Learning User Preferences for Wireless Services Provisioning

G. Lee, S. Bauer, P. Faratin, J. Wroclawski
M.I.T
Computer Science and Artificial Intelligence Laboratory
gjl@mit.edu, bauer@mit.edu, peyman@mit.edu, jtw@mit.edu

Abstract

The problem of interest is how to dynamically allocate wireless access services in a competitive market which implements a take-it-or-leave-it allocation mechanism. In this paper we focus on the subproblem of preference elicitation, given a mechanism. The user, due to a number of cognitive and technical reasons, is assumed to be initially uninformed over their preferences in the wireless domain. The solution we have developed is a closed-loop user-agent system that assists the user in application, task and context dependent service provisioning by adaptively and interactively learning to select the best wireless data service. The agent learns an incrementally revealed user preference model given explicit or implicit feedback on its decisions by the user. We model this closed-loop system as a Markov Decision Process, where the agent actions are rewarded by the user, and show how a reinforcement learning algorithm can be used to learn a model of the user's preferences on-line in the given allocation mechanism. We evaluate the performance and value of the agent in a series of preliminary empirical user studies.

1. Introduction

We have developed an adaptive personal agent called the Personal Router (PR) which is a small, portable hardware platform that provides wireless network connectivity for mobile users' data communication requirements, provisioning connectivity services to wireless access providers and continuously selecting the most appropriate service as applications, task and user requirements change [10, 5, 11, 18]. Interactions in such a system are naturally non-cooperative in nature. Therefore the principled design and engineering of such markets is theoretically possible using Mechanism Design (MD) tools, where the "Principle's" (the access provider) goal is to design a revelation mechanism that truthfully elicits the valuation of the users (the "agents") for services. However, this modeling task has been problematic for two related reasons: 1) the assumptions MD makes about the knowledge state of the players and 2) none-efficiency constraints faced in practical industrial organization of Internet. Firstly, MD is not a complete theory of design of an allocation *system*, because it provides a solution to only the preference aggregation problem and ignores the preference determination problem, a non-trivial problem in domains such as the Internet [20, 15, 19]. The mechanism/game form is said to induce a game once the preferences, or types, of the individuals are specified [6]. However, although a game form can model the possibility of individually varying preferences, the preferences specified in a game are assumed to be exogenous to the theory of MD. The theory is a theory of asymmetric information, requiring that only the "Principle" be uninformed and the "Agents" be informed (know their preferences) and the preferences be well structured [6]. Such knowledge and structural assumptions are reasonable for informed agents such as firms who invest time and effort to elicit, correct and model (using decision analysis tools such as conjoint analysis [7]) a well-defined closed-form utility function. These preferences are then (optimally, [3]) elicited and aggregated, by the Principle from strategic and informed agents, for computing solutions to some well defined multiagent distributed optimization problem [6]. Secondly, participation in an efficient mechanism is not always guaranteed in complex domains. In fact sellers may be motivated by technological and information, rather than efficiency, criteria and may therefore not have the incentive to participate in an efficient mechanism. Therefore, the mechanism designer may not actually be able to design an incentive mechanism to individually motivate all players to participate in the mechanism.

Conversely, we are interested in the design of allocation mechanisms when both side of the market are symmetrically uncertain of their valuations. The sub-problem we address in this paper is the design of a computational multiagent system for preference determination of end users or "*household*" consumers, who often, *in addition* to the Principle, are assumed to be initially uninformed due to high preference formation and enumeration costs. We begin by

AAMAS'04, July 19-23, 2004, New York, New York, USA.

