STATISTIC AND PROBABILISTIC MEASURES OF DIAGNOSIS LIKELIHOOD ON THE STATE OF SELF-IGNITION COMBUSTION ENGINES

Jerzy Girtler

Politechnika Gdańska
Wydział Oceanotechniki i Okrętownictwa
80-951 Gdańsk, ul. G. Narutowicza 11/12
tel. +48 58 347-2430, tel./fax: +48 58 347-1981, fax: +48 58 341-47-12
e-mail: jgirtl@pg.gda.pl; sek4oce@pg.gda.pl

Abstract

The paper presents the reasons for the need to differ the notions of: diagnosis likelihood and diagnosis rightness at making operating decisions. The formula of probability for formulating the right diagnosis, as the measure of diagnosis likelihood, has been herein derived. For deriving this formula the theory of semi-Markov processes and the Bayes' formula of conditional probability have been applied. Other probabilistic measures of diagnosis likelihood have also been provided. These measures have been referred to technical state of such important systems as e.g. main engines of sea-going ships. However, they can be useful for determining the technical state of other transport means.

1. Introduction

The diagnostic inference [1] enables, in the stochastic decision situation, formulating a diagnosis at determined likelihood. The knowledge of the diagnosis likelihood is necessary for taking a rational operating decision – this is such operating decision which has been worked out by using the optimum calculus [4, 7]. It is also known that each arbitrary decision should be taken only after analyzing the results of its performance. Deciding of which the result is the decision, should be understood as making a non-random choose at work, although, up to the moment of taking the decision, there are used probabilistic and stochastic measures of phenomena, events and processes which occur in the phase of operating the self-ignition combustion engines as e.g. main engines of sea-going ships, and measures of the diagnosis likelihood, too [3, 4, 5]. These measures are necessary to work out decision information which enables making the decision, e.g. the decision on whether determined self-ignition combustion engines such as main engines of sea-going ships can be used to realize a given task or whether their states need prior renovation after which they can be used to realize the task.

The diagnosis likelihood is determined in different ways [5]. It can be accepted that the diagnosis likelihood is:

- *in the descriptive meaning*, a characteristic of a diagnosis which describes the degree of identification of the real (past, actual, future) state (technical, energy – more generally, operating) of a self-ignition combustion engine acting as a diagnosed system (DNS), made by a diagnosing system (DGS).
- *in the valuing meaning*, a characteristic of a diagnosis which is determined by values of essential, in particular cases, indexes that describe the degree of identification of the state of a the self-ignition combustion engine (DNS) by the DGS.

The identification consists in that the DGS classifies the real state of DNS to the known class of diagnostic model states. Such action equals taking a diagnostic decision and in consequence of that – an operating decision.

The indexes that describe the degree of identification of the state of self-ignition combustion engines (DNS), are:

- probability of taking the right diagnostic decision (DG),
- ratio of the expected number of identified (in the fixed time) states of the DNS by the DGS to the expected number of really occurred DNS' states of the same kind (in the same time interval),
- expected value of relative frequencies of working the right diagnosis out.

Accepting that the essential index describing (determining) the degree of identification of the real state of the DNS by the DGS is the mentioned probability of working the right DG out, the diagnosis likelihood can be defined (in the value meaning) as follows: *"diagnosis reliability" is the probability of formulating a right diagnostic decision (the probability of formulating the right diagnosis), so the probability of classifying the supposed real state of the DNS (the state being identified by the DGS) to the class of model diagnostic states, which this real state belongs to and which it should be classified by the DGS to.*

2. Formulating the problem

In the operating practice, decisions concerning the operation of self-ignition combustion engines as main engines of sea-going ships are taken by a user of a diagnosis (a decision-maker) during different phases of the process of diagnosing. This results from the work of the DGS which, in case of the same type of self-ignition combustion engines, can be differently fitted to diagnostic inference. Among operated diagnosing systems (DGS), that are fitted to diagnosed self-ignition combustion engines (DNS) there are such ones which can be named: complex and local [5]. In the both cases the diagnostic inference consists of measuring, symptomatic, structural and operating inferences [1]. In each kind of the inference there are made mistakes which influence the diagnosis likelihood negatively.

