## 基于贝叶斯优化的代理建模优化软件平台开发

1. **研究背景及意义**
   1. **代理建模优化**

优化问题

普遍存在于科学研究和工程领域，如：设计控制器，辨识模型，选取神经网络训练的超参数，更为实际的例子，如：配置航天飞行器零部件，药物研究，短视频软件的推送内容等。

有些优化问题的目标函数复杂，没有可利用的结构性质，无法获得导数信息，甚至不存在明确的数学表达[1]，往往需要实验或仿真的方法搜索最优解。

但实验搜索法在求解时常面临以下的矛盾：

1. 一方面：变量寻优的空间常是连续且巨大的[2]，如在诊断阀门粘滞时，只考虑简单的由正向静摩擦、负向静摩擦、动摩擦、阀门初始位置四参数构成的模型，假设每个参数有5种取值，就会产生组合情况，而这只是连续空间中的少数采样值。
2. 另一方面：进行一次实验有时具有高昂的代价，如一次神经网络训练耗时耗能大，一次药物的临床实验可能造成病人死亡[2]。无法进行很多次实验。

基于上述的问题，如何科学设计实验，高效的求解复杂优化问题？解决方案之一：是建立便于计算和求解的代理模型（surrogate model），以近似原目标函数[3]，并通过求解代理模型的优化问题指导实验进行，最终得到近似全局最优解。

其中，由输入输出数据驱动建立响应面代理模型，并进行自适应采样的优化算法框架具有很高的扩展性，广为应用[3]，其算法描述如下：

首先，有一个难以分析和求解的目标函数

1. 通过实验法，采样得到了一系列历史数据，令。
2. 根据数据构建代理模型（神经网络、支持向量机、随机森林等），拟合、近似原目标函数。
3. 设计一个评估代理模型的采样函数，，求解采样函数的优化问题 ，得到一个新的采样建议，使得实验搜索不至于盲目进行，可取。
4. 在这个采样建议指导下做下一次实验，通过实验/仿真计算，并将添加进入历史数据，若仍未达到结束条件，令，回到

算法以流程图表示如下：

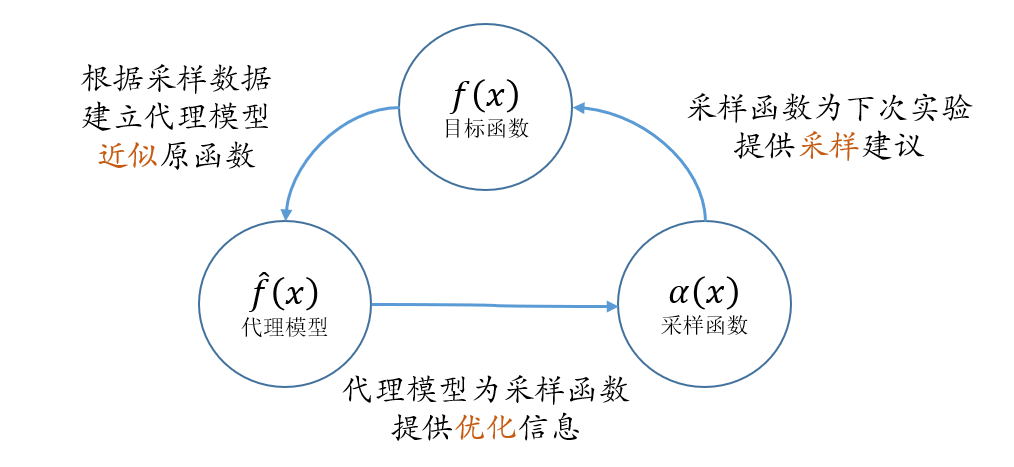


图1 典型代理建模优化算法

* 1. **典型代理建模优化的问题**

然而，在使用上述优化方法仍有以下几个疑问：

1. 在每一步建模代理模型时，如何衡量由局部采样带来的建模不确定性？
2. 观测值具有不确定性（噪声），如何在建模时考虑噪声的影响？[2]
3. 在采样下一个数据点时，如何对待未探索区域的不确定性？是否会出现只在确定性较大的区域内搜索，而忽略未探索区域的情况？
   1. **基于贝叶斯优化的代理建模优化**

