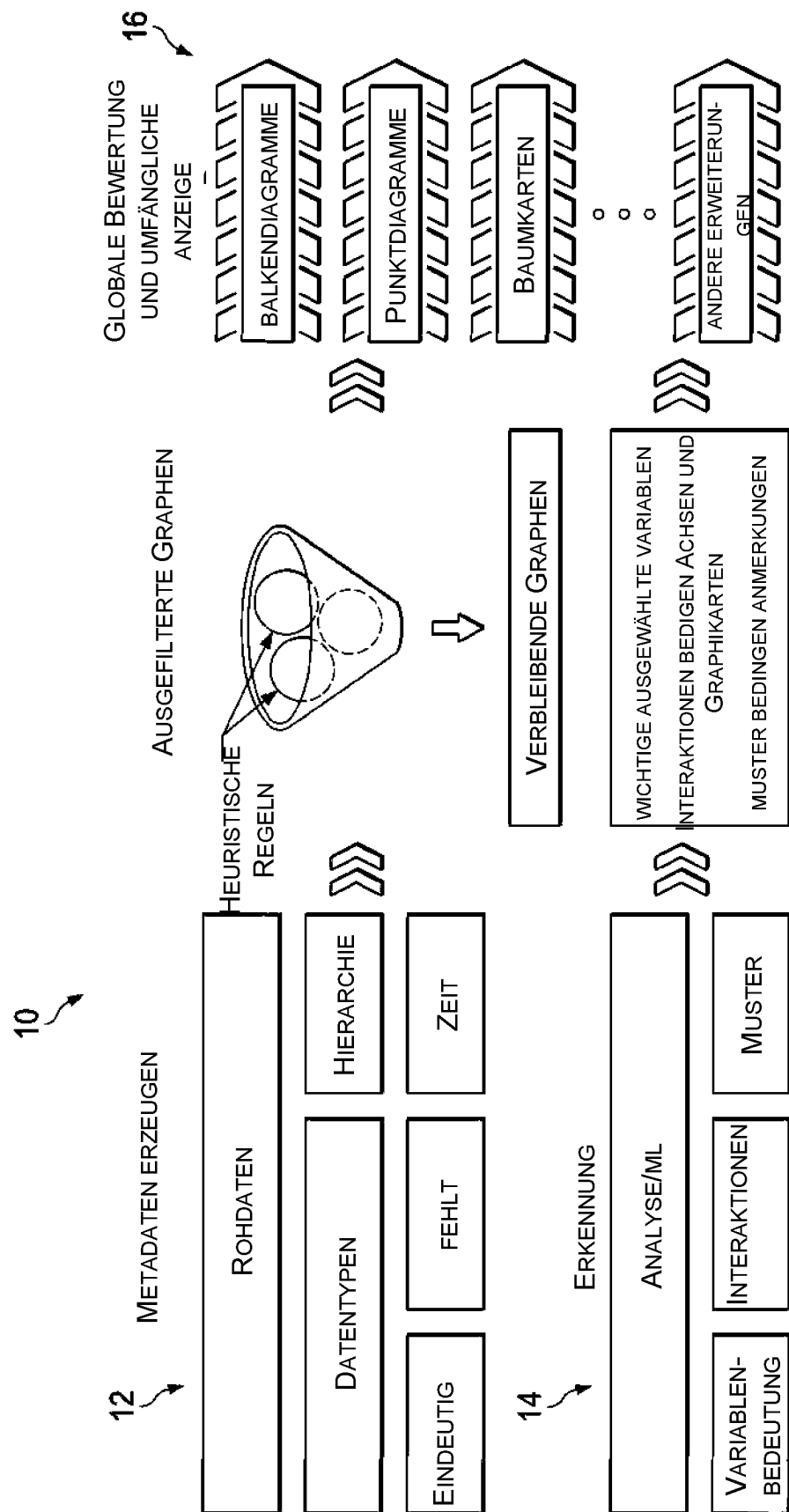


FIG. 1