assuming that in domains like the Internet what mechanism can be implemented is severely restricted because preferences cannot be assumed to be available to the mechanism, even ex-ante . In fact due to both seller and buyer uncertainties over the buyer's valuations the mechanisms actually implemented in the practice tradeoff optimality criteria [19]. The central goal in this paper is then to develop and engineer a generalizable and inductive model of how to learn the preferences of a "household" consumer and incrementally optimize a consumer's choices in the given approximate mechanism. We demonstrate this approach in a specific domain: dynamic wireless service selection problem in a market consisting of multiple access providers. We then motivate and present a commonly adopted take-it-or-leave-it "first hop" network access allocation mechanism. An adaptive agent who interfaces with and assists a user in the service selection decisions in this mechanism is then presented. The goal of the agent is to learn to select better context dependent services by incrementally learning the preferences of the user given minimal user feedback. This contrasts to most current agent-based market mechanisms where users are not part of the mechanism execution because their preferences are given apriori by a stationary and context independent axiomatic model which is then executed autonomously by their agent in the mechanism. Rather, the class of problems of interest in this paper are better characterized as choice-based (rather than model-based [13]), that involves approximate reasoning as opposed to optimization, where objectives and constraints are well defined prior to single/iterative optimal multiagent decision.

The intuition behind the problem is demonstrated in a scenario in section 2. The mechanism and the underlying mathematical optimization problem of the consumer are then presented in section 3. The adaptive optimization framework and the corresponding computational model, a Markov Decision Process, are then motivated and presented in sections 4 and 5 respectively. Select results from pilot user studies are then summarized in section 6. The conclusions and future work are presented in section 7.

2. The Service Selection Problem

Consider the following scenario that illustrate the technical system and cognitive difficulties. Imagine a user who is running some applications on a mobile device in a coffee shop. Further imagine each application requires different network performance guarantees for its basic operation (e.g. bandwidth, jitter, latency, etc.). Applications are run in some context to achieve some tasks (e.g arranging a meeting, transferring files, video-conferencing, etc.) and the user is likely to require some personal desires, such as urgency, that place further none application or task related constraints on the service the user requires. Assume services have profiles that describe minimally two axes of a service: 1) performance (bandwidth, jitter, latency) and 2) price (initial price and marginal usage price, often as micropayments). Further imagine, that the users represents and evaluates the utility of each service profile in higher level objectives such as quality and cost. The decision problem of the consumer *for a given context* is then which service profile to select that maximizes the quality and minimizes the cost for the set of application, task and user defined constraints. Now imagine, at some time step later the user is in a different context: different physical location, still running the same applications for the same task and requirements, but now facing supply from a different set of base stations offering different service profiles. In the worst case the user may in fact be in a more hurried state than before. Or maybe they have started another application, or terminated old ones.

The complexity and dynamicity of such domains are endless. Below we present an allocation mechanism and a closed-loop agent-user system that attempts to control for this complexity (see [10, 18] for solutions to more system oriented problems).

3. The Mechanism

The consequence of cognitive and dynamicity constraints of the domain means users are unlikely to have an incentive to participate in informationally and communicatively complex, but allocatively efficient, mechanisms such as bargaining or auctions. In fact, it is known that because of human factors users actually tradeoff efficiency for usability, preferring simple contracts for data/voice services [15, 20]. Likewise, it is unlikely that sellers have an incentive to participate in a complicated dynamic allocation mechanism because they are constrained by numerous factors, such as: a) an individual seller's contribution to the satisfaction of an end-user is only a portion of any end-to-end path and a fraction of what would have to be a globally coordinated distributed allocation mechanism [2, 19]; b) the cost structures, coordination costs, and expansion pressures tend to drive the industry to adopt simpler mechanisms over more complex but "efficient" mechanisms [15, 16]; and c) more fundamentally, the architectural principles that the Internet was founded upon dictate that complicated allocation mechanisms should not be embedded deeply into the network architecture [19]. For these seller and consumer constraints a commonly adopted end user access allocation mechanism is a take-it-or-leave-it mechanism.

