변분 베이지안 방법을 이용한 점집합의 오차제거

윤민철*+, Ioannis Ivrissimtzis+, 이승용*○
*포항공과대학교 컴퓨터공학과, +더람대학교 컴퓨터 공학과
*{mincheol, leesy}@postech.ac.kr, +ioannis.ivrissimtzis@dur.ac.uk

Point Set Denoising Using Variational Bayesian

Mincheol Yoon*+, Ioannis Ivrissimtzis+, Seungyoung Lee*○
*Dept. of Computer Science & Engineering, POSTECH
+ Dept. of Computer Science, Durham University.

제 1 절 서론

베이지안 통계는 곡면 복원, 모션 데이터 처리, 메쉬 처리 등에 적합한 방법으로 최근 많은 연구가 이루어지고 있다. 베이지안 방법은 확률식을 최대화 하여 문제를 해결하는 경우가 많다. 이때 확률식을 최대화하는 하나의 대표값을 구하는 점 추정 (point estimation)은 빠르고 간단한 문제 해결 방법이라고 할 수 있다. 그러나 이러한 점 추정은 하나의 값으로 확률 모델을 설명하고자 하므로 구한 값을 제외한 다른 값들을 설명해 주지 못한다는 근본적인 한계가 있다. 이러한 문제점을 해결하고자 본 논문에서는 [1]에서 기술한 변분 베이지안 방법을 사용한다. 이 방법에서는 구하고자 하는 파라미터를 하나의 확률 함수로 보고 실제 사후확률 (posterior)과 비슷한 확률 함수를 찾는다.

본 논문은 이러한 변분 베이지안 방법을 사용해서 점집합의 오차를 제거한다. 본 논문의 기여는 다음과 같다.

- 베이지안 확률식 최소화를 이용해서 높이 함수로 이루어진 곡면 근사 알고리즘을 개발
- 높이 함수를 점 집합의 오차제거에 응용할 수 있는 프레임워크 (framework) 제시
- 곡면 복원 문제에서 교차 검증 (cross validation) 없이 지역적 오버피팅 제어 (overfitting control)를 할 수 있는 방법 제시

제 2 절 곡면 근사 알고리즘

본 논문에서는 삼차원의 점 집합이 주어졌을 때, 곡면을 근사하는 알고리즘을 제안한다. 먼저 곡면 근사 알고리즘을 위한 모델은 B-스플라인 (B-spline) 높이함수로 한다. 따라서 스플라인의 기저함수(basis function)의 개수에 따라 모델이 변하게 된다. 그리고 스플라인의 계수와 그와 관련된 다른 파라미터(parameter)들이 알고리즘의 파라미터가 된다. 전체적인 높이함수를 구하는 알고리즘은 다음과 같다.

곡면 근사 알고리즘
입력: 3차원의 점집합
출력: 적합한 기저함수 개수, 곡면을 나타내는 파라미터

1. m행 m열 개의 기저함수를 가지는 스플라인 모델에 대해 곡면을 나타내는 파라미터를 구한다.
2. 구해진 파라미터를 이용해 비용함수를 구한다.
3. 기저함수의 개수를 m행 m열에서 $(m+1)$행 $(m+1)$열로 증가 한다.
4. 비용함수가 증가할 때까지 1부터 3까지를 반복한다.

본 논문에서는 다음과 같은 곡선 근사 모델을 사용한다.

$$D_i = \sum_{n=1}^{N} w_n f_{ni} + \nu_i \quad (1)$$

(단, D는 관측자료, w는 파라미터 벡터, x: 자료의 좌표 $f_{ni} = F_n(x_i)$, F_i는 B-spline 기저함수, ν_i는 오차)
이때 이 곡선의 가능도(likelyhood)는 다음과 같다.

$$P(D|w,\gamma,H) = \Pi_{i=1}^{l} g\left(D_i|\Sigma_{n=1}^{N} w_n f_{ni}\right) \quad (2)$$

(단, γ는 Gaussian분포의 역분산)
여기서 사전확률(prior)을 임의로 정하게 되면 사후 확률을 수식으로 나타낼수 있게 되고, 이 사후 확률과 Kullback-Leibler divergence가 최소인 확률분포를 구하게 된다. 이때 Kullback-Leibler divergence가 최소인 값을 구하기는 기술

적으로 어려워서 대신에 비용함수를 최소화한다. 이와같은 과정은 변분법을 이용해서 풀게되면 해석적인 식으로 구할 수 있다. 그리고 이식들을 반복적(iterative)으로 계산해서 원하는 사후확률분포를 구하게된다. 이렇게 구해진 사후확률분포를 이용하면 비용함수를 계산할 수 있다.

