Research Statement
Cem Tekin

Most of the modern engineering systems involve multiple informationally decentralized decision making agents that act in an unknown and changing environment, making decisions sequentially over time based on their history. Examples include Big Data systems where huge amounts of data being produced by numerous resources with time-varying characteristics is mined in real-time by distributed agents, social networks where agents share information and opportunities that arrive to them in a mutually beneficial fashion, or resource sharing networks such as cognitive radio networks where agents learn how to optimally share resources whose payoffs are unknown a priori. The salient features of all these systems are:

- Environment is unknown to the agents a priori and its probabilistic model can change over time.
- At each round, before making a decision an agent may observe a particular context (side information) which is related to the true state of the environment. For example, in a social network, context can be the personal information of the user that searched for a movie, in a cognitive radio network it can be a measurement of the noise level, and in a stream mining system it can be the size, location, type, etc. of the data stream. While many things can be part of the context, and while the space of possible contexts can be very large, only a small number of them may be the ones that are the most relevant to the underlying state of the environment.
- An agent gets online feedback only about the decisions it had made so far.
- Agents are informationally decentralized. They don't observe other agents' actions but they can make inferences about other agents' actions based on their own feedback.

In my research, I am developing an important methodology to formalize, analyze and solve such decentralized online learning problems. In particular, I answer the following fundamental questions:

- What is the price of learning? That is, how much utility is lost due to not knowing the underlying probabilistic model of the system?
- What is the price of decentralization? That is, how much utility is lost due to constraints on communication and coordination between the agents?
- Under what conditions the average loss due to learning and decentralization converges to zero, and what is the fastest convergence rate?
- When there are many types of contexts, how the agents can learn the most relevant ones?

In order to answer these questions, I develop new mathematical frameworks for both single-agent and multi-agent online learning problems, use tools from probability theory, stochastic control and machine learning, and develop new tools in these areas when necessary. Most importantly, in my research I apply the theory I developed to solve some of the most significant problems in data mining, social networks and communication networks, and also address application specific challenges including challenges arising in applications such as network security, real-time stream mining, smart healthcare and recommender systems. A summary of my past and current research along with some future research directions is given below.

Past Research

1) Online Learning in Multi-armed Bandit Problems - Applications in Communications, Resource Sharing Systems and Online Contract Selection Problems

In a multi-armed bandit (MAB) problem there is a set of arms, each of which when played by an agent yields some reward depending on its internal state which evolves stochastically over time. Within the context of this class of problems, agents who are initially unaware of the stochastic evolution of the environment (arms), aim to maximize a common objective based on the history of actions and observations. The classical difficulty in a bandit problem is the exploration-exploitation dilemma, which necessitates a careful algorithm design to balance information gathering and best use of available information to achieve optimal performance. The goal in these problems is to maximize the agent's total

expected reward (or global reward), which corresponds to minimizing the growth rate of regret, which is the difference between the expected total reward of agent's learning algorithm and the expected total reward of the best policy given complete information about the probabilistic model of the system. In my thesis I provided an agent-centric approach to derive optimal or approximately optimal learning algorithms by considering the computational power of the agents and the amount of information they can exchange with each other.

1-A) Markov MAB problems

For the single agent MAB problem, I studied the online learning problem when each arm evolves according to an unknown Markov process and the agent can only observe the state of the arm that it selects. These are called Markov MAB problems. Due to its computational intractability and due to the fact that the states of the arms keep evolving even when they are not selected by the agent, the optimal learning problem was previously solved only when the Markov process is an independent and identically distributed (IID) process. Markovian state evolution of the arms makes the problem equivalent to learning how to act optimally in an unknown partially observable Markov decision process (POMDP). I was able to prove the first logarithmic in time regret bound for this problem [1][12] which is the best one can prove even in the IID model. I also designed computationally simple heuristic online learning algorithms, whose performance loss with respect to the best strategy that always selects a fixed arm (i.e., weak regret) is bounded logarithmic in time [2], and a polynomial time algorithm which is approximately optimal for a subset of Markov MAB problems [3].

1-B) Dynamic resource sharing problems

I also proposed a MAB formulation for dynamic resource sharing problems [4], and investigated the best possible regret under different levels of communication and cooperation between the agents. The policies I developed for this setting are model free in the sense that they achieve optimal performance (in terms of weak regret) for both Markov and IID MAB problems. Applications include many resource sharing or resource allocation systems such as power control in wireless CDMA and opportunistic spectrum access.

1-C) Online contract selection problems

In an online contract selection problem, an agent offers a bundle of contracts to buyers arriving sequentially over time, whose types are drawn from an unknown distribution [5]. My results in this work can be used for offering wireless data plans to mobile consumers, offering spectrum contracts for wireless users on a secondary spectrum market and recommending bundles of items such as movie and news recommendations.

