Application of the Bayesian approach in an expert’s system for diagnostic fault in computing engineering

M G Dorrer, A A Popov and A K Ovsyankin
Reshetnev Siberian State University of Science and Technology,
Krasnoyarsk, prospect “Krasnoyarsky rabochy “31, Russia

E-mail: mdorrer@mail.ru, tolynbms@yandex.ru, dikoti1994@yandex.ru

Abstract. The problem of increasing the efficiency of maintenance and repair of computer equipment is solved by applying the method of probability theory in expert systems. An expert system is a set of programs that accumulates the knowledge of specialists in a specific subject area and replicates this empirical experience for consultations of less qualified specialists. A probabilistic analysis method of equipment operation is used in the expert system under development for diagnosing and preventing computer malfunctions. This article provides a detailed description of the expert system, diagnosing the causes of computer malfunctions, provides a structural diagram of the system, rules and facts of work are generated. Based on the well-known Bayesian formula, this article describes a method for using fault information to clarify initial probabilities.

1. Introduction

Research and analysis of existing strategies show that the planned maintenance and repair strategy, (the strategy according to the schedule), leads to a significant cost overrun of labor and material resources [1]. Moreover, this strategy does not reduce the probability of after-repair failures. The solution to this problem is the use of more flexible strategies for maintenance and repair (MRO), which will combine to ensure the reliability of computer technology through long-term diagnostics and forecasting - projective and predictive [2]. The application of such maintenance and repair strategies requires the support of information systems, including mobile ones. The purpose of such systems is to provide access for service engineers to the manufacturer’s documentation on equipment, credentials for its actual state, databases of typical problems and methods for eliminating them. The application of such maintenance and repair strategies provides troubleshooting until the moment when the system has obvious problems.

A preventative maintenance strategy is effectively and fully implemented if the staff has the knowledge, skills and time necessary to carry out the relevant activities.

A monitoring and troubleshooting method is proposed that uses information technology of engineering knowledge, which is practically implemented as an expert system to overcome the indicated difficulties.

The relevance of this theme is the number of failures in the work of computer technology is increasing, and modern methods for their detection do not provide timely diagnostics, it is required to find a new approach for detecting malfunctions. The proposed approach allows you to create a system that clarifies the probability of various failures according to the Bayes formula, with accumulated operating experience.
2. Application of expert systems

The expert system (ES) is a complex of programs that accumulates the knowledge of specialists in a specific subject area and copies this empirical experience for consultations of less qualified specialists [3].

ES can be presented in the form of applied information systems or it can be independent individual shells with universal functions in the role of a human expert or consultant who will propose solutions to the problem situation in a specific subject area. During the design of any information system (IS), an important point is to analyse the organization of the interface part, namely, to organize an IP dialogue with the user. It is especially important during the working an ES because the questions by the user must be understandable and arrive at the right when there is a definite situation. Therefore, the technique of dialogue should be carefully considered at the design stage of the EC.

A typical EC consists of the following main components: solver (interpreter), working memory (RP), also called database (DB), knowledge base (KB), knowledge acquisition components, explanatory and dialog components [4].

The expert system works in two modes: the acquisition of knowledge and problem solving (also called the consultation mode or the mode of using the ES).

Considering the mode of acquiring knowledge, dialogue with the ES is carried out with the help of a knowledge engineer. The problem situation is described by the expert in the form of a set of rules and facts. In any subject area presented to study a problem situation, there are objects that have their own characteristics and values - these are facts. Manipulating facts to resolve a problem situation in the subject area – are the rules. To solve a problem situation, an expert needs to acquire knowledge and fill the system with them. The number of identified facts and rules by an expert directly affects the correctness of the solution of the task from the problem area.

Consider a simple production system that diagnoses a malfunction in computer technology. The block diagram of the system is shown in figure 1.

Given a set of elements, \( F = \{f_{ij}\} \cup \{q_i\} = (f_1, f_2, ..., f_n) \) (1),

which consists of facts: \( F_0 \) - start of diagnosis; \( F_1 \) - capacitors are swollen; \( F_2 \) - capacitors are not swollen; \( F_3 \) - router firmware helped; \( F_4 \) - the firmware of the router did not help; \( F_5 \) - replacing the power supply helped; \( F_6 \) - replacement of the power supply did not help; \( F_7 \) - the case of the device is hot; \( F_8 \) - the case of the device is not hot. Elements \( f_{ij} \) define ordinary declarative knowledge from a specific subject area [5].

