# Extracting information from failure equipment notifications – use of fuzzy sets to determine optimal inventory

LES M. SZTANDERA
Computer Information Systems
Philadelphia University
School House Lane and Henry Avenues
Philadelphia, Pennsylvania 19144-5497
UNITED STATES OF AMERICA
sztanderal@philau.edu http://www.philau.edu

*Abstract:* - This paper addresses the use of a data analysis approach to extract information from a large number of failure equipment notifications. Based on that, a fuzzy system, capable of learning and optimizing the knowledge from historical evidence, is formed. Subsequently, its use as a guiding tool in decision making processes at the strategic level (estimation of the number of spare parts based on the warehouse location and type of failure), is outlined. To highlight its advantages, the fuzzy sets approach for spare parts allocation is compared with a probabilistic one.

*Key-Words:* - Computer repair parts; Optimal inventory; Data analysis; Fuzzy sets

## 1 Introduction

### 1.1. Spare parts allocation

Many businesses purchase expensive computer equipment to carry out their services. The computer equipment suppliers that sell these complex systems not only supply the businesses with the product, but must also ensure that the systems remain constantly functional.

However, computer equipment systems invariably fail, generating a demand for repair parts. This demand is extremely time-sensitive, as the computer equipment systems, and the businesses that depend on these systems, become crippled without the timely response of a repair vendor with the appropriate service parts. The consequences in the event of equipment failure or malfunction could cause significant losses to the company.

Policies regarding where and how many repair parts to store are far from obvious. For example, storing many and different types of repair parts close to the businesses' locations will virtually guarantee an extremely rapid response to a failure.

However, a computer equipment company would hardly be able to economically justify such a policy due to the cost of maintaining many warehouses, and of holding a generous number of repairable parts in inventory. In the latter case, holding parts "in storage" causes the computer equipment company to incur an opportunity cost, primarily a function of the company's inability to derive revenue-either directly from the sale of the repair part or indirectly through the part's inherent value-while the part sits in a warehouse unused (potentially becoming obsolete) [1].

### 1.2. Fuzzy sets

The concept of fuzzy sets was conceived at the University of California at Berkley in 1965, and presented as a way of processing data by allowing partial set membership rather than crisp set membership or non-membership. This approach to set theory was not applied until the 70's due to insufficient small-computer capability prior to that time. It was reasoned that people do not require precise, numerical information input, and yet they are capable of highly adaptive control. If feedback controllers could be programmed to accept noisy, imprecise input, or linguistic input, they would be much more effective and perhaps easier to implement. Unfortunately, U.S. manufacturers have not been so quick to embrace this technology while the Europeans and Japanese have been aggressively building real products around it.

In this context, a fuzzy system is a problem-solving system methodology that lends itself to implementation in systems ranging from simple, small, embedded micro-controllers to large, networked, multi-channel PC or workstation-based data acquisition and control systems. It can be implemented in hardware, software, or a combination of both. Fuzzy sets provide a simple way to arrive at a definite conclusion based upon vague, ambiguous, imprecise, noisy, or missing input information, often expressed in linguistic terms. Fuzzy sets approach to control problems mimics how a person would make decisions, only much faster.

Fuzzy sets incorporate a simple, rule-based IF X AND Y THEN Z approach to a solving control problem rather than attempting to model a system mathematically. The fuzzy sets model is empirically-based, relying on an operator's experience rather than their technical understanding of the system. For example, rather than dealing with numerical values in the measured attributes, terms like "IF (input 1 is adjective 3) AND (input 2 is adjective 3) THEN (output is adjective 2)" or "IF (input 2 is adjective 2) AND (input 4 is adjective 1) AND (input 5 is adjective 5) THEN (output is adjective 6)" are used. These terms are imprecise and yet very descriptive of what must actually happen. Fuzzy sets are capable of mimicking this type of behavior at very high rate.

Fuzzy set approach requires some numerical parameters in order to operate, such as what is considered significant error and significant rate-of-change-of-error, but exact values of these numbers are usually not critical unless very responsive performance is required in which case empirical tuning would determine them.

Fuzzy sets approach was conceived as a better method for sorting and handling data but has proven to be a excellent choice for many system applications since it mimics human logic. It can be built into anything from small, hand-held products to large computerized process control systems. It uses an imprecise but very descriptive language to deal with input data more like a human operator. It is very robust and forgiving of operator and data input and often works when first implemented with little or no tuning.

## 2 Spare parts stock level calculations – item approach

In order to appreciate the advantages of fuzzy rule based approach, we will first step through a set of examples detailed by Fukuda in [2] using Poisson probability process. The purpose of the spare computer parts stock level calculations – item approach is to describe a technique to calculate the spare parts quantity (for a given computer cluster size and inventory) taking into account item reliability (that can be initially modelled by Poisson probability process).