2) Atomic Congestion Games on Graphs and Their Applications in Networking

I proposed a new class of games in which players are connected through a graph network where each player's payoff is affected by its neighbors' strategies [6], and characterized the properties of these games. These games are ideal in modeling the behavior of strategic wireless users in a cognitive radio network, where interference depends on both the proximity and the actions of the wireless users.

Current Research

1) Cooperative Contextual Bandits for Online Learning in Big Data

In Big Data systems, huge amounts of data being produced by numerous resources in increasingly diverse formats such as sensor readings, documents, emails, transactions, videos etc., is mined in real-time by distributed agents, to assist numerous applications including patient monitoring, targeted advertisement, surveillance, network security etc. Hence, online data mining systems have emerged that enable such applications to analyze, extract knowledge and make decisions in real-time, based on the correlated, high-dimensional and dynamic data captured by multiple, distributed and heterogeneous data sources. I am developing an important methodology to formalize, analyze and solve such decentralized online learning problems, which is based on multi-armed bandits (MABs), called cooperative contextual bandits. In the considered systems, the distributed sources are processed by a distinct and heterogeneous

set of distributed online learners (agents) which determine on-the-fly how to classify the different data streams and make decisions based on this analysis. For this, the distributed learners either use their locally available classification functions or collaborate with each other to classify each other's data. Each learner is also provided with some context information (context vector) that comes along with the data which it can exploit to choose its classification actions smartly. Importantly, since the data is gathered at different locations, sending the data to another learner to process incurs additional costs such as delays, and hence this will be only beneficial if the benefits obtained from a better classification will exceed the costs. Since the prediction accuracy of the classifiers for various types of incoming streams is unknown and can change dynamically over time, this needs to be learned online. Hence, the learners need to continuously learn, exchange information, make accurate predictions and decisions over time.

I design distributed online learning algorithms whose long-term average rewards converge to the best distributed solution which can be obtained for the classification problem given complete knowledge of online data characteristics as well as their classification function accuracies and costs when applied to this data. The strength of my results comes from the fact that I make no distributional assumptions on how these context arrive to the learners, thus they hold even in the worst-case. I also address a lot of Big Data specific challenges including how cooperative contextual bandits can deal with concept drift, delay, missing labels, as well as how ensemble learning and learning over networks of learners can be performed using cooperative contextual bandits [8]. This is the first work to consider both heterogenous context information and different action spaces in a multi-agent learning environment from the MAB point of view. The significance of this theory comes from the fact that it provides a new perspective into many important prevailing problems in engineering, and it makes us possible to characterize the performance loss due to learning in a dynamic environment explicitly in terms of bounds that hold uniformly over time, rather than only providing asymptotic results.

There is a tradeoff between the asymptotic convergence and finite time performance of cooperative contextual bandits [7][8][13]. Most of the time, the context vector can be very high dimensional since it is not known a priori by the agents which contexts are the best to exploit. This slows down learning, and when the context is high dimensional, finite time performance can be bad although asymptotic convergence to the optimal payoff is guaranteed. In [9], I develop a method to learn the set of most relevant contexts which can be time varying based on the data characteristics. I prove that the convergence rate depends only to the dimension of the relevant contexts instead of the dimension of the entire context space. This implies that learning can be very fast when the set of relevant context is small.

2) Online Learning in Social Recommender Systems

In [10], I apply the idea of cooperative contextual bandits to social recommender systems. By forming a network, agents are able to share information and opportunities in a mutually beneficial fashion. For example, companies can collaborate to sell products, charities can work together to raise money, and a group of workers can help each other search for jobs. Through such cooperation, agents are able to attain much higher payoff than would be possible individually. But sustaining efficient cooperation can also prove extremely challenging. Firstly, agents operate with only incomplete information, and must learn the environment parameters slowly over time. Secondly, agents are decentralized and thus uncertain about their neighbor's information and preferences. Finally, agents are selfish and may opt not to reciprocate aid to their neighbors, favoring personal gain over social gain. I propose a class of mechanisms that effectively addresses all of these issues: at once allowing decentralized agents to take near-optimal actions in the face of incomplete information, while still incentivizing them to fully cooperate within the network.

