Mahalanobis Taguchi System (MTS) for Pattern Recognition, Prediction, and Optimization

Steven Holcomb
Process Improvement Project Manager
Reed Integration, Inc., Suffolk, VA
sholcomb@reedintegration.com

ABSTRACT

The Mahalanobis Taguchi System (MTS) uses multivariate data in a unique way, to enable recognition, diagnosis, and prediction of important conditions for a wide variety of applications in various industries. The modeling and simulation discipline could benefit from adding MTS pattern recognition to their tool set for analysis of simulation results and decision support. MTS is based on the combined works of Dr. P.C. Mahalanobis, Dr. Genichi Taguchi, Dr. Rajesh Jugulum, Dr. Soichi Teshima, and others. Although the MTS approach is not well known in the U.S., it has provided benefits worth considering in many diverse applications over the past couple of decades, including manufacturing, healthcare, and vehicle control systems. It has proven to be superior in several ways to other recognition and prediction approaches, such as regression analysis and neural networks. Unlike more complex methods like artificial neural networks, MTS can be successfully demonstrated and experienced by analysts, programmers, and engineers with simple analysis tools like Excel. This technique may be useful to modeling and simulation researchers, and professionals interested in a straightforward recognition and prediction approach, which resolves multi-dimensional problems into a simple, single measurement scale.

The paper will introduce 1) A concise ***explanation of the concepts*** behind the MTS and its various uses, 2) An introduction to the ***computational methods*** for MTS, including a simple live demonstration, 3) ***Optimization*** of the MTS approach to ensure reliable predictions with the fewest possible parameters, 4) Summary of several ***MTS case studies*** from various industries, and 5) Comparison of ***advantages and disadvantages*** between MTS and other recognition and prediction methods.

ABOUT THE AUTHORS

Steve Holcomb is a Process Improvement Project Manager with Reed Integration, Inc., in Suffolk, VA He is a PMP, a Jonah, a former Baldrige examiner, a senior engineer, and a performance excellence professional. Steve has over two decades of experience helping engineering, R&D, heavy manufacturing, military, and service organizations meet their most challenging performance goals. His qualifications include expert level training and experience in project / portfolio management, advanced process improvement, measurement, and data analysis for decision support, supply chain performance, organizational assessment, critical thinking, and strategic plan development and deployment.

His leadership and experience have consistently increased process throughput, shortened project cycle times, decreased work-in-process carrying costs, reduced process variation , and operating costs, all while honoring existing resource, policy , and regulatory constraints that typically affect organizations and systems. These benefits have translated into hundreds of millions of dollars of verified benefit to organizations' bottom line, velocity, and capacity.

In the past, Steve has served as a performance excellence professional at the Old Dominion University Business Gateway (small business consulting), Huntington Ingalls Industries (naval ship design, R&D, construction , and repair), Case New Holland (con/ag equipment design and manufacture), the American Supplier Institute (six-sigma consulting and training) , and the U.S. Coast Guard. He holds a BS degree in Ocean Engineering from the U.S. Coast Guard Academy and is a certified PMP.

. . . ***But***, if you ask Steve about himself he will tell you that he only knows five things, which can be useful to help organizations and individuals.

1. How to identify and understand their “wicked problems”
2. Get control and make better decisions
3. Execute better
4. Free up capacity and “Do More Good”
5. Accomplish challenging goals and enjoy success

Mahalanobis Taguchi System (MTS) for Pattern Recognition, Prediction, and Optimization

Steven Holcomb
Process Improvement Project Manager
Reed Integration, Inc., Suffolk, VA
sholcomb@reedintegration.com

EXPLANATION OF THE CONCEPTS BEHIND THE MTS AND ITS VARIOUS USES

Living creatures are marvels of "pattern recognition" capability.

