Reinforcement Learning for Customer Interaction

Zijie Xia zx2276

B11: Customer Interaction — Finance Product Sales & Marketing Strategy

2/12/20

1. Background (Review of Related Literature):

Targeted marketing is sequential in nature. It means marketing decisions are made over time, and the interactions among them affect the overall, cumulative profits. For example, services to be provided by a company will be affected by customers' actions before. Also, in many commercial applications, a company or organization interacts with many customers concurrently. It means interactions with different customers occur in parallel. As a result, it is imperative to learn online from partial interaction sequences, so that information acquired from one customer is efficiently assimilated and applied in subsequent interactions with other customers.

Our goal is to maximize the cumulative rewards for each customer using reinforcement learning, given their history of interactions with the company. This setting differs from traditional reinforcement learning paradigms, due to the sequential and concurrent nature of the customer interactions.

Reinforcement learning applied to sequential marketing problems is shown in [1]. It proposes three methods based on reinforcement learning, direct (Sarsa), indirect (Monte-Carlo) and semidirect methods. Semidirect methods are essentially direct methods but mirror some aspects of the indirect methods by means of sampling and use of estimated rewards. The experimental results show that semi-direct methods are effective in reducing the amount of computation necessary to attain a given level of performance, and often result in more profitable policies.

[2] supplements [1] by providing some sampling methods used in batch reinforcement learning to lighten the load of data size. Like in Q-sampling, only those states are selected that conform to the condition that the action taken in the next state is the best action with respect to the current estimate of the Q-value function.

Improvements about preprocessing are made in [3]. [1] uses multivariate linear regression tree method to implement function approximation before employing reinforcement learning methods, while [3] uses neural network to approximate value function in preprocessing progress.

However, concurrency is not realized in [1][2] and [3] although all of the models can be applied in sequential marketing. [4] presents a framework for concurrent reinforcement learning, using a variant of temporal-difference learning to learn efficiently from partial interaction sequences. It allows for interactions to occur asynchronously, then develops a concurrent variant of temporal difference (TD) learning which bootstraps online from partial interaction sequences. Especially when large numbers of interactions occur simultaneously, the feedback from one customer can be assimilated and applied immediately to other customers.

These algorithms are based on cost-sensitive learning. Basic information about cost-sensitive learning is discussed in the paper [5]. Algorithms about reinforcement learning can be referred to [6]. Some other methods to solve sequential customer interaction/segmentation problems can be found in these resources [7].

2. Introduction to the Project:

In a Marko Decision Process, the environment is assumed to be in one of a set of possible states, at any point in time. At each time clock, the environment is in some state, the learner takes

one of several possible actions, receives a finite reward (i.e., a profit or loss), and the environment makes a transition to another state. Our first goal is to maximize the cumulative rewards for each customer using reinforcement learning and find the optional action in each state, given customers historical interactions with the company.

To reach the goal, I will refer to [6] to research further on reinforcement learning methods, like Q-learning, Sarsa, Monte-Carlo and TD learning. A number of controlled experiments will be conducted to compare the performance of competing approaches of direct, semi-direct and indirect methods of reinforcement learning, in the domain of sequential targeted marketing. For example, the performance of the three methods will be compared in two set-ups: few features are accessible and all features are accessible. Then we can conclude the influences of the integrity of features to different models. Also, we can compare computational expenses of three models. When conducting the experiments, we need to choose a good base learning module using neural network.

The second goal is to realize the concurrency of models. We will evaluate Concurrent TD in a high dimensional commercial simulator against non-bootstrapping (Monte-Carlo), non-online (batch TD), and non-sequential (Contextual Bandit) algorithms respectively. By implementing and comparing these models, we will be curious about how important it is to learn online, and to bootstrap, when operating at different levels of concurrency (i.e. when interacting with different numbers of customers in parallel). Then, we will analyze how important it is to learn from delayed rewards, compared to learning from immediate rewards in a concurrent setting.

Finally, we can achieve a reinforcement learning model to solve sequential and concurrent customer interaction problems.

3. Introduction to the Dataset:

All of the training data we will use are generated from the donation data set from KDD Cup 1998. The training and validation data portion of the original data set contains data for approximately 100 thousand selected individuals and about 500 features. This data set concerns direct mail promotions for soliciting donations, and contains demographic data as well as promotion history of 22 campaigns, conducted monthly over an approximately two-year period. The campaign information contained includes whether an individual was mailed or not, whether he or she responded or not and how much was donated. Additionally, if the individual was mailed, the date of the mailing is available (month and year), and if the individual has responded, the date of the response is available.

Based on the campaign information in the data, we will generate a number of temporal features that are designed to capture the state of that individual at the time of each campaign. To reach the first goal, we will maximize the cumulative rewards for each customer and find the optional action in each state. Then the decision whether the next mail will be sent to some customer can be made. To discuss the concurrency of models, we will truncate the data set before implementing Concurrent TD.

4. Plan:

01/24/20 - 02/14/20: Define the selected topic; Do research on the topic, including reviewing literatures and searching datasets; Specify what to do and make a plan.
02/15/20 - 03/13/20: Do more research on reinforcement learning algorithms; Clean the dataset, Implement and compare direct, indirect and semi-direct models on the given dataset.
03/14/20 - 04/17/20: Do more research on concurrency reinforcement learning; Implement and evaluate Concurrent TD in a high dimensional commercial simulator against non-bootstrapping

(Monte-Carlo), non-online (batch TD), and non-sequential (Contextual Bandit) algorithms respectively.
04/18/20 - 05/08/20: Try some other methods, like SVM, LSTM, on sequential and even concurrent market; Arrange the contents and prepare the presentation.

Reference:
[1] Abe, N., Pednault, E., Wang, H., Zadrozny, B., Fan,W., and Apte, C. (2002). Empirical comparison of various reinforcement learning strategies for sequential targeted marketing. In International Conference on Data Mining, pages 3–10.
[2] Pednault, Edwin, Naoki Abe, and Bianca Zadrozny. "Sequential cost-sensitive decision making with reinforcement learning." Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. 2002.
[3] Gomez-Perez, G., Martin-Guerrero, J. D., SoriaOlivas, E., Balaguer-Ballester, E., Palomares, A., and Casariego, N. (2008). Assigning discounts in a marketing campaign by using reinforcement learning and neural networks. Expert Systems with Applications, (doi: 10.1016/j.eswa.2008.10.064)
[4] Silver, David, et al. "Concurrent reinforcement learning from customer interactions." International Conference on Machine Learning. 2013.
[5] Elkan, Charles. "The foundations of cost-sensitive learning." International joint conference on artificial intelligence. Vol. 17. No. 1. Lawrence Erlbaum Associates Ltd, 2001.
[6] Sutton, Richard S., and Andrew G. Barto. Reinforcement learning: An introduction. MIT press, 2018.
[7] Tsiptsis, Konstantinos K., and Antonios Chorianopoulos. Data mining techniques in CRM: inside customer segmentation. John Wiley & Sons, 2011.