杭州電子科技大学

硕　士　学　位　论　文

题　　目： 高斯过程回归模型在大数据上拓展的研究

研 究 生 周昌凯

专　　业 应用统计

指导教师 王文胜 教授

企业导师 许亦频

完成日期 2021 年 10 月

杭州电子科技大学硕士学位论文

高斯过程回归模型在大数据上拓展的研究

研 究 生：周昌凯

指导教师：王文胜　教授

2021 年 10 月

Dissertation Submitted to Hangzhou Dianzi University for the Degree of Master

Gaussian Process Regression Research for Big Data

Candidate：　Changkai Zhou

Supervisor：　Prof. Wensheng Wang

October，2021

杭州电子科技大学

学位论文原创性声明和使用授权说明

原创性声明

本人郑重声明： 所呈交的学位论文，是本人在导师的指导下，独立进行研究工作所取得的成果。除文中已经注明引用的内容外，本论文不含任何其他个人或集体已经发表或撰写过的作品或成果。对本文的研究做出重要贡献的个人和集体，均已在文中以明确方式标明。

申请学位论文与资料若有不实之处，本人承担一切相关责任。

论文作者签名： 日期： 年 月 日

学位论文使用授权说明

本人完全了解杭州电子科技大学关于保留和使用学位论文的规定，即：研究生在校攻读学位期间论文工作的知识产权单位属杭州电子科技大学。本人保证毕业离校后，发表论文或使用论文工作成果时署名单位仍然为杭州电子科技大学。学校有权保留送交论文的复印件，允许查阅和借阅论文；学校可以公布论文的全部或部分内容，可以允许采用影印、缩印或其它复制手段保存论文。（保密论文在解密后遵守此规定）

论文作者签名： 日期： 年 月 日

指导教师签名： 日期： 年 月 日

摘 要

近二三十年，机器学习中最活跃的研究方向之一是开发实用的贝叶斯方法来解决“学习”问题。高斯过程在机器学习领域的应用展现出了一种最重要的贝叶斯机器学习方法，其基于给定函数空间上先验分布的有效办法。同样，作为核方法，高斯过程为机器学习提供了一个有依据的、实用的、概率性的框架。长期的理论与应用的发展使高斯过程在解释性方面具有优势，并为学习和模型选择提供了一个有依据的框架，最终使得高斯过程模型在监督学习方面占据重要的地位。

然而，完整的高斯过程模型最突出的弱点在于其难以应用于大数据。给定包含 n 个样本的数据集，标准的高斯过程模型在训练过程中的时间复杂度为$\mathcal{O}\left(n^3\right)$，因为需要对$n \times n$的协方差矩阵求逆、求行列式；在预测过程中需要花费$\mathcal{O}\left(n^2\right)$，因为使用矩阵向量乘法去加速该过程。该弱点局限了标准的高斯过程模型难以应用于数据量大小大为$\mathcal{O}\left(10^4\right)$的数据集。

高斯过程模型在大数据集上的拓展形式是长久的需求，无论是基于模型本身的限制，还是大数据时代的背景。但由于该方向的研究在高斯过程模型被广泛应用以来一直是热门的领域，故本文在总结主流的拓展方法之后，一基于聚合模型的框架提出双层的在大数据集上可保持一致性的高斯过程模型，二基于分布式异方差稀疏高斯过程模型，研究如何添加诱导点使得近似模型能够保持原模型的精度。其中，一致性理论来源于高斯过程模型与克里金插值法的联系，并且高斯过程模型与其他模型的联系(如神经网络)揭露了高斯过程模型一些有趣的性质。

本文的实验基于玩具数据集及大量的现实数据集，在多方面评价改进模型的提升效果。实验结果显示：一双层聚合模型在大数据上能够保持预测的一致性，在聚合模型类中保持最优的预测精度；二为模型添加诱导点的方法能够还原完整模型的预测能力。

关键词：高斯过程回归，大数据，稀疏近似，多层模型，专家模型

Abstract

In the past two to three decades, one of the most active research directions in machine learning has been the development of practical Bayesian methods to solve "learning" problems. The application of Gaussian process in machine learning shows one of the most important Bayesian machine learning methods, which is an effective method based on the given prior distribution over a function space. Similarly, as a kernel method, the Gaussian process provides a principled, practical, and probabilistic framework for machine learning. The long-term development of theory and application has given the Gaussian process an advantage in terms of interpretability, and has provided a basis for learning and model selection. Finally, the Gaussian process model occupies an important position in supervised learning.

However, the most prominent weakness of the full Gaussian process model is that it is difficult to apply to big data. Given a data set containing n samples, the time complexity of the full Gaussian process model in the training process is $\mathcal{O}\left(n^3\right)$ because the $n \times n$ covariance matrix needs to be inverted and the determinant is calculated; it needs to be spent $\mathcal{O}\left(n^2\right)$ in prediction because the matrix vector multiply is used to speed up the process. This weakness restricts the full Gaussian process model for data sets with $\mathcal{O}\left(10^4\right)$ samples.

The expansion of the Gaussian process model on large data sets is a long-term demand, whether it is based on the limitations of the model itself or the background of the big data era. However, because the research in this direction has been a hot field since the Gaussian process model is widely used, this article summarizes the mainstream expansion methods. First, we based on the aggregation model proposes a two-tiered framework that can maintain consistency on large data sets. Gaussian process model. Second, based on the distributed heteroscedasticity sparse Gaussian process model, we study how to add inducing points so that the approximate model can maintain the accuracy of the original model. Among them, the consistency theory comes from the connection between the Gaussian process model and the Kriging method, and the connection between the Gaussian process model and other models (e.g., neural networks) reveals some interesting properties of the Gaussian process model.

The experiment in this paper is based on toy data sets and a large number of real data sets, and evaluates the improvement effect of the improved model in many aspects. The experimental results show that: a two-layer aggregation model can maintain the consistency of prediction on big data, and maintain the best prediction accuracy in the aggregation model category; second, the method of adding induction points to the model can restore the prediction ability of the complete model.