From the mentioned reasons it follows that after making *n*-diagnostic tests and inferences with formulated *n*-diagnoses on the state of self-ignition combustion engines (e.g. the technical state) by using the proper DGS, the right diagnoses can be obtained (so, it can be rightly stated that the state of the supposed self-ignition combustion engine is the same as its real state or belongs to the given class of the model states) in the quantity: $m < n$. The other quantity: $k = n - m$ tells that the diagnoses are not right. It means that the measure of formulating right diagnosis (the measure of likelihood), can be accepted as the following quantity:

$$h = \frac{m}{n} = 1 - \frac{k}{n} \tag{1}$$

Considering that at the known number n of expected (planned) diagnostic tests and inferences value taken by m is unknown, that's why the randomness of receiving the right diagnosis about the state of the self-ignition combustion engines should be taken into account while determining the likelihood. Therefore, it can be admitted that elaborating of the right diagnosis is a random event because during realization of a test and the diagnostic inference in determined conditions it may appear but it doesn't have to. When the event of formulating the right diagnosis was often observed in the past, one can assume that it exists a great chance of appearing it in the future (in the same conditions). This chance can be defined with help of the probability of taking the right diagnostic decision (formulating the right diagnosis) which can be, therefore, accepted as a measure of the diagnosis likelihood.

Considering the possibility of formulating the right diagnosis on the state of self-ignition combustion engines in the time interval [0, t] there can be tested two random variables: $N(t)$ which determines the number of possible-to-perform tests and diagnostic inferences, and $M(t)$ which determines the number of possible-to-formulate right diagnoses.

In case of such tests the expected values $E\{N(t)\}$ and $E\{M(t)\}$ can be defined, as well. Thus, the measure of the diagnosis likelihood can be also the quantity which can be called a likelihood index and defined by the formula:

$$w = \frac{E\{M(t)\}}{E\{N(t)\}} \tag{2}$$

The presented measures of the diagnosis likelihood do not reflect clearly the fact that identification of the state S of the self-ignition combustion engines by the DGS is done in the consequence of observing the adequate vector K of values of diagnostic parameters – being generated by the self-ignition combustion engines acting as DNS. They also do not reflect the fact that the DGS can, just like the DNS, be in different states which the diagnosis likelihood depends on. Thus, such a formula determining the probability $P(S/K)$, should be derived in way that would reflect these mentioned facts.

3. Solving the problem.

The process of using the diagnosing system (DGS) is the process $\{W(t): t \geq 0\}$ of which the values can be the elements of the set:

$$D = \{d_1,\ d_2,\ d_3\} \tag{3}$$

with interpretations as follows:

d_1 – state of active using (u) of the DGS (the state of this system's work), which is when the DGS is in the state of the full ability (s_1), thus, d_1 means diagnosing of the DNS state with the help of the DGS when the DGS is in the state s_1, so $d_1 = (u, s_1)$;

d_2 – state of active using (u) of the DGS, which is when the DGS does not stay in the state s_1, but in the state ~ s_1, so in the state of partial ability (s_2) or in the state of disability (s_3), that means the state which makes formulating the right diagnosis impossible, so the state $d_2 = (u, \sim s_1)$;

d_3 – state of active using (u) of the DGS, which is when the DGS stays in the state s_1 and in the same time it occurs the state s_0 of the self-ignition combustion engines (DNS), which has not been considered during diagnostic task performance, so it is the state which cannot be identified by the DGS, so $d_3 = (u, s_1, s_0)$.

