

Dynamical Systems and Nonlinear Kalman Filtering Applied in Classification

Wieslaw Pietruszkiewicz
SDART Ltd
One Central Park
Northampton Road
Manchester M40 5WW
Web: www.sdart.co.uk
Email: wieslaw.p@sdart.co.uk

Abstract—In this paper we present an application for the dynamical systems and nonlinear filtering in classification. Described here is a method which constitutes a new proposal for the classification of dynamical objects. This method does not require the expansion of dimensionality of the input vectors used in time series, compared to other methods which multiply inputs. The proposed classifier uses the dynamical systems as its core and nonlinear Kalman filtering as its learning algorithm. Presented in this paper are the results of performance tests to a real life problem – bankruptcy prediction, which is one of the fundamentals application of data mining in finance. The results achieved by the presented method were compared to other, very popular classifiers, and these experiments prove greater classification accuracy of the proposed method.

I. INTRODUCTION

The purpose of this paper is to show the possibility of the applications of the dynamical systems and nonlinear filtering in classification tasks. Simple construction of dynamical systems and the capability to deal with dynamical processes may allow a different approach to classification, where too often processes are simplified by treating them as static ones. While this is understandable for mathematic ease but it must be noted that it also causes lower accuracy in cases where cutting off time series means neglecting important characteristic of analysed objects. Dynamical models are widely used in different technical application, for example in control units or navigation systems. However their use in data mining or machine learning is very limited.

While Kalman filtering has many successful application since its invention in the sixties, its possibilities have been lately expanded by the invention of nonlinear Kalman filters like *UKF*, *SRKF* or *CDKF*. As these algorithms are able to filter nonlinear systems, they could be used in machine learning or data mining. Especially in cases where learning process deals with noisy data, Kalman filtering over perform common non noise aware algorithms.

The proposed method of classification was tested on the real life problem of bankruptcy prediction, which in terms of DM can be described as the classification of business entities to one of the groups of companies with similar financial standing. This prediction is very often used in business and mostly is being done with easy to calculate but inaccurate discriminant

functions.

II. UNSCENTED TRANSFORMATION AND UNSCENTED KALMAN FILTER

Dynamical systems can be described as systems with inertia or memory i.e. systems where the past dictates the present system's behaviour. They are opposite to static systems where the system's state can be calculated only on a base of its present inputs. Real life processes and objects in majority are examples of dynamical systems. However, sometimes simpler and easier static approaches to modelling are acceptable, while for other objects it could cause an unacceptable performance decrease. The basic notation for dynamical systems can be written in form of:

$$x_{k+1} = F(x_k, u_k, v_k) \quad (1)$$

$$y_k = H(x_k, n_k) \quad (2)$$

- x_k - state variables,
- u_k - inputs (financial ratios),
- Where: y_k - outputs,
- v_k - process noise,
- n_k - observation noise.

Practical applications of dynamical systems require knowledge of their mathematical form which can be done by analysis of ruling laws (where possible) e.g. laws of mechanic, chemistry or electrotechnical. In other cases – where systems structure is unknown or too complicated – system identification also called estimation of system's parameters is necessary. While identification or estimation of parameters are terms strongly connected with systems theory this process can be simply described as learning or in terms from artificial intelligence, machine learning or data mining.

Kalman filtering has many applications in different technical problems. In its original form (see [7]), due to linear construction, this filter cannot be used in identification of objects (or processes). As even identification of linear system transforms into estimation of state variables in nonlinear system, this is beyond capabilities of original Kalman filter. Limitation of KF only to linear systems caused development of Extended Kalman Filter, which can be also applied to

non-linear systems. While EKF makes it possible to estimate system's parameters, often such identification is not accurate enough as EKF's ability to work with nonlinear systems is ensured by simple linearization of estimated systems (done by evolving them into first order Taylor series).

