On Reasoning From Data

DAVID WALTZ

NEC Research Institute, Princeton, NJ 08540
⟨waltz@research.nj.nec.com⟩

SIMON KASIF

Department of Computer Science, Johns Hopkins University, Baltimore, MD 21218
⟨kasif@cs.jhu edu⟩

Introduction

Our society is currently entering a new phase in which gigabytes of information are becoming readily available for exploration over academic networks, digital libraries, and commercial information services as well as in proprietary commercial and governmental databases. This important technological development presents a substantial challenge, as future intelligent systems must be able to store very large streams of data, summarize and index this data using concise and efficient models, and subsequently perform very efficient retrieval and reasoning in response to real-time queries and updates. We informally refer to this challenging task as *reasoning from data*.

Most previous AI research and applications have concentrated on relatively simple operations, for example, highly constrained queries on relatively static, immutable systems of knowledge such as mathematics, chess, and hardware components inventories, where it is possible to abstract rules that can be viewed as true and valid. There are many other domains in which data changes more or less rapidly and in which abstract truths are at best temporary or contingent, for example, robot environments, software environments, demographic databases and public-health data, ecological and economics (eco)systems, chemical processes, marketing and point-of-sale databases, financial time series, and video and text databases. In addition, these domains are associated with a demand for very fast response to unanticipated queries and continuous updates over uncertain, dynamic, interactive, and rapidly changing environments.

These domains present a challenge for purely symbolic, rule-based approaches to AI. For instance, it appears to be difficult to give a formal logical specification of concepts such as an important electronic message, a fair scheduler, an urgent phone call, a good travel package to Hawaii, an intriguing new paper about Bayesian reasoning, a high-risk car, a good real-estate investment, or an interesting economic trend.

Statistical decision theory [Pearl 1988] provides a useful framework to model adaptive intelligent agents in stochastic and rapidly evolving domains. Moreover, it provides precise criteria (loss functions, expected utility) to evaluate the performance of such agents. However, when the environment is large, the process of fitting good models (finding maximum *a posteriori* models or even maximum likelihood models) to data generated by the environment is typically computationally intractable. When the environment is small, we often have trouble getting sufficient statistics. Thus the models we can devise effectively are rarely accurate, regardless of the size of

Permission to make digital/hard copy of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, the copyright notice, the title of the publication and its date appear, and notice is given that copying is by permission of ACM, Inc. To copy otherwise, to republish, to post on servers, or to redistribute to lists, requires prior specific permission and/or a fee.
© 1995 ACM 0360-0300/95/0900-0356 $03.50

the environment. Finally, equally disturbing is the fact that reasoning with general probabilistic models is intractable.

In this review we discuss a probabilistic (statistical) framework for *memory-based reasoning* (MBR) that marries the strengths of knowledge representation using probabilistic models and the computational advantages of case-based reasoning. We also argue for the importance of MBR as a paradigm for building modern intelligent applications in general. This paradigm has been successfully applied to information retrieval, robotics, classification, software agents, computational biology, and a number of other tasks.

Our main goals for MBR are (1) to approximate computationally efficient decision-theoretical intelligent agents in highly dynamic stochastic environments, and (2) to shift the burden from the slow and costly effort of hand-coding applications to the building of largely autonomous adaptive systems supported by rapidly decreasing-cost technologies such as powerful PCs, very large associative memories, and massively parallel and distributed parallel architectures.

Memory-Based Reasoning

The MBR approach attempts to combine the strengths of case-based reasoning (CBR) [Kolodner 1993] and probabilistic reasoning [Pearl 1988]. In the first phase, MBR procedures analyze data using relatively efficient algorithms in order to obtain a rough model of what the environment might look like. The model is subsequently used to define an adaptive model-based geodesic (distance metric) on the domain that in turn induces a transformation on instances in the original data space. The new transformed space can then be indexed using automated recursive partitioning methods over real-valued attribute spaces. Given a query, MBR retrieves a set of relevant instances (as judged by the rough model) and then uses local modeling techniques that interpolate the answer to the query on only a relatively small set of nearby (relevant) instances. Consequently, MBR procedures can use sophisticated (computationally intensive) local models.

The principal steps of MBR are outlined in the following:

—The user specifies a set of features that should be tracked by observing the environment (database). The user also specifies a set of probabilistic assumptions used to generate models (e.g., priors, independence assumptions specified as Bayes networks; see Pearl [1988]).

—The MBR agent incrementally generates a rough probabilistic model of the environment. The model may contain new hidden variables.

—The probabilistic model is used to induce an adaptive distance metric on the domain that induces an explicit transformation (*MBR transform*) on the static parts of data. Thus each data instance is transformed to a new instance that may include new attributes (e.g., hidden variables).

—A new, more efficient representation is mechanically derived in the transformed space.

—Local probabilistic models are learned over small neighborhoods. A local model is learned lazily only in response to a query. That is, once a specific data point is accessed during retrieval, MBR retrieves the most relevant set of instances to the query. It then performs local learning on this set, thereby producing a local model that is used to answer the specific query. Local modeling can be performed with locally weighted regression, piecewise linear separators, propositional formulae, or other learning methods.

—The local models are retained in short-term temporary memories and are cached if the number of queries to a given region is large.

The MBR hypothesis suggests that we not associate a static atomic symbol with an event. Each event is dynamically (incrementally) defined as an MBR-vector in a real-value space. For instance, Stanfill and Waltz [1986] and several later papers (e.g., Cost and Salzberg [1993], Rachlin et al. [1994], and Zhang et al. [1992]) assume conditional independence of features given a partitioning of the domain into classes. This simple probabilistic model (a two-layer causal tree [Pearl 1988]) induces a transformation that maps a symbol A associated with x_j to a discrete probability distribution $(p_1, p_2, \ldots, p_k)$ where p_i is the probability of the ith class given $x_j = A$. In planning domains, a given symbol is transformed into vector probabilities of accomplishing a set of tasks. Other transformations involving probabilistic entities such as mutual information (cross entropy with some event), log-likelihood ratios, and hidden-variables have been investigated by the authors.