基于贝叶斯优化的代理建模优化算法，给出了解决方案：通过贝叶斯统计的知识，考虑目标函数采样噪声和未采样区域的不确定性，建立了一个概率模型作为代理模型[1]，如高斯过程的模型。

通过贝叶斯统计的分析方法， 在指导采样时考虑未采样区域的不确定性，设计了一个考虑统计性能（均值与不确定性）的采集函数[1]。

将算法更新如下：

1. 通过实验法，采样得到一系列历史数据，令。
2. 根据数据构建**概率**代理模型，拟合、近似原目标函数。
3. 设计一个评估代理模型**统计性能**的采集函数，，求解采集函数的优化问题 ，得到一个新的采样建议，使得实验搜索不至于盲目进行，可取。
4. 在这个采样建议指导下做下一次实验，通过实验/仿真计算，并将添加进入历史数据，若仍未达到结束条件，令，回到

算法以流程图表示如下：

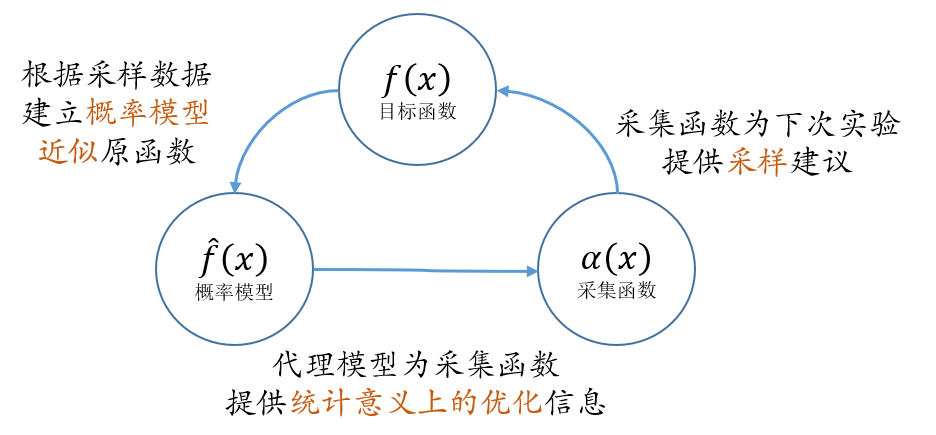


图2 基于贝叶斯优化的代理建模优化算法

具体实例如下图，可以看到，在概率模型和采集函数的指导下，实验进行了较为科学的采样，概率模型拟合目标函数最优值的精度不断提升。

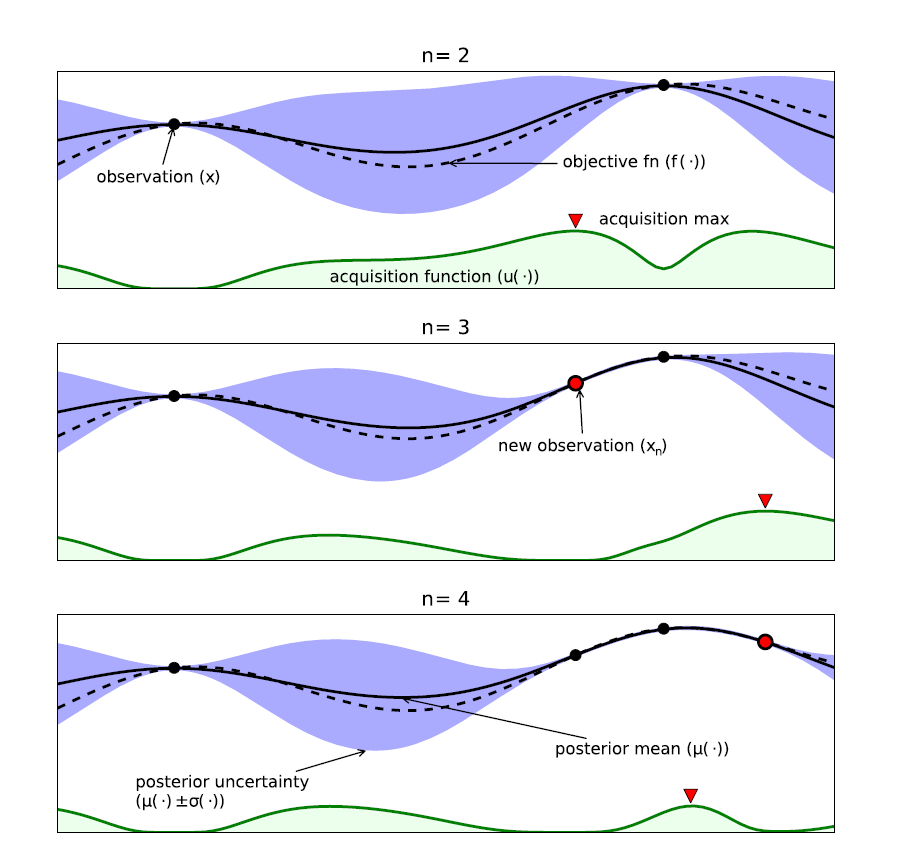


图3 贝叶斯优化算法可视化[4]

虚线为目标函数，实线代表概率代理模型的期望值，

紫色区域表示概率代理模型的不确定性，绿色为采集函数的值。

贝叶斯优化如今被广泛应用深度神经网络的超参数调整[5]、设计工程系统[6]、选择材料[7]和药物设计[8]中的流产实验、强化学习[4]等领域。

