AI Magazine Volume 11 Number 5 (1990) (© AAAI)

need to include members from a variety of communities such as, but not necessarily limited to, production systems or logic programming (engines and representations), data model building, machine learning, knowledge acquisition, natural language, hypermedia, modeling and simulation, and distributed software systems engineering. Finally, there will also need to be a team of behavioral practitioners responsible for designing the organizational-computer collaboration features such as direct manipulation metaphors, tutoring capabilities, group productivity enhancement techniques and knowledge sharing/reuse concepts.

But just hiring such team members alone will not assure the future of the FSKBS field. There are major gaps in our understanding the theory of ontologies, knowledge servers, collaboration, and knowledge reuse approaches. Researchers need not only to pioneer new theories for these and related FSKBS topics, but also to address the scale-up issues associated with applying the resulting techniques to large-scale KB problems. Researchers also need to evaluate current FSKBS applications to extract lessons learned and to compile design-related insights. Unfortunately, few researchers are prepared to think as broadly as or to work for the durations required to fill these FSKBS needs.

The costs and scale of effort of researching and developing FSKBS shells is intimidating. The potential benefits, however, seem even more dramatic. FSKBS shells portend the coming of the "knowledge age", an era where machines begin to truly understand our world and to collaborate with us as equals. The founding vision for the AI field is of machines that do our bidding without needing to be programmed. The advent of FSKBS shells will be a significant step in that direction.

While all at the workshop shared the common goal of reducing brittleness through the creation of larger scaled KBs, it seems safe to conclude two major paths are being taken toward that goal. On the one hand the advocates of the knowledge principle are pioneering in the areas of ontology development and national initiatives to launch man-centuries of KB coding effort. They are doing this in the belief that there are no short cuts to the creation of very large KBs. On the other hand, there are the application builders, knowledge server developers, and collaborative systems researchers who believe that FSKBSs must ultimately be multifaceted. It is worthwhile to attempt to advance and integrate a number of technologies that eventually will be needed by FSKBSs: e.g., knowledge acquisition, machine learning, natural language, database, and distributed agent frameworks to mention a few.

In the final analysis both approaches are necessary and complementary. No nation has the resources to continue large scale KB encoding by hand on an indefinate basis, and even if they did, hand coding would never keep up with the realtime problem solving, dynamic data flow, and collaboration requirements of current day FSKBSs. Similarly, the relative immaturity of current knowledge server technologies render most FSKBS developers to the position of crafting what they can as the need evolves. The concensus appears to be that in one to two decades, progress on both sides — manual knowledge coding and scale-up of knowledge servers — may both mature to the point where their merger could lead to an intelligent system that is both broad and able to solve real world problems. Until then, the two paths seem destined to be walked at first separately, and then increasingly in tandem.

Acknowledgements

The authors gratefully acknowledge the efforts of the workshop sponsors and attendees, and especially R. Gregory Wenig, workshop Co-chair, and Quasir Durrani, who helped with the organization and correspondence.

About the Authors

Barry G. Silverman is Director of the Institute for AI at the George Washington University and is author/editor of over 12 books or proceedings some of which include: *Expert Systems for Business, Military Applications of AI, AI Systems in Government,* and *Building Expert Critics* (pending).

Arthur J. Murray is Manager of Advanced Technology at McDonnell-Douglas in McLean, Virginia. He received his Ph.D. in engineering administration from the George Washington University in 1989.

1990 AAAI Spring Symposium Report:
Theory and Application of Minimal-Length Encoding

Edwin P. D. Pednault

This symposium was very successful and was perhaps the most unusual of the spring symposia this year. It brought together for the first time distinguished researchers from many diverse disciplines to discuss and share results on a particular topic of mutual interest. The disciplines included machine learning, computational learning theory, computer vision, pattern recognition, perceptual psychology, statistics, information theory, theoretical computer science, and molecular biology, with the involvement of the latter group having lead to a joint session with the AI and Molecular Biology symposium. The unifying topic was the problem of finding the shortest encoding of a body of information in a language rich enough to admit many possible encodings.

This particular problem arises in different forms and for different reasons in each of the disciplines mentioned above. In machine learning and computational learning theory, minimal encoding can be used to infer a general theory from a set of observations. If one employs a language for encoding observations in terms of general rules or properties, the shortest description will be an encoding of the desired theory together with a minimal set of additional information necessary to derive the observations from the theory.

In computer vision and pattern recognition, scene analysis can be equated with finding the shortest encoding of an image in a language for describing images in terms of objects and their visual properties. In perceptual psychology, this cod-

ing idea has been used to explain certain phenomena in human visual perception.

Statistical inference can likewise be accomplished by means of minimal encoding. In this case, data is described by first encoding a probabilistic model and then appending the Shannon code for the observations with respect to the model. Shannon coding is a technique that minimizes the expected length of code words for a given probabilistic model. The shortest such two-part encoding yields a model that fits the data well without overfitting.