Variables considered in this approach, after Fukuda [2], were:
1. Reliability of item to be spared (expressed as failure, removal, or replacement; or inversely as Mean Time Between Removals (MTBR) in usage hours)
2. Number of items installed per machine (indicated as A)
3. Required probability that a spare will available when needed, that is, the chance of having a spare part in inventory when required ($90\% \leq P \leq 95\%$), also called Fill Rate or confidence level
4. Number of machines to be supported N
5. Period to be supported as operational time or between initial and subsequent order (time T in months)
6. Average machine utilization (M in hours per month or day per machine). It may be expressed in %, e.g., 9.5 h/day = 9.5 h/24h = 39.58%

Spare parts are divided, for application of this approach, to repairable and non-repairable. For repairable parts, a stock level of spare parts was calculated to compensate items undergoing the process of repair.
7. For repairable items an average time it takes to repair (indicated as time between repairs RT) was considered instead of time of support T of point 5 above. Also, it is advisable to take into account an additional stock level of spare parts to compensate for unavoidable scrap of some repairable items.
8. The scrap rate of a repairable item can be estimated from historical evidence, for example, R = 0.15 or 1.5%.

Poisson distribution can be used in spare parts quantity determination as a forecasting technique with the use of reliability analysis [2]. The demand for spare parts for covering replacement of failed items, occurring as a result of maintenance action, is an event described as Poisson distribution process, assuming it occurs at a constant average rate and the number of events occurring in any interval is independent of the number of events occurring in any other time interval. These rather strict requirements are replaced in fuzzy systems approach.

For the purpose of reliability analysis, it is assumed

$$f(x;\lambda,t)=\frac{(\lambda t)^x e^{-\lambda t}}{x!}$$

where:
λ is a failure rate
t is time (or total operational period of all items)
x is a number of failures (or number of spare parts required)
λt is a mean value (number of failures in time t)
In terms of the probability of n or fewer failures in time t, we have [2]:

$$R(t)=\sum_{x=0}^{n}\frac{(\lambda t)^x e^{-\lambda t}}{x!}=e^{-\lambda t}\left[1+\lambda t+\dots\dots\dots\dots+\frac{(\lambda t)^n}{n!}\right]$$

## 2.1 Non-repairable items

After Fukuda [2], for these items the number of failures is equal to number of spare parts. The quantity of spare parts is the minimum value of n that satisfies the following condition as close as possible:

$$P=<\sum_{x=0}^{n}\frac{(\lambda t)^x e^{-\lambda t}}{x!}=e^{-\lambda t}\left[1+\lambda t+\dots\dots\dots\dots+\frac{(\lambda t)^n}{n!}\right]$$

where t is a total operating time for all items. Thus, attributing to Fukuda [2], the number of failures in time t can be expressed as:

$$\lambda t=\frac{1}{MTBR}t=\frac{AxNxMxT}{MTBR}=\frac{AxNxM}{MTBR}T$$

It is worth noting [2] that the above equation can be also directly used as a deterministic method for computing spare parts quantity. T may be mean resupply time, order, or production lead time.

Numerical example (non-repairable items)

Following Fukuda [2], let us calculate the spare parts quantity of a non-repairable item which is installed on 4 units per machine (A = 4EA), and having a mean time between removals of 7,500 usage hours (MTBR = 7, 500 h) for a computer cluster of 2 machines (N = 2) operating each one for 225 usage hours per month (M = 225 h/month/machine), and considering an initial period of 2 years (T = 24 months) to achieve a confidence level of 90% (P = 0.90).

$$\frac{AxNxMxT}{MTBR}=\frac{(4EA)x(2ac)x(225FH/ac/month)x(24month)}{7.500FH}=5.76EA$$

Proceeding recursively, we have the following:

for 0 spare parts, P = exp(-5.76) = 0.003 = 0.3% < 90%

for 1 spare part, P = 0.003(1+5.76) = 0.02 = 2% < 90%

for 2 spare parts, P = 0.003(6.76+16.6) = 0.07 = 7% < 90%

for 3 spare parts, P = 0.003(23.36+31.85) = 0.17 = 17% < 90%

for 4 spare parts, P = 0.003(55.21+45.9) = 0.303 = 30.3% < 90%

for 5 spare parts, P = 0.003(101.1+52.8) = 0.462 = 46.2% < 90%

for 6 spare parts, P = 0.003(153.91+50.7) = 0.614 = 61.4% < 90%

for 7 spare parts, P = 0.003(204.6+41.7) = 0.74 = 74% < 90%

for 8 spare parts, P = 0.003(246.3+30.1) = 0.83 = 83% < 90%

for 9 spare parts, P = 0.003(276.4+34.2) = 0.932 = 93.2% > 90%

thus obtaining the recommended quantity of 9 spare parts, that could be used in our decision making process at the tactical level (optimal inventory of non-repairable spare parts at a given warehouse).