It can be accepted that the work time of a DGS being in arbitrary state $d_i \in D$ (i = 1, 2, 3) is a random variable with the distribution $F_i(t) = P\{T < t\}$, continuous density $f_i(t)$ and positive expected value $E(T_i)$. It can be assumed that variables $T_i = (i = 1, 2, 3)$ are mutually independent [5]. In the time T_1 the DGS stays in the state d_1 from which, after finishing the time T_{12}, it can transform to the state d_2 at the probability p_{12} or after finishing the time T_{13} – to the state d_3 at the probability p_{13}. The state d_2 exists in the time T_2 and d_3 – in T_3. The diagnosing system (DGS) can change the state d_2 into the state d_1 in the case when a user finds that the DGS is damaged and immediately makes repairing on it. This change follows after the end of the time T_{21} at the probability p_{21}. The DGS can change from the state d_3 into the state d_1 when a user finds occurrence of the state s_0 of the self-ignition combustion engines as DNS (not identified earlier by the DGS) and immediately makes repairing on it. This change follows after the end of the time T_{31} at the probability p_{31}. In the time of the process $\{W(t)$: $t \geq 0\}$realization different random variables can be observed, that determine the moments $\tau_0 = 0$, τ_1, τ_2 ..., in which changes of the states of the process take place. At the known state of the process $\{W(t)$: $t \geq 0\}$ in the moment τ_n (n = 1, 2, ...), the time of lasting the current state and the state occurring in the moment τ_{n+1} can be identified as stochastically independent from the process states appeared in the moments τ_0, τ_1, τ_2, ... τ_{n-1} and of the time intervals of their duration.

Therefore the process $\{W(t)$: $t \geq 0\}$ can be accepted as the semi-Markov process.

According to [3]:

$$P_1 = \mathrm{E}(T_1)M^{-1} \tag{4}$$

in which: $M = \mathrm{E}(T_1) + p_{12}\mathrm{E}(T_2) + p_{13}\mathrm{E}(T_3)$

where:

p_{ij} – probability of changing the process $\{W(t)$: $t \geq 0\}$ from the state d_i into the state d_j (d_i, $d_j \in D$; i, j = 1, 2, 3; $i \neq j$),
$\mathrm{E}(T_j)$ – expected value of duration of the state $d_j \in D$(j = 1, 2, 3).

The probability P_1 is of the following interpretation:

$$P_1 = \lim_{t \to \infty} P\{W(t) = d_1\} \tag{5}$$

The probability P_1 can be considered as the probability of occurring the event A_1 which determines using a diagnosing system (DGS) in the time of lasting the state d_1 of the process $\{W(t)$: $t \geq 0\}$, so $P_1 = P(A_1)$; $A_1 = \{d_1\}$.

Any state of self-ignition combustion engines (which belongs to the set of the states enclosed in a diagnostic task) can be identified by the DGS when:

- the event A_1 occurs, being the event: when "the state d_1 of the process $\{W(t)$: $t \geq 0\}$is lasting";
- the event K occurs, which determines appearing of a vector of values of diagnostic parameters;
- occurrence of the event K is a consequence of occurrence of the event S which determines occurrence of an important (for a user) state of self-ignition combustion engines, enclosed in the diagnostic task and should be classified to the class of the model diagnostic states

Therefore, the diagnosis likelihood can be defined by the probability of occurrence of the events: A_1, S, K in the same time, according to the following dependences [5]:

$$P(A_1 \cap S \cap K) = P(A_1)P(S \mid A_1)P(K \mid A_1 \cap S) \tag{6}$$

$$P(A_1 \cap S \cap K) = P(K)P(S \mid K)P(A_1 \mid K \cap S) \tag{7}$$

From the equations: (6) and (7) it results that the probability P(S/K) as a diagnosis likelihood measure, can take the following form:

$$P(S|K) = \frac{P(A_1)P(S|A_1)P(K|A_1 \cap S)}{P(K)P(A_1|K \cap S)} \tag{8}$$

Occurrence of the event A_1 doesn't influence the probability of occurring the event S, what is obvious because the events: S and A_1 are independent. That means: $P(S \mid A_1) = P(S)$. If the DGS is reliable (if the process $\{W(t)\colon t \geq 0\}$ is always in the state $d_1 \in$ D), it always occurs the event which consists in occurring the event K, at the assumption that the event A_1 had occurred. In this situation the dependence: $P(K/A_1 \cap S) = P(K \mid S)$ is obtained. Apart from that, having reliable DGS, the event A_1 can be always observed at the assumption that the events: K and S have occurred at the same time. Therefore, it becomes obvious that, in the case of reliable DGS (such the DGS for which $P(A_1) = 1$) it should be taken into account, that $P(A \mid K \cap S) = 1$ and $P(K \mid A_1 \cap S) = P(K \mid S)$. Thus, at the assumption that $P(A_1) = 1$, the formula (8) can be reduced to the following form:

$$P(S|K) = \frac{P(S)P(S|K)}{P(K)} \tag{9}$$

what brings the diagnosis likelihood measure [2].

