

# 머신러닝을 위한 베이지안 방법론: 군집분석을 중심으로

김용대, 정구환  
서울대학교

## 요약

본고에서는 베이지안 기계학습 방법론에 대해서 간략히 살펴본다. 특히, 복잡한 자료들 사이의 관계를 규명하는 것이 목적이며 비지도학습(unsupervised learning)의 한 분야인 군집분석에서 베이지안 방법론들이 어떻게 사용되어지는지를 설명한다. 군집의 수를 사전에 아는 경우에 사용되는 모수적 베이지안 방법을 간단하게 설명하고, 군집의 수까지 추론 할 수 있는 비모수 베이지안방법에 대해서 자세하게 다룬다.

## I. 서론

통계학에는 빈도론(Frequentism)과 베이지안(Bayesian)이라는 두 개의 큰 패러다임이 있다. 이 둘은 자료의 무작위성을 바라보는 시각이 다르다. 빈도론은 자료를 생성하는 하나의 고정된 확률분포가 존재한다고 생각하기 때문에 이 분포를 정확하게 찾는 것을 궁극적인 목적이다. 반면에 베이지안은 자료를 생성하는 확률분포가 고정된 것이 아닌 랜덤이라고 생각하여 자료를 관측한 뒤 모수의 사후분포(posterior distribution)을 구하고, 이를 기반으로 의사결정을 한다.

베이지안 방법은 Markov chain Monte Carlo(MCMC, [1]) 방법의 개발로 인하여 복잡한 구조를 갖는 모형에서 빈도론적 방법보다 쉽게 모수를 추론할 수 있는 장점이 있어서 현재 기계학습의 여러 분야에서 사용되어지고 있다. 특히, 자료들 사이의 복잡한 구조를 파악하는 것이 목표인 비지도학습(unsupervised learning)에서 유용하게 사용되어지고 있는데, 본 논문에서는 비지도학습의 한 분야인 군집분석에서 베이지안 방법이 어떻게 사용되어지고 있는가에 대해서 간단히 살펴보고자 한다.

본고에서는 혼합모형을 이용한 베이지안 군집분석에 대해 다룬다. 혼합모형이란 관측치  $x$ 의 분포가  $K$ 개의 분포의 혼합으로 주어진 모형을 지칭한다.

$$\sum_{k=1}^K \pi_k p(\cdot | \phi_k)$$

이 경우  $K$ 는 군집의 수로, 각 분포는 각 군집에서의 관측치의 분포로 해석이 된다. 군집의 수  $K$ 가 알려진 경우에는 모수의 수가 유한개인 모수적 베이지안 모형이 되며 쉽게 사후분포를 구할 수 있다. 하지만 대부분의 경우 군집의 수를 알 수가 없으며, 경우에 따라서는 모수의 수가 무한대까지도 가능하다. 이러한 무한차원 모수의 효율적인 베이지안 추론을 위하여 비모수 베이지안 방법이 널리 사용되고 있다. 특히, 비모수 베이지안 방법을 사용하면 군집의 수  $K$ 에 대한 추론도 가능하여 새로운 군집의 생성이나 기존 군집의 소멸 등에 대해서도 추론이 가능하다는 매우 중요한 장점이 있다. 본고에서는 디리클레 과정을 이용하여 혼합모형의 추론을 하는 비모수 베이지안 방법에 대해서 살펴보고자 한다.

## II. 디리클레 과정 혼합모형

### 1. 디리클레 분포

디리클레 과정을 이해하기 위해서는 디리클레 분포를 이해하는 것이 필수적이다. 디리클레 분포는 베타 분포를  $K$ 차원으로 확장한 것으로 양의 상수  $\alpha_1, \dots, \alpha_K$ 를 모수로 갖는 디리클레 분포는  $D(\alpha_1, \dots, \alpha_K)$ 로 표기하고 확률 밀도 함수는 다음과 같이 정의된다.  $x_1, \dots, x_K$ 가 모두 양의 실수이고,  $\sum_{i=1}^K x_i = 1$ 을 만족할 때

$$f(x_1, \dots, x_K; \alpha_1, \dots, \alpha_K) = \frac{\Gamma(\sum_{i=1}^K \alpha_i)}{\prod_{i=1}^K \Gamma(\alpha_i)} \prod_{i=1}^K x_i^{\alpha_i - 1}$$

의 값을 갖고, 그 외의 경우는 0의 값을 가진다.