Keywords: Gaussian process regression, big data, sparse approximation, hierarchical model, mixture of experts

目录

参考文献

[1] 李金昌. 话说“回归”[J]. 中国统计，2020，68(10)：31-33.
[2] （奥）薛定谔. 生命是什么[M]. 海南：海南出版社，2017：1-97.
[3] （比）普里戈金. 从混沌到有序[M]. 上海：上海译文出版社，2005：1-314.
[4] 尼克. 人工智能简史[M]. 北京：人民邮电出版社，2017：1-255.
[5] Murphy, K.P.. Machine Learning: A Probabilistic Perspective[M]. Massachusetts: The MIT Press, 2012: 1-24.
[6] Bishop, C.M.. Pattern Recognition and Machine Learning[M]. New York: Springer Science+Business Media, 2007: 1-55.
[7] Hastie, T., Tibshirani, R., Wainwright, M.. Statistical Learning with Sparsity: The Lasso and Generalizations[M]. Los Angeles: CRC Press, 2015: 1-7.
[8] Shalev-Shwartz, S., Ben-David, S.. Understanding Machine Learning: From Theory to Algorithms[M]. Cambridge: Cambridge University Press, 2014: 60-66.
[9] Rasmussen, C.E., Williams, C.K.I.. Gaussian Processes for Machine Learning[M]. Massachusetts: The MIT Press, 2006: 1-218.
[10] Le, N.D., Zidek, J.V.. Statistical Analysis of Environmental Space-Time Processes[M]. New York: Springer Science+Business Media, 2006: 83-116.
[11] Cressie, N.. The origins of kriging[J]. Mathematical Geology, 1990, 22(03): 239-252.
[12] Chiles, J., Delfiner, P.. Geostatistics: Modeling Spatial Uncertainty (Second Edition)[M]. New Jersey: John Wiley & Sons, 2012: 31-238.
[13] Wahba, G.. Spline Models for Observational Data[M]. Pennsylvania: Society for Industrial and Applied Mathematics, 1990: 1-45.
[14] 牛文杰. 薄板样条法和泛克里金法在理论和应用方面的比较[J]. 工程图学学报, 2010, 31(04): 123-129.
[15] Bobrowski, A.. Functional Analysis for Probability and Stochastic Processes: An Introduction[M]. Cambridge: Cambridge University Press, 2005: 1-101.
[16] Brockwell, P.J., Davis, R.A.. Time Series: Theory and Methods (Second Edition)[M]. New York: Springer Science+Business Media, 2006: 1-76.
[17] Williams, D.. Probability with Martingales[M]. Cambridge: Cambridge University Press, 1991: 83-92.

[18] Banerjee, A., Guo X., Wang H.. On the optimality of conditional expectation as a Bregman predictor[J]. IEEE Transactions on Information Theory, 2005, 51(07): 2664-2669.

[19] Forrester, A.I.J., Sobester, A., Keane, A.J.. Engineering Design via Surrogate Modelling: A Practical Guide[M]. West Sussex: John Wiley & Sons, 2008: 33-153.

[20] Vazquez, E., Bect, J.. Pointwise consistency of the kriging predictor with known mean and covariance functions[A]. Giovagnoli A., Atkinson A., Torsney B., May C. (eds) mODa 9 – Advances in Model-Oriented Design and Analysis[C]. Berlin: Physica-Verlag HD, 2009: 221-228.

[21] Choi, T., Schervish, M.J.. Posterior consistency in nonparametric regression problems under Gaussian process priors[R]. Carnegie Mellon University, 2004.

[22] Neal, R.M.. Bayesian Learning for Neural Networks[D]. Ph.d. thesis, University of Toronto, 1995.

[23] Hanin, B.. Random neural networks in the infinite width limit as Gaussian processes[J]. arXiv preprint arXiv: 2107.01562, 2021.

[24] Lee, J., Bahri, Y., Novak, R., Schoenholz, S.S., Pennington, J., Sohl-Dickstein, J.. Deep neural networks as Gaussian processes[A]. Sixth International Conference on Learning Representations[C]. Vancouver: ICLR, 2018: 1-17.

[25] Williams, C.K.I.. Computing with infinite networks[A]. Advances in neural information processing systems 10[C]. Colorado: NeurIPS, 1997: 295-301.

[26] Cho, Y., Saul, L.K.. Kernel methods for deep learning[A]. Advances in Neural Information Processing Systems 22[C], Massachusetts: the MIT Press, 2009: 342-350.

[27] Matthews, A.G.D.G., Hron, J., Rowland, M., Turner, R.E., Ghahramani, Z.. Gaussian Process Behaviour in Wide Deep Neural Networks[A]. Sixth International Conference on Learning Representations[C]. Vancouver: ICLR, 2018: 1-15.

[28] Lee, C., Wu, J., Wang, W., Yue, X.. Neural Network Gaussian Process considering Input Uncertainty for Composite Structures Assembly[J]. IEEE/ASME Transactions on Mechatronics, 2020, doi: 10.1109/TMECH.2020.3040755.

[29] Pretorius, A., Kamper, H., Kroon, S.. On the expected behaviour of noise regularised deep neural networks as Gaussian processes[J]. Pattern Recognition

Letters, 2020, 138: 75-81.

[30] Garriga-Alonso, A., Rasmussen, C.E., Aitchison, L.. Deep convolutional networks as shallow Gaussian processes[A]. Seventh International Conference on Learning Representations[C]. Louisiana: ICLR, 2019: 1-16.

[31] Novak, R., Xiao, L., Lee, J., Bahri, Y., Yang, G., Abolafia, D.A., Pennington, J., Sohl-Dickstein, J.. Bayesian deep convolutional networks with many channels are Gaussian processes[A]. Third workshop on Bayesian Deep Learning (NeurIPS 2018)[C]. Canada: NeurIPS, 2018, 1-27.

[32] Wilson, A.G., Hu, Z., Salakhutdinov, R., Xing, E.P.. Deep kernel learning[A]. 19th International Conference on Artificial Intelligence and Statistics[C]. PRML: W&CP, 2016, 51: 370-378.

