# USING INDUCTIVE MACHINE LEARNING TO SUPPORT QUALITY MANAGEMENT

**L. Tsironis[1], N. Bilalis[1], V. Moustakis[1,2]**
*[1]Technical University of Crete*
*Department of Production & Management Engineering*
*[2]Institute of Computer Science, Foundation for Research and Technology, Heraklion, Greece.*
*E-mail: {loukas,moustaki}@logistics.tuc.gr; bilalis@dpem.tuc.gr*

## Abstract

This paper examines the enhancement of the available quality improvement tools. In real world environments there is not any way of monitor the whole production process and make decisions about its quality level. The use of machine learning and knowledge discovery techniques enable individual to monitor the progress of a production process by managing quality control data. The proposed method tested in collected data of production process of a large European Telecommunications company.

## Machine Learning

The field of Machine Learning studies computational methods for acquiring new knowledge, new skills and new ways to organize existing knowledge [Michalski 1983], [Shapiro, 1987]. When a computational system improves its performance without additional programming, then the system consider of having learning skills [Forsythe, 1986].

Michalski [1983] reports that the study and modeling of learning process through, computational systems (computer modeling), with its multiply occurrence composes the scientific field of machine learning.

Machine learning enables computer systems to draw knowledge from the studied data so that the decisions given by the system are based on the quality of data. Data analyzed in order to produce a decision. This analysis is based in the combination of statistics, advanced heuristic methods and artificial intelligence algorithms in order for the machine learning system to achieve its goal.

Learning systems are trying to understand their environment using simplifications of it. These simplifications are called *models*. The process of model construction is called *Inductive Learning*. During learning phase the learning system observe its environment and try to detect similarities between objects and facts. The system classifies similar objects in *classes* and makes predictions for the behavior of objects within classes through the production of *rules*.

Most machine learning systems search for relationships among attributes of an object and not among objects.

Machine learning system does not interact directly with its environment but it uses a subset of its environment, called *Training Set*. Production rules produced after an iterative search strategy. Training set contains examples and observations coded in an appropriate form in order the system to use them.

**From machine learning to knowledge discovery**

When the training set is a whole data base then the process of acquiring knowledge called knowledge discovery from databases.

Knowledge discovery in databases is the non-trivial process of identifying valid, novel, potentially useful and ultimately understandable patterns in data [Fayyad, 1996].

Data are a set of cases in database, pattern is an expression which describes the cases. The process of knowledge discovery includes several stages such as: data preparation, pattern search, knowledge evaluation, and knowledge refinement. The discovered patterns must be valid in new data with a degree of consistency. The consistency degree is a function of the pattern and the cases database contains.

Patterns must be useful that means that they should lead the user in useful decisions. The goal of the overall process should be the production of understandable from humans patterns in order to enforce their ability to understand the database.

Algorithm builds a model form database. The task of the algorithm is to produce rules from the relationships that govern data base objects.

The central goal of knowledge discovery is the same with the one of machine learning. With this point of view Matheus [1993] remarks that knowledge discovery contains techniques such as clustering, rule induction, finding dependency networks, analyzing changes. These techniques are also known as machine learning techniques and from this point of view there is no deference between knowledge discovery and machine learning.

The process of knowledge discovery is an iterative and interactive with the user process which contains a number of stages and decision that concern the user. In figure 1 Fayyad [1996] gives a practical view of the process.

