

158 of 172 DOCUMENTS

UNITED STATES PATENT AND TRADEMARK OFFICE PRE-GRANT
PUBLICATION

20030004958

(Note: This is a Patent Application only.)

[Link to Claims Section](#)

January 2, 2003

Platonic reasoning process

INVENTOR: Russell, Lucian - Alexandria, Virginia, United States (US)**APPL-NO:** 896601 (09)**FILED-DATE:** June 29, 2001**LEGAL-REP:** Greenberg and Lieberman Law Offices - 314 Philadelphia Ave., Takoma Park, Maryland, 20912**PUB-TYPE:** January 2, 2003 - Utility Patent Application Publication (A1)**PUB-COUNTRY:** United States (US)**US-MAIN-CL:** 707#100**CL:** 707**IPC-MAIN-CL:** [7] G06F 007#00**IPC-ADDL-CL:** [7] G06F 017#00**ENGLISH-ABST:**

The distinguishing feature of the present invention is that it provides a means to answer a query about a database when the data in the database is not complete or is not considered to be trustworthy. The adverb Platonic is used to describe the reasoning process because of Plato's metaphorical description of how human beings perceive reality. The metaphor was one of a fire in a cave. Plato said that human beings cannot perceive objects in the real world in their exact form. If an object were in a cave, a fire in the cave would cast a shadow of that object on the wall. That shadow, however, would alter shape and the edges would appear to flicker. A person in that cave facing the wall would not be able to see the true form of the object, only the shadows. However, by looking at those shadows it would be possible to get a good approximation of the shape of the actual object. That is the intent of the present invention, to process the data as so as to obtain a good approximation of the object in the real world that the data represents.

EXMPL-FIGURE: 4**NO-DRWNG-PP:** 8

SUMMARY:

[0001] Priority is hereby claimed to U.S. provisional patent application No. [lowbar][lowbar][lowbar][lowbar][lowbar][lowbar] filed on [lowbar][lowbar][lowbar][lowbar][lowbar][lowbar].

BACKGROUND OF THE INVENTION

[0002] There has been extensive work in reasoning about the data in databases going back over 40 years. This work encompasses monotonic reasoning, best known in the guise of automatic theorem proving, and non-monotonic reasoning of truth maintenance.

[0003] There has been extensive work in the area of applying Bayesian probability measures to complex situations, notably the work of Judea Pearl (Probabilistic Reasoning in Intelligent Systems : Networks of Plausible Inference), who maintains a bibliography of over 1000 references. Results are documented in many conferences on uncertainty reasoning. The only known practical example of the application of this formalism has been in text retrieval. The INQUERY system, developed at the University of Massachusetts at Amherst uses the relative occurrence of words in documents as an estimate of probability and then performs a Bayesian net inference on the data.

[0004] There has been extensive literature in fuzzy sets. For a baseline see Klir's "Fuzzy Sets and Fuzzy Logic: Theory and Applications", and "Uncertainty-Based Information : Elements of Generalized Information Theory." Again, the question arises on how to assign fuzzy measures. For a small number of measures of simple systems they can be assigned by a human being. This breaks down when the problem becomes too complex.

[0005] Baconian measures have not been applied mathematically and systematically except in a few academic settings. They have not been used as a technique to compute fuzzy measures.

[0006] The use of filters was discussed in great detail by Lucian Russell in "Posits and Rationales", a Ph.D. dissertation at George Mason University. It has been submitted as documentation for patents in 1998.

[0007] Computer representations of information are stored in a database. The following definitions are those generally used by Information Technology standards bodies. A database consists of data elements and their relationships and the range of values that the data representation is expected to assume. The data that describes other data is called metadata. The metadata structure of data in terms of groupings of data and linking of groupings is the schema. The metadata structure that describes the expected range of values is called the integrity constraints. A database that is organized into a set of tables, where linkages among tabular entries are represented by explicit data values present in those tables, can be called a relational database. This is because the tables are a representation of the mathematical construct called a relation. The term relation is mathematically defined whereas the term relationship has a more general meaning, and requires a context in which it is precisely defined.

[0008] A relational database has the property that any of its data elements can be reassembled into another table using a combination of three operations, select, project and join. A complex operation consisting of the use of these operations, perhaps multiple times, is a query. The use of a sequence of queries can also be used to create new tables. This process has been mathematically demonstrated to be identical, under three assumptions, to be the same as proving an equivalent mathematical theorem using that data. Because of this fact a process that queries data is a reasoning process. Therefore the Platonic Reasoning Process is also a Platonic Querying Process.

