Prediction Of Human Behavior

William C. Cave
Prediction Systems, Inc.
Spring Lake, NJ 07762

John H. Fikus, PhD
Prediction Systems, Inc.
Wall NJ, 07719

ABSTRACT

Prediction of human behavior can be approached in various ways. A recent study and overview of this topic is contained in the report to the National Research Council, [9]. Our goal was to determine appropriate methods that are suitable for modeling the varied characteristics of this behavior. For example, humans engage in extremely complex decision processes. Most of these involve discrete choices that are not easily described in common mathematical terms. We wanted to investigate methods that support characterization of these behaviors. We also wanted to determine their suitability for development of models that can evolve as we learn more about this ancient yet currently popular topic. This imposes considerations regarding model architectures that can be expanded. Lastly, recent research indicates that we must provide for incorporating knowledge by subject area experts who are schooled in neither modeling nor mathematics. Restated, our goal was to provide an intuitive environment for modeling that eases the burden of capturing expert knowledge while providing a framework for easily expanding models as this knowledge base is increased.

Keywords: Prediction, Modeling, Human Behavior, Control Theory, AI, Expert Intelligence.

STATEMENT OF THE PROBLEM

We are concerned with predicting behavior of an individual or group, where groups can be a small cell or large society. We treat each of these as a *system*.

A vector of observable response data, Z, can characterize the system behavior we want to predict. The components of this vector can be numbers or discrete characterizations, e.g., colors (red, yellow, green), sizes (large, medium, small), vehicles (train, bus, car, plane), etc. The set of possible states must be defined.

We can also observe influences on the system characterized by a vector U. The components of U can also be numbers or discrete characterizations as with Z.

To characterize the prediction problem, we will use a model of the system as shown in Figure 1. We denote the observable *system response* at a discrete time T by the vector Z.

$$Z(T) = [Z_1(T), Z_2(T), \ldots, Z_M(T)] \qquad (1)$$

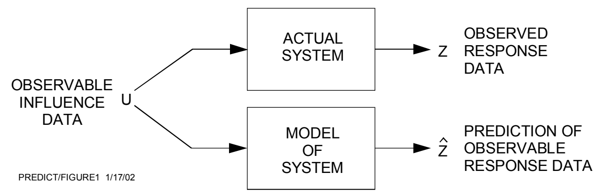

Figure 1. General form of a prediction model.

We call the vector, U, of influences at time T the *driving force* vector.

$$U(T) = [U_1(T), U_2(T), \ldots, U_K(T)] \qquad (2)$$

To be useful in predicting the future values of Z, the components of U must be directly observable, and must affect the response. Typically the driving force is unpredictable. Otherwise, it could be incorporated as a response to another driving force with a further lead, or be treated as a known function of time.

HISTORY AND TIME HORIZONS

Refer to Figure 2. Given the current time T, we want to predict the observable Z at time T+τ, i.e., $\hat{Z}(T+\tau)$, where τ takes on the values 1, 2, …, T_P, where T_P is the maximum time horizon for predictions ($T_P = 12\ \tau$ in the example).

Our goal is to develop a model that operates on the driving force vector, U, to produce sufficiently accurate predictions $\hat{Z}(T+\tau)$. The accuracy requirement is defined by the application. Its measure is defined, implicitly if not explicitly, by a set of probability statements. For example, if vector Z is to characterize an observable that takes on discreet values, then a prediction is composed of the probability that it will take on a particular discrete value. The components, z_1, z_2, and z_3, can be real numbers representing the probabilities that selected discrete values will occur. Depending upon the technique used to measure accuracy, it may be necessary to state the confidence in the probability statement.

If component z_1 takes on real values, then a prediction is composed of upper and lower limits on the value of z_1, shown in Figure 2, and the probability that z_1 will fall within those limits. Again, depending upon the measurement technique, a confidence statement may be necessary.

