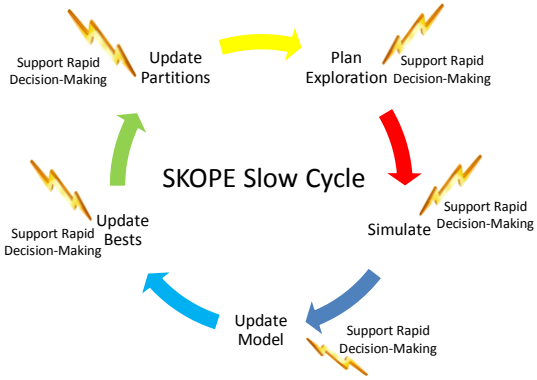
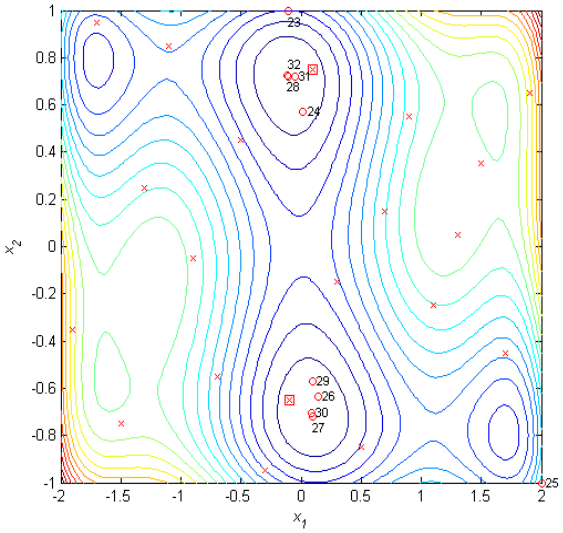
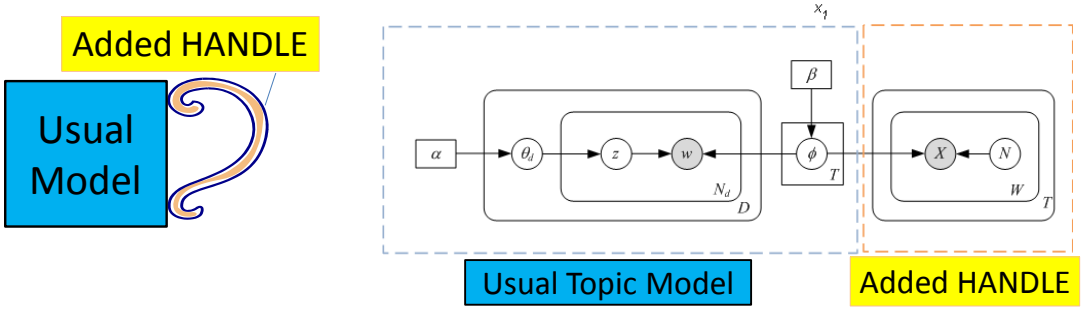
Bigish Data Analytics: Rapid Optimization and Fusion

White Paper
September 2014







Contact Information:

Theodore T. Allen, Ph.D.
The Ohio State University
1971 Neil Avenue – 210 Baker Systems
Columbus, Ohio 43221
(614) 292-1793 (office)

www.blying.com

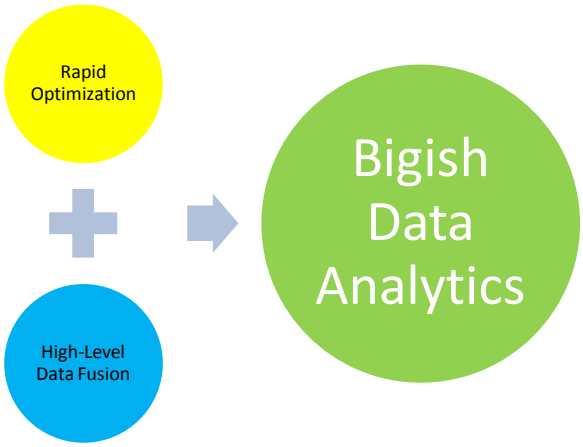


Figure 1. Elements of "bigish data" analytics.