### 2. 디리클레 과정

$H$ 가  $R^p$ 에서 정의된 분포함수이고,  $\alpha$ 가 양의 상수라 하자.  $R^p$

에서 정의된 랜덤확률분포  $P$ 가  $R^p$ 의 모든 가능한 유한개의 분할  $A_1, \dots, A_k$ 에 대해서  $(P(A_1), \dots, P(A_k)) \sim D(\alpha H(A_1), \dots, \alpha H(A_k))$  조건을 만족할 때 모수가  $(\alpha, H)$ 인 디리클레 과정을 따른다고 하고  $P \sim DP(\alpha, H)$ 로 표기한다. 즉, 어떠한 유한차원 주변분포도 디리클레 분포가 되는 확률과정을 디리클레 과정이라고 한다. 디리클레 과정은 [2]에 의해서 처음 제안된 이후로 비모수 베이지안 방법에서 유용한 사전분포로 널리 사용되고 있다.

디리클레 과정을 따르는 랜덤확률분포의 중요한 성질 중 하나는 랜덤확률분포가 이산형이라는 것이며, 이는 막대기 자르기 과정(stick-breaking process, [3])으로 쉽게 설명할 수 있다.  $\pi = (\pi_k)_{k=1}^{\infty} \sim GEM(\alpha)$ 이고  $\phi_1, \phi_2, \dots \sim H$ 일 때,  $P = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$ 는  $DP(\alpha, H)$ 를 따른다는 것이 [3]에서 규명되었다. 여기서  $\delta_{\phi}$ 는  $\phi$ 에서 확률이 1인 확률분포이고,  $GEM$ 은 Griffiths, Engen, McCloskey의 첫 글자들을 딴 것으로  $\pi = (\pi_k)_{k=1}^{\infty} \sim GEM(\alpha)$ 은  $\pi_1', \pi_2', \dots \sim Beta(1, \alpha)$ 일 때  $\pi_1 = \pi_1'$ 이고,  $k \geq 2$ 에 대하여  $\pi_k = \pi_k' \prod_{l=1}^{k-1} (1 - \pi_l')$ 를 의미한다.

### 3. 디리클레 과정 혼합모형

디리클레 과정 혼합모형은 비모수 베이지안 혼합모형의 한 종류로 디리클레 과정을 혼합분포의 사전분포로 사용한 것이다. 관측치  $x_i (i = 1, \dots, n)$ 는 모수가  $\theta_i$ 인 확률분포  $p(\cdot | \theta_i)$ 에서 추출되고,  $\theta_1, \dots, \theta_n$ 은 디리클레 과정을 따르는 랜덤확률측도  $P$ 에서 추출된다고 가정하는 것이다. 이를 수식으로 나타내면 다음과 같다.

$$\begin{aligned} P &\sim DP(\alpha, H), \\ \theta_i | P &\sim P, \\ x_i | \theta_i &\sim p(\cdot | \theta_i), \text{ for } i = 1, \dots, n. \end{aligned}$$

막대기 자르기 과정으로부터  $P = \sum_{k=1}^{\infty} \pi_k \delta_{\phi_k}$ 가 이산분포이므로  $\theta_i$ 는  $\phi_1, \phi_2, \dots$  중 오직 하나와 일치한다.  $\theta_i = \phi_{c_i}$ 를 만족하도록  $c_i$ 를 정의하면 디리클레 과정 혼합 모형을 다음과 같이 표현할 수 있다.

$$\begin{aligned} \pi &\sim GEM(\alpha), \\ \phi_k &\sim H \text{ for } k = 1, 2, \dots, \\ c_i &\sim H_{\pi}, \\ x_i | c_i, \phi &\sim p(\cdot | \phi_{c_i}) \text{ for } i = 1, \dots, n. \end{aligned}$$

여기서  $\phi = (\phi_1, \phi_2, \dots)$ 이고,  $H_{\pi}$ 는 자연수  $k$ 에 대해서  $H_{\pi}(\{k\}) = \pi_k$ 인 확률분포이다. 여기서  $c_i$ 는  $i$ 번째 관측치의 군집 라벨이 되며, 집합  $\{c_1, \dots, c_n\}$ 의 원소의 수가 군집의 수가 된다.