近些年来，贝叶斯优化也逐渐被应用到流程工业领域，如控制器的自动整定[9]等并取得了较好的效果，但整体研究较少，仍存在很大研究的空间。

1. **贝叶斯优化现有的研究进展**
   1. **概率模型**

概率模型分为以下两种[2]：

1. 参数模型：Beta-Bernoulli模型、线性模型、广义线性模型等
2. 非参数模型：高斯过程、随机森林

其中，非参数模型由于其对于决策变量弱假设性（仅需通过一致连续或利普希茨连续等局部平滑性弱假设）、函数高可扩展性[2]、不易发生“过拟合”[10]等性质被广泛应用。在非参数模型中，对于连续的目标函数，高斯过程概率模型广为应用。

高斯过程（Gaussian Processes）[11]由一个均值函数和一个半正定的协方差函数构成：

在已知历史数据，下一次采样点的前提下，可以获得下一次采样值的概率分布：

进一步假设目标函数的观测值带有噪声，并服从零均值高斯分布：

则的概率分布均值和方差可以解析的写成以下形式：

决定高斯过程的两个重要参数：

1. 均值函数

先验均值函数表示对于目标函数值期望的偏移量，在实际应用中，指定一个明确的、合理的先验均值函数十分困难[11]，一般假设其为常数，不过，该假设对后验准确性影响很小。

1. 协方差函数

协方差函数是高斯过程中计算两个数据点之间相似性的函数,它指定了未知目标函数的平滑性和振幅，对于对后验准确性影响大[12]，如下图所示：

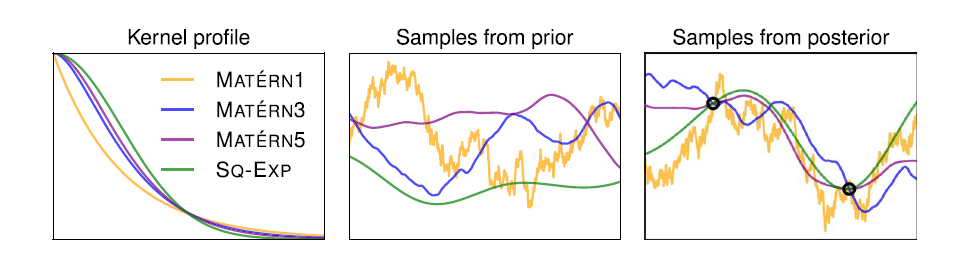


图4 不同高斯核的差别[12]

