모바일 광고 기술에서 Hawkes 프로세스 기반 생성 모델에 관한 연구

황상원*, 조인휘✉
*한양대학교 컴퓨터소프트웨어학과
e-mail : sangwonhwang@naver.com

Hawkes Process based Generative Model in Mobile AD Technology

Sangwon Hwang*, Inwhee Joe {@hanyang.ac.kr}
*Dept. of Computer and Software, Han-Yang University

요 약

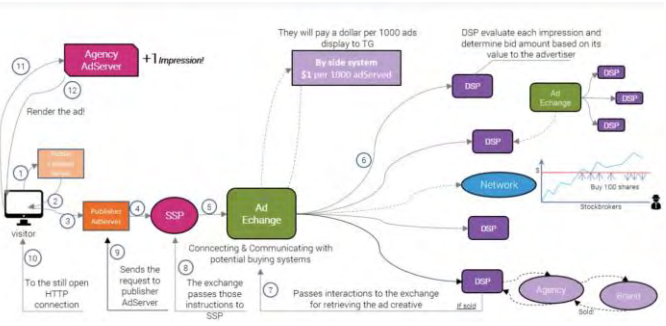
모바일 광고 기술(AD Tech)분야에서 기계학습은 사용자별 CTR(Click Through Ratio) 예측[1]에 국한되어 왔으며 2016 년도부터 시행된 GDPR(General Data Protection Regulation)[2] 정책으로 인하여 AD Tech 관련 사업자의 사용자 개인정보 접근이 제한되어왔다. 그 결과 Feature Data 에 기반한 사용자 타겟팅 및 Ad Tech 산업에서 기계 학습 기술을 적용하는데 현실적 어려움을 가져왔다. 본 논문에서는 기존 연구와 현실적 한계를 극복하기 위하여 사용자가 아닌 매체별 클릭 분포를 예측하는 Hawkes Process 에 기반한 생성모델을 제안하고 그 가능성을 살펴본다.

1. 서론

베르누이 실험의 성공 횟수를 확률 변수로 가지는 이항 분포에서 시행 횟수가 무한히 커지는 경우, 연산량의 증가로 인해 확률을 계산하는데 어려움이 있다. 이를 해결하기 위해 포아송 근사를 활용, 독립 사건의 확률을 계산하는데 포아송 분포를 많이 사용해왔다[3]. 하지만 포아송 분포는 각각의 시행이 독립적이며(Independent Increment) 확률은 변하지 않는다(Stationary Increment)는 성질로 인하여 과거 기록을 바탕으로하는 Survival Analysis 에는 한계를 보여왔다. Homogeneous Poisson Process 를 기반으로 한 모바일 광고/컨텐츠 추천 모델 또한 위의 두 속성으로 인해 단일 세션의 CTR(Click Through Rate)만을 고려, 분포를 모델링하는데 어려움이 있었다. 본 연구에서는 Non-Homogeneous Poisson Process 의 한 종류인 Hawkes Process 를 기반으로 매체별 click event 분포에 관한 생성모델을 제안한다.

이번 장에서는 AD Tech Network 구성 및 AD request 를 포함한 packet 의 구조 및 데이터 유실 가능성을 살펴보고 연구의 필요성에 대하여 알아본다. 2 장에서는 Hawkes Process 의 Hazard 함수를 수학적으로 유도하고 3, 4 장에서는 Hawkes Process 기반 생성모델의 결과를 보이고 앞으로의 활용방안 및 추후 연구과제에 대하여 논의한다.