### 4. 디리클레 과정 혼합모형의 추론

베이지안 방법의 주된 관심사는 사후분포인  $p(\theta_1, \dots,$

$\theta_n | x_1, \dots, x_n)$ 를 구하는 것이다. 하지만 정확한 사후분포를 찾는 것은 어렵기 때문에 사후분포를 근사하는 방법을 사용한다. 사후분포를 근사하는 대표적인 방법으로는 깁스 샘플러(Gibbs sampler) 방법과 변분 베이즈(variational Bayes) 방법이 있다.

#### 가. 깁스 샘플러

첫 번째 깁스 샘플러 방법([4]의 Algorithm 1)은  $\theta_i$ 를 제외한 다른 모수  $\theta_{-i} = \{\theta_1, \dots, \theta_n\} \setminus \theta_i$ 와 자료  $x = \{x_1, \dots, x_n\}$ 가 주어졌을 때의 조건부분포인  $p(\theta_i | \theta_{-i}, x)$ 에서 표본을 추출하여  $\theta_i$ 를 갱신하는 것을 모든 모수에 대해서 번갈아가며 여러 번 반복하는 방법이다. 디리클레 과정 혼합모형의 조건부분포는 다음과 같다.

$$p(\theta_i | \theta_{-i}, x) \propto \sum_{j \neq i} p(x_i | \theta_j) \delta_{\theta_j} + \alpha \int p(x_i | \theta) H(d\theta)$$

충분한 반복횟수 후에 얻어진  $\{\theta_1, \dots, \theta_n\}$ 을  $p(\theta_1, \dots, \theta_n | x)$ 의 표본으로 간주하여  $p(\theta_1, \dots, \theta_n | x)$ 를 추정한다.

이 방법은  $\theta_i$ 간의 상관관계가 크기 때문에 사후분포의 수렴이 느린 단점이 있다.  $\theta = (\theta_1, \dots, \theta_n)$ 는  $c = (c_1, \dots, c_n)$ 와  $\phi$ 로 결정되기 때문에  $c, \phi$ 에 대한 깁스 샘플러 방법([4]의 Algorithm 2)도 가능하다.

$$\begin{aligned} p(c_i = c | c_{-i}, \phi, x) &\propto \begin{cases} n_{-i,c} p(x_i | \phi_c) & \text{if } c = c_j \text{ for some } j \neq i \\ \alpha \int p(x_i | \phi) H(d\phi) & \text{o.w.} \end{cases} \\ \phi_c | c, x &\propto H(d\phi) \prod_{c_i = c} p(x_i | \phi) \end{aligned} \quad (1)$$

여기서  $n_{-i,c} = \#\{j : j \neq i \text{ and } c_j = c\}$ 이다.  $\phi$ 를 적분하여 제거하는 것을 통해 깁스 샘플러 방법의 수렴 속도를 향상시킬 수 있다([4]의 Algorithm 3).

$$\begin{aligned} p(c_i = c | c_{-i}, x) &\propto \begin{cases} n_{-i,c} \int p(x_i | \phi_c) H_{-i,c}(d\phi) & \text{if } c = c_j \text{ for some } j \neq i \\ \alpha \int p(x_i | \phi) H(d\phi) & \text{o.w.} \end{cases} \end{aligned}$$

여기서  $H_{-i,c}(\cdot)$ 는  $H_{-i,c}(d\phi) \propto H(d\phi) \prod_{c_j = c, j \neq i} p(x_j | \phi)$ 를 만족하는 확률분포이다.  $c_i$ 들을 번갈아가며 반복적으로 추출함으로써  $c$ 를 얻었다면  $\phi$ 는 식 (1)을 통해 추출할 수 있다. 따라서  $c, \phi$ 로부터  $\theta$ 를 얻을 수 있다. 이 방법은 붕괴 깁스 샘플러(Collapsed Gibbs sampler)라 불리고 수렴 속도가 가장 빠르다.