协方差函数分为非平稳协方差函数、平稳协方差函数，典型的平稳协方差函数由有：

* 伽马指数函数
* Matérn协方差函数
* 二次有理函数
  1. **采集函数**

常见的采集函数有4种；

1. 基于提升的策略：观测值获得提升的能力
2. 基于置信边界的策：对于未知区域持乐观态度
3. 基于信息的策略：评估可以对于最优点的求解带来多少信息上的提升

[17]

1. 采集函数组合策略：
   1. **贝叶斯优化在流程工业领域的应用挑战**

想要将贝叶斯优化的方法应用在流程工业领域，总结有以下几个挑战：

1. 概率模型、采集函数的选择对于贝叶斯优化的结果影响很大，没有普遍适用的具体算法。对于特定流程工业背景下的优化问题，如何根据问题的特性选择合适的（非）参数先验概率模型以及超参数，有待进一步研究。
2. 在问题维度高或样本量大时，贝叶斯优化更新概率模型需要高昂的计算量[2]，但工业领域普遍对问题求解的实时性要求高，需要研究、应用进一步的加速方法，如概率模型近似技术、采集函数并行化等。
3. 仿真模拟在流程工业过程控制与优化扮演非常重要的角色，贝叶斯优化技术可以辅助仿真模型的参数配置，仿真模型也可以为贝叶斯优化提供更精确的观测模型，但常用的流程模拟软件：UniSim、HYSYS 、ASPEN PLUS、Simulink等并没有方便调用和配置的贝叶斯优化算法模块，需要建立算法库与典型流程模拟软件的通讯接口。
4. **研究内容**

