The Usage of Probability Theory Concepts for Analysing Big Data Uncertainty Reasoning

[1]J. Lysa Eben, Research Scholar, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai. India. lysaeben@rediffmail.com

[2]Dr R. Renuga Devi, Assistant Professor, Vels Institute of Science, Technology & Advanced Studies (VISTAS), Chennai. India. nicrdevi@gmail.com

ABSTRACT - Partial observability, nondeterminism or a combination of the two develop the problem of uncertainty a common occurrence in big data. An agent is needed to handle this uncertainty. This paper aims to see how an agent can tame uncertainty with the degree of belief and to design an agent program that implements the agent function, the mapping from percepts to actions, especially in the field of big data where volumes of data needs to be handled. The agent program more often takes the current percept as input from the sensors and return an action. Uncertainty arises because of ignorance or volumes of data; The agent's lack to express the truth of the event in the sentence due to uncertainty that prevails, which can be expressed using probability. Probabilities summarises the agent's belief relative to the evidence and probability distribution is used to specify the probability that exist in assigning to any random variables. Partial observability of the world brings in unobserved aspects, these can be resolved by estimating the values using probability, that help better agent decision in any field including big data. The agent program come into being through learning methods. An agent is designed to form representations of a complex world, the world with huge voluminous data, use a process of inference to derive new representations about the world, and use these new representations to deduce what to do.

Key words: Big Data, Partial observation, Uncertainty, degree of belief, Probability Distribution.

I.   INTRODUCTION

An agent is needed to handle the uncertainty that develop due to partial observation, missing information, outliers, etc. that persist in every field that generates tremendous volumes of data. Here the agent may never know for certain what state it is in or where it will end up after a sequence of action. One principle of a problem-solving agent solving the problem of partial observability is using the key concept of belief state, representing the agent's current belief about the possible physical states it might be in, given the sequence of actions and percept up to that point. On the other hand, a logical agent or knowledge-based agent considers ever logically possible explanation for the observation, no matter how unlikely in the world of big data. This leads to impossible large and complex belief-state representations. A correct contingent plan that handles every eventuality can grow arbitrarily large and must consider arbitrarily unlikely contingencies. Sometimes there is no plan that is grantee to achieve the goal, yet the agent must act. It must have some way to compare the merits of plans that are not guaranteed. Suppose, for example,[1] an automated taxi has the goal of delivering a passenger to the airport on time, the sensor for this task generates huge volumes of data and the data even

more grows when the time is increased. The agent forms a plan, A_{90}, that involves leaving home 90 minutes before the plane departs, driving at a reasonable speed and the airport is 5 miles away, a logical taxi agent will not be able to conclude with certainty that "plan A_{90} will get to the airport in time". Instead, it reaches the weaker conclusion that plan A_{90} will get to the airport in time, as long as the car doesn't breakdown or run out of gas and don't get into an accident and the plane doesn't leave early and etc. None of these conditions can be deduced for sure, so the plans success cannot be inferred. Specifying all these exceptions is called the qualification problem. There is no complete solution within logic, system designers have to use good judgment in deciding how detailed they want to be in specifying their model and what details they want to leave out. Other plans, such as A180, might increase the agent's belief that it will get to the airport on time but also increase the likelihood of a long wait. The performance measure includes getting to the airport in time for the flight, avoiding a long, unproductive wait at the airport, and avoiding speeding tickets along the way. The agent's knowledge cannot guarantee any of these outcomes for A90, but it can provide some degree of belief that they will be achieved. So, the relational decision therefore depends

---

on both the relative importance of various goals and the likelihood that, and degree to which, they will be achieved. These ideas pay way for the development of the general theories of uncertain reasoning and rational decisions. Consider an example of uncertain reasoning: diagnosing a dental patient toothache. Diagnosis almost always involves uncertainty and decision making. Using proportional logic, the rule for dental diagnosis[1] approach breaks down as,

Toothache => Cavity

But not all patients with toothache have cavities some of them have gum disease and abscess, or of several other problems. Here the need for decision making with all available data occurs, which increases the sample size and complexity. In order to make the rule true, addition of almost unlimited possible problems had to be done.

