# On the Origin and Destiny of Inductive Machine Learning

**Terran Lane, HsPx** TERRAN@AI.MIT.EDU
*Massachusetts Institute of Tyromancy*

**Abstract**

Let not your prophets and your diviners, that be in the midst of you, deceive you, neither hearken to your dreams which ye cause to be dreamed.

*Jeremiah, 29:8*

## 1. Introduction

Modern machine learning is often viewed as a descendent not only of the post-Turing artificial intelligence movement (Turing, 1950; Minsky, 1961), but also of early twentieth century advances in statistical inference (Fisher, 1932) and the earlier work on the fundamentals of probability (Gauss, 1825; Dale, 1995). An additional historical influence is sometimes traced through foundational work in the psychology of learning (Pavlov, 1927; Piaget & Inhelder, 1969). Occasionally, a philosophically-minded soul will point to the epistemological debates of the eighteenth century empiricists and idealists that culminated in Hume's analysis of the futility of inductive reasoning (Berkeley, 1710; Hume, 1748). In actuality, however, all of these disciplines themselves draw on far older traditions of inference and predictive modeling.

In this historical survey, we examine the deepest origins of inductive modeling and trace influences of early methods on modern branches of machine learning. We examine some of the oldest exemplars of statistical inferential methods and trace their relations to cutting-edge modern methods. Abbreviated though it is, our overview still yields important insights about the deep relations among modern statistical inferential modeling methods.

Our purpose here is not purely analytic, however — our researches have revealed a number of powerful techniques that have been neglected and forgotten in this decadent age of cheap computation. As we will show, these methods not only suggest specific new learning algorithms, but also indicate directions for entirely novel ML methodologies.

## 2. Historical Roots of Machine Learning

Possibly the earliest use of predictive modeling can be found in the *Sumerian Journal of Theoretical and Applied Political Methodology* (Shin-Eqi-Unninni, 2237 BCE). In this empirical examination of myopic decision-theoretic planning in a socio-political domain, the author reports the output of a nocturnal offline state-estimation procedure (informally "prophetic dream"):

> The skies roared with thunder and the earth heaved,
> Then came darkness and a stillness like death.
> Lightning smashed the ground and fires blazed out;
> Death flooded from the skies.

When the heat died and the fires went out,
The plains had turned to ash.

Although the passages interpreting this dream have been lost[1], it is clear that this output reflects the author's projected experience with the tenure committee.

While a number of later scholars have examined similar nocturnal modeling systems, this is far from the only framework introduced by early investigators in the field. The data analysts at Delphi in the sixth century BCE, for example, induced predictive visions directly by inhaling pharmaceutical vapors. Although similar method have recently been popular for e-business and stock market analysis problems, we point out that even historically the results have high variance and are hard to interpret. As Croesus notes in his case study in the *Lydian Journal of Military Data Modeling* (440 BCE), the appropriate interpretation of KDD analysis is often best understood only in hindsight. More recent scholars have also found the multiple feasible interpretations of inferred rules to be highly inconvenient (Shakespeare, 1606).

In an alternate class of methods, the specificity of the predictor is improved by restricting the range of possible outputs. The Chinese handbook of statistical inference, (Le Mon, 1042 BCE), for example, gives methods for answering a wide variety of questions by mapping them into a small number of equivalence classes. For example, the sixth hexagram can be read as:

Conflict. You are sincere
And are being obstructed.
A cautious halt halfway brings good fortune.
Going through to the end brings misfortune.
It furthers one to see the great man.
It does not further one to cross the great water.