More specifically, we assume there are N heterogeneous wireless access providers and M consumers respectively at any instance of time. Consumers are located at some distinct location, and can detect a subset of sellers dictated by the

AAMAS'04, July 19-23, 2004, New York, New York, USA.

power transmission rates of their base stations. We assume the market contains producers with heterogeneous production capabilities, ranging from regional service providers to individually owned base stations who can only resell access and have no service negotiation power because they lack any service production capabilities. For this reason we adopt a mechanism where service proposers (the Principals) and the player who receives the proposal/s (the agents) are the sellers and buyers respectively. We assume, for reasons given above, that services are proposed on a take-it-or-leave-it basis: The seller $i \in N$ strategy space is given by the set of technologically feasible service profiles S_i (see section 3.1). An agent $k \in M$ strategy space is then to accept or reject the service profile $s_i \in S_i, \forall N, \forall S$ proposed by the Principals (see section 3.2). Rejection ends the relationship between the players. A contract is then setting the rules of a secondary (Quality of Service, or commitment) game to be played by the Principal and the agent, which we do not model in this paper.

3.1. Seller's Strategies: Services

We make no attempt to define or analyze the equilibrium of the sellers optimization problem given their strategy space. We omit such definitions and analysis because of the heterogeneity in the industrial organization among sellers and the uninformed nature of the consumers in such a dynamic spot market. Therefore, rather than analyzing equilibrium strategies for the sellers we instead enumerate the strategy space in the take-it-or-leave-it mechanism in this section. Note, we make no attempt to formally model the seller's constrained optimization due to technological factors in this reduced strategy space given these limitations.

More specifically, the current supply of services in the market, $S = \{\times_i s_i\}$, is determined by both the physical location of the user as well as the current production function of the sellers. A service $s_j \in S_i$ for seller i is assumed to be a multi-dimensional set of features (called a *service profile*, s_j) that describe two sets of necessary features: performance and price of a service. The performance of a service is specified through a two bucket profile describing a service in terms of its short term and long term burst characteristics, [21]. The other feature of a service profile s_j is the price of a service. Generally, in Internet pricing, service providers price their services based on tiered peak-rate schedules [9]. Other pricing strategies include usage based prices [12] or more complicated user-specific contracts [17]. Since the costs involved in contracting and competitive pricing is too high in a dynamic wireless domains we restrict the pricing strategies of the sellers to usage pricing only. Such a strategy minimally specifies just two price attributes: price per minute c_{min} and price per kilobyte c_{kb}. Therefore the strategy space of seller i, S_i, is some (constrained) continuous function of service performance guarantees it can make as well as the asking prices.

3.2. Buyer's Strategies

The strategy of each consumer $k \in M$ in this mechanism is then defined as:

$$\pi_k : v(s_j, g, w) \rightarrow \Pi \in \{A, R\},\ \forall i \in N, \forall s_j \in S_i \qquad (1)$$

where Π is the response of the consumer from the set, accept (A) or reject (R). g denotes the current context of the consumer and w represents the weighting over each objective. v in turn is the value function which is a representation induced by a set of rationality axioms, [7, 8]. We then describe *each* access decision problem of the user in context g as a constrained multi-objective maximization problem:

$$\begin{aligned} maximize\ v_m(\mathbf{s_j}, \mathbf{g}, \mathbf{w_m}),&\quad m = 1, 2, \ldots, M; \\ \text{s.t}\ \ g_o(\mathbf{s_j}) \geq 0,&\quad o = 1, 2, \ldots, O; \\ h_p(\mathbf{s_j}) = 0,& p = 1, 2, \ldots, P; \\ s_{il}^{min} \leq s_{il} \leq s_{il}^{max},& l = 1, 2, \ldots, n; \end{aligned} \qquad (2)$$

where m is the set of objectives $v_m : R^n \rightarrow R$ and w_m is the weight of the mth objective. $\mathbf{s_j} = [s_{j1}, \ldots, s_{jn}]$ is the vector of service profile features proposed by the seller in the mechanism, that belongs to the nonempty feasible region D and which is a subset of the seller's feasible service profile space R^n. We restrict ourselves to two objectives ($m = 2$) cost and quality. The set of such solutions to a Multi-Objective Optimization Problem (MOOP) is bounded by variable bounds that restrict each decision variable to some upper s_{jl}^{max} and lower s_{jl}^{min} bounds. For example, a user may have budgetary constraints and/or have thresholded bandwidth perception, etc. These bounds, possibly derived through a combination of psycho and econometric studies, then define the feasible *service profile space* D. Finally, service profiles/solutions may also need to satisfy not only variable bound constraints but also inequality and/or equality constraint functions (g_o and h_p respectively), such as minimal user aspiration levels.