Assuming that the diagnosing system (DGS) worked without any failure while diagnosing the states of self-ignition combustion engines being diagnosed systems (DNS), the diagnosis likelihood measures can be also expressed by quantities of dependences: (1), (2) and (8).

3. Diagnostic inference and diagnosis likelihood

In the operating practice it exists the necessity of formulating diagnoses on the states of self-ignition combustion engines (e.g. main engines of sea-going ships), as inferences which can be logically deduced from the premises being the values of diagnostic parameters which create the vector K, are recorded by the diagnosing system (DGS) and suggest existing (or just occurrence of) the state S of the mentioned self-ignition combustion engines being diagnosed systems (DNSs). This kind of inference (diagnostic inference) is called the non-deductive inference. Thus, significant becomes the answer to the questions*: in what degree can one trust the inferences being results of a non-deductive inference?, in what degree such inferences can be accepted as reliable and used for taking operating decisions*?

During formulation of the diagnosis (inference) on the state S of the DNS the sentence K is taken for a completely reliable premise. The sentence says that not any other but this vector (in this case, the vector K) of values of diagnostic parameters was recorded by the DGS. The sentence S says that that not any other but this DNS state (in this case, the state S), is the inference formulated on the basis of the sentence K, being the result of the finished non-deductive inference. In that case, the inference is the reductive one [5] which runs in the following schematic way: if the implication $S \Rightarrow$ K is true and its direct successor (K) is true, direct predecessor (S) of the implication is also true. It means that the presumed state of the DNS is taken for S because the vector K of values of diagnostic parameters has been recorded by the DGS.

In case, when the sentence S is the inference formulated on the basis of the sentence K (considered as a completely granted presume) in the process of the non-deduction inference it can be accepted that the sentence S is made probable by the sentence K. The measure of the probability can be the logical probability [8] of the sentence S for the sake of the sentence K. Accepting the sentence $K^*(n)$ as the following sentence: *K^* is the set of n – results of tests and diagnostic inferences of the state S of a diagnosed system,* and $S^*(m)$ - as the sentence: *S^* is confirmed by m- results of tests and diagnostic inferences*, the logical probability of the state S, considering the vector K, can be determined as:

$$P_L(S|K) = \frac{m}{n} = h; \; m \leq n \tag{10}$$

One can easily noticed that the formula (10) is the same as the formula (1) which determines the frequency of a random event. From the formula (12) it results that the degree of likelihood at which the diagnosis is accepted as reliable, can be bigger than the value of the logical probability *h* of the inference saying that the DNS is in the state *S* on the basis of the presume which is the observed vector *K* taken for completely right because of occurrence of (or existing) the state *S*. In case when n $\rightarrow \infty$, the determined by the formula (1) frequency of a random event tends to the statistic probability [8] that. in this case, can be determined by the formula:

$$P_S(S|K) = \lim_{n\to\infty} \frac{m_S}{n} \tag{11}$$

Determining the both probabilities: logical (P_L) and statistic (P_S), it should be taken into consideration that they concern the repeating events: *S* (that the state *S* has appeared) and *K* (that the vector *K* has appeared), which can occur only together. It means that, it is assumed, that the event *S* appears only when *K* occurs.

From the considerations above it results that the frequency *h* determined by the formula (1), can be considered as the logical probability (P_L) when the tests and diagnostic inferences are repeated many times, and when lots of tests and inferences (in the theory n $\rightarrow \infty$) are made it can be considered as the statistical probability (P_S).