基于上述问题，我的研究内容有以下三点。

* 1. **基于贝叶斯优化的代理建模优化算法库**

首先建立一个贝叶斯优化的代理建模优化算法库，如下图所示：

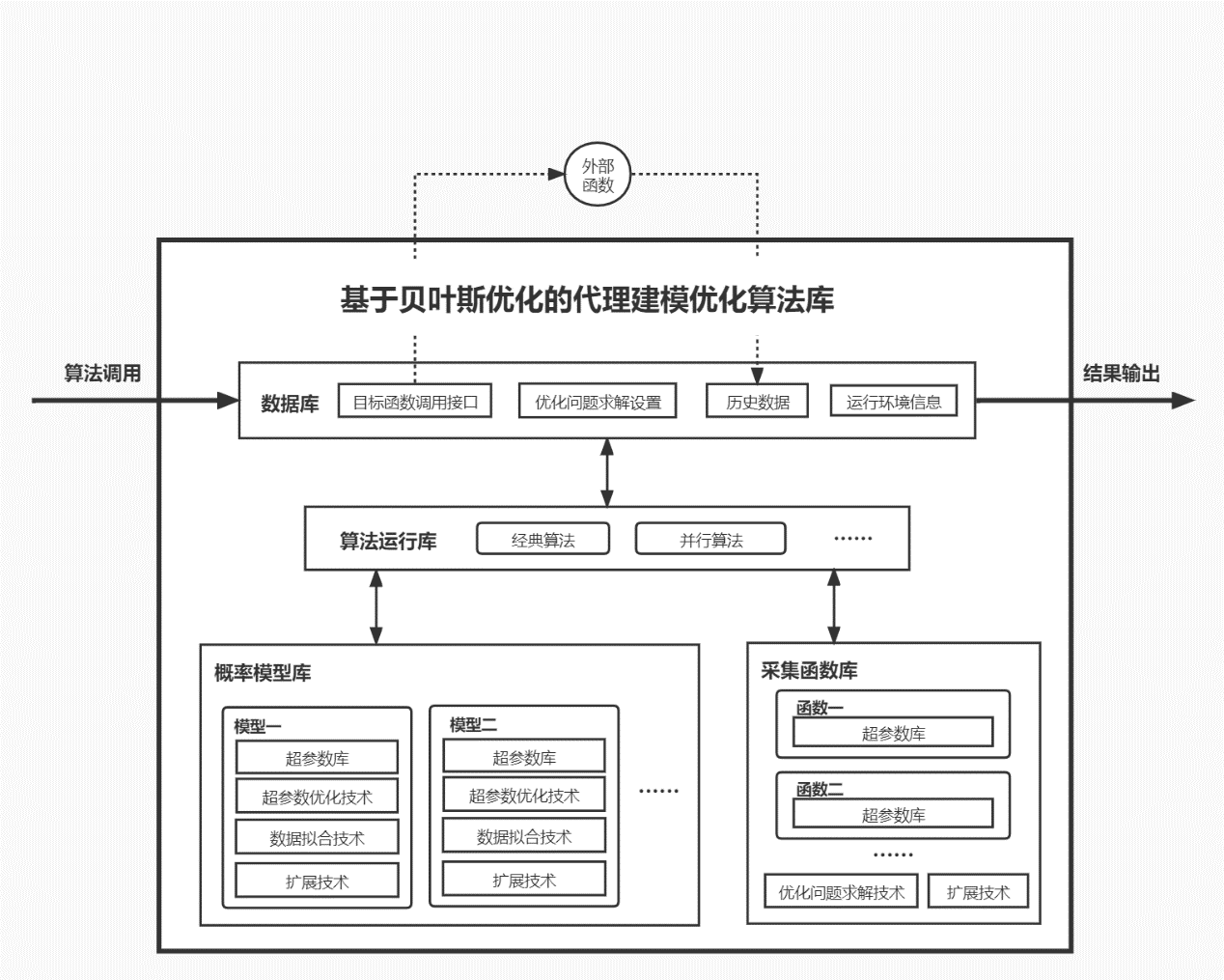


图5 不同高斯核的差别[12]

包含数据库，储存具体问题的目标函数调用接口、优化问题求解设置、运行环境信息以及历史数据。

算法运行库将根据数据库的信息调用概率模型和采集函数，并将计算得到的数据重新储存在历史数据中。

* 1. **基于贝叶斯优化的非线性全局优化问题求解方法**

贝叶斯优化在求解非线性全局优化问题时，其由于对于目标函数没有任何结构、导数的假设[1]，只把其当作黑箱处理；并且在面对复杂的评估代价大的目标函数时，贝叶斯优化将通过统计的方法，高效进行寻优采样，减少目标函数的评估次数以提升优化效率。

总的来说，贝叶斯优化有助于解决复杂的非线性全局优化问题，并且在理论上可以实现较快的求解，对于非线性优化问题，计划将重点研究以下内容：

1. 求解具体的流程工业过程控制与优化中的非线性全局优化问题，并分析该问题的特性，研究在不同要求下，贝叶斯优化具体的算法（如概率模型、采集函数）。
2. 对于高维决策变量、大样本的情况，研究加速求解技术：采集函数的并行求解技术[20][21][22]，概率模型近似技术等[23][24][25]。

在流程工业过程控制与优化中，有很多复杂的非线性全局优化问题。以阀门粘滞诊断问题为例，下阐述如何用贝叶斯优化框架进行阀门粘滞诊断：阀门粘滞是阀门在运动时受到了不可忽略静摩擦力影响，使得阀门的输入和输出（即图中的）呈现出滞环特性。

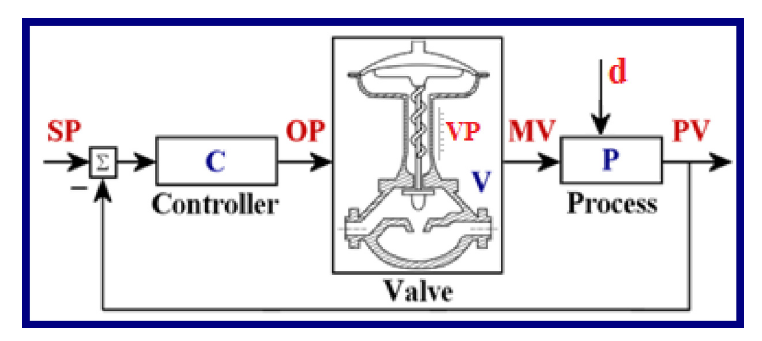


图6 控制回路的符号表示[19]

