Decision Learning in Data Science

Yan Chen[1], Chunxiao Jiang[1,2], Chih-Yu Wang[1,3], Yang Gao[1,4], and K. J. Ray Liu[1]

[1]Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742, USA.
[2]Department of Electronic Engineering, Tsinghua University, Beijing, 100084, P. R. China.
[3]Research Center for Information Technology Innovation, Academia Sinica, Taiwan.
[4]Facebook, USA.
E-mail:yan@umd.edu, chx.jiang@gmail.com, cywang@citi.sinica.edu.tw, yanggao@umd.edu, kjrliu@umd.edu

***Abstract*—With the increasing ubiquity and power of mobile devices, as well as the prevalence of social systems, more and more activities in our daily life are being recorded, tracked, and shared, creating the notion of "social media". Such abundant and still growing real life data, known as "big data", provide a tremendous research opportunity in many fields. To analyze, learn and understand such user-generated data, machine/social learning has been an important tool and various machine learning algorithms have been developed. However, since the user-generated data are the outcome of users' decisions, actions and their socio-economic interactions, which are highly dynamic, without considering users' local behaviors and interests, existing learning approaches tend to focus on optimizing a global objective function at the macroeconomic level, while totally ignore users' local interactions at the microeconomic level. As such there is a growing need in bridging machine/social learning with strategic decision making, which are two traditionally distinct research disciplines, to be able to jointly consider both global phenomenon and local effects to understand/model/analyze better the newly arising issues in the emerging social media with user-generated data. In this paper, we present the emerging notion of "decision learning", i.e. learning with strategic decision making, that involves users' behaviors and interactions by combining learning with strategic decision making. We will discuss some examples from social media with real data to show how decision learning can be used to better analyze users' optimal decision from a user's perspective as well as design a mechanism from the system designer's perspective to achieve a desirable outcome.**

***Index Terms*—Data science, big data, machine learning, game theory, social media, behavior analysis, mechanism design, decision learning.**

I. Introduction

With the rapid development of communication and information technologies, the last decade has witnessed a proliferation of emerging social systems that help to promote the connectivity of people to an unprecedentedly high level. Examples of these emerging systems can be found in a wide range of domains from online social networks like Facebook or Twitter; to crowdsourcing sites like Amazon Mechanical Turk or Topcoder where people solve various tasks by assigning them to a large pool of online workers; to online question and answering (Q&A) sites like Quora or Stack Overflow where people ask all kinds of questions; and all the way to new paradigms of power/energy systems like smart grid. Fig. 1 shows a few examples of such growing social systems.

Fig. 1. Examples of social systems.

Together with the increasing ubiquity and power of mobile devices, the prevalence of social systems, and the rise of global clouds, more and more activities in our daily life are being recorded, tracked, and shared, creating the notion of "social media". Such abundant and still growing real life data, known as "big data", provide a tremendous research opportunity in many fields, for example, behavior and sentiment analysis, epidemics and diseases propagation modeling, grid and network traffic management, financial market trends tracking, just to name a few.

To analyze, learn and understand such user-generated data, machine/social learning has been an important tool [1], [2]. Learning aims to use reasoning to find new, relevant information given some background knowledge through representation, evaluation, and optimization. However, there are some limitations and constraints. First the generalization assumption that the training set is statistically consistent with testing set is often not true because users behavior differently at different time under different setting. Second, the single objective function cannot cover users' different interests since users have different interests and thus different objective function. Besides, users are rational and thus naturally selfish - they want to optimize their own objective functions [3], [4]. Third, the data is the outcome of users' interaction, while learning algorithms cannot naturally involve users' individual local interest. Therefore, the knowledge contained in the data is difficult to be fully exploited from such a macroscopic view.

Existing learning approaches tend to focus on optimizing a global objective function at the macroeconomic level, while totally ignore users' local interactions/decisions at the microeconomic level. Indeed, user-generated data is the outcome of users' decisions, actions and their social-economic interac-

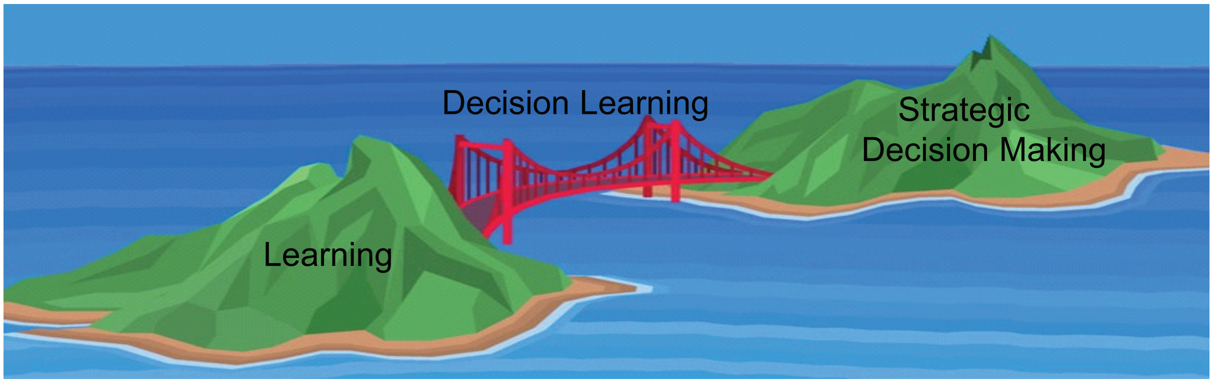


Fig. 2. Bridging learning with strategic decision making.

tions, which are highly dynamic, and thus the interactions of users and their decision making process should be taken into consideration. As such there is a growing need in bridging learning with strategic decision making to be more effective in mining, reasoning and extracting knowledge and information from “big data”.

Yet, there is a missing link. Traditionally, both learning and decision making are two distinct research disciplines. Success in bridging them allows us to jointly consider both the global phenomenon and local effects to better understand/model/analyze user-generated data from the emerging social media. Besides, learning is for making optimal decisions. In essence, learning and decision making are destined to couple due to the network externality, i.e., the influence of other users' behaviors on one user's reward [5]. In this paper, we describe the emerging research field of **“decision learning”**, i.e., learning with strategic decision making, that involves users' behaviors and interactions by combining learning with strategic decision making, as illustrated in Fig. 2. In decision learning, there are two major elements of data-driven issues: one is the modeling, analysis, and understanding, of user behaviors and their interactions, and the other is the design of mechanism to achieve the desired outcomes. The former considers the issues from user perspectives, while the latter motivates from system point of views.

Different from traditional networks and systems where users are mandated by fixed and predetermined rules, user interactions in social media/networks are generally self-enforcing [6], [7]. On one hand, users in these systems have great flexibilities in their actions and have the ability to observe, learn, and make intelligent decisions. On the other hand, due to the selfish nature, users will act to pursuit their own interests, which oftentimes conflict with other users' objectives and the system designer's goal. These new features call for new theoretical and practical solutions to the designs of social media/networks. How can system designers design their systems to resolve the conflicting interests among users? And given various and conflicting interests among users, how to achieve a desired system-wide performance?

The above questions motivate the study of user behaviors and incentive mechanisms in data science. Incentive mechanisms refer to schemes that aim to steer user behaviors through the allocation of various forms of rewards such as monetary rewards, virtual points and reputation status. Plenty of empirical evidences can be found in the social psychology literature that demonstrate user behaviors in social media/networks are indeed highly influenced by these rewards [9]–[14]. Although we can learn from the social psychology literature on what factors influence user behaviors and thus can be used as rewards, how to allocate these rewards to achieve desired user behaviors is still not well understood, which leads to ad hoc or poor designs of incentive mechanisms in many social media/networks in practice. How can we fundamentally understand user behaviors under the presence of rewards in social media/networks? Moreover, based on such understandings, how should a system developer design incentive mechanisms to achieve various objectives in a systematic way?

The focus of this paper is to open a discussion of an emerging field, termed as decision learning, that jointly combines learning with decision making towards a better fundamental understanding of user behaviors embedded under the tsunami of user-generated “big data”. In this paper, we present three game-theoretic frameworks to formally model user participation and interactions under various scenarios in social media/networks: decision learning with evolutionary user behavior, decision learning with sequential user behavior, and decision learning with mechanism design. On the evolutionary behavior, how information diffuses over online social networks using graphical evolutionary game is presented; on the sequential behavior, how customers learn and choose the “best” deals using Chinese restaurant game framework is considered; and on the mechanism design, how to design mechanism to collect high quality data with low cost from crowdsourcing is illustrated. Using these frameworks, we can theoretically analyze and predict user behaviors through equilibrium analysis. And, based on the analysis, one can optimize in a systematic way the design of incentive mechanisms for social media/networks to achieve a wide range of system objectives and analyze their performances accordingly.

The paper is organized as follows. In Section II, user behavior modeling and analysis will be considered. First a graphical evolutionary game framework is presented to tackle the repetitive/evolutionary user behavior, followed by the discussion of the Chinese restaurant game framework for the

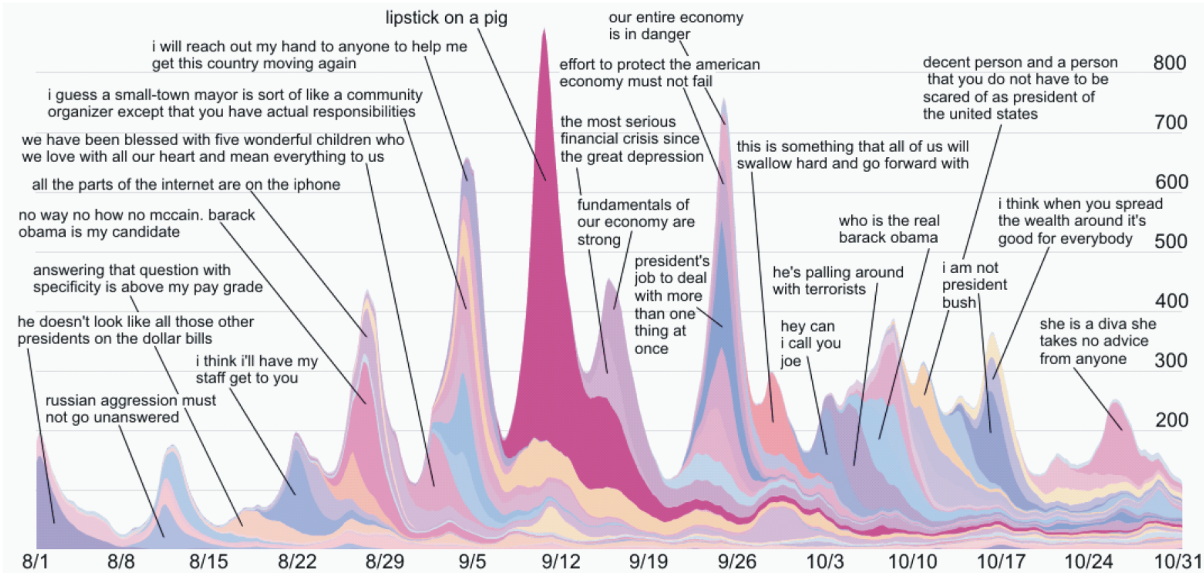


Fig. 3. Memetracker phrase cluster during 2008 US presidential election: spreading of comments and phrasing by candidates [8].

understanding of sequential user behavior. In Section III, the design of incentive mechanisms to achieve desirable goals is illustrated with an example on designing mechanism for obtaining high-quality data and for enforcing "good" behavior in crowdsouring. In Section IV, related works on combining learning with decision learning are discussed. Finally, conclusions are drawn and final thoughts are given in Section V.