Toothache => Cavity ∨ Gum problem ∨ Abscess …..

On turning the rule more casual:

Cavity => Toothache

But this rule is not correct either, not all cavities cause pain. So, to fix the rule is to make it logically exhausted. That is to augment the left-hand side with all the qualification required for a cavity to cause a toothache. Use of logic to cope with the domain like medical diagnosis fails for three main reasons. One is laziness that is too much work to list the complete set of antecedents or consequence needed to ensure and exception less rule. Second is theoretical ignorance, in adequate theory in medical science for this domain. Third, practical ignorance, even if all the rules are known, uncertainty about a particular patient can run. This is typical for judgemental domains. The agent's knowledge can at best provide only a degree of belief in the relevant sentences. One main tool for dealing with degree of belief is probability theory. For a sentence a logical agent can take three states, namely 1 a true state or 0 a false state or a state with no opinion, whereas if the agent considers the same sentence probabilistically, then numerical range between 0 (for sentences that are certainly false) and 1 (certainly true) may be assigned for degree of belief. Probability provides a way of summarising the uncertainty that comes from laziness and ignorance, there by solving the qualification problem. Probability statements represents a knowledge state, not a statement in view to the real world. Real world big data problems can be addressed using probability theory This paper is organised as follows: section-1 Introduction to uncertainty, section-2 Literature review, section-3 Uncertainty and rational decision, section-4 Basic probability notations, section-4 Language of propositions in probability assertion, section-5 Probability axioms and Inference using full joint distribution, section-6 conclusion and future work.

II. LITERATURE REVIEW

Thomas Vandel, et al, [2] [3] [4] In quantifying uncertainty with Bayesian deep learning focus on uncertainty in climate projections with respect to modelling precipitation of a day. As climatic systems are complex and dynamic a model that withstand discrete – continues distributions be proposed, focus mainly lies on precipitation estimation. Uncertainty with respect to noise parameter and heteroscedastic issues are handled using a model proposed with its main concentration of statistical down scaling. [5] Another mode that can work on discrete-continues data model with log normal likelihood, support the use of mode for fat-tailed skewed distribution, that commonly seems to arise in these applications. Apart from proposing a discrete-continues model that uses Bayesian deep learning for uncertainty quantification, a comparison of predictive accuracy and uncertainty calibration was carried out to understand the ground truth observation. Fadi Ai Turjman et al. [6] In quantifying uncertainty in internet of medical things and big data services using intelligence and deep learning address waste calculate the price incurred for big data transfer services provided over the crowd in the presence of dynamic uncertainty factors with respect to network topology, data transfer rate, acceptance of the node and computational complexity. In the era of communication delivery data at a best phase is considered important, to enhance quality, energy consumption and performance of cloud service but this cannot be carried out without considering the price attribute as cost effectiveness place a major role in deciding the performance of the model. Uncertainty in utility devices and communication link in cloud based IOT infrastructure need to be addressed by investigating different approaches to have cost effective optimised pricing model. The performance of the proposed framework was compared against genetic algorithm and simulated annealing algorithm. The comparison was tested for distributed and centralized version. Low running time was achieved for distributed algorithm.

FeiFei Yang et al [7] [8] Address quantifying uncertainty in machine learning based power outage prediction model training. A tool for sustainable strong restoration addresses the problem faced in the united states due to extreme weather conditions which has its impact through interruptions in the delivery of power. An outage prediction model (OPM) is proposed which considers into account two resilience components. The study model considers factors like environmental conditions, electric layout, weather conditions. Regression tree model, boosted gradian tree, Bayesian addictive regression tree are some models proposed earlier to predict power outages. The proposed model takes into consideration the issue of uncertainty that can occur due to the sample size of the available training data and the issue related to low impact, high impact factors namely over estimation and

under estimation respectively. Quantifying the uncertainty of the OPM was evaluated using [9] [10] absolute error, mean absolute percentage error, R-squared and nash-sutcliffe efficiency on training sample of extra tropical and thunderstorm events. The study for OPM attained an acceptable predicted performance.

