

# 기계학습과 정보검색

경북대학교 ■ 박성배\* · 태윤식

## 1. 서론

정보검색의 목적은 데이터베이스에 저장된 대량의 정보에서 사용자가 원하는 정보와 비슷한 정보를 추출하여 제공하는 것이다. 많은 경우에 여기서 이야기 하는 사용자는 특정한 사용자를 지칭하는 것이 아니라 여러 다양한 사용자 집단을 뜻한다. 또한, 일반적으로 정보가 반드시 텍스트로 되어 있을 필요도 없다. 즉, 정보는 텍스트 데이터일 수도 있고, 오디오나 비디오 데이터일 수도 있다. 하지만, 가장 일반적인 정보의 형태가 텍스트이므로, 본고에서는 논의를 텍스트 정보검색에만 국한하기로 한다.

정보검색은 사용자가 원하는 정보를 적절히 제공하기 위해서 텍스트 분석, 텍스트 저장, 저장된 데이터의 검색 등을 다룬다. 초창기의 정보검색에서부터 상당 부분을 입력 대신 자동화하는 방법이 시도되었다[1]. 하지만, 이 당시에는 텍스트 분석과 같이 고도의 지적 능력을 요구하는 분야에서는 기계화하는 것이 불가능할 것으로 믿어졌다[2]. 특히 정보검색은 도서관학으로부터 시작되었기 때문에, 대부분의 노력이 색인(index) 구조를 구축하는데, 그것도 수동으로 구축하는데 집중되었다. 책이나 논문에 대한 색인을 만드는 것이 가장 중요한 작업이었으며, 이는 대부분은 사람의 힘에 의해 구축되었다.

하지만, 1960년대 중반 이후 Salton 등이 벡터 공간 모델(vector space model)을 제안한[3] 이후부터 정보검색의 모든 부분이 자동화되기 시작했다. 즉, 벡터 공간 모델에 대한 수학 이론들이 개발되면서 대량의 말뭉치로부터 검색 프로세스를 시스템이 학습하는 방법이 널리 쓰이게 되었다. 정보검색의 많은 문제들이 수학적으로는 표기변환(notation transformation)으로 표현될 수 있다. 예를 들면, 전통적인 정보검색 문제는 사용자의 질의어라는 표기(notation)를 이와 관련된 문서라는 표기(notation)로 변화하는 것으로 해석할 수

있다.

정보검색 문제들이 표기변환으로 표현될 수 있다는 것은 이 문제들이 기계학습(machine learning)이 이 문제들에 적용될 수 있음을 의미한다. 기계학습은 정보검색과 같이 비정형 문제(ill-posed problem)를 푸는데 성공적으로 적용되어 왔으며, 대부분의 기계학습 방법들이 입력 데이터와 출력 데이터 사이의 변환 함수를 추정하는 것으로 해석될 수 있다. 즉, 정보검색의 표기변환의 대부분은 실제로는 기계학습의 분류문제(classification task)로 표현될 수 있으며, 이에 따라 1990년 중반이후에는 비록 외부로 들어나지는 않을지라도 거의 모든 정보검색 시스템이 기계학습 방법을 사용하고 있다[5].

정보검색의 표기변환을 성공적으로 수행하기 위해서 대부분의 정보검색 시스템은 어휘 분석 결과나 품사[6]와 같은 추가적인 약간의 정보를 사용하고 있다. 이와 같은 부가정보들은 모두 도구화되어 있는 시스템으로부터 쉽고 정확한 형태로 획득할 수 있다. 정보검색과 관련된 여러 가지 정보 및 학습 방법들이 도구화되면서, 최근에는 평가를 기반으로 한 빠른 시스템의 구축 및 검증이 가능하게 되었다. 이에 따라 정보검색의 범위가 단순한 질의어 검색에서 벗어나 교차언어 정보검색, 웹 검색, 정보 추출 등으로 확대되었다[7]. 정보검색에 있어서 기계학습 기법 사용의 확대는 문서의 내용을 충실히 파악하는 심화분석(deep analysis)을 어렵게 하지만, 지식의 자동 추출을 통하여 새로운 환경에 적응 가능한 시스템의 개발을 용이하게 하여 정보검색 시스템의 외연이 확대되었다.

