Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Predicting System Collapse: Two Theoretical Models

Pouyan Hosseinizadeh
Mechanical and Industrial Engineering Department
Ryerson University
Toronto, Canada
pouyan.hosseinizadeh@ryerson.ca

Aziz Guergachi
Ted Rogers School of Management
Ryerson University
Toronto, Canada
a2guerga @ryerson.ca

Vanessa Magness
Ted Rogers School of Management
Ryerson University
Toronto, Canada
vmagness@ryerson.ca

***Abstract*—Making precise predictions about the future behavior of a system such as a country's economy, a firm or a lake, or about the population of some species of animal has always been a challenge. While prediction methods and modeling procedures have been developed and used over the past decades, the high degree of uncertainty and complexity that underlie some systems makes it difficult, and in some cases impossible to exactly predict the next states of the system. The purpose of this paper is to present two approaches for identifying potential system Collapse. The first approach is inclination analysis, which examines the state of a system over several windows of time in an effort to predict the final inclination. The second one is based on Support Vector Machines and Kernel methods. Various applications of these approaches as well as their advantages and limitations are also discussed.**

***Keywords*—SVM, Inclination Analysis, modeling, prediction, Ecosystem , Financial Systems**

I. Introduction

Many research projects have attempted to increase the ability of the existing modeling tools to identify the mechanisms that underlie different systems, to analyze their interactions with other systems, and to predict the next state of the system. The objective of this prior research was to predict the terminal state of a system in terms of collapsing or surviving so as to avoid system Collapse in the future. Despite the availability of alternative tools for system analysis and modeling, it is often impossible to learn enough about the system so that models can be designed that product sufficient output. There is also the problem that considerable resources may be directed toward identifying system behavior which is itself in a state of flux. There is the risk that by the time the past behavior has been successfully modeled, that behavior has changed. This is particularly true for the more complicated systems. This paper discusses two approaches that aim to identify potential system Collapse. These approaches are

- Inclination Analysis, and
- Support Vector Machines.

Inclination Analysis [1] which was proposed by Kryazhimskii and Beck (2002) is a strategy to assess the Collapse potential of a system. The observation of historic local changes in the system's behavior is the key to this approach. Inclination Analysis asks the question: "Does access to a brief window of observed local changes indicate the system is tending toward Collapse or Survival in the future?" This method, which is discussed below in section II, has been used on both environmental [1] and financial [2] systems.

Statistical Learning Theory [3] has emerged as a useful tool for model construction in a situation of high uncertainty since the last decade. It has the advantage of concentrating on data sets that are made of system inputs and outputs for model construction and prediction. The modeling procedure treats the system under study as a black box. In this paper we discuss one of the major by-products of Statistical Learning Theory called *Support Vector Machines* (SVM). SVM is a popular supervised learning method [4] widely used for a variety of applications such as speaker verification [5], online handwriting recognition [6] & [7], and financial forecasting, [8]. We discuss SVM below in section III, along with Fisher kernel. In section IV we discuss the advantages and disadvantages of each method.

II. Inclination Analysis

A. Problem Definition

This method of analysis involves studying a system's local changes over sequential periods of time. The model uses a binary classification of desired, (+) or undesired, (−) behaviors to indicate whether the system is heading toward Collapse or Survival. The periods of time can be days, weeks or years,

978-1-4244-2794-9/09/$25.00 ©2009 IEEE

depending on the particular system. The system's final fate is assessed using the theory of path dependent stochastic processes. One must begin with a clear understanding of the characteristics of the system under study in order to identify the desired and undesired behaviors, and to understand the marginal behaviors that signify the positive or negative direction of change. It is assumed, for the purpose of this paper, that this knowledge has already been obtained. In general, there are two essential assumptions to be made about the system:

Assumption 1: The fate of the system in the distant future is either Collapse or Survival.

Assumption2: Local changes in time towards Collapse or Survival are positively correlated with changes in the immediate past.

