Detection of Anomalies in Manufacturing with Machine Learning

Introduction

High-volume manufacturers produce and ship large quantities of products to customers, and very high quality is expected in a competitive marketplace where customer satisfaction is an important metric. One key aspect of quality is uniformity in product characteristics over time. Uniformity can be especially important when the product is used as a component in a larger system; the compounding effects of small deviations in many components can have a large effect on ultimate system performance.

With this in mind, it can be important to put in place advanced analytical solutions to automatically detect product anomalies and prevent these products from being shipped out, even though their variation(s) may still be within specifications. The last chance to do this is at the final product test point before shipping.

This paper describes a multivariate machine learning solution that has been developed for semiconductor customers using the TIBCO Connected Intelligence Platform. The solution alerts when there are multiple clusters of product with distinctly different test characteristics. The approach outlined has applicability to detect anomalous equipment, processes, and products beyond this specific example.

Problem

Semiconductor manufacturing customers often place orders for many millions of integrated circuits (ICs) used in cars, mobile phones, satellites, and many other devices. These days, if it's electronic, it contains an IC, and it's very important that each one performs as expected so as not to compromise the systems they go into.

The last steps in the IC manufacturing process often involve two operations: assembly and final test. These high volume, high speed operations typically take place in two separate locations. Sometimes during assembly, similar types of ICs are accidentally mixed together. Because it is unknown when and how often this occurs, it is essential that mixed ICs are identified and separated into their correct groups during final test. However, individual test results of different products may overlap and readings fall within the allowed limits. Thus, it is impossible to determine if more than one type of IC is being tested so that mixed products that have polluted the production order go undetected and end up being shipped to the customer. This can cause failure of the electronics systems the ICs go into, resulting in scrap, rework, recalls, additional testing or failure analysis, and overall lowering of customer satisfaction, which can put future IC orders at risk.

Solution

TIBCO has set out to solve this problem by applying Data Science and Machine Learning (ML), including a solution that automatically identifies, detects, and alerts when products are being mixed.

The problem with a typical univariate approach is that looking at results of individual tests does not always work. They exhibit some differences for mixed products but there is not enough information to discriminate one product from another.

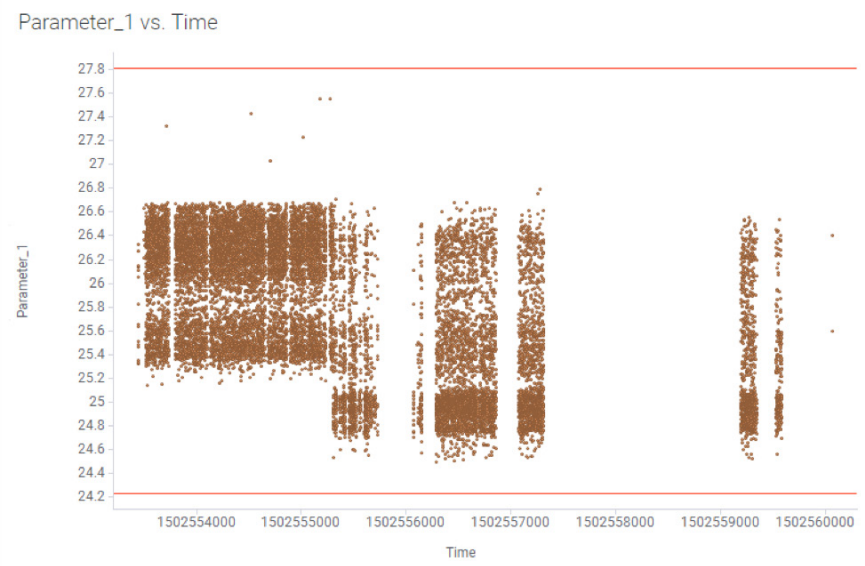

Figure 1. One test parameter that contains test results for two products, both within limits with overlapping distributions.

Figure 1 shows test results of one parameter for two products that have been mixed. The data shows some difference between the products, but both fall within allowed limits, so no alert is generated. In addition, these same results are within normal manufacturing variability; for example, the same products produced in different lots may produce these same results, which makes identification of mixed products even harder.

The question we've asked ourselves: Would we be able to identify mixed products if we take a multivariate approach, combining multiple test results in a different and more intelligent approach?

The idea was to intelligently combine various test parameters. Doing so would compound the slight differences that go unnoticed when looking at individual test parameters. It would also allow us to more easily identify multiple product signatures in the data.

The data science process:

Step 1: The dimensionality of the data is reduced by constructing several principal components explaining most of the variability of the data. This is important because the test results can contain hundreds or thousands of test parameters. Transformation of the data into a space with reduced dimensions also helps with visual inspection of suspicious behavior.

Step 2: Apply clustering on constructed components to identify groups of products with different behavior. In particular, the approach uses hierarchical clustering. The reason for doing this is that we have the whole hierarchy of clusters together with the information about linkage distances. This is quantification of how close or far away the clusters are that are being merged in each step of hierarchical clustering. (Figure 2 shows an example of hierarchy of clusters.)