본고는 기계학습이 정보검색 분야에서 최근에 어떻게 사용되는지를 알아보고, 이에 따라 최근 널리 쓰이는 기계학습 방법을 소개하는 것을 목적으로 한다. 2절에서는 최근 4년간의 ACM SIGIR을 분석하여 최근에 적용되고 있는 기계학습 기법을 파악한다. 3절에서는 정보검색에서 최근에 널리 쓰이는 기계학습 기법을 설명하고, 마지막으로 4절에서는 결론을 맺는다.

\* 종신회원

## 2. 최근 경향

정보검색에서는 이미 1990년 중반 이후부터 기계학습 방법의 사용이 일반화되었다. 본고에서는 2003년부터 2006년까지의 4년간의 ACM SIGIR 학회를 중심으로 하여 기계학습과 정보검색과의 최근 관계를 알아본다. 4년간의 ACM SIGIR 학회에서 발표되었던 논문들의 제목과 초록을 분석하였다.

2003년부터 2006년까지 ACM SIGIR 학회에 발표된 논문의 수는 표 1과 같다. 전체적으로 250편이 발표되었으며, 그 중 204편이 기계학습과 관련된 논문들이다. 따라서 ACM SIGIR 학회에 지난 4년간 발표된 논문들 중 약 82%가 기계학습 방법을 사용하였다. 전체 논문 수가 2003년 46편에서 2006년 74편으로 늘어나면서 기계학습과 관련된 논문의 수도 40편에서 58편으로 늘어나 기계학습과 관련된 논문의 비율은 거의 비슷하게 유지되었다.

표 1 지난 4년간의 ACM SIGIR 학회 논문 수

년도	전체 논문편수	기계학습 관련 논문 수	기계학습과 무관한 논문 편수	기계학습 관련 논문의 백분율
2006	74	58	16	78.4
2005	71	57	14	80.3
2004	59	49	10	17.0
2003	46	40	6	87.0
합계	250	204	46	81.6

그림 1은 ACM SIGIR 학회 논문에서 사용된 주요한 기계학습 방법들과 그 방법들의 사용 횟수를 보인다. 매우 다양한 기계학습 방법들이 정보검색 문제를 해결하기 위해 사용되었으며, 이 중 베이시안 방법론, EM 알고리즘, 그래피컬 모델(graphical model), 정보이론(information theory), SVM(support vector machine)을 포함한 커널 방법론(kernel method)이 가장 자주 쓰이는 방법들이다.

그림 2에서 그림 5까지는 위에서 열거한 주요 방법 중, 커널 방법론(그림 2), 그래피컬 모델(그림 3), 베이시안 방법론(그림 4), EM 알고리즘(그림 5)의 논문 수 변화를 보인 것이다. SVM을 포함한 커널 방법론의 논문 수가 거의 변화가 없는데 비해, 그래피컬 모델과 베이시안 방법론의 논문 수는 점차적으로 늘어나는 추세이다. EM 알고리즘 계열의 방법들도 조금씩 늘어나고 있다.

이에 비해, 기억기반 학습(memory-based learning)이나 결정트리(decision tree) 계열의 방법론들은 점차적으로 그 수가 줄어들어 2004년과 2005년에 1편씩 나타난 것을 제외하고는 사용되지 않았다. 이는 웹을 중심으로 하여 대량의 학습 데이터가 이용가능하게 됨에 따라 이로부터 비교적 안정적인 확률값을 얻는 것이 이제 가능해졌기 때문인 것으로 파악된다. 즉, 최소한 정보검색 분야에서는 확률 모델을 사용하거나 확률을 추정하여 모델을 학습하는 것이 최근에 가장 널리 쓰이는 방법이다.

