A model for failure nodes in wireless sensors networks (WSN) for military polygon monitoring

Petru JUNIE , Mihai TERTISCO, Cristian EREMIA

Politehnica University of Bucharest
junpetre2000@yahoo.com, tertisco_mihai@yahoo.com, cristian_valentin2003@yahoo.com

***Abstract*- The present paper presents the results of research on modelling and experimental identification systems with random binary events of failure nodes of a WSN which is monitored in a military polygon. For systems with random binary events model parameters estimation, a Monte – Carlo algorithm was forwarded and tested, using the maximum logarithmic likelihood criterion. The space of model parameters, for the point corresponding to maximum log likelihood function associated to the set of experimental data. The present paper proposes and tests an original method, to delimitate the search area, with Monte – Carlo, for the solution to the issue of logistic fuzzy- model parameters estimation in WSN.**

Keywords- Logistic Model, Monte Carlo method, maximum log likelihood criterion, node failure, WSN.

I. INTRODUCTION

A failed sensor node can not respond to his call to the base station or cluster of related neighbouring nodes and thus lose part about military polygon control centre monitored by WSN.

The main cause of failure of sensor nodes in WSN is considered, by most scholars, depletion battery power supply node in most scientific papers published, the authors propose the use of deterministic models to assess the dynamics of the process of exhaustion of supply related node.

Research whose results are presented in this paper led to the idea that the failure of nodes - the sensor is a random process caused primarily by random nature of exhaustion supply node expressed mainly by changing the supply voltage. At the same time failure event is a binary random event characterized the dependent variable (output) Y takes the value 1 when sensor node logic is failed and Y= 0 when the node is not failed.

For systems with discrete events characterized by discrete streams of operations and discrete activities accompanied by phenomena of blocking, non-synchronization and conflicts new modelling formalisms have been developed [2] .Classic models covered by conventional identification methods describe the dynamic behaviour of a single object from a collection of similar objects in which processes that are subject to physical and chemical laws occur. In the present paper we are concerned with models of systems with binary independent random events. Unlike traditional models, this particular type of models describes a homogeneous lot M of cardinality N consisting of two distinct entities. These entities can be separated into two classes. Each entity in this population is characterized by a dependent variable Y (output) and one or more independent variables (input) x. Variable Y can take only logical values: 1 or 0, yes or no, sick or healthy, etc. The independent variable can take logic values or can take values in the set of real numbers. Based on experimental testing for each entity. Entities can be divided into two classes: entities class with $Y = 1$ and entities class with $Y = 0$. The model in which we have only one independent variable x is called the logistic model SISO (single - input - single - output).

Problem description

In the case of the identification theory, the model that expresses the probabilistic interdependence between the dependent variable Y, binary type, and one or more independent variables x, is called logistic regression.

For example, the experimental data regarding sensors failure for the analysis of failure percent of a WSN node for monitor of a testing ground consist by a network with 23 sensor nodes.

Issues to be resolved is obtaining a model that expresses the value of a node failure probability $P(Y = 1 | x)$, knowing the current value of the WSN node X related analysis. Such a pattern exists in mathematical statistics is called the Logistic Regression model.

II. LOGISTIC REGRESSION STRUCTURES

The regression equation obtained in this case is of a type different from other known regressions, such as continuous, single dimensional, multidimensional, linear and nonlinear etc. Logistic model structure for a SISO, found in the literature of expertise. In the variant (1) the continuous size "*p*" is a nonlinear function of *x* and of two unknown parameters β o: and β 1:

$$P(Y=1\mid X)=\frac{\exp(\beta_0+\beta_1 X)}{1+\exp(\beta_0+\beta_1 X)}=p \qquad (1)$$

