Multi-Armed Bandits and Crowdsourcing

Submitted as part of Honours Project (Semester 2)

Debojit Das
Machine Learning Lab

Sujit P Gujar
Machine Learning Lab

Abstract

In this report, we talk about Multi-Armed Bandits and Crowdsourcing, and how they are related. We also study Diffential Game Theory. Then we summarise a few papers on these topics. We give a brief overview of these papers. Finally we talk about our problem statement idea.

Contents

1

Introduction

A crowdsourcing platform comprises of a lot of workers. Given a task, every worker has an expected return on the task. The return may vary from task to task (depending on factors like area of expertise, hardeness of task, etc.) for every person. Also, the return may vary from person to person for any given task, depending on factors like how often a person makes mistakes and what their area of expertise is. Interestingly, even the same worker doing a particular task may give varying returns due to the stochastic nature of humans. This is why it is justified to think of a crowdsourcing system as a MAB where every worker is an arm of the bandit.

We will now study some useful terms that will be useful in the course of the report.

1.1 Multi-Armed Bandits (MAB)

The MABs we consider consist of arms, with unknown *stochastic rewards*, and every turn we consider which arm to play so as to maximise the total reward at the end. We can choose to play an arm that has the highest expected reward at that point of time, thereby exploiting it; or we can choose to explore the other arms in hopes for finding their expected reward to be higher.

In its most basic formulation, a K-armed bandit problem is defined by random variables $X_{i,n}$ for $1 \leq i \leq K$ and $n \geq 1$, where each i is the index of a gambling machine (i.e., the arm of a bandit). Successive plays of machine i yield rewards $X_{i,1}$,$X_{i,2}$,... which are independent and identically distributed according to an unknown law with unknown expectation μ_i . Independence also holds for rewards across machines; i.e., $X_{i,s}$ and $X_{j,t}$ are independent (and usually not identically distributed) $\forall 1 \leq i < j \leq K \ \forall s, t \geq 1$.[3]

Given a MAB formulation, the goal is to come up with an algorithm that decides the order in which the arms should be pulled in order to maximise the total expected reward.

We define *regret* as the expected loss in reward on following our algorithm as compared to the reward we would have obtained by following some optimal (in terms of total expected reward) strategy. More formally, for a stochastic MAB, the regret $R(A)$ after n rounds on following algorithm A is given by

$$R_A(n) = \mu^* n - \sum_{i=1}^{K} \mu_i T_i(n)$$

, where $\mu^* = \max_{1 \leq i \leq K} \mu_i$, μ_i is the expected reward on pulling arm i, and $T_i(n)$ is the number of times arm i was pulled in n rounds. Our objective is essentially to minimise the expected regret after T rounds, i.e., choose A that minimises $R_A(T)$.

At any given point of time, we can decide to either play the arm which has the highest average reward till that time hence *exploiting* our current knowledge, or choose to *explore* the other arms to get a better estimate since on updation other arms may turn out to have higher expected reward.

1.2 Crowdsourcing

Crowdsourcing is a task allocation concept wherein a task can be outsourced to workers who are chosen from a crowd instead of being performed by a designated agent, which is often suitable for tasks that are trivial for humans but difficult for computers, such as classification tasks. Compared with traditional task allocation, the advantages of crowdsourcing include the following: faster completion speed, lower costs, higher accuracy, and completion of tasks that computers cannot perform.

1.3 MABs in Crowdsourcing

There are a lot of parallels between MABs and Crowdsourcing. Every arm of the bandit being pulled can be thought of as an agent performing a task. Every agent performs differently, similar to how every arm has a unique expected reward associated with it. This is why MABs find a lot of use in crowdsourcing settings.

1.4 Heterogeneous Tasks

Often the assumption that a task is identical to every other task will not be a fair assumption. In some settings, tasks can be of varying difficulty, or may require different skill-sets. We say that the tasks are heterogeneous.

1.5 Heterogeneous Workers

Similarly, often the assumption that a worker is identical to every other worker will not be a fair assumption. Workers may have different skills, and some may be more efficient than others at a given task. Then we say that the workers are heterogeneous.

1.6 Differential Game Theory (DGT)

In game theory, differential games are a group of problems related to the modeling and analysis of conflict in the context of a dynamical system. More specifically, a state variable or variables evolve over time according to a differential equation. Early analyses reflected military interests, considering two actors-the pursuer and the evader-with diametrically opposed goals.