III.   UNCERTAINTY ISSUE AND RATIONAL DECISIONS

Data being huge and complex to handle, relation decisions support an agent to come to a solution with optimality. To make choices of considering the plan A_{90} or plan A_{180}[1], give rise to different probable outcomes, choice of the plan is based on the preference an agent make. An outcome from the plan is a fully definite state, including such aspects of whether the agent reaches the airport on time and the wait time incurred at the airport. Utility theory can be used to maximize the expected utility with preferences. Utility theory on an average says that every given probability has a degree of usefulness, or utility, and it is the agent likeliness to prefer states with higher utility. A rational agent must try to possess all Utility or preference with probabilities that will help decision making agent to handle uncertainty called decision theory[1].

Decision theory = probability theory + utility theory

The fundamental idea of decision theory is that any rational agent must behave as if and only if it chooses the action that will yield the highest expected utility, averaged over all the possible outcomes of the action. This is called the principle of maximum expected utility (MEU). To note "expected" means the "average," or "statistical mean" of the outcomes, weighted by the probability of the outcome. Figure 2.1 sketches the structure of an agent that uses decision theory to select actions. An agent's belief state represented in decision theory is their probabilities and not just the possibilities for world states. With all possible belief state, it is the agent choice to make probabilistic predictions of the action outcomes and therefore select the action with expected utility to maximise.
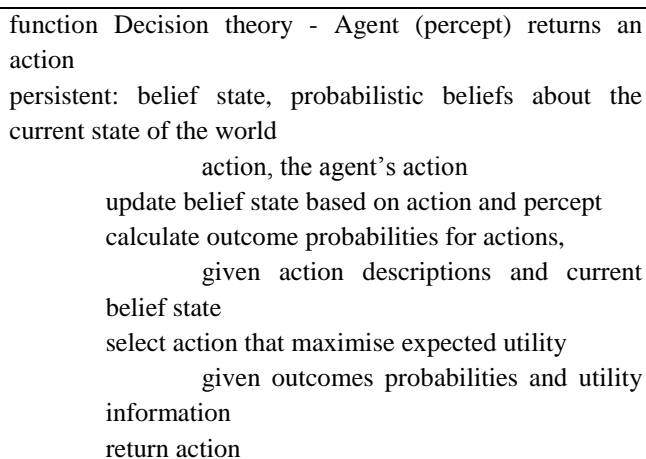
---

function Decision theory - Agent (percept) returns an action

persistent: belief state, probabilistic beliefs about the current state of the world

        action, the agent's action

     update belief state based on action and percept

     calculate outcome probabilities for actions,

        given action descriptions and current belief state

     select action that maximise expected utility

        given outcomes probabilities and utility information

     return action

Figure 2.1 A Decision Theory- Agent that selects rational actions.[1]

IV.   ELEMENTARY PROBABILITY NOTATION

Occurrence of an event, recording of the observations, representation of the problem includes probability notation. For an agent to represent and use probabilistic information,[1] a formal language is needed in data science. probabilistic assertions are about possible worlds. Probabilistic assertions talk about how probable the various worlds are. In probability theory, sample space is represented by set of all possible worlds. The possible worlds are mutually exclusive and exhaustive—two possible worlds cannot both be the case, and therefore cannot occur simultaneously. For example, if we are about to roll two (distinguishable) dice, there are 36 possible worlds to consider: (1,1), (1,2), ..., (6,6). The Greek letter Ω (uppercase omega) is used to refer to the sample space, and ω (lowercase omega) refers to elements of the space, that is, particular possible worlds. A fully specified probability model associates a numerical probability $P(\omega)$ with each possible world.1 The basic axioms of probability theory say that every possible world has a probability between 0 and 1 and that the total probability of the set of possible worlds is 1:

$$0 \leq P(\omega) \leq 1 \text{ for every } \omega \text{ and } \sum_{\omega \in \Omega} P(\omega)=1 \qquad (1)$$