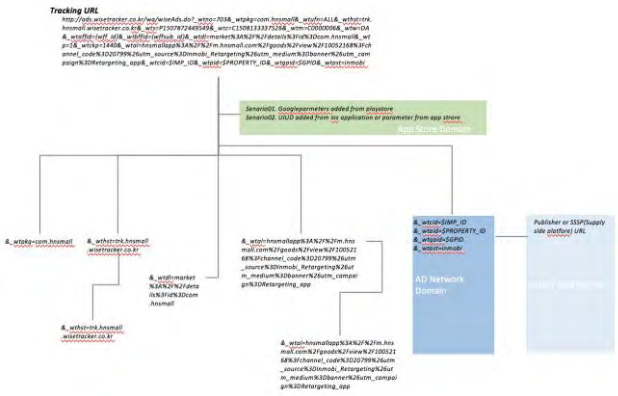
Advertising Technology 는 4 개 주체들간의 파라미터 교환을 통하여 이벤트를 Tracking 한다. 광고주(Advertiser), 매체(Publisher), 중계 네트워크(AD Network), Tracker 로 구성된다. 그림 2.1 은 Advertising Technology 주체간의 구조를 나타낸다.



(그림 1) AD Tech 구조 by Dr.Vytautas Bielinska

Ad Tech 분야에서의 conversion tracking 은 대게 광고 URL 을 통해 전달된다. 데이터 셋에 포함된 고유 아이디와 IP 주소, google referrer, iOS 의 경우에는 UDID 의 조합을 통하여 conversion 을 트래킹한다. 따라서 하나의 광고 URL 은 다수의 Deep Link 와 데이터셋, 고유아이디를 포함한 기타 파생링크를 통해 이루어진다. 따라서 광고주 혹은 AD Network 로부터 시작된 하나의 광고 링크는 cycle 이 있는 그래프 형태가 된다. 그림 2 에서는 첫번째 레벨까지만 기술한다. 문제점은 광고 URL 의

분기지점에서 발생하는 1)Network Error/Delay 로 인하여 각각의 데이터 셋으로 받은 고유 아이디가 불일치하는 부정 광고가 많이 발생하고 2)사용자 개인정보 접근 제한(GDPR)로 인하여 사용자 정보에 기반한 Feature Driven Machine Learning 기법의 적용이 어렵다는 사실이다.



(그림 2) URL 의 분기

2. Background on Hawkes Process

Hawkes Process 는 단위시간당 특정 이벤트의 발생 빈도가 가변하는 Point Process 이며 Poisson Point Process 의 변형된 형태이다. 이번 장에서는 Poisson 분포와 지수분포와의 관계와 각각의 질량함수에 근거하여 Hawkes Process 의 Hazard function 을 수학적으로 유도한다. 특정 사건이 발생한 순차적 시간을 원소로하는 집합(sequence)을 $T = \{t_i \mid t_i \in R^+ \ where \ i \in Z^+\}$로 둘때 t시간까지 발생한 사건의 수 Counting Process N_t 를 다음과 같이 정의한다.

$$N_t := \sum_{i \geq 1} 1\{t \geq t_i\} \quad (1)$$

분포의 단위시간을 t시간으로 확장할 때 t시간동안 평균 발생횟수는 $npt = \lambda t$이고 $\mathrm{X}_t \sim$ PoissonProcess (λt)를 따른다. 이때 확률변수 X_t는 t 시간까지 사건발생횟수이고 Counting Process N_t 와 동일하다.

따라서 집합 T 에 대한 inter-arrival time 을 $l_i = \{l_1, l_2, \cdots, l_n\}$로 두면 다음이 성립한다.

$$t_{n-1} \leq l_n < t_n \quad \Rightarrow l_n = [t_{n-1}, t_n) \quad (2)$$

j 번째 사건이 발생 시간을 t_j', 단위시간을 확률변수 두면 j 번째 사건이 t 시간 이후에 발생하는 확률은 t 시간까지 j 번째 사건이 발생하지 않는 확률과 같다.

$$f(\mathrm{X}_t = 0, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t} \quad (3)$$

$$P(t < t_j) = f(\mathrm{X}_t = 0, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^0}{0!} = e^{-\lambda t} \quad (4)$$

j 번째 사건이 발생할 확률은 $1 - S(t)$ 이고 이는 확률변수 t에 대한 누적분포함수이다,

$$\mathrm{P}(0 \leq t_j' \leq t) = F(t) = 1 - e^{-\lambda t} \quad (5)$$

확률변수 t 의 확률밀도함수(PDF)는 다음과 같다.