여러 가지 깁스 샘플러 방법 중에서 어떤 방법을 사용할 지를 결정하기 위해서는 데이터 종류, 크기 뿐 아니라 병렬처리 가능성 등 다양한 요소들을 고려하여야 한다.

#### 나. 변분 베이즈 (Variational Bayes, [5])

$q(\theta | v)$ 를  $v$ 가 모수인  $\theta$ 에 대한 확률분포라 할 때, 변분 베이

즈 방법은  $p(\theta|x)$ 와  $q(\theta|v)$ 가 가장 가까워지도록 하는  $\hat{v}$ 을 찾아서  $p(\theta|x)$ 를  $q(\theta|\hat{v})$ 으로 근사하는 것이다.

변분 베이지 방법에서  $p(\theta|x)$ 와  $q(\theta|\hat{v})$ 의 거리의 측도로는 콜백-라이블러 발산을 사용한다.

$$D_{KL}(q(\cdot|v)||p(\cdot|x)) = E_{q(\theta|v)}\{\log q(\theta|v) - \log p(\theta|x)\}$$

변분 베이지 방법에서는  $\hat{v}$ 를 구하기 쉽도록 변분 분포를 선택하는 것이 중요하다. 보통 변수간의 독립성을 가정한 분포를 변분 분포로 사용한다.

디리클레 과정 혼합모형에서는 막대 자르기 과정을  $P = \sum_{k=1}^K \pi_k \delta_{\phi_k}$ 로 근사하여  $\pi', \phi, c$ 에 대한 변분 분포를 다음과 같이 정의한다.

$$q(\pi', \phi, c|v) = \prod_{k=1}^{K-1} q_1(\pi_k | \gamma_k) \prod_{k=1}^K q_2(\phi_k | \tau_k) \prod_{i=1}^n q_3(c_i | \eta_i)$$

여기서  $q_1$ 은 베타 분포,  $q_2$ 는  $H$ 와 같은 지수족의 분포,  $q_3$ 는 다항분포이다. 변분 베이지 방법은 모수  $v = (\gamma, \tau, \eta)$ 를 아래 식들을 반복하여 추정하는 것이다.

$$\begin{aligned} \gamma_{k,1} &= 1 + \sum_{i=1}^n \eta_{i,k} \\ \gamma_{k,2} &= \alpha + \sum_{i=1}^n \sum_{t=k+1}^K \eta_{i,t} \\ \tau_{k,1} &= \lambda_1 + \sum_{i=1}^n \eta_{i,k} x_i \\ \tau_{k,2} &= \lambda_2 + \sum_{i=1}^n \eta_{i,k} \\ \eta_{i,k} &\propto \exp(S_k) \end{aligned}$$

여기서  $\lambda = (\lambda_1, \lambda_2)$ 는  $H$ 의 모수이고,  $S_k$ 는 [5]를 참고하라.

변분 베이지 방법은 계산이 빠른 이점이 있지만 실제 사후분포는 변분분포에서 많이 벗어나기 때문에 아무리 변분 모수  $v$ 를 잘 추정하더라도 사후분포를 정확하게 근사할 수 없다는 치명적인 단점이 있다.

### III. 디리클레 과정 혼합모형의 응용

이 장에서는 머신러닝 분야에서 사용되는 디리클레 과정 혼합모형의 다양한 응용에 대해 살펴본다.