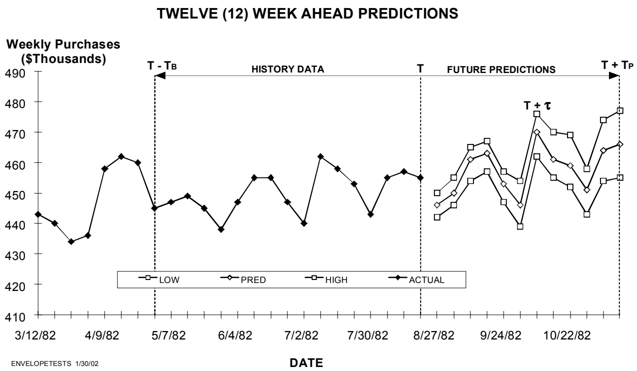

Figure 2. Illustration of prediction data.

To characterize the statistics required to produce the probability statements, the following definitions are offered.

T_P - Number of future time steps from the current time step to the future time horizon for which the system response is being predicted.

T_B - Number of past time steps from, and including, the current time step to the looking back horizon, used to define the probability statements.

N_S - Number of mutually exclusive "T_B" sample sets (ensembles) of history data available for testing the probability statement.

In other words if, T_T is the total number of sample points of history data, then

$$N_S = \frac{T_T}{T_B}. \tag{3}$$

Multiple sample sets are needed to identify model parameters used to maximize prediction accuracy. This is because, once a set of history data is used to optimize the model, it can no longer be used to measure prediction accuracy, see [2]. One must use a new “hidden” set to perform this measurement. Depending upon the model, some portion of the data may be needed to initialize the internal model memory. This makes it difficult to identify model parameters when there is very little data. This is especially important when dealing with neural net type models requiring reasonable amounts of training data.

It should be noted that we assume the systems we are modeling are bounded based on the concept described in [2]. This implies that measures consist of a finite number of sample points, taken at discrete times, whose values are bounded. We also assume that the system is stable, i.e., bounded inputs yield bounded outputs.

GENERAL APPROACHES TO MODELING

The prediction problem has been shown to be difficult, [2], and often misunderstood, [1]. Constructing models that accurately reflect system behavior generally fall into two categories: those that are naïve and those that are structured using the “physics” of the system. Naïve models use generic approaches, e.g., linear regression, and are not derived from physical properties of the system. The history data, and particularly the response data, is used to determine model coefficients or parameters using some form of optimization. In the case of neural nets, this is called a training period.

Structural models are built based upon prior knowledge of the internal mechanics and dynamics of the physical system. This knowledge comes from those contributing to the modeling process. They may not look at the data until the model is to be tested. The equations of physics are structural models. An example is Einstein’s model of light rays bending around the sun. His predictions were made years in advance of the data becoming available.

Structural models are very useful when there is very little data available. However, even structured models can have a naïve (empirical) component to account for behavior that cannot be characterized in terms of physically justifiable mechanics. This is typically introduced using parameters within the structure that are optimized to maximize prediction accuracy.

COMPARING MODEL PREDICTION ACCURACY

Models of a system can be compared in terms of their accuracy based upon measures of error. This can be accomplished using a measure of the sequence of differences between predictions and observed values of the response. A convenient measure uses the sequence of normalized residuals up to T,

$$RN[\hat{Z}(T+\tau)] = \frac{\hat{Z}(T+\tau) - Z(T+\tau)}{Z(T+\tau)}, \tag{4}$$

over the period from the looking back horizon, T_B. This measure is denoted by ε_Z:

$$\varepsilon_Z = E\left\{RN[\hat{Z}(T+\tau)]\right\} = \frac{1}{T_T - T_B} \bullet \sum_{T = T_B}^{T} \left| \frac{\hat{Z}(T+\tau) - Z(T+\tau)}{Z(T+\tau)} \right| \tag{5}$$

To compare model accuracies, one can compute the error statistics for the above measures using data that has not been used to build the models. If the data has been used to build the models, then one is comparing how well the model fits the history data, not how well it predicts the future, see [2].

A GENERIC MODELING FRAMEWORK

Physicists and engineers responsible for designing airplanes, power generators, and missile guidance systems have a common goal to “get it right”. They are aware of the potential cost of failure if they don’t. They are experienced in the modeling process, and how to ensure reduction of potential error in their designs as well as in system operation. The State Space framework has been evolved by these experts for characterizing dynamic systems.