$$f(t) = \frac{d}{dt}F(t) = \lambda e^{-\lambda t} \quad (6)$$

$f(t)$와 $s(t)$를 각각 확률밀도함수와 누적분포함수로 갖고 특정 사건이 발생하는데 걸리는 시간을 확률변수 t로 가질 때, 확률변수 t 는 지수분포를 따르고 다음이 성립한다.

$$\begin{aligned} \mathrm{S}(t) &= \mathrm{P}(t < t_j') = e^{-\lambda t} \\ \mathrm{F}(t) &= \mathrm{P}(0 \leq t_j' \leq t) = 1 - e^{-\lambda t} \\ f(t) &= \frac{d}{dt}F(t) = \lambda e^{-\lambda t} \end{aligned} \quad (7)$$

Non-Homogeneous Poisson Process 에서 Lambda Intensity 함수 $\lambda(t)$ 는 앞선 사건의 기록 H_t 가 일어난 후 Δt 동안 이벤트가 발생할 조건부 확률의 시간 변화량에 대한 미분 계수 $\lambda(t)$ 로 정의한다. 이를 Hawkes Process 의 Hazard Function 이라 한다.

$$\Delta t \lambda(t) = \lim_{\Delta t \to 0} Pr(t < T \leq t + \Delta t | T > t) = f(t)/S(t)$$

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{Pr(t < T \leq t + \Delta t | T > t)}{\Delta t} \quad (8)[4]$$

수식(8)의 계산은 평균($\lambda_0(t)$)과 kernel 의 합이며 다음과 같다.

$$\lambda(t) = \lambda_0(t) + \int_0^{t_i} \phi(t - T_i) \quad (9)$$

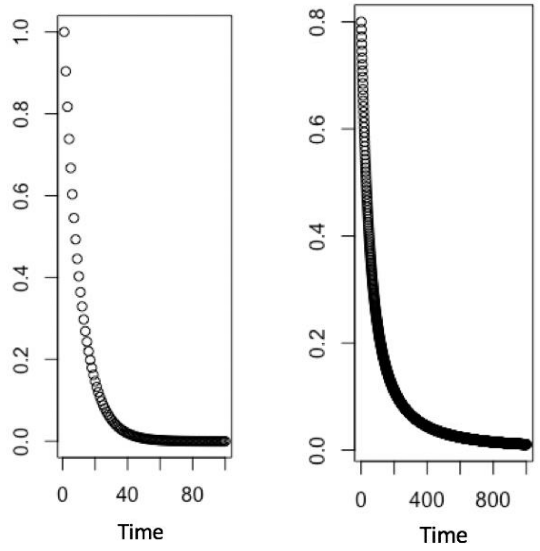
3. 제안하는 알고리즘

3.1. Type of Kernel

Hawkes Process 의 Hazard function 은 kernel 이라는 Hazard function 의 피적분 함수에 의하여 자유롭게 구현이 가능하다. 본 논문에서는 Hawkes 가 제안한 지수함수모델[5]과 사회 관계망 모델 설계에 적합한 Power-Law kernel[6]에 기반하여 생성 모델을 구현한다. 그림 3 은 시뮬레이션 결과를 보여준다.

$\phi(t-t_i) = \alpha e^{-\theta t}$ (α = 0.1, θ = 0.5)

$\phi(t-t_i) = \alpha(\tau + c)^{-1+\theta}$ (α = 0.8, c = 10, θ = 0.5)



(그림 3) Kernel Simulation

3.2. 파라미터 최적화 및 Gradient Hawkes

SGD(Stochastic Gradient Descent)방법을 활용하여 Negative Log Likelihood(10)를 최소화하는 방식으로 파라미터를 최적화한다. 이때 각 파라미터가 수렴할 때까지 SGD 를 반복하고 이후 Cumulative Intensity Function 의 값과 데이터의 오차를 Gradient Descent 방식을 통해 조정하는 방법으로 Hawkes 모델(11)을 최적화한다.