#### 1. 군집 분석(Clustering)

디리클레 과정의 불연속성에 의해  $\theta_i$ 의 값이 같은 자료들은 하나의 군집을 형성한다. 따라서 군집의 개수를 미리 지정하지 않아도 자료를 설명하는데 필요한 만큼의 군집 개수를 스스로

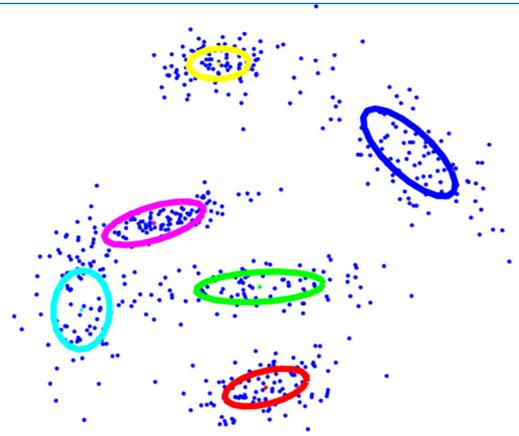


그림 1. 디리클레 과정 혼합모형을 이용한 군집분석 (이 그림은 Zoubin Ghahramani의 UAI 2005 tutorial 슬라이드에서 가져왔다.)

찾아서 군집을 만들어 준다. <그림 1>은 디리클레 과정 혼합모형을 이용하여 군집분석을 수행한 것의 결과를 나타낸다.

#### 2. 스파이크 분리(Spike sorting)

여러 뉴런에서 신호를 동시에 얻었을 때 이들을 각각의 뉴런별 신호로 분리해 내는 문제를 스파이크 분리라 한다. [6]에서는 디리클레 과정 혼합모형을 요인 분석(Factor analysis)에 사용하여 스파이크 분리에 응용하였다. 사용한 모형은 다음과 같다.

$$\begin{aligned} \sigma_d^2 &\sim IG(\beta, \beta w) \text{ for } d = 1, \dots, D, \\ P &\sim DP(\alpha, H), \\ \mu_i, \Lambda_i | P &\sim P, \\ x_i | \mu_i, \Lambda_i, \Psi &\sim N(\cdot | \mu_i, \Lambda_i \Lambda_i^T + \Psi) \text{ for } i = 1, \dots, n. \end{aligned}$$

여기서  $x_i$ 는 하나의 신호를 나타내고,  $D$ 는 신호의 차원,  $\Psi$ 는  $d$ 번째 대각원소가  $\sigma_d^2$ 인 대각행렬,  $H$ 는  $(\mu, \Lambda)$ 에 대한 사전분포로  $\mu$ 는 NIW(normal-inverse Wishart) 분포를 따르고  $\Lambda$ 는  $N(0, I/v)$ 를 따르는데 이 때  $v$ 는  $G(1/2, 1/2)$ 를 따른다.

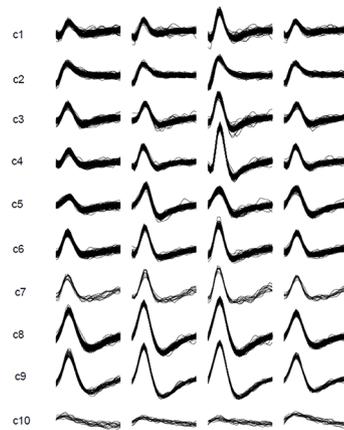


그림 2. 디리클레 과정 혼합모형을 이용하여 분리된 스파이크 파형 (이 그림은 [6]에서 가져왔다.)



그림 3. 하플로타입과 유전자형

G는 감마 분포를, IG는 역감마 분포를 의미한다.

<그림 2>는 10개의 군집에 대해서 반복횟수가 증가함에 따라 해당 군집에 속하는 스파이크들을 그림으로 나타낸 것이다.

### 3. 하플로타입(Haplotype) 추정

하플로타입은 부모 중 한명으로부터 받은 유전자 서열을 나타내고, 유전자형(genotype)은 부모로부터 각각 받은 두 하플로타입들의 조합이다.