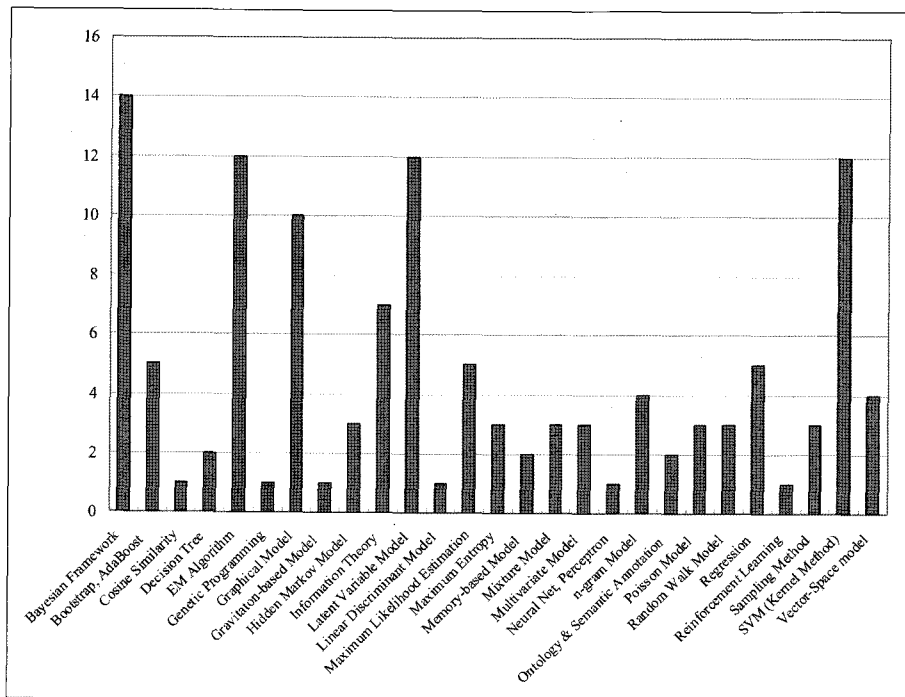


그림 1 ACM SIGIR 학회 논문에 적용된 여러 가지 기계학습 방법들의 사용 횟수

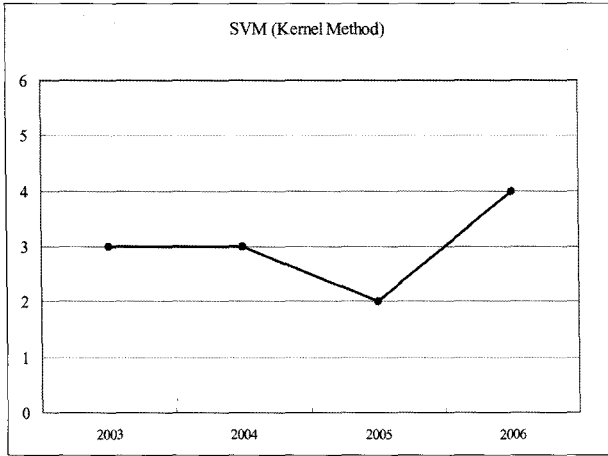


그림 2 커널 방법론의 논문 수 변화

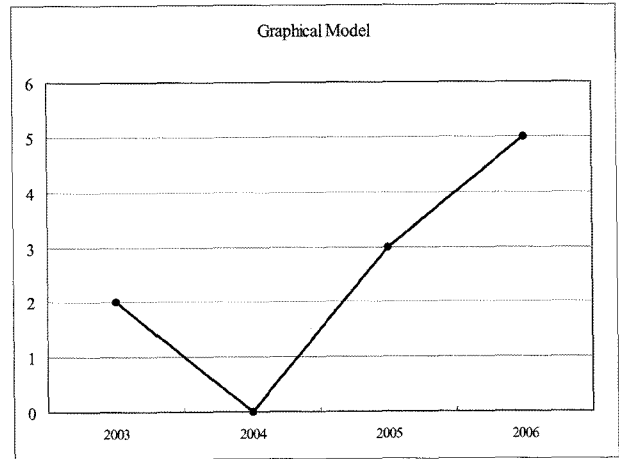


그림 3 그래피컬 모델의 논문 수 변화

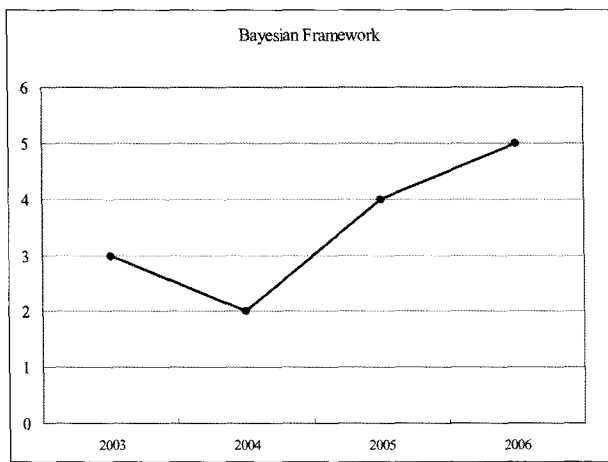


그림 4 베이지안 방법론의 논문 수 변화

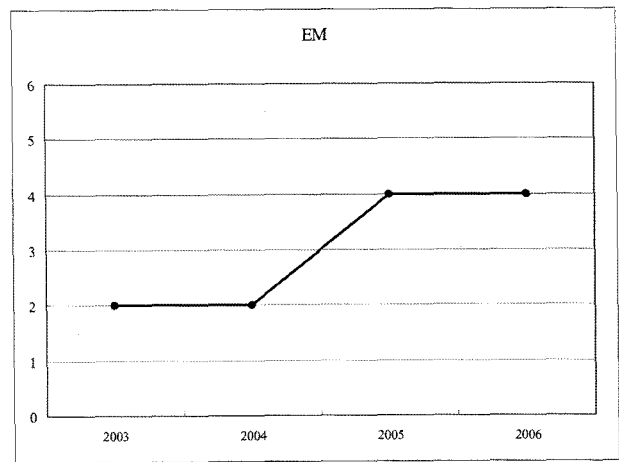


그림 5 EM 알고리즘 계열의 논문 수 변화

### 3. 정보검색을 위한 기계학습 방법

#### 3.1 Support Vector Machines

Support Vector Machine(SVM)은 90년대 중반에 등장하여 현재까지 널리 쓰여 지고 있으며 분류 문제에 가장 좋은 성능을 보여주고 있는 기계학습 방법 중 하나이다[8]. SVM은 기본적으로 이진 분류기이다. 비선형 데이터를 분류하기 위해, SVM은 데이터들을 커널 함수를 이용하여 고차원 공간에 사상시킨 후 support vector들로 이루어진 초평면(hyperplane)을 이용해 분류한다. 그림 6에 그 예를 보인다. 주어진 2차원 공간에서 선형으로는 분류 불가능한 데이터를 커널 함수  $\Phi$ 를 통해 3차원 공간으로 사상시키면 선형으로 분류할 수 있다.

