Inductive Reasoning and Chance Discovery*

AHMED Y. TAWFIK

School of Computer Science, University of Windsor, Windsor, Canada ON N9B 3P4; E-mail: atawfik@uwindsor.ca

Abstract. This paper argues that chance (risk or opportunity) discovery is challenging, from a reasoning point of view, because it represents a dilemma for inductive reasoning. Chance discovery shares many features with the grue paradox. Consequently, Bayesian approaches represent a potential solution. The Bayesian solution evaluates alternative models generated using a temporal logic planner to manage the chance. Surprise indices are used in monitoring the conformity of the real world and the assessed probabilities. Game theoretic approaches are proposed to deal with multi-agent interaction in chance management.

Key words: Bayesian confirmation, chance discovery, inductive reasoning

1. Introduction

A chance (opportunity or risk) can be characterized as a high impact event, situation or change. Typically, these situations are rare but their effects (payoff or loss) are so significant that it is advantageous to discover them as early as possible to try to avert the risks and exploit the opportunities (Ohsawa, 2001). The term chance discovery has been coined to refer to the process of discovering such situations using automated reasoning. From a reasoning perspective, chance discovery differs from knowledge discovery. Knowledge discovery extracts common patterns from data while chance discovery predicts future outcomes. For example, forecasting the market potential for a new product represents a form of chance discovery.

To illustrate how difficult and how far off human may be in discovering chance, consider the following examples: In 1943, Thomas Watson, then chairman of IBM Corporation, predicted a world market for about five computers. In 1970, Ken Olsen, founder of Digital Equipment Corporation is reported to have said that no one needed to have a personal computer at home. In both cases, it was difficult to assess the opportunities because traditionally, forecasting has relied on extrapolation. Extrapolation is a form of inductive reasoning that assumes that current trends would carry on into the future. Clearly, this approach does not work well with new types of products. This problem is not unique to extrapolation as a procedure but it is

inherently a problem in inductive reasoning that is closely related to Goodman's new riddle of induction (Goodman, 1955).

However, solving the problem of induction does not completely solve the chance discovery problem. The representational challenge is also as significant. Generally, our knowledge representations suffer from functional fixation, thus, hiding potential opportunities and risks. Thomas Watson's forecast implicitly assumes some function for computers that is rather limited. Certainly, in 1943, the range of computer applications imagined was rather limited. The same is true for personal computers in the 1970's. This work argues that finding the proper knowledge representation is of great importance for chance discovery.

Knowledge representation and reasoning frameworks typically favor 'the normal' and the 'common' to the 'rare' or 'exceptional'. Consequently, conventional frameworks will likely miss rare situations. Moreover, for these rare situations, it is necessary to distinguish cases that represent opportunities or risks from other rare changes. To identify these situations a decision theoretic approach for assessing such rare situations is needed.

The paper is organized as follows: Section 2 shows that chance discovery is a practical example of grue. As such, proposed solutions to the grue paradox are surveyed for clues that may help with chance discovery. The Bayesian approach seems to hold some promise. Section 3 discusses the use of entropy maximization to come up with probabilities. Section 4 presents a technique for managing chance. Section 5 suggests the need for chance monitoring. Section 6 discusses the impact of intent on probability and utility assessment.

2. Goodman's Riddle of Induction

The color of an emerald is grue (green then blue) if it is and has always been observed green until some future time (say year 2222) when it will turn blue. This notion presents a paradox to inductive reasoning because our observations support the statement that emeralds are green as well as the claim that they are grue (Goodman, 1955). This paradox, first proposed in 1955, has inspired arguments about the validity of induction. The essence of the problem lies in the inductive temporal uniformity assumption that implies that the future will look like the past. Many have contended that a correct solution would justify preferring green emeralds to grue ones. In the context of chance discovery, the correct solution would be one that minimizes grue predictions (i.e. by maximizing temporal uniformity) without missing any cases of grue (i.e. when temporal uniformity does not apply).

