Optimization of Mean and Standard Deviation of Multiple Responses Using Patient Rule Induction Method

Jin-Kyung Yang, Hanyang University, Seoul, South Korea
Dong-Hee Lee, Hanyang University, Seoul, South Korea

ABSTRACT
In product and process optimization, it is common to have multiple responses to be optimized. This is called multi-response optimization (MRO). When optimizing multiple responses, it is important to consider variability as well as mean of the multiple responses. The authors call this problem as extended MRO (EMRO) where both of mean and variability of the multiple responses are optimized. In this article, they propose a data mining approach to EMRO. In these days, analyzing a large volume of operational data is getting attention due to the development of data processing techniques. Traditional MRO methods takes a model-based approach. However, this approach has limitations when dealing with a large volume of operational data. The authors propose a particular data mining method by modifying patient rule induction method for EMRO. The proposed method obtains an optimal setting of the input variables directly from the operational data where mean and standard deviation of multiple responses are optimized. The authors explain a detailed procedure of the proposed method with case examples.

KEYWORDS
Data Mining, Design of Experiments, Desirability Function, Multi-Response Optimization, Operational Data, Patient Rule Induction Method, Process Optimization, Response Surface Methodology

1. INTRODUCTION
Response surface methodology (RSM) consists of a group of techniques used in empirical study between a response and a number of input variables. The researcher attempts to find the optimal setting for the input variables that either maximizes or minimizes the response (Myers et al., 2009). RSM assumes that variance of the response is constant and focuses on optimizing the mean of the response. However, the constant variance assumption might not be valid in practice. In such cases, not only the mean response, but also the standard deviation of the response should be considered in determining the optimum conditions for the input variables.

Dual-response optimization (DRO) attempts to optimize both the mean and standard deviation of the response. The conventional approach to DRO requires building statistical models for mean and
standard deviation of the response, by fitting response surface models (usually second-order models) to experimental data. Then, an optimal setting for the input variables is obtained by analyzing the statistical models (Lee et al., 2010; Lee & Kim, 2012).

Usually, RSM focuses on optimization of a single response. However, it is very common to consider several response variables in product and process development. This is called a multi-response problem. Some typical problems include the tire tread compound problem (Derringer & Suich, 1980), the injection molding of a washing machine agitator problem (Reddy et al., 1997), the rapid prototyping system problem (Palmer et al., 2006), and semiconductor manufacturing process problems (Lee & Kim, 2012; Lee & Kim, 2013). Similar to DRO, multi-response optimization (MRO) usually builds statistical models to analyze the means of multiple responses by fitting response surface models (usually second-order models) to experimental data. Then, an optimal setting for the input variables is obtained by analyzing the statistical models (Lee & Kim, 2012).

There have been many methods for MRO. However, most of them consider only the mean of multiple responses, like the traditional RSM. As previously mentioned, the constant variance assumption might not be valid in practice. Thus, methods for optimizing both mean and standard deviation of multiple responses are needed. We call this research area extended MRO (EMRO). Figure 1 shows the relationships among RSM, DRO, MRO, and EMRO. In spite of the importance of EMRO, limited studies have been performed.

One possible approach to EMRO is to utilize the conventional modeling methods of RSM, DRO, and MRO. However, these approaches have two practical limitations. First, efforts for conducting experiments can be a burden for process engineers. In order to build the statistical models, experiments are conducted and experimental data are gathered. In particular, in the case of DRO, a large number of experiments are usually conducted, because replicate runs at each design point are needed to fit the statistical model for standard deviation of the response, whereas replicate runs are not mandatory in RSM.

Second, the optimal settings for the input variables obtained from the conventional approaches have uncertainty (Lee & Lee, 2016). This is because the optimal setting is obtained from the prediction models. Care must be taken regarding uncertainty when the prediction capability of the models is poor. As the prediction model varies from the true model, the resulting optimal setting might be quite far from optimal (Xu & Albin, 2003). Also, it is reported that the prediction capability of the models is often poor when the models are fitted to operation data from manufacturing lines, whose volume is typically large (Lee & Kim, 2008).

At present, it is common to have operational data rather than experimental data, and as a result, analyzing the operational data becomes more important. This phenomenon has become very common.

Figure 1. Relationships among RSM, DRO, MRO, and EMRO
Algebraic and Graphic Languages for OLAP Manipulations
[www.igi-global.com/article/algebraic-graphic-languages-olap-manipulations/1798?camid=4v1a](www.igi-global.com/article/algebraic-graphic-languages-olap-manipulations/1798?camid=4v1a)

An Envisioned Approach for Modeling and Supporting User-Centric Query Activities on Data Warehouses
[www.igi-global.com/article/envisioned-approach-modeling-supporting-user/78288?camid=4v1a](www.igi-global.com/article/envisioned-approach-modeling-supporting-user/78288?camid=4v1a)