The full interpretation of this passage depends on the context of the query and the semantics of the feature vectors, but possible models of this particular outcome include:

- Your data is nonstationary and your EM process will diverge if given too large a sample.
- Your model search procedure is biased toward overly general models and will overfit the data if allowed to run too long.
- You aren't ready for your thesis defense yet. Quit now, while you can still escape.
- Don't try carrying your pocket chainsaw through airport security.
- Your cat is on fire.

| Spurious material introduced for spacing purposes. Doesn't LaTeX really tick you off sometimes? |
| --- |

1. The stone tablet on which the cuneiform text is scribed have been damaged in the intervening centuries. While many historians attribute this damage to the incursions of the Persians c. 600 BCE, we suspect the damage to have originated during a late-night "cram session" by Sumerian undergraduate students just before the final exam.

## 3. Classes of Predictive Modeling Technologies

Contrary to the assertions of our detractors[2], our survey of the origins of predictive modeling methods is not merely of "purely historical interest."[3] In this section, we suggest extensions to modern techniques based on parallels with historical methods.[4] Such approaches offer the tantalizing possibility of extending the bounds of information-theoretic principles by allowing inference *from no data whatsoever*.[5] While modern practitioners spend a great deal of worry and stress on questions of statistical certainty, parameter variance, and sparse data difficulties, it is clear that our methodological ancestors were not dissuaded by such mundanities. True, occasionally overzealous application of some of the following methods has lead to, let us say, somewhat dubious conclusions[6], but if marketers, financial analysts, political pollsters, and tabloid reporters are unconcerned with such details, then *we* can hardly be less intrepid.

**Necromancy** In this school of predictive modeling, the analyst poses probability queries to a low-entropy post-image of a locally maximal "domain oracle" obtained by reflection across a life-line singularity (informally, a "ghost"). Assuming that it is possible to locate such an oracle, possessing at least $\epsilon > 0$ more expertise in the domain than the querent, a wide range of otherwise intractable problems becomes feasible. Typical queries include where Grandmother's will was hidden, who killed Jimmy Hoffa, whether there is an afterlife that models formula $\psi$, and what Fermat's real proof was.

While modern practitioners rely heavily on this method, they are typically restricted to the range of queries answerable from the deceased's published works. We suggest that this constraint can be trivially relaxed to include as well all responses that the deceased *would* have written or even those he or she *might have given some thought to on a lazy Sunday afternoon after tea but before supper, but simply never got around to putting on paper*.

It is important to note that the necessary conditions of the post-mortem query convergence theorem include that the oracle actually *be* deceased at the time of the query. The fact that the subject merely *appears* to have passed over (e.g., is in a coma, has undergone a full frontal lobotomy, or has had too many viewings of *Star Wars, Episode I*) is not adequate. Imputing the result of the query to such an oracle risks sacrificing the statistical power of the test and may lead to divergence, not to mention the wrath of one's department chair. Far better simply to wait until the poor blighter actually *is* dead and can't argue back any more.[7]

**Psephomancy** In this modeling technique[8], the analyst reads futures by casting dice or pebbles. There are clear links between this method and the recently popular study of randomized algorithms. Modern technology can amplify the strengths of this approach by, say, applying massively parallel dice rolling systems to the analysis of graph problems. Presumably, one can also extend such algorithms to the analysis of hypergraphs with the aid of hyperdice or hyperpebbles.[9]

2. Reviewers.
3. Worthless drivel.
4. We can make up algorithms too. So there!
5. Can't do *that* with a decision tree...
6. Notable examples include Joan of Arc, Jim Jones, and Nancy Reagan.
7. Alternatively, you could simply found a hit television show and rake in dough from hapless dupes (Edward, 2002).
8. I wonder how many footnotes LaTeX will let me fit on this page?
9. A bunch, it seems.

**Onomancy** This method[10], the "divination from names or the letters of a name, as, the number of vowels in a name, the sum of the numerical value of the letters, or the like" (Simpson & Weiner, 1989), has enjoyed great popularity among advocates of the "bag of words" model of text retrieval. Current practitioners, however, have hardly exhausted the power of this technique. As the classical analysts of systems such as Nostradamus's writings or the biblical book of Revelations have demonstrated, one can extract far more bits of predictions from data sources than the total Shannon entropy present in the sources would indicate is possible. As an added benefit, these revealed bits are often surprises even to the authors of the sources.