Formulating the diagnosis about the state *S* of the SDN, the sentence *S* (which says that the DNS is in the state *S*) is a hypothesis and the sentence *K* is a result of a diagnostic test. The sentence *K** is the set of sentences *K* (the set of results of diagnostic tests), from which all are the tests confirming *m* – times or not confirming (falsifying) *n–m* - times the hypothesis *S*. The hypothesis about the state of the DNS can be, in this case, formulated in the following way: *the SDN is in the state S because the vector K of values of diagnostic parameters is observed.*

The suggested measures of the diagnosis likelihood are objective. When the measures cannot be applied because of different reasons (e.g. technical, economic, organizing) the diagnosis likelihood can be determined with the help of a psychological (subjective) probability [8]. The probability determines the degree of conviction (certainty) of the user of the diagnosis about the chances of coming such expectations true, that the state of the DNS, according to the formula included in the diagnosis, is the state *S*. Acceptance of the diagnosis as a reliable or not reliable one, by using this probability, is subjective because depends on the knowledge of the person who formulates (works out) the diagnosis. It differs from the objective probability at the fact that it reflects the subjective estimation of the SDN state, according to the relation: this state of the DNS is more probable than each other one or – this state of the DNS is the most probable, or it is the most probable that DNS stays in the state *S*, etc. The probability is indeed graduated but deprived of number measures which would determine the particular degree of acceptation. The diagnosis on the state *S* of the DNS can be considered only as more or less reliable. Therefore, the psychological probability (subjective) is not a good measure of diagnosis likelihood. From this reason, for operating practice, this should be of a limited application in case of technical kinds of transport, just like sea-going ships and aircrafts.

4. Summary

Occurrence of the event $A_1 = \{d_1\}$ is the necessary (but not sufficient) condition to be able to identify the state *S* of the given DNS. In case of using a complex DGS, the all states which the SDN can be in are considered at the diagnostic task and then $A_1 = \{ d_1, d_3\}$. Thus in the practice, the diagnosis likelihood (in case of employing the local DGS) can be bigger than $P(A_1) = P_1 < 1$, what follows from the formula (4).

The formula (8) which enables determining the diagnosis likelihood, has been formed by using the limiting distribution of the semi-Markov process $\{W(t)\colon t \geq 0\}$ and the Bayes' formula.

The presented semi-Markov model of the process $\{W(t)\colon t \geq 0\}$ is of the essential practical meaning because of the easiness of determining estimators of the probability p_{ij} as well as simply estimation of expected values $E(T_j)$ of random variables T_j, stating for the time of state duration $d_j \in D$ (j = 1, 2, 3) [7].

The presented measures of diagnosis likelihood are readable because in the extreme cases they can be assigned to by:

– the value 1 when the diagnosis is completely reliable,

– the value 0 when the diagnosis is completely unreliable.

In the cases, when it can be only stated that the diagnosis is reliable at a certain degree, this degree is needed to be précised by assigning a value from the non-negative real numbers interval $R_+ = (0, 1)$ to its likelihood.

References

[1] Bękowski, L., *Elementy diagnostyki technicznej*, Wyd. WAT Warszawa, 1992.

[2] Cempel, C., *Diagnostyka wibroakustyczna maszyn*, Wyd. Politechnika Poznańska, 1985.

[3] Girtler, J., Girtler J.: *Probabilistic measures of diagnosis' likelihood about the technical state of transport means*, Archives of Transport, Quarterly of Committee of Transport of Polish Academy of Sciences, vol. 9, iss. 3-4, Warsaw, 1997, pp.33-42.

[4] Girtler, J., *Zastosowanie bayesowskiej statystycznej teorii decyzji do sterowania procesem eksloatacji urządzeń*, XXII Zimowa Szkoła Niezawodności, Wyd. SPE KBM PAN – ITE, Szczyrk, 1994, s.55–62.

[5] Girtler, J., *Diagnostyka jako warunek sterowania eksploatacją okrętowych silników spalinowych*, Wyd. WSM, Studia Nr 28, Szczecin 1997.

[6] Girtler, J., *Zastosowanie wiarygodności diagnozy do podejmowania decyzji w procesie eksploatacji urządzeń*, Materiały V Krajowej Konferencji „Diagnostyka Techniczna Urządzeń i Systemów" DIAG'2003. WAT, PAN, Warszawa-Ustroń 2003, s.100-109.

[7] Grabski, F., *Teoria semi–markowskich procesów eksploatacji obiektów technicznyc*h, ZN WSMW (AMW), Nr 75A, Gdynia 1982.

[8] Pabis, S., *Metodologia i metody nauk empirycznych*, PWN, Warszawa 1985.