PNPACK: Algorithm and software development for probabilistic systems

Scope: A medium-term (3–5 year) project including basic research, software development and distribution, and creation of new classes of applications.

Investigators: (alphabetical order): Michael Jordan (MIT), Simon Kasif (Johns Hopkins), Daphne Koller (Stanford), Jitendra Malik (Berkeley), Judea Pearl (UCLA), Stuart Russell (Berkeley).

Industrial partners: Harlequin Inc. (Menlo Park), HRI (Berkeley).

An intelligent agent or system must reason and act in the real world. Over the last decade, there has been growing consensus that the ability to reason and act under uncertainty is fundamental to intelligent behavior and to successful performance in a wide variety of important application areas. After all, the available sensors are rarely accurate, there are many relevant factors that cannot be observed, the future can never be predicted with certainty, and actions do not always have the desired outcomes. There is now a broad consensus that systems based on probability, particularly in the Bayesian tradition, can provide the rigorous and effective tools needed for handling uncertainty. Such systems are based on well-understood techniques such as *conditioning* for incorporating new information, and *expected utility maximization* for making decisions.

For many years, probabilistic reasoning was only feasible for very small applications, since the representation of a probability distribution for realistic domains was too large to be practical. In recent years, techniques have evolved for decomposing complex problems into structured probability representations. This has led to rapid improvements in inference technology and application performance. For example, probabilistic networks (PNs) (also known as Bayesian belief networks) dominate the "Uncertainty in AI" and "AI in Medicine" communities. They have been applied to a wide variety of tasks, including natural language processing, information retrieval, image analysis, fault diagnosis, robot control, forecasting, medical expert systems, real-time sensor monitoring, software debugging, and various functions in Microsoft's Windows 95 operating system.

Other structured probability models include Markov random fields, hidden Markov models, and hierarchical mixtures of experts. Within their respective range of applicability, each of these representations dominates competing approaches. For example, hidden Markov models are currently the technology of choice for automated speech recognition and DNA modelling. Finally, the neural network community has come to recognize the value of probabilistic interpretations of neural networks and has begun to adopt probabilistic techniques for their training.

In summary, AI, computer science, and statistics have made very significant strides in the past decade in the area of computing with probabilities. Several large research communities are now engaged in investigating the application and further development of structured probabilistic models.

One very serious obstacle to the exploitation of these advances and to further scientific progress is the lack of a good infrastructure for the development and testing of new probabilistic inference algorithms and applications. Commercial software, such as HUGIN, is expensive ($8000 academic license), narrow, and inflexible because source code is not available. The most comprehensive available public domain software, IDEAL, is written in Lisp using inefficient data structures and algorithms and runs approximately 100 to 500 times slower than necessary. Since the mathematics and the algorithmic details are quite complex, some researchers avoid implementation altogether. Others have to devote a significant amount of research effort to reimplementing existing algorithms. The result is a large number of incompatible code fragments, which make it very difficult to compare between algorithms, to distribute new developments to researchers worldwide, and to construct systems that use more than one technique in solving a problem.

In addition to the problems for researchers and application developers, the lack of efficient, integrated software with source code availability also seriously impedes education and training.

Although textbooks are available that stress probabilistic methods, the lack of software suitable for educational purposes has proved to be a big problem for instructors, and is slowing down the process of training new generations of researchers and practitioners in this vital technology.

We propose a multi-year, multi-investigator project that will address these issues, thereby significantly advancing the state of the art in computing with probabilistic information. The project has three basic components:

- Basic research to develop more powerful representations for probabilistic information and more efficient algorithms for computing with those representations.
- Development and distribution of *PNPACK*—a suite of rigorously tested, efficient, portable, extensible software written in C++. The suite will embody the latest developments in probabilistic reasoning technology, and will be structured so as to allow easy incorporation of the new algorithms and representations that will be developed by the basic research component.
- Development of new prototype applications using PNPACK as a platform and taking advantage of the new algorithms and representations.

The project is modelled on the successful LAPACK project, which has revolutionized the quality, efficiency, and ease of development of scientific computing applications. PNPACK will serve as a conduit through which "suppliers"—the research community—can disseminate their advances to "customers"—application developers, other researchers, and educators. A key aspect of the proposal is the prior commitment of a large "customer base" to adopt the technology we produce.

Some of the extensions we will develop and incorporate into this project are:

- *Hybrid networks* containing both discrete and continuous variables.
- Improved *stochastic simulation* algorithms, which appear to be the best candidate for inference in very large networks or in general networks with continuous variables.
- *Dynamic probabilistic networks* for handling domains that change over time, including real-time monitoring.
- *Knowledge-based model construction* of belief networks tailored to specific circumstances from general-purpose probabilistic knowledge bases.
- *Learning methods* for constructing networks from data and prior knowledge.
- *Hierarchical probabilistic models* which permit representation and inference at different levels of abstraction, thereby making possible the modelling of very large domains.
- *Functional representation* of probabilities, and using them to speed up inference.
- *Dynamic data structures* that trade off the cost of processing evidence when it is obtained and the cost of answering queries.
- Algorithms that *dynamically refocus* attention on the most relevant aspects of the problem during the course of inference.
- *Parallel algorithms* for high-speed solution of hard problems.

The resulting system will extend the scope of problems that can be represented and solved using structured probabilistic models, and allow the development of more flexible and efficient algorithms for probability computations. It will also greatly increase the impact of basic research in this area on the development of real-world applications, eventually allowing the construction of probabilistic inference systems for large complex domains.