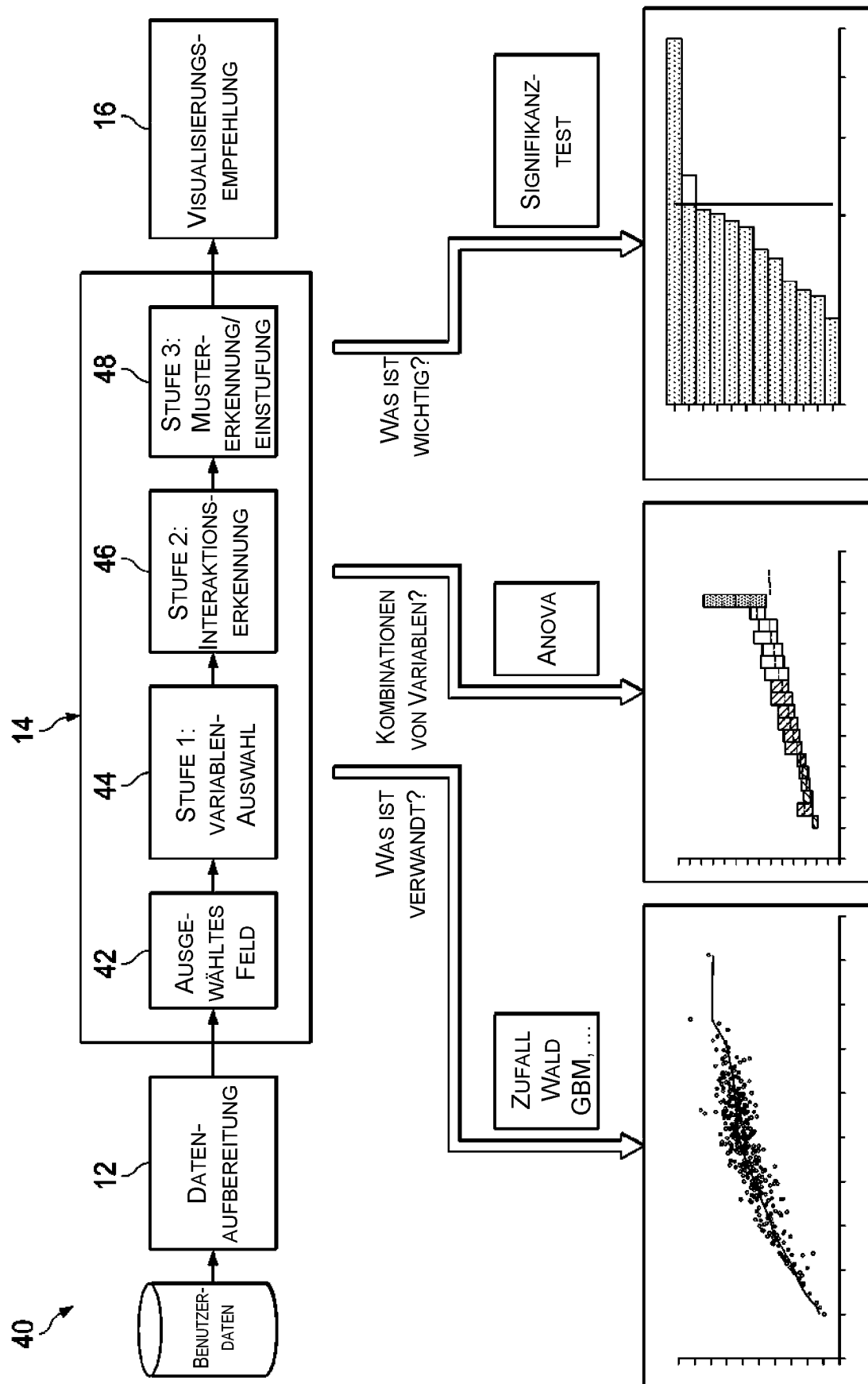
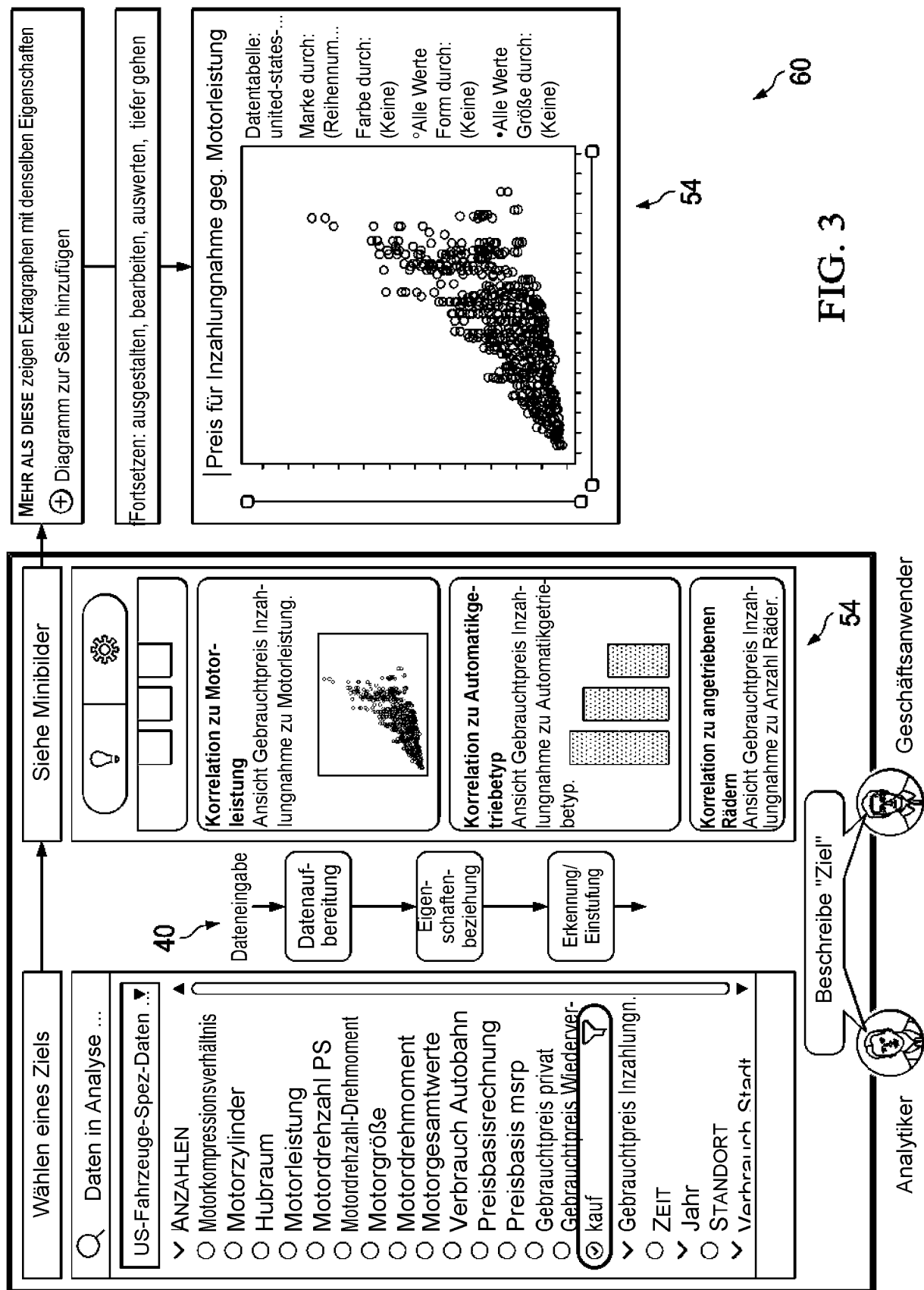