Each access decision problem of the user is then to simultaneously maximize m objectives/criteria by *searching* for a solution in the market that maximizes the m objectives. Solutions that satisfy all three types of constraints are referred to as *feasible solutions*. The objective and the feasible objective spaces are denoted by R^m and $Z(= v(D))$ respectively. Elements of Z are called objective vectors or criteria vectors and are denoted by $v(\mathbf{s_j}, \mathbf{g}, \mathbf{w})$ or $z = (z_1, \ldots, z_m)$, where $z_i = v_i(\mathbf{s_j}, \mathbf{g}, \mathbf{w}) \forall i = 1, \ldots, n$ are objective values or criterion values. Then the vector

AAMAS'04, July 19-23, 2004, New York, New York, USA.
Copyright 2004 ACM 1-58113-864-4/04/0007...$5.00

optimization problem is to select a single solution/service profile vector from the set of seller proposed solutions that optimally solves each objective simultaneously, $z^* = v^* = (v_1^*, \ldots, v_m^*)$. Furthermore, the vector $\mathbf{v}^* \in D$ chosen should be Pareto-optimal—there does not exist another decision vector $\mathbf{s_j} \in D$ such that $v_i(\mathbf{s_j}, \mathbf{g}, \mathbf{w}) \leq v_i(\mathbf{s_j^*}, \mathbf{g}, \mathbf{w}) \forall i = 1, \ldots, m$ and $v_j(\mathbf{s_j}, \mathbf{g}, \mathbf{w}) < v_j(\mathbf{s_j^*}, \mathbf{g}, \mathbf{w})$ for at least one index j.

4. Interactive Optimization

Since the consumer is assumed to be uninformed over their value function then most function evaluation search based mathematical programming solutions (such as a posteriori, a priori and interactive methods [14, 4]) which underlie the implementation of MD are inappropriate because they all assume the value function v is axiomatically given (is model-based) and well defined for the task of generating the Pareto profile set. However, the methodology of interactive methods permits learning of user preference functions because they firstly assume the user is available throughout the solution stages, then *iteratively* generate a subset of the Pareto-optimal set that is presented and evaluated by the decision maker.[1]. The learning framework we adopt is to tightly couple user to agent decisions in a closed looped system—interleaving allocation decision making with interactive search methods that continually learn an approximate user model and update this model sequentially and on-line. The problem then is learning an optimal strategy over sequences of decisions. Learning is a feasible option because the task-environment is forgiving, where the cost of errors (or regret, the difference between acting optimally and suboptimally) is not prohibitive to exploration since the monetary cost of a wrong service are in the order of cents. The interactive learning algorithm we propose maps the MOOP above from an objective space to a state space in a Markov Decision Process (MDP).

5. The Personal Router Agent

The problem of learning a user's preferences is framed as learning a strategy over a sequential decision problem. The search problem for an optimal strategy is solved through reinforcement learning, by incrementally learning an optimal control policy given an estimate of the utility of states, derived from user rewards, in a MDP [10]. Therefore, the agent (the controller) incrementally learns to select the optimal sequence of services given the user's (the system) feedback. Furthermore, preference data learnt in interactions with the reinforcement learner can be used as training data for regression algorithms, such as neural networks, to enable generalization and prediction of the utility of services not yet experienced by the user.