It may be necessary to first demonstrate that certain properties are really grue (Akeroyd, 1991). The statement that gold is soluble is grue (or falue – for false then true). For a long time in history, observations supported the

notion that gold cannot be dissolved until the invention of regal water. Scientific discoveries have generally challenged human conception of the universe that has proved to be grue in some ways. Grue phenomena challenge the notion of temporal uniformity. In this treatment, a phenomenon is grue if it involves an unexpected change. For example, the discovery of a treatment for a previously untreatable disease is grue. Similarly, we consider the eruption of a volcano or a strong earthquake in a historically stable area as grue phenomena. These phenomena present a new challenge to automated reasoning. Therefore, chance discovery is to some extent about discovering when our experiences mislead us. It is about identifying situations when grue is true.

Favoring green over grue for its simplicity is of limited relevance for two reasons. First, the definitions of green and blue in terms of grue and bleen (blue then green) are as simple as those of grue and bleen in terms of green and blue. In other words, measuring complexity by comparing message length can be an artifact of the representation. Therefore, trying to minimize message length does not necessarily resolve the problem. Minimum message length induction approximates a full Bayesian inference over the entire hypothesis space (Solomonoff, 1999). A better approximation is obtained if we use more terms corresponding to short codes. In chance discovery, whether the objective is to identify a risk or an opportunity, it is necessary to consider the more complex scenarios as long as they are possible no matter what measure of complexity is appropriate. Notions of simplicity such as Ockham's Razor would always miss some chances.

Preferring persistence (as in green) to change (as in grue) corresponds to the common sense law of inertia (McCarthy and Hayes, 1969). The commonsense law of inertia can be considered as a nonmonotonic circumscriptive assumption that minimizes change (Shanahan, 1997). This approach does not capture correctly many practical situations involving indigenous change or partially observable systems (Dean and Kanazawa, 1989). Moreover, chances will necessarily be missed. The commonsense law of inertia would prefer to assume that a volcano will not erupt, that a new product will not sell and in general that different new conditions will not occur.

Bayesian confirmation theory evaluates the probabilities of green and grue at future times based on prior and conditional probabilities incorporating evidence and background information (Horwich, 1982). As such the degree of belief in a particular outcome can be calculated provided an accurate theory exists to assess the probabilities and the causal dependencies. As such, Bayesian confirmation theory extends the use of probabilities as a measure of belief beyond frequency interpretation to include other interpretations such as subjective probabilities and propensities.

The Bayesian approach is to explicitly list all possibilities including all alternative models (all possible worlds), assess priors and conditional

probabilities, and calculate posterior probabilities given all available observations for the different models under consideration. Therefore, the probability of a statement (or conclusion) S given some evidence E is given by

$$P(S|E) = \frac{P(S)P(E|S)}{\sum_{j} P(E|h_j)P(h_j)}$$

where the h_j's represent all possible models consistent with E.

As chance discovery is the other side of the coin, a Bayesian approach might help. Consider for example that S represents the occurrence of a strong earthquake, and that E consists of a history of seismic activity in the region including small earthquakes. A number of competing theories h_i's are consistent with observations but differ in future predictions. The probability of a strong earthquake can be derived provided some prior and conditional probabilities. Typically, the number of possible theories or models can be very large. This large number of possible models in any practical situation presents a challenge to Bayesian chance discovery.

If Thomas Watson were to apply the above approach to assess the market for computers, he would have to consider a myriad of possible models including the one that actually happened. In hindsight, we know that the computer market became strong because computer prices went steadily down, performance increased exponentially and applications have been developed to fulfill a wide range of needs. This particular scenario was not very likely in 1943. Having been involved in the efforts to build some of the very first computers, Thomas Watson might have not seen the information revolution coming. Considering the complexity of using the early machines, their poor reliability, high prices, and limited performance as evidence, he predicted a very limited market for these machines. However, it is fair to assume that the need for computation in domains such as accounting, engineering design, banking, and planning, was evident even then. There may have been some speculations (theories) about ways to reduce prices and increase performance.