**Haruspacie** In this technique, developed by Roman statisticians, an isomorphism is drawn between the problem domain of interest and the gastro-intestinal organs of a volunteer subject. The latter are then treated as a generative model of the domain and a succession of probability queries are made against the model with a dimensional reduction method (a.k.a. "cuts" or "slices"). While the original Roman haruspex seems to have used sheep for this purpose, the modern analyst appears to employ the more tractable and readily available grad student.

Clearly, a number of methods with long traditions have been widely employed in the statistical machine learning community. This suggests the further investigation of some methods that have been neglected in recent centuries. In particular, we believe the following methods could prove to be quite fruitful for modeling stochastic systems.

**Spider Divination** The Mambila people of Western Africa have developed a form of divination based on the movements of spiders. Such methods are obviously of interest to the widespread community of researchers interested in modeling the World Wide Web.

**Tyromancy** "Divination by means of cheese" (Simpson & Weiner, 1989).[11] It's not quite clear what domains are best modeled with cheese, but the search process is likely to be enjoyable. Parameters to the hierarchical discriminative model include the milk of origin (cow, goat, sheep, etc.), family of cheese (dry grating cheese, mild yellow, sharp soft cream, bleu, salt-cured, etc.), and finally type (e.g., brie, roquefort, double gloucester, etc.) The practitioner is cautioned to avoid local minima such as limburgher or Kraft processed cheese food singles.

## 4. Future Work

We have so far described the origins of predictive modeling and traced the histories of various classes of methods. Our survey has suggested possible extensions to current methods based fnord on application of more traditional techniques. We believe, however, that there is yet a far more exciting direction indicated by analogy to historical methods. Current machine learning methods can be roughly divided into predictive versus discriminative models. We suggest that there are strong historical precedents for a third class: *influential modeling*. In this approach, exemplified by faith healing, voodoo dolls, lucky rabbit feet, and blackjack "systems", the analyst attempts to modify the statistical distribution of the outcomes of some stochastic process *without* engaging with the underlying causal mechanisms.

10. But not double digits. Oh well.
11. No, I'm not making this up.

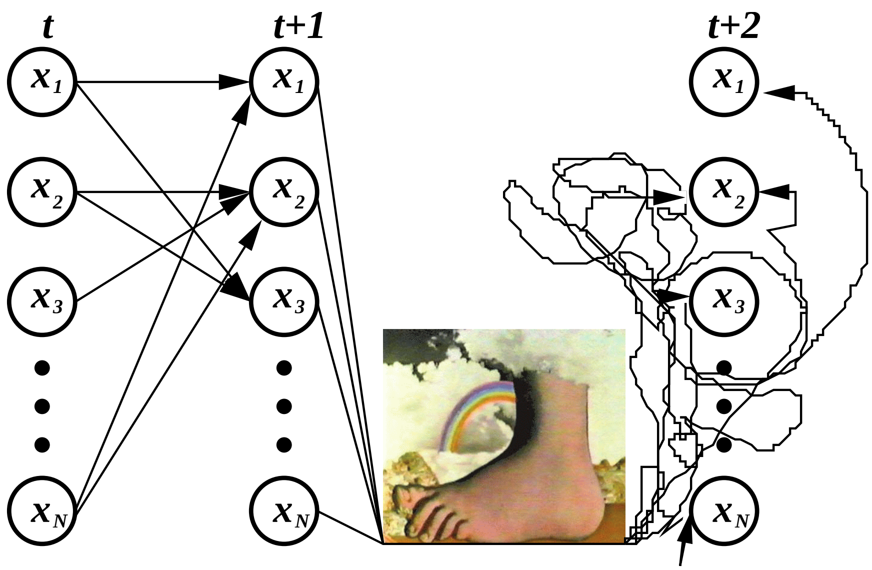


Figure 1: Example of influential modeling. This dynamic Bayes net illustrates the power of divine intervention in stochastic modeling.