We formally define the service selection problem addressed by the PR agent as a fully observable MDP, defined by a set of states S, a set of actions A, and a transition model. δ. [2] At each time step t, the agent receives an observation r_t from the user. The MDP model is defined by:

States (S): A state in our problem represents the set of available network services l and the context in which it is being used (the *user context* g). At time t, we define the state to be $S_t = (l_t, g_t)$.

Let $D = \{S_1, S_2, \ldots\}$ be the set of all feasible services sellers can produce and D_t be the set of services x_i available to the PR at time t. Let $s_i \in D_t$ be the service in use at time t.

The utility of a service also depends on the context in which it is used. Let G be the set of all user contexts. At time t, the agent determines the context g_t from the currently running application. The agent treats this context as exogenous and does not attempt to predict or affect it.

Agent Actions (A): The goal of the PR is to select network services for the user. At a time t, the agent action a_t is to choose the next service $s_j \in D_t$. Let $A = \{b_1, b_2, \ldots\}$ be the set of all agent actions, where action b_i is to select service s_i. The PR may choose from any of the available services D_t, hence the set of available agent actions at time t is $A_t = \{b_i | s_i \in D_t\}$.

Transition Function (δ): In general an MDP may have probabilistic transitions; however, in our problem the transitions are deterministic and the agent has full control over the services it receives. The transition function $\delta : S \times A \rightarrow S$ determines the next state the MDP enters given a state and agent action. That is, $\delta(s_t, b_i) = (c_t, x_i)$. The user context c may change exogenously at any time, however.

Reward Model (R): At each time step t, the agent determines its reward r_t from the user input. A typical user is unwilling or unable to provide detailed information about their preferences over network services. Therefore we allow only two user inputs: at a time t, the user may either indicate satisfaction with the current service ($r_t = sat$), dissatisfaction with the quality and a request for a higher quality service ($r_t = better$), or dissatisfaction with the cost and a request for a lower cost service ($r_t = cheaper$). Based on this input the agent calculates two values, Δq and Δc, The value $\Delta q = -1$ if $r_t = better$ and $\Delta q = 1$ if $r_t = sat$. Similarly, $\Delta c = -1$ if $r_t = cheaper$ and $\Delta c = 1$ if $r_t = sat$.

Action Value Estimation (V): For each service s_i and user context g that the PR observes, it maintains estimated quality and cost values $V_q(g, s_i) \in [0, 1]$ and $V_c(g, s_i) \in$

1 Note, the method still assumes the existence of a well defined value function to generate the Pareto set.

2 See [1] for a Partially Observable MDP treatment of the preference elicitation problem that, although has the same goal, represents states as utility function.

$[0,1]$. The estimated quality and cost are calculated by averaging the previous Δq and Δc inputs according to the exponential weighted average update rules

$$V_q(g,s_i) \leftarrow \alpha\frac{\Delta q+1}{2}+(1-\alpha)V_q(g,s_i) \quad (3)$$
$$V_c(g,s_i) \leftarrow \alpha\frac{\Delta c+1}{2}+(1-\alpha)V_c(g,s_i) \quad (4)$$

with learning rate $\alpha = 0.1$, where g is the current context and s_i is the currently selected service. If the PR's predictor has not yet been trained, then values of V_q and V_c are initialized optimistically with 0.9, encouraging the agent to try new services. Otherwise, they are initialized to the values $F_q(g,s)$ and $F_c(g,s)$ by the prediction algorithm described below.

The total price c_{total} of a service to the user is a linear time weighted combination of these price attributes where the weight is the duration of usage t and the quantity of data transferred y according to the equation $c_{total} = c_{min}t + c_{kb}y$.

Action Selection: Action selection then involves accepting the service that maximizes a weighted function of the estimated quality-cost estimates, where the weights are also an estimate of the user's quality/cost weight and are updated on user input according to the rule $w \leftarrow w + \Delta w$ and constrained so that $0 \leq w \leq 1$. We choose to simply use a linear action value function $v(s_i, g, w)$ as a starting point:

$$v(s_i,g,w) = wV_q(g,s_i) + (1-w)V_c(g,s_i) + c_{switch} \quad (5)$$

where $c_{switch} = 0.1$ if s is the currently selected service and 0 otherwise. We introduce the variable c_{switch} to account for switching costs in the system.