Differential games are related closely with optimal control problems. In an optimal control problem there is single control $u(t)$ and a single criterion to be optimized; differential game theory generalizes this to two controls $u_1(t), u_2(t)$ and two criteria, one for each player. Each player attempts to control the state of the system so as to achieve its goal; the system responds to the inputs of all players.[17]

This report is divided into three parts and appendix. In part 1, we summarize MAB-based papers. In part 2, we summarize crowdsourcing-based papers. In part 3, we summarize DGT-based papers.

Part 1: MAB-based Papers

2

Knapsack Based Optimal Policies for Budget–Limited Multi–Armed Bandits

[Read8]

2.1 Basic Details

This paper, authored by Long Tran-Thanh, Alex Rogers, Archie Chapman and Nicholas R. Jennings, was published in AAAI Conference on Artificial Intelligence in 2012.

2.2 Problem Addressed

In budget–limited MAB problems, the learner's actions are costly and constrained by a fixed budget. An optimal exploitation policy may not be to pull the arm with the highest reward repeatedly, as is the case in other variants of MAB, but rather to pull the sequence of different arms that maximises the agent's total reward within the budget. This difference from existing MABs means that new approaches to maximising the total reward are required. The paper considers the setting that pulling an arm has a cost, and both the exploration and exploitation phases are limited by a single budget.

2.3 Solution Proposed

The authors propose two new learning algorithms, called KUBE (for knapsack–based upper confidence bound exploration and exploitation) and fractional KUBE, that do not explicitly separate exploration from exploitation. They explore and exploit at the same time by adaptively choosing which arm to pull next, based on the current estimates of the arms' rewards.

At each time step, KUBE calculates the best set of arms that provides the highest total upper confidence bound of the estimated expected reward, and still fits into the residual budget, using an unbounded knapsack model to determine this best set. Since unbounded knapsack problems are NP–hard, the algorithm uses an efficient approximation method called the density–ordered greedy approach in order to estimate the best set.

Fractional KUBE also estimates the best set of arms that provides the highest total upper confidence bound of the estimated expected reward at each time step, and uses the frequency that each arm occurs within this approximated best set as a probability to randomly pull the arms. However, instead of using the density–ordered greedy to solve the underlying unbounded knapsack problem, fractional KUBE relies on a computationally less expensive approach, namely the fractional relaxation based algorithm.

2.4 Claims

- They claim that KUBE and fractional KUBE proveably achieve a O(lnB) theoretical upper bound on the regret, where B is the budget limit.
- They demonstrate that with an increased computational cost, KUBE out-performs fractional KUBE in the experiments. They also show that while both algorithms achieve logarithmic regret bounds, the buget-limited ϵ-first approaches fail to do so.
- They say that KUBE typically outperforms its fractional counterpart by up to 40%, however, this results in an increased computational cost (from O(K) to O($KlnK$)).

2.5 Previous Work

The standard MAB problem was originally proposed by Robbins (1952). In the standard MAB, this trade–off has been effectively balanced by decision–making policies such as upper confidence bound (UCB) and ϵ_n–greedy ([3]). A number of researchers have focused on MABs with budget constraints, where arm–pulling is costly and is limited by a fixed budget (Antos et al., 2008; [5]; Guha and Munagala, 2007). The theoretical best possible regret bound is typically a logarithmic function of the number of pulls. ([11])

2.6 Novel Ideas

In many settings, it is not only the exploration phase, but the exploitation phase that is also limited by a cost budget. They proposed a budget-limited ϵ-first approach for the budget–limited MAB. This splits the overall budget B into two portions, the first ϵB of which is used for exploration, and the remaining $(1-\epsilon)B$ for exploitation. They proposed KUBE and fractional-KUBE.

3

The Perils of Exploration under Competition: A Computational Modeling Approach

[Read1]

3.1 Basic Details

This paper, authored by Guy Aridor, Kevin Liu, Aleksandrs Slivkins and Zhiwei Steven Wu, was published in ACM Conference on Economics and Computation in June 2019.

3.2 Problem Addressed

When multiple systems are competing for the same market of users, exploration may hurt a system's reputation in the near term, with adverse competitive effects. In particular, a system may enter a "death spiral" (i.e., one firm attracts new customers at a lower rate than the other, and falls behind in terms of performance because the other firm has more customers to learn from, and this gets worse over time until most new customers go to the other firm), when the short-term reputation cost decreases the number of users for the system to learn from, which degrades the system's performance relative to competition and further decreases the market share.

While some bandit algorithms are traditionally considered better than others in the literature, competition may disincentivize the adoption of the better algorithms. This may be affected by the intensity of competition. They investigate these issues via extensive numerical experiments in a stylized duopoly model.

