Reinforcement Mechanism Design for Fraudulent Behaviour in E-Commerce

May 2020

1 Introduction

In e-commerce website, whenever a buyer searches for product, based on allocation algorithm, sellers are allocated slots/impression for their products related to the search keyword. Mostly this allocation algorithm is based on reputation of seller. Reputation Scores are given according to a seller's conversation rate, overall transaction, etc. The intention is to bring capable sellers forward and generate more revenue in long run.

However, seller use illicit ways to boost their reputation such as faking historical transactions, which increases their revenue but not for the system. Previous work involve :

1. Detecting such behaviour using Machine learning techniques as well as manual labour and to penalize them. As sellers increases, it becomes difficult to achieve
2. Designing a mechanism where dominant strategy is to act honestly such that seller aims to maximize their reputation given associated cost for eliciting fake reviews and transactions. This works game theoretic assumptions where agents are rational and intelligent , also we assume cost of eliciting fake behaviour.
3. Using Reinforcement Mechanism Design. Building some rationality model based on past data, then design a deep reinforcement learning algorithm that takes this model into account and optimize some objective function.

2 Problem Addressed

Goal is to prevent fraudulent behaviour in e-commerce, and allocate impression feasibly (i.e. not to allocate more impression that we actually have) in such a way that it maximize real revenue generated by the system, without assuming sellers' rationality and intelligence.

3 Solution

Build a deep learning network to learn the rationality of sellers i.e. extent to which they will manipulate their transactions. Using this rationality model, modelling an Markov Decision Process, and solve this MDP using deep reinforcement learning (Deep Deterministic Policy Gradient) to optimize objective (maximize real revenue). Compare it with existing heuristic methods.

4 Approach

4.1 Setting of the paper

There are m sellers each having one product. on day t, n^t buyers impression are allocated. On day t, given an impression the ability of seller i to facilitate a transaction is given by conversion rate $cr_i(t)$. On day t, history of seller i is $H_i(t) = (n_i^t, r_i^t, a_i^t, p_i^t)$ where :

- $n_i^t = (n_i(1), n_i(2), \ldots, n_i(t-1))$, $n_i(j)$ is number of real impression allocated to seller i on day j. There is upper bound on maximum number of impression that can be allocated to maintain feasibility of mechanism
- $r_i^t = (r_i(1), r_i(2), \ldots, r_i(t-1))$, $r_i(j)$ is the number of real transactions of seller i on day j

- $a_i^t = (a_i(1), a_i(2), \ldots, a_i(t-1))$ $a_i(j)$ is the number of fake transactions of seller i on day j
- $p_i^t = (p_i(1), p_i(2), \ldots, p_i(t-1))$ $p_i(j)$ is the price that seller i decides on day j

  So the behavior model is function f, $a_i(t) = f(H_i(t))$, takes input $H_i(t)$ before day t and outputs $a_i(t)$, number of fake transaction seller i will make on day t

  Further, defining record history $R_i(t) = (n_i^t, v_i^t, p_i^t)$, where $v_i^t = (v_i(1), v_i(2), \ldots, v_i(t-1))$ and $v_i(j) = r_i(j) + a_i(j)$, i.e. summation of number of real transaction and fake transactions.

  So the allocation algorithm $\mathcal{M}$, takes input $R(t)$ before round t and outputs number of impression allocated. $\mathcal{M}(R(t)) = (n_1(t), n_2(t), \ldots, n_m(t))$ and the revenue generated within T days by real transactions is $R(\mathcal{M}, T) = \sum_{i=1}^{m} \sum_{t=1}^{T} r_i(t) p_i(t)$

4.2 Building Rationality Model

So the behavior model is function f, $a_i(t) = f(H_i(t))$, takes input $H_i(t)$ before day t and outputs $a_i(t)$, number of fake transaction seller i will make on day t

The neural network will first do regression using square mean loss and apply softmax layer to the output and we get the probability predict of manipulation behaviour, and use cross entropy loss for classification. Input is H_i^t that contains 4 parameters. The paper tests on two types of networks - conventional Convolutional layers and ResNet blocks. Both networks perform quite similar. Variations :

- Classification is done for two classes - engages in fake transactions or not
- Classification is done for more than two classes - depending on parameters to which extent it involves in fake transactions
- Also, training the network keeping no of transactions as sum of real and fake. and still the accuracy is similar when the network knew the real and fake transactions

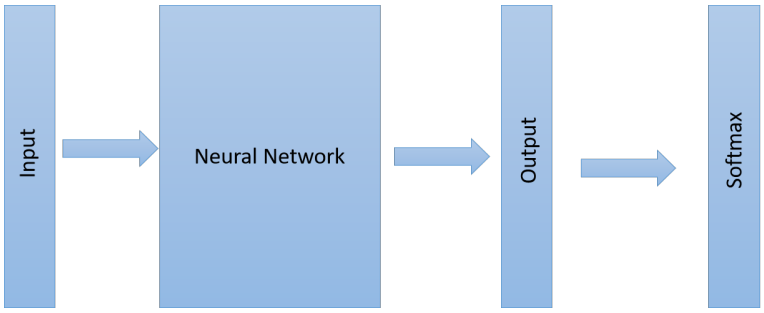

Figure 1: Behaviour Model

4.3 General Process of overall Mechanism

1. The allocation should be feasible at any day, number of impression allocation should not exceed
2. On day t , for each seller i, predict the number of fake transactions, estimate $p_i(t)$, real conversion rate $cr_i(t)$ and feigned conversion rate $fcr_i(t)$
   - For estimating price, we assume price to be random variable drawn from Gaussian distribution
3. Run the allocation algorithm $\mathcal{M}$ to allocate impression and calculate real revenue

4.4 MDP Formulation

- For each day t we have a state $St = (v_1(t), n_1(t), p_1(t), ..., v_m(t), n_m(t), p_m(t))$, where for each seller $i = 1, \ldots, m$, $v_i(t)$ is the number of total transactions(real and fake), $n_i(t)$ is the number of buyer impressions and $pi(t)$ is the price that the seller selects.
- Actions corresponding to this state is allocation of n^t impression among m sellers.

- Payoff of an action from a state is the real revenue corresponding to the impression allocation of that action.

Paper uses Deep Deterministic Policy gradient (DDPG) to solve the MDP (maximize the real revenue), treating impression allocation problem as division of a continuous resource

5 Experimental Results

Greedy, CVR, Mixed, Uniform Allocation, DDPG, Platform's implementation

- DDPG outperforms all of them in terms of generating real revenue
- *alpha* is a parameter to control maximum impression allocation to maintain feasibility. As α increases, real revenue also increases. As DDPG outperforms all.
- fake revenue results