### 2.2 Repairable Items

For these items the number of failure is different from number of spare parts. As indicated in Fukuda [2], it is a direct application of Palm's theorem. The stock level of spare parts has to compensate for repairable items undergoing the process of repair and could be expressed by [2]:

$$\lambda t = \frac{1}{MTBR} t = \frac{AxNxMxRT}{MTBR} = \frac{AxNxM}{MTBR} RT$$

To replace the very first failed item it is necessary to have an initial single spare part. Thus, substituting n by (n-1) in previous equation, we obtain the following:

$$P =< \sum_{x=0}^{n-1} \frac{(\lambda t)^x e^{-\lambda t}}{x!} = e^{-\lambda t}\left[1 + \lambda t + \ldots\ldots\ldots\ldots\ldots + \frac{(\lambda t)^{n-1}}{(n-1)!}\right]$$

<u>Numerical example (repairable items)</u>

As pointed out by Fukuda [2], considering the example of non-repairable items, we can calculate a number of spare repairable parts taking into account repair time of 3 months (RT = 3 months) [2]:

for 0 spare parts, P = exp(-0.72) = 0.487 = 48.7% < 90%

for 1 spare part, P = 0.487(1+0.72) = 0.837 = 83.7% < 90%

for 2 spare parts, P = 0.487(1.72+0.259) = 0.963 = 96.3% > 90%

thus obtaining the recommended quantity of (n-1) = 2 or n = 3 spare parts, that could be used again in our decision making process at the tactical level (optimal inventory of repairable spare parts at a given warehouse).

### 2.3 Repairable items with Scrap Rate

Repairable items returned to shop are sometimes scraped, that is, not repaired because through inspection, it is decided that some items were not economically feasible to repair. The scrap generally results in an increase of spare parts requirement. For this scenario, the approach could be a combination of repairable and non-repairable methods [2].

<u>Numerical example (scrap rate)</u>

Let us calculate the number of spare parts, if an item is considered a repairable with scrap rate of 10% (R = 0.10).

Previously (from the non-repairable item numerical example), we had 9 failures in the period requiring maintenance, thus 0.1 x 9 ~ 1 EA becomes scrap. From the repairable item numerical example, the total number of parts to support the operation was 3 EA. Adding 1 EA discarded now results in 4 spare parts [2].

<u>Note:</u> All the above described <u>item</u> approaches can be extended to <u>system</u> approaches, where the objective would be to solve the system maintenance problem determining the maximum total fill rate, or alternatively, minimum total expected back order in terms of repairable spare parts given a specific amount of allocated funds.

## 3 Fuzzy Rule-based System

The probability approach [2], as outlined above, requires that an event is described as Poisson distribution process, assuming it occurs at a constant average rate and the number of events occurring in any interval is independent of the number of events occurring in any other time interval. Thus, our research was focused on deriving an optimized fuzzy rule-based system from real life failure data, where those strict requirements are relaxed. Such a system would be a great asset to any decision maker dealing with the estimation of the number of spare parts based on the warehouse location and type of failure (strategic decisions).

Data used in the analysis were a courtesy of Prof. Garth Gibson, Computer Science Department at Carnegie Mellon University, and its Computer Failure Data Repository (CFDR) [3]. It was acknowledged that with the growing scale of today's IT installations, component failure was becoming a significant problem. Yet, very little data on failures in real systems were publicly available, forcing researchers working on system reliability to base their work on simulated, rather than empirical data.

The computer failure data repository (CFDR) [3] aims at accelerating research on system reliability by providing a collection of public data with detailed failure from a variety of large production systems.

In the reported research, memory hardware failure data collected on a 212 node server farm at internet services cluster ask.com from December 06 through February 07 were used. The data themselves were collected by the University of Rochester researchers who contributed to the repository [4]. The set consisted of 1,698 seven dimensional vectors. The dimensions were as follows:

• **Machine ID**: generic machine name

• **First seen**: when the error was first observed

• **Last seen**: when it was last observed

• **Times**: how many times the error in that error correction code (ECC) word was reported during the whole monitoring period

• **Address**: the system address of the error, that is physical memory address

• **Row/Column/Bank**: row/column/bank address of the error, translated from the system address

• **Syndrome**: error syndrome, which can be used to determine which bits are wrong within an error correction code ECC word

(Note: ECC (error correction code) allows data that are being read or transmitted to be checked for errors).