커널 함수는 SVM에서 가장 중요한 요소이며, 여러 문제에 적합한 커널을 만들어 사용할 수 있다. 선형 커널, 다항 커널, Gaussian 커널 등이 많은 문제에서 주로 사용되고 있다.

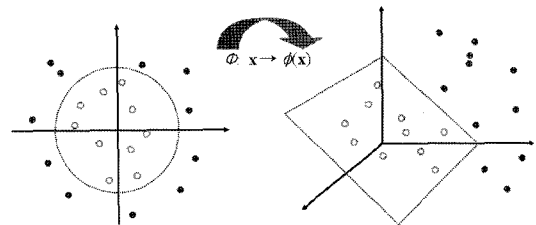


그림 6 커널 함수를 이용한 SVM의 이진 분류

선형으로 주어진 데이터들을 분류할 수 있는 초평면의 수는 무한히 많은데, SVM은 마진을 가장 크게 하는 최적의 초평면을 찾는다. 그림 7에 2차원 실수 공간에서 데이터 포인트들을 두 개의 클래스로 분류하는 초평면이 나타나 있다. 여기서 마진  $d$ 는 두 클래스 사이의 간격을 가리키며 다음과 같이 나타낼 수 있다.

$$d = 2 / \|w\|$$

따라서 SVM 학습은 최적화 문제로 생각될 수 있다.

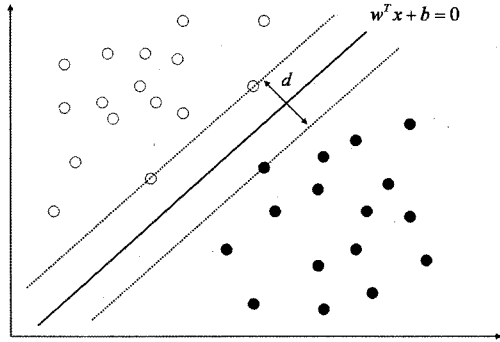


그림 7 Support Vector와 초평면

모든 데이터가 초평면에서의 거리가 최소한 1이라고 가정한다면 학습 데이터 집합에 대해 다음과 같은 두 가지 제약이 따른다.

$$w \cdot x + b \geq 1 \quad \text{if } y_i = 1$$

$$w \cdot x + b \leq -1 \quad \text{if } y_i = -1$$

즉,  $y_i(w \cdot x + b) \geq 1$ 이라는 제약 하에서  $\|w\|$ 가 최소가 되도록  $w$ 와  $b$ 를 찾는다. SVM의 장점은 이론적으로 매우 깨끗하며 대부분의 문제에서 또한 데이터의 양이 적을 때에도 좋은 성능을 보여준다는 것이다.