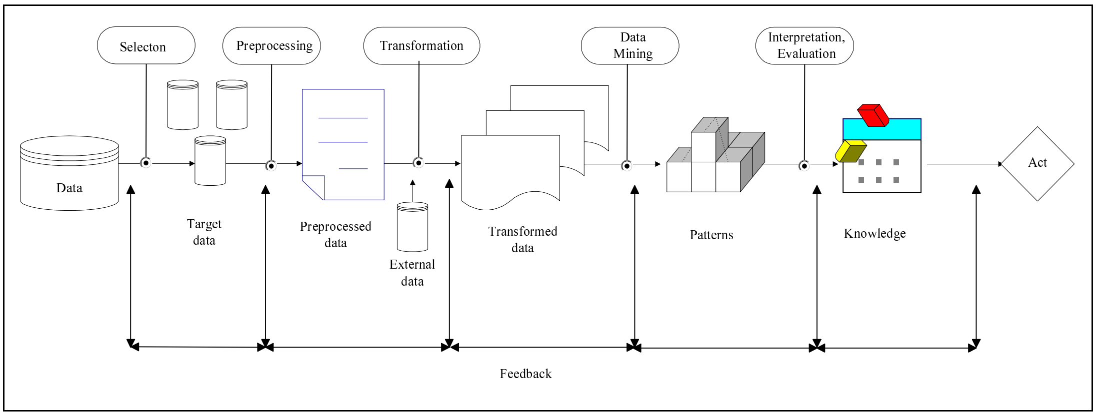


**Figure 1: The process of knowledge discovery from databases**

## Manufacturing process control with Machine Learning and Data Mining Techniques

The traditional way to assure the quality within a manufacturing process is the application of SPC. SPC contains tools such as affinity diagrams, cause and effect diagrams, pareto charts, systematic diagrams, control charts and so forth. SPC enables someone to monitor the stability of the process and proceed to corrective actions when a quality characteristic is outside of specifications. In general SPC tools are problem solving techniques which aims the humans to:

1. Locate the defectives
2. Locate the problem
3. Finding the causes of the problem
4. Evaluate the proportion of the damage which each cause concern
5. Monitor the progress of the quality characteristics

When the manufacturing process is more complicated the degree of uncertainty is bigger. In situations like this humans with great skills and knowledge in statistical techniques are necessary to prevent or face a problem. An important fact is that during the progress of the manufacturing process knowledge is producing. This knowledge is the experience that human needs in order to reinforce their ability to understand the process and solve the problems arise. Under these conditions its very difficult for the humans to acquire this knowledge. One way to solve this problem is to replace humans with intelligent computer programs. The international literature shown that Artificial Intelligence (AI) has the ability to solve problems in complex and uncertain manufacturing environments. AI has gain its popularity because of the ability of its algorithms to learn through experience and to manage the uncertainty. An important advantage is that a naive user can use easily such programs and can produce accurate results and be able to solve problems.

## Applications of artificial intelligence systems in process control

Kloesgen, [1996] reports that in a chemical company monitor quality parameters (temperature, pressure, chemical reaction) of the process and try to determine the quality of the final product.

Zhang [1992], examine statistical variability in electronic circuits. During the continuous size reducing circuits are sensitive in statistical variation. A methodology based in a neural network predicts this variation of the overall manufacturing process.

Khanchustambham [1992] presents a decision support system for on-line monitoring of manufacturing processes. This approach based on the interpretation of a sensor signal. A neural network analyzes this signal and makes decisions for the appropriate control action.

Pan [2000] uses neural networks in pyro-metallurgy processes. Neural networks give good results not only in behavior prediction but in control and optimization.

Leech [1986] describes an inductive algorithm used in control a chemical process. The algorithm induces decision trees and produces rule which guide to solve problems concerning the achievement of the process tasks.

Vivarelli [1998] depicts a model based in a Bayesian neural network. Neural net used for the modeling of control variables of antibiotic cephalosporin concentration in a fermentation process. The identification of input relationships allows the understanding of the quality of the information that supply the neural network and the intrpretation of facts that concern the chemical process.

## The application domain

The application domain is a large European telecommunications company. Specifically we apply data mining in a terminal devises assembly process. The flow chart of the process presented in figure 2.

The overall process concerns of placement parts on a card, parts are either surface mount device (SMD), radial and axial. The process is an automatic assembly except for the manual stage, where the bigger parts (capacitor) placing manually by humans.