II. User Behavior Modeling and Analysis in Decision Learning

In this section, we will address decision learning from user point of view. Both the evolutionary and sequential user behaviors are commonly exhibited in social systems. How learning with strategic decision making may arise from both settings will be illustrated first with information diffusion over online social networks using graphical evolutionary game framework from Twitter and Memetracker data, and then with optimal restaurant strategy using Chinese restaurant game framework from both Groupon deals and Yelp rating, respectively.

A. Evolutionary User Behavior: Graphical Evolutionary Game Framework

One typical user behavior in social systems is the repetitive and evolutionary decision making. A good example is that users repetitively decide whether to post/forward information or not on online social networks. Fig. 3 shows the top 50 threads in the news cycle with highest volume for the period Aug. 1 - Oct. 31, 2008, where each thread consists of all new articles and blog posts containing a textual variant of a particular quoted phrases. The five large peaks between late August and late September corresponding to the Democratic and Republican National Conventions illustrate the spreading of comments and phrasing by candidates. Notice that the information forwarding is often not unconditional. One has to make a decision on whether or not to do so based on many factors, such as if the information is exciting or if his/her friends are interested on it, etc. Other examples include repetitive online purchasing and review posting.

We find that in essence the repetitive/evolutionary decision making process on social systems follows similarly the evolution process in natural ecological systems [15]. It is a process that evolves from one state at a particular instance to another when information is shared and decision is made. Thus, the evolutionary game is an ideal tool to model and analyze the social system users' repetitive and evolutionary behavior. Evolutionary game theory is an application of the mathematical theory of games to the interaction dependent strategy evolution in populations [15]. Arising from the realization that frequency dependent fitness introduces a strategic aspect to evolution, evolutionary game theory becomes an essential component of a mathematical and computational approach to biological contexts, such as genes, viruses, cells, and humans. Recently, evolutionary game theory has also become of increased interest to economists, sociologists, anthropologists, and social scientists. Here, we show how the evolutionary game theory is deployed to study users' repetitive and evolutionary behavior in social systems.

In the setting of our consideration, the social system user topology can be treated as a graph structure and the user with new decision can be regarded as the mutant. By considering the decision making process as the mutant spreading process (to forward or not to forward when an event (mutant) takes place), the graphical evolutionary game provides us with an analytical means to find the evolutionary dynamics and equilibrium of user behavior.

1) Graphical Evolutionary Game Framework: In evolutionary game theory, the utility of a player is referred to as "fitness" [16]. Specifically, the fitness Φ is a linear combination of the baseline fitness (B) representing the player's inherent property and the player's payoff (U) which is determined by the predefined payoff matrix and the player's interactions with

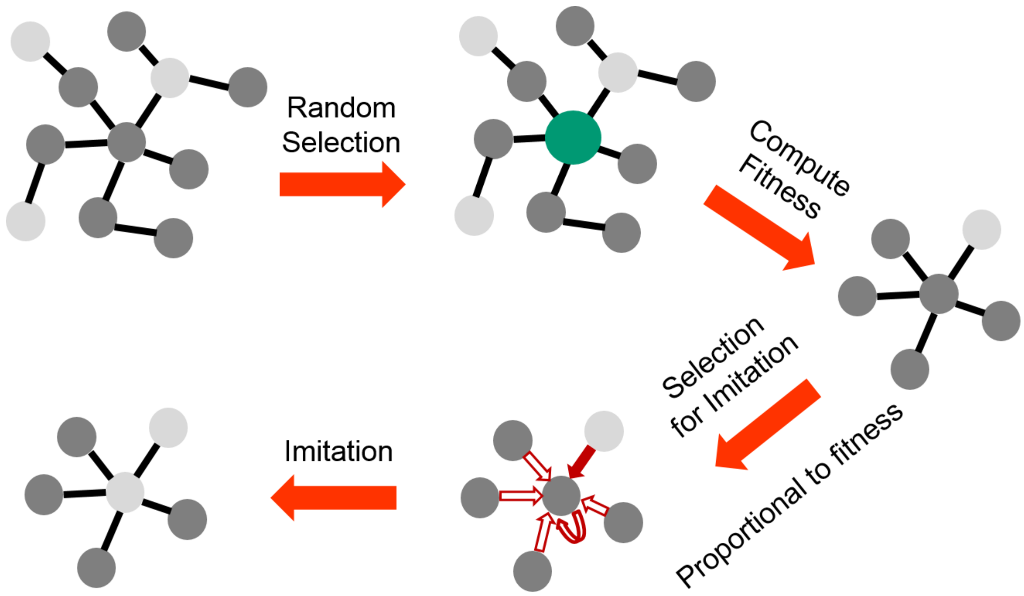


Fig. 4. Imitation strategy updating rule.

others as follows

$$\Phi = (1-\alpha)B + \alpha U, \tag{1}$$

where the combining weight α is called as the selection intensity. One can interpret that one's fitness is not only determined by one's own strength, but also from one's environment affecting with a selection intensity α. The case that $\alpha \to 0$ represents the limit of weak selection [17], while $\alpha \to 1$ denotes strong selection. The selection intensity can also be time varying, e.g., $\alpha = \beta e^{-\epsilon t}$, which means that the contribution of game interaction decreases along with time.

With the fitness function, the evolutionary game theory studies and characterizes how a group of players converges to a stable equilibrium after a period of strategic interactions. Such a final equilibrium state is called as the Evolutionarily Stable State (ESS), which is "a strategy such that, if all members of the population adopt it, then no mutant strategy could invade the population under the influence of natural selection" [15]. In other words, even if a small fraction of players may not be rational and take out-of-equilibrium strategies, ESS is still a locally stable state. How to find the ESSs is an important issue in evolutionary game theory. One common approach is to find the stable points of the system state dynamic, which is known as replicator dynamics. The corresponding underlying physical meaning is that: if adopting a certain strategy can lead to a higher fitness than the average level, the proportion of population adopting this strategy will increase and the increasing rate is proportional to the difference between the average fitness with this strategy and the average fitness of the whole population. Note that when the total population is sufficiently large and homogeneous, the proportion of players adopting a certain strategy is equivalent to the probability of one individual player adopting such a strategy, i.e., the strategy distribution over the whole population can be interpreted as each player's mixed strategy and the replicator dynamics can be interpreted as each player's mixed strategy update.

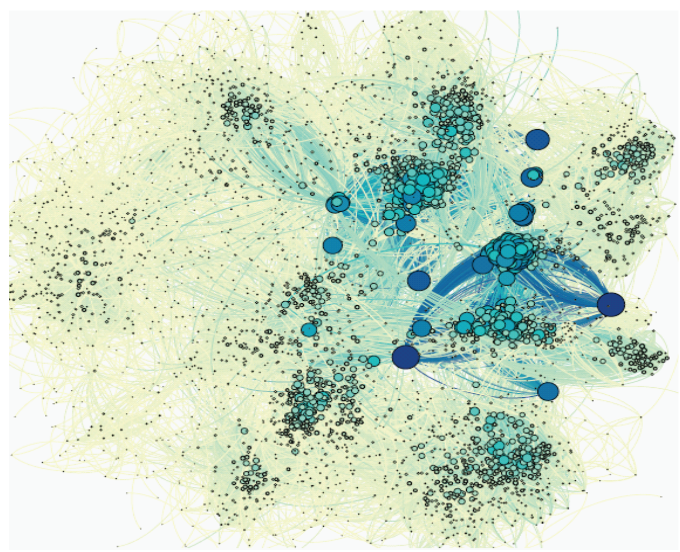


Fig. 5. A Facebook subnetwork.

Graphical evolutionary game theory is to study the strategies evolution in such a structured population [18]. In the graphical evolutionary game theory, in addition to the entities of players, strategy and fitness matrix, each game model is associated with a graph structure, where the vertexes represent players

Graphical EGT	Social Network
Graph structure	Social network topology
Players	Users in the social network
Strategy	S_f : forward the information S_n: not forward the information $\begin{array}{ccc} & S_f & S_n \\ \begin{array}{c} S_f \\ S_n \end{array} & \multicolumn{2}{c}{\begin{pmatrix} u_{ff} & u_{fn} \\ u_{fn} & u_{nn} \end{pmatrix}} \end{array}$
Fitness	Utility from forwarding or not
ESS	Stable information diffusion state

Fig. 6. Information diffusion as a graphical evolutionary game.

and the edges determine which player to interact with. Since the players only have limited connections with others, each player's fitness is locally determined from interactions with all adjacent players.

The commonly used strategy updating rules [19] are originated from the evolutionary biology field and used to model the resident/mutant evolution process. Fig. 4 illustrates the detailed evolution procedures of the imitation (IM) strategy update rule. In the first step, a user is randomly chosen from the population for imitation. Then, the fitness of the chosen user and all corresponding neighbors is computed. Finally, the user will, in probability, either be imitated by one of the neighbors or remain with his/her current strategy, with the probability being proportional to fitness. There are also other rules such as birth-death (BD) strategy update rule and death-birth (DB) strategy update rule, but through theoretical analysis [20], we find that these rules are equivalent when the network degree is sufficiently large.

2) Information Diffusion Formulation and Analysis: A social network is usually illustrated by a graph, e.g., a Facebook sub-network is shown in Fig. 5, where each node represents a user and the edge represents the relationship between users. When some new information is originated from one user, the information may be propagated over the network depending on other users' actions: to forward the information or not. For each user, whether to forward the information is determined by several factors, including the user's own interest on this information and his/her neighbor's actions in the sense that if all his/her neighbors forward the information, the user may also forward the information with a relatively high probability. In such a case, the users' actions are coupled with each other through their social interactions. This is very similar to the player's strategy update in the graphical evolutionary game, where players' strategies are also influenced with each other through the graph structure. In graphical evolutionary game, a user's strategy can influence one of his/her neighbors when the fitness of adopting this strategy is high. Similarly, in the information diffusion process, when forwarding the information can bring a user more utility, the user's neighbors may also be influenced to forward the information in the near future. Therefore, the information diffusion process can be well modeled by the graphical evolutionary game as illustrated in Fig. 6.