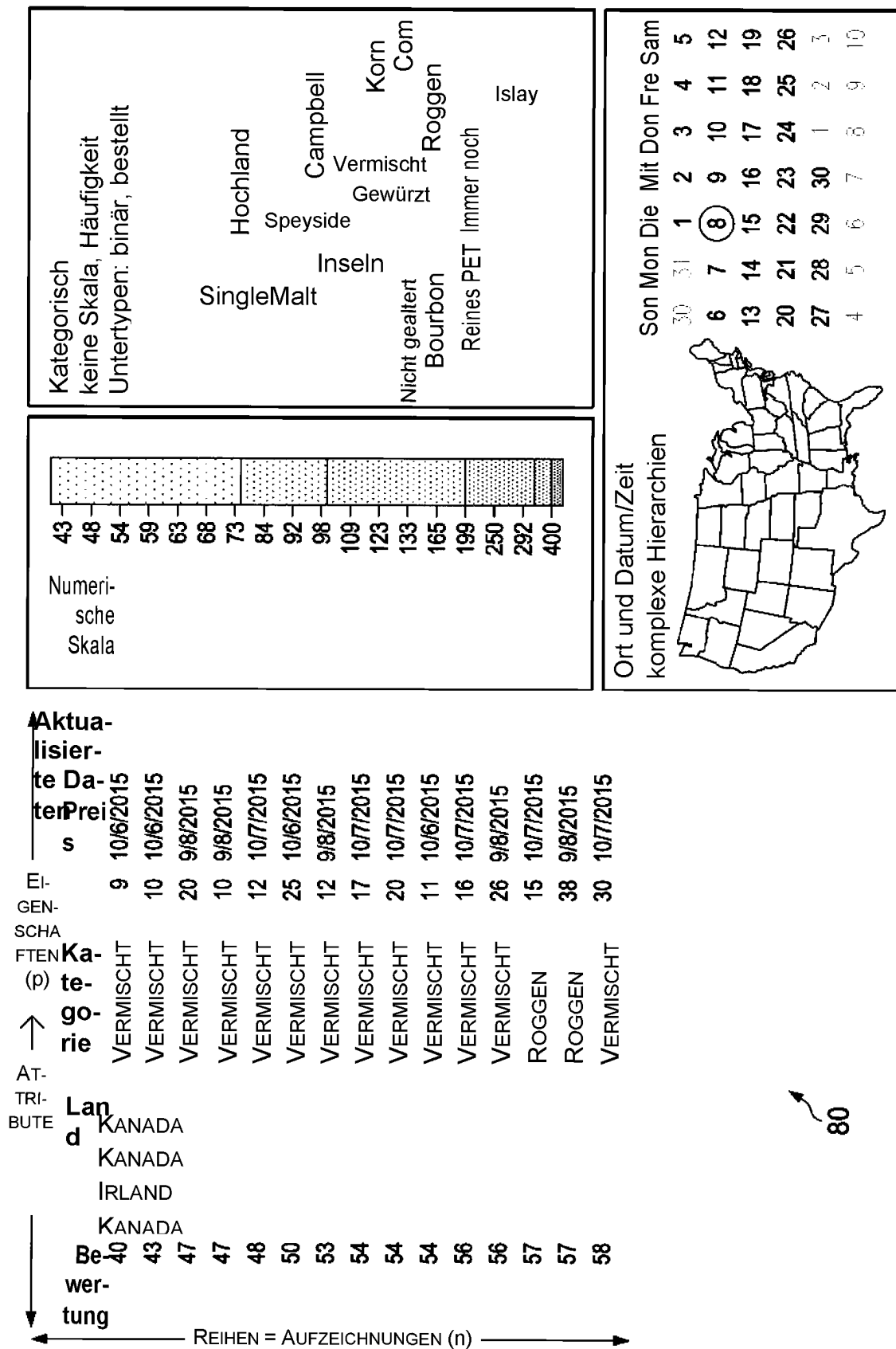