Most are able to efficiently process hundreds of billions of pieces of sensory information per second. Then, in just milliseconds, they recognize and act upon familiar patterns. This is apparently done without laborious processing of vast, detailed data models. Figure 1 shows a few examples, including the following:

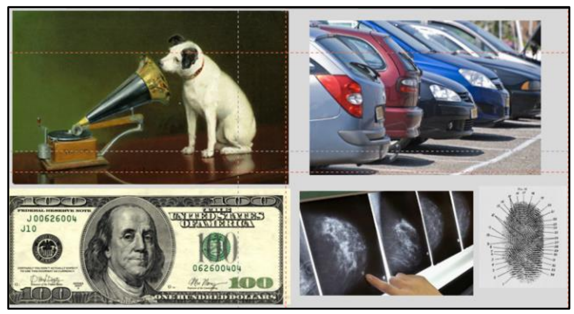

Figure 1. Some Examples of Pattern Recognition

- Finding your car in a crowded parking lot by seeing just "a sliver" of a fender or roof line
- An FBI agent easily detecting a counterfeit $100 bill because of deep familiarity with "the real thing"
- The famous painting called "His Master's Voice", by artist Francis Barraud, of the terrier Nipper.

The Mahalanobis Taguchi System (MTS) is a straight forward method for quickly recognizing, diagnosing, and predicting conditions in a wide variety of applications whose systems are defined by multivariate data. It uses a vector analysis approach that consolidates data from many parameters of a system into a single "score" or dimension. This dimensional analysis enables several useful capabilities, such as these listed below:

- Defining a multi-dimensional space or "bubble", that contains all possible matches to a particular pattern (the Mahalanobis space or reference space)
- Determining if each new individual encountered is inside or outside this Mahalanobis (reference) space
- Evaluating the degree of pattern match or mismatch by calculating the "distance" of each new individual encountered from the center of the reference space (the Mahalanobis Distance)
- Comparing individuals using their calculated Mahalanobis distances
- Monitoring trends of an individual's Mahalanobis distances over time and predicting future states
- Optimizing each recognition and prediction system to minimize data required for accurate decision support

To illustrate these capabilities of the MTS approach, a simple example will be used, followed by some brief case studies of more complex, real-world examples. In the simple example, four physical characteristics of a person's body, shown in Figure 2, are used to ***anonymously*** judge whether or not individuals fall within the reference (Mahalanobis) space considered to be "healthy" by physicians. When individuals fall outside the "healthy" space, this example will show how MTS measures the degree of "unhealthiness" of their conditions, and how this kind of information can be used for decision support.

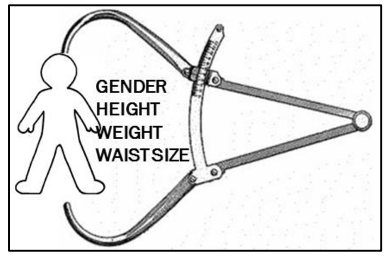

Figure 2. A "Simple" Multivariate Recognition Problem

Figure 3 shows an illustration using just two of these parameters; ***Height*** and ***Weight***. Healthcare organizations and practitioners generally agree on the "reference space" that defines a "healthy" condition in this simple two dimensional example. This "healthy" space is depicted by the red ellipse in Figure 3. If an individual's height and weight are plotted on the chart, a vector distance from the center of the "healthy" space can be calculated, as shown by the blue arrow in the figure. This blue arrow represents the degree of mismatch or "unhealthiness" of an individual when compared to the "healthy" reference space. For a growing child, calculation of the length of that vector over time could provide a single "score", useful for comparing individuals and monitoring trends as that child's height and weight increase. The length of this blue arrow is analogous to the Mahalanobis Distance that will be calculated as a multidimensional vector for recognition and prediction using MTS. In our more complicated four dimensional example, we will add ***gender*** and ***waist size*** as system parameters to make the problem more challenging. In actual multivariate pattern recognition problems, the Mahalanobis space and individual Mahalanobis distances might be calculated using hundreds of parameters. In this simple example, we will define the four dimensional "healthy" or "normal" reference space, and calculate individual distance vectors from that reference space to evaluate health and predict trends.