SUMMARY OF THE INVENTION

[0009] The distinguishing feature of the PRP is that it provides a means to answer a query about a database when the data in the database is not complete or is not considered to be trustworthy. The adverb Platonic is used to describe

the reasoning process because of Plato's metaphorical description of how human beings perceive reality. The metaphor was one of a fire in a cave. Plato said that human beings cannot perceive objects in the real world in their exact form. If an object were in a cave, a fire in the cave would cast a shadow of that object on the wall. That shadow, however, would alter shape and the edges would appear to flicker. A person in that cave facing the wall would not be able to see the true form of the object, only the shadows. However, by looking at those shadows it would be possible to get a good approximation of the shape of the actual object. That is the intent of the PRP, to process the data as so as to obtain a good approximation of the object in the real world that the data represents.

[0010] The most general type of schema structure is called an Ontology. It provides a structure, frames containing slots, with inter-frame linkages, that is so general that it suffices to represent all the different varieties of metadata that have been found to be useful, and it can be extended to encompass new types of metadata. The Ontology provides the description of the data about the real world that, if available, represents the object of set or objects (1) about which data is collected and (2) queries made. In other words the Ontology can represent the database that describes the ideal description, the one that we would wish to obtain. The Ontology, because of its generality, can also describe the limitations that existed and constrain the completeness and accuracy and validity of the data that was actually collected in the real world.

[0011] Therefore the Ontology is the means whereby we (1) organize the computer representations of information together with the representation of the presumed relationships that exist among elements of that information, and (2) applies and records corresponding measures of completeness and correctness of that information.

[0012] The third step in the process is to generate new measures. The process is valuable because it is a means to generate "fuzzy" measures. The discipline of "Fuzzy Sets" and "Fuzzy Logic" has been well established for over 35 years; it is the mathematics that results from generalizing the binary function describing the membership relation of subsets of a set to a more general function. A binary relation is a mapping of an element "a" of a set "A" and a subset "S" of "A" to one of two numbers, "1" or "0" depending on whether "a" is said to be a member of the subset "S". This mapping is called the membership function. When the mapping is not just to "1" or "0" but to numbers in between then the membership function is "fuzzy".

[0013] Fuzzy measures and the resulting fuzzy logic has been well accepted in the United States because of a linguistic prejudice against the word "fuzzy". The mathematics, however, exists independent of linguistic conventions, and has been applied successfully to many industrial processes in Japan, where patents have been granted. In general the strength of the methods based on "fuzzy logic" is that they are used to create a simpler mathematical model of the control processes that need to be managed, which enable the creation of more efficient techniques for controlling the automatic operation of machinery.

[0014] Although the usefulness and applicability of fuzzy mathematics is well demonstrated for nearly two decades, the extension of these methods to more general, non mechanistic problems has been halted by the difficult in setting forth general criteria by which to assign the measures to real world data. A similar problem exists in using probabilistic techniques that are called "Bayesian". These require the generation of a large number of initial probability values that must be assigned to data, and the task of generating these probabilities has proved to be too complex for the technique to be useful to real problems, with the exception of the Computer Science discipline of Information Retrieval.

[0015] The PRP, however, provides a mechanism for generating fuzzy measures. These are the "new measures" that provides a measure of a new category of information that summarizes the information available concisely. The latter summarization is what is done when fuzzy measures are applied in mechanical control systems. The process allows for two techniques to be applied, linguistic variable mapping and precision through abstraction.

[0016] 1. Linguistic variables are words assigned to ranges of data values. An example is the risk of an investment in a company's stock where the chance that you would make a 100% profit on your investment would be characterized by one of the words "certainly, probably, plausibly, possibly, conceivably and inconceivably".

[0017] 2. Abstraction occurs when detail is omitted. If values for three variables are needed to describe an object, but only two are certain, one may define a new object that is described by just the two variables. For example to designate an area as a "mountain" one needs the height of the object, whereas a "mountain range" is a flat area on the map that has both mountains and non-mountainous areas. Lacking the height coordinate the latter more abstract term is appropriate.

[0018] Using a-priori principles is one way of assigning the value of fuzzy measures. The PRP however, uses a more novel technique whereby the measures are induced by the data. There are several meanings for the word induction, but the one that is intended here is the one from electromagnetic theory. In this theory an electric charge moving through a conductor induces a magnetic field, and if another conductor is placed in the same field a current in the opposite direction. The data that is actually loaded into the database creates a series of objects which generate an description of those objects with respect to the description loaded in the Ontology.