FIG. 2





		Attribute		Eigenschaften	
Reihen = Aufzeichnungen	Bewertung	Land	Kategorien	Preis	Aktualisierte Daten
	40	Kanada	Vermischt	9	10/6/2015
	43	Kanada	Vermischt	10	10/6/2015
	47	Irland	Vermischt	20	9/8/2015
	47	Kanada	Vermischt	10	9/8/2015
	48	Kanada	Vermischt	12	10/7/2015
	50	Irland	Vermischt	25	10/6/2015
	53	Kanada	Vermischt	12	9/8/2015
	54	Irland	Vermischt	17	10/7/2015
	54	USA	Vermischt	20	10/7/2015
	54	Kanada	Kanada	11	10/6/2015
	56	Kanada	Kanada	16	10/7/2015
	56	Kanada	Kanada	26	9/8/2015
	57	USA	Roggen	15	10/7/2015
	57	USA	Roggen	38	9/8/2015
	58	Schottland	Vermischt	30	10/7/2015

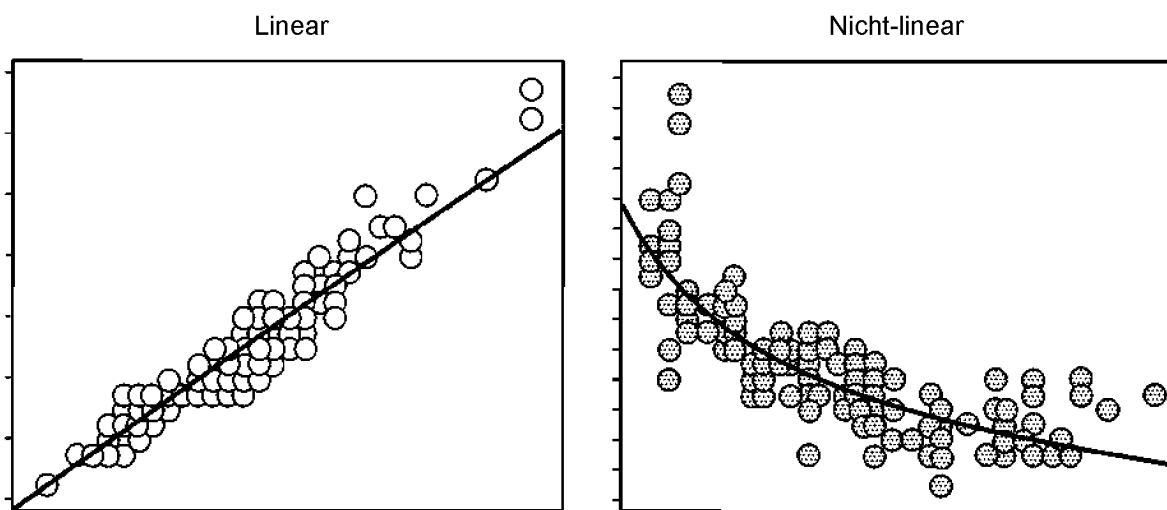


FIG. 5

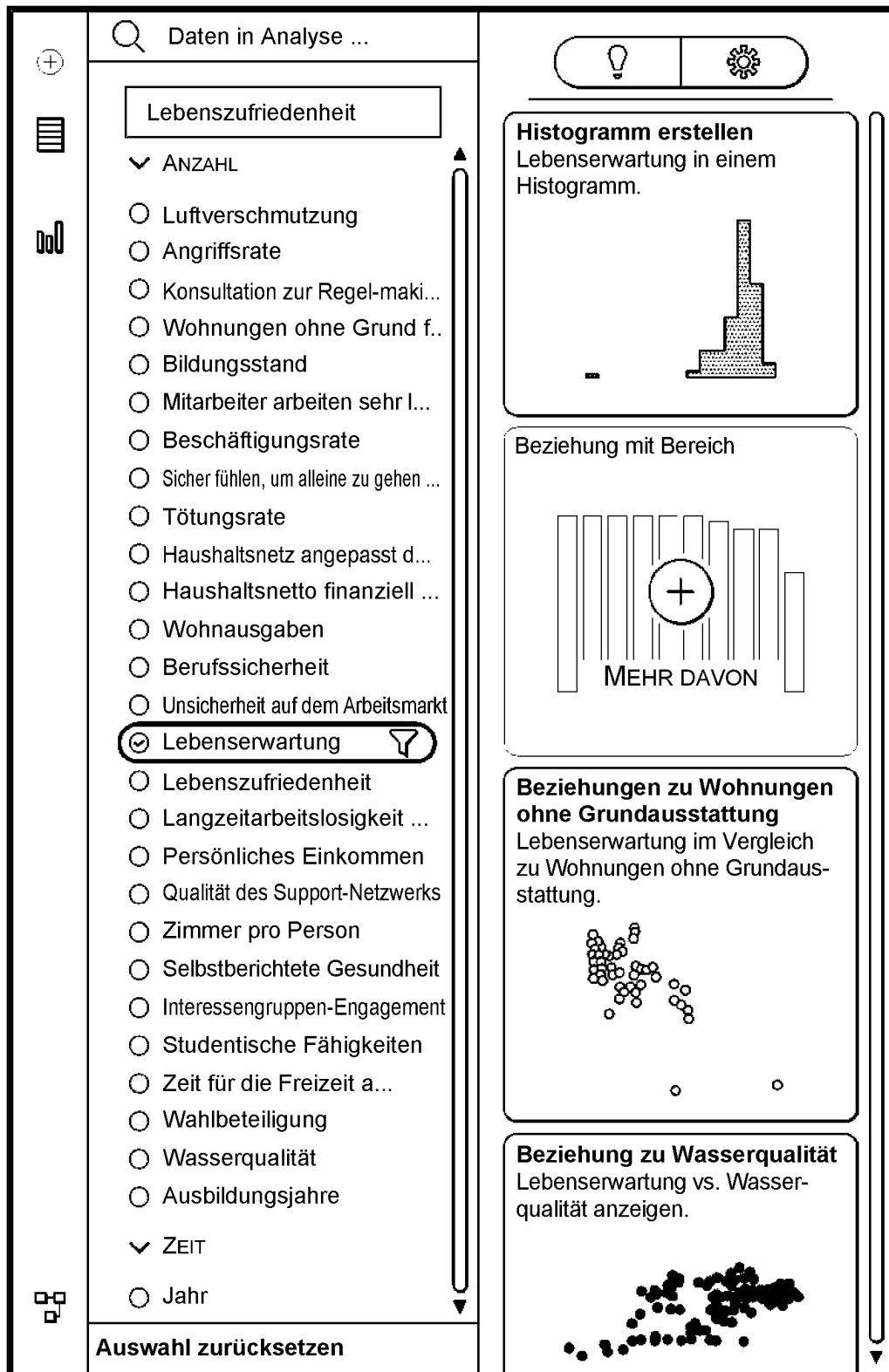


FIG. 6