[7]에서는 유전자형이 주어졌을 때 하플로타입을 추정하는 문제에 디리클레 과정 혼합모형을 사용하였다.

$$\begin{aligned}
 P &\sim DP(\alpha, H) \\
 \text{Ancestral templates : } \theta_1^{(p)}, \dots, \theta_n^{(p)} | P &\sim P \\
 \theta_1^{(m)}, \dots, \theta_n^{(m)} | P &\sim P \\
 \text{Paternal haplotype : } H_i^{(p)} | \theta_i^{(p)} &\sim f(\cdot | \theta_i^{(p)}) \\
 \text{Maternal haplotype : } H_i^{(m)} | \theta_i^{(m)} &\sim f(\cdot | \theta_i^{(m)}) \\
 \text{Genotype : } x_i | H_i^{(p)}, H_i^{(m)} &\sim g(\cdot | H_i^{(p)}, H_i^{(m)})
 \end{aligned}$$

### 4. 계층적 디리클레 과정(Hierarchical Dirichlet processes, [8])

계층적 디리클레 과정은 그룹화된 자료를 다루기 위해 제안되었다. j번째 그룹의 랜덤확률측도를 P<sub>j</sub>라 할 때 계층적 디리클레 과정을 다음과 같이 정의한다.

$$\begin{aligned}
 P &\sim DP(\gamma, H), \\
 P_j &\sim DP(\alpha, P) \text{ for } j = 1, \dots, J.
 \end{aligned}$$

P는 디리클레 과정을 따르기 때문에 이산분포이다. 따라서 P가 기본분포인 디리클레 과정을 따르는 P<sub>j</sub>들은 모두 P와 받침(support)이 같다. 따라서 그룹간의 의존성을 반영한 군집분석이 가능하다.

계층적 디리클레 과정을 사전분포로 사용하는 계층적 디리클레 과정 혼합모형은 다양한 분야에서 응용되고 있다.

#### 가. 토픽모형 (Topic model, [8])

문서 자료는 단어들이 문서 속에 포함되어 있는 전형적인 그룹화된 자료이다. 문서 자료에서 주제를 찾아내고자 하는 토픽모형은 계층적 디리클레 과정 혼합모형이 가장 많이 사용되고 있는 분야이다.

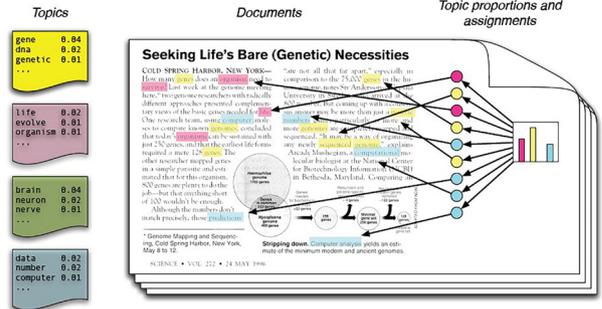


그림 4. 토픽모형 (이 그림은 [9]에서 가져왔다.)

$$\begin{aligned}
 P &\sim DP(\gamma, H), \\
 P_j &\sim DP(\alpha, P) \text{ for } j = 1, \dots, J, \\
 \theta_{ji} | P_j &\sim P_j \\
 x_{ji} | \theta_{ji} &\sim p(\cdot | \theta_{ji}) \text{ for } i = 1, \dots, n_j.
 \end{aligned}$$

여기서 J는 문서의 수, n<sub>j</sub>는 j번째 문서에 있는 단어의 개수, x<sub>ji</sub>는 j번째 문서의 i번째 단어를 의미한다.

#### 나. 시각 장면 분석(Visual scene analysis, [10])

이미지 자료를 가지고 이미지 속의 물체들의 위치를 찾아내는 것을 시각 장면 분석이라 한다. 이미지에 포함된 물체들의 종류에 따라 이미지 자료를 그룹화 할 수 있다. 이미지는 여러 부분들의 조합으로 볼 수 있고, 이러한 부분들은 많은 이미지에서 공통으로 나타나기 때문에 계층적 디리클레 과정 혼합모형을 통해 이미지 자료를 분석할 수 있다. 공통된 부분이라도 이미지 자료마다 다른 위치에 나타나기 때문에 [10]에서는 이를 반영할 수 있도록 계층적 디리클레 과정을 발전시킨 변형 디리클레 과정(transformed Dirichlet processes)을 이용해 분석하였다.

<그림 5>는 시각 장면 분석을 통해 네 가지 이미지를 분할(segmentation)한 결과이다. 자동차, 빌딩, 길을 각각 빨간색, 초록색, 파란색으로 나타내었다.