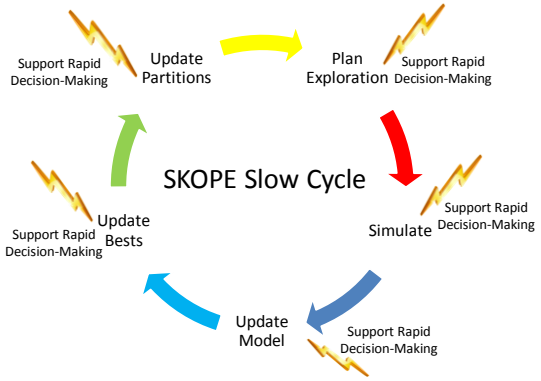


Figure 2. The SKOPE slow cycle involves potentially continual metamodel improvement while maintaining constant readiness.

1. Overview

MapReduce and other systems for processing truly large data sets involve mapping tasks into parallelizable portions and then reducing or assembling the results. Such methods can perform tabulations and even matrix operations with potentially trillions of documents. Such cases involve truly "big data". By "bigish data" we mean situations involving thousands or even millions of data points. By staying relatively small, we seek to develop and apply more complicated operations such as sequential Kriging optimization (SKO) and subject matter expert refined topic models (SMERT) to much larger problems than was previously possible (Figure 1). SKO and SMERT and other machine learning methods have impressive benefits in terms of providing rapid optimizations and impressive "sense-making" for the problems involving tens or hundreds of data points to which they have been previously applied

2. Rapid Simulation-Based Re-Optimization with "Metamodels"

Many organizations build discrete event or finite element simulations to aid in decision-making. Yet, such codes can take minutes or even days to evaluate a set of inputs. Sequential Kriging Optimization (SKO) facilitates optimization using these slow simulations by building a global "metamodel" to predict outputs rapidly for any combination of inputs. The metamodel is not as accurate as the simulation but it helps in deciding where to perform future slow simulation runs and for rapid re-optimization as decision problems arise. Figure 1 shows the on-going extension of SKO to address bigish data problems. The extended method is called sequential Kriging optimization via partitioned envelopes (SKOPE).

$$E[I(\mathbf{x})] \equiv \underbrace{E\left[\max\left(\hat{Y}(\mathbf{x}^{**}) - Y_p(\mathbf{x}), 0\right)\right]}_{\text{Expect "betterment" over current known "best"}} \cdot \underbrace{\left(1 - \frac{\sigma_\varepsilon}{\sqrt{s^2(\mathbf{x}) + \sigma_\varepsilon^2}}\right)}_{\text{A factor accounting for "diminishing" contribution of a replicate}}$$

$$Y_p(\mathbf{x}) \sim N\left[\hat{Y}(\mathbf{x}), s(\mathbf{x})^2\right]$$

Figure 3. Planning exploration balancing improvement and search objectives.

As indicated in Figure 2, the developing SKOPE methods perform exploration, launch slow simulations automatically, and update themselves in a slow cycle. They also feature the management of partitions which both facilitate parallelization and ensure that updating computations have bounded costs.

The SKOPE cycle in Figure 2 is "slow" because the simulations it involves remain relatively time-consuming. Yet, like SKO, SKOPE maintains readiness for rapid decision-support and even re-optimization with different goals because of the constant availability of global metamodels for all outputs.

Figure 3 shows the planning formulas for SKO and SKOPE which balance exploitation (to achieve improvement of a single objective) with exploration (which gains information about all outputs). Our past innovation in SKO related to an extension for noisy problems of previous methods. In the years since the publication of SKO, the performance of the formulas in Figure 3 has apparently stood the test of time as indicated in Figure 4.