Actions are selected stochastically to encourage exploration according to a Gibbs softmax distribution of the estimated utility of the available action/services, where the probability of selecting an action/ service s_i with value $v(s_i)$ from the set of available services S is given by the expression

$$\frac{e^{v(s_i,g,w)/\tau}}{\sum_{x\in S} e^{v(x,g,w)/\tau}} \quad (6)$$

where τ is called the annealing temperature.

Action Value Predictor Values of actions/services which have not yet been experienced by the user are initiated using the current action value estimates as training data for a multi-layer neural networks (MNN). The (non-linear) regression output from the MNN is then used as a predicted value of actions/services (see [10, 11] for a more detailed exposition of the MNN).

6. User Experiments

The developed agent's learning performance was tested in static and dynamic service environments in a series of exploratory controlled user experiments. A detailed treatment of the experimental goals, hypothesis and results of previous experiments are given in [10, 11].

6.1. Experimental Setup: Network and Users

The set of available services were generated using traffic shaping in a software router. Services were defined by three features: Average Data Rate (ADR), Cost per Minute (CPM), and Cost per Kilobyte (CPB). The values of these features were chosen to mimic a range of realistic network services, from inexpensive low quality services to expensive high quality services. We chose to use seven quality levels, corresponding to bandwidth levels commonly encountered by users in current 802.11b and broadband, modem, and cellular data networks: 11Mbps, 1Mbps, 384Kbps, 128Kbps, 56Kbps, 28.8Kbps, and 9600bps. Costs were set so that the user must choose services carefully to avoid expending all their credits.

Eight services were available in three simulated locations. For each location there was exactly one optimal service that allowed the user to complete the experiment objectives. All other services were either too costly or too slow. Eight services were chosen because it is a large enough number to make the search task of correct service selection hard enough for the subjects while enabling the PR to learn service values within the time frame of the experiment.

Subjects were 17 students and staff of the MIT Computer Science and Artificial Intelligence Laboratory. Subjects were rewarded for their participation with $10 to $20 based on their performance.

6.2. Procedure

In order to evaluate the ease of use and effectiveness of the agent and user interface the performance of the PR was benchmarked against a manual selection policy. Subjects were randomly assigned to one of two groups: 1) the control group where subjects had to choose between services manually by selecting from a menu displaying the available services and their features and 2) the test group where subjects used the PR to select between services, requesting services using the *better* and *cheaper* buttons described earlier. The final distribution of subjects to groups was 8 control and 9 test conditions.

An experiment consisted of three phases. The first phase controlled for task learning effects. Subjects were given ten minutes to become familiar with the user interface, the procedure and the available services. The second phase consisted of a *static* configuration of all eight services for a particular location and tested how well the PR can learn an estimate of user preferences (estimation tests). The third phase was identical to the second phase but tested for the adequacy of the selected choices when the set of available services *changed* from the set available in the second phase (prediction tests). Within each phase the subjects had to complete one or more tasks. For each task, the subjects had to use a web browser to fully load a series of ten web pages within five minutes while selecting a service using the mechanism designated for their group. Each web page contained four large image files. We chose this task because it approximates the network usage of a typical subject shopping or looking for information on-line and makes use of the network service in a realistic and familiar way. Subjects were instructed that they would be charged for their network usage based on the cost of the current service. The task performance of the subjects was given by the dependent variable *score*, measured as the number of credits expended during that task.

In each phase the simulated location and the set of available services changed. In Phase 1 the subject was placed in Location 1 and was asked to perform their downloading task twice for practice. In Phase 2, the PR was reset and the subject chose services from Location 2. The subject was given two attempts at their task and was instructed to try to minimize their score. Finally in Phase 3, the user attempts to minimize their score in Location 3.