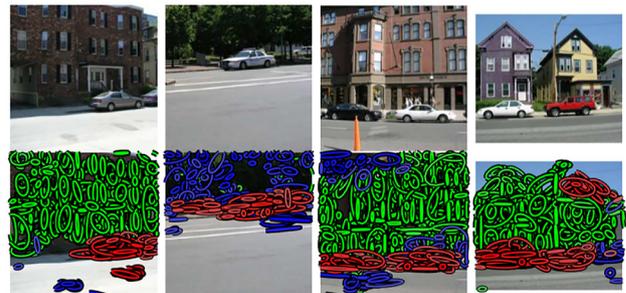


그림 5. 변형 디리클레 과정을 이용한 시각 장면 분석 결과 (이 그림은 [10]에서 가져왔다.)

## IV. 결론

본고에서는 군집분석을 위한 비모수 베이지안 방법 및 다양한 기계학습에서의 응용을 살펴보았다. 베이지안 방법의 가장 큰 장점은 군집의 수에 대한 추론이 가능하다는 것이며, 이는 여러 가지 기계학습 분야에서 매우 유용하게 사용될 수 있다.

본고에서 살펴보지 못한 주제로는 디리클레 과정을 대체할 수 있는 다양한 비모수 사전분포들(예: Pitman-Yor process, [11]), 거대자료에서 효율적인 사후분포 계산을 위한 병렬알고리즘([12]), 시간에 따라 변하는 모형에 대한 추론([13]) 등이 있다.

## 참고 문헌

[1] Gelfand, Alan E., and Adrian FM Smith. "Sampling-based approaches to calculating marginal densities." *Journal of the American statistical association* 85,410 (1990): 398-409.

[2] Ferguson, Thomas S. "A Bayesian analysis of some nonparametric problems." *The annals of statistics* (1973): 209-230.

[3] Sethuraman, Jayaram. "A constructive definition of Dirichlet priors." *Statistica sinica* (1994): 639-650.

[4] Escobar, Michael D., and Mike West. "Bayesian density estimation and inference using mixtures." *Journal of the american statistical association* 90,430 (1995): 577-588.

[5] Blei, David M., and Michael I. Jordan. "Variational inference for Dirichlet process mixtures." *Bayesian analysis* 1,1 (2006): 121-144.

[6] Görür, Dilan. "Nonparametric Bayesian Discrete Latent Variable Models for Unsupervised Learning." (2007).

[7] Xing, Eric P., Michael I. Jordan, and Roded Sharan. "Bayesian haplotype inference via the Dirichlet process." *Journal of Computational Biology* 14,3 (2007): 267-284.

[8] Teh, Yee Whye, et al. "Hierarchical dirichlet processes." *Journal of the american statistical association* (2012).

[9] Blei, David M., Andrew Y. Ng, and Michael I. Jordan.

"Latent dirichlet allocation." *Journal of machine Learning research* 3,Jan (2003): 993-1022.

[10] Sudderth, Erik B., et al. "Describing visual scenes using transformed objects and parts." *International Journal of Computer Vision* 77,1-3 (2008): 291-330.

[11] Pitman, Jim, and Marc Yor. "The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator." *The Annals of Probability* (1997): 855-900.

[12] Williamson, Sinead, Avinava Dubey, and Eric P. King. "Parallel Markov Chain Monte Carlo for Nonparametric Mixture Models." *ICML* (1), 2013.

[13] Blei, David M., and John D. Lafferty. "Dynamic topic models." *Proceedings of the 23rd international conference on Machine learning*. ACM, 2006.

## 약 력



김 용 대

1991년 서울대학교 학사  
 1993년 서울대학교 석사  
 1997년 Ohio State University 박사  
 1997년~1999년 NIH 연구원  
 1999년~2001년 한국외국어대학교 조교수  
 2001년~2004년 이화여자대학교 조교수  
 2004년~현재 서울대학교 교수

관심분야: Survival Analysis, Bayesian Statistics, Machine Learning, Model Selection, Deep Learning



정 구 환

2011년 포항공과대학교 학사  
 2012년~현재 서울대학교 석박사 통합과정 재학  
 관심분야: Bayesian Statistics, Machine Learning