

계층적 가우시안 프로세스 회귀모델

박선호⁰, 최승진

포항공과대학교 컴퓨터 공학과

titan@postech.ac.kr, seungjin@postech.ac.kr

Hierarchical Gaussian process model for regression

Park sunho⁰, Seungjin Choi

Department of Computer Science,

Pohang University of Science and Technology (POSTECH)

가우시안 프로세스 회귀분석 모델은 [1] (Gaussian process regression, GPR) 대표적인 베이시안 비모수 알고리즘으로 (Bayesian nonparametric methods) 비선형 함수 예측 문제, 즉 회귀분석에 (regression) 많이 적용된다. GPR은 높은 예측 성능을 보일 뿐 아니라, 예측결과에 대한 확률적 해석이 가능하고 비교적 단순한 행렬연산 만으로 구현이 가능하여 최근 많은 각광을 받고 있다. 그러나 GPR은 계산의 효율성 면에서 여타 다른 회귀분석 모델에 비해 복잡도가 크다. 만약 학습 데이터의 수가 N 일 경우, 필요한 메모리 공간이 $O(N^2)$ 이고 계산의 복잡도가 $O(N^3)$ 이다. 그러므로 만약 대규모의 데이터를 다룰 경우, GPR의 사용에 많은 어려움이 있다. 이의 해결을 위해 많은 근사화 방법이 연구되고 있다.

지금까지 GPR의 근사화 (approximation) 위한 여러 방법들이 제안되어 왔는데, 대부분의 근사 방법들이 sparse approximation [3, 또는 여기에 수록된 참고문헌들]에 속한다. 이는 전체 공분산 행렬에 낮은 계수 근사법 (low-rank approximation)을 적용하는 방법인데, 전체 N 개의 잠재변수를 그 보다 매우 작은 R 개의 inducing variables에 사영하여 (projection) 근사화 하는 방식을 취한다. 이 경우 GPR의 필요한 메모리 공간은 $O(RN)$, 계산상의 복잡도는 $O(R^2N)$ 으로 줄어든다. 이 sparse approximation 방법이 일정 성능이상으로 동작하기 위해서는 inducing variables의 선택이 매우 중요하다. 이들은 주어진 학습 데이터에서 뽑아내거나 [4], 최적화 방법을 통해서 학습 할 수 있다 [5]. 보통 최적화방법을 통해 inducing variables을 학습하는 경우가 높은 성능과 안정적인 동작을 보장한다. 그러나 높은 차원의 입력 데이터의 경우, 이 계산 과정이 쉽지 않으며 과적합 문제(over-fitting) 야기 할 수도 있다.

이에 본 논문은 GPR의 효율적인 계산을 위해 앞서의 sparse approximation 방법들과는 다른 방식으로 공분산 행렬을 근사화 한다. 특정 군집화방법으로 (clustering) 입력 데이터를 분할한 후, 같은 partition에 속하는 데이터들 간의 공분산은 정확히 계산하고 서로 다른 partition들에 속하는 데이터들 간의 공분산은 각 partition의 프로토 타입 벡터 (center or prototype) 간의 공분산으로 근사화하는 방법을 제안한다. 이는 2층 구조의 계층적 가우시안 프로세스 회귀분석 모델로 (Hierarchical Gaussian process regression model - HGPR) 표현 된다 (그림 1 참조). 즉, 상위 계층에 (upper layer) 정의된 함수는 전체 함수를 대략적으로 (coarse level) 모델링하는 함수로 각 partition들의 프로토타입들을 입력 인자로 가진다. 이에 비해, 하위 계층에 (lower layer) 정의된 함수는 전체 함수를 상세히 (fine level) 모델링하는 함수로 각 partition의 데이터 포인트를 입력 인자로 가진다. 이 두 계층의 함수는 각각 GP로 모델링 되는데, 하위 계층 GP prior의 평균함수를 상위계층의 함수 값으로 정의함으로써 계층적 관계가 성립되도록 한다. 상위 계층의 잠재변수를 적분하여 제거하면, 대각블록 (diagonal block)은 각 partition의 데이터 간의 공분산 값으로 정확히 계산되고 비 대각 블록(off-diagonal block)은 프로토타입들 간의 공분산 값으로 근사화되는 블록 공분산 행렬을 얻게된다.

이 블록 공분산 행렬은 원래 GPR 계산 문제를 보다 작은 복잡도를 가지는 여러 작은 문제들로 분리한다. 각 문제들은 각 partition에 속하는 입력 데이터의 수에 비례한 복잡도를 가진다. 만약 모든 partition들이 \tilde{N} 의 입력 데이터를 가지게 된다면 전체 계산상의 복잡도는 $O(\tilde{N}^2N)$, 필요 메모리 공간은 $O(\tilde{N}N)$ 으로 줄어들어 sparse approximation와 동일한 복잡도를 가진다. 또한 본 방법은 전체 문제를 각 partition별로 나누어 풀 다음 다시 통합하는 multi-strategy [6] 방법으로 이해 할 수 있다. 이는 각 partition들이 가지는 특성을 보존하면서 전체 계산을 근사화 할 수 있다. 그러므로 각 partition별로 다른 특성을 가지는 데이터를 분석하는데 유용하다. 또한 본 방법은 최적의 inducing variables을 선택 할 필요가 없으므로, 특히 고차원의 입력데이터를 다룰 때 유용하다.