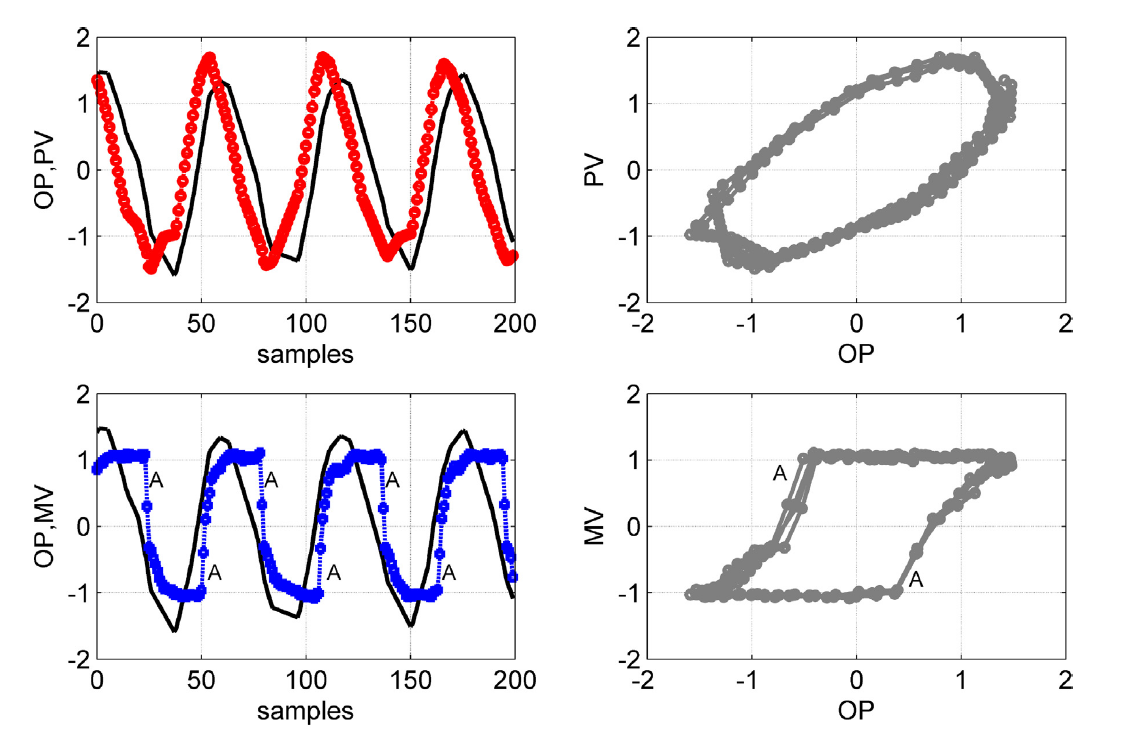


图7 阀门粘滞的输入和输出[19]

然而在实际中，通常是不可测或难以获得的，通常测量的量是和，而被控对象常常是非线性的，所以阀门粘滞诊断问题常常转化为一个模型辨识问题。

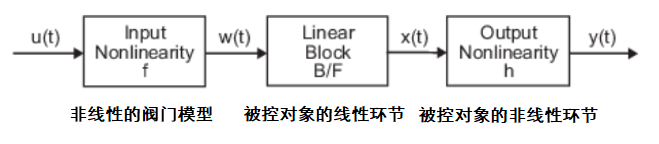


图8 阀门粘滞回路的模型

进一步，模型辨识时，由于辨识参数的范围难以界定，所以转化为一个非线性全局优化问题进行求解。所以可以将阀门粘滞的辨识问题纳入到贝叶斯优化的求解框架下，设置目标函数为建模的误差，优化变量为模型的参数，进行优化求解。

* 1. **基于贝叶斯优化的流程模拟变量优化方法**

流程模拟的变量优化时，由于变量搜索空间大，且互相影响，且一次仿真的计算量大，所以靠人工经验优化难度大，但贝叶斯优化可以对于多变量优化问题进行联合优化[12]，可以在仿真结果评估函数的帮助下，基于仿真结果实现闭环自动优化，避免人为调试。