The State Space Framework

The State Space framework, shown in Figure 3, is commonly used in engineering and physics, [1], [3], [4], [6], [7], [15]. It has been shown to encompass the most general modeling problem, see for example Gelb, [5], or Schweppe, [12]. It provides an excellent framework upon which to develop and evaluate models. This framework has evolved to describe discrete event systems as described in the next section.

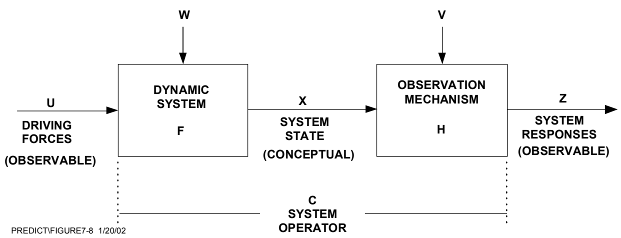

Figure 3. The State Space framework.

We start with basic definitions. The *state* of a system is defined as a set of properties that, along with the input driving forces to the system, are sufficient to describe the dynamic behavior of the system:

$$X(T) = [X_1(T), X_2(T), \ldots, X_N(T)] \tag{6}$$

where X(T) is a vector valued function of time in some n-dimensional space. Some of these properties may be observable, but *none* need be. The important criterion is to select a set of properties which simplifies the modeler's conceptual view of the internal "mechanics" of a system, i.e., how its components operate, causing the system to change with time.

We note that W and V are covariance matrices used to represent unknown random variations in the model and observation mechanism. These can be characterized using a Kalman filter, [2], [5], [6], [7], [12], to provide a maximum likelihood estimate of the current state, X(T), of the system. For the purpose of this discussion, we can ignore these.

In many cases, the conceptual properties of the system cannot be measured, at least for economic reasons. For example, we can envision a market as being composed of a mass of people who enter the "market place" upon making a decision to buy or sell. Upon striking a deal which satisfies their desire to buy or sell, they leave the market place. We can write the "equations of motion" which describe their rate of entry, their number at any time, and rate of departure based on external influences. Whether we can observe these properties directly is unimportant, as long as we can relate them to things we can observe, such as high price, low price, and volume of trading for the time period of interest. The objective is to predict $X(T+\tau)$, the *state* of the system at a future time step.

To this end, a dynamic model of the form

$$X(T+1) = F[X(T+1), X(T), U(T), T] \tag{7}$$

is used. Thus, the next state of the system can depend upon itself X(T+1) (i.e., it is nonlinear), the current state X(T), the external influences or driving forces U(T), and directly upon time, T.

Once a set of attributes representing the state of a system has been selected, the modeler can describe the state transition process in terms of its causes and effects. To do this, the modeler must describe the conceptual relationships perceived to exist in the system. These relationships, which must be described by the modeler, typically represent significant additional information about the structure of a system, which can lead to a corresponding improvement in model accuracy.

The convenience of using the state space framework comes about by a separation of the observation mechanism, H, from the conceptual dynamics of a system. It is this separation of conceptual variables, X, from observations, Z, which affords the modeler a powerful tool for mathematically formulating his conceptual knowledge about the structure of a system.

If the observation vector, Z, can be derived from the state vector, X, at any time T via a relationship of the form:

$$Z(T) = H[X(T), T], \tag{8}$$

then, given the prediction of X(T+1) from our dynamic model Eq. (7), we can determine $\hat{Z}(T+1)$. In addition to being a general formulation for dynamical systems, experience has shown that this separation of observation from concept allows the modeler to more easily translate his knowledge of system structure into an algorithmic representation.

In future sections we will have cause to view Eq. (7) and Eq. (8) as a single transformation, C, denoting the relationship between the driving force vector at time T, and the observation vector at time $T+\tau$.

$$\hat{Z}(T+\tau) = C[X(T), U(T)] \tag{9}$$

We will refer to C as the *system operator*, reference Figure 3.