Future Research

1) A Unifying Online Discovery Framework: From Stream Mining to Personalized Medicine

The basic principle behind online learning is that after an action is taken, its payoff/reward is immediately observed. However such a sequential structure may not fit into every problem. For example, consider personalized medical treatments, where medicine is provided to a patient based on his/her

specific health history and tests. Personalized medicine can both significantly improve the success rate of the treatments and reduce the operating costs of healthcare. Usually a sequence of actions is taken during a treatment and the success of the treatment depends on the initial condition (initial context) of the patient as well as the final condition (final context). Intermediate observations (intermediate contexts) between the start and end of the treatment such as blood tests, x-rays, etc. may be available or not. Basically, the agent (hospital, doctor, etc.) should take a sequence of actions (possibly there are costs associated with actions), and should take a final action only after which the reward is revealed. Standard MAB methods are inefficient in this setting because the strategy set which is the set of all sequences of actions that the agent can take is huge. Moreover, the agent is not only interested in the total expected reward, but there are also constraints on how good the selected sequence of actions should be in each round. This is a unifying framework that captures both stochastic and contextual MAB problems as special cases. Specifically, I would like to address the following questions:

- Given a context, what sequence of actions should be chosen such that the probability of success is maximized while the costs are minimized?
- What is the tradeoff between the total expected reward and minimum instantaneous reward, and can they be maximized jointly?
- What are the necessary and sufficient structural assumptions on the context evolution process which guarantees fast learning?

I have given partial answers to some of these questions in [11], where I considered the specific problem of classification using cascades of classifiers for a stream mining system. As a final remark, stream mining and personalized medicine are just two examples of this broadly applicable framework.

2) Strategic Pricing and Investment in an Unknown Environment

Economists have studied the pricing and investment strategies of firms competing in a dynamic environment where the quality of the product of a firm depends on its investment history while the consumer's willingness to pay for a product depends on its perceived quality. They have used game theory to characterize the equilibrium behavior of the firms. However, this characterization is heavily dependent on strong assumptions such as complete information, prior beliefs and public signals about the product quality and its evolution. I plan to apply tools from both online learning theory and game theory to derive approximately optimal pricing and investment strategies for the firms, without any of these assumptions. The fundamental question is: Is it possible that the firms can learn to play their equilibrium strategies, when they learn about their quality and the effectiveness of their investment strategies online. A positive answer to this question will have far reaching consequences which will bridge the gap between prior-free online learning and game theoretic analysis. The results of this can be applied to engineering problems involving strategic agents such as femtocell networks, social networks and electronic commerce.

References

[1] C. Tekin and M. Liu. Learning of uncontrolled restless bandits with logarithmic strong regret. *IEEE Trans. on Information Theory*, second round of revision, 2013.

[2] C. Tekin and M. Liu. Online learning of rested and restless bandits. *IEEE Trans. on Information Theory*, 58(8): 5558 – 5611, August 2012.

[3] C. Tekin and M. Liu. Approximately optimal adaptive learning in opportunistic spectrum access. In *Proceedings of the 31st IEEE International Conference on Computer Communications (INFOCOM)*, May 2012.

[4] C. Tekin and M. Liu. Performance and convergence of multiuser online learning and its application in dynamic spectrum sharing. In *Mechanisms and Games for Dynamic Spectrum Access.* Cambridge University Press, available from January 2014.

[5] C. Tekin and M. Liu. Online learning in a contract selection problem, preprint: http://arxiv.org/abs/1305.3334, 2013.

[6] C. Tekin, M. Liu, R. Southwell, J. Huang, and S. Ahmad. Atomic congestion games on graphs and its applications in networking. *IEEE/ACM Trans. on Networking*, 20(5):1541 –1552, October 2012.

[7] C. Tekin and M. van der Schaar. Distributed online learning via cooperative contextual bandits. *IEEE Trans. on Signal Processing*, under review, preprint: http://arxiv.org/pdf/1308.4568.pdf, 2013.

[8] C. Tekin and M. van der Schaar. Decentralized online big data classification - a bandit framework. *IEEE Trans. on Signal Processing*, under review, preprint: http://arxiv.org/pdf/1308.4565.pdf, 2013.
[9] C. Tekin, L. Canzian, and M. van der Schaar. Context adaptive big data stream mining. *IEEE Trans. on Emerging Topics in Computing (TETC) Special Issue on Big Data*, under review, preprint: http://medianetlab.ee.ucla.edu/papers/TETCadaptive.pdf, 2013.
[10] C. Tekin, S. Zhang, and M. van der Schaar. *Distributed online learning in social recommender systems*. *IEEE Journal of Special Topics in Signal Processing, Signal Processing for Social Networks (JSTSP-SPSN)*, accepted for publication, December 2013.
[11] J. Xu, C. Tekin, and M. van der Schaar. Learning optimal classifier chains for real-time Big Data mining. In *Proceedings of the 51st Allerton Conference on Communication, Control, and Computation*, October 2013.
[12] C. Tekin and M. Liu. Adaptive learning of uncontrolled restless bandits with logarithmic regret. In *Proceedings of the 49th Annual Allerton Conference on Communication, Control, and Computation*, September 2011
[13] C. Tekin, and M. van der Schaar. Distributed online Big Data classification using context information. In *Proceedings of the 51st Allerton Conference on Communication, Control, and Computation*, October 2013.