베이즈 정리와 MLE, MAP

작성자 : 조현제(Retriever)

작성일 : 2012.04.24

소속 : NewHeart

E-mail : retriever89@gmail.com

1. 서론

Bayesian Theorem (베이즈 정리)은 Machine Learning, Artificial Intelligence, Computer Vision, Data Mining 등의 분야에서 가장 기본이 되는 개념이다. 이 정리를 응용하여 추출된 데이터를 통해서 어떤 값에 대한 추정치를 계산하는 Maximum Likelihood Estimation (MLE) 기법과, Maximum a Posteriori (MAP) 기법 역시 자주 등장하는 이론이며, 핵심적인 이론이다. 때문에 정리를 해야 할 필요성을 느꼈고, 각각 이론에 대해 간단한 예를 들어 설명할 것이다.

2. 개념 설명

2.1 Bayesian Theorem (베이즈 정리)

Bayesian Theorem 에는 크게 다음 두 가지 요소가 존재한다.

- P(A) : A의 사전확률 (a prioi). 어떠한 사건에 대한 정보가 없을 때의 확률.
- P(A|B) : B에 대한 A의 사후확률 (posteriori). B라는 정보가 주어졌을 때의 확률.

그리고 Bayesian Theorem은

$$\mathrm{P(A|B)} = \frac{P(B|A)P(A)}{P(B)} \qquad (\mathrm{Eq.\,1})$$

이다. 유도 과정은 이 문서에서 생략하겠다. ([1] 참조)

내 친구 A가, 자신이 카페에 갔었는데 모르는 사람이 자기 번호를 알아갔다고 자랑을 하였다. 이 때, 그 번호를 알아 간 사람이 여자일 확률 P(W)는 0.5(남자일 확률 P(M) = 0.5) 이다. 그런데 친구가 그 번호를 알아간 사람의 머리카락 길이가 어깨 아래까지 길었다(머리가 어깨 아래까지 길었다는 사건: L)고 말하였다. 아무 조건이 없이 누군가 번호를 알아갔다고 할 때의 P(W)는 0.5 이지만, 여기에 한가지 정보가 추가되었을 때의 확률 P(W)는 당연히 변화한다. 조건부 확률 계산 방식에 의해 다음과 같이 계산이 가능하다.

$$\mathrm{P(W|L)} = \frac{P(W \cap L)}{P(L)}$$

하지만 위 식에서 $P(W \cap L)$을 계산하기 위해서는 어려움이 있다.[1] 이러한 과정을 줄이기 위해

[1] 위 예제에선, 사전확률의 변수가 하나이기 때문에 $P(W \cap L)$을 계산 하는 것이 간단하지만, 변수가 많아지면 직접 구하는 것이 어렵다.

Bayesian Theorem 을 적용하면,

$$\mathrm{P(W|L)} = \frac{P(L|W)P(W)}{P(L)}$$

로 바뀐다. 위 식에서는 $P(L|W)$만 알게 되면 $\mathrm{P(W|L)}$를 계산 할 수 있기 때문에 더 효율적이다.

2.2 Maximum Likelihood Estimation (MLE)

Likelihood는 이미 주어진 표본적 증거에 비추어 보았을 때, 모집단에 관해 어떠한 통계적 추정이 그럴듯한(Likelihood) 정도를 말해주는 것을 가리킨다. 다시 말하면 어떤 가설을 전제하였을 때, 그 전제하에서 우리에게 주어져 있는 증거가 얼마나 나타날 수 있는가에 대한 정도이다[2]. 때문에 가설 T가 전제 되었을 때 증거 E 가 등장할 확률인 $\mathrm{P(E|T)}$ 에 비례한다. 이를 $\mathrm{L(T|E)}$라고 하며, 증거 E가 관측되었을 때 이론 T와 그럴듯한 정도라 표현한다. 앞에서 언급하였듯이, $\mathrm{L(T|E)}$는 $\mathrm{P(E|T)}$에 비례하기 때문에 $\mathrm{L(T|E)}$의 값을 최대화 하는 과정은 $\mathrm{P(E|T)}$를 최대화 하는 과정과 동일하다. 간단한 동전 던지기 예[출처: 3]를 보자.

우리는 일반적으로 동전 하나를 던졌을 때 앞 또는 뒤가 나올 확률은 같다고 가정하여 0.5라고 생각한다. 하지만 이 0.5라는 확률 값은 정확한 값이 아니라 우리가 가정한 값이다. 때문에 몇 번의 수행 결과로 동전의 앞면이 나올 확률 P(H)를 정하고자 한다.

만약 100번의 동전던지기를 수행했을 때, 앞면이 56번 나왔다면 '동전을 던졌을 때 앞면이 나올 확률'은 얼마라고 얘기 할 수 있을까? 이 문제에 대한 해답을 구하는 것이 Maximum Likelihood Estimation 이다. 좀 더 자세하게 설명하면 앞에서 언급한 용어 $\mathrm{L(T|E)}$에서, 증거 E는 앞면이 56번 나왔다는 사실 이고, 이론 T를 변화시키면서 어느 이론이 가장 그 확률이 높은지 찾는 과정이다.

$\mathrm{L(T|E)}$는 $\mathrm{P(E|T)}$ 를 최대화 하는 과정이기 때문에 이론 T가 'P(H) = 0.5' 일 때를 계산하면 다음과 같다.

$$\mathrm{L(P(H) = 0.5|E)} = \frac{100!}{56!\,44!}0.5^{56}0.5^{44} \approx 0.0389$$

이다. 이와 같은 방법으로 몇몇 이론 T에 대해 값을 구해보면 다음 표와 같다.