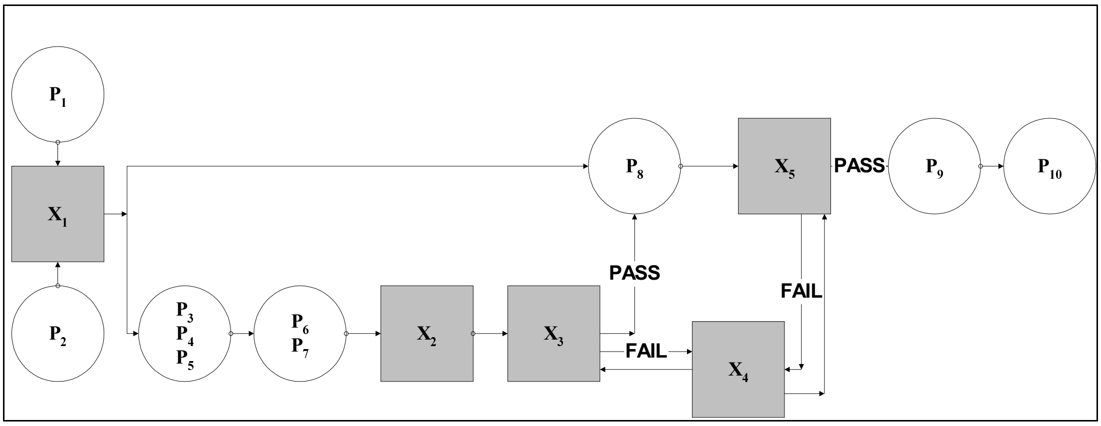


**Figure 2: Production process flowchart**

The quality characteristics concerned under monitoring are: Right solder placement, Right part placement, Absence of a part, Right soldering, Right direction of a part, Faulty part, Waste part, Power supply, Voice encoding, Signal measurement, Voltage, Short circuit

There are exact specifications for the above characteristics. The values of the specifications are qualitative. When a value is find to be outside of the specifications then we have the production of defects, which detected from control stations.

The goal is to eliminate defects. For that purpose we used two data mining techniques, classification rules and association rules.

A famous program that produces classification rules is Quinlan's C5, based on the ID3 algorithm [Quinlan, 1983], [Quinlan, 1986]. C5 produces classification rules from description of application domain and from a set of examples (cases). Descriptions of application domain are specific values of defects, the control station that detects the defect and the exact part which was defected. The number of classes is equal to the number of production stations. The decision made from a classification rule is concern the production station which is responsible for the defect occurrence.

A general classification rule is of the form:

IF control station X detect a defect Y in part Z THEN production station K is responsible. Each rule consists of:

A rule number -- this is quite arbitrary and serves only to identify the rule.

Statistics (*n*, lift *x*) or (*n/m*, lift *x*) that summarizes the performance of the rule. Similarly to a leaf, *n* is the number of training cases covered by the rule and *m*, if it appears, shows how many of them do not belong to the class predicted by the rule. The rule's accuracy is estimated by the Laplace ratio *(n-m+1)/(n+2)*. The lift *x* is the result of dividing the rule's estimated accuracy by the relative frequency of the predicted class in the training set.

One or more conditions that must all be satisfied if the rule is to be applicable.

A class predicted by the rule.

A value between 0 and 1 that indicates the confidence with which this prediction is made. (Note: If boosting is used, this confidence is measured using an artificial weighting of the training cases and so does not reflect the accuracy of the rule.)

C5 has the option of cross-validation. This is one way to get a more reliable estimate of predictive accuracy. The cases are divided into *f* blocks of roughly the same size and class distribution. For each block in turn, a classifier is constructed from the cases in the remaining blocks and tested on the cases in the hold-out block. In this way, each case is used just once as a test case. The error rate of a classifier produced from all the cases is estimated as the ratio of the total number of errors on the hold-out cases to the total number of cases.

For the purpose of defects elimination used a second technique called association rules. An association rule represents an association relationship among a set of patterns (values) in a database [Agrawal, 1996]. The program we used called Magnum Opus. Magnum Opus produces association rules from data. An association rule identifies a combination of attribute values or items that occur together with greater frequency than might be expected if the values or items were independent of one-another.

A Magnum Opus association rule has two parts, a Left Hand Side (LHS) and a Right Hand Side (RHS). The LHS is a set of one or more attribute values or items. The RHS is a single attribute value or item. Each association rule indicates that the frequency with which the RHS occurs in the data is higher among examples that have the LHS attribute values or items than among those that do not.

The association rules resulted is on the form:

IF station X detects a defect Y in part Z THEN the responsible station for the defect is station K.