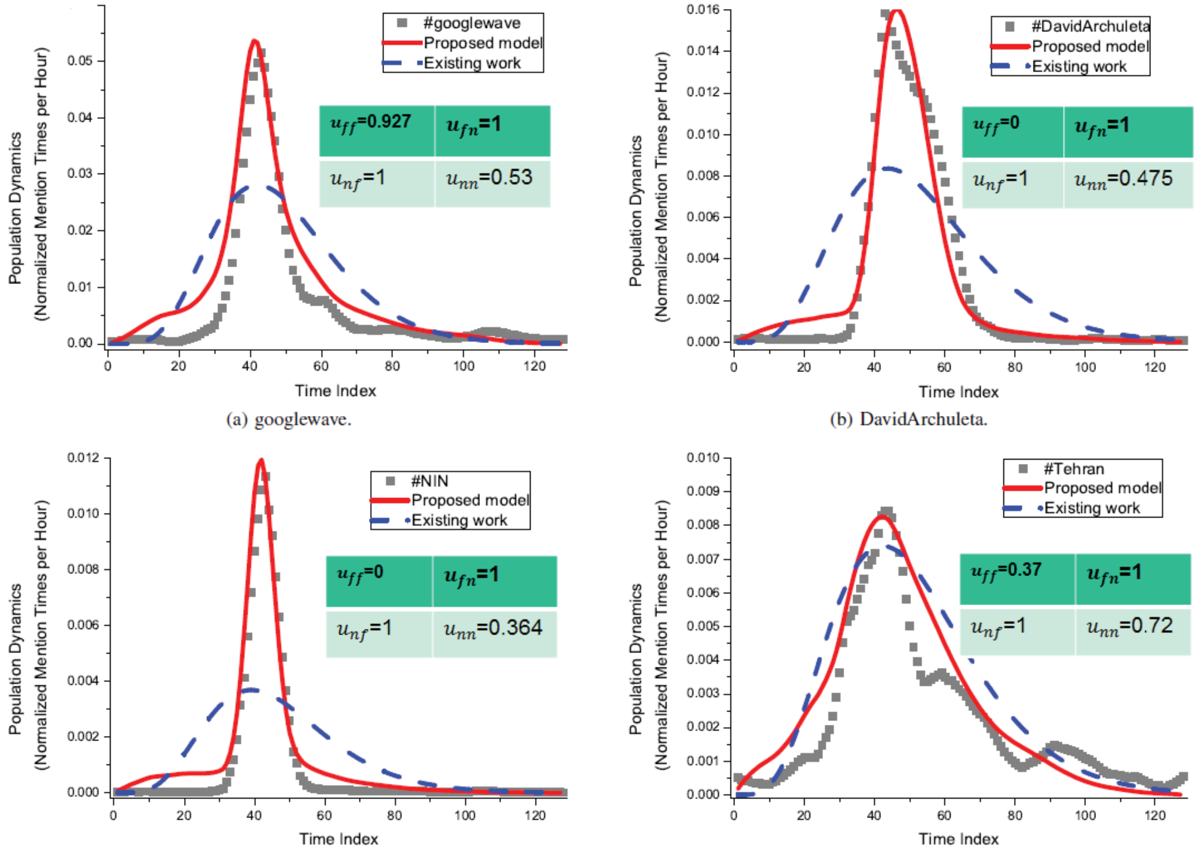
There are two possible actions for each user, i.e., to forward (S_f) or not forward (S_n), and the corresponding users' payoff matrix can be written as

$$\begin{pmatrix} u_{ff} & u_{fn} \\ u_{fn} & u_{nn} \end{pmatrix} \tag{2}$$

where a symmetric payoff structure is considered, i.e., when a user with strategy S_f meets a user with strategy S_n, each of them receives the same payoff u_{fn}. Note that the payoff matrix is related to the fitness in the graphical evolutionary game according to (1). The physical meaning of the payoff can be either the popularity of a user in a social network or the hit rate of a website. These three parameters will be "learned" from the data and then used for "decision making". Under different application scenarios, the values of the payoff matrix may be different. For example, if the information is related to recent hot topics and forwarding of the information can attract more attentions from other users or website, the payoff matrix should have the following characteristic: $u_{ff} \geq u_{fn} \geq u_{nn}$. According to (1), the fitness of forwarding is larger and thus the probability of forwarding will be higher. On the other hand, if the information is about useless advertisements, the payoff matrix would exhibit $u_{nn} \geq u_{fn} \geq u_{ff}$, i.e., the fitness of not forwarding is higher and thus users tend not to forward information. Furthermore, if the information is supposed to be shared only within a circle, i.e., a small group with same interest, the payoff matrix could exhibit $u_{fn} \geq u_{ff} \geq u_{nn}$.

Since the player's payoff is determined by both of its own strategy and the opponent's strategy, in order to characterize the global population dynamics, we need to first derive the local influence dynamics, as well as the corresponding influence equilibria. We find in [20] that the local network states, i.e., the neighbors' strategy distribution given a player's strategy, evolve with a rate of order 1, while the global network state, i.e., the strategy distribution of the whole population, evolves with a rate at the order of the selection intensity α, which is much smaller than 1 due to the weak selection [17]. In such a case, the local network states will converge to equilibria in a much faster rate than the global network state. This is because the dynamics of local network states are only in terms of a local area, which contains only the neighbors. In such a small scale, the local dynamics can change and converge quite fast. On the other hand, the dynamics of global network state are associated with all users, i.e., the whole networks, the dynamics would be much slower. Therefore, the global network state can be regarded as constant during the convergence of influence dynamics. By doing so, the equilibria of the local influence dynamics can be obtained, which are found to be linear functions of the global network state.

With the equilibria of the local influence dynamics, the global population dynamics can be derived through analyzing the strategy updating rules specified in the graphical evolutionary game [19]. It is found that the global population dynamics



(a) googlewave. (b) DavidArcheleta. (c) NIN. (d) Tehran.

Fig. 7. Experimental results of the evolutionary population dynamics.

```
2  8  we're not commenting on that story i'm afraid   2131865
   3  3  we're not commenting on that    489007
      2008-08-18 14:23:05  1  M  http://business.theage.com.au/business/bb-chief-set-to-walk-plank-20080818-3xp7.html
      2008-11-26 01:27:13  1  B  http://sfweekly.com/2008-11-26/news/buy-line
      2008-11-27 18:55:30  1  B  http://aconstantineblacklist.blogspot.com/2008/11/re-researcher-matt-janovic.html
   5  2  we're not commenting on that story      2131864
      2008-12-08 14:50:18  3  B  http://videogaming247.com/2008/12/08/home-in-10-days-were-not-commenting-on-that-story-says-scee
      2008-12-08 19:35:31  2  B  http://jplaystation.com/2008/12/08/home-in-10-days-were-not-commenting-on-that-story-says-scee
```

Fig. 8. An example of Memetracker prase cluster dataset [8].

can be represented as a two-parameter third order polynomial function of the global network state [20]

$$\dot{p}_f(t) = \frac{\alpha(\bar{k}-1)(\overline{k^2}-2\bar{k})}{(\overline{k^2}-\bar{k})^2} p_f(t)\left[1-p_f(t)\right]\left[ap_f(t)+b\right], \tag{3}$$

where $p_f(t)$ is the proportion of population forwarding the information and $\dot{p}_f(t)$ is the corresponding dynamics, $\bar{k} = E[k]$ is the average degree of the network, $\overline{k^2} = E[k^2]$ is the second moment of the degree of the network, a and b are two parameters determined by the payoff matrix shown in (2).

From (3), we can see that given the characteristic of the network, i.e., the average degree $\bar{k}$ and the second moment of the degree $\overline{k^2}$, the evolution dynamics of the information diffusion can be modeled by a simple two-parameter third order polynomial function where the two parameters a and b are determined by the payoff in the payoff matrix, i.e., u_{ff}, u_{fn} and u_{nn}. Therefore, by learning the payoff from the data, we are able characterize the evolution dynamics of information diffusion using the evolutionary game-theoretic framework.

By evaluating the global population dynamics at the steady state, the global population equilibria can be found [21], which is 0 (no user shares the information to the neighbors), 1 (all

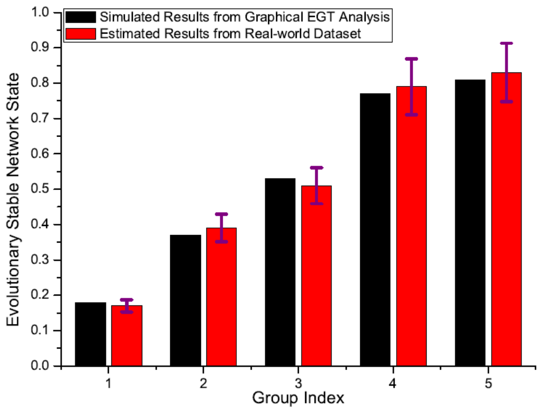


Fig. 9. Experimental results of the evolutionarily stable strategy.

users share the information to their neighbors), or only a portion of users share the information to their neighbors where the amount of such users is purely determined by the payoff matrices as follow

$$p_f^* = \begin{cases} 0, & \text{if } u_{nn} > u_{fn} > u_{ff}; \\ 1, & \text{if } u_{ff} > u_{fn} > u_{nn}; \\ \frac{(\overline{k^2}/\bar{k}-2)(u_{fn}-u_{nn})+(u_{ff}-u_{nn})}{(\overline{k^2}/\bar{k}-2)(2u_{fn}-u_{ff}-u_{nn})}, & \text{else}. \end{cases} \tag{4}$$

From (4), we can see that neither user forwarding the information can gain the most payoff while both forwarding gains the least payoff, $p_f^* = 0$. This is corresponding to the scenario where the released information is useless or negative advertisement, forwarding which can only incur unnecessary cost. On the contrary, when both users forwarding the information can gain the most payoff while not forwarding gains the least payoff, $p_f^* = 1$. This is corresponding to the scenario where the released information is an extremely hot topic, forwarding which can attract more attentions. For other cases, p_f^* lies between 0 and 1. For this third ESS, some approximations can be made as follows:

$$\begin{aligned} p_f^* &= \frac{(\overline{k^2}/\bar{k}-2)(u_{fn}-u_{nn})+(u_{ff}-u_{nn})}{(\overline{k^2}/\bar{k}-2)(2u_{fn}-u_{ff}-u_{nn})}, \\ &\doteq \frac{1}{1+\frac{u_{fn}-u_{ff}}{u_{fn}-u_{nn}}}, \end{aligned} \tag{5}$$

where the last approximation is due to $\overline{k^2}/\bar{k} \geq \bar{k}$ and the assumption that the average network degree $\bar{k} \gg 2$ in real social networks. We can see that when average network degree $\bar{k}$ is sufficiently large, the information diffusion result is independent of the network scale, i.e., there is a scale-free phenomenon for the information diffusion equilibrium.

3) Experiments with Real-World Datasets: The real-world datasets are used to validate the proposed model. We first use the Twitter hashtag dataset to validate the evolutionary population dynamics [20]. Specifically, we learn the payoff matrix in (2) by fitting the real temporal dynamics with the evolution dynamics in (3), and generate the corresponding evolution dynamics based on the estimated payoff matrix. The Twitter hashtag dataset contains the number of mention times per hour of 1000 Twitter hashtags with corresponding time series, which are the 1000 hashtags with highest total mention times among 6 million hashtags from June to December 2009 [22]. We compare our results with one of the most related existing works using data mining method [8]. Fig. 7 shows the comparison results, where the vertical axis is the dynamics and the mention times of different hastags per hour in the Twitter dataset are normalized within interval [0, 1] and denoted by solid gray square. From the figure, we can see that the game-theoretic model can fit very well the real-world information diffusion dynamics, better than the data mining method in [8] since the users' interactions and decision making behaviors are taken into account.