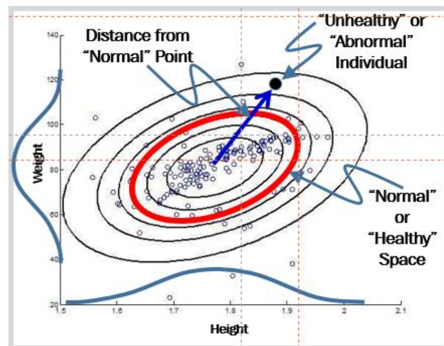

Figure 3. Simple recognition and prediction example; height & weight

AN INTRODUCTION TO THE COMPUTATIONAL METHODS FOR MTS, INCLUDING A SIMPLE LIVE DEMONSTRATION

Using the simple four parameter example described above, the computational process will be illustrated in four steps:

1. Generate the "normal" or "healthy" reference space
2. Confirm discrimination power
3. Identify critical parameters and optimize
4. Implement the optimized measurement system

In this simple example, computation methods will be shown for calculating the reference Mahalanobis space and the Mahalanobis Distances for each individual.

Step 1: Generation of Normal ("healthy") Space

1.1 – A normal ("heathy" or “reference”) population is selected for use in defining the reference space (Mahalanobis Space) for recognition and prediction.

1.2 – Calculation of the Mahalanobis Distances (MD) for each sample in the normal group is accomplished through the five sub-steps, a – e, shown here:

a - Computing the mean & standard deviation for each parameter using all individual parameter values (Table 1, leftmost table).

b - Normalizing individual parameter values (height, weight, etc.) for each member of the reference group using the previously calculated means and standard deviations (Table 1, rightmost table).

c - Computing the correlation matrix of the normalized “healthy” data set (left half of Table 2).

d - Compute the inverse matrix of the correlation matrix for the “healthy” group (right half of Table 2)

e – And finally calculating the normalized Mahalanobis Distance ($\mathbf{D^2}$) for each “healthy” individual in the reference group using equation 1.

Table 1. Normalize the Reference Group Data Using Mean and Standard Deviation

"HEALTHY" GROUP SAMPLES						"HEALTHY" GROUP SAMPLES (x variables) NORMALIZED BY MEAN & STDEV (into z variable, where z = (x-xbar)/stdev x)				
mean	6.4	68.1	155.1	30.3			0.0	0.0	0.0	0.0
stdev	4.5	4.6	23.7	4.3			1.0	1.0	1.0	1.0
n	Gender (1=F, 10=M) x1	Ht (in) x2	Wt (lb) x3	Waist (in) x4	k = 4	n	Gender (1=F, 10=M) z1	Ht (in) z2	Wt (lb) z3	Waist (in) z4
	1	2	3	4			1	2	3	4
1	10	73	178	32		1	0.796	1.060	0.979	0.382
2	1	58	127	28		2	-1.194	-2.185	-1.197	-0.571
3	10	71	162	28		3	0.796	0.628	0.282	-0.633
4	10	67	160	33		4	0.796	-0.238	0.214	0.622
5	1	69	137	32		5	-1.194	0.195	-0.753	0.340
6	1	64	129	21		6	-1.194	-0.887	-1.093	-2.135
7	10	70	179	27		7	0.796	0.411	0.992	-0.697
8	10	75	201	30		8	0.796	1.493	1.927	-0.023
9	10	71	168	36		9	0.796	0.628	0.525	1.236
10	1	70	151	30		10	-1.194	0.411	-0.177	0.021
11	10	65	140	34		11	0.796	-0.671	-0.620	0.756
12	1	68	148	26		12	-1.194	-0.022	-0.313	-0.959
13	1	65	136	26		13	-1.194	-0.671	-0.799	-0.964
14	10	76	191	39		14	0.796	1.709	1.517	2.092
15	10	69	174	32		15	0.796	0.195	0.789	0.391
16	10	66	150	34		16	0.796	-0.454	-0.219	0.868
17	10	72	179	36		17	0.796	0.844	0.993	1.348
18	1	59	110	27		18	-1.194	-1.969	-1.925	-0.722
19	1	66	133	26		19	-1.194	-0.454	-0.947	-0.890
20	10	68	151	28		20	0.796	-0.022	-0.174	-0.463