A number of values are listed with each association rule:

Coverage: The proportion of examples covered by the LHS.

Support: The proportion of examples jointly covered by the LHS and RHS.

Strength: The proportion of examples covered by the LHS that are also covered by the RHS.

Lift: The strength divided by the proportion of all examples that are covered by the RHS.
Leverage: The proportion of additional examples covered by both the LHS and RHS above those expected if the LHS and RHS were independent of each other.

**Data**

The set of data collected from reports of the control station. It concerns defects in parts. One card may have several defected parts because on it there are hundreds of them. When a card is defected it leads to the repair station and a report is formed for the corresponding card from the control station.
A set of 600 defected cards are collected and used in this study. Data are stored in a relational table and coded in an appropriate form for Magnum Opus can uses them.

**Results**

The classes are the production stations. That means that the cases classified according to a production station which produces the corresponding defect. The production station in this case is the decision variable. The decision variable was decided from the company.
C5 produced 146 rules from almost 1000 cases classified in 7 classes. A representative set of rules is given in figure 3.

1. Rule 1: (4, lift 59.8) quality control station X2, defect Pl1A, part AP3 = > class P3 [0.833]
2. Rule 10: (21, lift 11.9) quality control station X2, defect Pl11A, part AP8 = -> class P4 0.957]
3. Rule 39: (9, lift 18.6) quality control station X3, defect Pl2S, part SP26 = -> class P6 [0.909]
4. Rule 57: (131, lift 2.3) quality control station X1, defect G8S, part SP11 = > class P1 [0.992]
5. Rule 97:(4, lift 37.8) quality control station X4, defect G4S, part SP31 = > class P2 [0.833]
6. Rule 109: (51/1, lift 9.0) quality control station X2, defect D25A, part AP1 = > class Pext [0.962]
7. Rule 121: (164/1, lift 3.4) quality control station X4, defect G8S, part SP36 = > class P5 [0.988]

**Figure3: Set of classification rules**

Table 1 gives the results of 10 cross validation folds.

| **Folds** | **Rules** | |
|---|---|---|
| | **No** | **Errors** |
| 0 | 146 | 4.20% |
| 1 | 140 | 13.70% |
| 2 | 146 | 6.30% |
| 3 | 142 | 12.60% |
| 4 | 144 | 7.30% |
| 5 | 143 | 11.50% |
| 6 | 142 | 11.50% |
| 7 | 145 | 7.30% |
| 8 | 144 | 7.30% |
| 9 | 144 | 10.40% |
| **Mean** | 143.6 | 9.2% |
| **SE** | 0.6 | 1.0% |

**Table 1: Summarization of error rate after 10 cross-validation folds**

Table 1 estimates the error rate of both the decision tree and ruleset classifiers produced from the 1000 cases. The *SE* figures (the standard errors of the means) provide an estimate of the variability of these results.

Table 2 is the *confusion matrix* that pinpoints the kinds of errors made. For example in the second row 54 cases are classified correct when one case misclassified in class C 20 misclassifed in class D and 2 in class F.

| A | B | C | D | E | F | G | Classified as |
|---|---|---|---|---|---|---|---|
| **4** | | | 10 | | | | Class A |
| | **54** | 1 | 20 | | 2 | | Class B |
| | 1 | **32** | 14 | | | | Class C |
| | 1 | | **413** | | 1 | 4 | Class D |
| | | | 9 | **12** | | | Class E |
| | 2 | | 6 | | **94** | | Class F |
| | | | 17 | | | **259** | Class G |

**Table 2: Confusion matrix**

Figure 4 gives the representative association rules from Magnum Opus.