B. Model parameters

A number of parameters must be defined before the binary model can be constructed. Time periods over which the data are available are denoted by k. Local changes over a single period of time are either $+$ (toward Survival) or $-$ (toward Collapse). These changes are called *transitions*. As stated earlier, it is assumed that procedures are available to identify the $+$ or – transitions. In each period k, a finite string s of fixed length L is used to characterize those features of the system's past that have impact on the system's local change between period k and $k+1$. In fact, the model has the Markov chain property, which means that, at each period k, the current state of the system only (not the past states) is enough to make decision about the next transition to period $k+1$. String s is a combination of $-/+$ transitions realized sequentially between each period from $k-L$ to $k-1$. String s is defined as the state of the model that gives rise to the model's transition from $k-1$ to k. Therefore, the model's state is the L-long moving window always preceding the latest – or $+$ transition. This model is illustrated in figure 1.

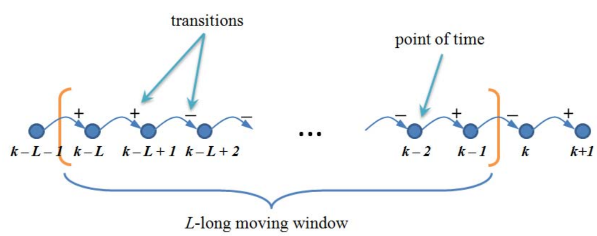

Figure 1. ***L***-long moving window indicating the system's state between period $\boldsymbol{k-L}$ and $\boldsymbol{k-1}$

Based on Hypothesis 2 a strong dominance of the $-$ or $+$ transitions in the string s causes a $-$ or $+$ at the next transition. The probability of having $-$ for the next transition is defined to be r^-. Clearly r^-can accept any value between 0 and 1. The probability of the next transition being $+$ is defined by r^+. Since r^- and r^+ are complementary ($r^-+r^+=1$), r^- is considered in operations. Table 1 illustrates the relation between the $-$ or $+$ dominance and the probability of having $-$ or $+$ as the next transition.

$-$ or $+$ dominance	Probability of having $-$ (r^-)	Probability of having $+$ (r^+)
$n^- \geq m^-$	$r^- = 1$	$r^+ = 0$
$n^+ \geq m^+$	$r^- = 0$	$r^+ = 1$
$n^- < m^- \& n^+ < m^+$	$r^- = r$	$r^+ = 1-r$

TABLE 1. n^- and n^+ are the number of – and + in one state respectively. m^- and m^+ are some critical numbers. r^- and r^+ are the probability of next transition being $-$ and $+$ respectively

Once L, the state's length, is fixed with respect to the binary model, then in each period of time k the system can be in one and only one of the 2^L different possible states. So the matrix of transition probabilities is defined as Z and is a square matrix of size $2^L \times 2^L$ as follows:

$$Z = \begin{array}{ccccc|c} ++\cdots++ & ++\cdots+- & \cdots & --\cdots-+ & --\cdots-- & \\ r & 0 & \cdots & 0 & 0 & ++\cdots++ \\ 1-r & 0 & \cdots & 0 & 0 & ++\cdots+- \\ 0 & r & \cdots & 0 & 0 & \vdots \\ 0 & 1-r & \ddots & 0 & 0 & \vdots \\ 0 & 0 & \cdots & r & 0 & \vdots \\ 0 & 0 & \cdots & 1-r & 0 & \vdots \\ 0 & 0 & \cdots & 0 & r & --\cdots-+ \\ 0 & 0 & \cdots & 0 & 1-r & --\cdots-- \end{array} \quad (1)$$

The elements of the Z matrix reflect the probability of having – transition, i.e. r^- from one possible state to another. To assess the final fate of the system by having it operate over a very large number of time periods, the calculation of Z^∞ is required. This probability matrix then can be expressed as

$$Z_{inf} = \lim_{k\to\infty}(Z)^k \quad (2)$$

if the limit exists. When the system is in period k, the probability of it being in each of the possible states can be shown as a $2^L \times 1$ matrix as

$$P(k) = Z^{k-1}q = \begin{bmatrix} P_{++\cdots++}(k) \\ P_{++\cdots+-}(k) \\ \vdots \\ P_{--\cdots-+}(k) \\ P_{--\cdots--}(k) \end{bmatrix} \quad (3)$$

where q is the initial probability of the state s at $k=0$. It can be shown as a $2^L \times 1$ matrix:

$$q = \begin{bmatrix} P_{++\cdots++}(0) \\ P_{++\cdots+-}(0) \\ \vdots \\ P_{--\cdots-+}(0) \\ P_{--\cdots--}(0) \end{bmatrix} = \begin{bmatrix} q_{++\cdots++} \\ q_{++\cdots+-} \\ \vdots \\ q_{--\cdots-+} \\ q_{--\cdots--} \end{bmatrix} \quad (4)$$