3.3 Solution Proposed

This is an experimental paper. The authors consider a *permanent duopoly* in which both firms start at the same time, as well as temporary monopoly: a duopoly with a first-mover. The intensity of competition in the model varies from *permanent monopoly* (just one firm) to *incumbent* (the first-mover in temporary monopoly) to permanent duopoly to *entrant* (late-arriver in temporary monopoly).

They focus on three classes of bandit algorithms, ranging from more primitive to more sophisticated:

- greedy algorithms that do not explicitly explore
- algorithms that separate exploration and exploitation
- algorithms that combine the two

3.4 Findings

They find that in the permanent duopoly, competition incentivizes firms to choose the greedy algorithm, and even more so if the firm is a late arriver in a market. This algorithm also prevails under monopoly, simply because it tends to be easier to deploy. Whereas the incumbent in the temporary monopoly is incentivized to deploy a more advanced exploration algorithm. As a result, consumer welfare is highest under temporary monopoly. They find strong evidence of the "death spiral" effect mentioned earlier; this effect is strongest under permanent duopoly.

They investigate the *first-mover advantage* phenomenon in more detail. Being first in the market gives free data to learn from (a *data advantage*) as well as a more definite, and possibly better reputation compared to an entrant (a *reputation advantage*). They run additional experiments so as to isolate and compare these two effects. They find that either effect alone leads to a significant advantage under competition. The data advantage is larger than reputation advantage when the incumbent commits to a more advanced bandit algorithm.

They also investigate how algorithms' performance in isolation (without competition) is predictive of the outcomes under competition. They find that mean reputation is sometimes not a good predictor.

3.5 Previous Work

The study of competition vs. exploration has been initiated in [14]. The setting is also closely related to the dueling algorithms framework [8]. In *dueling bandits* (e.g., [18]), an algorithm sets up a duel between a pair of arms in each round, and only learns which arm has won. The interplay between exploration, exploitation and incentives has been studied in other scenarios: incentivizing exploration in a recommendation system, e.g., [7], dynamic auctions, online ad auctions, e.g., [4], human computation [15], and repeated auctions.

Their work is also related to a longstanding economics literature on competition vs. innovation, e.g., [1]. Whether data gives competitive advantage has been

studied theoretically and empirically. The first-mover advantage has been well-studied in other settings in economics and marketing, see survey.

3.6 Novel Ideas

The authors vary the intensity of competition while studying the effect of data advantage as well as recommendation advantage. While these have been studied separately, this is the first time a paper has combined them.

4

Fiduciary Bandits

[Read2]

4.1 Basic Details

This paper, authored by Gal Bahar, Omer Ben-Porat, Kevin Leyton-Brown and Moshe Tennenholtz, was published in ArXiv in 2019.

4.2 Preliminaries

Incentive Compatible (IC): A mechanism is incentive compatible (IC) implies that following its recommendations constitutes an equilibrium. So, when given a recommendation and given that others follow their own recommendations, an agent's best response is to follow her own recommendation.

Ex-Ante Individually Rational (EAIR): A mechanism is EAIR if any probability distribution over arms that it selects has expected reward that is always at least as great as the reward of the default arm, both calculated based on the recommender's knowledge.

Ex-Post Individually Rational (EPIR): The EPIR condition requires that a recommended arm must never be a priori inferior to the default arm given the planner's knowledge.

4.3 Problem Addressed

Recommendation systems face exploration-exploitation tradeoffs as the system can only learn about the desirability of new options by recommending them to some user. Such systems can be modeled as MAB settings. However, users are self-interested and cannot be made to follow recommendations. The paper addresses whether exploration can be performed in a way that respects agents' interests-i.e., by a system that acts as a fiduciary. The authors introduce a model in which a recommendation system faces an exploration-exploitation tradeoff under the constraint that it can never recommend any action that it knows yields lower reward in expectation than an agent would achieve if it acted alone.

4.4 Solution Proposed

They develop an MAB algorithm, Fiduciary Explore & Exploit (FEE), to address the problem. Then they modify it to M_{EPIR} for stronger properties.

4.5 Claims

Claims about FEE:

- FEE is Incentive Compatible (IC) and Ex-Ante Individually Rational (EAIR)
- FEE obtains the highest possible social welfare by an EAIR mechanism up to a factor of $o(\frac{1}{n})$, where n is the number of agents

The authors claim that M_{EPIR} is asymptotically IC and Ex-Post Individually Rational (EPIR).

4.6 Previous Work

Kremer et al. [10] is the first work that investigated the problem of incentivizing exploration. Cohen and Mansour [6] extended this optimality result to several arms under further assumptions. This setting has also been extended to regret minimization, social networks, and heterogeneous agents.

Liu et al. [12] aim at treating similar arms similarly and Joseph et al. demand that a worse arm is never favored overt a better one despite a learning algorithm's uncertainty over the true payoffs.