Table 2. Correlation Matrix & Inverse Matrix for the Normalized “Health” Group Dataset

CORRELATION MATRIX OF THE NORMALIZED "HEALTHY" GROUP MATRIX (EXCEL>DATA > DATA ANALYSIS > CORRELATION)				
r	**1**	**2**	**3**	**4**
1	1.00	0.58	0.75	0.62
2	0.58	1.00	0.90	0.52
3	0.75	0.90	1.00	0.56
4	0.62	0.52	0.56	1.00

INVERSE MATRIX OF THE "HEALTHY" GROUP CORRELATION MATRIX (EXCEL > DATA > fx > MINVERSE > F2 > cntl + shift + enter)				
a	**1**	**2**	**3**	**4**
1	3.07	1.53	-3.17	-0.93
2	1.53	5.84	-6.02	-0.65
3	-3.17	-6.02	8.61	0.32
4	-0.93	-0.65	0.32	1.73

$$D^2 = \sum (a_{ij} z_i z_j) / k \qquad (1)$$

At the completion of step one, a reference space of “healthy” or normal individuals has been constructed. This four dimensional reference space is shown graphically in the bar chart of Figure 4 as the dashed horizontal line with an average D^2 of 1.0. This is the result of “normalizing” the reference group, as shown in Table 1 above. At the conclusion of step 1, we have created a new measurement system for recognition and prediction for this multivariate system, with ***two important components***, listed here:

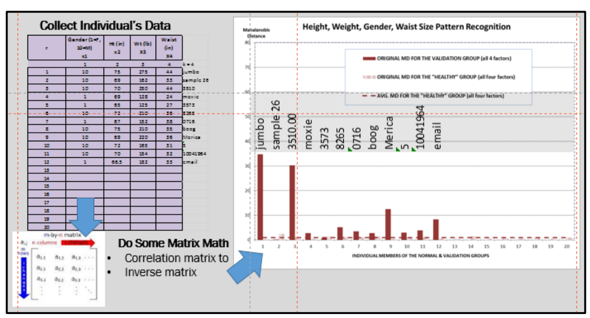

Figure 4. Results of Computing a Reference Space & Individual “Distance” Values for Comparison

1. A ***reference point*** from which to begin measurement and evaluation of the “health” or “match” of each new individual we encountered.

2. A ***measurement scale*** that we can now use to determine vector distance for each individual from the reference point. This is a multidimensional vector calculated from the four separate parameters of our example. This provides a single numerical “score” for each individual, for easy evaluation, comparison, and prediction.

Although the math involving matrix manipulation may seem intimidating, most simulation and analysis software can handle this smoothly. These examples were created easily in Excel using standard functionality.

Step 2: Confirm Discriminating Power

In the bar chart of Figure 4, additional D^2 values have been calculated and charted for 12 individuals, shown as vertical bars with "nicknames" above them for anonymity (i.e. jumbo, sample 26, 3500.00, moxie, etc.). These "distances" from the "healthy" reference group were calculated using the height, weight, waist size, and gender information provided by the individuals for comparison. The calculation methods for each individual are similar to those shown in Step 1 for the reference group.

Discriminating power of our new measurement system is confirmed by evaluating the calculated individual distance values (D^2) to see if they provide accurate comparisons of individuals to the reference group and to each other.