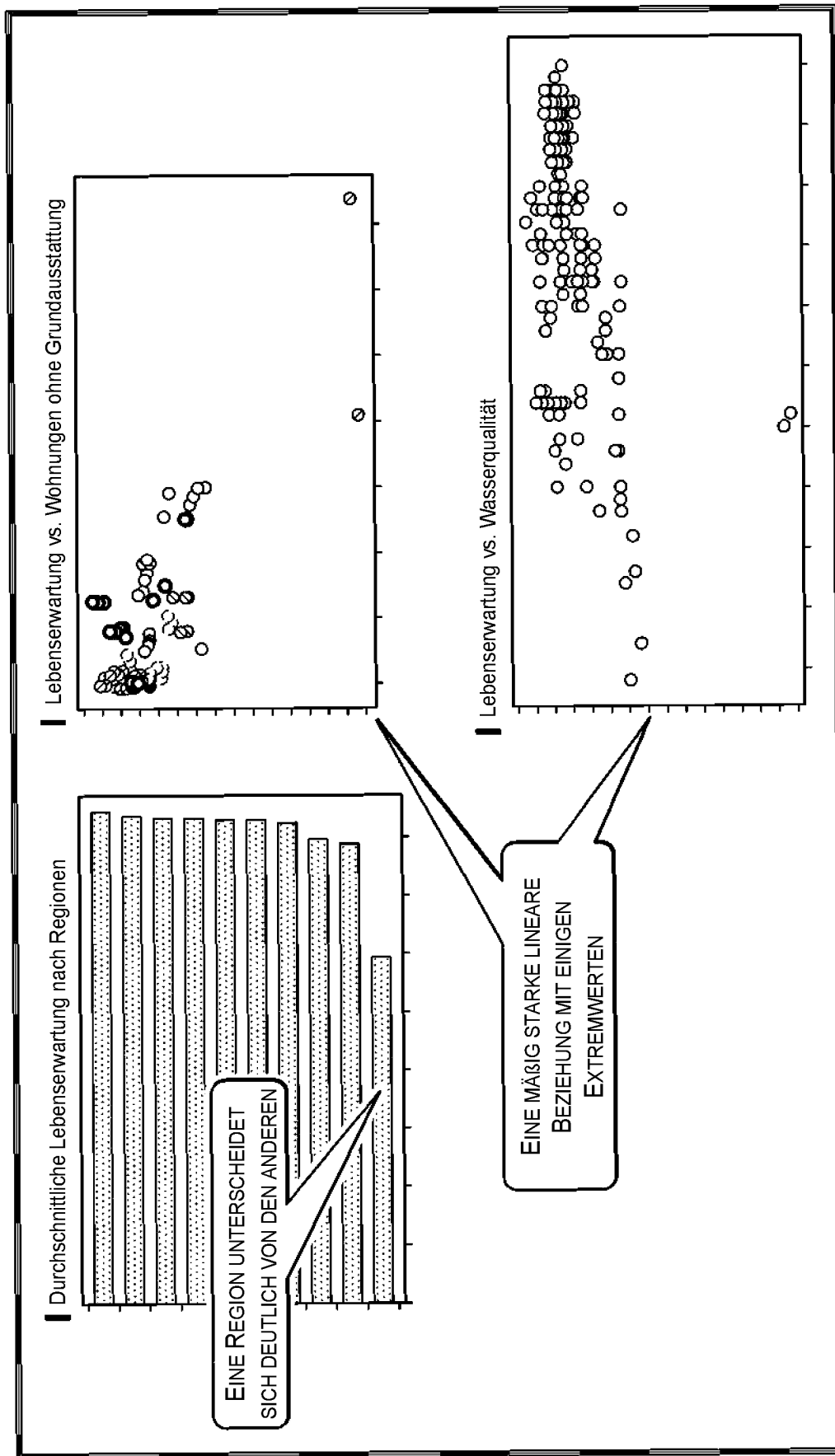


FIG. 7

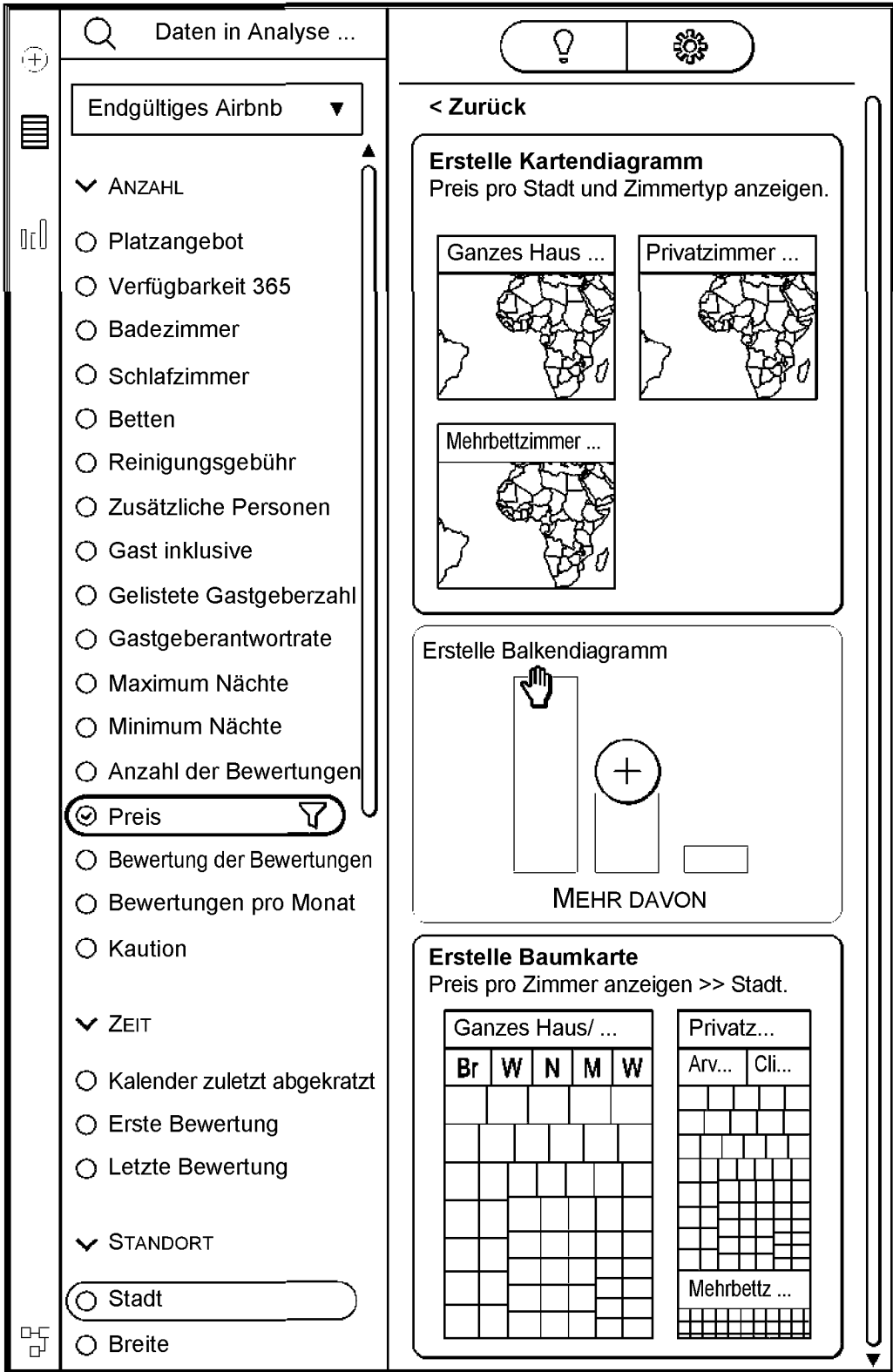


FIG. 8A

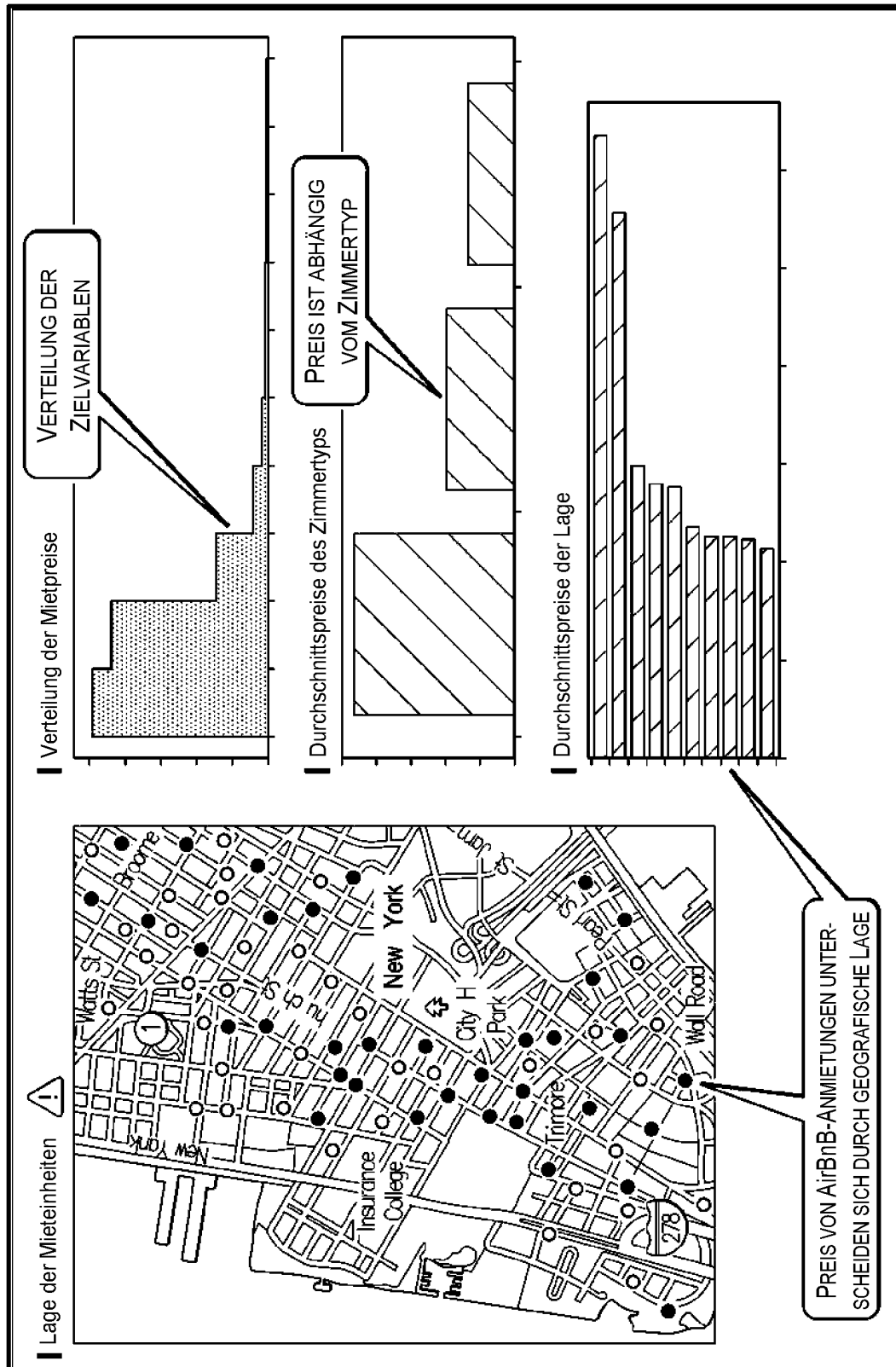


FIG. 8B

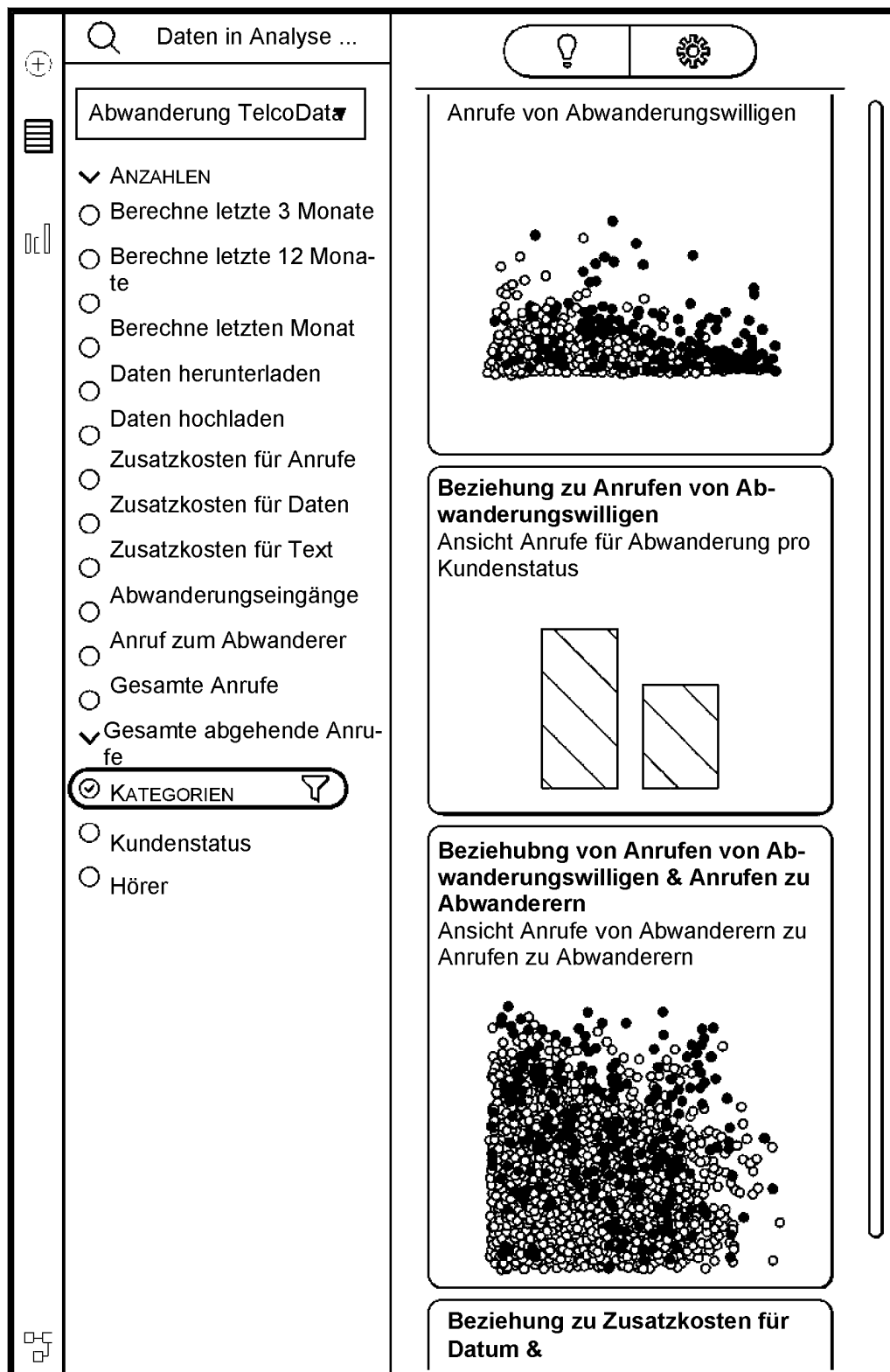
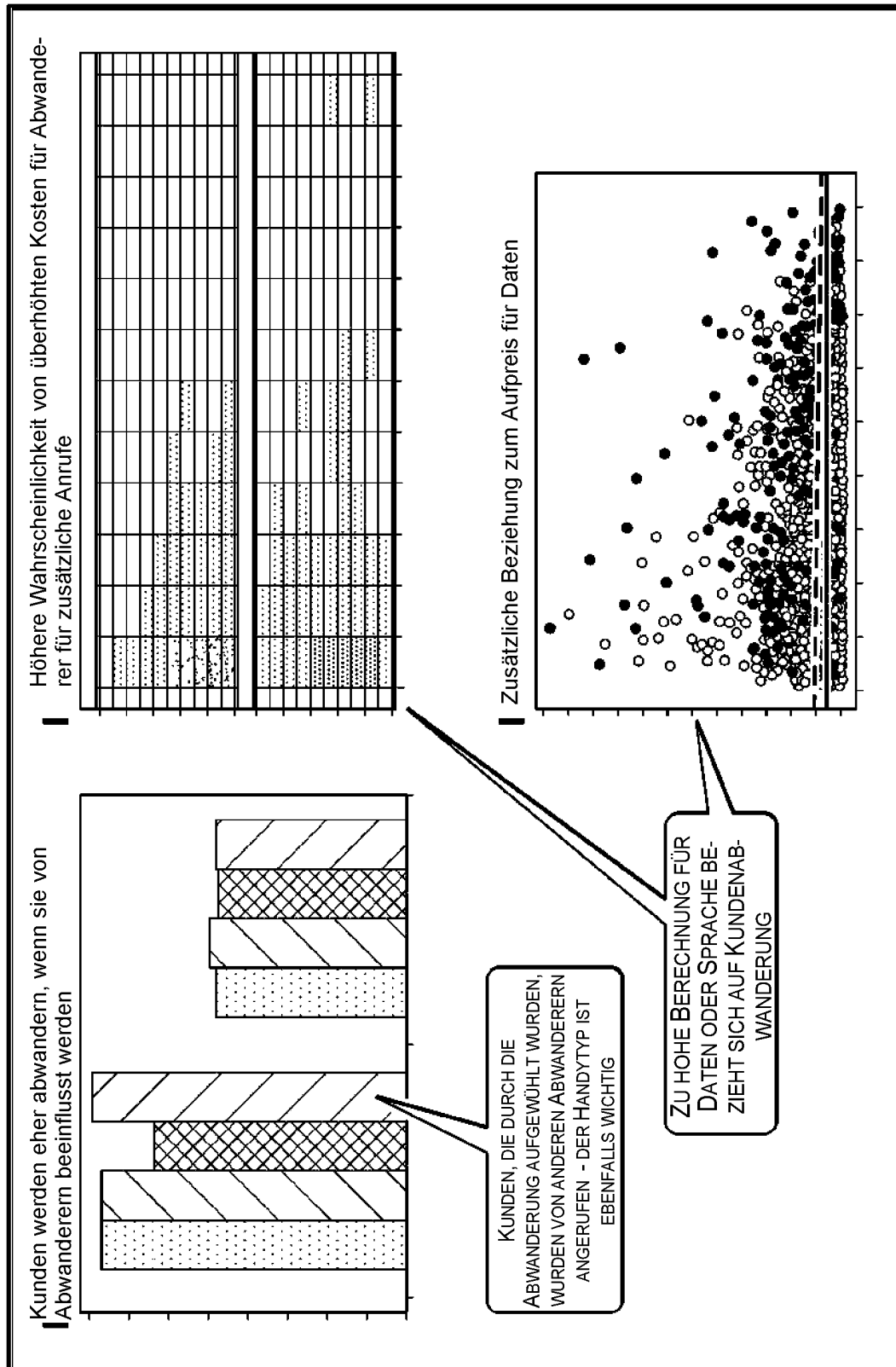