在用贝叶斯优化框架进行流程模拟变量优化时，可以设置目标函数为评估结果（如经济效益），优化变量为模型的参数，进行优化求解。

这部分的研究重点

1. 建立算法库与典型流程模拟软件的通讯接口；
2. 通过贝叶斯优化算法库实现流程模拟变量的闭环自动优化。

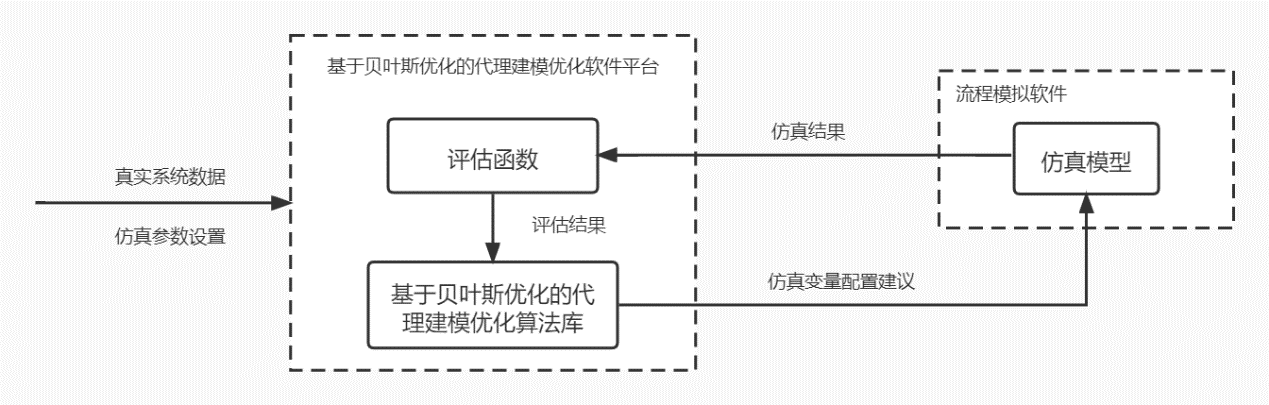


图10 流程模拟变量的闭环自动优化

1. **工作计划**
2. 秋季学期13-17周，阅读文献，学习典型的贝叶斯实现，以c#实现典型的贝叶斯优化算法库。
3. 春季学期1-3周，阅读文献，求解典型复杂非线性优化问题。
4. 春季学期4-8周，阅读文献，研究并实现贝叶斯优化的加速算法 。
5. 春季学期9-12周，实现算法库与典型仿真软件的通讯，基于贝叶斯优化进行仿真参数优化，连接数据库实现过程记录。
6. 春季学期13-16周，撰写毕业论文。
7. **参考文献**
8. Frazier P I. A tutorial on Bayesian optimization[J]. arXiv preprint arXiv:1807.02811, 2018.
9. Cui JX, Yang B. Survey on Bayesian optimization methodology and applications. Ruan Jian Xue Bao/Journal of Software, 2018,29(10):3068-3090 (in Chinese). http://www.jos.org.cn/1000-9825/5607.htm
10. Chen Wang,Qingyun Duan,Wei Gong,Aizhong Ye,Zhenhua Di,Chiyuan Miao. An evaluation of adaptive surrogate modeling based optimization with two benchmark problems[J]. Environmental Modelling and Software,2014,60:
11. E. Brochu, V. M. Cora, and N. de Freitas,‘‘A tutorial on Bayesian optimization of expensive cost functions, with application to active user modeling and hierarchical reinforcement learning,’’ Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, Tech. Rep. UBC TR-2009-23, 2009
12. J. Snoek, H. Larochelle, and R. P. Adams. Practical Bayesian optimization of machine learning algorithms. F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, eds. Advances in Neural Information Processing Systems 25. Curran Associates, Red Hook, NY, 2951–2959, 2012.
13. A. Forrester, A. S´obester, and A. Keane. Engineering Design via Surrogate Modelling: A Practical Guide. John Wiley & Sons, Chichester, UK, 2008.
14. P. I. Frazier and J. Wang. Bayesian optimization for materials design. T. Lookman, F. J. Alexander, and K. Rajan, eds. Information Science for Materials Discovery and Design. Springer International, Cham, Switzerland, 45–75, 2016.
15. D. M. Negoescu, P. I. Frazier, and W. B. Powell. The knowledge gradient algorithm for sequencing experiments in drug discovery. INFORMS Journal on Computing 23(1):46–363, 2011.
16. M. Neumann-Brosig, A. Marco, D. Schwarzmann and S. Trimpe, "Data-Efficient Autotuning With Bayesian Optimization: An Industrial Control Study," in IEEE Transactions on Control Systems Technology, vol. 28, no. 3, pp. 730-740, May 2020, doi: 10.1109/TCST.2018.2886159.
17. Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature, 2015,521:452—459.
18. Rasmussen C E. Gaussian processes in machine learning[C]//Summer school on machine learning. Springer, Berlin, Heidelberg, 2003: 63-71.