In addition, there are \( q_i \) elements that determine the interaction with the external environment, in particular, questions to the user with possible alternatives: \( q_1 \) - did capacitors swell? \( q_2 \) - did the router firmware help?; \( q_4 \) - is the power supply OK ?; \( q_5 \) - Is the device case hot?

In the designed ES in the dialog box, the user is provided with questions concerning the specific problem situation that has arisen. Depending on the choice the user can agree with this fact or not, but the EC will provide the following question: \( q_i = f_{t1}, ..., f_{tk} \).

Products in this system are \( f_{ij} \rightarrow q_m = \{f_{m1}, ..., f_{mk}\} \) (3).

Many facts and products are collected in a certain system, represented as a diagnostic OR column with terminal vertices \( q_5, q_6, q_7, q_8, q_9 \) which contain malfunctions of computing equipment: \( q_5 \) - replace capacitors; \( q_6 \) - the router firmware is out of date; \( q_7 \) - wear of the power supply; \( q_8 \) - overheating of the device; \( q_9 \) - possibly increased load of the router. Figure 1 shows a fragment of such a graph.

In the “OR” column, the orientation of the arcs shows the direction of the output. The natural division of the graph vertices into tiers reflects the depth of the output.
The principle of operation of the designed ES is we get to the top $q_1$ of the initiating question to the user in the form of the corresponding dialog menu after the user accesses the ES: 

$q_1 = \{f_1, f_2\}$. 

Suppose, to the question of the system: “Which of the facts $f_1, f_2$ takes place?” The user answered: $f_1$. As a result, we follow the link $f_1$ to the working field and get to the new vertex - question $q_2$: 

$q_2 = \{f_1, f_2\}$ where this procedure is repeated. 

We find ourselves in one of the terminal vertices in the end, where the user receives a message about the result. 

However, in practice, there are much more frequent cases where the decision maker does not have complete information about the situation, and an assessment of the reliability of decisions is required. One way to evaluate this reliability is to use the Bayesian probabilistic approach [6][7][8].

The knowledge base on a diagnosing ES with this approach contains two types of records: 

VERSION 1: knowledge of specific hypotheses (malfunctions) - $H_1, ..., H_N$; 

VERSION 2: knowledge of evidence (facts) - $F_1, ..., F_M$. 

The entry describing the specific hypothesis $H_i$ in FORMAT 1 has the following structure.

$NAME: P; S; (k_1; p_1^+; p_1^-) ; ...; (k_j; p_j^+; p_j^-)$ 

Here:

Name - the name of the hypothesis (malfunction) $H_i$; 

$P = (H_i)$ - a priori probability of this hypothesis; 

$S$ - the number of evidence (facts) for this hypothesis; 

$k_j$ - certificate number; 

$p_j^+ = P(F_k / H_i)$ - probability of fulfillment of evidence $F_k$, if the hypothesis $H_i$ is true; 

$p_j^- = P(F_k / H_i)$ - the probability of fulfilling the evidence if the hypothesis $H_i$ is not true. 

Knowledge of certificates (facts) $F_k$ must be presented in FORMAT 2 in the following form: 

$F_k$ - certificate number; name of certificate; asked question.
A key element of this methodology is the calculation of the price of certificates $F_k$ by the formula:

$$C(F_k) =$$

$$= \sum_{i=1}^{N} \left[ \frac{P(F_k / H_i) \cdot P(H_i)}{P(F_k / H_i) \cdot P(H_i) + P(F_k / \overline{H_i}) \cdot P(\overline{H_i})} - \frac{(1 - P(F_k / H_i)) \cdot P(H_i)}{(1 - P(F_k / H_i)) \cdot P(H_i) + (1 - P(F_k / \overline{H_i})) \cdot P(\overline{H_i})} \right]$$

(7)

Here $k = 1,\ldots,S$ - numbers of certificates $F_k$ specified in format 2; $i = 1,\ldots,N$ - numbers of hypotheses $H_i$ indicated in format 1, for confirmation of which certificate $F_k$ is required.