### 3.2 Conditional Random Field와 Markov Random Field

CRF(conditional random field)는 그래피컬 모델(graphical model) 중의 하나로서, 정보검색 분야에서는 일련의 시퀀스 데이터를 레이블링하는데 주로 사용되는 방법이다[9]. CRF는 HMM(hidden Markov model)이나 MEMM(maximum entropy Markov model)과 같은 방향성 그래프가 아니라 무방향성 그래프로 나타내며(그림 8), Markov 가정에 따라 각 정점(vertex)는 그 주위의 이웃 정점들과 의존 관계를 가지게 된다.

주어진 입력 데이터의 시퀀스를  $X$ 라고 하고 이에 해당하는 레이블을  $Y$ 라고 하면, 입력 데이터가 주어졌을 때 레이블의 시퀀스가 나타날 확률  $p(Y|X)$ 를 정의하게 되며, CRF에서는 이를 데이터의 자질 벡터  $f$ 와 각 자질의 가중치 벡터인  $\lambda$ 로 나타낼 수 있다. 예를 들어 품사 태깅 문제를 CRF로 모델링한다고 하면,  $X$ 는 단어들의 시퀀스가 되며  $Y$ 는 품사 태그의 시퀀스가

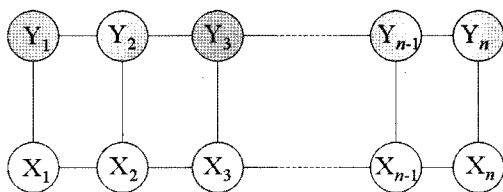


그림 8 시퀀스 데이터인 경우 CRF의 구조

된다. 일반적으로 입력 시퀀스와 레이블시퀀스를  $x = x_1 \cdots x_n$ 와  $y = y_1 \cdots y_n$ 라고 했을 때, 각각의 자질은 상태 자질  $s(y, x, i)$ 와 전이자질  $t(y, \hat{y}, x, i)$ 의 두 가지로 나누어 표현할 수 있다. 이 두 자질을 합쳐  $f(y, x, i)$ 라고 하면 전체 자질벡터는 아래와 같이 나타낼 수 있다.

$$F(y, x) = \sum_i f(y, x, i)$$

일반적인 그래피컬 모델에서는 위의 자질벡터를 이용해 조건부 확률분포를 아래와 같이 나타낸다.

$$p_\lambda(Y|X) = \frac{\exp \lambda \cdot F(Y, X)}{Z_\lambda(X)}$$

이 식에서  $Z_\lambda(x)$ 는 정규화(normalization)를 위한 분할 함수(partition function)이다.

CRF는 마코프 가정을 따르므로 조건부 확률분포를 아래와 같이 표현할 수 있다.

$$p(Y_i|Y_j, X, j \neq i) = p(Y_i|Y_{i-1}, Y_{i+1}, X)$$

이 식은 자질벡터를 이용한 조건부 확률분포 식으로 쉽게 나타낼 수 있게 된다. 따라서 가장 확률이 높은 시퀀스  $\hat{y}$ 는 아래와 같은 식에 의해 구할 수 있다.

$$\hat{y} = \operatorname{argmax}_y p_\lambda(y|x) = \operatorname{argmax}_y \lambda \cdot F(y, x)$$

MRF(Markov random field)는 CRF와 같은 무방향성 그래프로 나타내는 그래피컬 모델의 하나로서 결합 분포(joint distribution)를 효과적으로 모델링할 수 있으므로 기계학습 분야에서는 널리 쓰이는 방법 중의 하나이다. 그래프 상에서 이웃하는 노드들과만 의존 관계를 가진다는 접근방법은 같지만 MRF에서는 그래프 상에서 클릭(clique)을 정의하게 되고 클릭들 사이의 조건부 확률을 구하게 된다. CRF보다는 복잡하지만, 두 모델 모두 무방향성 그래피컬 모델로서 학습을 통해 potential function을 구하여 조건부 확률분포를 구한다는 큰 맥락은 같다.