[33] Wilson, A.G., Hu, Z., Salakhutdinov, R., Xing, E.P.. Stochastic variational deep kernel learning[A]. 30th International Conference on Neural Information Processing Systems[C]. New York: Curran Associates Inc., 2016, 2594-2602.

[34] Bui, T.D., Hernandez-Lobato, D.H., Li, Y., Hernandez-Lobato, J.M.. Deep Gaussian processes for regression using approximate expectation propagation[J]. arXiv preprint arXiv: 1602.04133, 2016.

[35] Damianou, A., Lawrence, N.D.. Deep Gaussian processes[A]. Sixteenth International Conference on Artificial Intelligence and Statistics[C]. PMLR, 2013, 31: 207-215.

[36] Jain, A., Srijith, P.K., Khan, M.E.. Subset-of-data variational inference for deep Gaussian-processes regression[A]. 37th Conference on Uncertainty in Artificial Intelligence[C]. PMLR, 2021, (in press): 1-15.

[37] Pleiss, G., Cunningham, J.P.. The limitations of large width in neural networks: a deep Gaussian process perspective [J]. arXiv preprint arXiv: 2106.06529, 2021.

[38] Duvenaud, D., Rippel, O., Adams, R.P., Ghahramani, Z.. Avoiding pathologies in very deep networks[A]. 17th International Conference on Artificial Intelligence and Statistics[C]. PRML: W&CP, 2016, 33: 202-210.

[39] Dunlop, M.M., Girolami, M.A., Stuart, A.M., Teckentrup, A.L.. How Deep Are Deep Gaussian Processes?[J]. Journal of Machine Learning Research, 2018, 19(54): 1-46.

[40] Garnelo, M., Rosenbaum, D., Maddison C.J., Ramalho, T., Saxton, D., Shanahan, M., Teh, Y.W., Rezende D.J., Eslami S.M.A.. Conditional neural processes[A]. 35th International Conference on Machine Learning[C]. PMLR, 2018, 80:

1704-1713.

[41] Garnelo, M., Schwarz, J., Rosenbaum, D., Viola, F., Rezende D.J., Eslami S.M.A., Teh, Y.W.. Neural Processes[A]. the ICML 2018 workshop on Theoretical Foundations and Applications of Deep Generative Models[C]. ICML, 2018, 1-11.

[42] Kim, H., Mnih, A., Schwarz, J., Garnelo, M., Eslami, A., Rosenbaum, D., Vinyals, O., Teh, Y.W.. Attentive neural processes[A]. Third workshop on Bayesian Deep Learning (NeurIPS 2018)[C]. NeurIPS, 2018, 1-17.

[43] Friedrich, J.. Neuronal Gaussian process regression[A]. Advances in Neural Information Processing Systems 34[C]. Vancouver: NeurIPS, 2020, (in press): 1-11.

[44] Hastie, T., Tibshirani, R., Friedman, J.. The Elements of Statistical Learning: Data Mining, Inference, and Prediction (Second Edition)[M]. Springer, 2017, 139-189.

[45] Raket, L.L.. Differential equations, splines and Gaussian processes[J]. arXiv preprint arXiv: 2102.03306, 2021.

[46] Alvarez, M., Luengo, D., Lawrence, N.D.. Latent force model[A]. 12th International Conference on Artificial Intelligence and Statistics (AISTATS)[C]. JMLR: W&CP, 2009, 5: 9-16.

[47] McDonald, T.M., Alvarez, M.A.. Computational modeling of nonlinear dynamical systems with ODE-based random features[J]. arXiv preprint arXiv: 2106.0596, 2021.

[48] Lindgren, F., Rue, H., Lindstrom, J.. An explict link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differenial equation approach[J]. Journal of the Royal Statistical Society: Series B, 2011, 73(04): 423-498.

[49] Rasmussen, C.E.. Evaluation of Gaussian processes and other methods for non-linear regression[D]. Ph.d. thesis, University of Toronto, 1999.

[50] Liu, H., Cai, J., Ong, Y., Wang, Y.. Understanding and comparing scalable Gaussian process regression for big data[J]. Knowledge-Based Systems, 2019, 164:324-335.

[51] Quinonero-Candela, J., Rasmussen, C.E.. A Unifying View of Sparse Approximate Gaussian Process Regression[J]. Journal of Machine Learning Research, 2005, 6: 1939-1959.

[52] Williams, C.K.I., Seeger, M.. Using the Nystrom Method to Speed Up Kernel Machines[A]. Advances in Neural Information Processing Systems 13[C]. MIT Press, 2001.

[53] Snelson, E., Ghahramani, Z.. Local and global sparse Gaussian process approximations[A]. Eleventh International Conference on Artificial Intelligence and Statistics[C]. PMLR, 2: 524-531.

[54] Low, K.H., Yu, J., Chen, J., Jaillet, P.. Parallel Gaussian process regression for big data: low-rank representation meets Markov approximation[A]. Association for the Advancement of Artificial Intelligence 2015[C]. AAAI Press, 2015, 1-10.

[55] Snelson, E., Ghahramani, Z.. Sparse Gaussian Process Using Pseudo-inputs[A]. Advances in Neural Information Processing Systems 18[C], MIT Press, 2006, 18: 1257--1264.

[56] Walder, C., Kim, K.I., Scholkopf, B.. Sparse multiscale Gaussian process regression[A]. 25th International Conference on Machine Learning[C]. ICML, 2008, 1112-1119.

[57] Bui, T., Turner R.. Tree-structured Gaussian Process Approximations[A]. Advances in Neural Information Processing Systems 26[C]. NeurIPS, 2014, 3: 2213-2221.

[58] Figueirasvidal, A., Lázarogredilla M.. Inter-domain Gaussian Processes for Sparse Inference using Inducing Features[A]. Advances in Neural Information Processing Systems 21[C], MIT Press, 2009, 21: 1087--1095.

[59] Lázaro-Gredilla, M., Quiñonero-Candela, J., Rasmussen, C.E., Figueiras-Vidal A.R.. Sparse Spectrum Gaussian Process Regression[J]. Journal of Machine Learning Research, 2010, 11(9):1865-1881.