### 3.1 Dimensionality reduction

Dimensionality reduction techniques (data analysis algorithms) were employed to identify the most contributing dimensions from the set [5].
One of the approaches used was regression analysis, and it is highlighted below. Yet another approach could be a count of the repeating fuzzy rules to determine the retained ones.
Though all the independent attributes contribute in the regression equation towards estimation of the number of spare parts, some of the attributes contribute more than others. If the attributes that contribute the least to the prediction ability of the regression equation are eliminated, the overall dimensionality of the data set comes down. Using stepwise regression analysis, one can reduce the dimensionality of the data, keeping only the most influential properties while eliminating properties contribute minimally to the prediction ability of regression equation. In addition to using Beta standardized coefficients as indicators while determining the importance of each attribute in the regression equation, partial correlation coefficients can also be used as indicators as they are measures of association between an attribute and the number of spare parts when the influences of the rest attributes are removed. Similar to the Beta coefficients, their absolute magnitudes reflect their relative degree of association with estimation of the number of spare parts. The absolute magnitude of partial coefficient ranges from 0 to 1. A partial correlation coefficient with a value close to 1 is an indication that the attribute is strongly associated with the number of spare parts, and the opposite is true for partial correlation coefficients with values close to 0.
It was found out that the most important features were: location (machine ID), error type (error syndrome), and count (how many times the error in that error correction code (ECC) word was reported during the whole monitoring period).

### 3.2 Generation of fuzzy rules

Subsequently, fuzzy rule generation algorithm was used to generate a knowledge base for decision making processes. Correlation analysis approach (as opposed to data point rules) was used to come up with an optimized knowledge base. In practical terms it means clipping rule consequent fuzzy sets by the degree of truth of the rule's antecedent.

Thus, instead of generating an impractical set of 1,698 rules, the system came up with an optimized knowledge base with 12 rules that could be used in the decision making processes at the strategic level (estimation of the number of spare parts based on the warehouse location and type of failure). The obtained knowledge base is listed below.

Fuzzy Rules (optimized):

IF location IS location_M1 AND error IS error_type_0x0 THEN count IS around_6;

IF location IS location_M1 AND error IS error_type_0x1 THEN count IS around_60;

IF location IS location_M1 AND error IS error_type_0x2 THEN count IS around_3;

IF location IS location_M1 AND error IS error_type_0x3 THEN count IS around_50;

IF location IS location_M6 AND error IS error_type_0x0 THEN count IS around_3;

IF location IS location_M6 AND error IS error_type_0x1 THEN count IS around_1;

IF location IS location_M6 AND error IS error_type_0x2 THEN count IS around_3;

IF location IS location_M6 AND error IS error_type_0x3 THEN count IS around_3;

IF location IS location_M8 AND error IS error_type_0x0 THEN count IS around_1;

IF location IS location_M8 AND error IS error_type_0x1 THEN count IS around_3;

IF location IS location_M8 AND error IS error_type_0x2 THEN count IS around_1;

IF location IS location_M8 AND error IS error_type_0x3 THEN count IS around_1;

## 4 Conclusion

Because repair parts can be extremely valuable, even minor reductions in repair parts inventory can result in significant savings for the computer equipment company. Therefore, research investigating optimal inventory policies is important. In addition to determining the stocking levels for each part at each warehouse, questions concerning the number and location of warehouses must also be answered. Different types of warehouses must be considered. Some only hold inventory, while others hold inventory and repair broken parts.

A final level of complexity presents itself in determining the way in which warehouses interact, e.g., whether one warehouse ships parts to another stocked out warehouse [6, 7].

*References:*


[1] B. Schroeder and G. Gibson, "Disk Failures in the Real World", in Proceedings of FAST'07, 5th USENIX Conference on File and Storage Technologies, San Jose, CA, 2007, pp. 1-16.

[2] J. Fukuda, "Reliability, Excel, Kettelle's Algorithm, and Inventory of Repairable Spare Parts", February, 2008.

[3] G. Gibson and B. Schroeder, "Computer Failure Data Repository (CFDR)", http://cfdr.usenix.org/, 2009.

[4] X. Li, M. C. Huang, K. Shen, and L. Chu, "An Empirical Study of Memory Hardware Errors in a Server Farm", HOTDEP Workshop, 2007.

[5] L. Sztandera, "Identification of the Most Significant Properties Influencing Tactile Fabric Comfort Using Regression Analysis", WSEAS Transactions on Computers, Vol. 2(8), 2009, pp. 302-311.

[6] L. Sztandera, "Optimal Inventory of Computer Repair Parts – A Fuzzy Systems Approach", in Proceedings of PRDC'10, Tokyo, Japan, pp. 225-226.

[7] L. Sztandera, "Spare Parts Allocation – Fuzzy Systems Approach", in Proceedings of the 15th WSEAS International Conference on Computers, Corfu, Greece, pp. 245-249.