본 논문에 수록된 실험에서는 본 제안 방법과 기존의 sparse approximation 방법들인 DTC[4], FITC

[5], PITS [3] PIC [2]와 비교하였다. 공정한 비교를 위해, 위 근사화 방법들의 경우 총 R 개의 inducing variables를 사용하도록 설정하였으며 구현을 위해 <http://www.cs.manchester.ac.uk/~neill/software.html>에서 제공하는 GP toolbox를 이용하였다. 그리고 근사화를 적용하지 않은 full-GP를 성능의 척도로 사용하였다. 모든 경우에 squared exponential kernel function을 공분산 함수로 사용하였다. 각 방법의 예측 성능을 평가하기 위해 normalized mean squared error (NMSE)를 사용하였다:

$$NMSE = \langle (y^* - \bar{f}(\mathbf{x}_*))^2 \rangle / \langle (y^* - \bar{y})^2 \rangle, \quad (18)$$

여기서 $\langle \cdot \rangle$ 은 테스트 데이터 셋에 대한 평균을 의미하고, \bar{y} 는 전체 테스트 데이터의 평균값이다.

수록된 실험들은 크게 두 관점에서 여러 방법들을 비교 평가한다. 실험 1은 본 제안방법과 동일하게 입력 데이터의 partition 정보를 이용하는 PIC [10]와 비교가 목적이며, 이는 inducing variables을 최적화 방법으로 구하는 방법들이 고차원의 입력데이터를 다루는데 한계가 있음을 확인하였다. 또한 실험 2에서는 몇 가지 실제 대규모 데이터에서의 각 근사화 방법들의 예측 성능을 비교하였다. 위 두 실험을 통해, 본 제안 방법이 다른 근사화 방법들과 비견되는 예측 성능을 보이며 각 partition별로 다른 특성을 가지는 데이터를 분석하는데 유용하다는 점을 확인하였다.

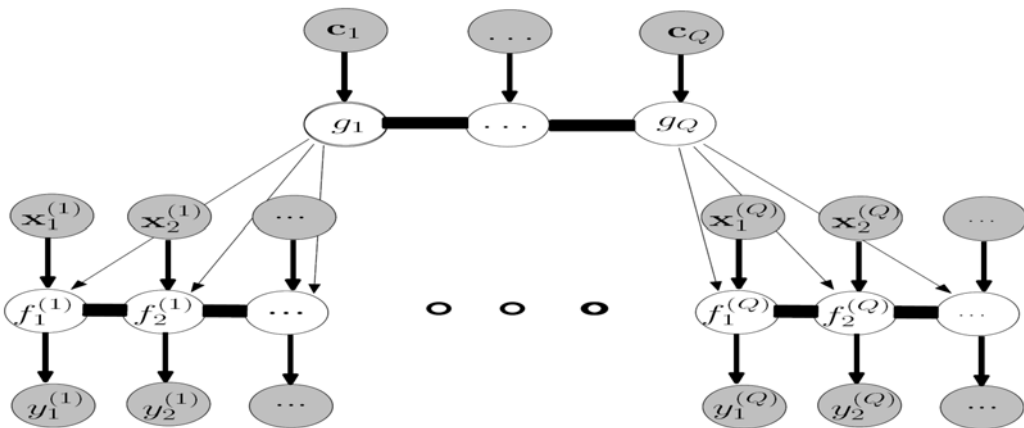


그림 1 : 계층적 가우시안 프로세스 회귀모델. 입력 데이터들이 Q 개의 partition으로 분할되고 \mathbf{c}_j 가 각 partition의 프로토 타입 벡터라 할 때, 함수 g 는 $\{\mathbf{c}_j\} \mapsto \mathbb{R}$ 로 정의 되었으며 함수 f 는 각 partition에 속한 데이터에서 정의된 함수이다, 즉 $X_j \mapsto \mathbb{R}$. 함수 $g(\mathbf{c}_j)$ 가 함수 $f_j(\mathbf{x})$ 의 평균함수로 쓰이는 2층 구조의 계층적 구조를 가진다.

참고 문헌

[1] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[2] N. Lawrence. Learning for larger datasets with the Gaussian process latent variable model. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, San Juan, Puerto Rico, 2007.

[3] J. Quiñero-Candela and C. E. Rasmussen. A unifying view of sparse approximate Gaussian process regression. *Journal of Machine Learning Research*, 6:1939-1959, 2005.

[4] M. Seeger, C. K. I. Williams, and N. D. Lawrence. Fast forward selection to speed up sparse Gaussian process regression. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2003.

[5] E. Snelson and Z. Ghahramani. Sparse Gaussian processes using pseudo-inputs. In *Advances in Neural Information Processing Systems (NIPS)*, volume 18. MIT Press, 2006.

[6] R. S. Michalski and G. Tecuci. *Machine Learning: A Multistrategy Approach*. Morgan Kaufmann, San Francisco, CA, 1994.