APPROACHES TO MODELING SYSTEM BEHAVIOR

Although we have described our model of a dynamic system using an equation, Eq. (7), there are many ways to obtain the transformation of driving forces, U, into the next state, X. For example, one can write algorithms containing rules: IF this occurs… , THEN do that … . Such algorithms can even be in the form of neural networks. In general, the dynamic system, F, can be a large model, composed of many submodels, working together to produce the desired transformation. This approach does not need a constant discrete time-base, but can move ahead in time based upon discrete events, see for example [10], and [13]. An illustration of interconnected models is shown in Figure 4. Each model can run independently, sharing data.

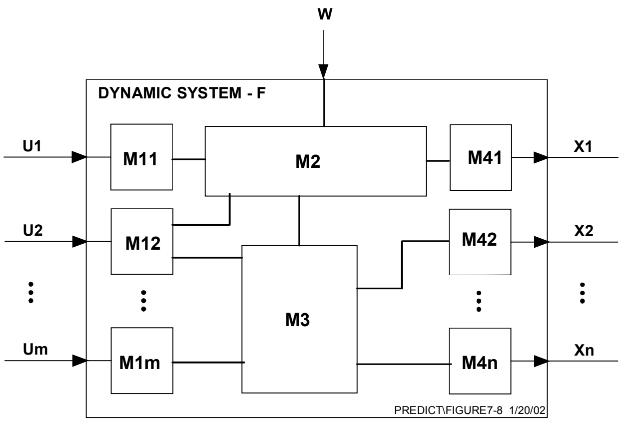

Figure 4. Illustration of a dynamic model.

Expert Intelligence (EI) Models

An exploratory simulation was built (details available separately) that can be used to predict the aggregate behavior of consumers purchasing food. This simulation is built using a discrete event environment based upon the state space framework, but is more general. It provides for vector spaces containing discrete states that can be described by words as well as numbers. Transformations can be generic rules, not restricted to mathematical operators. This permits an IF *this* …., THEN *that* …., ELSE …. format for decision rules. This format supports direct translation of subject expert knowledge into model process rules. The use of subject area experts is stressed in recent literature on modeling human behavior, particularly in the military environment.

Prediction Systems, Inc. has developed a Computer-Aided Design (CAD) technology, the General Simulation System (GSS), for building large-scale simulations. GSS provides an environment where subject area experts can read the rules written in a process language without knowledge of computer programming. As opposed to a Program Design Language (PDL) or Universal Modeling Language (UML) approach, this language is translated into machine code. Models are built and maintained directly in this environment.

The GSS "consumer" simulation contains two models and an instrument that measures the results as the simulation runs. The models were derived based upon the modeler's basic knowledge of how decisions are made by both the consumer and the store manager. We note that there are no outside influences on the simulation. This is defined as a homogeneous model, implying there is no driving force vector, U.

If we wanted to investigate the effect of losing a store, e.g., due to a fire or terrorist attack, we could create an external force that closes a store at some time during the scenario. This would be an external force. Likewise, we could investigate what would happen if a store closed and reopened over different periods of time due to labor strikes or other external forces. These would result in nonhomogeneous models.

One can consider the consumers and stores to be part of a much larger community. In this case, the consumer model is just one of a larger set of models of community members that makes observations, makes decisions, and takes actions on many other aspects of life. One can see how other aspects of a community can be modeled by having people who are experts in those aspects contribute to the model parameters and rules that implement the decision process.

Artificial Intelligence (AI) Models

We now consider the use of Artificial Intelligence (AI) approaches. Figure 5 illustrates a neural net using driving forces to produce a prediction. This approach falls into the general category of pattern recognition. Neural nets can be "trained" to recognize patterns. This has been shown to be a fast way to identify objects, moving in 3-D, by their shapes. It can work well even when object images are fuzzy. If patterns are recognized in the input, U, over some observation period, then a signal is produced indicating what patterns occurred.