Value $C(F_k)$ determines the total amount of the maximum possible changes in the probabilities for all $N$ hypotheses available in the database due to evidence $F_k$.

The inference algorithm consists of the following steps.

1. Calculate all $C(F_k)$, $k = 1,\ldots,S$ select certificate $F_m$ with the highest score:

$$F_m : m = \arg \max_k C(F_k).$$

2. We ask the user a question stored in FORMAT 2 of the selected certificate.

The answer may be: YES, NO, DO NOT KNOW.

3. Depending on the answer, we recalculate the probabilities of all hypotheses related to evidence $F_k$. Recalculation is carried out according to the Bayes formula.

If the answer is YES:

$$P(H_i / F) = \frac{P(F / H_i) \cdot P(H_i)}{P(F / H_i) \cdot P(H_i) + P(F / \overline{H_i}) \cdot P(\overline{H_i})}$$

(8)

In case of a NO answer:

$$P(H_i / \overline{F}) = \frac{(1 - P(F / H_i)) \cdot P(H_i)}{(1 - P(F / H_i)) \cdot P(H_i) + (1 - P(F / \overline{H_i})) \cdot P(\overline{H_i})}$$

(9)

in case of an answer I DO NOT KNOW:

$$P = (H_i / F) = P(H_i)$$

(10)

4. After that, we replace in the array of records about hypotheses of probability $P(H_i)$ by

$$P = (H_i / F) , i = 1,\ldots,N.$$ 

5. Repeat the described procedure (paragraphs 1–4) for all evidence.

6. After interviewing all the evidence, we get the final values of the probabilities of hypotheses, this is the final result.

Consider an example of diagnosing the state of the router based on the above theory.

VERSION 1

Hypotheses $H_i$:

$H_1$: swollen capacitors; 0.1; 5; (1; 0; 0.99); (2; 0.7; 0.05); (4; 0.2; 0.5); (5; 0; 0.99); (6; 1; 0.01)

$H_2$: firmware is outdated; 0.05; 2; (2; 1; 0.01); (6; 0.9; 0.02)

$H_3$: the power supply has broken; 0.01; 3; (3; 0.9; 0.1); (4; 0.25; 0.5); (6; 0.9; 0.02)

$H_4$: overheating of the device; 0.01; 2; (4; 0.01; 0.5); (6; 0.9; 0.02)

VERSION 2
Testimonies (malfunctions) and questions to the user.
Capacitors swell; capacitors swollen?
The firmware did not help; did the router firmware help?
The power supply is not working; Is the power supply operational?
The case of the device is not hot; Is the device case hot?
A large load on the router; Is there a big load on the router?
The router does not work properly; Is the router working properly?

Based on this knowledge base and formula (7), it is possible to form an array of initial prices of evidence (table 1), as well as the probability of hypotheses (table 2) at each step of the decision-making procedure.

| Testimonies (symptoms) $E_k$ | Hypotheses $H_i$ | Step 1 $C_1^k$ | Step 2 $C_2^k$ | Step 3 $C_3^k$ | Step 4 $C_4^k$ | Step 5 $C_5^k$ |
|-----------------------------|-----------------|----------------|----------------|----------------|----------------|----------------|
| $F_1$ Capacitors swelled    | $H_1$           | 0.917          | 0.999          | 0.999          | 0              | 0              |
| $F_2$ The firmware is out of date | $H_1$           | 1.415          | 1.213          | 0              | 0              | 0              |
|                              | $H_2$           | 0.082          | 0.755          | 0.755          | 0.755          | 0              |
| $F_3$ The power supply is not working | $H_3$           | 0.137          | 0.815          | 0.597          | 0.597          | 0              |
|                              | $H_4$           | 0.917          | 0.999          | 1.000          | 0              | 0              |
| $F_5$ High load on the router | $H_1$           | 2.238          | 0              | 0              | 0              | 0              |
| $F_6$ The router does not work properly | $H_1$           | 0.917          | 0.999          | 1.000          | 0              | 0              |