1. quality control station X3, defect G8S, part SP36 => production station = P4 [Coverage=0.204 (159); Support=0.203 (158); Strength=0.994; Lift=3.14; Leverage=0.1385 (107)]
2. quality control station X2, defect G8S, part AP11 => production station = P1 [Coverage=0.096 (75); Support=0.096 (75); Strength=1.000; Lift=7.70; Leverage=0.0839 (65)]
3. quality control station X6, defect D25A, part AP1 => production station = Pext [Coverage=0.067 (52); Support=0.066 (51); Strength=0.981; Lift=9.19; Leverage=0.0584 (45)]
4. quality control station X7, defect P2S, part SP32 => production station = P1 [Coverage=0.013 (10); Support=0.013 (10); Strength=1.000; Lift=19.45; Leverage=0.0122 (9)]
5. quality control station X8, defect P1A, part AP3 => production station = P3 [Coverage=0.005 (4); Support=0.005 (4); Strength=1.000; Lift=55.57; Leverage=0.0050 (3)]
6. quality control station X4, defect P11A, part SP23 => production station = P3 [Coverage=0.009 (7); Support=0.009 (7); Strength=1.000; Lift=11.11; Leverage=0.0082 (6)]
7. quality control station X5, defect P32S, part SP30A => production station = P2 [Coverage=0.008 (6); Support=0.001 (1); Strength=0.167; Lift=21.61; Leverage=0.0012 (0)]

**Figure 4: Association rules**

**Conclusions**

In this paper presented a real world application of machine learning techniques as a quality improvement tool. Machine learning program constructs rules that can use as a problem-solving tool. Rules can guide decision makers to the source of defect producing in an automatic way.

The proposed method enhance the available SPC tools rather than eliminate them. Machine learning is a fast and intelligent way of monitoring a production process. The method collects quality control data from quality control stations. Data stored in a relational database. The machine learning algorithm analyze these data and produce decisions about the specific stage of the process that is responsible for the variation of the specifications.

In Total Quality Management environments a fundamental principle is to be able to monitor a production process and its progress in terms of a number of quality laws. For that reason the process should analyze and should surround with the appropriate and trained people whose task is to monitor its progress through the experience they gain. The proposed method manage this knowledge, derived from experience, and guide people to make decisions about quality improvements within the process.

## References


**Agrawal** R., Mannila H., Srikant R., Toivonen H, Verkamo A., 1996, "Fast discovery of association rules", in Fayyad U., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (Eds), "Advances in Knowledge Discovery and Data Mining", AAAI Press / MIT Press

**Fayyad** U., Piatetsky-Shapiro G., Smyth P., Uthurusamy R. (Eds), 1996, "Advances in Knowledge Discovery and Data Mining", AAAI Press / MIT Press

**Forsythe** R., Rada R., 1986, "Machine Learning: Applications in Expert Systems and Information Retrieval", Ellis Horwood Limited, a Division of John Willey & Sons

**Khanchustambham** R., Zhang G., 1992, "A neural network approach to on-line monitoring of a tuning process", Technical Report (TR 92-93)

**Kloesgen** W., 1996, "Tasks, methods and applications of knowledge extraction", In New techniques and technologies for statistics, Kloesgen W., Nanopoulos P., Unwin Q., (Eds.) IOS Press, Amsterdam

**Matheus** C., Chan P., Piatetsky-Shapiro G., 1993, "Systems for Knowledge Discovery in Databases, IEEE Transactions on Knowledge and Data Engineering, V. 5, No. 6

**Michalski** R., Carbonell J., Mitchell T. (Eds), 1983, "Machine Learning: An Artificial Intelligence Approach", Tioga Publishing Company, California

**Pan** X., Livneh B., " Neural networks to predict and control pyro-metallurgical processes", http://www.kbe.co.za/articles/nnets/pyrometal/pyrometal.htm

**Quinlan** R., 1983, "Learning efficient classification procedures & their applications to chess endgames", in Michalski R., Carbonell J., Mitchell T. (Eds), "Machine Learning: An Artificial Intelligence Approach", Tioga Publishing Company, California

**Quinlan** R., 1986, "Induction of decision trees", Machine Learning, V. 1, pp 81-106

**Shapiro** S. (Ed), "Encyclopedia of Artificial Intelligence", 1987, V. 2, Willey Interscience, USA,

**Vivarelli** F., 1998, "Studies on generalization in Gaussian processes and Bayesian neural networks", PhD Thesis, Neural Computing Research Group, Aston University, Birmingham, Unitwed Kingdom

**Zhang** W., 1992, "A neural network based approach for surveillance and diagnosis of statistical parameters in the IC manufacturing process", Master's degree thesis, University of Maryland.