Some Methodological Considerations on Chance Discovery

Helmut Prendinger and **Mitsuru Ishizuka**
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
E-mail: {helmut,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract

This paper investigates the methodological foundations of a new research field called *chance discovery*, which aims to detect future opportunities and risks. By drawing on concepts from cybernetics and system theory, it is argued that chance discovery best applies to *open systems* that are equipped with regulatory mechanisms to approximate an 'ideal' state. We will motivate *anticipation* as an additional (regulatory) mechanism that 'creates' new future alternatives for open systems by human initiative. Our framework is applied to chance discovery in enterprises and scientific research programmes.

Introduction

Several researchers within the Knowledge Discovery in Databases (KDD) community (e.g., (Ohsawa 2000)) questioned whether the methods of this research field are able to find what they call 'chances'. Chances refer to phenomena that will have a (high) impact to the scientific (and human) society or an enterprise in the future. High impact is intended to have two complementary readings: on the one hand it refers to *opportunities*, i.e., the possibility to bring about desirable effects; on the other it refers to *risks*, i.e., possible threats to an enterprise or society. The notion of *chance discovery* has been coined to cover both aspects. Finding future features is seen in contrast to prediction (e.g., in KDD), the scientific activity to derive phenomena that appear at some future time point. By contrast, chance discovery explicitly integrates human initiative into the discovery process. Although there might be some interesting interactions with the *probabilistic* notion of chance, this reading is not intended in chance discovery. Likewise, chance discovery is not concerned with discovery *by* chance, such as the discovery and isolation of penicillin by Alexander Fleming.

We will discuss the following topics. In the following section, the notion of *open system* is explicated in terms of cybernetics and system theory, and the possibility of prediction is discussed for both nature and open systems. The next section discusses chance discovery in open systems. In particular, the notion of 'anticipation' is introduced as a mechanism for chance discovery and exemplified by examples. After that, we explicate notions underlying the possibility of chance discovery: uncertainty and freedom. In the following section, chance discovery is contrasted with KDD. Finally, we briefly discuss and conclude the paper.

Nature vs. Open Systems

To clarify the application field of chance discovery, we draw a broad distinction about the object of investigation: nature vs. open systems (Schurz 1999). Whereas nature is governed by natural laws, open systems are typically modeled abstractly by cybernetics (Ashby 1964) and system theory (v. Bertalanffy 1979). Examples of open systems include 'living' systems such as human beings, scientific communities and companies, and artificial (or technical) systems, e.g., cars and power plants. Both kinds can be described by the following system-theoretical ($S1-2$) and cybernetical ($C1-2$) features (Schurz 1999):

- $S1$ Open systems are physical ensembles placed into an environment significantly larger than themselves. There is a continuous exchange of energy between system and environment. The environment may satisfy the system's 'needs' (see $C1$) or 'destroy' the system (see $C2$).
- $S2$ Open systems preserve a relative identity through time, called their *dissipative* state.
- $C1$ The identity in time is abstractly governed by *ideal states* (or norm states) which the system tries to approximate, given its actual state.
- $C2$ Regulatory mechanisms compensate disturbing influences of the environment, i.e., they continuously try to counteract influences that move the system apart from its ideal state. If the external influences exceed a 'manageable' range, the system is destroyed.

For our present discussion, the regulatory mechanisms of open systems are of central concern since they can actively interfere with the evolution of the system, by bringing about (an approximation of) the ideal state, or avoid the destruction of the system. Later, we will introduce a new kind of mechanism, called 'anticipation', that has the potential to significantly influence the systems evolution and most closely corresponds to our notion of chance discovery.

Prediction in the Natural Sciences

Nature is governed by the laws of physics, e.g., Newton's second axiom (the *total force* law). Obviously, in the physics domain there is no way to influence the natural laws. So even if we predict a phenomenon of high impact to society, such as a giant meteorite approaching the earth at high speed, all we can do is to evacuate the area the meteorite is predicted to hit.

Since it is not possible to change the course of nature, chance discovery here means to take appropriate (supportive, preventive) measures to minimize damage or maximize benefit.

Prediction in Open Systems

Open systems are characterized by system laws. (Schurz 1999) argued that we are *theoretically unable* to determine the exact numerical values corresponding to system laws, because the systems are open and hence described by nonlinear differential equations. In the extreme case, if external influences exceed the manageable (or critical) range of the system, nonlinear dynamics becomes effective and leads to *chaotic* behavior. Due to the sensitivity of open systems to external influences, prediction is a difficult matter. Below we will argue that in open systems, the activity of regulatory mechanisms is of major importance, rather than prediction.

Chance Discovery in Open Systems

Enterprise Example

Let us first give an illustrative example. *Enterprises* (companies) can be viewed as open systems that consist of subsystems (branches, sections, and individuals), and operate in an environment, the so-called 'economic market'. This environment typically satisfies the companies 'needs', e.g., customers demand the company's products. Under unfortunate circumstances, the company may run into the risk of being 'destroyed', e.g., by the appearance of a strong competitor (cf. $S1$). In spite of that, companies preserve identity through time (cf. $S2$). A company constantly tries to approximate an ideal state where, for instance, increasing profits are made and the economic situation of the company is stable. This is achieved by the company's subsystems that perform certain functions, including good production and distribution, and marketing (cf. $C1$). A company is typically confronted with a multitude of 'disturbing' influences in the form of, e.g., cheaper and better products of other companies and changing customer needs. At this point, the regulatory mechanisms of the company come into force, e.g., to lower production costs by increasing the efficiency of the production cycle. It is well-known that companies go bankrupt when a critical range is exceeded (cf. $C2$).