| Hypothesis                      | Step 0 | Step 1 | Step 2 | Step 3 | Step 4 | Step 5 |
|---------------------------------|--------|--------|--------|--------|--------|--------|
| $H_1$: Swollen capacitors       | 0.1    | 0.917  | 0.994  | 0      | 0      | 0      |
| $H_2$: The firmware is out of date | 0.05   | 0.703  | 0.959  | 1.0    | 1.0    | 1.0    |
| $H_3$: The power supply has broken | 0.01   | 0.312  | 0.312  | 0.312  | 0.312  | 0.330  |
| $H_4$: Device overheating       | 0.01   | 0.312  | 0.312  | 0.312  | 0.312  | 0.327  |
3. Discussion
According to the initial a priori probabilities, the prices of certificates are calculated and a dialogue with the system begins. The certificates with the highest price are highlighted (Table 1). As a result of the obtained values in Table 1, we see that the probability estimates at each step changes, which indicates a change in the degree of faith in a particular hypothesis

Step 1. The certificate has the highest price $F_1$: The router is malfunctioning. The question is: Is the router working properly? When answered YES, the probabilities of hypotheses and the prices of evidence are recalculated.

Step 2. The certificate has the highest price $F_2$: The firmware is outdated. Asked: Is the firmware out of date? When answered YES, the probabilities of hypotheses and the prices of evidence are recalculated.

Step 3. The highest price has a certificate $F_3$: A large load on the router. The question is asked: Is the load on the router heavy? When answered YES, the probabilities of hypotheses and the prices of evidence are recalculated.

Step 4. The certificate has the highest price $F_4$: The power supply is defective. The question is asked: Is the power supply OK? When I answer, I DO NOT KNOW the probabilities do not change.

Step 5. The only evidence with a non-zero price $F_5$: The device case is hot. With the answer NO, the probabilities of hypotheses $H_3$ and $H_4$ change.

Diagnosis: Hypothesis $H_1$ is rejected because the posterior probability $P(H_1 / F) = 0$ is less than the a priori probability. This suggests that the capacitors are in order.

Hypothesis $H_2$ is confirmed, since $P(H_2 / F) = 1$. This indicates that the firmware is outdated.

Hypothesis $H_3$ has a probability of $P(H_3 / F) = 0.330$ because it is not known whether the power supply is operational.

Hypothesis $H_4$ has a probability of $P(H_4 / F) = 0.327$ because the device is not hot.

From this we can conclude that the observed evidence fully confirms the hypothesis, which suggests that the malfunction of the router is in an outdated version of the software, which may cause the device to malfunction.

4. Conclusion
Bayesian approach can be applied to all areas of activity related to malfunctions of computing equipment.

Calculations carried out on the available experimental data show that as a result of recalculation, the posterior probability of a particular malfunction can either increase or decrease. After several steps, this algorithm leads to the fact that some faults, the posterior probabilities of which have become very small, are discarded (no longer taken into account), and others are suggested to be corrected.

The proposed relatively simple method will significantly accurately assess the reliability of decisions made by an engineer on computer malfunctions, in contrast to other existing methods. Since often the engineer does not have complete information about the situation, and an assessment of the reliability of decisions is required.

Continuous automated recalculation of a priori probabilities based on the incoming information will significantly improve the quality of forecasting, as well as increase the efficiency of maintenance and repair.

References
[1] "The pros and cons of different maintenance strategies," 2019. [Online]. Available: https://www.hansfordsensors.com/the-pros-and-cons-of-different-maintenance-strategies/. [Accessed 2 June 2019]
[2] Gustav F and Hanna L 2012 *An analysis of maintenance strategies* (Sweden) pp 14-24

[3] Doshina A 2016 Expert system. Classification. Overview of existing expert systems *Young Scientist* B 125 756

[4] Aderonke A 2013 An integrated knowledge base system *I.J. Inf. Tech. Com. Sci* B 1 74-85

[5] Ovsyankin A, Popov A and Dorrer M 2019 Expert system on the mobile platform for diagnosis of faults in the work of IT-equipment *Mater. Sci. Eng. B* 537

[6] Dorrer G 2016 *Methods and systems decision making* (Krasnoyarsk) pp 59-65

[7] Morris D 2016 *Bayes theorem: A visual introduction to beginners* (Blue Windmill Media)

[8] John K 2016 *Doing Bayesian data analysis* (Amsterdam: Elsevier)