

Detección de outliers multivariados para el control de la variabilidad en un artículo ensamblado

Georgina Solís Rodríguez¹, Ricardo Reyes Torres²

Fernando Ortiz Flores³, Eugenia Rodríguez Contreras⁴, Eduardo Marroquín Prado⁵

¹Instituto Tecnológico de Saltillo, Saltillo, Coahuila, México, gsolis@its.mx

²Instituto Tecnológico de Saltillo, Saltillo, Coahuila, México, jrrt384@gmail.com

³Instituto Tecnológico de Orizaba, Orizaba, Veracruz, México, ferchilo@prodigy.net.mx

⁴Instituto Tecnológico de Saltillo, Saltillo, Coahuila, México, erc7304@gmail.com

⁵Instituto Tecnológico de Saltillo, Saltillo, Coahuila, México, emarro53@hotmail.com

ABSTRACT

In this study there is carried out an analysis for the detection of multivariate outliers. This is done by using different methodologies in the problem of identifying problematic variable control systems applied to the process of assembly of an item, where the main features are; the presence of a considerable number of input and output variables. It is common to see atypical or outliers within a set of data points. The outliers, regarded as errors or noise, are actually data that may contain important information. The detection of outliers is only possible when performing a multivariate analysis using the appropriate methods. The following methods for detecting outliers apply; detection of fraud credit cards, diversity of clinical studies, analysis of irregularities in voting, intrusions into networks, predictions of severe weather, geographical information systems and the analysis of performance of athletes and other fields of data mining.

Keywords: outliers, multivariate statistical analysis, traditional PCA, PCA robust, dimensional variability.

RESUMEN

En este trabajo se realiza un análisis para la detección de outliers multivariados mediante la aplicación de diferentes metodologías en el problema de identificación de variables problemáticas aplicado al sistema de control del proceso de ensamble de un artículo, donde las características principales son; la presencia de un considerable número de variables de entrada y salida. Es común observar dentro de un conjunto de datos, puntos atípicos u outliers. Los outliers se consideran como errores o ruido, son datos que pueden contener información importante. La detección de outliers solamente es posible cuando se realiza un análisis multivariado con métodos apropiados. Los métodos de detección de outliers se aplican en: detección de fraudes en tarjetas de crédito, diversidad de estudios clínicos, análisis de irregularidades en votaciones, intrusiones en la red, predicciones graves del clima, sistemas de información geográficos, análisis de rendimiento de

los deportistas y otros campos de la minería de datos.

Palabras Clave: outliers, análisis estadístico multivariado, ACP tradicional, ACP robusto, variabilidad dimensional.