OPTIMIZATION OF THE MTS APPROACH TO ENSURE RELIABLE PREDICTIONS WITH THE FEWEST POSSIBLE PARAMETERS

Continuing with the third step of the MTS recognition and prediction computational process;

Step 3: Identify Critical Parameters and Optimize

Often, complex systems (e.g. vehicle control & collision avoidance systems) offer huge numbers of parameters that might be considered for pattern recognition and decision support. To consider all available parameters could require high processing capacity, lengthy run times during computer analysis, and vast data storage requirements. Instead, it is useful to know and only use those parameters which actually promote effectiveness of our new recognition and prediction system. There are three possibilities for each parameter in our multivariate system, based on their impact on sensitivity, accuracy, and usefulness of the D^2 measure. These three possibilities are as follows:

- Parameters that ***should be included*** because they ***promote*** sensitivity and accuracy
- Parameters that ***could be excluded*** because they ***neither promote nor degrade*** sensitivity and accuracy.
- Parameters that ***must be excluded*** because they actually ***degrade*** sensitivity and accuracy.

The analysis for identifying critical parameters, and optimizing the recognition and prediction system uses Dr. Taguchi's Quality Engineering approach. Figure 5 shows the outcomes of that analysis for this simple system. The line chart in the left side of Figure 5 shows individual impact of each of the four factors on sensitivity of the recognition and prediction system. The red circled parameter, "height", actually degrades sensitivity when included with the other parameters of weight, gender and waist size, and must be excluded for better sensitivity and accuracy. The green circled parameters, "gender" "weight" and "waist size" promote sensitivity and accuracy, and should be included in the MTS analysis.

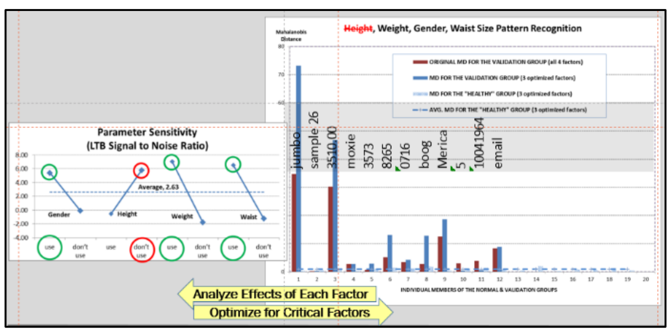

Figure 5. Results of the Process of Identifying Critical Parameters to Optimize Recognition & Prediction

A visual comparison of the bar chart at the right side of Figure 5 shows the individual calculated D^2 values under two different conditions. The brown vertical bars include all four parameters (height, weight, gender, waist size) in the calculation of D^2, while the blue vertical bars include only three (weight, gender, waist size), with "height" excluded. A visual comparison of the bar chart shows a marked improvement in the sensitivity of this multivariate measurement system when "height" is excluded from the analysis. The benefits of this type of analysis include 1) improved decision support and 2) savings associated with reducing the quantity of data to analyze and store for decision making.

Step 4: Implement the optimized measurement system

Typically, a project plan is developed and executed to accomplish implementation of the new, optimized measurement system using MTS. This implementation plan usually addresses issues related to people, processes, policies and technology.

SUMMARY OF SEVERAL MTS CASE STUDIES FROM VARIOUS INDUSTRIES

Examples of MTS applications from various industries include the following, and provide a glimpse into the full MTS solution, as indicated in Figure 6.

