

58b Overcoming False Positives (Type-I Errors) While Monitoring of Transient Operations Using Principal Component Analysis

Yew Seng Ng and Rajagopalan Srinivasan

Introduction

Principal Component Analysis (PCA) is a commonly used approach for process monitoring [1;2]. SPE (Q statistic) and Hotelling's T^2 statistics [3] are the commonly used metrics for detecting deviations. While they are adequate for steady-state operations, these statistics are prone to Type-I errors (false positives) when applied to transient operations, such as batch processes and startups, shutdowns, grade change operations etc. in continuous processes. This is because the transient operations violate the basic assumption the statistics are built upon, ie: the normal density distribution of the source data.

Proposed Methodology

In this work, a new process supervision technique called Adjoined Dynamic Principal Component Analysis (ADPCA) is proposed to overcome the Type-I errors associated with PCA-based monitoring methods. The proposed technique combines clustering methodology with dynamic-PCA. Multiple overlapping DPCA models are used for efficient multiphase-based monitoring. The operating data of normal process transitions is first collected from the plant historian and after preprocessing (filtering and autoscaling), clustered into groups through fuzzy c-means clustering. Clustering of process states based on historical data can be used to differentiate multiple modes of operations in these temporal signals for building different DPCA models for monitoring purposes. The DPCA models constructed overlap with the neighboring PCA models to guarantee complete coverage of operations space. The operations information from different stages/phases can be extracted and assessed based on the clusters obtained. One of the challenges in adjoined DPCA model development is the selection of the number of clusters, i.e., the number of PCA models to be used. Here we propose a novel model-validation approach that compares the DPCA models resulting from the clustering using a *similarity factor analysis*[4]. The similarity factor measures the similarity between PCA models based-on the angles between the PCs subspace. Our proposed algorithm for PCA model validation integrates the similarity factor analysis with evolutionary technique by allowing the number of PCA models to evolve within a pre-specified number of generations. The PCA models of the final generation will be always optimized in terms of *distinctness* and *resemblance*. These conflicting goals whereby distinctness requires the models to be dissimilar while resemblance requires the data within each PCA models to show high similarity, enables the selection of the optimal number of clusters.

During online monitoring, the best-fit DPCA model is selected at every instant using the lowest *combined discriminant similarity factor* [5], which evaluates the distance between the current online signals and the DPCA model bank. This best-fit model is used for monitoring. Detection of a process fault is based on the Hotelling's T^2 and SPE statistics generated from the best-fit DPCA model selected, DPM_{opt} . In addition to this, the sequences of DPCA models taken by the current process, together with the process dwell-time are also monitored to detect process anomaly.

Benefits and Case Study

The application of the proposed technique to the cultivation of a fed-batch penicillin process shows that the method gives better performance and robustness in comparison to multiway-PCA [6] and dynamic-PCA [7]. The proposed method is better in terms of sensitivity and accuracy as it detects process disturbances at earlier stages than other methods. The gain in sensitivity is not at the expense of false positives since the proposed method is less prone to Type-I errors as compared to multiway-PCA and

dynamic-PCA. The proposed adjoined DPCA monitoring technique has a sound theoretical basis for monitoring of transient multiphase processes as each local DPCA model satisfies the assumption of normal data distribution. Also, differentiating transitions/batch information into several small and manageable phases allows phase-specific control and monitoring rules to be incorporated easily.

References

- [1] Qin, S. J., (2003). Statistical process monitoring: basics and beyond, *Journal of Chemometrics* 17, 480-502.
- [2] Kourti, T., (2002). Process analysis and abnormal situation detection: From theory to practice, *IEEE Control System Magazine*, October Issue 2002.
- [3] Jackson, J.E., and Mudholkar, G., (1979). Control procedures for residuals associated with principal component analysis, *Technometrics* 21, 341-349.
- [4] Krzanowski, W.J., (1979). Between-groups comparison of principal components, *Journal of American Statistical Association* 74, 703-707.
- [5] Raich, A., and Çinar, A., (1997). Diagnosis of process disturbances by statistical distance and angle measures, *Computers & Chemical Engineering* 21, 661-673.
- [6] Nomikos, P., and MacGregor, J.F., (1995). Multivariate SPC charts for monitoring batch processes, *Technometrics Vol. 37(1)*, p41-59.
- [7] Srinivasan, R., Wang, C., Ho, W. K., Lim, K. W., (2004a). Dynamic PCA based methodology for clustering process states in agile chemical plants, *Industrial and Engineering Chemistry Research* 43, 2123 – 2139.