4.7 Novel Ideas

The authors introduce the idea of individual rationality to the previous works on incentivizing exploration. They develop FEE which is IC and EAIR. They also extend the idea of fairness amongst arms to EAIR.

5

Stochastic Bandits with Delay-Dependent Payoffs

[Read3]

5.1 Basic Details

This paper, authored by Leonardo Cella and Nicolò Cesa-Bianchi, was published in International Conference on Artificial Intelligence and Statistics in March, 2020.

5.2 Problem Addressed

The paper addresses the problem of designing a recommendation system for a music streaming platform. In that case, the expected reward of an arm depends on the number of rounds that have passed since the arm was last pulled.

5.3 Solution Proposed

They introduce a simple nonstationary stochastic bandit model, B2DEP, in which the expected reward of an arm is a bounded nondecreasing function of the number of rounds that have passed since the arm was last selected by the policy. They assume each arm has an unknown baseline payoff expectation (equal to the expected payoff when the arm is pulled for the first time) and an unknown delay parameter. Hence, if the arm was pulled recently, then the expected payoff may be smaller than its baseline value.

B2DEP is different from UCB because UCB performs a lot of switches if the size of the gaps in the expected rewards is low. Switching may be expensive in some scenarios. B2DEP provides a bound on the number of switches.

They also run simulations of B2DEP against UCB for comparing performance given the size of the gap in their expected rewards.

5.4 Claims

Finding an optimal policy for the problem is NP-hard even when all model parameters are known.

This algorithm has a distribution-free regret bounded by $\sqrt{kT}$. It also has a bound $O(klnlnT)$ on the number of switches, irrespective of the size of the gaps in their expected rewards. Hence, it works better than UCB when the gap is sufficiently small.

5.5 Previous Work

Their model can be compared to nonstationary models, such as rested bandits (Gittins, 1979) and restless bandits (Whittle, 1988). Their setting is a variant of the model introduced by Kleinberg and Immorlica (2018). Pike-Burke and Grunewalder (2019) consider a setting in which the expected reward functions are sampled from a Gaussian Process with known kernel. A special case of the model is investigated in the work by Basu et al. (2019). Similarly to (Radlinski et al., 2008; Kveton et al., 2015a; Katariya et al., 2016) the strategies in the paper learn rankings of the actions.

5.6 Novel Ideas

In this paper, the block length is unknown, and the greedy strategy is not defined in terms of the agent's delay configuration, unlike previous work. Also, in this paper, the optimal number of elements in the ranking are also learned.

6

Dueling Bandits: From Two-dueling to Multi-dueling

[Read4]

6.1 Basic Details

This paper, authored by Yihan Du, Siwei Wang and Longbo Huang, was published in AAMAS in May, 2020.

6.2 Problem Addressed

The paper studies a general multi-dueling bandit problem, where an agent compares multiple options simultaneously and aims to minimize the regret due to selecting suboptimal arms. It starts with the two-dueling bandit setting and proposes efiicient algorithms.

6.3 Solution Proposed

- The authors improve the Doubler algorithm to propose DoublerBAI algorithm for best arm identification in a two-dueling problem setting and give a finite-time regret analysis
- They propose MultiSBM-Feedback algorithm, which is an improvement on the MultiSBM algorithm, using Single Bandit Machine for a two-dueling problem setting and give a finite-time regret analysis
- They propose MultiRUCB for the general multi-dueling bandit problem, with the key idea of exploiting as much information as possible from one pull, and give a finite-time regret analysis
- Based on both synthetic and real-world datasets, the paper empirically demonstrates that the proposed algorithms outperform existing algorithms.

6.4 Claims

- DoublerBAI achieves a $O(lnT)$ regret bound
- MultiSBM-Feedback has an optimal $O(lnT)$ regret and also reduces the constant factor by almost a half compared to benchmark results
- MultiRUCB also achieves an $O(lnT)$ regret bound and the bound tightens as the capacity of the comparison set increases

6.5 Previous Work

The dueling bandits problem [19] has been used in applications involving implicit or subjective (human) feedback, such as information retrieval and recommendation systems. This multi-dueling bandit model has been used in information retrieval, ranking algorithms, online ranker evaluation methods.

The algorithms DoublerBAI and MultiSBM-Feedback build upon the Doubler and MultiSBM algorithms resp. in [2], which reduces the dueling bandits problem to the conventional stochastic MAB problem.