[60] Hoang, Q.M., Hoang, T.N., Pham, H., Woodruff, D.P.. Revisiting the sample complexity of sparse spectrum approximation of Gaussian processes[A]. 34th Conference on Neural Information Processing Systems[C]. NeurIPS, 2020, 1-11.

[61] Wilson, A.G., Adams, R.P.. Gaussian process kernels for pattern discovery and extrapolation[A]. 30th International Conference on Machine Learning[C]. JMLR: W&CP, 2013, 28: 1067-1075.

[62] Rahimi, A., Recht, B.. Random features for large-scale kernel machines[A]. 20th International Conference on Neural Information Processing Systems[C]. New York: Curran Associates Inc., 2007, 1177-1184.

[63] Solin, A., Särkkä, S. Hilbert space methods for reduced-rank Gaussian process

regression[J]. Statistics and Computing, 2019, 30: 419–446.

[64] Bengio, Y., Vincent, P., Paiement, J.. Spectral clustering and kernel PCA are learning eigenfunctions[R]. CIRANO Working Papers, 2003, 2003s-19.

[65] Wilson, A.G., Dam, C., Nichisch, H.. Thoughts on massively scalable Gaussian processes[J]. arXiv preprint arXiv: 1511.0187, 2015.

[66] Titsias, M.K.. Variational learning of inducing variables in sparse Gaussian processes[A]. 12th International Conference on Artificial Intelligence and Statistics[C]. JMLR: W&CP, 2009, 5: 567-574.

[67] Titsias, M.K.. Variational Model Selection for Sparse Gaussian Process Regression[R]. 2009.

[68] Hoang, T.N.. A distributed variational inference framework for unifying parallel sparse Gaussian process regression models[A]. 33rd International Conference on International Conference on Machine Learning[C]. PLMR, 2016, 48: 382-391.

[69] Gal, Y., Mark, V., Rasmussen, C.E.. Distributed Variational Inference in Sparse Gaussian Process Regression and Latent Variable Models[A]. Advances in Neural Information Processing Systems 2014[C]. NeurIPS, 2014, 1-9.

[70] Liu H., Ong Y., Cai, J.. Large-scale heteroscedastic regression via Gaussian process[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 32(02): 708-721.

[71] Hensman, J., Fusi, N., Lawrence, N.D.. Gaussian processes for big data[J]. arXiv preprint arXiv: 1309.6835, 2013.

[72] Hoang, T.N., Hoang, Q.M., Low, K.H.. A unifying framework of anytime sparse Gaussian process regression models with stochastic variational inference for big data[A]. 32nd International Conference on International Conference on Machine Learning[C]. ICML, 2015, 37: 569-578.

[73] Bui, T. D., Yan, J., Turner, R.E.. A unifying framework for Gaussian process pseudo-point approximations using power expectation propagation[J]. Journal of Machine Learning Research, 2016, 18: 1-72.

[74] Liu H., Ong, Y., Shen, X., Cai, J.. When Gaussian Process Meets Big Data: A Review of Scalable GPs[J]. IEEE Transactions on Neural Networks and Learning Systems, 2020, 31(11): 4405-4423.

[75] Hernandez-Lobato, D., Hernandez-Lobato, J.M., Li, Y., Bui, T., Turner, R.E.. Stochastic expectation propagation for large scale Gaussian process classification[J]. arXiv preprint arXiv: 1511.03249, 2015.

[76] Matthews, A.G.D.G., Hensman, J., Turner, R.E., Ghahramani, Z.. On sparse variational methods and the Kullback-Leibler divergence between stochastic processes[A]. 19th International Conference on Artificial Intelligence and Statistics[C]. JMLR: W&CP, 2016, 41: 231-239.

[77] Hensman, J., Matthews, A.G.D.G., Filippone, M., Ghahramani, Z.. MCMC for variationally sparse Gaussian processes[A]. Advances in Neural Information Processing Systems 2015[C]. NuerIPS, 2015, 1648-1656.

[78] Huggins, J.H., Campbell, T., Kasprzak, M., Broderick, T.. Scalable Gaussian Process Inference with Finite-data Mean and Variance Guarantees[A]. 22nd International Conference on Artificial Intelligence and Statistics[C]. PMLR, 2019, 89: 1-20.

[79] Hensman, J., Durrande, N., Solin, A.. Variational fourier features for Gaussian processes[J]. The Journal of Machine Learning Research, 2017, 18(01): 5537-5588.

[80] Gal, Y., Turner, R.. Improving the Gaussian process sparse spectrum approximation by representing uncertainty in frequency inputs[A]. 32nd International Conference on International Conference on Machine Learning[C]. JMLR: W&CP, 2015, 37: 655-664.

[81] Bauer, M., Wilk, M.V.D., Rasmussen, C.E.. Understanding probabilistic sparse Gaussian Process approximations. Advances in Neural Information Processing Systems 2016[C]. NeurIPS, 2016, 1533-1541.

[82] Gibbs, M., Mackay, D.. Efficient implementation of Gaussian processes[R]. 1997.

[83] Gray, A.. Fast kernel matrix-vector multiplication with application to Gaussian process learning[R]. Carnegie Mellon University, 2004.

[84] Pleiss, G., Jankowiak, M., Eriksson, D., Damle, A., Gardner, J.R.. Fast matrix square roots with applications to Gaussian processes and Bayesian optimization[A]. 34th Conference on Neural Information Processing Systems[C]. NeurIPS, 2020, 1-14.

[85] Quinonero-Candela, J., Ramussen, C.E., Williams, C.K.I.. Approximation methods for Gaussian process regression[R]. MIT Press, 2007.

[86] Murray, I.. Gaussian processes and fast matrix-vector multiplies[A]. Workshop on numerical mathematics at the 26th International Conference on Machine Learning[C]. ICML, 2009, 1-4.

[87] Vanhatalo, J., Vehtari, A.. Modelling local and global phenomena with sparse Gaussian processes[J]. arXiv preprint arXiv: 1206.3290, 2012.

[88] Wilson, A.G., Nickisch, H.. Kernel interpolation for scalable structured Gaussian processes (KISS-GP)[A]. 32nd International Conference on International Conference on Machine Learning[C]. PRML, 2015, 37: 1775-1784.

