Contextual Bandit for Recommendation System

KrishnaKant Singh,Abhishek Jain,Varun Mishra

September 6, 2016

Task description

Personalisation is important for making recommendation to the user.Many web servies try to personalise their search results according to the intrest of the users one of the most popular technique for this is Collabrative Filtering. But CF approaches face many problems like dynamic changes like addition or deletion of the items , cold start problem and time varying popularity of items.
One of the most popular solution to this the Multi Armed Bandit problem(Described Later).This solution suffers greatly because it dose not uses the context information at all.A variation of the multi armed bandit called the contextual multi armed bandit is much better alternative.
Using techniques form Contextual Bandit try to build a online learning system that can rank the items on the fly given diffrent contexts (diffrent users). [**?**] [**?**]

Background

Multi Armed Bandit The problem more generally is called the k-armed bandit problem defined as follows on each round
1.A policy choses arm a from 1 of k arms 2.The world reveals the reward r_a of the choosen arm As information is accumulated over multiple rounds,a good policy might converge on a good choice of arm. More formally ,

$$Arms \in 1, 2, ...k$$

$$Actions \in a_1, a_2, ...a_k$$

$$Rewards \in r_{a1}, r_{a2}, ..., r_{ak}$$

The rewards can be defined as Expectation of distribution on r_{a1} Objective function is

$$Max \sum_{t=0}^{T} r_t$$

If we have some intial distribution of rewards over the arms then a greedy strategy can be to always select the arm with highest reward to get the maximum reward but we can be stuck in a local optimum.We need a policy that can explore and exploit in some way to get the maximum reward.There are several strategies for this the ε greddy strategy and Upper confidence Bound strategy.These strategies are shown to converge to a global optimum. A major problem with this is we do not care for context for eg in the medical treatment recommendation problem,the medical treatment with highest recommendation is prescribed for everybody without taking into account the symptos of the patients.
Contextual Bandit A contextual MAB can be defined as following [**?**] [**?**] The world produces some context

$$x_t \in X$$

The learner chooses an action

$$a_t \in 1.....K$$

The world reacts with reward

$$r_t(a_t) \in [0,1]$$

A policy can be defined as follow

$$\prod = \{\pi : X -> 1..., K\}$$

The objective function is defined as

$$Regret = max_{\pi \in \prod} \sum_{t=1}^{T} r_t(\pi(x_t)) - \sum_{t=1}^{T} r_t(a_t)$$

This can be interpreted as chossing actions at each step to minimize regret which is defined as diffrence in the reward between selected action and optimal action at step t. Contextual bandits systems have gained a lot of intrest of late in the Information Retrival Community of late. For solving the contextual bandit problem Linear Upper Bound Confidence Interval and Linear Upper Bound Confidence Internal with Hybrid linear models are popular approaches.Thompson sampling is also another popular alternative.[**?**] [**?**] [**?**]

Data and evaluation

At present the most popular data sources for the evaluation Contextual Bandits are **Yahoo!Today Module** and **KDD 2012 advertising challenge** Evaluation methodology for Bandit problems are much harder than classification problems in machine learning.Still 2 popular techniques for evaluation are Replayer method [**?**] and the simulator method [**?**][**?**]. We intent on using the replayer method as it is an offline evaluation method and can be shown to perform better than the simulator method as bias is introduced inherently in a simulator.

References

[1] Adversarial bandits and the exp3 algorithm — math programming. `https://jeremykun.com/2013/11/08/adversarial-bandits-and-the-exp3-algorithm/`. (Accessed on 09/06/2016).

[2] agrawal13.pdf. `http://jmlr.org/proceedings/papers/v28/agrawal13.pdf`. (Accessed on 09/06/2016).

[3] Contextual bandits machine learning (theory). `http://hunch.net/?p=298`. (Accessed on 09/06/2016).

[4] An introduction to contextual bandits - the stream blog. `http://blog.getstream.io/introduction-contextual-bandits/`. (Accessed on 09/06/2016).

[5] Learning for contextual bandits. `http://hunch.net/~exploration_learning/main.pdf`. (Accessed on 09/06/2016).

[6] Multi-armed bandits. `https://dataorigami.net/blogs/napkin-folding/79031811-multi-armed-bandits`. (Accessed on 09/06/2016).

[7] Personalization with contextual bandits - richrelevance engineering blog : Richrelevance engineering blog. `http://engineering.richrelevance.com/personalization-contextual-bandits/`. (Accessed on 09/06/2016).

[8] Lihong Li, Wei Chu, and John Langford. A Contextual-Bandit Approach to Personalized News Article Recommendation. pages 661–670, 2010.

[9] Lihong Li, Wei Chu, John Langford, Taesup Moon, and Xuanhui Wang. An Unbiased Offline Evaluation of Contextual Bandit Algorithms with Generalized Linear Models. *Stat.Berkeley.Edu*, 1:1–18, 2011.

[10] Chunqiu Zeng, Qing Wang, Shekoofeh Mokhtari, and Tao Li. Online Context-Aware Recommendation with Time Varying Multi-Armed Bandit. 2016.