Algorithm iterator fit

1. ***input*** : event times T_i with flow capacity $\lambda_0(t)$
2. ***output*** : parameters of exponential kernel $\mu, \alpha, \theta, \gamma$
3. ***while*** parameters get converged
   (a) for each time unit out of eval points with T_i
   (b) $distance = (evaluation\ point_i - t_i)$
   $i++$
   (c) optimize parameters which minimize negative log likelihood $l(\theta)$
   (d) return $\mu, \alpha, \theta, \gamma$
4. ***input*** : CIF values on time line T_i
5. ***output*** : parameter γ
6. ***while*** parameter γ gets converged
   (a) optimize γ which minimize MSE
   (b) return γ
7. update γ

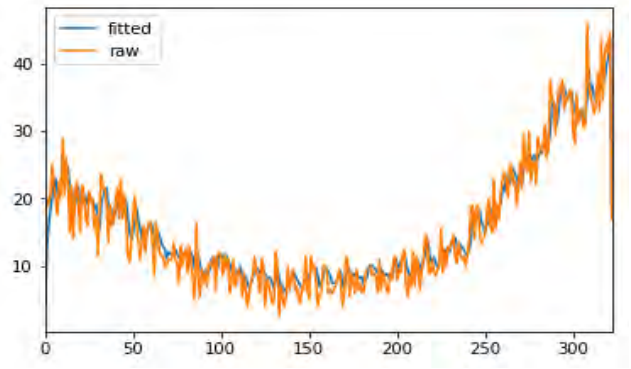
(그림 4) 파라미터 최적화

$$-l(\theta) = \int \lambda(t,x,\theta)dtdx - \int \log \lambda(t,x,\theta)N(dt,dx) \quad (10)$$

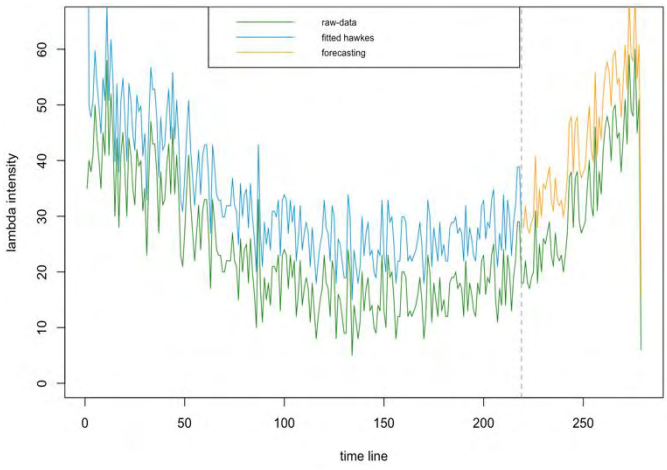
$$\lambda(t) = \gamma\lambda_0(t) + \gamma \int_0^{t_i} \phi(t-T_i) \quad (11)$$

4. 학습 및 예측 결과

본 연구에서 sampling 과정이 이상적으로 진행되었다고 가정한다. 앞서 제안한 Gradient Hawkes 를 기반으로 하는 생성 모델그래프를 그려보고 그 오차를 알아본다. Kaggle 대회에서 사용된 Criteo 사의 Click 데이터[7]를 학습 데이터로 사용, 생성모델을 구현한다. 예측에 적용된 데이터는 전체 7000 Time Stamps 중 2000 Time Stamps 이다. 그림 5 는 5000 Time Stamps 를 학습한 결과이고 그림 6 는 10000 Time Stamps 를 학습한 이후 원데이터와의 비교 그래프이다. 그림 5 에서는 처음 5000 Time Stamps 에서 학습을 진행하고 다음 2000 Time Stamps 을 예측한 결과를 나타낸다. 220 구간부터는 예측 구간이며 MSE 는 그림 8 과 같다.