The Limits of Regulatory Mechanisms

Regulatory mechanisms are the system's means to approximate the system's ideal state. Those mechanisms are mainly active to compensate disturbing influences by *reacting* to them. Although regulatory mechanisms are usually able to guarantee the identity of an open system, they come into force only if confronted with 'threats' from the environment. For instance, if a company's sales decrease, the CEO might decide to shrink the company, thereby making a number of people unemployed.

In the next section we will argue that in addition to regulatory mechanisms, open systems need mechanisms of *anticipation* to cope with the complexities and influences of the environment.

Chance Discovery as Anticipation

In a recent report to the *Club of Rome*, (Botkin, Elmandjra, & Malitza 1998) introduce the term "anticipation" as a key feature of innovative learning that emphasizes human initiative. It is described as follows (Botkin, Elmandjra, & Malitza 1998, p. 25):

> [...] anticipation is not limited to simply encouraging desirable trends and averting potentially catastrophic ones: it is also the "inventing" or creating of new alternatives where none existed before.

Anticipation is contrasted to prediction, since the former focuses on the *creation* of possible and desirable futures, and plans to bring them about. The notion of anticipation shares the intuition of Alan Kay's phrase "The best way to predict the future is to invent the future".

Promotion In philosophy of science, the term "self-fulfilling prophecy" describes situations such as the following. Newspapers write articles about the morbidity of a bank institute. As a consequence, many customers of this institute withdraw their money and other commitments. In effect, the bank institute gets into serious trouble. A recent 'real' example is the success of the so-called *New Economy* (internet and telecommunication related shares). Since many people believed in its success, it became a great success (at least for some time).

Chance discovery as anticipation in this context means the promotion of a trend desired by New Economy companies. As a result of promotion, the desired trend could be effected. Similar forms of promotion are daily practice in companies: certain products are advertised with the hope that they actually trigger a desire in customers. The detection of 'latent' customer desires will be briefly discussed in the next section.

Collaboration In business there is a lot of talk about 'mergers'. Recently, for instance, a large Japanese and a large German company decided to collaborate in car industry. Collaborations are also seen in scientific research programs. We will briefly describe the field of Quantum Computation.

(Deutsch 1985) is reported to be the first to explicitly ask whether it is possible to compute more efficiently on a quantum computer. For a long time, this possible collaboration of quantum theory (physics) and artificial intelligence (computer science) remained a curiosity. However, there are already some indications of 'killer applications' for quantum theory. For instance, (Spector *et al.* 1999) report on problems that take polynomial time on a quantum computer but exponential time on a classical computer.

In academics, possibilities for collaborations are ubiquitous, and sometimes realized, e.g., in genome analysis, artificial intelligence and biology collaborate. What might chance discovery as anticipation mean here? In particular, how can we anticipate the success of a certain kind of collaboration? We cannot provide a working methodology here. In the case of quantum computation, the chance was 'discovered' by (Feynman 1982) who observed that classical systems cannot effectively model quantum mechanical systems. This observation suggests that computers based on the laws of quantum mechanics (instead of classical physics) could be used to efficiently model quantum mechanical systems, and possibly even solve classical problems such as database search in a highly efficient way.

Given that Quantum Computation will indeed be successful, how could we have known 10 years ago? One method would be to track the history of 'conjectures' (ideas, observations) formulated by various insightful researchers, and evaluate their feasibility in the light of current knowledge in possibly quite different research areas. The availability of huge amounts of information on the Web might facilitate such an endeavor.

Chance Discovery, Uncertainty, Freedom

One of the tacit assumptions underlying chance discovery is that the future is uncertain, and hence there is freedom to change is course of action.

Entropy Following (Suppes 2000), we propose entropy as the measurement of freedom. Specifically, the measurement of freedom is phenomenologically rather than procedurally oriented. The freedom of a set A of alternatives is measured by the entropy H of the actual chosen proportions, i.e.,

$$H(A) = -\sum_{i \in A} p_i \log p_i$$

where log is to the base 2, $p_i \geq 0$ and if $p_i = 0$ then $0 \log 0 = 0$. Accordingly, we may say that chances exist if there are (almost) evenly distributed alternatives. On the other hand, a market with, say, one dominant provider and high customer loyality has low entropy.

Let us repeat the aforementioned open system situation, that features a high degree of uncertainty, and formulate it as a problem for chance discovery.