식 1에서 곡면의 경우는 곡선의 경우와 달리 2차원의 기저함수(basis function)를 갖는다. 따라서 2차원의 계수를 가지며, 입력 자료 x도 2차원이다. 본 곡면 근사 알고리즘에서서는 이러한 2차원 자료를 1차원의 식으로 바꾸어서 곡선 근사 알고리즘과 같은 방법으로 계산하게 된다. 예를 들면 기저함수의 개수가 m행 n열 일때 입력자료 F'_{ij}는 F_{m*i+j}의 형태로 바꾸어서 2차원 행렬을 1차원 배열의 형태로 바꾼다.

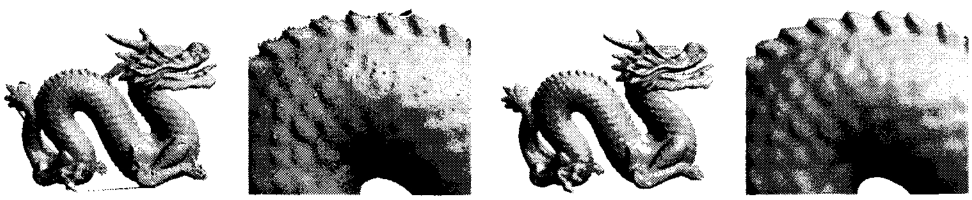
제 3 절 점 집합의 오차 제거

점집합 오차제거 알고리즘은 다음과 같다.

점집합 오차제거 알고리즘
입력: 오차를 포함한 점집합
출력: 오차가 제거된 점집합

1. 점 집합에서 하나의 점 p_i를 뽑는다.
2. 입력 점 집합으로 p_i의 주변점 N_i를 구하고 PCA를 이용해서 p_i의 노말 벡터를 구한다.
3. p_i의 노말 벡터를 이용해서 지역좌표계를 잡는다. (노말 방향이 z축이 되게한다.)
4. 지역 좌표계를 이용하여 제 3절에서 설명한 방법으로 p_i의 주변점들 n_k를 근사하는 곡면을 만든다.
5. 주변점 n_l들을 곡면에 대해 z축 방향으로 선을 내려 만나는 점n_l^p를 구한다.
6. n_l와 n_l^p사이에 벡터를 구해서 n_l에 저장한다.
7. 1번 과정부터 6번과정 까지를 일정한 기준이 될때까지 반복한다.
8. 모든 점에 저장된 거리 벡터를 종합하여 정점을 이동한다.

본 논문의 알고리즘의 결과를 시각적으로 보기 위해서 실제 스캐닝 데이터에 본 논문의 알고리즘을 적용했다. 입력으로 주어진 노이즈를 포함하는 실제 스캐닝 데이터를 Tight Cocone알고리즘으로 곡면 복원하고 입력 점집합을 본 논문의 알고리즘으로 노이즈를 제거한 후 Tight Cocone알고리즘으로 곡면 복원한 후에 결과를 비교하였다. 그림 1에서 볼수 있듯이 용의 비늘 모양은 거의 변하지 않고 노이즈가 효과적으로 제거되었음을 알 수 있다.



(a) (b)

그림 1: 용: (a) 원래 데이터를 Tight Cocone으로 복원 (b) 노이즈 제거후에 Tight Cocone으로 복원

제 4 절 결론

본 연구에서는 점 추정 방법이 가지는 한계를 인식하고 변분 베이지안 방법을 이용해 점 집합의 오차를 제거했다. 이를 위해서 변분 베이지안 방법을 이용해서 2차원의 정의역을 갖는 높이 함수로 3차원 데이터를 근사하는 방법을 개발 했다. 그리고 지역 좌표계와 높이 함수를 이용해서 실제 점 집합의 오차를 제거하는 방법도 개발했다.

본 연구의 방법은 교차 검증 없이 지역적 오버피팅이 가능하다는 장점이 있다. 본 방법은 이러한 사용자의 입력 없이 자동적으로 기저함수의 개수를 결정해 준다. 그리고 기저함수의 개수를 모델의 지역적인 특징에 따라 다르게해서 더 적은 오차를 가지게 한다.

감사의 글

Dragon 모델은 Stanford 그래픽스 연구실에서 구한 것입니다. 본 연구는 과학재단 국제협력연구(F01-2007-000-10208-0)의 지원을 받았습니다.

참고 문헌

[1] James W. Miskin. *Ensemble Learning for Independent Component Analysis.* PhD thesis, Department of Physics, University of Cambridge, 2001.