Step 3: If the increment in linkage distance for the last steps of hierarchical clustering is bigger than a specific threshold, it is possible to conclude with a specific degree of certainty that a mixing problem occurred for data groups with highly different behavior.

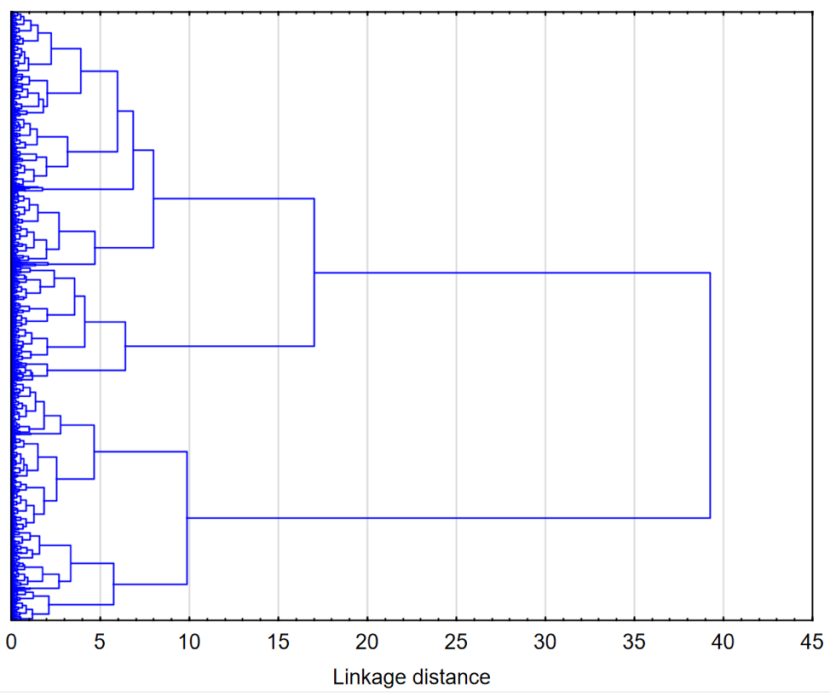

Figure 2. Hierarchical clustering representation of product mixing case. Last merge has large increase in linkage distance, which means that these two groups of products are highly different. In this case, the reason is that two separate product types have been mixed into one test batch.

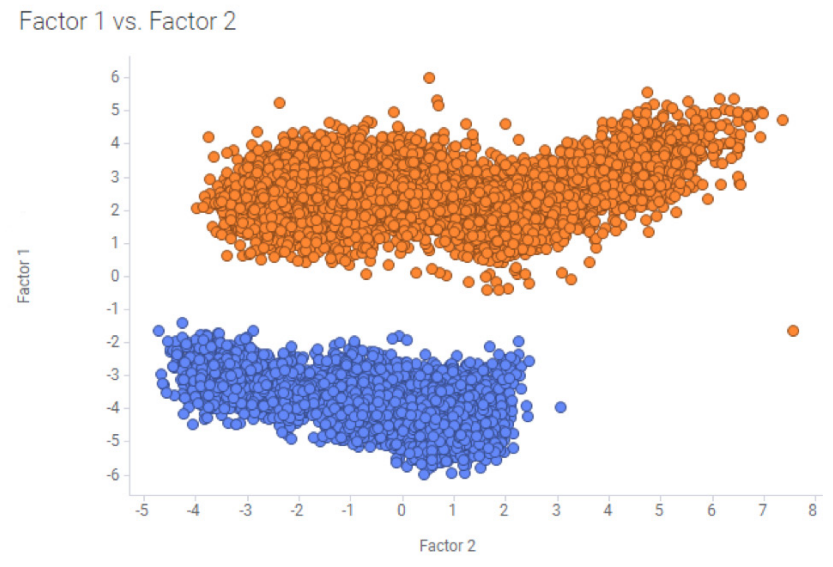

Figure 3. Test data after applying principal component analysis (PCA). Two separate groups/clusters are clearly visible and show the product mixing problem. Such graphs are typically displayed to engineers after an automated alarm is raised for exceeding a threshold of cluster linkage distance.

The solution proactively alerts test engineers when ICs have been mixed before being shipped to the end customer. The test engineers are empowered with a collaborative analytics dashboard that allows them to immediately visualize production issues, double check and evaluate the scale of the problem, and decide what action to take.

Figure 4. Interactive visual investigation of the suspicious lot (no product mixing detected in this example). Computed index and cluster sizes together with visual inspection in factor space of principal components is depicted. Cluster sizes and yield loss visualizations help to diagnose whether the problem is only with several anomalous units or if it is a systematic mixing of two distinct populations. The operator can drill down to particular parameters or switch to lot-to-lot comparisons for further diagnosis.

How Does TIBCO Help High Tech Manufacturers Solve the Problem?

Typically high tech manufacturers have a number of requirements for their data driven solution:

Data amounts are very large and increasing. How can data science be efficiently used on big data?

High tech manufacturing production lines generate gigabytes of data each day. Cost effectively analyzing this amount of data clearly requires a big data platform. The data science solution built on the TIBCO Connected Intelligence platform can leverage the scalability, security, and cost effectiveness of a big data platform.