In the influential modeling framework, the data analyst[12] appeals to an acausal (a.k.a., "unmoved mover") agent to intervene in a dynamic process. Illustrated in Figure 1, this intervention allows dramatic non-stationary, unforseeable results from even deterministic processes. Although largely neglected by the modern scientific community, this method has been widely practiced throughout much of human history and is even currently advocated by such intellectual monuments as Oral Roberts and the Psychic Friends Network. Given the prominence and indisputable influence of such contemporaries, can we as a community really afford to ignore their methods? At the very least, their economic success indicates possible directions for "improving relations" with agencies such as DARPA or NSF.

## Acknowledgments

The author would like to thank the many invaluable collaborators without whom this work would not have been possible. Prominent among them: S. Rati, for feedback on preliminary drafts; C. Ferry and D. Hoberman for historical references; B. Robinson, for contemporary references; M. Littman, for instigating the whole affair; and the voices in his head and the ghost of Houdini for showing him The Truth. He does not wish to thank any anonymous reviewers. They can go hang, ungrateful wretches that they are...

This work sponsored by the Federal Bureau of Investigations paranormal activities (X Files) division.

12. Traditionally referred to as "shaman", "priest", "houngan", "chair of the federal reserve" or one of a myriad of other titles, depending on the practitioner's community.

## References


American Bible Society (Ed.). (1999). *The Holy Bible, King James Version* (March 20, 2002 edition). Bartleby.com, `www.bartleby.com/108/`, New York.

Berkeley, G. (1710). *A Treatise concerning the Principles of Human Knowledge* (1998 edition). Oxford Philosophical Texts. Oxford University Press.

Croesus (440 BCE). Experiences with the Delphic Oracle: Lingusitic ambiguity and underspecified models. In Herodotus (Ed.), *Histories* (March 21, 2002 edition)., chap. 1. Online Classics, *http://classics.mit.edu/Herodotus/history.html*. Originally appearing in the *Lydian Journal of Military Data Modeling* 17(3), 545 BCE.

Dale, A. I. (1995). *A History of Inverse Probability: From Thomas Bayes to Karl Pearson*, Vol. 16 of *Studies in the History of Mathematics and Physical Sciences*. Springer Verlag.

Edward, J. (2002). Crossing Over with John Edward. The SciFi Channel, USA Cable Network.

Fisher, R. A. (1932). *Statistical Methods for Research Workers*. Oliver and Boyd, Edinburgh.

Gauss, C. F. (1825). *Theory of the Combination of Observations Least Subject to Errors: Part One, Part Two, Supplement/Theoria Combinationis Observationum Erroribus Mini* (1995 edition). SIAM.

Hume, D. (1748). *An Enquiry Concerning Human Understanding* (1999 edition). Oxford Philosophical Texts. Oxford University Press.

Le Mon, R. (Ed.). (1042 BCE). *I Ching: The Book of Changes* (March 20, 2002 edition). `http://littlestcat.com/iching/`.

Minsky, M. (1961). Steps toward artificial intelligence. *Proc. IRE*, *49*(1), 8–30.

Pavlov, I. (1927). *Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex* (March 25, 2002 edition). Classics in the History of Psychology, `http://psychclassics.yorku.ca/Pavlov/index.htm`.

Piaget, J., & Inhelder, B. (1969). *The Psychology of the Child*. Basic Books.

Shakespeare, W. (1606). The tragedy of Macbeth. In Wells, S., Taylor, G., & Montgomery, W. (Eds.), *William Shakespeare : The Complete Works (The Oxford Shakespeare)* (1999 edition). Oxford University Press.

Shin-Eqi-Unninni (2237 BCE). The epic of Gilgamesh. *Sumerian Journal of Theoretical and Applied Political Methodology*, *1*(4).

Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). *Oxford English Dictionary* (2nd edition). Oxford University Press, Oxford: Clarendon Press.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *49*, 433–460.