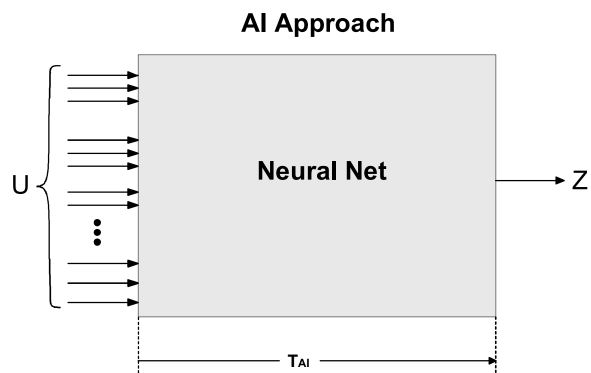

Figure 5. Illustration of an AI approach.

This approach has been tried in schemes for predicting the stock market. The simplest case is when the history of Z itself is used to predict its future path. This implies that the data contains recognizable pattern repetitions. This, in turn, implies that a recognizable component of Z is stationary. From a mathematical standpoint, stationary or quasi-stationary systems can be represented by a homogeneous model, see [2].

Another good example is encryption. Cracking codes is a pattern recognition problem. AI approaches have been applied for years. However, recent encryption techniques apparently render the required effort to be economically overwhelming.

The more interesting case is recognition of nonstationary patterns in other data sets, U, that occur in advance of patterns in Z. These could be predictive, but require a nonhomogeneous model. This has been tried in the stock market where daily data has been recorded for publicly traded stocks for over 70 years. An enormous database of history exists for training neural nets. One would think this to be a powerful approach. However, it has not been shown to be successful in any scientific literature, and to the best of our knowledge, companies that have touted the potential future of these approaches have not demonstrated any extra-ordinary success over the long term.

EI - AI Hybrid Models

On the other hand, experience has shown that, using small amounts of data with no repeating patterns, models can be built that provide very accurate predictions, see [2]. These are developed using expert knowledge of the mechanics of how driving forces affect the system. Small amounts of data are sufficient to characterize one or two model parameters. In these cases, the systems can be highly nonlinear. Yet, using nonlinear models, one can provide accurate predictions. We will consider the use of such models using a hybrid approach, see Figure 6.

Prediction Systems, Inc. (PSI) has participated in a number of AI projects, applying various techniques to manage huge databases, perform nonlinear optimization, and support jammer management. One of the projects entailed the use of neural nets to detect intrusions in the US Army Tactical Internet (TI). Briefly described, the approach maps AI sub-blocks into network components as they are organized, hierarchically, distributing the computational load. The hierarchy reduces the bandwidth required to communicate accurate and timely views of the network at all tiers.

More importantly, PSI's experience with this application, as well as neural nets and adaptive algorithms in other projects indicates that the learning and processing times for adaptive algorithms, T_{AI} in Figure 6, can be reduced considerably through smart preprocessing of information known in advance. This is done using the EI models. EI models are quite simple and fast relative to AI approaches, rendering T_{EI} very small compared to T_{AI}. This greatly reduces the overall processing to be done in real time. It also reduces the amount of observable data and training required relative to the pure AI approach.

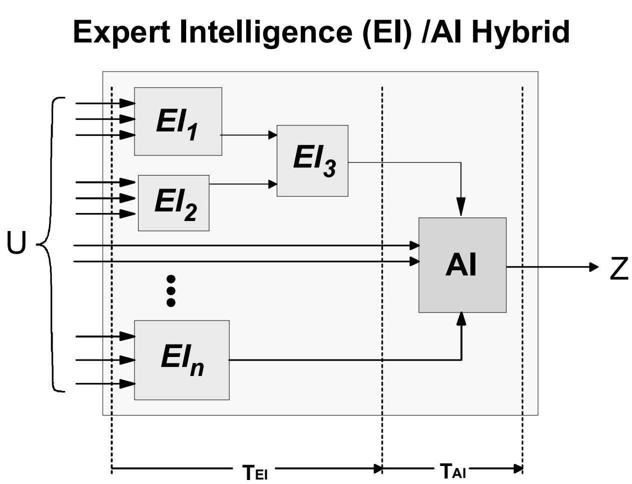

Figure 6. Illustration of a hybrid EI - AI approach.