A solution to this problem can be Unscented Transformation (UT in abbrev. – see [8] or [18]) which is a core of UKF – *Unscented Kalman Filter*. UT's main idea is to calculate statistics of random variables by propagating a set of specially chosen points (sigma points) via the *real* model of the system and calculating their statistic parameters, not by *approximation* achieved from inaccurate system's simplification. As UKF belong to the family of Kalman filters it can be described as a recurrent estimation of variables, with a major loop being prediction–correction steps. The detailed steps of UKF are (more in [9] or [19]):

Step I – Initialization

$$\hat{x}_0 = E[x_0] \quad (3)$$

$$\chi_{k|k+1}^x = F [\chi_{k-1}^x, u_{k-1}, \chi_{k-1}^\nu] \quad (4)$$

$$\hat{x}_0^a = E[x^a] = [\hat{x}_0^T \ 0 \ 0]^T \quad (5)$$

$$P_0^a = E[(x_0^a - \hat{x}_0^a)(x_0^a - \hat{x}_0^a)^T] = \begin{bmatrix} P_0 & 0 & 0 \\ 0 & R^\nu & 0 \\ 0 & 0 & R^n \end{bmatrix} \quad (6)$$

For $k \in \{1, \dots, \infty\}$

Step II – Calculation of sigma points:

$$\chi_{k-1}^a = \left[\hat{x}_{k-1}^a \quad \hat{x}_{k-1}^a + \tau \sqrt{P_{k-1}^a} \quad \hat{x}_{k-1}^a - \tau \sqrt{P_{k-1}^a} \right] \quad (7)$$

Step III – Time update:

$$\chi_{k|k-1}^x = F [\chi_{k-1}^x, u_{k-1}, \chi_{k-1}^\nu] \quad (8)$$

$$\hat{x}_k^- = \sum_{i=0}^{2L} W_i^{(m)} \chi_{i,k|k-1}^x \quad (9)$$

$$P_k^- = \sum_{i=0}^{2L} W_i^{(c)} \left[\chi_{i,k|k-1}^x - \hat{x}_k^- \right] \left[\chi_{i,k|k-1}^x - \hat{x}_k^- \right]^T \quad (10)$$

$$Y^{k|k-1} = H \left[\chi_{k|k-1}^x, \chi_{k-1}^n \right] \quad (11)$$

$$\hat{y}_k^- = \sum_{i=0}^{2L} W_i^{(m)} Y_{i,k|k-1} \quad (12)$$

Step IV – Measurement update equations:

$$P_{\hat{y}_k, \hat{y}_k} = \sum_{i=0}^{2L} W_i^{(c)} \left[Y_{i,k|k-1} - \hat{y}_k^- \right] \left[Y_{i,k|k-1} - \hat{y}_k^- \right]^T \quad (13)$$

$$P_{x_k, y_k} = \sum_{i=0}^{2L} W_i^{(c)} \left[\chi_{i,k|k-1}^x - \hat{x}_k^- \right] \left[Y_{i,k|k-1} - \hat{y}_k^- \right]^T \quad (14)$$

$$K_k = P_{x_k, y_k} P_{\hat{y}_k, \hat{y}_k}^{-1} \quad (15)$$

$$\hat{x}_k = \hat{x}_k^- + K_k (y_k - \hat{y}_k^-) \quad (16)$$

$$P_k = P_k^- - K_k P_{\hat{y}_k, \hat{y}_k} K_k^T \quad (17)$$

Where:

$$x^q = [x^T \ \nu^T \ n^T]^T \quad (18)$$

$$\chi^a = \left[(\chi^x)^T (\chi^\nu)^T (\chi^n)^T \right]^T \quad (19)$$

$$\tau = \sqrt{l + \lambda} \quad (20)$$

Where used symbols mean:

- λ – scaling parameter,
- L – dimensionality of extended state vector,
- R^ν – process noise covariance,
- R^n – measurement noise covariance,
- W_i – weights.

As numerous experiments prove better performance of UKF over standard EKF e.g. [14], this filter was selected to be the learning algorithm in further experiments. Apart of better accuracy UKF is also simpler to apply as it does not require to compute matrix of partial derivatives (the Jacobian).