[89] Fritz, J., Neuweiler, I., Nowak, W.. Application of FFT-based algorithms for large-scale universal Kriging problems[J]. Mathematical Geosciences, 2009, 41(5): 509-533.

[90] Shen, Y., Ng, A.Y., Seeger, M.. Fast Gaussian process regression using KD-Trees[A]. 18th International Conference on Neural Information Processing Systems[C]. NeurIPS, 2005, 1225-1232.

[91] Wang, K.A., Pleiss, G., Gardner, J.R., Tyree, S., Weinberger, K.Q., Wilson, A.G.. Exact Gaussian processes on a million data points[A]. 33rd Conference on Neural Information Processing Systems[C]. NeurIPS, 2019, 1-13.

[92] Yang, C., Duraiswami, R., Davis, L.. Efficient kernel machines using the improved fast Gauss transform[A]. Advances in neural information processing systems17[C] NeurIPS, 2004, 1561-1568.

[93] Chen J., Wang, L., Anitescu, M.. A fast summation tree code for Matern kernel[J]. SIAM Journal On Scientific Computing, 2014, 36(01): 289-309.

[94] Pleiss, G., Grandner, J.R., Weinberger, K.Q., Wilson, A.G.. Constant-time predictive distributions for Gaussian processes[A]. 35th International Conference on Machine Learning[C]. PMLR, 2018, 80: 4114-4123.

[95] Chalupka, K., Williams, C.K.I.. A framework for evaluating approximation methods for Gaussian process regression[J]. arXiv preprint arXiv: 1205.6326, 2012.

[96] Nguyen, D., Filippone, M., Michiardi, P.. Exact Gaussian process regression with distributed computations[A]. 34th ACM/SIGAPP Symposium on Applied Computing[C]. New York: Association for Computing Machinery, 2019, 1286-1295.

[97] Dietrich, C.R., Newsam, G.N.. Fast and exact simulation of stationary Gaussian processes through circulant embedding of the covariance matrix[J]. SIAM Journal on Scientific Computing, 1997, 18(4): 1088-1107.

[98] Ambikasaran, S., Foreman-Mackey, D., Greengard, L., Hogg D.W. O’Neil, M.. Fast direct methods for Gaussian processes[J]. IEEE Transactions on Pattern

Analysis and Machine Intelligence, 2016, 38(02): 252-265.

[99] Xu, Y., Yin, F., Zhang, J., Xu, W., Cui, S., Luo, Z.. Scalable Gaussian process using inexact admm for big data[A] 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. IEEE, 2019, 1-5.

[100] Chen, H., Zheng, L., Kontar, R.A., Raskutti., G.. Stochastic gradient descent in correlated settings: a study on Gaussian processes[A]. 34th Conference on Neural Information Processing Systems[C]. NeurIPS, 2020, 1-12.

[101] Almosallam, I.A., Jarvis, M.J., Roberts, S.J.. GPz: non-stationary sparse Gaussian processes for heteroscedastic uncertainty estimation in photometric redshifts[J]. Monthly Notices of the Royal Astronomical Society, 2016, 462(1): 726-739.

[102] Raissi, M.. Parametric Gaussian process regression for big data[J]. arXiv preprint arXiv: 1704.03144, 2017.

[103] Sarkka, S., Solin, A., Hartikainen J.. Spatiotemporal Learning via Infinite-Dimensional Bayesian Filtering and Smoothing: A Look at Gaussian Process Regression Through Kalman Filtering[J]. IEEE Signal Processing Magazine, 2013, 30(4): 51-61.

[104] Hartikainen, J., Särkkä, S.. Kalman filtering and smoothing solutions to temporal Gaussian process regression models[A]. 2010 IEEE International Workshop on Machine Learning for Signal Processing[C]. IEEE, 2010, 379-384.

[105] Samo, Y.K., Roberts, S.. String and membrane Gaussian processes[J]. The Journal of Machine Learning Research, 2016, 17(01): 4485-4571.

[106] Choudhury, A., Nair, P.B., Keane, A.J.. A data parallel approach for large-scale Gaussian process modeling[A]. 2020 Society for Industrial & Applied Mathematics (SIAM) International Conference on Data Mining[C]. SDM, 2002, 95–111.

[107] Gramacy, R.B., Lee, H.K.. Bayesian Treed Gaussian Process Models With an Application to Computer Modeling[J]. Journal of the American Statistical Association, 2008, 103(483): 1119-1130.

[108] Park, C., Huang, J.Z., Ding, Y.. Domain Decomposition Approach for Fast Gaussian Process Regression of Large Spatial Data Sets[J]. Journal of Machine Learning Research, 2011, 12(4): 1697−1728.

[109] Park, C., Huang, J.. Efficient Computation of Gaussian Process Regression for Large Spatial Data Sets by Patching Local Gaussian Processes[J]. Journal of

Machine Learning Research, 2016, 17(1): 6071-6099.

[110] Park, C., Apley, D.. Patchwork Kriging for large-scale Gaussian process regression[J]. Journal of Machine Learning Research, 2018, 19(1): 269-311.

[111] Urtasun, R., Darrell, T.. Sparse probabilistic regression for activity-independent human pose inference[A]. 2008 IEEE Conference on Computer Vision and Pattern Recognition[C]. IEEE, 2008, 1-8.

[112] Moore, D., Russell, S.J.. Gaussian Process Random Fields[A]. Advances in Neural Information Processing Systems 28[C]. NeurIPS, 2015, 21(3): 763-772.

[113] Das, S., Roy, S., Sambasivan, R.. Fast Gaussian process regression for big data[J]. Big Data Research, 2018, 14: 12-26.

[114] Chen, T., Ren, J.. Bagging for Gaussian process regression[J]. Neurocomputing, 2009, 72(7-9): 1605-1610.

[115] Zhu, J., Jiang, M., Peng, G., Yao, L., Ge, Z.. Scalable soft sensor for nonlinear industrial big data via bagging stochastic variational Gaussian processes[J]. IEEE Transactions on Industrial Electronics, 2021,68(08): 7594-7602.