Probabilistic assertions and queries are not usually about particular possible worlds, but about sets of them. For example, on rolling two dice the cases where the two dice add up to 11, the cases where doubles are rolled, and so on. In probability theory, these sets are called events,an event is a subset of a sample space, the sets are always described by propositions in a formal language. For each proposition, the corresponding set contains just those possible worlds in which the proposition holds. The probability associated with a proposition is defined to be the sum of the probabilities of the worlds in which it holds:

$$\text{For any proposition } \varphi, P(\varphi) = \sum_{\omega \in \varphi} P(\omega) \qquad (2)$$

Probabilities such as P (Total = 11) and P(doubles) are called unconditional or prior probabilities (and sometimes just "priors" for short); they refer to degrees of belief in propositions in the absence of any other information. Most of the time, however, some information, usually called evidence, already been revealed. For example, the first die may already be showing a 5 In this case, the unconditional probability of rolling doubles, but the conditional or posterior probability (or just "posterior" for short) of rolling doubles given that the first die is a 5. This probability is written P (doubles | Die1 = 5), where the " | " is pronounced "given." Mathematically speaking, conditional probabilities are defined in terms of unconditional probabilities as follows: for any propositions a and b, we have

$$P (a \mid b) = \frac{P(a \wedge b)}{P(b)} \quad (3)$$

which holds whenever P(b) > 0. For example,

$$P \text{ (doubles | Die1 = 5)} = \frac{P(doubles \wedge Die = 5)}{P(Die = 5)}$$

The definition makes sense if on remembering that observing b rules out all those possible worlds where b is false, leaving a set whose total probability is just P(b). Within that set, the a-worlds satisfy a ∧ b and constitute a fraction P(a ∧ b)/P(b). The definition of conditional probability, Equation (3), can be written in a different form called the product rule:

$$P \text{ (a} \wedge \text{b)} = P \text{ (a | b) P(b)}$$

It comes from the fact that, for a and b to be true, we need b to be true, and we also need a to be true given b.

V.   THE LANGUAGE OF PROPOSITIONS IN PROBABILITY ASSERTIONS

In this section, [1] propositions describing sets of possible worlds are written in a notation that combines elements of propositional logic and constraint satisfaction notation. A factored representation, in which a possible world is represented by a set of variable/value pairs. Variables in probability theory are called random variables and their names begin with an uppercase letter. Thus, in the dice example, Total and Die are random variables. Every random variable has a domain—the set of possible values it can take on. The domain of Total for two dice is the set {2,..., 12} and the domain of Die is {1,..., 6}.   A Boolean random variable has the domain {true, false}. By convention, propositions of the form A = true are abbreviated simply as a, while A = false is abbreviated as ¬a. the domain of Weather might be {sunny, rain, cloudy, snow}. When no ambiguity is possible, it is common to use a value by itself to stand for the proposition that a particular variable has; thus, sunny can stand for Weather = sunny. Variables can have infinite domains, too—either discrete (like the integers) or continuous (like the reals). For any variable with an ordered domain, inequalities are also allowed, such as Number of Atoms in Universe ≥ 10^{70}. Finally, these sorts of elementary propositions (including the abbreviated forms for Boolean variables) can be combined by using the connectives of propositional logic. For example, we can express "The probability that the patient has a cavity, given that she is a teenager with no toothache, is 0.1" as follows [1]:

$$P \text{ (cavity | ¬toothache} \wedge \text{teen)} = 0.1$$

When talking about the probabilities of all the possible values of a random variable. It can be written as:

P (Weather = sunny) = 0.6
P (Weather = rain) = 0.1
P (Weather = cloudy) = 0.29
P (Weather = snow) = 0.01