III. FINANCIAL CONDITION EVALUATION AS A CLASSIFICATION TASK

The bankruptcy forecast is one of the basic tasks of company management, where assessment of the current and further company condition allows one to predict its ability to survive on the market. From data mining perspective this problem can be described as a classification task where evaluated business entities are being recognised as belonging to one group of companies with similar financial profiles. There are many groups interested in such prediction e.g. company owners, managers, creditors, stock market investors or business partners.

The first quantitative model assessing probability of company bankruptcy was a model worked out by E.I. Altman [1]. His pioneer approach to this problem is currently used by many other researchers [10] or [15], who have also used different kinds of discriminant analysis. Another popular prediction methods are neural networks [2], [11], [16] or [17]. There are also researchers who have used other methods of classification like Nearest Mean or Nearest Neighbour [6]. In all those cases classifiers were used as black-box's where inputs are financial ratios and output from a box is the classification result.

A common limitation to all these approaches is the assumption that business entity is a static system and its current financial

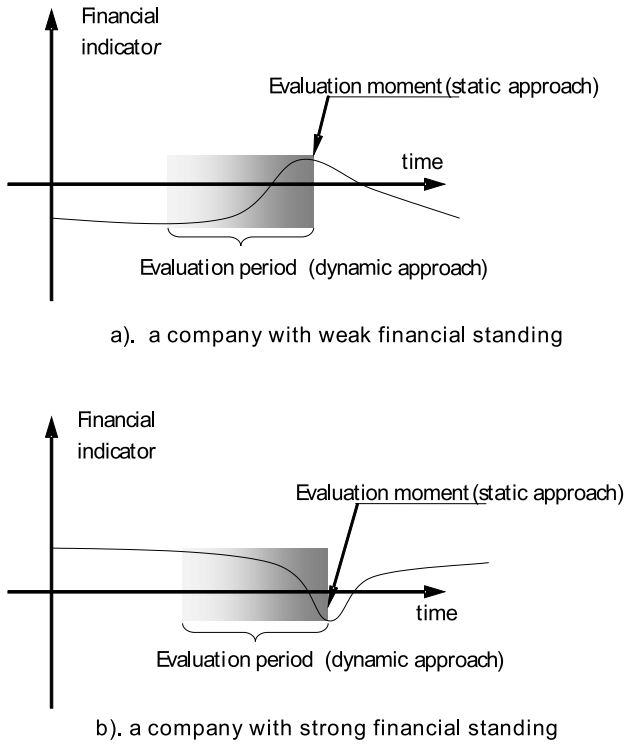


Fig. 1. Graphic explanation of static vs. dynamic classification

standing can be assessed on its current financial ratios. This assumption can cause errors in cases where current (at the moment of evaluation) company financial statements differ from ordinary situations in the past (see Fig. 1). This results in misrecognized companies. In real life application each non-recognized bankruptcy causes losses for capital donors and each non-recognized strong company is a loss of income generating client or profitable investment.

IV. PROPOSED DYNAMICAL CLASSIFIER

Structure of dynamic classifier was presented on Fig.2 . As one can see size of vector U – being system's input – can vary due to idea build flexible structure that will be able to classify bankruptcy with different number of financial ratios. Observed system's output contains a variable $Y_{n+1} \in \langle -1 \dots 1 \rangle$ which is bankruptcy prediction, where n denotes a number of used financial inputs. Values ≥ 0 mean that company was not classified as a bankrupt in nearest future, while values < 0 are signals of very likely bankruptcy.

Nodes W_i are parameters to be estimated (or learnt) by the UKF filter. A progress of learning these parameters was shown in Fig.3. To ensure faster process convergence at the beginning and limited fluctuations at the end, the estimation was done with simulated annealing of the noise matrix.