We then use the "MemeTracker" dataset to validate the ESS [21]. The dataset contains more than 172 million news articles and blog posts from 1 million online sources [8]. When a site publishes a new post, it will put hyperlinks to related posts in some other sites published earlier as its sources. And later, the site will also be cited by other newer posts as well. An example is shown in Fig. 8. In such a case, the hyperlinks between articles and posts can be used to represent the spreading of information from one site to another site. We extract 5 group of sites, where each group includes 500 sites. Each group is regarded as a complete graph and each site is considered as a user. We divide the dataset into two halfs, where the first half is used to train the payoff matrix and the second half is used for testing. Fig. 9 shows the results using the proposed model and the results from the real-world dataset, from which we can see they match well with each other. We also depict the variances of the estimated results in Fig. 9, which shows that the simulated results are always in the variance interval of the corresponding estimated results. Fig. 9 also reveals the cohesiveness of different group. We can see that the sites in Group 5 behave cohesively or share major common interests, while the sites in Group 1 share relatively little common interests. This is in particular interesting to advertisement or advocation scenarios where certain cohesive focus groups need to be mined to target with high return value.

B. Sequential User Behavior: Chinese Restaurant Game Framework

Another distinguish feature of social systems is that users often contribute/participate sequentially at their own time and space. For example, users sequentially visit question and answering (Q&A) sites like Yahoo!Answers and Stack Overflow, and decide whether to provide an answer, to vote an existing answer, or not to participate. Other examples include online reviews where customers write reviews for the product they purchase, and social news sites where online users post and promote stories under various categories.

The existence of network externality [5] in a social group dictates that users' actions/decisions influence each other. The network externality can be either positive or negative. When it is positive, users will have higher utilities when making the same decisions. On the contrary, when negative, users tend to make different decisions from others to achieve higher utilities.

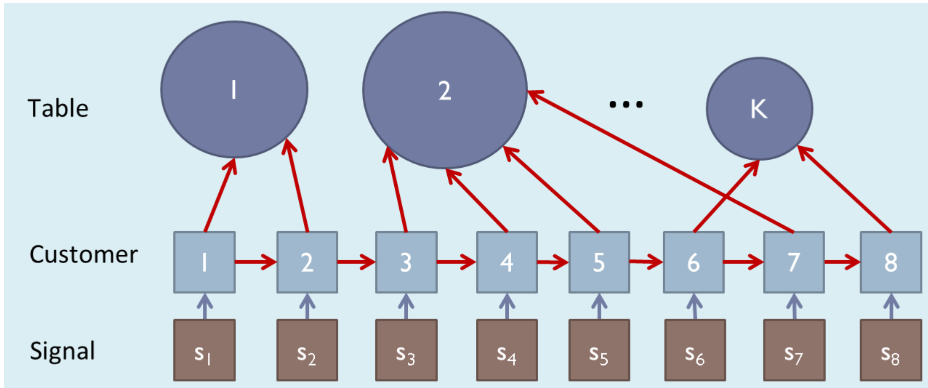


Fig. 10. System model of Chinese restaurant game.

To achieve better performance, users should take into account the effect of network externality when making decisions.

On the other hand, users' decisions are also influenced by their knowledge on the system. In general, a user's knowledge on the system may be very limited due to the uncertainty in observations. This limitation reduces the accuracy of the user's decision and thus the overall system performance. The phenomenon of limited knowledge can be overcome through learning [23]–[26]. Users can learn from their previous experiences through machine learning technique and/or from other users' decisions and observations through social learning. All such information can help users to construct a belief, which can be probabilistic, on the unknown system states. In most cases, the accuracy of users' decisions can be greatly enhanced by taking into account the belief.

Therefore, to achieve the best utilities, users need to consider the effects of both learning and network externality when making decisions. While there are some existing works on combining positive network externality with learning [27]–[29], few works have been done on combining negative network externality with learning in the literature mainly due to the difficulty of the problem, where a user has to consider the previous users' decisions and predict the subsequent users'. Furthermore, the information leaked by a user's decision may eventually impair the utility the user can obtain. However, in practice, negative network externality commonly exists in social systems where users share and/or compete with resources and contents. To address this issue, we have developed a joint learning-decision making framework, called Chinese Restaurant Game [30], [31], to study users' sequential learning and decision-making behavior in social systems.

1) Chinese Restaurant Game Framework: The well-known Chinese restaurant process has been used in various fields including machine/social learning, speech recognition, text modeling, and object detection in images and biological data clustering [32]. It offers an ideal structure to jointly formulate the decision making problems with negative network externality. The Chinese restaurant process is a non-parametric learning method for unbounded number of objects in machine learning. In a Chinese restaurant process, a restaurant has infinite number of tables and customers arrive the restaurant sequentially. When a customer enters, he/she either joins one of the existing tables or requests a new table with a predetermined probability. However, there is not yet any notion of strategic decision-making in Chinese restaurant process.

By introducing the strategic behavior into the non-strategic Chinese restaurant process, we proposed a new framework, called Chinese Restaurant Game [30], [31], to study the learning and decision-making problem with negative network externality. To illustrate the framework, let us start with a Chinese restaurant with fixed number of tables, and customers sequentially come in requesting for seats from these tables. Each customer may request a table to sit. Since tables are available to all customers, there may be multiple customers requesting to sit at the same table, which thereafter incurs the negative network externality. We can imagine the more personal space a customer has, the more comfortable in dining experience. Moreover, when the table sizes are unknown to the customers (before arriving the restaurant), each of them may resort to some "signals" (e.g. through advertisements or previous customers) about the table sizes. By observing previous actions or signals, a user can exercise a learning process to make up the shortcoming of limited knowledge. With the proposed Chinese Restaurant Game, we are able to develop an analytical framework involving the learning and decision-making with negative network externality.

As shown in Figure 10, in the Chinese Restaurant Game, there is a Chinese restaurant with K tables numbered $1, 2, ..., K$ and N customers labeled with $1, 2, ..., N$. The table sizes are determined by the restaurant state $\theta \in \Theta$ and the table size functions $\{R_1(\theta), R_2(\theta), ..., R_K(\theta)\}$. When customer i arrives, he/she receives a signal s_i about the state θ and makes a decision which table to sit such that he/she can maximize his/her utility, based on what was observed and the prediction to future customers' decisions. The prior distribution of the state information is assumed to be known by all customers. The signal is generated from a predefined distribution. Since there are uncertainties on the table sizes, customers who arrive

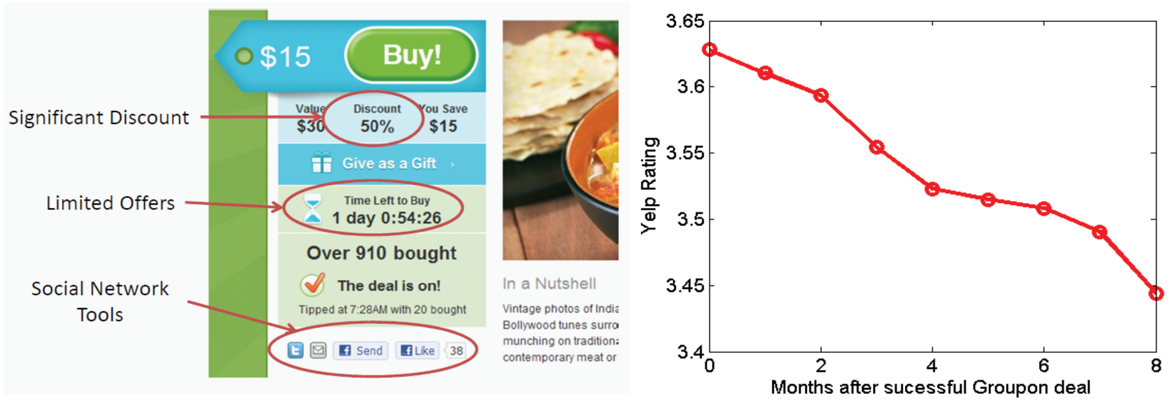
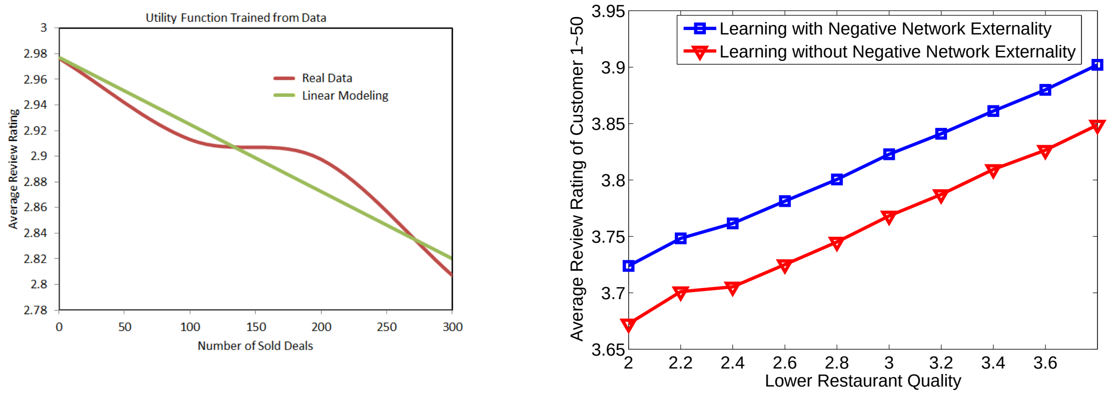


Fig. 11. Yelp star rating declines after a successful Groupon deal.



(a) (b)

Fig. 12. (a) Utility function modeling using real data from Groupon and Yelp; (b) The performance comparison of our method with the learning method without negative network externality.

first may not choose the right tables, due to which their utilities may be lower. On the other hand, customers who arrive later may eventually have better chances to get the better tables since they can collect more information to make the right decisions. In other words, when signals are not perfect, learning can help to result in higher utilities for customers choosing later. Therefore, there is a trade-off between more choices when playing first and more accurate signals when playing later. To study this trade-off, some questions need to be answered: How can customers learn from their own signals and the information revealed by other customers? How can customers predict the decision of future customers and what are the best strategies of customers?

To study how customers learn from the revealed information from others and their own signals, we first introduce the concept of belief to describe customers' uncertainty about the system state. One customer's belief on the system state is the conditional probability of the system state given all the information observed by the customer as follows

$$\mathbf{g}_i = \{g_{i,l} | g_{i,l} = P(\theta = l | \mathbf{h}_i, s_i, \mathbf{g}_0), \forall l \in \Theta\}, \tag{6}$$

where $\mathbf{h}_i = \{s_1, s_2, ..., s_{i-1}\}$ is the signals observed by customer i and $\mathbf{g}_0 = \{g_{0,l} | g_{0,l} = P(\theta = l), \forall l \in \Theta\}$ is the prior distribution.

With Bayesian learning [25], rational customers use Bayes rule to find the optimal estimate about the system state and update their belief on the system state as follows

$$g_{i,l} = \frac{g_{0,l} P(\mathbf{h}_i, s_i | \theta = l)}{\sum_{l' \in \Theta} g_{0,l'} P(\mathbf{h}_i, s_i | \theta = l')}. \tag{7}$$

Due to the rationality and selfish nature, customers will choose their strategies to maximize their own utilities. In such a case, considering the incomplete information about the future customers, the best response of a customer is to maximize his expected utility based on all the observed information as follows

$$\mathbf{BE}_i(\mathbf{n}_i, s_i, \mathbf{h}_i) = \arg\max_j E\left[U_i(R_j(\theta), n_j^*) | \mathbf{n}_i, s_i, \mathbf{h}_i, x_i = j\right], \tag{8}$$

where n_j^* is the final number of customers choosing table j, $U_i(R_j(\theta), n_j^*)$ is the utility of customer i choosing table j, $\mathbf{n}_i = \{n_{i1}, n_{i2}, ..., n_{iK}\}$ is the grouping observed by customer i with n_{ik} being the number of customers choosing table k before customer i, and x_i is the action of customer i.