If the event Y = 1, then this event's occurrence takes place with the probability *P(Y=1|x) =p*. Unlike classical linear regression, in the case of logistic regression (1), instead of dependent variable Y, which may take the binary value Y = 1→ "failure" or Y = 0 → "non-failure", it is used a continuous variable p, which takes values ranging from 0 to 1. A value of p is interpreted as the probability of obtaining a "failure" (*Y* = 1), subject to the independent variable value *x*. Then the opposite event *Y* = 0 has a probability of occurrence *P (Y = 0) = 1-p*. This type of regression provides information about the importance of variables *x* in the differentiation of each entity within a given set of n entities sorting them into categories. In the WSN field these entities may be n nodes investigated by direct testing. The existence of these categories is determined by the variables *x* which are often referred to as categorical or predictor variables. The result of these experimental investigations constitutes the necessary data for the estimation of logistic parameters model. It can immediately be seen that, for an observation from the set of experimental observations (experimental data), if p> 0.5, then it is more likely for the observation to belong to the group characterized by Y = 1

The data are obtained by successive direct observations on this set of n nodes resulting in the following sequence of n pairs of experimental data:

X=[**66,70,69,68,67,72,73,70, 57,63, 70,,78,67,53,67,75, 70,81,76,79,75, 76,58**]-Data vector of the *x and,*

Y=[**0;1;0;0;0;0;0;0;1;1;1;0;0;1;0;0;0;0;0;0;1;0;1**]- Data vector of the output (2)

For logistic regression parameters determination there are used the n experimental data pairs which characterize the n entities (events) with binary characteristics of the set for which the SISO logistic type model is determined. The recommended criterion by systems theory identification for evaluating the matching degree between data and model, in such cases is the *likelihood function* [4].

III. LOG LIKELIHOOD FUNCTION FOR SISO LOGISTIC SYSTEM

These data are direct successive observations of that particular set of n entities in which each entity i is characterized by the pair of values (Yi and xi). Based on these n pairs of experimental data those values of vector parameters Θ need to be determined so that the model obtained (having the structure (1) can best describe the experimental data and to ensure a high level of generality, in the sense of being able to correctly describe the specific logistics process behaviour in other points of the experimental data (2). Among these points from the experimental data set there are some in which Y = 1 and others in which Y = 0. Since the output of the process is a logistic variable which within the experiment takes the values Y1, Y2 ,..., Yn then the output of the model in the n experimental points is expressed by the probabilities sequence (p (Yi = 1 | Xi) or p (Yi = 0 | Xi) = 1 - p (Yi = 1 | Xi)). The probabilistic description of the entire set of n independent random events of logistic type is expressed by the product of n random probabilities related to observed binary random events:

$$P_n=\prod_{i=1}^{n} p_i \qquad (3)$$

Within this product there are two types of terms: terms corresponding events for which *Yi = 1 pi = p = Pr (Yi = 1, xi, parameters)* and terms related to the events for which *Yi = 0 pi = 1-p*. Under these conditions the relation (5) becomes:

$$P=\prod_{i=1}^{n} p_i^{Y_i}(1-p_i)^{1-Y_i} \qquad (4)$$

In [5], the probability function (4) is marked ***L (data parameters)*** and is called the *likelihood function of SISO logistic regression.*

If the case of logistics processes identification the problem is to find those values for model parameters that will ensure the maximum likelihood function. These values, in the case of a SISO model logistics are noted: $\hat{\beta}_0$ *şi* $\hat{\beta}_1$ and constitute the so-called model parameter estimates for the purposes of maximum likelihood. The problem of maximum likelihood estimates for a logistic SISO regression model :

$$\hat{\theta}=\arg\max_{\beta_0,\beta_1}\prod_{i=1}^{n}\frac{[\exp(\beta_0+\beta_1 X_i)]^{Y_i}}{1+\exp(\beta_0+\beta_1 X_i)}\ldots(5)$$

Where $\hat{\theta} = [\hat{\beta}_0\, \hat{\beta}_1]^T$ is the parameters estimation vector of the SISO logistic model.