The FBR parameter requires a short explanation. The whole learning process was divided into 2 steps. At the first step, Kalman filter was used to estimate parameters where the feedback loop was left open ($FBR = 0$) causing this model to act as a form of Minimum Square Error classifier (see

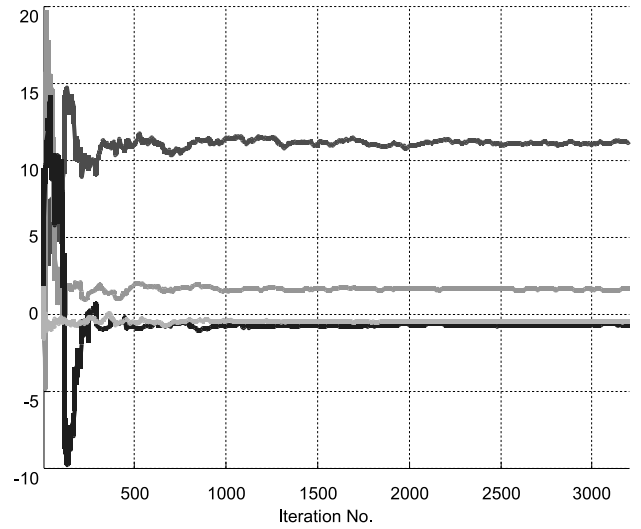


Fig. 3. The progress of parameters estimation

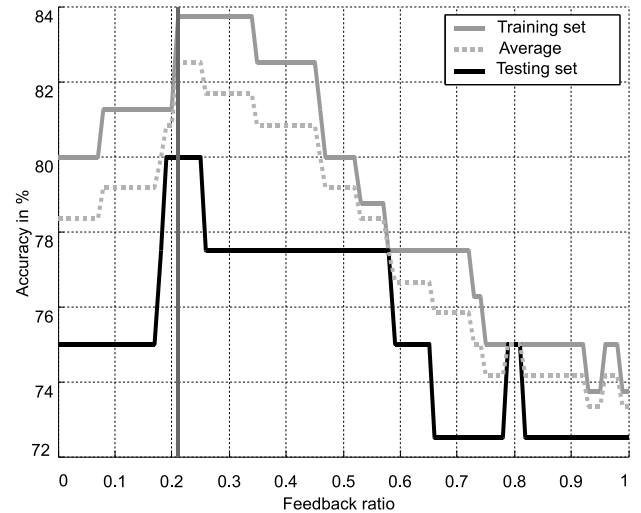


Fig. 4. Classification accuracy with different FBR values

at [4] for more information about SME classifier) – denoted further as DSS. During the second step, the FBR value was adjusted from 0 to 1 (with step 0.01) to select optimal value, where the best accuracy on the training set can be achieved. After feedback loop closure the DSS transforms into the DDS classifier (see Tab.I). Fig.4 shows relation between FBR value and classification accuracy for both sets: training and testing.

V. THE RESULTS

The described method of the dynamical classification was tested on a financial dataset containing 240 samples – 112 from bankrupt Polish companies and 128 from existing companies (at least at the moment of dataset collection). Prediction accuracy was tested using 3-fold cross validation for several popular classification methods (see Tab. I for their names and abbreviations).

To ensure that the dataset was long enough to be reliable testing material it was necessary to check how the length of