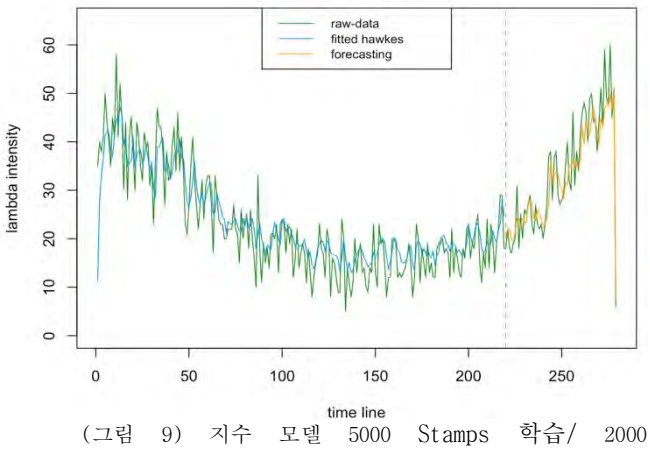


(그림 6) 10000 Time Stamps 학습



(그림 7) Power Law 모델 5000 Stamps 학습/ 2000 Stamps 예측

	MSE	Min	1st Qu	Median	Mean	3rd Qu	Max
Raw Data	0	5.00	17.00	22.00	25.09	33.00	60.00
Power Law	96.787444203194	14.98	26.92	31.89	35.29	42.84	134.80
Exponential	7.5030165214	7.503	18.280	22.630	25.590	33.220	50.240



(그림 9) 지수 모델 5000 Stamps 학습/ 2000 Stamps 예측

5. 결론 및 향후 연구

본 논문에서는 그동안 사용자 개인의 CTR 예측 및 Feature Data 에 기반한 기계학습에 국한되어 있던 광고 기술의 인공지능 분야에서 Hawkes Process 를 통해 시간 데이터 입력만으로도 매체별 클릭 분포에 관한 생성모델의 구현이 가능함을 보였다. 또한 단일 입력 데이터만으로도 높은 예측율과 연산 효율성을 보일 것으로 예상되기 때문에 AD Tech 분야 뿐만 아니라 Feature Data 기반 기계학습이 어려운 분야에서 활용 가능할 것으로 예측된다. 하지만 Hawkes 모델이 갖는 과적합(over-fitting) 문제와 sampling 관련 연구가 추가적으로 필요하다. Conversion Rate(Number of Conversion / Number of Click) 예측은 광고 기술 분야에서 중요한 주제이다. Hawkes 기반 생성모델과 신경망을 조합한 Hybrid Model 을 바탕으로 conversion 분포 및 Conversion Rate 을 예측하는 연구를 향후 연구의 목표로 한다.

(그림 8) 예측 구간 지표

6. 참고문헌

[1]Field-aware Factorization Machines for CTR Prediction

[2]REGULATION (EU) 2016/679 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

[3]Improvement of the LPWAN AMI backhaul's latency thanks to reinforcement learning algorithms

[4]]Rodríguez, G. (2007). Lecture Notes on Generalized Linear Models. URL: http://data.princeton.edu/wws509/notes

[5]A. G. Hawkes. 1971. Spectra of some self-exciting and mutually exciting point processes. Biometrika, pp. 89-90.

[6] Rizoiu, M.-A., Xie, L., Sanner, S., Cebrian, M., Yu, H., & Van Hentenryck, P. Expecting to be HIP: Hawkes Intensity Processes for Social Media Popularity. In 26th International Conference on World Wide Web - WWW ' 17, pp. 735-744, Perth, Australia, 2017. doi: 10.1145/3038912.305265

[7]http://labs.criteo.com/wp-content/uploads/2014/07/criteo_conversion_logs.tar.gz