Applying the natural logarithm of the likelihood function (*L*) events results in the function log likelihood (LL) binary logistic model with random shit. This function denoted LL (βo, β1) has the expression:

$$LL(\beta_0, \beta_1) = \ln[L(\beta_0, \beta_1) =$$

$$= \sum_{i=1}^{n} Y_i (\beta_0 + \beta_1 X_i) -$$

$$- \sum_{i=1}^{n} \ln[1 + e^{(\beta_0 + \beta_1 X_i)} \ldots\ldots\ldots\ldots\ldots(6)$$

IV. EXPERIMENTAL DETERMINATION OF THE SEARCH FIELD BORDERS (SFB)

Classical Monte-Carlo (fig.1)algorithm(CMCA) is random testing of the Log Likelihood surface, using two test sequences of random numbers S1 and S0 of finite length, (one sequence for each parameter β_1 and β_0). These sequences are cut from infinite strings of random numbers uniform probability distribution in the band- plan under investigation in the two parameters area.

Step 1: Generate a pair of random numbers [S0 (k = 1), S1 (k = 1)] with these values and existing experimental ***data*** *(12) is calculated log Likelihood(1)=LL(1) and is stored in memory M:*

*Step 2**:** increment by one count variable k = k +1 a number of tests and generates a new pair of random numbers that are calculated* ***LL (S0 (k), S1 (k), data) = LL (k)***

Step 3: Compare the L (k) with M from the previous step:

IF,

LL(k)>M

THEN

replaced the old content is LL (k)→M and return to Step 2

OTHERWISE

return to Step 2, M preserving the previous value.

Fig.1 Classical Monte-Carlo algorithm (CMCA)

The two random sequences obtained from two random number generators in Matlab CMCA algorithm of random search of the maximum log likelihood LL (βo, β1) simulation, in the SISO case, involves the execution of three steps[5] The three steps described above are performed within SFB

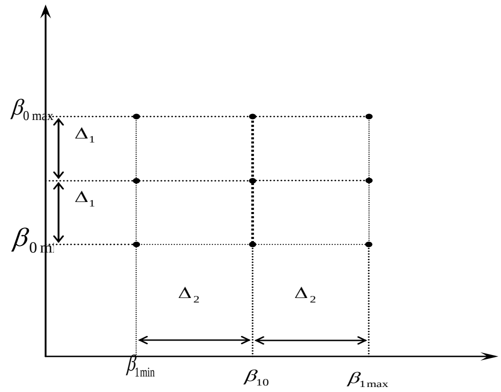

Fig. 2. Experimental determination of the SFB: Δ_1, Δ_2 -limits of variation of parameters in the search process; β_{10}, β_{00} - SFB centre coordinates in the plane parameters

The node from the network centre coincides with the SFB centre only if LLF values in the 8 peripheral nodes are lower than the LLF values in the centre of the network. In the opposite case, when in one of peripheral nodes the SFB value is greater than the value of SFB in the centre then the network is moved placing the node in the centre in the point with the highest value of LLF. The search continues in the same manner until the greatest value of LLF is in centre network. In the case study shown in table 2 the parameters variation limits $\Delta_1 = .5$ also $\Delta_2 = 2$ and the initial coordinate's dorm the centre of the network $(\beta_{10}, \beta_{00}) = (0,0)$ are arbitrary (Fig. 2).

CMCA application results in case 1 (for N = 50 steps) for searching illustrates how the maximum LLF point in the parameters plane (point coordinates, β0 =- 0.10127, β1 =- 3.043 ') was found after only 20 practical steps. And in the second case for N = 500, the coordinates point, β0 =- 0.103, β1= - 3.003 was found after 25 steps.

Limit values of the parameters (b0min, b0max, b1min, b1max) determines SFB. These limits are settled by means of pre explorations of the LLF values, made in a 9-node network shown in Figure2.

If the sequence of 1500 random numbers can be imagined as consisting of 100 consecutive segments of the same length N = 15 random numbers. Each of these segments can be used for repeated searching of he maximum LLF with CMCA.