6.3. Results

Previous experiments confirmed that the PR can learn user preferences and select services effectively [10, 11]. The goal of these experiments was to assess the subjects' preferences for services by analyzing the observed ordinal rankings over the set of services. Direct conjoint analysis was not possible because asking pairwise questions over even such a small service set was, in addition to their treatment task, too costly in time and attention since subjects would have to specify 72 pairwise preference relationships after experiencing each service for their task. The observed duration spent on a service s_i ($t(s_i)$) was instead used as a heuristic about the satisfaction of the user with a service's cost and quality. The preference heuristic to compare two services s_i and s_j was $s_i \succ s_j$ *iff* $(t(s_i) \geq t(s_j))$. Usage duration is a reasonable dependent variable because the attention of the users was controlled for in each treatment condition by the task. Therefore, the subjects should have behaved, spent time using a service, "as though" they were maximizing the value of their scores. The correlation between the observed service usage durations and the PR estimates of utility of that service was computed to derive the relationship between the usage duration and the agent estimated orderings. We observed that indeed the usage duration was linearly and positively correlated (with coefficient of $r^2 = 0.736$) with the estimated utility (with $w = 0.625$)

S	1	2	3	-	-	-	-	8
1	1	-	3	-	-	-	-	-
2	-	-	3	-	5	6	-	8
3	1	2	3	-	-	-	7	8
4	-	-	3	-	-	6	-	-
5	1	2	-	4	5	-	-	-
6	1	2	3	4	-	-	7	8
7	1	2	3	-	5	-	-	-
8	-	2	-	-	-	-	-	-
9	1	2	-	-	-	-	-	8
G	1	2	3	-	-	-	-	8

Table 1. Final Pareto Set Induced by the Actual Services, the Users and the Group

of that service by the agent.

We first compared the *final* set of weakly Pareto optimal services induced by each subject's PR estimate of quality and cost of all services (PO_i) with the Pareto optimal set induced by the eight actual experimental services (PO_S). Weakly Pareto optimal services were identified by pairwise comparison of all of the services, $(s_i, s_j), \forall i, j$ and selecting those that satisfied the condition: $(adr(s_i) \geq adr(s_j)) \wedge (cpm(s_i) \leq cpm(s_j))$. That is, service s_i weakly Pareto dominates s_j because it provides at least as much quality at equal or less cost.

Table 1 shows the weak Pareto optimal set for the actual experimental services (PO_S, row S), compared to the final induced set by the PR estimates of each individual subject (PO_i, rows 1 to 9), together with the set induced by the aggregate estimates of the group of subjects (row G). The results show that 5 out of 8 experimental services weakly Pareto dominated others objectively: $PO_S = \{s_1, s_2, s_3, s_6, s_8\}$.[3] By observation we can see that the final Pareto dominating set induced by the PR estimates of the group (PO_G) is similar to PO_S. As suspected more variations in similarity between the actual and PR estimated Pareto set were observed at the level of each individual subject, but nonetheless some similarities obtained.

We also measured the frequency of times the PR ordered pairs of services correctly by either quality or cost alone or both. The observed frequency ratio of each subject for each ordering condition (out of 28 service comparisons) is shown in Table 2. For example, the first subject was observed to correctly order almost 80% of the services correctly using the either the quality or cost criteria alone. However, the accuracy of *all* subjects responses was lower when both cost

3 We observe such a large Pareto set size because we chose to experiment with service profiles whose quality and cost profiles were not random but instead were similar to actual current services.