Therefore, to discover a chance, it is important to incorporate relevant evidence within possible models. This requirement adds additional challenges to the challenge of coming up with prior and conditional probabilities usually encountered in Bayesian approaches.

3. Finding the Probabilities

A knowledge representation suitable for chance discovery has to be able to concisely encode a possibly very large number of models (possible worlds). To achieve this representation efficiency, models can be grouped in

equivalence classes or categories such that a group of models belonging to the same class K share the same $P(S, E|h_k)$ for all h_k in K. The formation of these equivalence classes can be based on the structural similarities of the models or on the propensity of the models to support particular evidence (Bacchus et al., 1996). For example, the probability of increased sales of computers in the future may be the same according to a theory projecting future educational applications and another projecting more business applications. In this case, both theories belong to the same equivalence class.

In Equation 1 above, $P(S|E)$ is inversely proportional to $\Sigma_i P(E|h_i)P(h_i)$. The other terms in the expression do not depend on the particular model, and are constants for any particular combination of any statement S and evidence E. However, the choice of priors $P(h_i)$ and conditional probability distributions $P(E|h_i)$ has to reflect the background knowledge or the lack thereof (ignorance). Choosing the values that maximize entropy reflects ignorance (Jaynes, 1968). This information theoretic approach to the determination of priors has an advantage over other subjective approaches in the case of ignorance.However, entropy maximization, like many other probabilistic inference procedures, is representation dependent (Halpern and Kollar, 2004). It appears that all non-trivial probabilistic inference procedures are representation dependent to some extent. The entropy is given by

$$\text{Entropy} = -\sum_i P(h_i) \log P(h_i)$$

In the context of chance discovery, the determination of conditional probabilities for $P(E|h_i)$ has to rely on a background theory. Edis (2000) suggests that, in the absence of any background theory, the evidence Esupports all competing models equally as long as they are consistent with it. In other words, if we do not have a background theories allowing us to prefer the model of green emeralds, all observations of green emeralds also support the model of grue emeralds to the same extent. This results in equal weights for all alternatives. For example, in the absence of any information to guide the assessment of probabilities, scenarios representing computers becoming more expensive are as likely as scenarios representing computers getting cheaper and so on.

4. Model Formation

The treatment so far assumed the availability of three elements: a chance statement S, some related evidence E, and a set of models $\{h_1, h_2, \ldots, h_N\}$. All three elements are hardly ever readily available in a chance discovery context.

Formally, the chance discovery problem can be represented by a Kripke structure (Kripke, 1963). The structure $M = (W, \Phi, \pi, R)$ represents a chance

discovery Kripke structure. W denotes a set of worlds. Each world is described using truth assignment π defined for a set of propositions Φ. An accessibility relation R determines the set of worlds reachable from a particular world. Each world w occurs with a probability $\mu(w)$. The probability of a proposition ϕ is given by

$$P(\varphi) = \sum_{w|=\varphi} \mu(w)$$

A chronicle C is a path between a start world w_0 and a final world w_f such that for any two consecutive worlds along the path w_i and w_j, w_i follows w_j if and only if $w_j \in R(w_i)$. The evidence set E is a subset of Φ such that E holds in a temporally constrained set of worlds in all chronicles. Depending on the nature of E and W, temporal constraints may be ordering constraints over points and intervals or clock constraints. The chance states S constitute another subset of Φ such that a utility function U when applied to $s \in S$ in some world(s) $w \in W$ results in a significant chance. In chance discovery, unlike in decision theory, utilities are assessed for worlds irrespective of their probabilities in order to detect rare chances. Each accessibility relation $r \in R$ between a pair of worlds encodes a set of assumptions, actions, or events. A model h is the conjunction of all the assumptions actions or events along a chronicle $c \in C$.