[0019] The foundation of this technique is the application of implicit functions. In mathematics a function is a table of values, like a relation, with some restrictions. Usually people describe functions "intensionally", by a formula, especially if the formula has an infinite number of values like $x^2=y$. An alternative is to describe functions "extensionally", as a table where all values are explicitly recorded. Any given relational database can be described as a set of extensionally defined functions. These can in turn generate self-describing measures, of which one familiar example is quartiles (quartiles are four groupings of numerical values within a set of numbers from the lowest to the highest). By applying measures created by self-describing functions to the database it is possible to create fuzzy measures. These measures are then applied to the data in the database. The data and their fuzzy measurements enable the use of reasoning, the application of rules of the form "If X then Y" where the logical expressions X and Y contain terms that include the fuzzy measures.

DETDESC:

DETAILED DESCRIPTION

[0020] The PRP is realized by adding a set of processes, realized in software, into a system that performs reasoning on data. The PRP is the process of reasoning about the set-up of the data that will be used by process that reasons on exact data. Some basic processing functions are common to both types of reasoning, and so the software will share some programs in common. The baseline system components perform reasoning on exact data. As explained above, reasoning about a hypothesis using data in a database is the same thing as running a set of programs that perform a query on the database. That is because a hypothesis that can be validated or disproved on a data base has the form "there exist data elements in the database form Y that match the description X" can be proven by the query "select all data from D that match the description X". If no such data exists the query returns no data and thereby disproves the hypothesis.

[0021] The system will be shown in terms of a software Architecture diagram, one which identifies software components with special functions that interface with one another. The Architecture of the system is shown in FIG. 1. It is a 3-tier architecture; the terms layer and tier will be used interchangeably. The first tier contains a user interface, a software client, that sends data to and receives it from software on the second tier, the middleware layer. At the bottom of the Figure, Tier 3, are Database Management Systems (DBMSs) accessing databases containing the original information that the system must access.

[0022] The baseline reasoning process without any PRP functions is simple. In the Client Tier there is a box that says "Reasoning Chains and Filter Editor" and one that says "Display Results". The Reasoning Chain is the statement of the hypothesis to be proved, which may also be seen as a set of queries. There are no filters in this case. The results of the queries are displayed where indicated, if there is data to be displayed, and a "no data found" message is displayed

otherwise. The software interacts along the path shown as the narrow, black think lines with triangular solid arrowheads. Note that the PRP Plug-In is bypassed and the SQL 3 Engine is directly accessed. The Object-Relational Database Management System is a standard middleware software Commercial Off the Shelf (COTS) software component, used to overcome formatting differences among the different data sources.

[0023] The description in FIG. 1 covers the case where the PRP is used.

[0024] Tier 1

[0025] 1. The Ontology Editor is the interface to the Ontology Builder (an example is available at the Knowledge Sharing Laboratory, www-ksl.stanford.edu at Stanford University). An Ontology is a general representation format for systems that represent knowledge about data items and their inter-relationships, including the various elements of metadata that describe the data. When an instance of an Ontology is built it will contain descriptions of (1) the ideal collection of descriptive data, (2) the data that is actually available, (3) transformations that are admissible between the two formats, (4) functions used to filter data that will be transformed, and screens that support the generation of new measures and objects that are the result of the PRP. This interface supports the first part of the PRP: it organizes computer representations of information together with the representation of the presumed relationships that exist among elements of that information.

[0026] 2. The Results Display Engine is an interface to COTS products on the client platform that are used to view the results of performing reasoning using the PRP.

[0027] 3. The User Interface for Reasoning Chains: This user interface is to two things, build the hypothesis and its associated reasoning chain (or query sets) and assign uncertainty management conditions. This supports the first half of the second part of the PRP: it records the measures of correctness corresponding of the information that will be used. It also feeds rules to any reasoning engine that may be needed to enhance the functionality of the Object Relational Database.

[0028] 4. The Screens Editor is the interface to the Ontology builder that is used to support the second half of the second part of the PRP: record the measures of completeness corresponding to the information that will be used.

[0029] In addition there are system administrators who need interface with the system through COTS product interfaces. These are:

[0030] 5. The Command Center: The Object Relational Database's Interface

[0031] 6. Math Function Editor: This is an environment for building programs that are compiled and inserted into an executable program library for use in the Object Relational database by the PRP.