Note that the best response is determined by the final grouping, which depends on the subsequent customers' decisions. Since the decisions of subsequent customers are unknown to a customer when the customer is making the decision, a

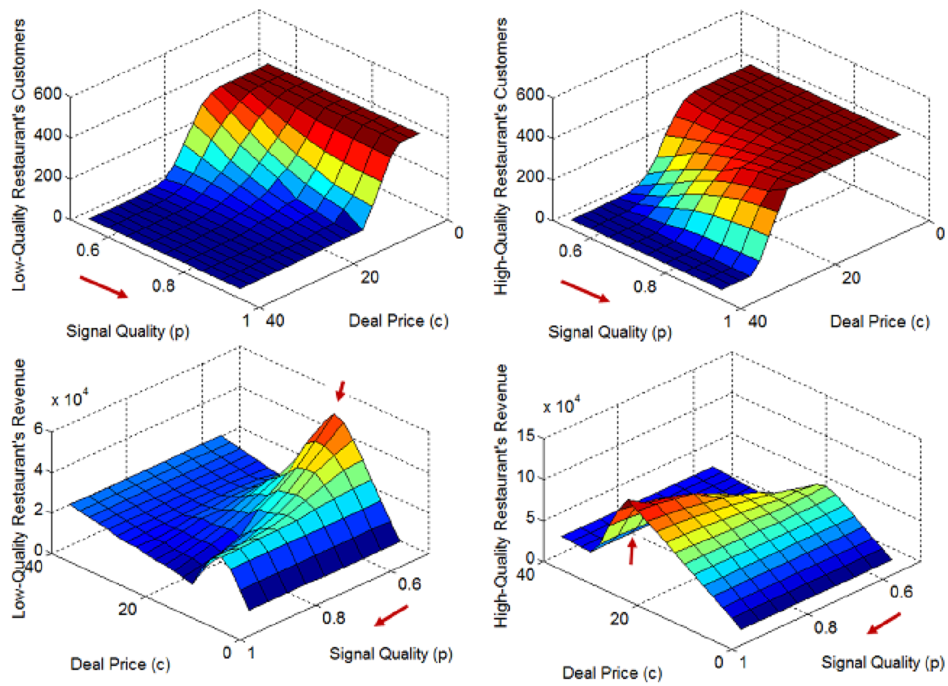


Fig. 13. A new restaurant's strategy: the upper row shows the number of customer choosing the restaurant with low quality and with high quality, respectively, while the lower row shows the result of the revenue of the restaurants.

closed-form solution to the best response function is generally impossible and impractical. To find the best response for each customer, a recursive method based on backward induction is designed [31]. The key idea is to use next customer's best response $\mathbf{BE}_{i+1}(\mathbf{n}_{i+1}, s_{i+1}, \mathbf{h}_{i+1})$ to derive current customer's best response $\mathbf{BE}_{i}(\mathbf{n}_{i}, s_{i}, \mathbf{h}_{i})$.

With the best response function, a customer's optimal decision is purely determined by the received signal given the grouping, i.e., the number of customers on each table, and the information revealed by other customers. Therefore, we can partition the signal space into subspaces where within each subspace, the customer will choose a specific table. By integrating the signal over each subspace, we can derive a recursive form of the probability mass function for the final grouping, i.e., the final number of customers on each tables. With the recursive form of the final grouping, the expected utility of each customer can be computed and the best response of all customers using backward induction can be derived.

From previous discussions, we can see that the learning and decision making in the Chinese Restaurant Game framework are interweaved. On one hand, customers learn the system state from the information revealed by previous customers for better decision makings. On the other hand, the decisions and information revealed by the customers will affect subsequent customers' learning and decision making processes. Moreover, before any decision making, the utility function in (8) needs to be learned from the real data.

2) Experiments with Real Data from Social Systems: We use the deal selection on Groupon as an example to illustrate the Chinese Restaurant Game framework. Many have the experiences that some deals on Groupon look pretty good but eventually turn out to have poor quality due to the overwhelming number of customers showing up at the same time, i.e., the negative network externality is at work here. By collecting the data on Groupon and Yelp around the Washington D.C. area for eight months, we indeed observe the decline of Yelp review rating after some successful Groupon deals, as depicted in Fig. 11. One can see a nonlinear decline function in review rating. Let us use the real Yelp rating data to train the utility function of customers by approximating as a linear model, as shown in Fig. 12-(a). Then, based on (8), we evaluate the average utility, which is the average review rating customers can obtain. A comparison was made between the decision learning method, denoted as "learning with negative network externality", and that does not consider negative network externality, denoted as "learning without negative network externality". Note that the "learning with negative network externality" considers the interplay between the learning and decision making while "learning without negative network externality" only considers the learning of system state but totally ignoring the influence among customers' decision makings. The results are shown in Fig. 12-(b). One can see that by combing learning with negative network externality, the proposed method can achieve much better utility for customers.

We further study the best pricing and promotion strategy

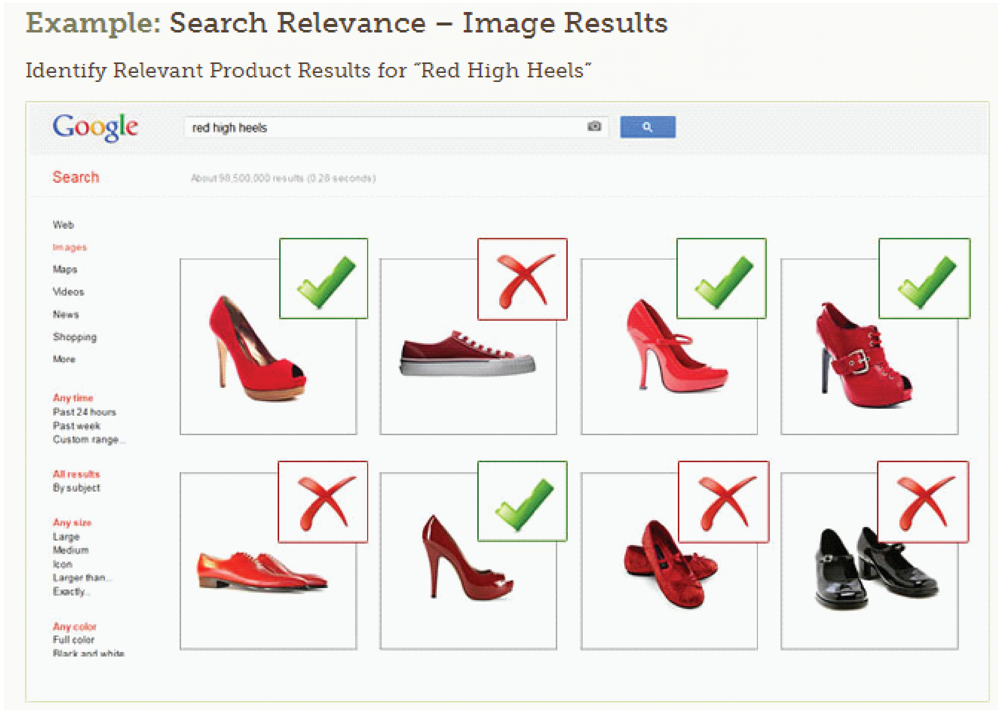


Fig. 14. An example of microtask crowdsourcing.

of a new restaurant under the Chinese Restaurant Game framework. Let us consider two restaurants, one is always of high quality and the other is a new restaurant which could be of low or high quality. The same utility function trained from the real data in the above experiment is used to infer the strategy. The results are shown in Fig. 13. One can see that if the new restaurant is of low quality, then the number of customers choosing the new restaurant decreases as signal quality increases, and vice versa. One can also see that the optimal deal price of the high-quality restaurant is higher than that of the low-quality restaurant. Therefore, high quality restaurant should try every effort to increase the signal quality, while low quality restaurant should hide the quality information and use a low deal price to attract customers to increase the revenue. This offers a vivid example of utilizing data to learn and come out with an optimal strategy.

3) Extension to Chinese Restaurant Game Family: We have discussed the Chinese Restaurant Game under a fixed population setting, i.e., there is a finite number of customers choosing the tables sequentially. However, in some applications, customers may arrive and leave the restaurant at any time, which results in the dynamic population setting. Examples include cloud storage service selection, deal selection on Groupon, and WiFi access point selection in a conference hall [33]. In such a case, the utilities of customers will change from time to time due to the dynamic number of customers on each table. To tackle this challenge, we have extended the Chinese Restaurant Game to the dynamic population setting [34], [35], where we consider the scenario that customers may arrive and leave the restaurant with, for example, a Poisson process. With such a dynamic population setting, each new coming customer not only learns the system state according to the information received and revealed by former customers, but also predicts the future customers' decisions to maximize the utility.

The Chinese Restaurant Game is proposed by introducing the strategic decision-making into the Chinese restaurant process, where each customer can choose one table to maximize the utility. However, in some applications, users may want to simultaneously choose multiple resources. For example, mobile terminals may access multiple channels, cloud users may have multiple cloud storage services, and students may take multiple online courses. To further generalize the setting, we have introduced the strategic decision-making into another well-known random process, Indian buffet process [36], and develop a new framework, called Indian Buffet Game, to study the learning and decision-making problem with negative network externality under the scenario that customers can have multiple choices [37]. In the Indian Buffet Game framework, we also consider multi-slot interactions where customers can interact and make decision repeatedly and partial information reveal where customers only reveal beliefs instead of full signals to others. We use the non-Bayesian social learning to learn from each other to improve the knowledge of the system and thus make better decisions. Similar extension can be applied to multi-armed bandit problems by introducing a decision making processing into its formulation.

III. Mechanism Design in Decision Learning

In this section, we will address decision learning from system point of view, i.e., can we design mechanisms for users to learn the desired behavior and thus achieve goals of the system designer? An example from microtask crowdsourcing is shown to illustrate how to design mechanism for obtaining high-quality data for data analytics.

One key factor for the success of supervised/semi-supervised learning is the large scale labeled dataset [1], [2]. In general, a larger scale dataset will lead to a more accurate model and thus better performance. However, large scale annotation is very expensive, which often becomes one of the bottlenecks of supervised/semi-supervised learning. To address this challenge, microtask crowdsourcing, with the access to large and relatively cheap online labor pool, is a promising way since it can generate large volume of labeled data in a short time at a much lower price compared with traditional in-house solutions. An example of microtask crowdsourcing is illustrated in Fig. 14.

On the other hand, due to the lack of proper incentives, microtask crowdsourcing suffers from quality issues. Since workers are paid a fixed amount of money per task they complete, it is profitable for them to provide random or bad quality solutions in order to increase the number of submissions within a certain amount of time or effort. It has been reported that most workers on Mturk, a leading marketplace for microtask crowdsourcing, do not contribute high quality work [38]. To address this issue, a common machine learning solution is to either add a data curation phase to filter out low quality data or to modify the learning algorithm to accept noisy labels [39]–[43].