This deductive approach to chance discovery has some complexity and feasibility limitations. In realistic domains, it is difficult to encode all possible combinations of events, actions, and assumptions as well as all their consequences. Moreover, as chances are rare, a chance discovery procedure that expands all future worlds would waste a tremendous amount of computational resources, seldom discovering chances. However, this last observation suggests that backward chaining is a more efficient solution if there is a proper characterization of risks and opportunities. Accordingly, the chance discovery process proceeds from a chance statement S. Similarly, McBurney and Parsons (2001) start with a statement and proceed with building a chance discovery dialogue between collaborative agents. The Bayesian analogue to the backward reasoning approach is to consider the probability of the model given the evidence E and the chance S.

$$P(h|S, E) = \frac{P(S, E|h)P(h)}{P(S, E)}$$

The purpose of the above equation is to measure if there is a model that explains both E and S well. The model h that maximizes the probability in Equation (4) above is the model (sequence of actions) to follow or prevent most depending on our interpretation of S as opportunity or risk respectively.

Thus far, the development does not provide any insights into how to enumerate the models compatible with given S and E. Assuming that the

number of models satisfying the constraints imposed by S and E is finite, the list $h_1, h_2, \ldots, h_N$ represents these models (or equivalence classes of such models). Recent advances in planning (Bacchus and Kabanza, 2000), allow us to build efficient backward chaining planners that guide their search by incorporating temporal domain dependent constraints. Each plan thus generated corresponds to a model.

The use of planning for model formation leaves one last challenge: how to enumerate possible actions, events, and assumptions? This is a knowledge representation issue. The challenge stems from the fact that many threats and opportunities result from unusual or innovative use or interaction of previously defined objects. Largely, our knowledge representations suffer from functional fixation and it is imperative that the actions, events, and assumptions available be as general as possible. One way to achieve this generality is using object hierarchies and generic actions. For example, a saltshaker ought to be defined as a rigid container that has holes and that may contain salt. Actions that may involve the saltshaker include those applicable to rigid objects such as using it to drive a nail, as well as actions for containers (e.g. filling it) and those specific to saltshakers (e.g. pouring salt).

This approach may discover many common situations such as crossing a street or driving to work as risks and other daily occurrences may be labeled as opportunities. In risk and opportunity discovery, we have to assume that such common risks and opportunities have already been addressed and the focus is on discovering unusual and rare risks and opportunities. Here, the expedient of setting a threshold on $P(S)$ to qualify as a chance avoids this problem at the risk of missing some chances.

5. Chance Monitoring

As the Bayesian approach to chance discovery relies on subjective probability assessment, it is important to monitor the chance exploitation plan execution to verify that these assessments conform to reality. Moreover, our chance discovery algorithm may still miss some chances because of an inadequate representation or limited observations. Both of these concerns can be addressed using chance monitoring. Given some observations during the execution of the chance exploitation plan, how to identify new chances that may come about or detect deviations preventing proper chance exploitation? For example, how to tell if a particular technology is no longer a good investment opportunity or if another new one is promising?

Chance monitoring relies on surprise measures to detect deviations between expected behavior and observed behavior. From a Bayesian perspective, all alternative scenarios must be considered and the probabilities are used to determine potential risks and opportunities. This approach is

methodologically sound but of limited practical value because it is nearly impossible to enumerate all alternatives in a given situation. A considerable simplification results from enumerating some common subset of alternative scenarios and using a surprise measure to detect other situations. Surprise measures reflect the degree of incompatibility of observed data and enumerated models (Bayarri and Berger, 1997).

A particular observation is surprising if its probability is small in comparison with the probability of other possible results (Weaver, 1948). The occurrence of an event such as 'someone won the lottery' is not surprising despite its small probability because all other alternatives are equally unlikely. However, the person who wins the lottery would be surprised because the alternative event (i.e. not winning) is much more probable. Weaver (1948) uses the ratio of the expected probability value to the probability of the observed event as a surprise measure.

$$\text{Surprise} = \frac{\sum_i P^2(x_i)}{P(x_{\text{Obs}})}$$

The numeric value of this surprise index is less than 1 as long as the more likely event takes place. It gets higher the less likely the event is compared to the alternatives. Other surprise indices that differ in their generality, mathematical properties, and ease of use have been proposed[1] including logarithmic forms (Good, 1983).