The apparent random nature of the topography as well as of the estimates can be noticed in all 100 cases (fig. 3 where ε - random-noise with zero average value and dispersion $\sigma_{\cdot\varepsilon}$ **).**

$$\tilde{\beta}_0(1), \tilde{\beta}_0(2), \tilde{\beta}_0(3), \tilde{\beta}_0(4), \dots, \tilde{\beta}_0(100),$$

$$\tilde{\beta}_1(1), \tilde{\beta}_1(2), \tilde{\beta}_1(3), \tilde{\beta}_1(4), \dots, \tilde{\beta}_1(100),$$

$$\tilde{\beta}_i(k) = \beta_i + \varepsilon_i(k) \; ; \; i=0,1 \;\; ;k=1,2,\dots,$$

Fig.3

These observations lead us to the simple idea that instead of applying the CMCA only once on a long string of N random numbers we should apply CMCA by N repeatedly, = N / No of times on shorter sequences formed of No random numbers. Then, based on the results (10) we determine the parameter estimates through mediation. In the case of a MISO type logistics model with *n*+1 parameters and Nr, we have short sequences repetitions:

$$\hat{\beta}_i = \frac{1}{N_r}\sum_{k=1}^{N_r}\tilde{\beta}_i(k) = \beta_i + \frac{\sum_{k=1}^{N_r}\varepsilon_i(k)}{N_r} \quad (7)$$

Where i = 0,1,2 ,..., n
From (7) results,

$$M(\hat{\beta}_i - \beta_i)^2 = \frac{N_r^2\sigma^2\varepsilon_i}{N_r} = \frac{\sigma_{\varepsilon_i}^2}{N_r} = \sigma_i^2 \quad (8)$$

where M is mathematical expectancy operator and σ_i is the dispersion estimates of the parameters.

$$\hat{\sigma}_{\varepsilon_i} \approx \left[\frac{1}{N_r - 1}\sum_{k=1}^{N_r}(\tilde{\beta}_i(k) - \hat{\beta}_i)^2\right]^{1/2} \quad (9)$$

For N, >10 the noise dispersion can be approximated by the following relation (9). If samples of length n from a population are extracted, then for values of n> 10 the sample averages are distributed (approximately) normally (according to the central limit theorem [7]). Given (11) it results that the distribution of random values probabilities of the estimates $\hat{\beta}_i$ are asymptotically Gaussian. Thus, you can apply the well known rule of the "three sigma" for determining the estimate probability $P(|\beta_i - \hat{\beta}_i|)$:

$$P\left(|\beta_i - \hat{\beta}_i| < 3\frac{\sigma_{\varepsilon_I}}{\sqrt{N_r}}\right) \cong 0.997 \quad (10)$$

From (10) results that the probability value is very close to one, for the inequality to be fulfilled ,

$$\left(|\beta_i - \hat{\beta}_i| < 3\frac{\sigma_{\varepsilon_i}}{\sqrt{N_r}}\right) \quad .$$

V. CONCLUSIONS

The paper highlights the models of logistic processes particularities with random binary events and presents a technique for identifying these processes. In order to estimate the logistic model parameters, it is necessary to apply the statistical criterion maximum likelihood. Original Contributions:

1. A Monte Carlo method is put forward in order to estimate logistic model parameters using the maximum log likelihood criterion;
2. Statistical analysis of parameters estimate.
3. In conclusion we consider that the theory of systems modelling and identification should be extended to the forecast for nodes failure on the basis of battery energy and others independent variables like sensor age expressed in time interval from putting in action of the battery, etc.

REFERENCES

1. *Tertişco, M , Stoica , P.* Identificarea asistată de calculator a sistemelor , Editura Tehnică, Bucureşti , 1985.
2. *Antyufeev, V. S., and Marshak, A. L.* Monte Carlo method and transport equation in plant canopies, *Remote Sens. Environ.* 31, pp.183-191