This statistical coding criterion, known as the minimum description-length principle, is also useful in vision and learning applications when noise and other sources of uncertainty are an issue. It was also illustrated at the joint session with the AI and Molecular Biology symposium, where results were presented showing how the minimum description-length principle can be used to help discover hidden properties of DNA.

While the above uses focus on the content of minimal encodings, the length of the shortest encoding also has meaning in and of itself. If a collection of data exhibits a significant amount of structure and regularity, then this structure can be exploited to obtain a compact encoding. Conversely, the less the apparent structure is, the longer the encoding. The length of the shortest encoding therefore measures the amount of complexity (i.e., lack of structure) in the data. When data is encoded in terms of computer programs that generate the data as output, the length of the shortest encoding is known as the Kolmogorov complexity of the data.

Kolmogorov complexity has been used in statistics and information theory to help define the notion of intrinsic randomness by equating randomness with high Kolmogorov complexity. In theoretical computer science, generalized Kolmogorov complexity measures have been developed that restrict the amount of computation used by the generating programs. These measures have been found useful in exploring issues ranging from pseudorandom number generation and the existence of one-way functions, to the P=NP problem and the relationship between other computational complexity classes.

The papers presented at the symposium provided a broad view of the various uses of minimal encoding described above. In the machine learning and computational learning theory sessions, Sharad Saxena presented experimental results demonstrating that the coding length of a target concept is a good indicator of the learning performance of Quinlan's ID3 program, with more compact encodings yielding better performance. Ray Solomonoff presented a scheme for incremental learning based on his pioneering work on algorithmic probability measures. These measures are related to the Kolmogorov complexity of the observed data. Ming Li and Paul Vitanyi presented results relating Solomonoff's algorithmic probability measure to the Valiant learning model. They showed that the number of concept classes that are polynomially learnable in the Valiant model can be increased by restricting the probability distributions that are considered to ones dominated above by Solomonoff's measure. David Aldous and Umesh Vazirani presented other extensions to the Valiant model in which examples are drawn according to a Markov random process. Sanjeev Kulkarni and Sanjoy Mitter presented results on the relationship between the notion of metric entropy studied by Kolmogorov, the Vapnik-Chervonenkis dimension, and the learnability of concept classes. Osamu Watanabe presented preliminary work on a formal theory of learning that allows a "pupil" to query a "teacher" with yes/no type questions, with the teacher possibly supplying a counterexample in the case of a no response.

In the computer vision and pattern recognition sessions, Peter van der Helm presented joint work with Emanuel Leeuwenberg on deriving a psychologically plausible language for encoding visual patterns, with the goal of explaining various phenomena in human perception. Pascal Fua and Andy Hanson presented a framework for extracting polygonal features (i.e., buildings, roads, etc.) from aerial images, where the number, locations, and shapes of the polygons are determined using a minimal encoding criterion. Kenneth Keeler presented a scheme for encoding subdivisions of an image plane using results from topology. He then showed how this scheme can be applied to image segmentation. Yvan Leclerc presented refinements to his image segmentation technique in which the discrete search space of possible encodings is approximated by a continuous function. The function is then minimized using numerical analysis techniques. Alex Pentland and Trevor Darrell presented a technique for segmenting images into objects by first estimating the locations and positions a large set of potential object parts. A modified Hopfield-Tank neural network is then used to select the subset of potential parts that minimizes the overall coding length of the image. Arthur Sanderson and Nigel Foster presented a feature-oriented approach to model-based vision that handles missing, extra, and noisy features by employing a minimal encoding criterion to match image features with model prototypes. Mati Wax presented an acoustic signal processing application. In his application, a minimal encoding criterion is used to detect and localize one or more acoustic sources based on signals received by a sensor array.

In a lively joint session with the AI and Molecular Biology Symposium, Peter Cheeseman presented a minimal encoding approach for reconstructing evolutionary trees for families of related genes based on a model of how genetic codes mutate over time. Marla Babcock presented joint work with Wilma Olson and Edwin Pednault on the use of minimal encoding to identify structurally and/or functionally significant regions in DNA by detecting local statistical variations in the DNA sequences. Aleksandar Milosavljevic, David Haussler, and Jerzy Jurka presented a minimal encoding technique for clustering DNA sequences into classes, with the aim of discovering new biologically-significant classes. Lloyd Allison, Chris Wallace, and C. N. Yee presented a DNA sequence alignment technique based on minimal encoding. Their technique produces not only the best alignment according to their model but also a pictorial representation of the most probable competing alignments, which is useful information for biologists.