The speed at which data is being created is very fast and accelerating. How can data science solutions be applied so they avoid building a backlog?

Because of the high speed of production and data generation, the solution must be fast enough to prevent a backlog and also have the ability to quickly scale to meet even greater speeds and volumes. The TIBCO Connected Intelligence platform joins, prepares, and analyzes data from multiple sources using a high performance, horizontally and vertically scalable in-memory data grid. Its performance achieves 20 times the speed of native Apache Spark. Data can be swapped in and out of this grid rapidly. When data is in the grid, visual and advanced analytics can be completed within milliseconds.

Test engineers are typically not data scientists. How can data science be operationalized so test engineers can benefit from its capabilities?

There is an increase in professionals who require data to make decisions. Many are data savvy, but do not necessarily have a data science degree. These professionals are often referred to as citizen data scientists. TIBCO Data Science software provides capabilities that allow designing rather than coding of data science pipelines. On a comprehensive and collaborative platform, citizen data scientists can collaborate, design, and deploy data science workflows. These workflows can be operationalized easily because the data engineering works behind the scenes, abstracted from the end user view. It is because of this simplicity that the solution can be used and maintained by citizen data scientists. Figure 5 shows the data science design tool interface.

There are additional challenges to address using data-driven decisions. Can the data science solution be extended to do more when the need arises?

Exploring ways to make better decisions using data requires multiple team players coming together. For that reason, a solution should be tailored to the skills of those engaged in such projects. The TIBCO Connected Intelligence platform brings people in different roles with different skills together. Each persona works on a particular part of the solution, and the solution works as a whole. It also supports open source capabilities like Python, R, Jupyter Notebooks, etc. Learn more about how the TIBCO Connected Intelligence platform helps manufacturers.

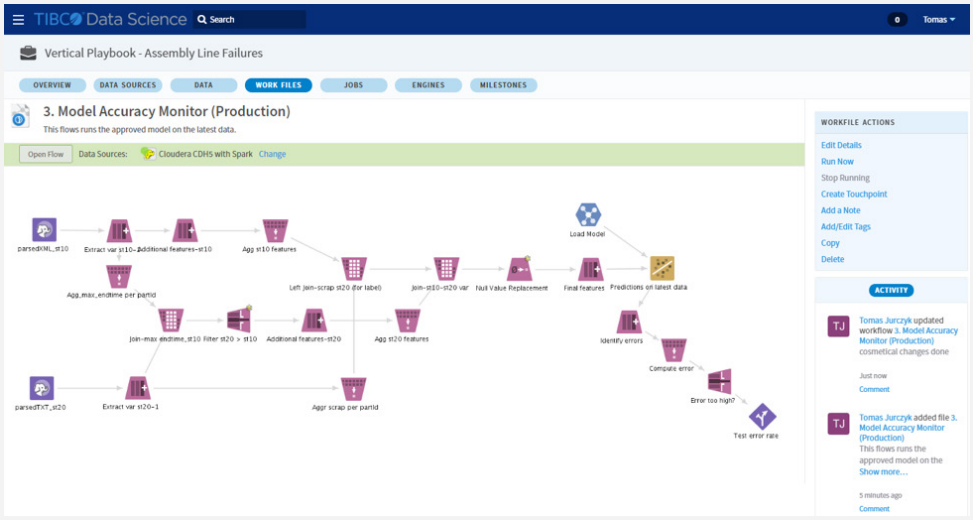

Figure 5. Collaborative data science design interface.

Summary

High tech manufacturers produce products at very high speeds and in large volumes. This generates lots of real-time data. Detecting quality issues like product mixing problems requires a cost effective, scalable, easy-to-use platform. It is because of these requirements that the product mixing solution has been built using the TIBCO Connected Intelligence platform. With it, you can perform high-performance data science cost effectively on big data, and seamlessly integrate with an easy-to-use BI platform. The visual data science workflows are easy to maintain by (citizen) data scientists on a collaborative platform. The general approach outlined has broad applicability for detecting anomalous equipment, processes, and product.

To learn more about TIBCO's anomaly detection solutions, visit https://www.tibco.com/solutions/anomaly-detection

Global Headquarters
3307 Hillview Avenue
Palo Alto, CA 94304
+1 650-846-1000 TEL
+1 800-420-8450
+1 650-846-1005 FAX
www.tibco.com

TIBCO Software Inc. unlocks the potential of real-time data for making faster, smarter decisions. Our Connected Intelligence platform seamlessly connects any application or data source; intelligently unifies data for greater access, trust, and control; and confidently predicts outcomes in real time and at scale. Learn how solutions to our customers' most critical business challenges are made possible by TIBCO at www.tibco.com.

©2020, TIBCO Software Inc. All rights reserved. TIBCO and the TIBCO logo are trademarks or registered trademarks of TIBCO Software Inc. or its subsidiaries in the United States and/or other countries. All other product and company names and marks in this document are the property of their respective owners and mentioned for identification purposes only.

20Feb2020