6.6 Novel Ideas

They improve upon the existing Doubler and MultiSBM algorithms and provide DoublerBAI and MultiSBM-Feedback algorithms resp. They also provide regret analysis for both. While there have been previous work on the multi-dueling bandit problem, this is the first work to provide a finite-time regret analysis for the general multi-dueling bandit problem. They also conduct experiments based on both the synthetic and real-world datasets.

Part 2: Crowdsourcing-based Papers

7

Understanding Crowdsourcing Systems from a Multiagent Perspective and Approach

[Read6]

7.1 Basic Details

This paper, authored by Jiuchuan Jiang, Bo An, Yichuan Jiang, Donghui Lin, Zhan Bu, Jie Cao and Zhifeng Hao, was published in ACM Transactions on Autonomous and Adaptive Systems in July 2018.

7.2 Problem Discussed

A multiagent system is a computing system that is composed of a set of agents that perform tasks. The multiagent perspective mainly considers what multiagent approaches can offer to real systems (including crowdsourcing systems) and how multiagent technologies help analyze these systems.

This is a review paper and discusses the following things:

1. The multiagent perspective can be used for conducting a comprehensive survey on the state of the art of crowdsourcing
2. The multiagent approach can bring about concrete enhancements for crowdsourcing technology and inspire future research directions that enable crowdsourcing research to overcome the typical challenges in crowdsourcing technology
3. The advantages and disadvantages of the multiagent perspective by comparing it with two other popular perspectives on crowdsourcing: the *business perspective* and the *technical perspective*

7.3 Solution Proposed

This paper is a general survey on the state of the art of a comprehensive set of crowdsourcing systems from a multiagent perspective and future research directions for overcoming existing typical technology challenges by using a multiagent approach, which can correlate the research on crowdsourcing and multiagent systems and inspire an interdisciplinary research between the two areas. For crowdsourcing researchers, this article presents a new viewpoint for understanding and investigating crowdsourcing systems. For multiagent system researchers, this article motivates them to apply multiagent technologies to solve real problems in crowdsourcing systems. It provides a systematic review of the association between key elements and key processes in crowdsourcing systems.

7.4 Novel Ideas

The authors presents a novel multiagent perspective and approach to understanding crowdsourcing systems, which can be used to correlate the research on crowdsourcing and multiagent systems and inspire possible interdisciplinary research between the two areas. They compare their multiagent perspective of crowdsourcing with two other prevalent perspectives: the business perspective and technical perspective. The business perspective mainly considers the business behavior and business principles in crowdsourcing markets. The technical perspective considers the technologies for the development of efficient crowdsourcing systems.

7.5 Previous Work

Most previous surveys have mainly reviewed a single aspect of crowdsourcing or the application of crowdsourcing in one specific domain, such as the survey of tasks in crowdsourcing [13], the survey of crowdsourcing for the data mining domain, the survey of crowdsourcing on the world-wide-web, the survey of the future of crowdsourcing [9], the survey of the difference between crowdsourcing and human computation, and the survey of crowdsourcing in software engineering. Although a few surveys have

attempted to present a more general review of crowdsourcing [20], they have only reviewed the definitions of crowdsourcing and typical crowdsourcing systems.

7.6 Key Contributions

The paper provides a general and macroscopic review of various elements and processes of crowdsourcing systems.

Each crowdsourcing application includes the following key elements:

- tasks
- requesters
- system platforms
- workers

When a crowdsourcing system wants to perform a task, the following processes are necessary from the time the task comes to the system to the time the task is completely finished:

- pre-execution process
- execution process
- post-execution process

It draws the following comparison between crowdsourcing systems and multiagent systems:

- Tasks in crowdsourcing systems are similar to tasks in multiagent systems
- Requesters in crowdsourcing systems are similar to hosts in multiagent systems
- System platforms in crowdsourcing systems are similar to system mechanisms in multiagent systems
- Workers in crowdsourcing systems are similar to agents in multiagent systems
- Pre-execution process in crowdsourcing has the same job that task analysis and allocation has in multiagent systems
- Execution process in crowdsourcing has the same job that task execution has in multiagent systems
- Post-execution process in crowdsourcing has the same job that task feedback has in multiagent systems

8

Mechanism Design for Crowdsourcing Markets with Heterogeneous Tasks

[Read5]

8.1 Basic Details

This paper, authored by Gagan Goel, Afshin Nikzad and Adish Singla, was published in AAAI Conference on Human Computation and Crowdsourcing in 2014.

8.2 Problem Addressed

The paper mainly talks about how one can design market mechanisms for crowdsourcing when the tasks are heterogeneous and workers have different skill sets.