T	Likelihood
P(H) = 0.48	0.0222
P(H) = 0.50	0.0389
P(H) = 0.52	0.0587
P(H) = 0.54	0.0739
P(H) = 0.56	0.0801
P(H) = 0.58	0.0738
P(H) = 0.60	0.0576
P(H) = 0.62	0.0378

Table 1 가정 T의 변화에 따른 Likelihood 의 변화

위 표에서 가장 높은 Likelihood를 가지는 이론 T는 P=0.56 일 때이다. 때문에 우리는 동전의 앞면이 나올 확률이 0.5 라는 것을 전혀 모르는 상황에서 위와 같은 증거가 있을 때에는, "동전을 던져서 앞면이 나올 확률은 0.56 이다" 라고 말할 수 있다.

이렇게 P(H)에 대한 확률을 모르는데, 어떠한 데이터가 주어진 경우, 이 데이터를 통해서 확률 P(H)를 추정하는 과정을 Maximum Likelihood Estimation 이라 한다.

(본 예제의 출처는 [3]이다.)

2.3 Maximum a Posteriori (MAP)

Maximum a Posteriori를 설명하기 전에 먼저 Posteriori 에 대한 설명이 필요하다.

일반적으로 동전을 던졌을 때, 질량의 분포가 균등하게 이루어져 있다면 앞면이 나올 확률은 0.5 라고 알고 있다. 이 말이 검증 하기 위해 동전을 100번 던졌다. 그 결과로 70번의 앞면이 나왔는데, 이 경우에 Maximum Likelihood Estimation 에서는 앞면이 나올 확률 T 를 0.5 혹은 0.7 과 같이 가정하여 그 가정에 대한 확률을 구하였다. 하지만 이렇게 추측하였을 때에는 이전에 우리가 알고 있던 사실인 '앞면이 나올 확률은 0.5 이다' 가 전혀 반영되지 않고 순수하게 측정 된 결과로만 확률을 추측한다. 반면에 Posteriori를 구할 때에는 다음 식(Eq. 1에 대입)을 이용하여 사전에 우리가 알고 있는 정보를 활용한다.

$$P(T|E) = \frac{P(E|T)P(T)}{P(E)}$$

알고자 하는 확률 '100번 동전을 던진 시행에서 70번의 앞면이 나왔을 때, 동전의 앞면이 나올 확률' 은 $P(T|E)$로 표현이 된다. 이 확률은 Bayesian Theorem 을 통해서 우변과 같은 유도가 가

능하다. 이론 T에 특정 가정을 대입하고, 그에 대한 확률을 계산하는 방식이다. 만약 T = 0.5 라는 가정을 기준으로 계산하면,

$$P(T = 0.5|E = 0.7) = \frac{P(E = 0.7|T = 0.5)P(T = 0.5)}{P(E = 0.7)} \qquad (Eq. 2)$$

이 된다. 여기서 $P(E = 0.7|T = 0.5)$는 Likelihood Function 에서(각주 1 참조) 쉽게 구할 수 있고, $P(E = 0.7)$는 변수에 특정 값을 대입했을 때 나온 값이 아닌 상수이기 때문에, $P(T = 0.5)$, 즉 우리가 이전에 알고 있던 '동전을 던졌을 때 앞면이 나올 확률은 0.5 이다' 라는 명제에 대한 사전 확률만 주어진다면 Posteriori 를 계산 할 수 있다.

이제 (Eq. 2) 에서 가정을 변수 x라고 생각을 하면 다음과 같이 식이 바뀐다.

$$P(T = x|E = 0.7) = \frac{P(E = 0.7|T = x)P(T = x)}{P(E = 0.7)} \qquad (Eq. 3)$$

(Eq.3) 과 같이 변수 x 에 대해 식을 표현하면, x에 대한 함수가 되어, x의 변화에 따른 최대값을 구할 수 있다. 이때 최대값을 구하는 과정이 Maximum a Posteriori 이다.

3. 결론

간단한 예를 통해서 Bayesian Theorem, Maximum Likelihood Estimation, Maximum a Posteriori 에 대해 알아보았다. 이 개념들은 컴퓨터 분야에서 다양하게 활용되는 이론이므로 반드시 알아 둘 필요가 있다. 이제 이 개념들을 Markov Random Fields[4] 에 적용시켜 Image Segmentation 을 수행하는 프로그램을 작성 할 것이다.

4. 참고 문헌

[1] http://ko.wikipedia.org/wiki/베이즈_정리

[2] 전영삼, 전문가 시스템을 위한 우도의 활용, 고려대 철학연구 14 권, 1990

[3] http://statgen.iop.kcl.ac.uk/bgim/mle/sslike_3.html

[4] http://newheart.kr/xe/4131

[5] http://www.ws.binghamton.edu/fowler/fowler%20personal%20page/EE522_files/EECE%20522%20Notes_26%20Ch_11B.pdf

[6] http://www.robots.ox.ac.uk/~pawan/eccv08_tutorial/index.html

[7] http://stats4you-textcube.blogspot.com/2009/10/우도의-개념과-최대우도측정법 maximum-likelihood.html

[8] http://www.bioen.utah.edu/wiki/index.php?title=ML_and_MAP