[116] Cao, Y., Fleet, D.J.. Generalized product of experts for automatic and principled fusion of Gaussian process predictions[J]. arXiv preprint arXiv: 1410.7827, 2015.

[117] Hinton, G.E.. Training products of experts by minimizing contrastive divergence[J]. Neural Computation, 2002, 14(08): 1771-1800.

[118] Deisenroth, M.P., Ng, J.W.. Distributed Gaussian processes[A]. 32nd International Conference on Machine Learning[C]. PMLR: W&CP, 2015, 37: 1481-1490.

[119] Tresp, V.. A Bayesian committee machine[J]. Neural computation, 2000, 12(11): 2719-2741.

[120] Liu, H., Cai, J., Wang, Y., Ong, Y.S.. Generalized robust bayesian committee machine for large-scale Gaussian process regression[A]. 35th International Conference on Machine Learning[C]. PMLR, 2018, 80: 3131-3140.

[121] Li, N., Gao, Y., Li, W., Jiang Y., Xia S.. H-GPR: a hybrid strategy for large-scale Gaussian process regression[A]. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. IEEE, 2021, 2955-2959.

[122] Rullière, D., Durrande, N., Bachoc, F. Chevalier, C.. Nested Kriging predictions for datasets with a large number of observations[J]. Statistics and Computing, 2018, 28: 849–867.

[123] Bachoc, F., Durrande, N., Rullière, D., Chevalier, C.. Some properties of nested Kriging predictors[J]. arXiv preprint arXiv: 1707.05708, 2017.

[124] Ng, J.W., Deisenroth, M.P.. Hierarchical mixture-of-experts model for large-scale Gaussian process regression[J]. arXiv preprint arXiv: 1412.3078, 2014.

[125] Da, B., Ong, Y., Gupta, A., Feng, L., Liu, H.. Fast transfer Gaussian process regression with large-scale sources[J]. Knowledge-Based Systems, 2019, 165: 208-218.

[126] Gao, Y., Li, N., Ding, N., Li, Y., Dai, T., Xia, S.. Generalized local aggregation for large scale Gaussian process regression[A]. 2020 International Joint Conference on Neural Networks (IJCNN)[C] IEEE, 2020, 1-8.

[127] 刘晓芳，刘策，刘露咪，程丹松．重叠局部高斯过程回归[J]．哈尔滨工业大学学报, 2019, 51(11): 22-26.

[128] Jacobs, R.A., Jordan, M.I., Nowlan, S.J., Hinton, G.E.. Adaptive mixture of local experts[J]. Neural computation, 1991, 3: 79-87.

[129] Rasmussen, C.E., Ghahramani, Z.. Infinite mixtures of Gaussian process experts[A]. 14th International Conference on Neural Information Processing Systems: Natural and Synthetic[C]. NeurIPS, 2001, 881-888.

[130] Meeds, E., Osindero, S.. An alternative infinite mixture of Gaussian process experts[A] Advances in Neural Information Processing Systems 18[C]. NeuraIPS, 2005, 1-8.

[131] Nguyen, T.N.A., Bouzerdoum, A., Phung, S.L. Variational inference for infinite mixtures of sparse Gaussian processes through KL-correction[A]. 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. IEEE, 2016, 2579-2583.

[132] Nguyen, T.V., Bonilla, E.V.. Fast allocation of Gaussian process experts[A]. 31st International Conference on International Conference on Machine Learning[C]. JMLR, 2014, 32: 145-153.

[133] Shi, J.Q., Murray-Smith, R., Titterington, D.M.. Hierarchical Gaussian process mixtures for regression[J]. Statistics and computing, 2005, 15(1): 31-41.

[134] Nguyen-Tuong, D., Seeger, M., Peters, J., Koller, D., Bottou, L.. Local Gaussian process regression for real time online model learning and control[A]. Advances in Neural Information Processing Systems 21[C]. NeurIPS, 2008, 1193-1200.

[135] Nguyen, T.N.A., Bouserdoum, A., Phung, S.L.. Scalable hierarchical mixture of Gaussian processes for pattern classification[A]. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)[C]. IEEE, 2018, 2466-2470.

[136] Nguyen, T.N., Bouzerdoum, A., Phung, S.L.. Stochastic variational hierarchical mixture of sparse Gaussian processes for regression[J]. Machine Learning, 2018, 107(12): 1947-1986.

[137] Trecate, G.F., Williams, C.K.I., Opper, M.. Finite-dimensional approximation of Gaussian processes[A]. Advances in Neural Information Processing Systems 11[C]. MIT Press, 1999, 218-224.

[138] Zhu, H., Williams, C.K.I., Rohwer, R., Morciniec, M.. Gaussian regression and optimal finite dimensional linear models[J]. Neural networks and machine learning, 1997, 167-184.

[139] Dereziński, M., Khanna R., Mahoney, M.W.. Improved guarantees and a multiple-descent curve for the column subset selection problem and the Nystrom method[A]. Advances in Neural Information Processing Systems 34[C]. Vancouver: NeurIPS, 2020, 1-12.

[140] Burt, D.R., Rasmussen, C.E., Wilk, M.V.D.. Rates of convergence for sparse variational Gaussian process regression[A]. 36th International Conference on Machine Learning[C]. PMLR, 2019, 97: 862-871.

[141] Raykar, V.C., Duraiswami, R.. Fast large scale Gaussian process regression using approximate matrix-vector products[R]. 2007.

[142] Smola, A., Bartlett, P.. Sparse Greedy Gaussian process regression[A]. Advances in Neural Information Processing Systems 13[C]. MIT Press, 2000, 1-7.

[143] Gramacy, R.B., Apley, D.W.. Local Gaussian process approximation for large computer experiments[J]. Journal of Computational and Graphical Statistics, 2015, 24(2): 564-578.

[144] Schreiter, J., Englert, P., Nguyen-Tuong, D., Toussaint, M.. Sparse Gaussian process regression for compliant, real-time robot control[A]. 2015 IEEE International Conference on Robotics and Automation[C]. IEEE, 2015, 2586-2591.