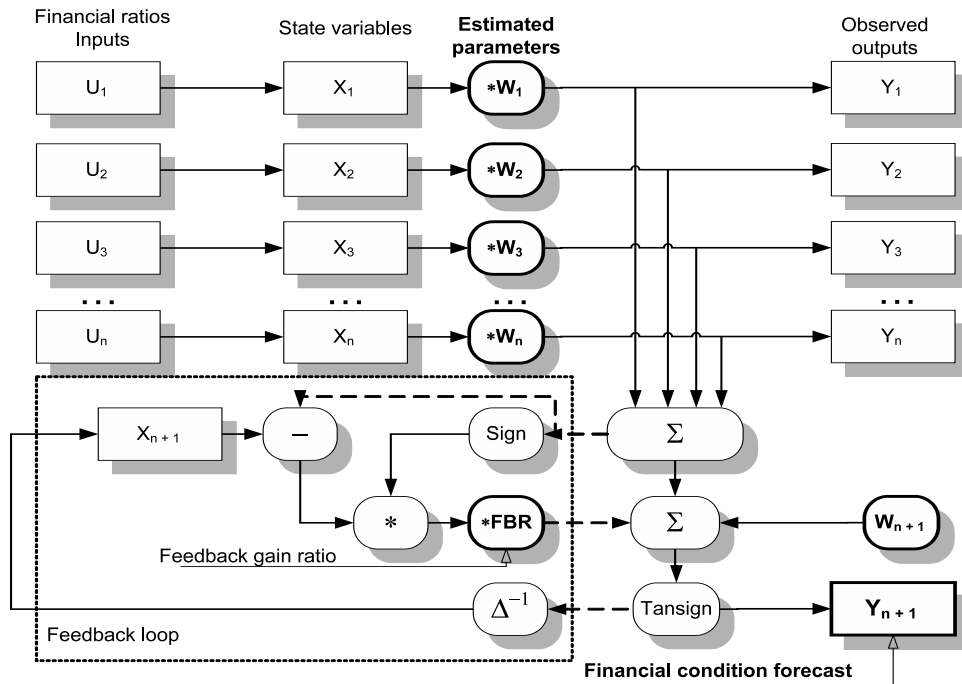


Fig. 2. Schema of the discrete dynamical classifier

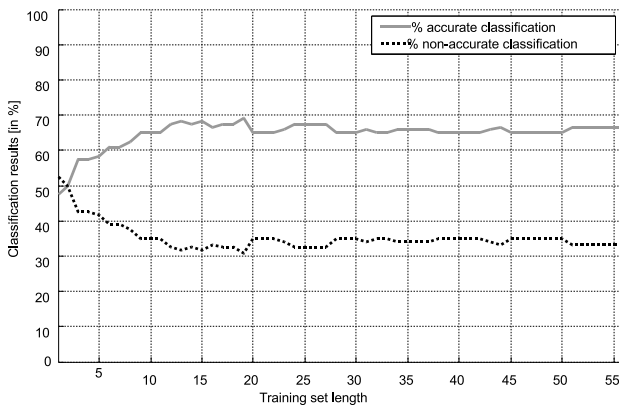


Fig. 5. The reliability of training dataset

training dataset effects on prediction accuracy. The results shown on Fig.5 prove that increasing dataset length does not lead to an accuracy increase, thus the dataset was reliable (compare with [3]).

At the beginning, a full set of attributes contained 30 financial ratios being possible bankruptcy indicators. To select most relevant attributes a new selection recursive algorithm was performed, where its core is recursive weighting of attributes significant by their reversed correlation coefficients (to ensure high significance of attributes but removing highly inter-correlated attributes). After this selection a subset of 5 ratios was chosen (see Tab.II) to further experiments.

Abbreviation	Classification method
NN	Nearest Neighbour
NM	Nearest Mean
ANN	Artificial Neural Network
DA	Discriminant Analysis
DSS	Discrete Static System
DDS	Discrete Dynamical System

TABLE I
USED METHODS OF CLASSIFICATION AND THEIR ABBREVIATIONS

Attribute	Ratio
	Selected
X_9	net profit ÷ total assets
X_1	cash ÷ current liabilities
X_{24}	net profit ÷ sales
X_6	working capital ÷ sales
X_2	cash ÷ total assets
	Rejected
X_{11}	net profit ÷ sales
X_{15}	net profit ÷ (equity + long term liabilities)
X_6	working capital ÷ sales
X_7	sales ÷ inventory
X_{21}	liabilities ÷ total income

TABLE II
INPUT ATTRIBUTES AFTER SIGNIFICANCE ANALYSIS (IN DECREASING ORDER OF SIGNIFICANCE)