It is assumed that at $k = 0$, for which there is no available observation, the system is in a stable cycle. This helps one to specify the values of the matrix q elements. Finally, the binary model can be shown as

$$P = Z^{\infty} q \quad (5)$$

where P is the probability vector with each element indicating the probability of being in one of the possible states after a large number of time periods. As the system operates over infinite periods of time and with respect to the first hypothesis, one of the two states of $(- - \cdots - -)$ and $(+ + \cdots + +)$ dominates after some period, and the probability of being in any other possible state approaches to zero. So the probabilities of Collapse and Survival are defined to be $P^- = P_{--\dots--}$ and $P^+ = P_{++\cdots++}$ respectively. Due to the fact that the Collapse and Survival probabilities are complementary the following relation holds:

$$P^- + P^+ = 1 \quad (6)$$

Up to this point the model-based analysis is completed without having access to past observations. Now suppose there exists an observation such as $g = (- \, +)$, at some period. Hence, the equation (5) will change to

$$P[- +] = Z^{\infty} q[- +] \quad (7)$$

which indicates the conditional probability vector with respect to g as an observation. Probabilities of Collapse and Survival then change to

$$P^-[- +] = P_{--\dots--}[- +], \quad (8)$$

$$P^+[- +] = P_{++\cdots++}[- +] . \quad (9)$$

Now, by comparing the results from model-based analysis with historical observations, we can investigate the following:

If $P^-[- +] > P^-$, one claims that the observed $(- +)$ sequence of local changes in the system increases the probability of Collapse in the model. Hence, the model registers an inclination toward Collapse.

If $P^-[- +] < P^-$, one claims that the observed $(- +)$ sequence of local changes in the system decreases the probability of Collapse in the model. Hence, the model registers an inclination toward Survival.

If $P^-[- +] = P^-$, the model registers no inclination. This finding is relatively rare.

As stated before, Inclination Analysis is based on stochastic processes theory. Other methods are based on statistical learning theory. These other methods can be used to deal with systems that are subject to a high level of uncertainty such as ecosystems and financial systems. We discuss one of these other methods in the following section of this paper.

III. Support Vector Machines

A. Basic Principles

Support Vector Machines (SVM) is a method that is used to train a machine to recognize patterns in a system. The training uses a set of examples consisting of input and output data in cases where no human expert is able to model the system dynamics, or where the system under investigation is subject to rapid change, such as the stock market. While there are some problems with this learning methodology, which are discussed in section IV, SVM has been used in a wide variety of applications such as speaker verification [5], handwriting recognition [6] &[7], and financial time series forecasting [8] in the few years since its introduction by Vapnik and Co-workers [3], [14]. These prior researchers have been able to address many of those problems. The applications where SVM has been successfully employed fall into two main categories:

- Classification
- Regression

In this paper we focus on the classification category.

B. SVM for classification

Suppose we have a set of data points, each point belonging to one of two predefined classes. Given a new data point, we want to determine the class to which this new point belongs. To carry out the training, one must first assign each available data point to its appropriate class. This is accomplished by identifying the best classifier line. When we have classes of data that are linearly separable, there are many possible classifier lines (Fig. 2), but there is only one *best* line. This line should be as far away as possible from the data of both classes (Fig. 3). In order to identify this line we must solve an optimization problem. This is called the Maximal-margin Decision Boundary. We do not discuss the calculations in this paper. For a complete discussion refer to [5], [14], and [15].