EI Models With Optimization

Another approach to modeling system behavior uses optimization to identify parameters in an EI model. This is illustrated in Figure 7. The optimization process can be applied off-line, or adaptively in real time. This is the approach found to be most useful when modeling decision processes for which expert knowledge can be introduced.

Using this approach, expert human knowledge is incorporated into the EI models shown in Figure 7. Unknown parameters are used to account for lack of knowledge. However, these parameters are selected judiciously. Their placement in the model is determined based upon where information is lacking. Often, one has reasonable knowledge of ranges on these parameter values. Any piece of additional information that can be used to create the model cuts down on the size of the unknown space, leading to a faster solution.

First hand experience on many projects clearly demonstrates that experts are aided significantly when observing the models as they behave in a simulated environment. This generally leads to significant improvements in representation of the decision process.

The optimization system used by PSI is built into the GSS environment. It uses an EI approach, having adaptive algorithms that automatically formulate ensembles of data to generate distributions used to update the search process. This system has been used to solve a wide variety of highly nonlinear problems, such as finding the best location for antennas in a hilly environment under threat jamming, or finding optimal flight paths for ELINT or SIGINT collections, accounting for threat air defense systems.

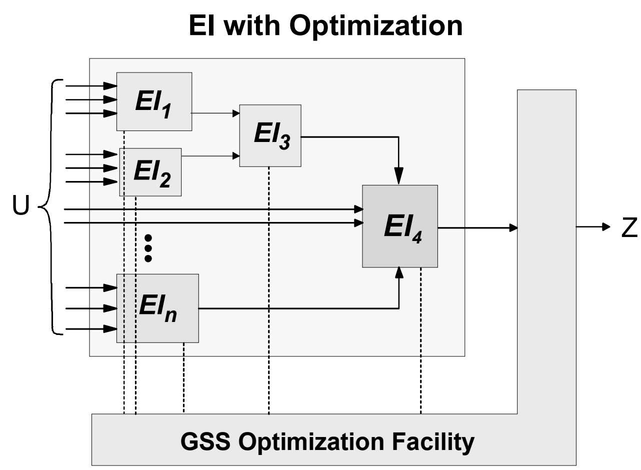

Figure 7. Illustration of an optimized EI approach.

As an example, the decision models for the consumers and store managers in the reference simulation can be optimized to meet their objectives. The consumer generally wants to minimize his cost of purchases. The store manager generally wants to maximize his profits. Alternatively, given that one has observations on the quantities purchased or inventoried, parameters in the model can be optimized to match the observed data.

Parametric And Sensitivity Analysis To Support The Modeling Process

Another means of preprocessing to develop models using Expert Intelligence is by using simulation to support model analysis. For example, we can generate distributions of responses by running a sufficient number of simulations while varying parameters to determine if model results fall inside sensible ranges.

ACCURACY CONSIDERATIONS

As indicated above, additional accuracy can be obtained by modeling the effects of driving forces. To be useful, they must be observable and "lead" the response. Otherwise they contribute no additional information for improving prediction accuracy. In particular, we are concerned with the description of models which relate system responses to nonstationary driving forces, see [2]. These relations can be highly nonlinear and difficult to model if one does not understand the mechanics of the system. Incorporation of these effects can lead to significant improvements in model accuracy.

These two aspects of a model,

- Expression of the structural properties of a system
- Modeling the effects of driving forces

represent *additional information* that is generally *not contained in the response data*. This is particularly true when a system is nonlinear.

Finally, we want to develop complex models without violating rules of parsimony, see Tukey, [20]. When using methods where the structure of a system is ignored, and a naive approach is pursued for model identification, then unknown parameters are used merely to *fit* the response data. In this case, the modeler should be concerned about parsimony. This is because additional coefficients add no information to condition the probability statement and contribute nothing to accuracy. In fact, if they contain noise, they can decrease accuracy. However, if a model is enhanced by the benefit of additional knowledge of the structure of the mechanics, then these additions will serve to condition the probability statement so as to be more accurate, and the constraints of parsimony do not apply.