Tables III–VII show classification accuracy for different number (1–5) of input vector. The best result for each stage of experiments is in bold. A number in parenthesis at ANN position denotes number of neurons at the hidden layer. Cost

Method	The results		
	Accurate	Type I errors	Type II errors
NN	66.667%	19.583%	13.750%
NM	69.583%	16.667%	13.750%
ANN (9)	70.417%	15.417%	14.167%
DA	69.583%	12.083%	18.333%
DSS	69.167%	17.083%	13.750%
DDS	73.333%	15.000%	11.667%

TABLE III
THE RESULTS OF CLASSIFICATION WITH 1 INPUT

Method	The results		
	Accurate	Type I errors	Type II errors
NN	67.500%	15.417%	17.083%
NM	65.417%	3.333%	31.250%
ANN (8)	77.083%	10.000%	12.917%
DA	72.500%	5.833%	21.667%
DSS	76.667%	10.417%	12.917%
DDS	79.833%	9.750%	10.417%

TABLE IV
THE RESULTS OF CLASSIFICATION WITH 2 INPUT

Method	The results		
	Accurate	Type I errors	Type II errors
NN	71.667%	11.667%	16.667%
NM	67.917%	3.333%	28.750%
ANN (6)	74.583%	10.833%	14.583%
DA	73.333%	6.250%	20.417%
DSS	75.000%	11.667%	13.333%
DDS	80.833%	10.000%	9.167%

TABLE V
THE RESULTS OF CLASSIFICATION WITH 3 INPUT

Method	The results		
	Accurate	Type I errors	Type II errors
NN	61.250%	20.833%	17.917%
NM	68.750%	6.250%	25.000%
ANN (9)	72.917%	12.083%	15.000%
DA	70.000%	9.167%	20.833%
DSS	72.917%	13.333%	13.750%
DDS	76.667%	10.000%	13.333%

TABLE VI
THE RESULTS OF CLASSIFICATION WITH 4 INPUT

Method	The results		
	Accurate	Type I errors	Type II errors
NN	59.583%	21.667%	18.750%
NM	68.333%	6.250%	25.417%
ANN (4)	69.167%	14.583%	16.250%
DA	71.250%	7.500%	21.250%
DSS	71.250%	15.000%	13.750%
DDS	74.167%	12.500%	13.333%

TABLE VII
THE RESULTS OF CLASSIFICATION WITH 5 INPUT

of I (misrecognized bankrupt) and II type (misrecognized nonbankrupt company) errors were set to be equal.

At all experiments stages the DDS achieves the best classification accuracy, which leads to the conclusion that process's dynamic is an important factor in the case of bankruptcy prediction and proves that dynamical systems and nonlinear

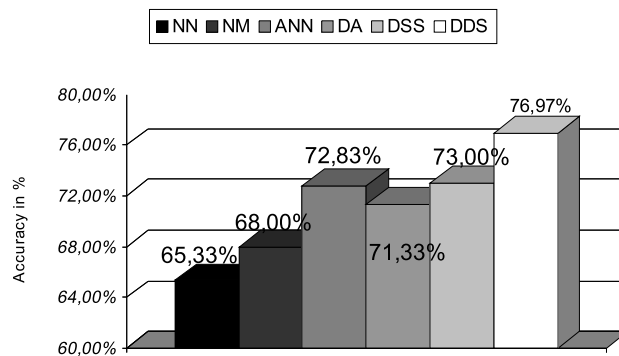


Fig. 6. Average classification accuracy for all tested methods

Kalman filters can be used as a dynamical method of classification. An average classification accuracy was presented on Fig.6.

VI. CONCLUSION

As it was shown in this paper, dynamical classification achieved better results than previously known methods of bankruptcy prediction. The presented method of objects classification, based on models of dynamical systems and nonlinear Kalman filtering is a proposition for a new application of these methods in data mining, machine learning or artificial intelligence. Even if it is possible to incorporate time series in the other classification methods by multiplication of attributes vector shifted to previous periods, this still leads to increase of attributes dimensionality. Additional experiments done for other kinds of nonlinear Kalman filters like *SRKF* or *SPKF* gave very similar results to *UKF*.