### 3.3 잠재 변수 모델(Latent Variable Model)

실제 우리가 접하는 문제들에서는 직접 관찰 가능한 확률변수들만으로는 효과적으로 모델링할 수 없는 문제들이 많다. 이러한 문제들에는 직접 관찰할 수는 없지만 어떠한 현상들에 영향을 미치는 잠재 변수(latent variable) 혹은 은닉 변수(hidden variable)가 있다고 가정할 수 있다. 잠재 변수 모델(latent variable model)은 이러한 가정을 바탕으로 문제를 해결하는 접근방법이다. 따라서 LVM을 사용하면 실제 관찰된 확률

변수들의 통계적인 속성들을 잠재 변수로 설명할 수 있다.

일반적으로 통계적 모델에 잠재 변수를 사용하면 일반적으로 아래와 같이 표현할 수 있다.

$$p(X) = \int p(X, Z) dZ = \int p(X|Z)p(Z) dZ$$

여기서,  $X = (x_1, x_2, \dots, x_n)$ 는 관찰 변수의 벡터를 나타내며  $Z = (z_1, z_2, \dots, z_m)$ 는 잠재 변수의 벡터를 나타낸다. 즉, 관찰 변수  $X$ 가 잠재 변수  $Z$ 로부터 영향을 받으므로  $p(X)$ 를 잠재변수와 결합 확률 분포로 나타내는 것이다. 위의 식에서 알 수 있듯이, 구해야 할 확률  $p(X)$ 를 알기 위해서는  $p(X|Z)$ 와  $p(Z)$ 를 구해야 한다.

LVM으로 가장 잘 알려진 모델로는 mixture model과 factor analysis 등이 있다. Factor analysis는 데이터를 설명할 수 있는 요소(factor)들을 각 가중치의 행렬을 이용해 선형함수의 형태로 만들어 분석하는 방법이다. 그림 9에 factor analysis 모델이 간단히 나타나 있다.  $X$ 는  $n$ 차원의 데이터 벡터이며,  $Z$ 는  $m$ 차원의 잠재 변수 벡터이다. 이들 간의 관계는  $m \times n$ 차원의 행렬  $W$ 로 나타낼 수 있다. 이를 식으로 나타내면 아래와 같다.

$$X = Wz + e$$

Factor analysis는 일반적으로 SVD(singular value decomposition)를 통해 이루어진다.

### 3.4 Expectation Maximization Algorithm

EM 알고리즘은 손실 정보(missing information)가 있는 문제에서 아주 폭 넓게 사용되고 있는 기계학습 방법이다. EM 알고리즘은 실제 관찰 가능한 데이터만으로는 직접적으로 MLE(maximum likelihood estimation)를 구할 수 없을 때 반복적으로 파라미터들을 조금씩 변경해가면서 MLE를 구하는데 널리 사용되는 방법이다. 관찰된 데이터의 벡터  $X$ 를 불완전 데이터, 손실 정보를 포함하는 완전한 데이터의 벡터를  $Y$ , 실제 구해야 할 파라미터들의 벡터를  $V = (v_1, \dots, v_d)^T$

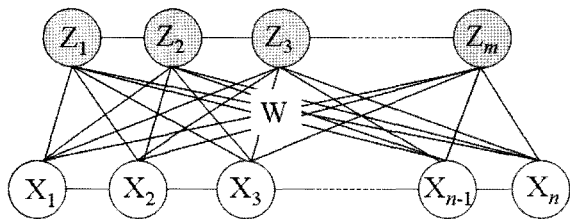


그림 9 기본적인 factor analysis 모델

라고 하면, 완전 데이터 벡터인  $Y$ 가 주어졌을 때의 log-likelihood는 아래의 식과 같이 표현할 수 있다.

$$\log L_c(V) = \log p_c(y; V)$$

하지만  $Y$ 가 주어지지 않고 실제로는 불완전 데이터  $X$ 만이 주어지기 때문에 완전 데이터의 log-likelihood를 직접 미분하여 likelihood가 최대가 되는 파라미터를 구하는 것은 불가능하다. 따라서 EM에서는 주어진  $X$ 를 이용하여 현재 파라미터  $V(k)$ 에 대한 조건부 확률의 기대값을 구하게 된다.  $k$ 는 반복횟수를 가리키는 인덱스이다.

EM 알고리즘은 크게 데이터  $y$ 가 주어졌을 때의 log-likelihood의 기대값을 구하는 E-step과 E-step을 통해 구한 likelihood의 기대값이 최대가 되도록 하는 파라미터  $V$ 를 찾는 M-step으로 나누어진다.