19. Shahriari B, Swersky K, Wang Z, Adams RP, Freitas ND. Taking the human out of the loop: A review of Bayesian optimization.Proc. of the IEEE, 2016,104(1):148—175.
20. H. J. Kushner, ‘‘A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise,’’ J. Fluids Eng., vol. 86, no. 1, pp. 97–106, 1964.
21. J. Mocˇkus, V. Tiesis, and A. Z ˇ ilinskas, ‘‘The application of Bayesian methods for seeking the extremum,’’ in Toward Global Optimization, vol. 2, L. Dixon and G. Szego, Eds. Amsterdam, The Netherlands: Elsevier, 1978.
22. N. Srinivas, A. Krause, S. M. Kakade, and M. Seeger, ‘‘Gaussian process optimization in the bandit setting: No regret and experimental design,’’ in Proc. Int. Conf. Mach. Learn., 2010, pp. 1015–1022.
23. J. M. Herna´ndez-Lobato, M. W. Hoffman, and Z. Ghahramani, ‘‘Predictive entropy search for efficient global optimization of black-box functions,’’ in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 918–926.
24. P. Auer, N. Cesa-Bianchi, Y. Freund, and R. E. Schapire, ‘‘Gambling in a rigged casino: The adversarial multi-armed bandit problem,’’ in Proc. Symp. Found. Comput. Sci., 1995, pp. 322–331.
25. B. Shahriari, Z. Wang, M. W. Hoffman, A. Bouchard-Coˆte´, and N. de Freitas, ‘‘An entropy search portfolio,’’ in Proc. NIPS Workshop Bayesian Optim., 2014.
26. di Capaci R B, Scali C. Review and comparison of techniques of analysis of valve stiction: From modeling to smart diagnosis[J]. Chemical Engineering Research and Design, 2018, 130: 230-265.
27. D. Lizotte, R. Greiner, and D. Schuurmans, ‘‘An experimental methodology for response surface optimization methods,’’ J. Global Optim., vol. 53, pp. 1–38, 2011.
28. F. Hutter, H. H. Hoos, and K. Leyton-Brown, ‘‘Parallel algorithm configuration,’’ Learning and Intelligent Optimization, Berlin, Germany: Springer-Verlag, 2012, pp. 55–70.
29. D. Jones, ‘‘A taxonomy of global optimization methods based on response surfaces,’’ J. Global Optim., vol. 21, no. 4, pp. 345–383, 2001.
30. M. Seeger, C. Williams, and N. Lawrence, ‘‘Fast forward selection to speed up sparse Gaussian process regression,’’ in Proc. Artif. Intell. Stat. 9, 2003, pp. 1–8.
31. E. Snelson and Z. Ghahramani, ‘‘Sparse Gaussian processes using pseudo-inputs,’’ in Proc. Adv. Neural Inf. Process. Syst., 2005, pp. 1257–1264.
32. M. La´zaro-Gredilla, J. Quin˜nonero-Candela, C. E. Rasmussen, and A. R. Figueiras-Vidal, ‘‘Sparse spectrum Gaussian process regression,’’ J. Mach. Learn. Res., vol. 11, pp. 1865–1881, 2010.