There is a requester who has a set of heterogeneous tasks and a limited budget. For each task, there is a fixed utility that the requester achieves if that task gets completed. To do the tasks, there is a pool of workers. Each worker has certain skill sets and interests which makes her eligible to do only certain tasks, and not all. Each worker has a cost, which is the minimum amount she is willing to take for doing a task. This minimum cost is assumed to be a private information of the worker, and is same for all the tasks. The goal is to design an auction mechanism that is:

- incentive compatible in the sense that it is truthful for agents to report their true cost
- picks a set of workers and assigns each to a task such that the utility of the requester is maximized
- budget feasible, i.e., the total payments made to the workers does not exceed the budget of the requester

8.3 Solution Proposed

The authors begin by designing a deterministic mechanism which they call as TM-UNIFORM (i.e. Truthful Matching using Uniform Rate).

To improve the approximation guarantees of their mechanism, they make a connection of a subroutine in TM-UNIFORM to the problem of online bipartite-matching and Adwords allocation problem. They use this connection to design a randomized mechanism TM-RANDOMIZED.

Finally, they carry out extensive experimentation on a realistic case study of Wikipedia translation project using Mechanical Turk workers. Their results demonstrate the practical applicability of their mechanism. They also do simulations on synthetic data to evaluate the performance of our mechanisms on various parameters of the problem.

8.4 Claims

They claim that even though TM-UNIFORM is not fully truthful, it satisfies truthfulness in a weaker form, which they call oneway-truthfulness. By this property, workers only have incentive to report costs lower than their true

cost. It achieves an approximation factor of 3. Then, they design a new payment rule for TM-UNIFORM, that makes it fully truthful.

They claim that TM-RANDOMIZED has an approximation factor of 2.58. However this mechanism was only shown to be truthful in large markets, that is, the incentive to deviate goes down to zero as the market grows larger.

They say that these mechanisms easily extend to work for many-to-many matchings as well. In the many-to-many setting, they can handle the case when the utility of doing a task is a non-decreasing concave function of the number of times that the task is done.

8.5 Previous Work

The most similar work to that of theirs is the design of budget-feasible mechanisms, initiated in Singer (2010). Subsequent research in this direction (Chen, Gravin, and Lu 2011; Bei et al. 2012; Singer 2011; Singla and Krause 2013a) has improved the current results and extended them to richer models and applications. The problem of knapsack utility functions with matching constraints has been studied by Singer (2010).

Motivated from crowdsourcing settings, budget-limited multi-armed bandits have been studied (Badanidiyuru, Kleinberg, and Slivkins 2013; [16]).

8.6 Novel Ideas

Their model generalizes the results of budget feasible mechanism design by extending them to problems with matching constraints, though they consider a simpler utility functions (they consider knapsack and nondecreasing concave utility functions). They make use of the mathematical structure of matchings in bipartite graphs and the assumption of large markets to design polynomial-time deterministic and randomized mechanisms with much better approximation ratios as compared to what is given by the current known results.

Part 3: DGT-based papers

9

Sequential Linear Quadratic Method for Differential Games

[Read7]

9.1 Basic Details

This paper, authored by H. Mukai, Akio Tanikawa, Ilker Tunay and I.N. Katz, was published in IEEE in 2000.

9.2 Problem Addressed

The authors present a numerical method for computing a Nash solution to a zero-sum differential game for a general nonlinear system based on a sequential linear-quadratic(SLQ) approximations. In the scenario that a force chooses to deviate from Nash solution, the authors also present an algorithm that the other force can follow for better results.

Each unit on either force has two objectives:

- to reach its specified fixed target
- to reduce the number of enemy platforms while preserving the number of its own

9.3 Solution Proposed

They propose the SLQ iterative algorithm which uses Ricatti equations and gives us the control and trajectory for both forces (red and blue) in the case of Nash Solution.

In the case that a force (say, red) chooses to deviate from the Nash solution, they propose the "Game Theoretic Controller" method which provides controls and trajectories for different time intervals, and these time intervals depend on how much deviation from Nash solution we can tolerate.

They also provide simulations of the performance of the SLQ iterative algorithm for different game states. They also experimentally show how the performance of the Game Theoretic Controller method under noisy observations and model mismatch.

9.4 Claims

- They claim that SLQ iterative algorithm gives Nash solution
- The optimum cost is only slightly sensitive to noise. The numbers of platforms and the respective distances to the final targets are even less sensitive
- The nonlinear controller is more realistic than the linear controller

9.5 Previous Work

This paper does not talk about any previous papers. It only cites some books for using their equation.

9.6 Novel Ideas

The authors develop algorithms SLQ iterative algorithm and Game Theoretic Controller method. They also run simulations to show the performance of SLQ method for different game states, and to show the performance of the Game Theoretic Controller under noisy observation and model mismatch.