보통 처음 파라미터  $V(0)$ 는 임의의 값으로 정하게 되며 log-likelihood의 기대값은 아래와 같이 계산한다.

$$Q(V; V(0)) = E_{V(0)}\{\log L_c(V) | x\}$$

M-step에서는 위의 식을 통해 계산된 기대값이 최대가 되도록 하는 새로운 파라미터  $V(1)$ 를 구하게 되며 아래와 같이 나타낼 수 있다.

$$V(1) = \arg \max_V Q(V; V(0))$$

위의 두 과정을 파라미터의 값이 수렴하거나 그 차이가 미리 정해진 임계값보다 작아질 때까지 반복하게 된다. 따라서 E-step과 M-step을 다시 정의하면 아래와 같다.

E-step:

$$Q(V; V(k)) = E_{V(k)}\{\log L_c(V) | x\}$$

M-step:

$$V(k+1) = \arg \max_V Q(V; V(k))$$

M-step에서는 위의 식과는 다르게 log-likelihood의 기대값이 최대가 되도록 하는 파라미터를 구하는 것이 아니라 아래의 조건을 만족하는 새로운 파라미터를 구해도 된다.

$$Q(V(k+1); V(k)) \geq Q(V; V(k))$$

EM을 이용해 풀 수 있는 가장 대표적인 예로 3.3절에서 언급한 Gaussian Mixture model을 들 수 있다. Gaussian mixture model에서는 둘 이상의 정규 분포 중에서 어느 정규분포로부터 데이터가 나왔는지를 나

타내는 변수가 missing data가 된다. 여기서 관찰된 데이터  $X = (x_1, x_2, \dots, x_n)$ 는 아래와 같이 나타낼 수 있다.

$$f(x, \mu, \sigma, w) = \sum_{j=1}^m w_j g_j(x, \mu_j, \sigma_j)$$

여기서  $\mu_j, \sigma_j$ 는 각각 Gaussian의 파라미터인 평균과 분산을 나타내며  $w_j$ 는 각 Gaussian에 대한 가중치를 나타낸다. 즉, 찾아야 하는 파라미터가 세 개이다. 위의 식이 나타내는 GMM의 log-likelihood 함수는 아래와 같이 나타낼 수 있다.

$$P(X|M, \Sigma) = \sum_{i=1}^n \ln \sum_{j=1}^m P(j) p(x_i | j; \mu_j, \sigma_j)$$

하지만 위의 식을 최대화 하는 파라미터들을 찾기 위해서는 주어진 데이터가 어떤 Gaussian으로부터 나왔는지 추정하기가 어려우므로 이를 바로 계산하기는 불가능하다. 따라서 위 식에 지시 변수  $z_{ij}$ 를 넣어 아래의 식과 같이 변형할 수 있다.

$$P(X|Z, M, \Sigma) = \sum_{i=1}^n \sum_{j=1}^m z_{ij} \ln p(x_i | z_i; \mu_j, \sigma_j) P(z_i)$$

위에서 변경된 likelihood 함수로는 직접 계산이 불가능하기 때문에 EM 알고리즘을 이용해  $Z$ 의 분포에 관해 기대값이 최대가 되도록 한다.

$j$ 번째 Gaussian의 파라미터인 평균과 분산을 간단히  $v_j$ 라고 나타내었을 때 처음 파라미터의 값  $v(0)$ 와  $w(0)$ 는 일반적으로 임의로 주어지게 되며 이를 EM 알고리즘에 의해 반복하는 과정을 거쳐 파라미터들이 수렴할 때까지 아래와 같은 과정을 거치게 된다.

E-step:

log-likelihood의 기대값을 아래와 같이 계산한다.

$$Q(v; v(k)) = \sum_{i=1}^n \sum_{j=1}^m \pi_{ij} (\ln g_j(x_i, v_j(k)) + \ln w_j(k))$$

$$\pi_{ij} = E[z_{ij} | x_i] = \frac{w_j(k) g_j(x_i, v_j(k))}{\sum w(k) g(x, v(k))}$$

여기서,  $\pi_{ij}$ 는  $i$ 번째 데이터가  $j$ 번째 Gaussian으로부터 나왔을 확률이므로, 이것이  $z_{ij}$ 에 대한 기대값이 된다.