[145] Herbrich, R., Lawrence, N., Seeger, M.. Fast sparse Gaussian process methods: the informative vector machine[A]. Advances in Neural Information Processing

Systems 15[C]. MIT Press, 2002, 1-8.

[146] Seeger M., Williams, C.K.I., Lawrence, N.D.. Fast forward selection to speed up sparse Gaussian process regression[A]. Ninth International Workshop on Artificial Intelligence and Statistics[C]. AISTATS, 2003, 1-8.

[147] Wang, W., Zhou, C.. A two-layer aggregation model with effective consistency for large-scale Gaussian process regression[J]. Engineering Applications of Artificial Intelligence, 2021, 106(2021): 104449.

[148] Jakkala, K.. Deep Gaussian processes: a survey. arXiv preprint arXiv: 2106.12135, 2021.

[149] Havasi, M., Hernandez-Lobato, J.M., Murillo-Fuentes, J.J.. Inference in Deep Gaussian Processes using Stochastic Gradient Hamiltonian Monte Carlo[A]. 32nd Conference on Neural Information Processing Systems[C]. NeurIPS, 2018, 1-11.

[150] Rasmussen, C.E., Nickisch, H.. Gaussian Processes for Machine Learning (GPML) Toolbox[J]. Journal of Machine Learning Research, 2010, 11: 3011–3015.

[151] Hamidieh, K.. A data-driven statistical model for predicting the critical temperature of a superconductor[J]. Computational Materials Science, 2018, 154: 346-354.

[152] Dua, D., Graff, C., 2019. UCI Machine Learning Repository[DB/OL]. http://archive.ics.uci.edu/ml.

[153] Neshat, M., Alexander, B., Sergiienko, N.Y., Wagner, M.. New insights into position optimization of wave energy converters using hybrid local search[J]. Swarm and Evolutionary Computation, 2020, 59: 100744.

[154] Neshat, M., Alexander, B., Wagner, M., Xia, Y.. A detailed comparison of meta-heuristic methods for optimising wave energy converter placements[A]. 2018 Genetic and Evolutionary Computation Conference[C]. GECCO, 2018, 1318–1325.

[155] Bertin-Mahieux, T., Ellis, D.P.W., Whitman, B., Lamere, P.. The million song dataset[J]. 12th International Society for Music Information Retrieval Conference[C]. ISMIR, 2011, 24–28.

[156] Burgués, J., Jiménez-Soto, J.M., Marco, S., 2018. Estimation of the limit of detection in semiconductor gas sensors through linearized calibration models[J]. Analytica Chimica Acta, 2018, 1013: 13–25.

[157] Burgués, J., Marco, S., 2018. Multivariate estimation of the limit of detection

by orthogonal partial least squares in temperature-modulated MOX sensors[J]. Analytica Chimica Acta, 1019: 49–64.

[158] Brooks, C., Burke, S., Persand, G.. Benchmarks and the accuracy of garch model estimation[J]. International Journal of Forecasting, 2001, 17: 45–56.

[159] Charles, A., Darne, O., 2019. The accuracy of asymmetric garch model estimation[J]. International Economics, 2019, 157: 179–202.

[160] Pereira, F.C., Antoniou, C., Fargas, J.A., Ben-Akiva, M.. A metamodel for estimating error bounds in real-time traffic prediction systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 15: 1310–1322.

[161] Urban, S., Ludersdorfer, M., Patrick, V.D.S.. Sensor calibration and hysteresis compensation with heteroscedastic gaussian processes[J]. IEEE Sensors Journal, 2015, 15: 6498–6506.

[162] Munoz-Gonzalez, L., Lazaro-Gredilla, M., Figueiras-Vidal, A.R.. Divisive gaussian processes for nonstationary regression[J]. IEEE Transactions on Neural Networks and Learning Systems, 2014, 25: 1991–2003.

[163] Munoz-Gonzalez, L., Lazaro-Gredilla, M., Figueiras-Vidal, A.R.. Laplace approximation for divisive gaussian processes for nonstationary regression[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38: 618–624.

[164] Saul, A.D., Hensman, J., Vehtari, A., Lawrence, N.D.. Chained gaussian processes[A]. 19th International Conference on Artificial Intelligence and Statistics[C]. PMLR, 2016, 51: 1431-1440.

[165] Goldberg, P.W., Williams, C.K.I., Bishop, C.M.. Regression with input-dependent noise: A gaussian process treatment[A]. Advances in Neural Information Processing Systems 10[C]. NeurIPS, 1997, 493-499.

[166] Kersting, K., Plagemann, C., Pfaff, P., Burgard,W.. Most-likely heteroscedastic gaussian process regression[A]. 24th International Conference on Machine Learning[C]. Association for Computing Machiner, 2007, 393–400.

[167] Heinonen, M., Mannerstrom, H., Rousu, J., Kaski, S., Lhdesmki, H.. Non-stationary gaussian process regression with hamiltonian monte carlo[A]. 19th International Conference on Artificial Intelligence and Statistics[C]. PMLR, 2016, 51: 732-740.

[168] Binois, M., Gramacy, R.B., Ludkovski, M.. Practical heteroscedastic gaussian process modeling for large simulation experiments[J]. Journal of Computational

and Graphical Statistics, 2018, 27: 808– 821.

[169] Zhang, Q.H., Ni, Y.Q.. Improved most likely heteroscedastic gaussian process regression via bayesian residual moment estimator[J]. IEEE Transactions on Signal Processing, 2020, 68: 3450–3460.

[170] Lázaro-Gredilla, M., Titsias, M.K.. Variational heteroscedastic gaussian process regression[A]. 28th International Conference on International Conference on Machine Learning[C]. Omnipress, 2011, 841-848.

[171] Menictas, M., Wand, M.P.. Variational inference for heteroscedastic semiparametric regression[J]. Australian & New Zealand Journal of Statistics, 2015, 57: 119–138.

[172] Munoz-González, L., Lázaro-Gredilla, M., Figueiras-Vidal, A.R.. Heteroscedastic gaussian process regression using expectation propagation[A]. 2011 IEEE International Workshop on Machine Learning for Signal Processing[C]. IEEE, 2011, 1-6.