10

Brief Overview

Paper	Appeared in	Problem Statement	Solution Proposed
Knapsack Based Optimal Policies for Budget–Limited Multi–Armed Bandits	AAAI Conference on Artificial Intelligence, 2012	Finding an algorithm that gives a sequence of arms pulled, where each arm has a cost, such that it minimises regret within budget constraints	They propose algorithms KUBE and fractional-KUBE
The Perils of Exploration under Competition: A Computational Modeling Approach	ACM Conference on Economics and Computation, 2019	While some bandit algorithms are traditionally considered better than others in the literature, does competition incentivize the adoption of the better algorithms? How is this affected by the intensity of competition?	This is an experimental paper. They consider a permanent duopoly in which both firms start at the same time, as well as temporary monopoly. The intensity of competition in the model varies from permanent monopoly to incumbent to permanent duopoly to entrant
Fiduciary Bandits	ArXiv, 2019	They introduce a model in which a recommendation system faces an exploration-exploitation tradeoff under the constraint that it can never recommend any action that it knows yields lower reward in expectation than an agent would achieve if it acted alone	They propose algorithms FEE and M_{EPIR}
Stochastic Bandits with Delay-Dependent Payoffs	International Conference on Artificial Intelligence and Statistics, 2020	Designing a recommendation system for a music streaming platform where the expected reward of an arm depends on the number of rounds that have passed since the arm was last pulled	They propose algorithm B2DEP
Dueling Bandits: From Two-dueling to Multi-dueling	AAMAS, 2020	Proposing algorithms for two-dueling and multi-dueling bandits and providing finite-time regret bound analysis	They propose algorithms Doubler-BAI and MultiSBM-Feedback for two-dueling bandits, both having regret bound $O(lnT)$. They propose algorithm MultiRUCB which also has a regret bound $O(lnT)$.

Table 1: **MAB-based Papers**

Paper	Appeared in	Problem Statement	Solution Proposed
Understanding Crowdsourcing Systems from a Multiagent Perspective and Approach	ACM Transactions on Autonomous and Adaptive Systems, 2018	The multiagent perspective can be used for conducting a comprehensive survey on crowdsourcing. The multiagent approach can bring about enhancements for crowdsourcing technology and inspire future research directions that enable crowdsourcing research to overcome challenges in crowdsourcing technology	This paper is a survey on crowdsourcing systems from a multiagent perspective and future research directions for overcoming existing technology challenges by using a multiagent approach, which can correlate the research on crowdsourcing and multiagent systems and inspire an interdisciplinary research between the two areas.
Mechanism Design for Crowdsourcing Markets with Heterogeneous Tasks	AAAI Conference on Human Computation and Crowdsourcing, 2014	The paper mainly talks about how one can design market mechanisms for crowdsourcing when the tasks are heterogeneous and workers have different skill sets	They propose algorithms TM-UNIFORM and TM-RANDOMIZED

Table 2: **Crowdsourcing-based Papers**

Paper	Appeared in	Problem Statement	Solution Proposed
Sequential Linear Quadratic Method for Differential Games	IEEE, 2000	Finding optimal control for a non-linear system in a zero-sum differential game	They propose SLQ iterative algorithm and Game Theoretic Controller method

Table 3: **DGT-based Papers**

Acknowledgements

We acknowledge Mr. Kumar Abhishek (Machine Learning Lab) for discussions on MAB and crowdsourcing which greatly helped with the pacing of our work. We also acknowledge Ms. Manisha Padala for discussions on DGT which helped us build an understanding of the topic.

List of Papers Summarized

[Read1] Guy Aridor, Kevin Liu, Aleksandrs Slivkins, and Zhiwei Steven Wu. Competing bandits: The perils of exploration under competition. *CoRR*, abs/1902.05590, 2019.

[Read2] Gal Bahar, Omer Ben-Porat, Kevin Leyton-Brown, and Moshe Tennenholtz. Fiduciary bandits. *CoRR*, abs/1905.07043, 2019.

[Read3] Leonardo Cella and Nicolò Cesa-Bianchi. Stochastic bandits with delay-dependent payoffs, 2019.

[Read4] Yihan Du, Siwei Wang, and Longbo Huang. Dueling bandits: From two-dueling to multi-dueling. In *Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '20, page 348–356, Richland, SC, 2020. International Foundation for Autonomous Agents and Multiagent Systems.

[Read5] Gagan Goel, Afshin Nikzad, and Adish Singla. Mechanism design for crowdsourcing markets with heterogeneous tasks. In *HCOMP*, 2014.