But when abbreviated

P (Weather) = {0.6, 0.1, 0.29, 0.01},

where the bold P indicates that the result is a vector of numbers, and assumes a predefined ordering {sunny, rain, cloudy, snow} on the domain of Weather. Here P statement defines a probability distribution for the random variable Weather. The P notation is also used for conditional distributions: P (X | Y) gives the values of P (X = xi | Y = yj) for each possible i, j pair. For continuous variables, it is not possible to write out the entire distribution as a vector, because there are infinitely many values. Instead, the probability that a random variable takes on some value x as a parameterized function of x can be defined. For example, the sentence P (Noon Temp = x) = Uniform[18C,26C] (x)

expresses the belief that the temperature at noon is distributed uniformly between 18 and 26 degrees Celsius. This is called a probability density function. In addition to distributions on single variables, notation is needed for distributions on multiple variables. Commas are used for this. For example, P (Weather, Cavity) denotes the probabilities of all combinations of the values of Weather and Cavity. This is a 4 × 2 table of probabilities called the joint probability distribution of Weather and Cavity. Variables can also be mixed with and without values; P (sunny, Cavity) would be a two-element vector giving the probabilities of a sunny day with a cavity and a sunny day with no cavity. The P notation makes certain expressions much more concise than they might otherwise be. For example, the product rules for all possible values of Weather and Cavity can be written as a single equation [1]:

P (Weather, Cavity) = P (Weather | Cavity) P (Cavity)

instead of as these 4 × 2=8 equations (using abbreviations W and C):

P (W = sunny ∧ C = true) = P (W = sunny | C = true) P (C = true)
P (W = rain ∧ C = true) = P (W = rain | C = true) P (C = true)
P (W = cloudy ∧ C = true) = P (W = cloudy | C = true) P (C = true)
P (W = snow ∧ C = true) = P (W = snow | C = true) P (C = true)
P (W = sunny ∧ C = false) = P (W = sunny | C = false) P (C = false)
P (W = rain ∧ C = false) = P (W = rain | C = false) P (C = false)
P (W = cloudy ∧ C = false) = P (W = cloudy | C = false) P (C = false)
P (W = snow ∧ C = false) = P (W = snow | C = false) P (C = false)

Now the syntax for propositions and probability assertions is defined and semantics given. To complete the semantics, it is needed to say what the worlds are and how to determine whether a proposition holds in a world. This part is taken directly from the semantics of propositional logic, as follows. A possible world is defined to be an assignment of values to all of the random variables under consideration. It is easy to see that this definition satisfies the basic requirement that possible worlds be mutually exclusive and exhaustive. For example, if the random variables are Cavity, Toothache, and Weather, then there are 2 × 2 × 4 = 16 possible worlds. Furthermore, the truth

of any given proposition, no matter how complex, can be determined easily in such worlds using the same recursive definition of truth as for formulas in propositional logic. From the preceding definition of possible worlds, it follows that a probability model is completely determined by the joint distribution for all of the random variables— the so-called full joint probability distribution. For example, if the variables are Cavity, Toothache, and Weather, then the full joint distribution is given by P(Cavity, Toothache, Weather). This joint distribution can be represented as a $2 \times 2 \times 4$ table with 16 entries. Because every proposition's probability is a sum over possible worlds, a full joint distribution suffices, in principle, for calculating the probability of any proposition.

VI.   PROBABILITY AXIOMS AND INFERENCE USING FULL JOINT DISTRIBUTION

The basic axioms of probability (Equations (1) and (2)) imply certain relationships among the degrees of belief that can be accorded to logically related propositions. [1] For example, we can derive the familiar relationship between the probability of a proposition and the probability of its negation:

$$P(\neg a) = \omega \in \neg a\ P(\omega) \qquad\qquad \text{by Equation (2)}$$
$$= \omega \in \neg a\ P(\omega) + \omega \in a\ P(\omega) - \omega \in a\ P(\omega)$$
$$= \omega \in \Omega\ P(\omega) - \omega \in a\ P(\omega)\ \text{grouping the first two terms}$$
$$= 1 - P(a) \qquad\qquad \text{by (1) and (2).}$$

One argument for the axioms of probability is as follows; If an agent has some degree of belief in a proportion a, then the agent should be able to state odds at which it is indifferent to a bet for or against a. probabilistic inference ie the computation of posterior probabilities for query proportions given observed evidence. The full joint distribution is used as the knowledge base from which answers to all questions may be derived. On examining with a simple example, a domain consisting of just three Boolean variable toothache, cavity and catch. The full disjoint distribution is a 2x2x2 table as shown in the figure [1] below.