[0032] Tier 2, the Middleware

[0033] Middleware is used to assemble data from multiple data sources, shown as databases in Tier 3. The following components constitute the middleware:

[0034] 1. Object Relational Database: the Object-Relational database management system (ORDBMS) that is a general-purpose data management system that contains the ability to store relational data and other types of data. Although many of its functions could be managed purely be a relational database doing so would make for more complex application programs that access and use that data. In the diagram it will build databases of filters and screens because any person using the PRP will run a problem, decide on some changes and want to store the prior working assumptions in a database for reuse.

[0035] 2. The Reasoning Engine: the Reasoning Engine executes rules that impact both the Ontology and the data

in the results database. It is likely to be part of the ORDBMS, but is shown separately just in case a more powerful reasoning system is required.

[0036] 3. PRP Application Server: the activities of the client user interfaces is coordinated by this application. It will interface with the ORDBMS to load screen and filters, and initiate the access to raw data and initiate the use of rule sets in order to generate results. This is the third part of the PRP: it generates a new, novel and useful measure that provides a useful new description of the information available that summarizes it concisely.

[0037] Tier 3

[0038] This is the data that is input to the system.

[0039] Refining Hypotheses is the term used describe the user's activity of interacting with the data as follows:

[0040] 1. The user come to the database with an initial hypothesis which he/she wants to validate using the data in the database.

[0041] 2. The user examines the data and comes up with a chain of reasoning, a set of steps during which the data in the database will be accessed, transformed, intermediate results created and finally a result generated.

[0042] 3. The results are examined and the process is repeated with either a reformulated hypothesis, a change in the scope and/or transformations of the data that is to be examined or both. The new results are examined.

[0043] 4. The process is repeated until the user is satisfied.

[0044] This interaction can be done with or without using the PRP. The PRP provides a more powerful way for the user to use the data available

[0045] We assume that the user is familiar with the specialized field of inquiry to which the data in the database is relevant. That means specifically that due to the user's personal training and experience he/she knows what types of hypotheses may be propounded and validated with the data. For purposes of illustration we assume that the user is processing data to assess data describing the current situation in which it is suspected their may be a threat to resources, e.g. business, medical, or military assets. The data in the database may admit to multiple interpretations. Each can be formulated as an initial decision hypothesis: as follows: "the current situation X poses a threat to my resources Y at Z". After the data is analyzed one of the following may be inferred:

[0046] 1. Contradiction: "the current situation X does not pose a threat to my resources Y at Z"

[0047] 2. Alternatives: "the current situation X poses a threat to my resources Y at W" OR "the current situation X poses a threat to my resources Y[prime] at Z"

[0048] 3. Ambiguity: "the current situation cannot be assessed with sufficient certainty to support any hypothesis".

[0049] To use the data in the computer the hypothesis must be formulated in a specific manner. Because R. Reiter proved in 1984 that a query to a relational database was mathematically equivalent to a proof that the data supported a hypothesis the form of the hypothesis can be that of a query. The conversion from the verbal human statement to the computer format of the hypothesis is a two step process (the notation[mdash]backwards E means "there exists" and inverted A means "for all"):

[0050] 1. Expand the Meanings:

[0051] ([Thereexists](a,b,c . . .) with relationships r1, r2, . . .)whenever situation X exists.

[0052] ([forall](a,b,c . . .) with relationships S1,s2, . . .) define the my resources Y.

[0053] The area Z is defined by criteria (A,B,C, . . .)

[0054] 2. Restate the verbal hypothesis as a query: "the evidence available in the database shows X, Y exist and Y meets criteria (A,B,C,. . .)".

[0055] The word used in step 2, however, was not "data" but "evidence". Once the conversion is made, the evidence will either (1) support the assumptions of a decision hypothesis H1 or (2) contradict it by supporting its contradiction H1C, or (3) not be relevant at all. Data is not necessarily evidence, and a method of relevance determination is provided to convert the stream of data to evidence. One of its actions is to defined exactly when redundant data is present. Such data need not be considered (eliminating redundant data solves the info-glut problem). This is one of the explicit PRP process steps, converting data to evidence by building screens. The useful advantage of this step potentially reduces the massive volume of data (also known as "infoglut") and makes both the reasoning process and results more amenable to being effectively reviewed by a human.