Different from existing machine learning solutions, we tackle such a problem by incentivizing the high quality data from the first place [44], e.g., from the workers. This problem is challenging due to the inherent conflict between incentivizing high quality solutions from workers and maintaining the low cost advantage of microtask crowdsourcing for requesters. On one hand, requesters typically have a very low budget for each task in microtask crowdsourcing. On the other hand, the implementation of incentive mechanisms is costly as the operation of verifying the quality of submitted solutions is expensive [39]. Such a conflict makes it extremely challenging to design proper incentives for microtask crowdsourcing. Therefore, it motivates us to ask the following question: what incentive mechanisms should requesters employ to collect high quality solutions in a cost-effective way? In a general sense, the core problem is how to design mechanism for obtaining "good" data?

To answer the question above, we first study and model the behavior of workers. Specifically, let us consider a model with strategic workers, where the action of a worker is the quality of the solution $q \in [0, 1]$, and the primary objective of a worker is to maximize his own utility, defined as the reward he/she will receive minus the cost of producing solutions of a certain quality $c(q)$. Based on this model, we analyze two basic mechanisms that are widely adopted in existing microtask crowdsourcing applications: reward consensus mechanism M_c and reward accuracy mechanism M_a [44].

A. Reward Consensus Mechanism M_c

With this mechanism, a task is assigned to multiple workers. Only the same answer that is submitted by the majority of workers will be chosen as the correct solution, and the workers whose solution agrees with the correct one will receive a positive reward. Through analyzing this mechanism, we find that there exists a minimum mechanism cost per task in order to obtain high quality solutions [44],

$$C^*_{M_c} = 3c'(1), \tag{9}$$

where $c'(1)$ is the first order derivative of the cost function $c(q)$ evaluating at the desired solution quality $q = 1$.

B. Reward Accuracy Mechanism M_a

This mechanism assigns each task only to one worker. The requester evaluates with a certain probability the quality of submitted solutions directly, where each validation incurs a constant cost d. The validation can be erroneous with a probability of ϵ. The workers whose solution is not evaluated or evaluated and confirmed as correct solution will receive a positive reward. Through analyzing this mechanism, we again find that there exists a minimum mechanism cost per task in order to obtain high quality solutions [44],

$$C^*_{M_a} = \begin{cases} 2\sqrt{\frac{c'(1)d}{1-2\epsilon}} - \epsilon\frac{c'(1)}{1-2\epsilon}, & \text{if } d \geq \frac{c'(1)}{1-2\epsilon}; \\ \frac{c'(1)(1-\epsilon)}{1-2\epsilon} + d, & \text{otherwise.} \end{cases}, \tag{10}$$

From (9) and (10), we can see that in order to obtain high quality solutions using the two basic mechanisms (M_c and M_a), the unit cost incurred by requesters per task is subject to a lower bound constraint, which is beyond the control of requesters. In case that the budget of the requester is lower than the minimum cost constraint, it becomes impossible for the requester to achieve desired quality solutions with these two basic mechanisms. In other words, neither of these two basic mechanisms is cost-effective.

C. Incentive Mechanism via Training M_t

To tackle this challenge, we design a cost-effective mechanism by employing quality-aware worker training as a tool to stimulate workers to provide high quality solutions [44]. Different from current microtask crowdsourcing applications where training tasks are usually assigned to workers at the very beginning and are irrelevant to the quality of submitted solutions, we use the training tasks in a more effective way by assigning them to workers when they perform poorly. That is when a worker performs poorly, he/she will be enforced to enter a training session without reward to regain accreditation to be able to go back to perform in the regular session with reward.

With the introduction of quality-aware training tasks, there will be two system states in our proposed mechanism: the working state and the training state. The working state is for production purpose where workers work on standard tasks in return for reward, while the training state is an auxiliary state where workers do a set of training tasks to gain qualifications for the working state. The state transition diagram is shown in Fig. 15, where $P_w(\tilde{q}_w, q_w)$ represents the probability of a solution with quality q_w being accepted in the working state when other submitted solutions from working state are of quality $\tilde{q}_w$, and $P_t(q_t)$ is the probability of a worker who

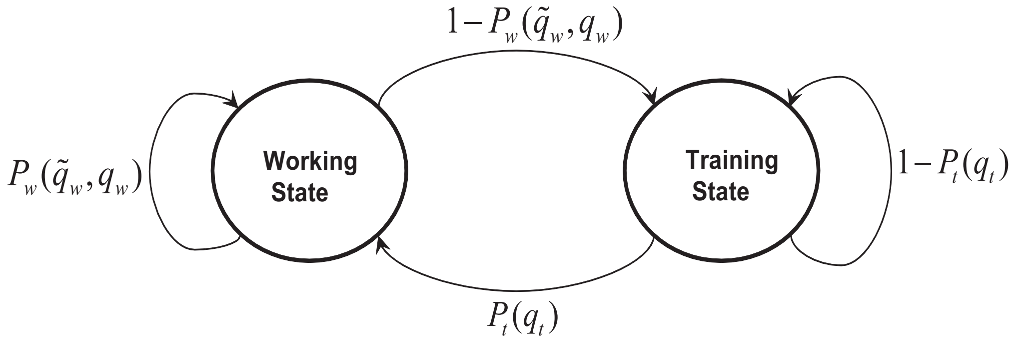


Fig. 15. The state transition diagram of the incentive mechanism M_t.

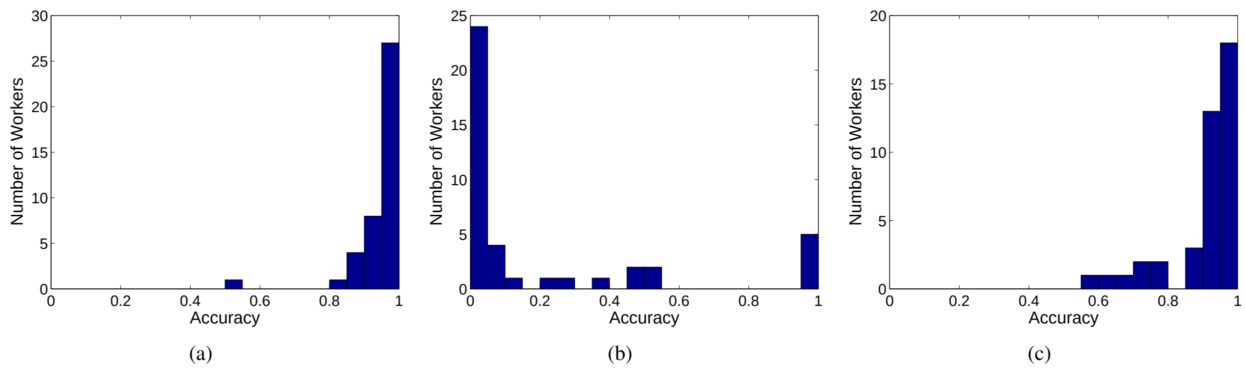


Fig. 16. Histogram of accuracy: (a) the results using the reward accuracy mechanism M_a with sampling probability 1; (b) the results using the reward accuracy mechanism M_a with sampling probability 0.3; (c) the results using the incentive mechanism M_t with sampling probability 0.3.

produces solutions of quality q_t at the training state being allowed to enter the working state next time.

From Fig. 15, we can see that the current action of a worker will affect the future system state of the worker. In other words, the quality of a worker's solution to one task will affect not only the worker's immediate utility but also his future utility due to the possible change of the system state. Such a dependence provides requesters with an extra degree of freedom in designing incentive mechanisms and thus enables them to collect high quality solutions while still having control over their incurred costs.

To find the optimal action, each worker must solve a Markov Decision Process (MDP) according to the state transition diagram shown in Fig. 15, and the MDP faced by each worker also depends on other workers' actions. In essence, this is a challenging game-theoretic MDP problem [44]. Through analyzing the incentive mechanism M_t, we find that as long as the number of training tasks is large enough, there always exists a desirable equilibrium where workers submit high quality solution at working state. In other words, given any parameters in the working state, one can always guarantee the existence of desirable equilibrium through the design of the training state. When the desirable equilibrium is adopted by all workers by floowing a certain design procedure, the minimum mechanism cost is theoretically proved to be 0 [44], i.e.,

$$C^*_{M_t} = 0, \tag{11}$$

which means that one can collect high quality solutions with an arbitrarily low cost. In other words, given any pre-determined budget, the incentive mechanism M_t enables the requester to collect high quality solutions while still staying within the budget.

Notice that one can easily achieve better learning purposes with the high quality data collected by the incentive mechanism M_t. Therefore, through modeling and analyzing users' strategic decision making process, one can design mechanisms from system point of view to steer users' strategic behaviors to obtain better quality data for better learning.

D. Real Behavioral Experiments

A set of behavioral experiments are conducted to test the incentive mechanism M_t in practice. We evaluate the performance of participants on a set of simple computational tasks under different incentive mechanisms. We compare the incentive mechanism M_t with the reward accuracy mechanism M_a where the quality of submitted solutions is evaluated with a certain probability.

There are 41 participants in our experiments, most of whom are engineering graduate students. We use the accuracy of each participant as an indicator to the effectiveness of incentive mechanisms, and the results are shown in Fig. 16, where (a) shows the results of the reward accuracy mechanism M_a with sampling probability 1, i.e., every submitted solution is evaluated; (b) shows the results of the reward accuracy mechanism M_a with sampling probability 0.3, i.e., 30% of the submitted solutions are evaluated; and (c) shows the results using the incentive mechanism M_t with the sampling probability as that in (b).

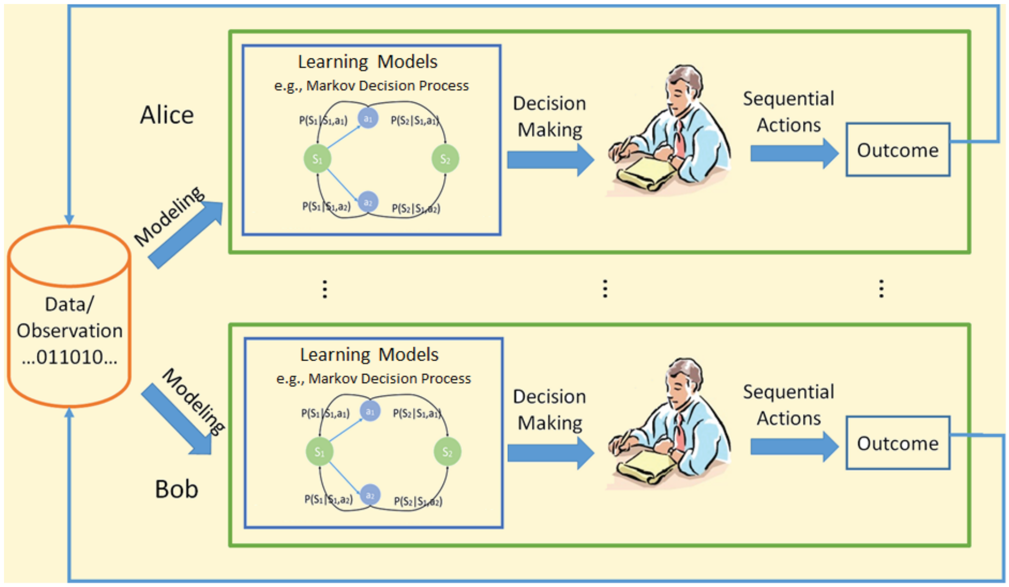


Fig. 17. The Coming Big Data Tsunami.