[Read6] Jiuchuan Jiang, Bo An, Yichuan Jiang, Donghui Lin, Zhan Bu, Jie Cao, and Zhifeng Hao. Understanding crowdsourcing systems from a multiagent perspective and approach. *ACM Trans. Auton. Adapt. Syst.*, 13(2), July 2018.

[Read7] H. Mukai, Akio Tanikawa, Ilker Tunay, I.N. Katz, H. Schattler, P. Rinaldi, I.A. Ozcan, G. Wang, L. Yang, and Yuichi Sawada. Sequential linear quadratic method for differential games. 12 2000.

[Read8] Long Tran-Thanh, Archie Chapman, Alex Rogers, and Nicholas Jennings. Knapsack based optimal policies for budget-limited multi-armed bandits. *Proceedings of the National Conference on Artificial Intelligence*, 2, 04 2012.

References

[1] Philippe Aghion, Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt. Competition and Innovation: an Inverted-U Relationship*. *The Quarterly Journal of Economics*, 120(2):701–728, 05 2005.

[2] Nir Ailon, Thorsten Joachims, and Zohar Shay Karnin. Reducing dueling bandits to cardinal bandits. *CoRR*, abs/1405.3396, 2014.

[3] Peter Auer, Nicolò Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Mach. Learn.*, 47(2–3):235–256, May 2002.

[4] Moshe Babaioff, Robert D. Kleinberg, and Aleksandrs Slivkins. Truthful mechanisms with implicit payment computation. *CoRR*, abs/1004.3630, 2010.

[5] Sébastien Bubeck, Rémi Munos, and Gilles Stoltz. Pure exploration in multi-armed bandits problems. In *Proceedings of the 20th International Conference on Algorithmic Learning Theory*, ALT'09, page 23–37, Berlin, Heidelberg, 2009. Springer-Verlag.

[6] Lee Cohen and Yishay Mansour. Optimal algorithm for bayesian incentive-compatible exploration. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, page 135–151, New York, NY, USA, 2019. Association for Computing Machinery.

[7] Peter Frazier, David Kempe, Jon Kleinberg, and Robert Kleinberg. Incentivizing exploration. In *Proceedings of the Fifteenth ACM Conference on Economics and Computation*, EC '14, page 5–22, New York, NY, USA, 2014. Association for Computing Machinery.

[8] Nicole Immorlica, Adam Tauman Kalai, Brendan Lucier, Ankur Moitra, Andrew Postlewaite, and Moshe Tennenholtz. Dueling algorithms. In *Proceedings of the Forty-Third Annual ACM Symposium on Theory of Computing*, STOC '11, page 215–224, New York, NY, USA, 2011. Association for Computing Machinery.

[9] Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton. The future of crowd work. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, CSCW '13, page 1301–1318, New York, NY, USA, 2013. Association for Computing Machinery.

[10] Ilan Kremer, Yishay Mansour, and Motty Perry. Implementing the "Wisdom of the Crowd". *Journal of Political Economy*, 122(5):988–1012, 2014.

[11] T.L Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.*, 6(1):4–22, March 1985.

[12] Yang Liu, Goran Radanovic, Christos Dimitrakakis, Debmalya Mandal, and David C. Parkes. Calibrated fairness in bandits. *CoRR*, abs/1707.01875, 2017.

[13] Nuno Luz, Nuno Silva, and Paulo Novais. A survey of task-oriented crowdsourcing. *Artif. Intell. Rev.*, 44(2):187–213, 2015.

[14] Yishay Mansour, Aleksandrs Slivkins, and Zhiwei Steven Wu. Competing bandits: Learning under competition. *CoRR*, abs/1702.08533, 2017.

[15] Adish Singla and Andreas Krause. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, page 1167–1178, New York, NY, USA, 2013. Association for Computing Machinery.

[16] Long Tran-Thanh, Archie Chapman, Alex Rogers, and Nicholas R. Jennings. Knapsack based optimal policies for budget-limited multi-armed bandits. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, AAAI'12, page 1134–1140. AAAI Press, 2012.

[17] Wikipedia contributors. Differential game — Wikipedia, the free encyclopedia, 2020. [Online; accessed 26-May-2020].

[18] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *J. Comput. Syst. Sci.*, 78(5):1538–1556, September 2012.

[19] Yisong Yue, Josef Broder, Robert Kleinberg, and Thorsten Joachims. The k-armed dueling bandits problem. *J. Comput. Syst. Sci.*, 78(5):1538–1556, September 2012.

[20] M. Yuen, I. King, and K. Leung. A survey of crowdsourcing systems. In *2011 IEEE Third International Conference on Privacy, Security, Risk and Trust and 2011 IEEE Third International Conference on Social Computing*, pages 766–773, 2011.