[0056] Returning to the point about alternative hypotheses: whenever evidence is not relevant to one hypothesis about a situation it might be relevant to another. Thus this step of looking at alternative hypotheses can be applied to situations where multiple explanations for data are possible. A user may start out with one hypothesis, and monitor the situation, looking for new evidence that suggests that an initial hypothesis is now negated. This means that when the OPRP is used for monitoring it provides a useful advantage as all the plausible hypotheses by be compared against the evidence compared.

[0057] In FIG. 2 we see that Hypothesis1 predicts two negative events and three positive ones, but has 4 events unaccounted for. The others have different coverage. Under the circumstances Hypothesis2 is the best. The goal of the step in the PRP is to enable such a comparison.

[0058] The process steps are:

[0059] 1. Initialize problem space: specify

[0060] Decision: a hypothesis OR a set of hypotheses.

[0061] Data sources, objects, algorithms etc.

[0062] 2. Generate missing data

[0063] 3. Establish the Chain of Reasoning: data transformations needed to gather the data neededd to test the hypothesis or hypotheses.

[0064] 4. Set thresholds of uncertainty for valid data.

[0065] 5. Create screens to convert data to evidential objects.

[0066] 6. Run hypotheses verification & compare results.

[0067] 7. Adjust the abstraction level to encompass evidence.

[0068] 8. Potentially Perform Data Mining to improve results.

[0069] The first step is to state the subjective decision hypothesis or a set of competing hypotheses as a query. This means reformulating each hypotheses as a logical combination of one or more statements of the form:

[0070] "There exist objects whose properties [Pi] have value ranges [Ri]"

[0071] "All objects of type X have properties [Pi] within value ranges [Ri]."

[0072] In logic these are the generalizations of statements containing "OR" conditions and "AND" conditions.

[0073] The second step is to use the Ontology editor to describe the

[0074] 1. Data Sources: form and constraints for all levels of available data inputs, and the

[0075] 2. Objects: complete description of all properties of objects that are detectable, of sensors, and abstract concepts (Ontology). This includes screen objects described below Further one should specify any additional

[0076] 3. algorithms, i.e. any new techniques embodied in programs that must be added to the PRP to enable filtering of data or eliminating uncertainty. This includes filter algorithms and thresholds described below.

[0077] These may be submitted also as input to the object relational database management system as needed.

[0078] Missing Data is generated after a subjective judgement by the PRP user. It may occur as an initial step, or during re-iteration of previous steps. Some sources

[0079] Projection of prior data about objects no longer visible,

[0080] Assumptions based on knowledge of enemy doctrine, or

[0081] Simulated data based on the data at hand.

[0082] Multiple hypotheses may reflect multiple guesses at the missing data and its values. This data will be necessary for making projections when not all of the area with possible data is observed.

[0083] Every data manipulation is the formal equivalent of reasoning step, so the total is called the Chain of Reasoning. The Chain of Reasoning is therefore the set of transformations from the raw data to the data used in the query based solely upon the meaning and form of the data. This is the set of steps that one uses to go from the data to the hypothesis. It is the data upon which the query representing that hypothesis is run. The Ontology will contain precise descriptions of all the data formats needed, and as needed these will be loaded into the ORDBMS.

[0084] FIG. 3 is a visualization of four sources of data that will be used to generate the Final View, data used to test the hypothesis. Source 1 is input to other data streams and is therefore some standard immutable description like a terrain map. Other processing steps combine or fuse data from the different sources to make intermediate data sets. The "F" symbol stand for filtering action described below.

[0085] The chain of reasoning must be constructed with the knowledge that not all of the data is equally valid. Whenever tests can be devised to eliminate data that is too questionable for use they will be incorporated as a filter. The filter admits some data and excludes other data. To use one, however, a threshold value must be set. That is the user's tolerance for data uncertainty. In the PRP this sub-process is explicit and may be reset to a different value if the hypotheses need to be looked at a later time with a different tolerance for uncertainty. Specifically, although traditional Pascalian Probability measures can be used, as well as Baconian Probability Measures, explained below, fuzzy measures can be used as well, these can take the form of explicit assignments of values, or the values may be inferred from the Pascalian and Baconian measures, a KEY feature of the PRP.