As shown in Fig. 16-(a), with the highest sampling probability, most participants respond positively by submitting solutions with very high qualities. There is only one participant who had relatively low accuracy compared with others in that he was playing the strategy of "avoiding difficult tasks" according to our exit survey. When a much lower sampling probability of 0.3 is used, it becomes profitable to increase the number of submissions by submitting lower quality solutions, as most errors will simply not be detected. This explains why the majority of participants had very low accuracies as shown in Fig. 16-(b). Noteworthily, a few workers, 5 out 41, still exhibited very high accuracies in this case. Our exit survey suggests that their behaviors are influenced by a sense of "work ethics", which prevents them to play strategically to exploit the mechanism vulnerability. With the incentive mechanism M_t, as the introduction of training tasks makes it more costly to submit wrong solutions, participants need to re-evaluate their strategies to achieve a good tradeoff between accuracy and the number of submitted tasks. From Fig. 16-(c), we can see that the accuracy of participants with the incentive mechanism M_t has a very similar distribution as that using the reward accuracy mechanism M_a with the highest sampling probability. Therefore, through the use of quality-aware worker training, the incentive mechanism M_t can greatly improve the effectiveness of the basic reward accuracy mechanism M_a with a low sampling probability to a level that is comparable to the one that has the highest sampling probability.

IV. Related Works

Albeit not termed as decision learning, there has been a growing body of literature in recent years on the intersection of learning and strategic decision making, mostly from the computer science community. One class of related works is learning to understand how human beings make strategic decisions from real data. For example, classical machine learning techniques are used in [45] to predict how people make and respond to offers during the negotiation and how they reveal information and their response to potential revelation actions by others. Their results showed that the strategies derived from machine learning algorithms, even still not optimal, can beat that of the real human beings [45]. The study of year-long empirical data shows that an experienced human being in a repeated game will be more cooperative, but turn the table more definitely when he is betrayed by the opponent [46]. Additionally, the study in [47] shows that human beings behave as having very limited memory space and computation capability, which limits the optimality of their decisions. It has also been shown in [48] that a dynamic belief model by ignoring the older signals in constructing the belief works best in predicting human being decisions. Through empirically analyzing the purchase history on Taobao, a large-scale online shopping social network, Guo et al. revealed that a real human values purchase experiences shared by his friend and would pay higher price for trustful vendors [49]. Nevertheless, in such a complicated system, it is still difficult to predict the purchase decisions which accuracy over 50% using traditional machine learning algorithms [49]. In [50], how users make decisions on social computing systems is learnt from real data, and used to guide the design of mechanism for the systems.

Another class of related works is finding equilibrium through learning. Finding Nash equilibrium is critical yet challenging in most game models since its difficulty has been shown to be PPAD in general settings or even NP-complete in specific problems [51]. Given that a general and exact solution is intractable, it is a natural choice to design proper learning algorithms to find the solutions. No-regret learning, for instance, has been shown to be a practical candidate. It has been applied in extensive-form game to reduce the number of subgame trees to explore [52]. The sufficient conditions for such type of learning algorithms to converge in the selfish routing problem [53] are also theoretically studied. Reinforcement learning is another candidate since its action-reward structure naturally forms the best response dynamic in game theory. Since traditional Q-learning may fail to converge if directly applied in a game, especially when the Nash equilibrium is not unique, maxmin Q-learning is proposed in [54] to find the Nash equilibrium in two-player zero sum game. The objective of maxmin Q-learning is modified from pure reward maximization into a maxmin problem with opponent's actions in mind. Nash Q-learning, a more general Q-learning algorithm is proposed in [55] to handle multi-player game with non-zero sum. The objective of Nash Q-learning is replaced with equilibrium conditions defined in game theory. The experiment results show that Nash Q-learning can help identify better Nash equilibrium than traditional Q-learning algorithm. Learning has also been used to reduce the complexity in finding the subgame-perfect Nash equilibrium in a sequential game [56], which is PSPACE-hard in general. In [57], MDP and Monte Carlo simulation is used to reduce the complexity in identifying the optimal bidding strategy in sequential auctions.

There have also been some related works that formulate the training problem in machine learning as a game. For instance, it has been shown in [58] that a class of online learning algorithms can be modeled as a drifting game with both trainer and the system as players. The learning algorithm in such a formation becomes the best response of the trainer to the system's reply to each training problem. Another application is maintaining fairness in multi-agent sequential decision problem. Given that the objective of the system is max-min fairness, one may model the learning model as a two-player game, where the first player aims at maximizing the utility of the target agent who is chosen by second player, while the second player chooses the agent with lowest utility as the target agent [59].

Active learning is another related field [60]. Through actively choosing which data to learn from, active learning algorithms have the potential to greatly reduce the amount of labeling effort in machine learning algorithms. Active learning with explicit labeling cost has been widely studied [61]–[64]. It is known that active learning algorithms degrade quickly as the noise rate of labels increases [65]. To address the quality issue in labeled data collection, a variety of approaches have been proposed to filter low quality labels and to increase the robustness of machine learning algorithms [39]–[43]. In [66], a game-theoretic dynamics was proposed to approximately denoise the data to exploit the power of active learning. Incentive mechanisms have also been utilized to improve the quality of collected data. In [12], [67], [68], all-pay auctions are applied to incentively high quality user contributions. In [69], Shaw et al. conducted an experiment to compare the effectiveness of a collection of social and financial incentive mechanisms. A reputation-based incentive mechanism was proposed and analyzed for microtask crowdsourcing in [70]. In [71] and [72], Singer and Mittal proposed an online mechanism for microtask crowdsourcing where tasks are dynamically priced and allocated to workers based on their bids. In [73], Singla and Krause proposed a posted price scheme where workers are offered a take-it-or-leave-it price offer.

A non-strategic but related problem is the multi-armed bandit problem [74]. In the multi-armed bandit problem, a bandit with multiple arms is provided to a gambler. The gambler may have different levels of rewards by playing different arms each time. Thus, the gambler may try and learn in each play in order to maximize his collected rewards. Liu and Zhao extended this model by considering multiple agents and including the network externality in [75]. They studied how agents learn the expected payoff and other agent's choice by estimating the regrets after choosing different arms based on his current belief. A multi-armed bandit problem with costs in observations is discussed in [76]. Nevertheless, traditional study in multi-armed bandit problem is generally non-strategic. They assume agents will always follow the learning rule designed by the system designer. Combining strategic thinking with multi-armed bandit problem has gained more attentions recently mainly due to a popular and practical application: website ad auction. The ad slot on a website is usually sold through auction. The value of the ad depends on two factors: the value of the product in the ad, and the expected number of clicks on the ad. The former one is known by the advertiser and can be collected through truthful auction such as Vickery auction. Nevertheless, the expected number of clicks, or the Click-Through-Rate (CTR) of the ad, is unknown to both the website owner and the advertiser. No-regret algorithm in multi-armed bandit problem can be used to learn CTR while maintaining the truthfulness of the auction [77] [78]. In [79], it is shown that increasing number of explore stages will push the buyers to reveal their true valuation more, with fewer exploit stages for sellers to gain extra revenue from the learned valuation in return.

V. Conclusion and Final Thought

Decision learning is learning with strategic decision making that can analyze users' optimal behaviors from users' perspectives while design optimal mechanisms from system designers' perspectives. In this paper, we have used three social media examples to highlight the concepts of decision learning. Specifically, information diffusion over online social networks was used to illustrate how to learn users' utility functions from real data for understanding/modeling strategic decision making; deal selection on Groupon/Yelp was used to discuss how to learn from each other's interactions for better strategic decision making; and microtask crowdsourcing was used to discuss how to design mechanism to steer users' strategic behaviors to obtain better quality data for better learning. Besides the three examples discussed in this paper, there can be other forms of joint learning with strategic decision making, including those discussed in section IV. In essence, in the coming big data tsunami, when a large volume of data is available, users can learn better models to improve their own decision making, as depicted in Fig. 17. On the other hand, their actions result in changes/modifications of the data pool, which consequently affects the models learned by the users.

In summary, users' decisions/actions affect each others in an ever changing fashion for user-generated data applications. Decision learning is an emerging research area to bridge learning from large volumes of data with strategic decision making that models/understands user behaviors.

REFERENCES

[1] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.

[2] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*. The MIT Press, 2012.

[3] D. Fudenberg and J. Tirole, *Game Theory*. The MIT Press, 1991.

[4] M. J. Osborne, *An Introduction to Game Theory*. New York: Oxford Univ. Press, 2004.

[5] M. Katz and C. Shapiro, "Technology adoption in the presence of network externalities," *J. Political Economy*, pp. 822–841, 1986.

[6] Y. Chen and K. J. R. Liu, "Understanding microeconomic behaviors in social networking: An engineering view," *IEEE Signal Processing Magazine*, vol. 29, no. 2, pp. 53–64, 2012.

[7] H. V. Zhao, W. Lin, and K. J. R. Liu, *Behavior Dynamics in Media-Sharing Social Networks*. Cambridge University Press, 2011.

[8] J. Leskovec, L. Backstrom, and J. Kleinberg, "Meme-tracking and the dynamics of the news cycle," in *International Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 497–506, 2009.

[9] M. Burke, C. Marlow, and T. Lento, "Feed me: motivating newcomer contribution in social network sites," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2009.

[10] K. Ling, G. Beenen, X. Wang, K. Chang, D. Frankowski, P. Resnick, and R. Kraut, "Using social psychology to motivate contributions to online communities," *Journal of Computer Mediated Communication*, vol. 10, no. 4, 2005.

[11] J. Yang, L. A. Adamic, and M. S. Ackerman, "Crowdsourcing and knowledge sharing: strategic user behavior on taskcn," in *Proceedings of the 9th ACM conference on Electronic commerce (EC)*, 2008.

[12] D. DiPalantino and M. Vojnovic, "Crowdsourcing and all-pay auctions," in *Proceedings of the 10th ACM conference on Electronic commerce (EC)*, 2009.

[13] Y. R. Tausczik and J. W. Pennebaker, "Participation in an online mathematics community: differentiating motivations to add," in *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW)*, 2012.

[14] A. Anderson, D. Huttenlocher, J. Kleinberg, and J. Leskovec, "Steering user behavior with badges," in *Proceedings of the 22nd international conference on World Wide Web (WWW)*, 2013.

[15] J. M. Smith, *Evolutionary and the theory of games*. Cambridege University Press, 1982.

[16] M. A. Nowak and K. Sigmund, "Evolutionary dynamics of biological games," *Science*, vol. 303, pp. 793–799, 2004.

[17] H. Ohtsuki, M. A. Nowak, and J. M. Pacheco, "Breaking the symmetry between interaction and replacement in evolutionary dynamics on graphs," *Phys. Rev. Lett.*, vol. 98, no. 10, p. 108106, 2007.

[18] P. Shakarian, P. Roos, and A. Johnson, "A review of evolutionary graph theory with applications to game theory," *Biosystems*, vol. 107, no. 2, pp. 66–80, 2012.

[19] H. Ohtsukia and M. A. Nowak, "The replicator equation on graphs," *J. Theor. Biol.*, vol. 243, pp. 86–97, 2006.