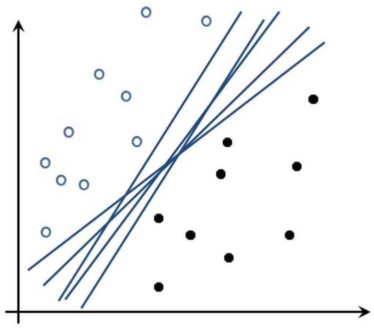

Figure 2. Possible decision boundaries to separate two classes of linearly separable data

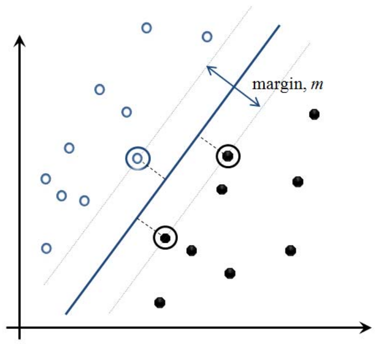

Figure 3. Maximal-margin Decision Boundary

In most real world problems the classes of data are not linearly separable. The Soft Margin method can be used to address this particular challenge. The Soft Margin method allows for classification error by building in a slack variable (Fig. 4).

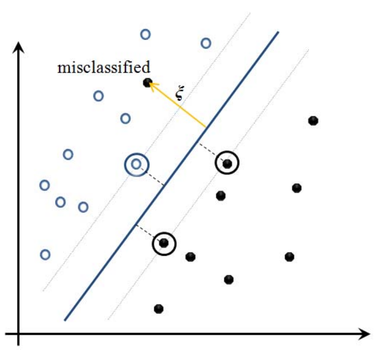

Figure 4. Soft Margin method where ξ is the slack variable

The original optimal hyperplane algorithm is a linear classifier. As discussed above, alternative approaches are needed to address most real-world situations because the classes are not likely to be linearly separable. A second approach was suggested by Boser et al [9]. Their key idea is to transform the data from input space to a higher dimensional space called Feature Space (Fig. 5). Computation in the feature space can be costly because it is high dimensional (the feature space is typically infinite-dimensional). This is where the kernel trick is used.

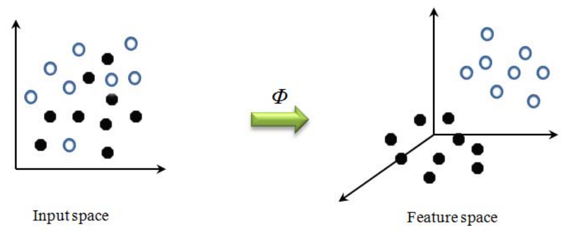

Figure 5. Input space: the space where the data points "$\boldsymbol{x_i}$" are located. Feature space: the space of $\emptyset(\boldsymbol{x_i})$ after transformation

In a kernel transformation

- Linear operation in the Feature space is equivalent to non-linear operation in Input space, and
- Classification can become easier with a proper transformation.

With respect to the optimization problem, the data points only appear as inner products. So one can calculate the inner product in the feature space and there is no need to map data explicitly. Therefore, by defining the kernel function K as

$$K(x_i, x_j) = \emptyset(x_i)^T \emptyset(x_j), \qquad (10)$$

carrying out $\emptyset(x_i)$ explicitly is avoided. The use of kernel functions is an attractive computational short-cut. Many kernel functions have been introduced and it is possible to design kernels depending upon the nature of the system nonlinearities and dynamics [10], [11].

C. Fisher kernel

In this paper we explain Fisher kernel [12] since it was shown to perform well in many applications and can process data that is not of the vector type [16]. Generative models such as Gaussian Mixture Models (GMM) and Hidden Markov Models (HMM) [13] have been used to model the observed data. Fisher kernel attempts to extract from a generative model more information than simply its output probability. The goal is to obtain internal representation of the data items within the probability model that describes the system's input to handle the variable length data. The model needed to be smoothly parameterized so that derivatives of the model with respect to the parameters can be computed. Suppose we have a generative model which is a probability density function of input points x, say, $M = P(x, \theta_0)$, of data with parameters $\theta_0 = \{\theta_{0i}\}_{i=1}^{n}$ where n is the number of parameters. Therefore, the probability of each data point x can be computed using the generative model M with the parameters θ set to θ_0. This probabilistic value with respect to the generative model M is defined to be the likelihood of that specific data point. So for an input space