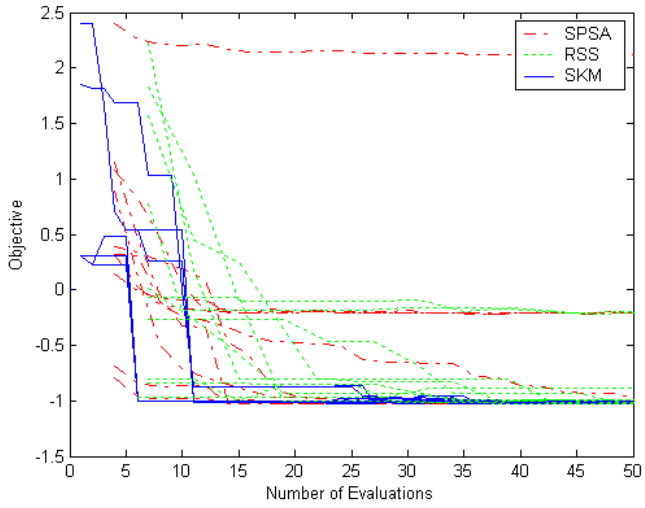


Figure 4. SMAC plot using real data for critical hosts under a single cost structure.

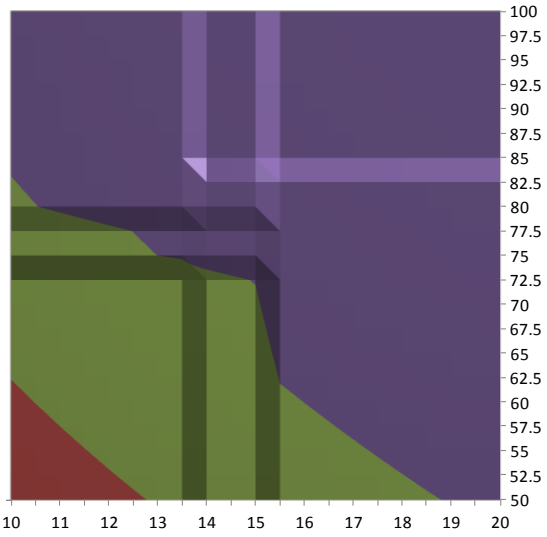


Figure 5. MCRAD control chart for monitoring.

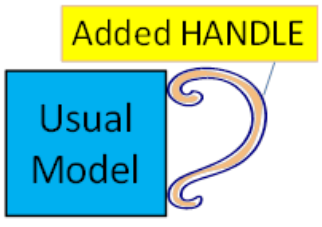


Figure 6. The concept of a model with a handle that makes it more directable by experts.

4. SKOPE Objectives

The structure of SKO is promising but as the number of runs increases past a few hundred, the computational overhead and the costs of simulations often are prohibitive. The on-going research on SKOPE seeks to measure and bound the overhead and permit bigish datasets and effectively continual or on-line optimization.

Figure 5. Shows a surface plot from a SKOPE implementation showing how the metamodels can be discontinuous across partition boundaries. Managing these boundaries and maintaining efficient convergence to global optimal solutions is a major challenge.

5. High-Level Data Fusion

Constraints imposed by expert-known physical laws are often not utilized in empirical modeling procedures. For example, SKOPE relates to modeling and optimization combined. Yet, the Kriging metamodel in SKOPE derives only from the data and there is limited ability for the user to improve the model short of inserting fabricated data, i.e., simulation outputs.

Another class of models involves “high-level” data from subject matter experts (SMEs). These models relate to a view that (some) people can be trusted (and honest) to influence models in desirable ways. This data is fused with ordinary “low-level” data. There have been many related ideas in the literature but much about the subject of putting a “handle” on a model and allowing experts to manipulate it is new. This is the key idea of subject matter expert refined topic (SMERT) models which were invented by SEAL researchers. Topic model outputs sometimes contain pairing of words within topic definitions that make no physical sense.