Explanation Versus Prediction *Assume as given a theory that explains why a particular phenomenon turned out to be a chance (opportunity or risk), as observed by its high (positive or negative) impact. Given a similar phenomenon, can we employ this theory to predict high impact under comparable circumstances?* Of course, the notions of *similar* and *comparable* warrant further explication. In order to clarify the problem, consider the case of unstable or chaotic systems that support explanations without predictive value. Assume an ideal ball exactly on top of another ideal ball. Here, we cannot predict in which direction the ball will roll down, but after it rolled down, we can explain it by an unmeasureably small disturbance in the direction in which the ball rolled down (Schurz 1995).

Thus, the 'explanation vs. prediction' problem raises the fundamental question about which systems support the predictive use of chance discovery results. Straightforward answers seem to be ruled out by the fact that human initiative is essential to take opportunities or avoid risks, and the complexity of systems such as the web or financial markets.

Chance Discovery vs. KDD

(Fayyad, Piatetsky-Shapiro, & Smyth 1996) characterize Knowledge Discovery in Databases (KDD) as

> [...] the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data.

The discovery goal in KDD can be divided into a descriptive and a predictive part. In *description* the system seeks for patterns (or models) in order to present them to the user in an intelligible way; in *prediction* the system finds patterns so that the future behavior of some entity can be predicted. There exist a number of established (mostly statistical) data mining methods to achieve those goals, such as classification, regression, clustering, summarization, dependency modeling, and change and deviation detection (Fayyad, Piatetsky-Shapiro, & Smyth 1996).

Chance discovery may use the knowledge extracted by data mining methods to detect future features (opportunities and risks). For instance, by Web usage mining, i.e., the clustering of Web users based on their browsing activities, potential customer groups can be identified, and specifically addressed by companies. Here the interplay of data mining—describing correlations between users' interests—and chance discovery—actively promoting a possibility—is of crucial importance.

One may ask whether, e.g., data mining already *is* a form of chance discovery. Our answer is "no". Data mining can summarize or predict trends, but leaves out the rôle of human interference. Anticipation as a mechanism of an open system, on the other hand, 'matches' a desired (or predicted) trend with the system's goals (typically human 'desires') and accordingly takes supportive or preventive measures.

Discussion and Conclusion

In this paper, we explicate our take on a new research area called 'Chance Discovery'. The notion of 'open system', as characterized in cybernetics and system theory, serves as a framework to embed the activity of Chance Discovery. In particular, anticipation is introduced as a mechanism that may perform the rôle of detecting chances in open systems. The anticipating mechanism is explained in the context of promotion in New Economy and collaboration in the Quantum Computation research programme. Chance Discovery is contrasted to KDD and mutually beneficial aspects are explained. We identify human initiative as a distinguishing feature of Chance Discovery (as opposed to KDD), e.g., to actively initiate and foster a trend by promotion or to actively explore the (practical) feasibility of a theoretical conjecture.

Unlike the practical methods for data mining, we only described a methodology for Chance Discovery. A method for Chance Discovery might analyze 'success stories', i.e., cases where features of high impact for the future were successfully identified and accordingly promoted by human initiative. This retrospective analysis might be framed and processed by means of *KeyGraph* (Ohsawa, Benson, & Yachida 1998)), a smart indexing method originally developed for information retrieval. We have to leave this promising avenue for future research.

References

Ashby, W. R. 1964. *An Introduction to Cybernetics.* London.

Botkin, J. W.; Elmandjra, M.; and Malitza, M. 1998. *No Limits To Learning. Bridging The Human Gap.* A Report to the Club of Rome. Pergamon Press.

Deutsch, D. 1985. Quantum theory, the Church-Turing principle and the universal quantum computer. In *Proceedings of the Royal Society of London*, 97–117.

Fayyad, U.; Piatetsky-Shapiro, G.; and Smyth, P. 1996. Knowledge discovery and data mining: Towards a unifying framework. In *Proceedings 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96).*

Feynman, R. 1982. Simulating physics with computers. *International Journal of Theoretical Physics* 21:467–488.

Ohsawa, Y.; Benson, N. E.; and Yachida, M. 1998. KeyGraph: automatic indexing by co-occurrence graph based on building construction metaphor. In *Proceedings Advanced Digital Library Conference (IEEE ADL-98)*, 12–18.

Ohsawa, Y., ed. 2000. *Workshop on Chance Discovery and Management. In conjunction with the Forth International Conference on Knowledge-based Intelligent Engineering Systems and Allied Technologies (KES'2000).* Brighton, UK: IEEE, Inc.

Schurz, G. 1995. Scientific explanation: A critical survey. *Foundations of Science* 3:429–465.

Schurz, G. 1999. Normic laws as system laws: Foundations of nonmonotonic reasoning. In *Proceedings 4th Dutch-German Workshop on Nonmonotonic Reasoning Techniques and Their Applications (DGNMR-99).*

Spector, L.; Barnum, H.; Bernstein, H. J.; and Swamy, N. 1999. Quantum computation and AI. In *Proceedings 16th National Conference on Artificial Intelligence (AAAI-99).* Invited Talk.

Suppes, P. 2000. Freedom and uncertainty. In Natke, H., and Ben-Haim, Y., eds., *Uncertainty: Models and Measures, Mathematical Reasearch.* Academie Verlag. 69–83.

v. Bertalanffy, L. 1979. *General System Theory.* New York.