$X = \{x_j\}_{j=1}^{N}$ with N as number of data points in a set of training data points, the model parameter θ_0 can be learnt by adapting all the points to maximize the likelihood of the training set. So basically one needs to maximize

$$\prod_{j=1}^{N} \mathcal{L}_\theta(x_j) \qquad (11)$$

where $\mathcal{L}$ represents the likelihood of the training data. In order to get rid of the complicated calculations, in terms of dealing with several multiplications, in practice the log-likelihood of the data points x_j with respect to the generative model is used as follows:

$$log \prod_{j=1}^{N} \mathcal{L}_\theta(x_j) = \sum_{j=1}^{N} log\ \mathcal{L}_\theta(x_j). \qquad (12)$$

Consider the vector gradient of the log-likelihood

$$g(\theta, x_j) = \left(\frac{\partial\, log\, \mathcal{L}_\theta(x_j)}{\partial \theta_1}, \frac{\partial\, log\, \mathcal{L}_\theta(x_j)}{\partial \theta_2}, \dots, \frac{\partial\, log\, \mathcal{L}_\theta(x_j)}{\partial \theta_n}\right). \qquad (13)$$

Hence for each data point x_j, $g(\theta_0, x_j)$ is defined as its Fisher score with respect to the generative model for the given set of parameters θ_0. The Fisher score gives an embedding into the feature space R^F and therefore, immediately suggests a possible kernel which is called Fisher kernel as

$$K(x_i, x_j) = g(\theta, x_j)'\ I_M^{-1} g(\theta, x_i) \qquad (14)$$

where I_M called the Fisher information matrix which is usually approximated by identity. So the practical Fisher kernel is defined as

$$K(x_i, x_j) = g(\theta, x_j)'\ g(\theta, x_i). \qquad (15)$$

IV. Discussion

Although we have found only two prior applications of Inclination Analysis at the time we wrote this paper, we believe the method has thus far shown itself to be capable of dealing with different levels of uncertainty [1], [2]. This method was first used to predict the disappearance of some types of rodents in the Ukraine. That study indicated that Inclination Analysis would have predicted the extinction of rodents even before the Chernobyl disaster happened in 1986. The only data required for that application was the rodent population size between 1981 and 1985. Based on this finding, we suggest that Inclination Analysis could be used when dealing with environmental systems. Inclination Analysis was also used to predict the recession in Russia. This research shows how by having access to only the USD/Ruble exchange rates from January 1997 to August 1998, just before the recession, Inclination Analysis could have predicted the Collapse of the Ruble and the recession in Russia in August 1998. For this reason we suggest that Inclination Analysis could also be applied successfully to an economic system. Both rodent population size and the USD/ Ruble exchange rate are subject to great uncertainty. In fact it is the level of uncertainty that makes it difficult to effectively model these systems using conventional methods. Moreover, the information requirements are very moderate for applications of Inclination Analysis. For the Russian economic crisis application, for example, there are many factors that impact the system, and yet the USD/ Ruble exchange rates were the only information required for the application. These rates are very easy to obtain.

While we are excited that these two applications suggest that Inclination Analysis may prove useful in studies of financial and environmental sustainability, we realize that a lot more study is required. In theory, Inclination Analysis should distinguish a system that is in decline and yet robust enough to recover and Survive, from one that is headed irretrievably for Collapse. For example, Lake Erie was pronounced dead many years ago. However it is clear today, that the Lake is not dead. A re-analysis of the Lake Erie data would be an excellent subject for future study. We also note that both of the applications discussed earlier examined systems which had already Collapsed. We are unaware of any applications of the model where the objective was to predict a future state. It might be necessary to refine the model's dynamic in such a way that it would predict a Collapse that has not already occurred.

The problem of transcribing and classifying local change observation values into the Collapse, "$-$", or Survival, "$+$" categories must also be addressed. We suggest that SVM may prove useful for this particular task. Also, one needs to decide how many periods of observations in the systems history are affecting the system's behavior in next period. In fact, as the state's length grows, the number of possible states increases exponentially and consequently the calculations take more time. However, the good news is that different L-values can be easily evaluated to find the best states length for the system. There are some other issues regarding the tuning of other parameters such as $q, r, m^-$and m^+. One needs to decide what would be the best stable dynamic of the system in order to make a decision about the probability of being in each of the possible states at period $k = 0$. Choosing proper values for critical number of $+$ or $-$ in a state is another challenge. Due to simple implementation of the model, it provides a good flexibility to design the model to operate and be evaluated for various parameters.

