References from the talk "How Bayes re-invented Artificial Intelligence"

John Mark Agosta*

1 Early AI, 1980-1994

Artificial Intelligence in the 1980's embraced logical methods over quantitative and probabilistic ones. Expert systems were a fad. Their utility fell way short of their promise, and a backlash ensued, as anticipated in the prescient first article listed here.

But at the same time, several authors pursued a probabilistic agenda that today has become the mainstream. The remaining articles in this section are some of the notable sources of that time.

1. Drew McDermott, M. Mitchell Waldrop, Roger Schank, B. Chandrasekaran, John McDermott "The Dark Ages of AI: A Panel Discussion at AAAI-84," AI Magazine Volume 6 Number 3 (1985).

   > "To sketch a worst case scenario, suppose that five years from now the strategic computing initiative collapses miserably as autonomous vehicles fail to roll. The fifth generation turns out not to go anywhere, and the Japanese government immediately gets out of computing. Every startup company fails....
   >
   > "I don't think this scenario is very likely to happen, nor even a milder version of it. But there is nervousness, and I think it is important that, we take steps to make sure the "AI Winter" doesn't happen—

   Of course, it all came true.

2. Peter Cheeseman, "In Defense of Probability," IJCAI 1985.

   In this paper, it is argued that probability theory, when used correctly, is sufficient for the task of reasoning under uncertainty.

3. J. Pearl "Fusion, Propagation, and Structuring in Belief Networks" Artificial Intelligence 29 (1986) 241-288

   The local propagation method Pearl developed later became the subject of his book: Pearl, J., "Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference," (San Mateo, CA: Morgan Kaufmann, 1988)

   These original sources on Belief Networks are eminently readable.

4. Shachter, R., D., "Evaluating Influence Diagrams," Operations Research, vol 33, No. 6, (1986).

   An extension of Belief networks with decision and value variables, and alternate view of network updates by reductions of the network rather than by local propagation.

5. Whittle, P., "Probability via Expectation", (Springer-Verlag, 1992).

   An alternate, more intuitive derivation of the probability axioms.

6. Cox, R. T. "Algebra of Probable Inference," Johns Hopkins University Press (2001).

   First published in the '40s, this is an elegant derivation to show why any reasonable assignment of the weight of evidence to a number must be equivalent to probability.

7. See also E. T. Jaynes and G. L. Bretthorst, "Probability Theory: The Logic of Science" (Cambridge, 2003). for a comprehensive discussion of probability from a Bayesian view point.

8. D. Heckerman, "Probabilistic Interpretations for Mycin's Certainty Factors," Workshop on Uncertainty and Probability in AI, UCLA, pp. 9-20, (1985).

   From the first UAI workshop in 1985, the inconsistencies in certainty factors revealed in this work became the argument for abandoning them for probability based reasoning.

9. Agosta, J.M., T. Gardos, "Bayes Network 'Smart Diagnostics'" Intel Technology Journal, Vol. 8 No. 4 (November 2004).

   A review of research published in the UAI conference on Bayes network diagnostic modeling.

10. Boutilier, C., "The Influence of Influence Diagrams on Artificial Intelligence," Decision Analysis, Vol.2, No. 4, (December, 2005), pp. 229-231.

    An insightful retrospective on the origins of probabilistic AI.

2 AI and Machine Learning 1995-2004

Statistical approaches to Machine Learning became popular in the 1990s. Bayes networks, rechristened *Probabilistic Graphical Models* by the statistics community, earned a place alongside other tools in the field.

1. David Heckerman "A Tutorial on Learning with Bayesian Networks," March 1995, Revised November 1996,Technical Report MSR-TR-95-06.

   This watershed paper did two things: It placed Bayes networks squarely within the field of Bayesian statistics, and it introduced Bayes networks as a tool for Machine Learning.

2. A textbook that covers Bayes networks, both as part of early AI, and in their use for Machine Learning.
   Cowell., R. G., A. P. Dawid, S. L. Lauritzen, D. J. Spiegelhalter, "Probabilistic Networks and Expert Systems," (Information Science and Statistics, 2003) .

3. Ng, A. and M. Jordan, "On discriminative vs. generative classifiers: A comparison of logistic regression and naïve Bayes," in Advances in Neural Information Processing Systems 14. Cambridge, MA: MIT Press, 2002.

   By definition, PGMs are generative models that are inverted by inference to serve as discriminative models. The generative–discriminative distinction points out a primary feature the Bayesian approach.

4. This paper introduced *Tree-Augmented (Bayes) Network classifiers* (TAN) by combining naïve Bayes with a spanning tree network among features, borrowing the idea from previous work on Chow-Liu trees. This improved the performance of naïve Bayes to be competitive with other Machine Learning techniques.
   Friedman, N., D Geiger, M. Goldszmidt "Bayesian Networks Classifiers" Machine Learning, 29, 131–163 (1997).

5. Scoring methods for selecting models are the Bayesian approach to learning Bayes network structure. This problem arises generally, and the conventional statistical approach of goodness-of-fit tests fails, as this quote describes:

> "GoF tests have some serious limitations. If we reject H_0 then we conclude we should not use the model. But if we do not reject H_0 we cannot conclude the model is correct.[1]

Bayes Factors as an alternative resolve the problem, as this tutorial paper explains.

Kass, R. E., and A. E. Raftery, "Bayes Factors" Journal of the American Statistical Association, Vol. 90, No. 430. (Jun., 1995), pp. 773-795.

3 A Recent Application to "Query-Based Diagnosis"

Here is some recent work extending Bayes networks, by observing expert users as they solve problems and inferring the structure of the network from their actions. In concept the approach to probability is slightly different, treating the user's actions as constraints on the model probabilities rather than uncertain knowledge about their values. This quote makes the distinction:

> ".. a Bayesian is a probabilist, a person who sees making up the mind as a matter of either adopting an assignment of judgmental probabilities or adopting certain features of such an assignment, e.g., the feature of assigning a higher conditional probability to [survival of diagnosis of one cancer than on diagnosis of another].[2]

1. Agosta, J. M., Thomas Gardos and Marek J. Druzdel, "Query-based Diagnostics," The Fourth European Workshop on Probabilistic Graphical Models (PGM 08) Hirtshals, Denmark, September 17-19, 2008.

2. Agosta, J. M., Omar Zia Khan and Pascal Poupart, "Evaluation Results for a Query-Based Diagnostics Application," The Fifth European Workshop on Probabilistic Graphical Models (PGM 10) Helsinki, Finland, September 13-15, 2010.

3. Omar Zia Khan, Pascal Poupart and J. M. Agosta, "Automated Refinement of Bayes Networks' Parameters based on Test Ordering Constraints," Neural Information Processing Systems 2011 (NIPS 11), Granada, Spain, December 12-15, 2011.

[1] Wasserman "All of Statistics" (Springer Texts in Statistics, 2004) p.168.
[2] p.2, R. Jeffrey, Probability and the Art of Judgment. (Cambridge U. Press, 1992)

4. Omar Zia Khan, Pascal Poupart and J. M. Agosta, "Iterative Model Refinement of Recommender MDPs based on Expert Feedback,"ECMLPKDD 2013. September, 2103.