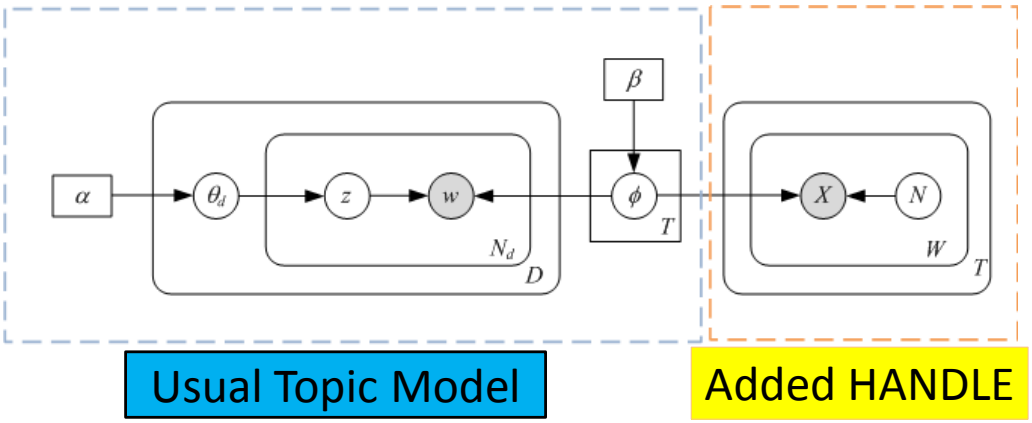


Figure 7. SMERT model showing ordinary Latent Dirichlet Allocation and the handle.

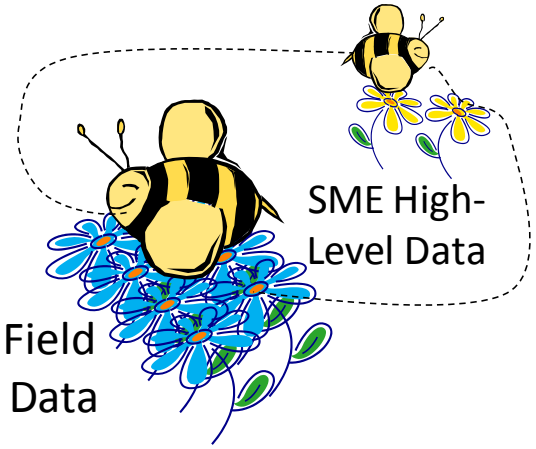


Figure 8. MCMC Bee cross pollinating.

$$P(w,z) = \int_\theta P(\theta|\alpha)P(z|\theta)d\theta \times \int_\phi P(\phi|\beta)P(w|z,\phi)P(x|N,\phi)d\phi$$
$$\propto \int \prod_{d=1}^{D} \frac{1}{B(\alpha)} \prod_{t=1}^{T} \theta_{d,t}^{\alpha_t - 1} \prod_{n=1}^{N_d} \theta_d^{z_{d,n}} d\theta$$
$$\times \int \prod_{t=1}^{T} \frac{1}{B(\beta)} \prod_{c=1}^{WC} \phi_{t,c}^{\beta_c - 1} \prod_{d=1}^{D} \prod_{n=1}^{N_d} \phi_{z_{d,n}}^{w_{d,n}}$$
$$\times \prod_{t=1}^{T} \prod_{c=1}^{WC} \binom{N_{t,c}}{x_{t,c}} \phi_{t,c}^{x_{t,c}} (1-\phi_{t,c})^{N_{t,c}-x_{t,c}} d\phi$$

$$P\left(z_{a,b} \middle| z_{-(a,b)}, w, \alpha, \beta\right)$$
$$\propto \left(C_{z_{a,b},a,*}^{-(a,b)} + \alpha_{z_{a,b}}\right) \times \ \ldots$$

Draw topic conditioned on Assumptions for all other quantities

Generalizes LDA collapsed Gibbs Griffiths and Stuyvers (2004)

Our formula is approximate in some cases.

Figure 9. SMERT model formulation.

The SMERT model form is pictured in Figure 7. The handle permits the SME to critique the high-level topic definitions (f) with data from experiments (which can be though experiments). Yet, SMERT models like topic models are currently too slow for big or even bigish data sets.

It is perhaps true that the majority of modeling approaches relevant to big data involve applying ad hoc, artificial rules or minimizing deviations. These approaches result in fitted models that have no standing in statistics, i.e., they are neither Frequentist nor Bayesian. Their errors are generally not understood and diagnostic information is severely limited.

Contrast this with SMERT models which are fully Bayesian. The high-level data enters like a prior for subsequent evaluations. If Gibbs sampling is used, there is an analogy to a Markov Chain Monte Carlo natural bee buzzing from the high to the low level data cross-pollinating (Figure 8). In ordinary SMERT models, the actions of the so-called bee relate to the probabilities indicated in Figure 9. Extending this framework to large datasets for improved sense making is an important challenge.