In the context of chance monitoring, the probabilities used for evaluating the surprise index are model probabilities. The frequent occurrence of surprising events signals model inadequacy. By adding the new surprising observations as evidence and revisiting the model selection stage, it is possible to adjust the probabilities as well as choosing a better plan.

6. Intent-Based Chance Management

The chance discovery software agent relies on utilities to assess the desirability of a situation. These utilities express a form of intent. The chance discovery process typically involves interactions between many intelligent entities with converging or diverging intentions. These intentions guide responses to challenges and opportunities forcing a defined structure when a random response would be expected. For example, knowing that a business aims at making profits implies that it will not try to exploit a chance in such a way to maximize its losses. Moreover, chronicles inconsistent with the intention are highly unlikely. The bias introduced by intentions can be very significant. For example, Schelling (1960) reports that a group of individuals

asked to choose head or tail with the intent of matching the choice of another unknown person, have predominately (86%) chosen head. A preponderance of participants (90%), asked to divide 100 with a rival to increase the chance of a match, have divided the sum equally. These observations clearly suggest that intentions can strongly bias otherwise random choices. Incorporating motives and intentions within the proposed framework would improve the probability assessments. By ascribing motives to agents, the behavior of these agents becomes more deterministic. Multi-agent systems that exploit this phenomenon, also known as focal points, can achieve a certain level of coordination without communications (Fenster et al., 1995).

The perception of value depends on the intentions of the agent. Some agent's trash could be another's treasure. The utilities or values assignments depend to a great extent on the intent of an agent. In a chance management process involving humans and intelligent agents, a proper formulation would require a game theoretic framework to account for multi-agent conflict, coordination, and cooperation. The interaction between two agents may involve strict competition, strict cooperation, or a combination of cooperation, coordination, and a degree of conflict. The latter results in a non-zero sum game (Schelling, 1960). In such situations, proper chance management requires building on common interests and resolving conflicts. Humans tend to accept conflict resolution compatible with the notion of focal points. These focal points are zones within a solution space that possess some specially appealing features like uniqueness, symmetry, simplicity, or precedence.

Formally, a game theoretic (Osborne and Rubinstein, 1994) formulation would describes the interaction of a set of agents with a Markov environment in which they all receive some payoffs for reaching their intended goals. The game consists of a tuple $\langle S, \pi, G, T, r \rangle$ where S is a discrete state space that corresponds roughly to the set of world states in Section 4; π is a probability distribution over the initial state; G is a collection of agents, each described by defining three sets: A, O, B where A is a discrete action space, O is a discrete observation space, and B is a set of mappings from the observation space to a probability distribution denoting the world state corresponding to an observation; T is a mapping from states of the environment and actions of the agents to probability distributions over states of the environment.; and r : $S \times A_G \rightarrow R$ is the payoff function, where A_G is the joint action space of the agents. Upon observing an observation o, an agent would try to deduce the corresponding world state, and act according to a strategy. The objective of each agent's strategy is to maximize its reward. Algorithm for optimal and suboptimal strategy development has been proposed (Boutilier, 1999; Peshkin et al., 2000). While, this game theoretic formulation is very similar to the techniques described in previous sections, it is a necessary extension to account for multi-agency in chance discovery contexts.

7. Conclusion

The chance discovery process described here relies on a background theory to build a plan for managing and exploiting chances. By combining abductive and deductive reasoning, the Bayesian treatment of chance discovery complements the Bayesian solution to the Goodman's riddle of induction. The role of probability in chance management is fundamental. Attempts to reap the rewards of discovery plans are not likely to be productive unless the chance exploitation plan can reasonably improve the probability of opportunities and reduce that of risks. A game theoretic approach may be necessary to exploit chances in a multi-agent environment.

Note

[1] Consult Bayarri and Berger (1997) for a survey.

References

Brown, J.S. and Burton, R.R. (1978), 'Diagnostic Models for Procedural Bugs in Basic Mathematical Skills', *Cognitive Science* 2(2), pp. 155–192.