Regarding the prediction of system collapse, the Fisher kernel (equation 15) can be used to transform the non-linearly separable data from the input space to the feature space where it is possible to find a separating hyperplane. Also, it is usually possible to transform the data from one space to another in using multiple kernels in order to reach a better classification. The data are usually some features or indices extracted from the phenomenon that is being modeled. During the time of data collection, the data points have to be labeled, which means that one should know the (predefined) classes to which each data point belongs. Once the classification is preformed and the model parameters are identified, the model can be used for testing new data points for potential collapse prediction. For example, by extracting the same types of features from a sufficient number of bankrupted (class 1) and not-bankrupted (class 2) companies and training the model using the corresponding data, potential collapse prediction for an unknown company can be made. The accuracy of prediction highly depends on the correlation between the extracted features and bankruptcy of a company. Although predicting bankruptcy for a non-bankrupted company may be seen as a wrong decision, it can be considered as a warning to that specific company in order to avoid future bankruptcy.

SVM has shown very accurate results with regard to the classification task. The introduction of kernels has reduced the computational complexity of this method. Combining Fisher kernel with SVM makes it even more powerful when dealing with datasets of varying lengths. Using SVM, one needs only a set of examples or datasets which are already labeled in order to train the model. Once training is finished, the classification of the new data is very fast. However, problems arise if an insufficient number of examples of the system under study are available. With respect to the Statistical Learning theory, SVM uses a batch processing method to train the model [3]. So a sufficient number of examples must be available for training purposes. SVM seems to have difficulties if there are insufficient labeled examples.

Acknowledgment

Sincere thanks to NSERC, CFI, OIT and Ryerson University for funding the research.

References

[1] A.V. Kryazhimskii, M.B.Beck; "Identifying the inclination of a system towards a terminal state from current observation"; Environmental Foresight and Models: a Manifesto, Edited by M. B. Beck; Elsevier 2002

[2] Sergei Kryazhimskii, "Economic crisis in Russian, August 1998. Testing the Inclination approach", July 2002, unpoblished

[3] V. Vapnik, Statistical Learning Theory, Wiley-Interscience, New York, 1998.

[4] S. B. Kotsiantis, "Supervised machine learning: A review of classification techniques", Overview paper, July 2007

[5] Douglas A. Reynolds, T. F. Quatieri, and R. Dunn, "Speaker verification using adapted Gaussian mixture models", Digital Signal Processing, vol. 10, no. 1-3, pp. 19.41, 2000.

[6] Claus Bahlmann, Bernard Haasdonk and Hans Burkhardt, "On-line Handwriting Recognition using Support Vector Machines - A kernel approach". In Int. Workshop on Frontiers in Handwriting Recognition (IWFHR) 2002, Niagara-on-the-Lake, Canada, August 2002

[7] Abdul Rahim Ahmad, Marzuki Khalid, Christian Viard-Gaudin, Emilie Poisson, "Online handwriting recognition using support vector machine", ©2004 IEEE

[8] Kyoung-jae Kim, "Financial time series forecasting using support vector machines", 2003 Elsevier

[9] Bernhard E. Boser, Isabelle M. Guyon, Vladimir N. Vapnik, "A training algorithm for optimal margin classifiers", 1992

[10] N. Cristianini and J. SH. Taylor, "An introduction to support vector machines and other kernel-based learning methods". United kingdom: Cambridge University Press, 2000.

[11] J. SH. Taylor and N. Cristianini, "Kenel methods for pattern analysis", United kingdom: Cambridge University Press, 2004.

[12] T.S. Jaakkola and Haussler, "Exploiting generating models in discriminanative calssifiers", Advances in neural information processing systems, 1998.

[13] Lawrence R. Rabiner, "A tutorial on Hidden Markov Models and selected applications in speech recognition". Proceedings of the IEEE 77 (2): 257-286, 1989

[14] Vladimir Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, 1995.

[15] J. A. K. Suykens, T. Van Gestel, J. De Brabanter, B. De Moor, J. Vandewalle, "Least squares support vector machines", World Scientific Pub. Co., Singapore, 2002

[16] NICOTRA, L., A. MICHELI and A. STARITA, "Fisher kernel for tree structured data". Neural Networks, 2004 IEEE