In the statistical inference sessions, Andrew Barron and Tom Cover presented convergence results for the minimum description-length principle. They showed that the principle enables one to converge on the cor-

rect probabilistic model at a rate that is bounded above by an expression involving the relative entropy distance between the correct model and all other models in the hypothesis space. Laszlo Gerencser presented results showing how a variant of the minimum description-length principle known as stochastic complexity can be applied to parameter estimation in discrete-time linear systems, which is useful for such things as adaptive control in robotics. Jacob Sheinvald, Byron Dom, and Wayne Niblack presented two methods for detecting redundant features in a classification feature space. One method uses the minimum description-length principle, the other uses stochastic complexity, with the stochastic complexity method performing just slightly better in the experiments they conducted. Using results from statistical mechanics, Naftali Tishby, Esther Levin, and Sara Solla showed that if we interpret parametric learning in which one minimizes an additive error function (e.g., learning in neural networks) as being equivalent to maximizing the posterior probability of the parameters, then the posterior probability will be given by the Gibbs distribution. In addition, the average prediction error will equal the stochastic complexity and will be given by a derivative of the free energy. Jacob Ziv and Neri Merhav presented a method based on the Lempel-Ziv data compression algorithm for estimating the number of states in a finite-state random process. The approach minimizes the probability of underestimating the number of states given a user-defined upper bound on the probability of overestimation. In a separate talk, Neri Merhav also presented results showing that the Lempel-Ziv algorithm asymptotically minimizes the probability of the code length exceeding a given threshold for any threshold. He also showed that the minimum description-length principle achieves an even faster rate of convergence to this minimum probability for a particular subclass of finite state random processes.

In addition to the six statistical inference talks described above, four other presentations were made on the specific topic of clustering and classification. Three of these considered the classical problem of grouping feature vectors into a suitable number of classes. Jorma Rissanen and Eric Ristad presented one approach that first uses histogram techniques to approximate the overall distribution of the feature vectors. The local maxima in the histogram are then used as seed points to find the set of classes that minimizes the stochastic complexity of the data. Chris Wallace presented a different approach based instead on the minimum description-length principle. His approach uses a split-merge technique to group feature vectors into a set of classes that minimizes the overall coding length of the data. Richard Wallace and Takeo Kanade also employed the minimum description-length principle, but their approach involves first constructing a cluster tree using more conventional statistical techniques, and then deciding where to break the tree to minimize the overall coding length. Instead of considering feature vectors, Jakub Segen presented work on grouping labeled graphs into classes and then matching new graphs to these classes using the minimum description-length principle. Particular attention was given to the heuristics needed to overcome the inherent complexity of clustering and matching.

In the information theory and theoretical computer science sessions, Vladimir Uspensky presented joint work with Kolmogorov and other colleagues at Moscow Lomonosov University on intrinsic randomness and its relationship to Kolmogorov complexity and other descriptional complexity measures. Jack Lutz presented results on a measure-theoretic definition of intrinsic pseudorandomness. His results combine work initiated by Andy Yao on resource-bounded pseudorandom generators with work by Martin-Lof on measure-theoretic definitions of intrinsic randomness. Luc Longpre presented several results on space-bounded Kolmogorov complexity, where a space bound is placed on the computations needed to reconstruct observations from their encodings. In particular, he related space-bounded Kolmogorov complexity to Martin-Lof randomness tests, showing among other things that space-bounded Kolmogorov random strings will pass all Martin-Lof tests whose computations can be performed in less space, but not tests that use slightly (about four times) more space. Leonid Levin, Russell Impagliazzo, and Michael Luby showed that the existence of one-way functions in cryptography is necessary and sufficient for the existence of pseudorandom generators, with the proof introducing the notion of computational entropy. Eric Allender presented a survey of results that he and others have obtained in which time-bounded Kolmogorov complexity is used to address the E=NE problem (exponential time equals nondeterministic exponential time), as well as issues regarding pseudorandom generators, tally sets, and circuit complexity. Ron Book presented a survey of results in which the notion of a set with small information content is introduced and defined in terms of a slightly different time-bounded Kolmogorov complexity measure. He showed how sets with small information content can be used to address the P=NP problem as well as issues in circuit complexity.

From these presentations, it is apparent that minimal encoding concepts are having a broad impact in several different fields. Prior to the symposium, there had not been much interaction among researchers across fields. Since the symposium, minimal encoding has become a featured topic at the 1991 Workshop on Computational Learning Theory, and machine learning has become a topic at the 1991 IEEE Computer Society Data Compression Conference. The symposium has clearly succeeded in stimulating interdisciplinary interaction and in heightening the awareness among all attendees of the many aspects of minimal encoding. It is hoped that the future will show the meeting helped inspire further interaction and collaboration, as well as a broadening of this area of study.

About the Author

Edwin Pednault is presently a research scientist at AT&T Bell Laboratories. He obtained his Ph.D. in electrical engineering from Stanford University in 1987 for his thesis on the mathematical foundations of automatic planning. He received an M.S. in computer science from Stanford in 1980 and a B.Eng. in electrical engineering from McGill University in 1979. His research interests include automatic planning, knowledge representation, minimal encoding, and the application of minimal encoding to machine learning, computer vision, and molecular biology.