Dynamical systems and Kalman filters are too rarely used in data mining or machine learning, while they have proven their excellent performance and capabilities in other areas like control units or estimation devices. Moreover, nonlinear Kalman filtering can be used separately as a learning algorithm in other classifiers e.g. neural networks.

The method presented in this paper is being further developed by SDART Ltd. as a part of financial decision support system.

REFERENCES

- [1] E.I. Altman, *Financial Ratio, Discriminant Analysis and Prediction of Corporate Bankruptcy*, Journal of Finance, August 1968.
- [2] A.F. Atiya, *Bankruptcy Prediction for Credit Risk Using Neural Networks: A Survey and New Results*, IEEE Transactions on Neural Networks, Vol. 12, No. 4, July 2001.
- [3] B. Back, T. Laitinen, J. Hekanaho, K. Sere, *The Effect of Sample Size on Different Failure Prediction Methods*, Turku Centre for Computer Science, TUCS Technical Report No 155, December 1997.
- [4] R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, New York, United States: John Wiley & Sons, 2001.
- [5] J.H. Gove, D.Y. Hollinger, *Application of a dual unscented Kalman filter on simultaneous state and parameter estimation in problems of surface-atmosphere exchange*, Journal of Geophysical Research 111(d8): D08S07, 2006
- [6] H.A. Guvenir, S. Altingovde, I. Uysal, E. Erel *Bankruptcy Prediction Using Feature Projection Based Classification*, Proceedings of SCI/ISAS'99, United States:Orlando, 1999.

- [7] R.E. Kalman, *A New Approach to Linear Filtering and Prediction Problems*, Transaction of the ASME - Journal of Basic Engineering, March 1960.
- [8] S.J. Julier, J.K. Uhlmann, *A General Method for Approximating Nonlinear Transformations of Probability Distributions*, Robotics Research Group, University of Oxford, Technical Report, October 1996.
- [9] R. van der Merwe, E. Wan, *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*, Proceedings of the Workshop on Advances in Machine Learning, Montreal 2003.
- [10] J. Ohlson, *Financial ratios and the probabilistic prediction of bankruptcy*, Journal of Accounting Research Vol. 18, No. 1 1980.
- [11] C. Serrano, *Self Organizing Neural Networks for Financial Diagnosis*, Decision Support Systems Vol. 17 1996.
- [12] C.Y. Shirata, *Financial Ratios as Predictors of Bankruptcy in Japan, An Empirical Research*, Journal of Risk and Management Vol.23 1998.
- [13] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches* 2006), New York, United States: John Wiley & Sons, 2006.
- [14] A. Sitz, U. Schwarz, J. Kurths, H.U. Voss, *Estimation of parameters and unobserved components for nonlinear systems from noisy time series*, Physical Review E 66/2002.
- [15] P. Theodossiou, E. Kahya, *Predicting Corporate Financial Distress: A Time-Series CUSUM Methodology*, Review of Quantitative Finance and Accounting Vol.13 1999.
- [16] W.L. Tung, C. Quek, P.Y.K. Cheng, *Bank Failure Prediction and Financial Data Reconstruction using Novel Soft-Computing Approaches*, Nanyang Business School Finance Seminar, Nanyang 2003.
- [17] E.K. Tyree, J.A. Long, *Assessing Financial Distress with Probabilistic Neural Networks*, Working Paper, City University of London 1994.
- [18] J.K. Uhlmann, S. Julier, M. Csorba, *Nondivergent Simultaneous Map Building and Localization using Covariance Intersection*, Proceeding of American Control Conference 1995.
- [19] E.A. Wan, R. van der Merwe, *The Unscented Kalman Filter*, in: *Kalman Filtering and Neural Networks*, ed. Haykin S., New York, United States: John Wiley & Sons, 2001.