[173] Tolvanen, V., Jylanki, P., Vehtari, A.. Expectation propagation for nonstationary heteroscedastic gaussian process regression[A]. 2014 IEEE International Workshop on Machine Learning for Signal Processing[C]. IEEE, 2014, 1-6.

[174] Hartmann, M., Vanhatalo, J.. Laplace approximation and natural gradient for gaussian process regression with heteroscedastic student-f model[J]. Statistics & Computing, 2019, 29: 753–773.

[175] Gittens, A., Mahoney, M.. Revisiting the Nystrom method for improved large-scale machine learning[A]. 30th International Conference on Machine Learning[C], PMLR, 28(3): 567-575, 2013.

[176] Alaoui, A.E., Mahoney, M.W.. Fast randomized kernel ridge regression with statistical guarantees[A]. 28th International Conference on Neural Information Processing Systems[C]. MIT Press, 2015, 775–783.

[177] Burt, D.R., Rasmussen, C.E., Mark, V.D.W.. Convergence of sparse variational inference in gaussian processes regression[J]. Journal of Machine Learning Research, 2020, 21: 1–63.

[178] Ji, C., Shen, H.. Stochastic variational inference via upper bound. arXiv preprint arXiv:1912.00650, 2019.

[179] Dieng, A.B., Tran, D., Ranganath, R., Paisley, J., Blei, D.M.. Variational inference via $\mathcal{X}$-upper bound minimization[A]. 31st International Conference on Neural Information Processing Systems[C]. Curran Associates Inc., 2017,

2729-2738.

[180] Brooks, F.T., Pope, D.S., Marcolini, A.M.. Airfoil self-noise and prediction[R]. Technical Report, NASA-RP-1218, 1989.

[181] Kaya, H., Tüfekci, P.. Local and global learning methods for predicting power of a combined gas & steam turbine[A] Intertational Conference on Emerging Trends in Computer and Electronics Engineering[C]. ICETCEE, 2012, 13-18.

[182] Tüfekci, P.. Prediction of full load electrical power output of a base load operated combined cycle power plant using machine learning methods[J]. International Journal of Electrical Power & Energy Systems, 2014, 60: 126-140.

[183] Kaya, H., Tüfekci, P., Uzun, E.. Predicting co and noxemissions from gas turbines: novel data and abenchmark pems[J]. Turkish Journal of Electrical Engineering & Computer Sciences, 2019, 27: 4783–4796.

[184] MacKay, D.J.. Neural Networks and Machine Learning[M]. Springer-Verlag, 1998, 133-165.

[185] Chen, X., Wang, B.. How priors of initial hyperparameters affect Gaussian process regression models[J]. Neurocomputing, 2018, 275: 1702-1710.

[186] Teng, T., Chen, J., Zhang, Y., Low, B.K.H.. Scalable Variational Bayesian Kernel Selection for Sparse Gaussian Process Regression[A]. 34th Conference on Artificial Intelligence[C]. AAAI, 2020, 34(4): 5997-6004

致谢

本人于杭州电子科技大学度过本科加硕士接近七年的时间，得到了许多帮助，借此机会表达诚挚的感谢：

首先，感谢导师王文胜教授，王老师的学识渊博本身就是良好的学习对象，在初入研究生期间指点了这个极具研究价值的模型，并且讨论期间提供的建议具备良好的合理性与参考价值，也在研究外提供了非常大的帮助。感谢家人们，为本人创造了非常优秀的学习条件。

同时，感谢经济学院的老师们，尤其是斯介生老师的建议与解答，和郑静老师在讨论班上的问题与建议。也感谢师姐师兄师弟师妹们，感谢研究生同班以及同学院的小伙伴们，以及本科时期授过课的理学院老师们。

其次，感谢浙江华为公司提供的实习机会，并且本文的第二章与第三章内容中大部分工作皆于实习工作之余完成，变相得到资助。同时感谢组内李滔与苏宝星在工作上提供的帮助。

然后，感谢本人读研期间在网络上认识的陈平、温铁军、卢麒元、艾跃进等老师，其思想深刻，犹如高屋建瓴，当为本人的学习榜样。具体的，本人于 2019 下半年开始接触陈平老师主持的“眉山论剑”栏目，当时关于“修例风波”的见解着实使本人耳目一新，之后便一直关注该老师，也翻阅了老师的论文集及出版物，引起了本人对生命科学与复杂系统的兴趣，颇具拨云见日之效。另一位同样使本人产生“熠熠生辉”感觉的好友曹宁于今年在莫斯科大学进修力学数学博士，在此祝愿他顺利毕业，也感谢他在本科与硕士期间和本人分享、交流观点。

再次，感谢华红光教练从大一开始提供的帮助以及其培养的锻炼习惯，终身受益。感谢刘海涛教授扎实的工作，本人硕士期间的研究内容主要基于其工作。感谢吴安琪教授，虽仅有一信交流，且本人也未满足 GT 的申请条件，但于 21 年 6 月关注到您的信息及研究领域，是本人完成本文第二章与第三章的主要动力。

感谢好友黄家杨、孔祥昊、周尹茜于论文的语言方面修改提供的帮助，促使本人发表了人生中第一篇论文。感谢好友尹卓立、曹宁在申请方面提供的帮助。感谢好友吴雨璇在语言学习方面提供的帮助。感谢好友卢从安在考研时提供的帮助。同样感谢其他在本科与硕士期间认识的小伙伴们。

最后，感谢百忙之中评阅论文和参加答辩的各位专家、教授。也感谢阅读本文的学生、学者们，因本人水平有限，故在文章细节处理方面不是非常到位，但也愿效抛砖引玉之力，希望对您的研究有所帮助。

附录 作者在读期间的主要学术成果

[1] Wang, W., Zhou, C.. A two-layer aggregation model with effective consistency for large-scale Gaussian process regression[J]. Engineering Applications of Artificial Intelligence, 2021, 106: 104449.

[2] Wang, W., Zhou, C.. Variational model selection of inducing points in sparse heteroscedastic Gaussian process regression [J]. Knowledge-Based Systems, 2021, Under Review.