MBR obviously admits an "embarassingly parallel" low-communication mapping to parallel and distributed systems and has been implemented successfully on a number of platforms. Alternatively, MBR can utilize data structures for associative retrieval (such as KD-trees or R-trees) that retrieve points in expected logarithmic time.

Discussion

Traditional AI systems are based on reasonable and intuitively appealing principles that have resulted in numerous successful practical demonstrations. Despite their apparent differences, most conventional symbolic AI systems use symbolic pattern matching over rigid symbolic expressions and categorical variables. They also typically rely on laborious handcoding, have difficulty coping with uncertainty and change, and rarely use statistical evaluation of the quality of specification in terms of its match to the observed data.

Memory-based reasoning generalizes traditional nearest-neighbor (NN) procedures used in pattern recognition (see Dasarathy [1991]) and provides a practical framework for reasoning from stored data. Note that conventional NN methods typically use neither adaptive distance metrics nor local interpolating functions. The special case of MBR used in control applications often uses local functional approximation methods such as locally weighted regression (see Moore et al. [1995]). These methods typically do not use sophisticated model-based adaptive distance functions.

Recently, several researchers in statistics have independently suggested adaptive kernel methods that bear a strong similarity to MBR, another indication of the usefulness of the paradigm. This work in statistics is primarily concerned with classification and function approximation rather than general query answering or data mining. However, the motivation is similar.

Another recent relevant development is the emergence of case-based decision theory in economics. Case-based decision theory defines the "subjective utility" of a state as a statistical kernel function approximation to expected utility and, like MBR, is motivated by the need to reconcile decision theory and traditional theories of rationality with the practical reality of resource-bounded intelligent agents, in this case humans. Finally, database researchers also are beginning to realize the potential of this methodology for data-mining applications.

To summarize, the memory-based reasoning approach attempts to marry the strengths of probabilistic reasoning with the computational advantages of case-based reasoning. Although conventional case-based reasoning methods require hand-coding of cases, have not been rigorously evaluated subject to statistical criteria, and are unlikely to scale up as the complexity of the domain is increased, MBR relies on adaptive model-induced distances that are known to improve the statistical performance and scalability of NN methods.

Memory-based reasoning has been used successfully in a number of domains such as classification of news articles [Masand et al. 1992], classification of census data [Creecy et al. 1992], software agents

[Maes and Kozierok 1993], computational biology [Zhang et al. 1992; Cost and Salzberg 1993], and robotics [Moore et al. 1995], and a number of other applications.

This review has focused on application-oriented MBR rather than its cognitive interpretation. Discussion of a cognitive perspective is precluded by the length requirements of this review, but see Waltz [1990]. The authors urge readers to contrast Allen Newell's "symbol hypothesis" with MBR's nature—incremental, model-based, constantly mutating in response to current beliefs, goals, and experiences.

Much research must still be done to understand fully the relative capabilities and limitations of the outlined methodology. The MBR proposal shares the goals of machine learning, namely, shifting the focus in AI research to activities in which most of the burden can be put on hardware and algorithms for adaptive domain-independent reasoning from data that do not require significant input and costly labor from human experts and programmers. Currently available preliminary experiments with MBR give an optimistic prognosis for the success of this a paradigm.

REFERENCES

CREECY, R., MASAND, B., SMITH, S., AND WALTZ, D. L. 1992. Trading MIPS and memory for knowledge engineering. *Commun. ACM 35*, 8, 48–64.

COST, S. AND SALZBERG, S. 1993. A weighted nearest-neighbor algorithm for learning with symbolic features. *Mach. Learn. 10*, 1, 57–78.

DASARATHY, B. V., Ed. 1991. *Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques*. IEEE Computer Society Press, Los Alamitos, CA.

KOLODNER, J. 1993. *Case-Based Reasoning*. Morgan-Kaufmann, San Mateo, CA.

MOORE, A., ATKESON, C., AND SCHAAL, S. 1995. Memory-based learning for control. *Artif. Intell. Rev.* (to appear).

MAES, P. AND KOZIEROK, R. 1993. Learning interface agents. In *Proceedings of the Eleventh National Conference on Artificial Intelligence* (Washington, D.C.), MIT Press, Cambridge, MA, 459–465.

MASAND, B., LINOFF, G., AND WALTZ, D. L. 1992. Classifying news stories using memory-based reasoning. In *Proceedings of the SIGIR*, 59–65.

PEARL, J. 1988. *Probabilistic Reasoning in Intelligent Systems*. Morgan-Kaufmann, San Mateo, CA.

RACHLIN, J., KASIF, S., SALZBERG, S., AND AHA, D. 1994. Towards a better understanding of memory-based and Bayesian classifiers. In *Proceedings of the Eleventh International Conference on Machine Learning* (New Brunswick, N.J., July), 242–250.

STANFILL, C. AND WALTZ, D. 1986 Toward memory-based reasoning. *Commun. ACM 29*, 12, 1213–1228.

WALTZ, D. L. 1990. Eight principles for an intelligent robot. In *SAB-90: Simulations of Animal Behavior*, S. Wilson and J. Meyer, Eds., MIT Press, Cambridge, MA, 462–465.

ZHANG, X., MESIROV, J. P., AND WALTZ, D. L. 1992. A hybrid system for protein secondary structure prediction. *J. Molecular Biol. 225*, 1049–1063.