[0086] Baconian Probability is not as well known as Pascalian Probability, although de-facto it is the basis of scientific induction and is used extensively albeit informally. Let $B(H)$ be the monadic (i.e. True or False) Baconian Probability of a hypothesis H and $B(H,E)$ the dyadic conditional probability of H given E ; then these are the formal

mathematical properties [Schum, The Evidential Foundations of Probabilistic Reasoning, pp254-255]:

[0087] (1) Ordinal Property:

[0088] monadic case: $B(H1) \geq B(H2)$ or $B(H2) \geq B(H1)$

[0089] dyadic case: $B(H1, E1^*) \geq B(H2, E2^*)$ or $B(H2, E2^*) \geq B(H1, E1^*)$

[0090] (2) Negation Property:

[0091] monadic case IF $B(H) > 0$ then $B(\neg H) = 0$

[0092] dyadic case: IF $B(H, E^*) > 0$ then if $B(\neg E^*) = 0$, then $B(\neg H, E^*) = 0$

[0093] (3) Conjunction Rule:

[0094] monadic case: IF $B(H1) \geq B(H2)$ THEN $B(H1 \cap H2) = B(H2)$

[0095] dyadic case: IF $B(H1, E^*) \geq B(H2, E^*)$ THEN $B(H1 \cap H2, E^*) = B(H2, E^*)$

[0096] (4) Disjunction Rule:

[0097] monadic case: IF $B(H1) \geq B(H2)$ THEN $B(H1 \cup H2) = B(H1)$

[0098] dyadic case: IF $B(H1, E^*) \geq B(H2, E^*)$ THEN $B(H1 \cup H2, E^*) = B(H1, E^*)$

[0099] (5) Contraposition:

[0100] dyadic case: $B(H, E^*) = B(\neg E^*, \neg H)$

[0101] The system has been of great interest to the many people who have become aware of it, but heretofore no practitioners have seen a way to interpret it to be applicable to real-world situations. The PRP provides this assignment by a novel mechanism.

[0102] The Baconian probability associated with an object is determined by (1) the object observed, (2) the question asked of it and (3) the number of tests of values of relevant variables. The innovation is to applying the probability to all terms or slots of the Ontological description of the object, not just the static attributes. For example, a jeep may have 25 possible data values or properties that describe it, but only six of them are needed for identification. Let the hypothesis H be that object X is a jeep. Suppose the data contains 4 relations each of which of six variables, and together they encompass 12 terms of the 25, but also together they account for only 5 of the 6 identifying attributes. Then the Baconian probability of the (H) equals 5. This corresponds to the data passing five out of six possible Baconian existence tests on the different identifying property.

[0103] This process is totally new and is a major invention within the PRP. The problem is relevance of data and the value of having multiple instances of the "same" data. This must be addressed because future technology will enable better, more accurate, and more numerous data to be collected. The challenge is how then to use them without being overwhelmed. The step is the conversion of data to evidence.

[0104] The key step is to provide a precise description of the distinctions made in the discipline of Evidential Reasoning. Let H1 be the decision hypothesis, and D1 a datum that supports it. If D1 supports a decision hypothesis H1 does the next datum D2 support the hypothesis more, less, the same. Is it relevant at all. In Evidential Reasoning the following definitions are provided:

[0105] Directly Relevant Data: data that is used to infer the decision hypothesis H.

[0106] Corroborating Data: data that strengthens the decision hypothesis H.

[0107] Redundant Data: data that duplicates what is already known about H.

[0108] Contradictory Data: data that supports the negation of H.

[0109] Conflicting Data: data either confirms or negates H_i but does so for H_j .

[0110] Using the PRP technology the user:

[0111] 1. enables the system to use relevant data as direct evidence,

[0112] 2. fuses corroborating data,

[0113] 3. screens out redundant data relevant to the competing hypotheses H_j and their contradictions \tilde{H}_j or H_j^c , and

[0114] 4. assigns conflicting data to the same process steps for competing hypotheses.

[0115] This is made possible by the use of the Ontology.

[0116] Hypothesis comparison requires defining the above terms precisely, and relating them to multiple hypothesis comparison. First the PRP user creates an Ontology for all objects $[X_1, \dots, X_M]$, i.e. a complete logical description of the terms (a.k.a "slots) for each object i.e. properties (attributes), functions, relations and axioms. Let object X have terms (t_1, t_2, t_3, \dots) that may be any of the types of slots mentioned. A given object may be identifiable, however, by a subset of those terms, and there may be more than one such subset. For object X_i let the subsets of identifying terms be $IT_{i1}, IT_{i2}, IT_{i3}, \dots, IT_{iM}$ and let IT_i be the set of these subsets. Note that if XT_i is the set of all X_i 's terms, and $2XT_i$ is its power set, the set of all subsets, then the set of sets of identifying terms $IT_i[Subset]2XT_i$.