	Toothache		¬ Toothache	
	Catch	¬ Catch	Catch	¬ Catch
Cavity	0.108	0.012	0.072	0.008
¬ Cavity	0.016	0.064	0.144	0.576

The probabilities in the full joint distribution sum to 1 as required by the axioms of probability. The equation (2) gives us a direct way to calculate the probability of any proposition simple or complex, simply identify those possible worlds in which the proposition is true and add up their probabilities. The above example is explained for three Boolean variables and can be implemented for any Big Data problems

VII.   CONCLUSION AND FUTURE WORK

Big Data uncertainty problem can be approached and solved with Probability theory. This paper suggests probability theory as a suitable foundation for uncertain reasoning and an introduction to its use. Uncertainty arises due to partial observable environments and when data happens to be voluminous. Probabilities summarise the agent's belief relative to the evidence. Decision theory combines the agent's belief and desires defining the best action as the one that maximises expected utility. Basic probability statement includes prior probabilities and conditional probabilities over simple and complex propositions. The axioms of probability constrain the possible assignments of probabilities to propositions. Probability is a rigorous formalism for uncertain knowledge Joint probability distribution specifies probability of every atomic event and queries can be answered by inference Rules allow unknown probabilities to be computed form known conditional probabilities. Quantifying uncertainty can be applied in machine learning, deep learning, Internet of Things, fields of climatic research, medical diagnosis and many decision seeking problems.

REFERENCES

[1] Russell, S., & Norvig, P. (2005). AI a modern approach. Learning, 2(3), 4.

[2] Vandal, T., Kodra, E., Dy, J., Ganguly, S., Nemani, R., & Ganguly, A. R. (2018, July). Quantifying uncertainty in discrete-continuous and skewed data with Bayesian deep learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (pp. 2377-2386).

[3] Dong, C., Loy, C. C., He, K., & Tang, X. (2014, September). Learning a deep convolutional network for image super-resolution. In European conference on computer vision (pp. 184-199). Springer, Cham.

[4] Gal, Y. (2016). Uncertainty in deep learning. University of Cambridge, 1, 3.

[5] Katz, R. W. (2002). Techniques for estimating uncertainty in climate change scenarios and impact studies. Climate research, 20(2), 167-185.

[6] Al-Turjman, F., Zahmatkesh, H., & Mostarda, L. (2019). Quantifying uncertainty in internet of medical things and big-data services using intelligence and deep learning. IEEE Access, 7, 115749-115759.

[7] Wang, G., Li, W., Zuluaga, M. A., Pratt, R., Patel, P. A., Aertsen, M., ... & Vercauteren, T. (2018). Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE transactions on medical imaging, 37(7), 1562-1573.

[8] Yang, F., Wanik, D. W., Cerrai, D., Bhuiyan, M. A. E., & Anagnostou, E. N. (2020). Quantifying Uncertainty in Machine Learning-Based Power Outage Prediction Model Training: A Tool for Sustainable Storm Restoration. Sustainability, 12(4), 1525.

[9] Tang, X. S., Li, D. Q., Cao, Z. J., & Phoon, K. K. (2017). Impact of sample size on geotechnical probabilistic model identification. Computers and Geotechnics, 87, 229-240.

[10] He, J., Wanik, D. W., Hartman, B. M., Anagnostou, E. N., Astitha, M., & Frediani, M. E. (2017). Nonparametric tree- based predictive modeling of storm outages on an electric distribution network. Risk Analysis, 37(3), 441-458.