- Complex waveform analysis for recognition and prediction, applicable to many problems.
- Medical Diagnosis – liver disease detection, blood factor indications
- Manufacturing – automated quality inspection, process monitoring, yield prediction
- Measurement Systems – infrared absorption (chemical)
- Fire Detection – public buildings
- Identity verification – biometrics, handwriting, counterfeit currency
- Fault Analysis – space guidance system
- Candidate Selection – college application, high potentials, loans, job skills
- Earthquake Prediction (Japan)
- Automotive Airbag Deployment
- Automotive Collision Avoidance

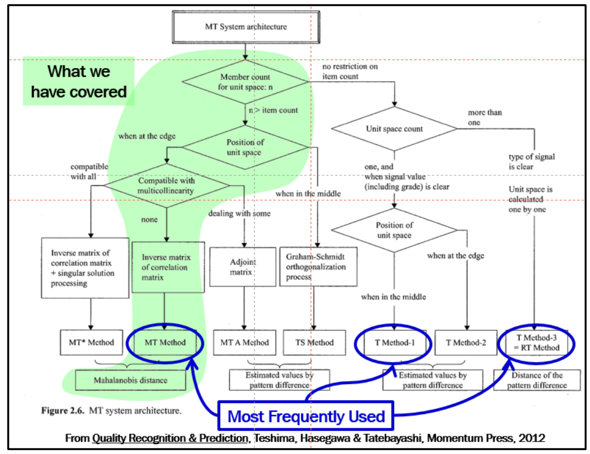

Figure 6. The Full MTS Solution for Recognition & Prediction in Multivariate Systems

COMPARISON OF ADVANTAGES AND DISADVANTAGES BETWEEN MTS AND OTHER RECOGNITION AND PREDICTION METHODS

A high level comparison of various recognition and prediction methods, with their advantages and disadvantages, is shown in Table 3. MTS offers some distinct advantages over other, more traditional recognition and prediction approaches. Most other approaches are founded in statistical analysis, rather than dimensional analysis used by MTS. This can make other techniques complex and subject to issues of sample size for learning new pattern. One of the greatest advantages of MTS is the relative simplicity to demonstrate the approach with little more than an Excel spreadsheet. Frequently, specialized skills and software are not required to try out the MTS method.

Table 3. Simple Comparison of Various Recognition and Prediction Methods to MTS

Requirement	Principal Component Analysis (PCA)	Discrimination & Classification Method	Regression Analysis	Stepwise Regression	Multivariate charts	Artificial Neural Network (ANN)	Mahalanobis-Taguchi System (MTS)
Dimensionality Reduction of datasets required.	X			X		X (problems learning new patterns, forgetting old patterns; highly complex)	X
Finds "normals & abnormals"		X	X				X
Measures degrees of abnormality			X (highly complex)				X
Monitor & control process conditions					X (abnormality based on probabilistic limits)		X (abnormality based on quality loss function)
Data Analytic approach (doesn't require assumptions about data distribution)						X	X
Can find the relationship between inputs & outputs							X
Widely Known & Practiced in U.S.	yes	yes	yes	yes	yes	yes	no

ACKNOWLEDGEMENTS

Steve Holcomb gratefully acknowledges the training and guidance in application of the Mahalanobis Taguchi System, provided by Dr. Genechi Taguch, Shin Taguchi, Yuin Wu, and Alan Wu of American Supplier Institute in the U.S., and Dr. Soichi Teshima, Executive Dir. of Angle Try Associates in Japan. He also acknowledges the support of Becky Reed, CEO and Steve Waddell, Vice President of Reed Integration, Inc. in providing the time and resources to offer this presentation to MODSIM WORLD 2016.

REFERENCES

Teshima, S., Hasegawa, Y., & Tategayashi, K. (2012). *Quality Recognition and Prediction; Smarter Pattern Technology with the Mahalanobis-Taguchi System*. New York, NY: Momentum Press.

Taguchi, G., Chowdhury, S., & Wu, Y. (2004). *Taguchi's Quality Engineering Handbook*. Hoboken, NJ: Wiley-Interscience.

Taguchi, G., & Jugulum, R. (2002). *The Mahalanobis-Taguchi Strategy: A Pattern Technology System*. New York, NY: John Wiley & Sons.

Taguchi, G., Chowdhury, S., & Wu, Y. (2000). *The Mahalanobis-Taguchi System*. New York, NY: McGraw-Hill Professional.