FIG. 9A



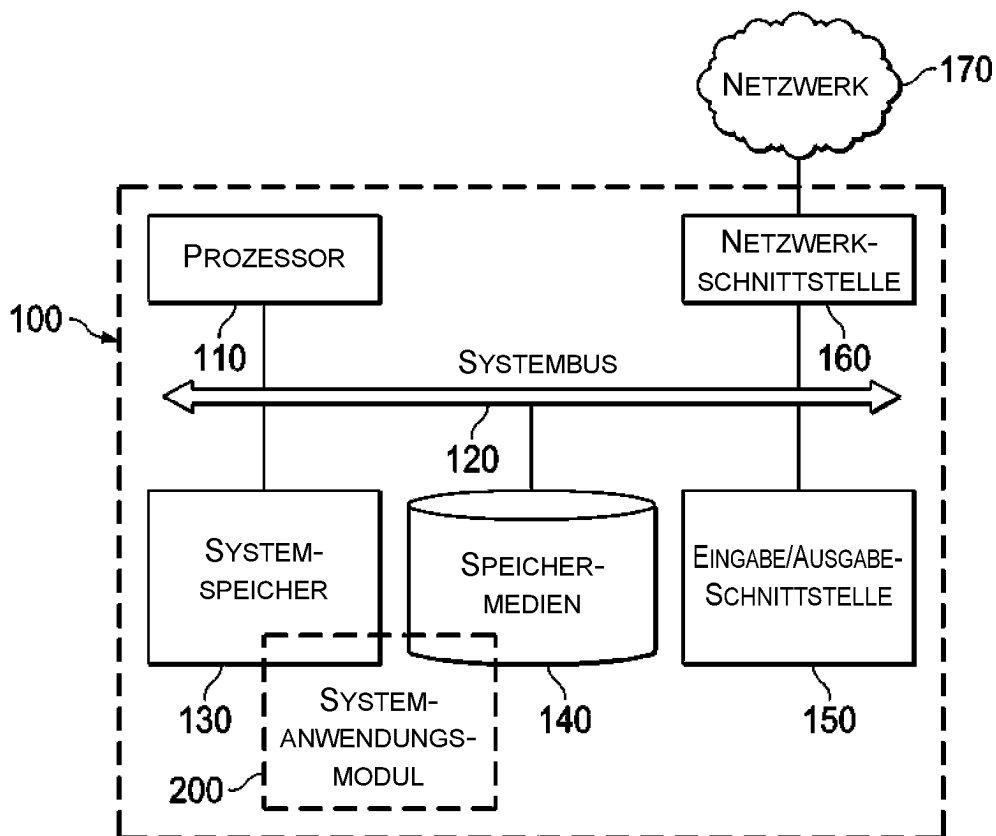


FIG. 10