[20] C. Jiang, Y. Chen, and K. J. R. Liu, "Evolutionary dynamics of information diffusion over social networks," *IEEE Transactions on Signal Processing*, vol. 62, no. 17, pp. 4573–4586, Sept. 2014.

[21] C. Jiang, Y. Chen, and K. J. R. Liu, "Graphical evolutionary game for information diffusion over social networks," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 4, pp. 524–536, August 2014.

[22] J. Leskovec, *Stanford large network dataset collection*. [Online]. Available: http://snap.stanford.edu/data.

[23] V. Bala and S. Goyal, "Learning from neighbours," *Rev. Econ. Studies*, vol. 65, no. 3, p. 595, 1998.

[24] B. Golub and M. O. Jackson, "Naive learning in social networks and the wisdom of crowds," *Amer. Econom. J.: Microeconomics*, vol. 2, no. 1, pp. 112–149, 2010.

[25] D. Acemoglu, M. A. Dahleh, I. Lobel, and A. Ozdaglar, "Bayesian learning in social networks," *Rev. Econom. Studies*, vol. 78, no. 4, pp. 1201–1236, 2011.

[26] D. Acemoglu and A. Ozdaglar, "Opinion dynamics and learning in social networks," *Dyn. Games Appl.*, vol. 1, pp. 3–49, 2011.

[27] J. Wit, "Social learning in a common interest voting game," *Games Econ. Behav.*, vol. 26, no. 1, pp. 131–156, 1999.

[28] M. Battaglini, "Sequential voting with abstention," *Games Econ. Behav.*, vol. 51, no. 2, pp. 445–463, 2005.

[29] A. Dasgupta, "Coordination and delay in global games," *J. Econ. Theory*, vol. 134, no. 1, pp. 195–225, 2007.

[30] C. Y. Wang, Y. Chen, and K. J. R. Liu, "Chinese restaurant game," *IEEE Signal Processing Letters*, vol. 19, no. 12, pp. 898–901, 2012.

[31] C. Y. Wang, Y. Chen, and K. J. R. Liu, "Sequential chinese restaurant game," *IEEE Trans. on Signal Processing*, vol. 61, no. 3, pp. 571–584, 2013.

[32] J. Pitman, *Combinatorial Stochastic Processes*. Lecture Notes for St. Flour Summer School. Springer-Verlag, New York, NY, 2002.

[33] Y. H. Yang, Y. Chen, C. X. Jiang, C. Y. Wang, and K. J. R. Liu, "Wireless access network selection game with negative network externality," *IEEE Trans. on Wireless Communications*, vol. 12, no. 10, pp. 5048–5060, 2013.

[34] C. Jiang, Y. Chen, Y.-H. Yang, and K. J. R. Liu, "Dynamic chinese restaurant game: Theory and application to cognitive radio networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 1960–1973, April 2014.

[35] C. Y. Wang, Y. Chen, H. Y. Wei, and K. J. R. Liu, "Optimal pricing in stochastic scalable video coding multicasting system," in *IEEE INFOCOM Mini-Conference*, April 2013.

[36] T. L. Griffiths and Z. Ghahramani, "The indian buffet process: An introduction and review," *Journal of Machine Learning Research*, vol. 12, pp. 1185–1224, 2011.

[37] C. Jiang, Y. Chen, Y. Gao, and K. J. R. Liu, "Indian buffet game with negative network externality and non-bayesian social learning," *to appear in IEEE Transactions on Systems, Man and Cybernetics: Systems*.

[38] P. Wais, S. Lingamnei, D. Cook, J. Fennell, B. Goldenberg, D. Lubarov, D. Marin, and H. Simons, "Towards building a high-quality workforce with mechanical turk," in *NIPS Workshop on Computational Social Science and the Wisdom of Crowds*, 2010.

[39] P. G. Ipeirotis, F. Provost, and J. Wang, "Quality management on amazon mechanical turk," in *HCOMP 10: Proceedings of the ACM SIGKDD Workshop on Human Computation*, pp. 64–67, 2010.

[40] J. Whitehill, P. Ruvolo, T. Wu, J. Bergsma, and J. Movellan, "Whose vote should count more: Optimal integration of labels from labelers of unknown expertise," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 2035–2043, 2009.

[41] P. Welinder, S. Branson, S. Belongie, and P. Perona, "The multidimensional wisdom of crowds," in *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[42] V. C. Raykar, S. Yu, L. H. Zhao, G. H. Valadez, C. Florin, L. Bogoni, and L. Moy, "Learning from crowds," *Journal of Machine Learning Research*, vol. 11, pp. 1297–1322, 2010.

[43] V. C. Raykar and S. Yu, "Eliminating spammers and ranking annotators for crowdsourced labeling tasks," *Journal of Machine Learning Research*, vol. 13, pp. 491–518, 2012.

[44] Y. Gao, Y. Chen, and K. J. R. Liu, "On cost-effective incentive mechanisms in microtask crowdsourcing," *to appear in IEEE Transactions on Computational Intelligence and AI in Games*.

[45] N. Peled, Y. K. Gal, and S. Kraus, "An Agent Design for Repeated Negotiation and Information Revelation with People ," in *The 27th AAAI Conference on Artificial Intelligence*, 2013.

[46] W. Mason, S. Suri, and D. J. Watts, "Long-run learning in games of cooperation," in *Proceedings of the fifteenth ACM conference on Economics and computation (EC)*, pp. 821–838, June 2014.

[47] D. Fudenberg and A. Peysakhovich, "Recency, records and recaps: learning and non-equilibrium behavior in a simple decision problem," in *Proceedings of the fifteenth ACM conference on Economics and computation (EC)*, pp. 971–986, Jun. 2014.

[48] S. Zhang and A. J. Yu, "Forgetful Bayes and myopic planning: Human learning and decision-making in a bandit setting," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 2607–2615, 2013.

[49] S. Guo, M. Wang, and J. Leskovec, "The role of social networks in online shopping: information passing, price of trust, and consumer choice," in *Proceedings of the 12th ACM conference on Electronic commerce (EC)*, pp. 157–166, 2011.

[50] Y. Gao, Y. Chen, and K. J. R. Liu, "Understanding sequential user behavior in social computing: to answer or to vote?," *submitted to IEEE Transactions on Network Science and Engineering*.

[51] C. Daskalakis, P. W. Goldberg, and C. H. Papadimitriou, "The complexity of computing a nash equilibrium," *SIAM Journal on Computing*, vol. 39, no. 1, pp. 195–259, 2009.

[52] N. Burch, M. Lanctot, D. Szafron, and R. G. Gibson, "Efficient Monte Carlo Counterfactual Regret Minimization in Games with Many Player Actions," in *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[53] W. Krichene, B. Drighes, and A. Bayen, "On the convergence of no-regret learning in selfish routing," in *The 31st International Conference on Machine Learning (ICML)*, 2014.

[54] M. L. Littman, "Markov games as a framework for multi-agent reinforcement learning," in *The 11th International Conference on Machine Learning (ICML)*, pp. 157–163, 1994.

[55] J. Hu and M. P. Wellman, "Nash Q-learning for general-sum stochastic games," *The Journal of Machine Learning Research*, vol. 4, pp. 1039–1069, Dec. 2003.

[56] T. Kalinowski, N. Narodytska, T. Walsh, and L. Xia, "Strategic Behavior when Allocating Indivisible Goods Sequentially ," in *The 27th AAAI Conference on Artificial Intelligence*, 2013.

[57] A. Greenwald, J. Li, and E. Sodomka, "Approximating Equilibria in Sequential Auctions with Incomplete Information and Multi-Unit Demand," in *Advances in Neural Information Processing Systems (NIPS)*, 2012.

[58] H. Luo and R. E. Schapire, "A Drifting-Games Analysis for Online Learning and Applications to Boosting ," in *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[59] C. Zhang and J. A. Shah, "Fairness in Multi-Agent Sequential Decision-Making ," in *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[60] B. Settles, "Active learning literature survey," in *Technical report, University of Wisconsin-Madison*, 2010.

[61] D. D. Margineantu, "Active cost-sensitive learning," in *Proceedings of IJCAI*, 2007.

[62] B. Settles, M. Craven, and L. Friedlan, "Active learning with real annotation costs," in *Proceedings of the NIPS Workshop on Cost-Sensitive Learning*, 2008.

[63] S. Sabato, A. D. Sarwate, and N. Srebro, "Auditing: Active learning with outcome-dependent query costs," in *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[64] L. Yang and J. Carbonell, "Buy-in-bulk active learning," in *Advances in Neural Information Processing Systems (NIPS)*, 2013.

[65] A. Beygelzimer, S. Dasgupta, and J. Langford, "Importance weighted active learning," in *ICML*, 2009.

[66] M. F. Balcan, C. Berlind, A. Blum, E. Cohen, K. Patnaik, and L. Song, "Active learning and best-response dynamics," in *Advances in Neural Information Processing Systems (NIPS)*, 2014.

[67] S. Chawla, J. D. Hartline, and B. Sivan, "Optimal crowdsourcing contests," in *ACM-SIAM Symposium on Discrete Algorithms(SODA)*, 2012.

[68] R. Cavallo and S. Jain, "Efficient crowdsourcing contests," in *International Conference on Autonomous Agents and Multiagent Systems(AAMAS)*, 2012.

[69] A. Shaw, J. Horton, and D. Chen, "Designing incentives for inexpert human raters," in *Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW)*, 2011.

[70] Y. Zhang and M. van der Schaar, "Reputation-based incentive protocols in crowdsourcing applications," in *IEEE INFOCOM*, 2012.

[71] Y. Singer and M. Mittal, "Pricing tasks in online labor markets," in *Proceedings of HCOMP11: The 3rd Workshop on Human Computation*, 2011.

[72] Y. Singer and M. Mittal, "Pricing mechanisms for crowdsourcing markets," in *Proceedings of the 22nd international conference on World Wide Web (WWW)*, 2013.

[73] A. Singla and A. Krause, "Truthful incentives in crowdsourcing tasks using regret minimization mechanisms," in *Proceedings of the 22nd international conference on World Wide Web (WWW)*, 2013.

[74] D. Berry and B. Fristedt, *Bandit problems: sequential allocation of experiments*. Chapman and Hall London, 1985.

[75] K. Liu and Q. Zhao, "Distributed learning in multi-armed bandit with multiple players," *IEEE Transactions on Signal Processing*, vol. 58, no. 11, pp. 5667–5681, 2010.

[76] Y. Seldin, P. Bartlett, K. Crammer, and Y. Abbasi-Yadkori, "Prediction with Limited Advice and Multiarmed Bandits with Paid Observations ," in *The 31st International Conference on Machine Learning (ICML)*, 2014.

[77] N. R. Devanur and S. M. Kakade, "The price of truthfulness for pay-per-click auctions," *Proceedings of the 10th ACM conference on Electronic commerce (EC)*, 2009.

[78] N. Gatti, A. Lazaric, and F. Trovò, "A truthful learning mechanism for contextual multi-slot sponsored search auctions with externalities," in *Proceedings of the 13th ACM Conference on Electronic commerce (EC)*, pp. 605–622, 2012.

[79] K. Amin, A. Rostamizadeh, and U. Syed, "Learning Prices for Repeated Auctions with Strategic Buyers," in *Advances in Neural Information Processing Systems (NIPS)*, pp. 1169–1177, 2013.