A purchaser of predictions will judge one model to be superior to another if it provides him with consistently more accurate predictions of the future. On this basis, there are many examples in engineering (e.g., modeling of integrated circuit chips) where models have been carefully constructed based on knowledge of the physical structure and mechanics. The complexity of these models would appear to violate rules of parsimony as advocated by many statistical forecasters. Nevertheless, these models have provided outputs yielding excellent consistency with test results long after model development.

SUMMARY

Recent literature on modeling human behavior stresses the use of subject area experts to devise decision rules for modeling that behavior. This leads to the desire for direct translation of rules described in a natural language. The use of discrete event systems and simulation technology provides the framework for satisfying this goal. This approach is termed Expert Intelligence (EI).

Also investigated are Artificial Intelligence (AI) approaches. They are compared to the EI approach. AI techniques are also combined to form an EI-AI Hybrid, and compared to the AI and EI approaches. Finally, an EI Optimization approach is discussed where EI techniques are built into the optimization process.

A Computer-Aided Design (CAD) technology for building large scale simulations has been used to model human behavior. This technology, the General Simulation System (GSS), allows subject area experts to read rules written in a process language without knowledge of computer programming. As opposed to a Program Design Language (PDL) or Universal Modeling Language (UML) approach, the English-like rules are automatically translated into machine code. GSS also contains built-in optimization facilities that make the EI Optimization approach easy to use. Optimization criteria can be automatically derived from graphical interfaces tailored to the problem being addressed.

The EI approach with optimization supports the requirements for incorporating subject area expertise into the modeling process. It also provides a framework for building independent model architectures that can grow large, hierarchically, as more knowledge of the behavior mechanisms is incorporated into the models.

Clearly this investigation merely scratches the surface on modeling to predict human behavior. However, the acceleration of technological advances in related areas should serve to aid in this immense endeavor.

REFERENCES

[1] Athans, M. and Kendrick, D., "Control Theory and Economics: A Survey, Forecast and Speculation", **IEEE Trans, Automatic Control**, pp. 518-524, Oct. 1974.

[2] Cave, W.C., Prediction Theory - A Framework for Modeling, Prediction Systems, Inc., Spring Lake, NJ, 2001.

[3] Cave, W.C. and Rosenkranz, E., "A Stochastic State Space Model for Prediction of Product Demand," **AFIPS-Conference Proceedings, Volume 48**, AFIPS Press, pp. 67-72, 1979.

[4] Cave, W.C. and Guilfoyle, R.H., "A General Approach to Supply/Demand Market Prediction Using the State Space Framework". **First International Forecasting Symposium, Quebec**, CD, 1981.

[5] Gelb, A., Editor, **Applied Optimal Estimation**, MIT Press, Cambridge, MA, 1974

[6] Kalman, R., "A New Approach to Linear Filtering and Prediction Problems", **Trans. ASME, Series D: Journal Basic Engineering 82**, pp. 35-45, 1960.

[7] Mehra, R., "Kalman Filters and Their Applications to Forecasting", **TIMS Studies in the Management Sciences 12**, North-Holland, 1979.

[8] Papoulis, A., Probability, **Random Variables, and Stochastic Processes**, McGraw-Hill, New York, NY, pp. 554, 1965.

[9] Pew, R. et al, **Modeling Human and Organizational Behavior**, National Academy Press, Washington, DC, 1998.

[10] Wong, K.C. and W.M. Wonham, "Modular Control and Coordination of Discrete Event Systems", **Discrete Event Dynamic Systems: Theory and Applications**, 6:241-273, July 1996.

[11] Rosenkranz, E., "Robust Estimation of Time Varying Parameters", **Third International Symposium on Forecasting Program**, Philadelphia, PA. pp. 150, 1983.

[12] Schweppe, F., **Uncertain Dynamic Systems**, Prentice Hall, Englewood, NJ, 1973.

[13] Stadter, P.A., Discrete Event Command and Control Systems, **13th Annual AIAA/USU Conference on Small Satellites**, 1999.

[14] Weinberg, G., **An Introduction to General Systems Thinking**, John Wiley & Sons, NY, NY, 1975.

[15] Zadeh, L.A. and Desoer, C.A., **Linear System Theory: The State Space Approach**, McGraw-Hill, NY, 1963.