[0117] Then consider a set $[d_0, \dots, d_n]$ of data items that establish that an instance of object X_i has been observed. The first one d_0 establishes that X_i exists. This means that values for the instance of X_i may be filled in for its property values, function and relation values, and the designation that certain axioms have been shown to apply. Then the rest of the elements of the set d_1, \dots, d_n that also establish it are split into two subsets, the corroborating evidence $[c_1, \dots, c_{n1}]$ and redundant evidence $[r_1, \dots, r_{n2}]$. The distinction is that d_j is redundant unless it provides a value for a term that d_0 did not. Corroborating data is data that adds information about the object X_i . Although it does not change the evidence that X_i exists the new term values may be useful to subsequent queries that access the database. Otherwise, as they add no information, they are redundant. Formally:

[0118] Let d_1 and d_2 be relational data that consists of attributes $(t_{11}, t_{12}, \dots, t_{1n})$ and $(t_{21}, t_{22}, \dots, t_{2m})$ respectively.

[0119] Let d_1 directly identify object X_i : it provides values for the subset of terms IT_{i1} , that is an element of IT_i . Then d_2 is:

[0120] Redundant if it confirms the values of IT_{i1} ,

[0121] Corroborative is it confirms values using IT_{i2} [notequalto] IT_{i1} ,

[0122] Contradictory if it does not have all the values of one of the IT_i elements, and

[0123] Conflicting if it identifies elements IT_{j1} of another object X_j at the same time and place as X_i .

[0124] The above technique of identifying IT_i is called creating and using a screen. One advantage is that it reduces

the data glut to a manageable flow of information. It also separates out what data accrues to what decision hypothesis about the existence of an object. Sensor fusion is now also easy to define. Every sensor is modeled as an object as well with its own Ontology. There is a line of reasoning from the output of that sensor to terms in the Ontologies of the objects that it identifies. In IT1 the infrared detector will see that the hood of a jeep is bright, meaning the jeep is operational. The logic of this association is stored in the Ontology. The fusion is done at the ontological level. The screen is now illustrated in detail in FIG. 4.

[0125] Continuing the example, a jeep may have an ontology entry yielding a set S of 25 properties, of which any of 4 subsets S1, S2, S3 or S4 of 6 properties might identify it (the ITij's). These are sensor-identification objects (SIO) (new[excl]). Then if filtered data item d1 identifies X1 according to S2 we have an identification object (IDO) (new[excl]). The value d, becomes direct evidence of the existence of X1, and S2 together with any additional sensed properties becomes the identification object (new process[excl]). Then when data item di is detected it is redundant if it uses S2 and corroborative otherwise, say using S3, as is datum d17 in FIG. 8. Adding this information updates the IDO. The IDO is the screen (new concept) with the properties in S2 [xcirc]S4 together with all other properties of d1 and d17. The screen eliminates infoglut because once enough observations are made to establish all 25 properties, then, as long as these values hold all additional data with the same values is redundant and not considered for decision making purposes: its not relevant. The IDO contains a binary array that tells which of the properties in the Ontology is present and which not by a 0/1 coding. This is used for probability filtering.

[0126] This is shown in FIG. 4. Evidence of the existence of an object and its lack of existence are both results of evidential reasoning. Evidence that exists but cannot be assigned to objects is unassigned. No hypothesis accounts for it. It can either be rejected as outlier data, or erroneous data or a new more expansive hypothesis can be generated that accounts for it.

[0127] The term of "abstraction" has been used in computer science in a number of sub-disciplines for some time. In this context it refers to the amount of detail that is provided for objects. A jeep categorized differently from other objects in that it is a vehicle that is (1) small, (2) motorized, (3) open, and is (4) all-wheel drive. The evidence may be insufficient to identify the jeep from a small convertible, but for purposes of supporting one or more the decision support hypotheses it may be sufficient to use only properties (1) and (2). In exact terms this means that a more abstract object, "small motorized vehicle" exists and a jeep is also an instance of it. The Ontology shows all such classes and allows queries to be made at any level of abstraction. If this is done, then more evidence may "come in from the cold" and be associated with the hypotheses with striking results, as shown in FIG. 4.

[0128] This is shown in FIG. 4. Evidence of the existence of an object and its lack of existence are both results of evidential reasoning. Evidence that exists but cannot be assigned to objects is unassigned. No hypothesis accounts for it. It can either be rejected as outlier data, or erroneous data or a new more expansive hypothesis can be generated that accounts for it.