Akeroyd, A. (1991), 'A Practical Example of Grue', *British Journal for the Philosophy of Science* 42(4), pp. 535–539.

Bacchus, F. Grove, A., Halpern, J. and Koller, D. (1996), 'From Statistical Knowledge to Degrees of Belief', *Artificial Intelligence* 87, pp. 75–143.

Bacchus, F. and Kabanza, F. (2000), 'Using Temporal Logics to Express Search Control Knowledge for Planning', *Artificial Intelligence* 116(1–2), pp. 123–191.

Bayarri, M. and Berger, J. (1997), 'Measures of Surprise in Bayesian Analysis', Duke University Institute of Statistics and Decision Sciences Working Paper No. 97-46, Durham, North Carolina.

Boutilier, C. (1999), 'Sequential Optimality and Coordination in Multiagent Systems', in *Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence*.

Dean, T. and Kanazawa, K. (1989), 'A Model for Reasoning about Persistence and Causation', *Computational Intelligence* 5(3), pp. 142–150.

Edis, T. (2000), 'Resolving Goodman's Paradox: How to Defuse Inductive Skepticism', Unpublished Manuscript. http://www2.truman.edu/~edis/.

Fenster, M., Kraus, S. and Rusenschein, J. (1995), 'Coordination without Communications: Experimental Validation of Focal Point Techniques', in *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, San Francisco, CA.

Good, I.J. (1983), *Good Thinking: The Foundation of Probability and its Applications*, Minneapolis: University of Minnesota Press.

Halpern, J. and Koller, D. (2004), 'Representation Dependence in Probabilistic Inference', *Journal of Artificial Intelligence Research* 21, pp. 319–356.

Horwich, P. (1982), *Probability and Evidence*, Cambridge: Cambridge University Press.

Jaynes, E.T. (1968), 'Prior Probabilities', *IEEE Transactions on System Science and Cybernetics* 4, pp. 227–241.

Kripke, S. (1963), 'Semantic Analysis of Modal Logic I: Normal Modal Propositional Calculi', *Zeitschrift f. Math. Logik und Grunlagen d. Math.* 9, pp. 67–96.

McBurney, P. (2001), 'Review of: First International Workshop on Chance Discovery', *Knowledge Engineering Review* 16(2), pp. 215–218.

McBurney, P. and Parsons, S. (2001), 'Chance Discovery Using Dialectical Argumentation', in Y. Ohsawa, ed., *Proceedings of the First International Workshop on Chance Discovery*, Matsue, Japan, pp. 37–45.

McCarthy, J. and Hayes, P. (1969), 'Some Philosophical Problems from the Standpoint of Artificial Intelligence', *Machine Intelligence* 4, pp. 463–502.

Goodman, N. (1955), 'The New Riddle of Induction', *Fact, Fiction and Forecast*, Cambridge, MA: Harvard University Press.

Ohsawa, Y. (ed.) (2001), *Proceedings of the First International Workshop on Chance Discovery* Matsue, Japan: Japanese Society for Artificial Intelligence.

Osborne, M. and Rubinstein, A. (1994), *A Course in Game Theory* Cambridge, MA: MIT Press.

Peshkin, L., Kim, K., Meuleau, M. and Kaelbling, L. (2000), 'Learning to Cooperate via Policy Search', *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*.

Prendinger, H. and Ishizuka, M. (2001), 'Some Methodological Considerations on Chance Discovery', in Y. Ohsawa, ed., *Proceedings of the First International Workshop on Chance Discovery*, Matsue, Japan, pp. 1–4.

Schelling, T. (1960), *A Strategy of Conflict*, Cambridge, MA: Harvard University Press.

Shanahan, M. (1997), *Solving the Frame Problem*, Cambridge, MA: MIT Press.

Solomonoff, R. (1999), 'Two Kinds of Probabilistic Induction', *The Computer Journal* 42(4), pp. 251–259.

Weaver, W. (1948), 'Probability, Rarity, Interest and Surprise', *Scientific Monthly* 67, pp. 390–392.