M-step:

E-step에서  $k$ 번째 파라미터들을 이용하여 구한 기대값이 최대가 되도록 하는 파라미터  $v(k+1)$ 과  $w(k+1)$ 을 구한다.

$$w_j(k+1) = \frac{1}{n} \sum_{i=1}^n \pi_{ij}$$

$$v_j(k+1) = \arg \max_{v_j} \sum_{i=1}^n \pi_{ij} \ln g_j(x_i, v_j(k))$$

## 4. 결론

본고에서는 기계학습이 정보검색 분야에서 어떤 역할을 하고 있는지 살펴보았다. 특히, ACM SIGIR 학회를 중심으로 하여 어떤 기계학습이 최근에 정보검색 분야에서 자주 사용되고 있는지 알아보았으며, 자주 쓰이는 기계학습에 대한 간단한 소개를 하였다. 이와 같은 내용을 정리하면, 아래와 같다.

우선 지난 4년간의 ACM SIGIR 학회 발표 논문을 조사해 보았을 때, 전체 논문 중 약 82%가 기계학습 방법을 사용하고 있고 정보검색에 사용된 기계학습 방법의 종류가 매우 많음을 확인할 수 있었다. 또한, 어떤 기계학습 방법들의 사용이 늘어나고 있고, 어떤 방법들의 사용이 줄어들고 있는지에 대한 경향을 논문 분석을 통해 알아보았다. 마지막으로, 최근에 자주 쓰이는 네 가지 기계학습 방법에 대한 간단한 소개를 통해, 정보검색 분야에서 새로이 기계학습 방법을 적용하고자 하는 연구자들에게 도움이 되도록 하였다.

## 참고문헌

- [1] G. Salton, "Recent Trends in Automatic Information Retrieval," In Proceedings of the 9th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 1-10, 1986.
- [2] Y. Hillel, "Theoretical Aspects of the Mechanization of Literature Searching," in Digitale Informationswandler, pp. 406-443, 1962.
- [3] G. Salton and M. Lesk, "Computer Evaluation of Indexing and Text Processing," Journal of the ACM, Vol. 15, No. 1, pp. 8-36, 1968.
- [4] J. Ponte and B. Croft, "A Language Modeling Approach to Information Retrieval," In Proceedings of the 21st Annual International ACM SIGIR Conference on Research and Development in Informaiton Retrieval, pp. 275-281, 1998.
- [5] B. Croft, "Machine Learning and Information Retrieval," Invited Talk in ICML '95, 1995.

- [6] E. Brill, "Transformation-Based Error-Driven Learning and Natural Language Processing: A Case Study in Part-of-Speech Tagging," Computational Linguistics, Vol. 21, No. 4, pp. 543-565, 1995.
- [7] D. Oard, "Interactive Cross-Language Information Retrieval," SIGIR Forum, Spring 2001, Vol. 35, No. 1, 2001.
- [8] V. Vapnik, Statistical Learning Theory, Springer, 1995.
- [9] J. Lafferty, A. McCallum and F. Pereira, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data," In Proceedings of the 18th International Conference on Machine Learning, pp. 282-289, 2004.
- [10] D. Metzler and W. Bruce Croft, "A Markov Random Field Model for Term Dependency," In Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 472-479, 2005.
- [11] Christopher M. Bishop, "Latent Variable Models," in Learning in Graphical Models, M. Jordan (Ed), MIT Press, pp. 371-403, 1999.



**박성배**

1994.2 한국과학기술원 전산학(학사)  
 1996.2 서울대학교 컴퓨터공학(석사)  
 2002.8 서울대학교 전기컴퓨터공학(박사)  
 2004.3~현재 경북대학교 컴퓨터공학과 조교수  
 관심분야 : 기계학습, 자연언어처리, 정보검색, 생물정보학  
 E-mail : sbpark@sejong.knu.ac.kr



**태윤식**

2005. 2 경북대학교 컴퓨터공학(학사)  
 2005. 3~현재 경북대학교 컴퓨터공학 석사과정  
 관심분야 : 정보검색, 자연언어처리, 기계학습  
 E-mail : ystae@sejong.knu.ac.kr

**프로그래밍언어연구회 춘계학술발표회**

- 일 자 : 2007년 4월 28일
- 장 소 : 숙명여자대학교
- 내 용 : 학술발표 등
- 주 최 : 프로그래밍언어 연구회
- 상세안내 : <http://www.sigpl.or.kr>