[0129] The PRP also allows an alternative use of fuzzy logic. Whereas it is possible to assign a measure to a data that provides an ordinal scale mapping to values like "certainly, Probably, etc." a novel technique of applying measures results in an exact abstraction with a name that incorporates the fuzziness. The evidence on hand is describes exactly, and the fuzziness is refected in the name of the object created[mdash]it is not as exact as one would wish to have, but it summarizes precisely what is known.

[0130] The term of "abstraction" has been used in computer science in a number of sub-disciplines for some time. In this context it refers to the amount of detail that is provided for objects. A jeep categorized differently from other objects in that it is a vehicle that is (1) small, (2) motorized, (3) open, and is (4) all-wheel drive. The evidence may be insufficient to identify the jeep from a small convertible, but for purposes of supporting one or more the decision support hypotheses it may be sufficient to use only properties (1) and (2). In exact terms this means that a more abstract object, "small motorized vehicle" exists and a jeep is also an instance of it. The Ontology shows all such classes and allows queries to be made at any level of abstraction. If this is done, then more evidence may "come in from the cold"

and be associated with the hypotheses with striking results, as shown in FIG. 4.

[0131] Sensor data coming into the system is directly mappable to a set of DIDs[mdash]data ID arrays. Each sensor has its own particular types of data that it gathers in its spectrum, and some sensors may output interpretations of what has been observed. This means that there are a number of different types of attributes that can be found. Let the input be a X-coordinate, a Y-coordinate and three values V1 through V3. These are then put in the "dimensional baskets", as shown in FIG. 5.

[0132] The input needs then to be connected with the type of decision that must be made. The objects that may be present are represented in the Ontology. The Ontology provides ranges that the variables V1 though V3 may have and be consistent with the presence of the particular object that is being defined. This is shown for Object 1 of three possible objects in FIG. 6. This will result in three different ranges for the variables V1 to V3 depending on what object they identify (e.g. truck, jeep, tank). This means that the table of grouped data can be pivoted. In FIG. 7 we see it graphically.

[0133] This becomes a table upon which a clustering process of association (e.g. data mining) can proceed: groups of data values for different objects in the same 2,3 or 4-dimensional area will be compared to see to what objects they correspond. The value of this approach is that it works hand in glove with the abstraction technique. The pivoted table provides a summary of all of the input data, transformed into evidence. It is then possible to dynamically adjust the level of abstraction to take advantage of what data exists. However, even when inadequate data is present this can be useful.

[0134] The level of abstraction of data available "small vehicle" may lack sufficient detail, i.e. is it motorized or not to be useful directly. However, the filtering mechanism may be invoked here as well. The filter would normally exclude a piece of evidence for a small vehicle for being considered a jeep or a car if it only had two out of four identifications. On the other hand, the user could have a hypothesis of a "worst case" scenario and assume that all of the "small vehicles" were jeeps just to see what threat would be possible in this case. This could be compared to the traditional approach where an object is not part of a threat until it is identified as one.

[0135] The system must allow the PRP user to change parameters for uncertainty via filters and for the objects to be fused via Ontology entries. Summarized, the requirements of such a system are that the user must be able to specify (1) a set of data sources, or a process such as a search engine query that supplies them. For each data source the user must specify (2) the processing filter to be used to include or omit the data. Each filter consists of transformations, including (3) the user supplied parameters for screening out any data with too high an uncertainty. The filter at the final data level should include (4) the last (higher) level of detail at which information is to be processed and (5) how frequently the data is to be updated. In addition the user must specify (5) the Final View's filter for combining the final sources of data. This section discusses the fusion of data using objects and the application of Evidential Reasoning, the use of an Ontology and screens. The PRP results are displayed to the user and changes like the above are made to explore the possible interpretations of the evidence.

[0136] The various embodiments and modifications of the present invention are not just those which have been heretofore described, but also all those within the scope of the following claim.

ENGLISH-CLAIMS:

Return to Top of Patent

I claim:

1. A process for reasoning about the real world' s representation as a database stored on a computer, comprising: organizing computer representations of information together with the representation of the presumed relationships that exist among elements of said information; applying and recording corresponding measures of completeness and correctness of said information; and generating a new measure that provides a useful new description of said information that summarizes it concisely.

LOAD-DATE: April 9, 2006