	Quality	Cost	Quality and Cost
1	0.79	0.79	0.68
2	0.82	0.75	0.68
3	0.93	0.79	0.75
4	0.93	0.75	0.68
5	0.82	0.79	0.64
6	0.57	0.89	0.54
7	0.89	0.75	0.68
8	0.61	0.64	0.46
9	0.82	0.86	0.79
G	0.71	0.89	0.79

Table 2. Actual v.s PR Estimated Quality-Cost Consistency

1	0.57
2	0.86
3	0.86
4	0.86
5	0.57
6	0.71
7	0.29
8	1.00
9	0.57
G	0.86
manual	1.00

Table 3. Usage Duration and Locally Pareto Optimal Ratios

and quality was considered. A value of 0.5 would correspond to random estimates. These results indicate that the quality and cost estimates learned by the PR are fairly consistent with the actual quality and cost of services.

Next we measured if Pareto dominant services were used more by the subjects, in the *process* of arriving to the final service estimations. To support such an inference we used the concept of Local Pareto dominance as the dependent variable. If for s_j there exists another service s_i such that $||s_i - s_j|| \leq \epsilon$, where ϵ is a small positive number, and s_i dominates s_j, then s_i is locally Pareto optimal. In other words, s_i is simply a more optimal service in the *neighborhood* of a service s_j, where s_i does not neccessairly lie on the Pareto frontier (i.e is not the global optimal).

Table 3 shows for each PR user the fraction of service pairs (s_i, s_j) that obey the relation $s_i \succ s_j$ over the set of services such that s_i weakly locally Pareto dominates s_j. The fractions are also given for the aggregate PR and manual selection durations. The table shows that ordering services based on duration of usage is consistent with a preference ordering based on Local Pareto optimality. When presented with two services, one of which weakly Pareto dominates the other locally, the user almost always uses the locally Pareto dominant service for a longer duration than the dominated service. [4]

Finally, we analyzed the subject's search dynamics by analyzing performance of the PR in terms of the timing of user inputs. We define *satisfaction* to be when a service is in use for more than 15 seconds without any button presses. Each time the user pressed a button following a state of satisfaction, we counted the number of button presses before satisfaction was again achieved. Figure 6.3 shows the cumulative probability distribution of the number of these button presses. Shown in the figure are the average results for users of the manual interface and for the PR interface. In addition, we show the results following *better* and *cheaper* button presses in the PR. We observe that the average number of button presses required to satisfy the PR users is less than the average number of presses required under manual selection up to about the 95th percentile. On average, about 65% of the time the PR satisfies the user after just one button press, whereas after just one button press manual selection users were satisfied less than 50% of the time. We also see that PR users as satisfied more quickly after pressing the *cheaper* button than when they press the *better* button. This may be because users are more likely to be satisfied with lower cost services than higher quality services, or perhaps because the PR's selections are more accurate when users request cheaper services.

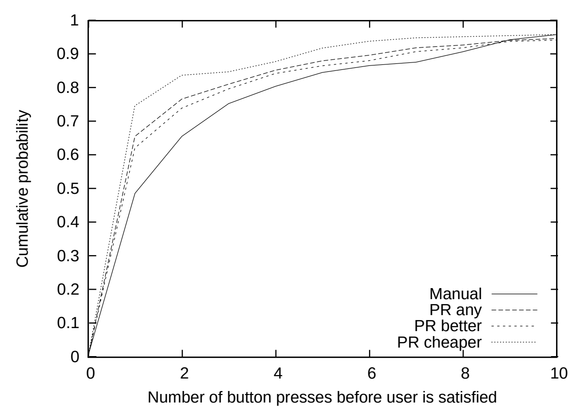

Figure 1. Satisfaction Distribution

4 Recall, that since the observed correlation between the usage duration of the service and the estimated utility of the agent is positive and large we can also assume these results can be applicable to preference relation and estimated utilities.

7. Conclusion

We motivated, presented and evaluated a state-based model of an interactive on-line agent that learns the user's preferences for wireless services in a dynamic and competitive environment. Decisions were made to optimize estimated preferences of users in a take-it-or-leave-it allocation mechanism, a mechanism that was argued to be feasible given real economic, technical and cognitive cost constraints.

Central to our approach is an inductive methodology, because we are interested in engineering rather than analysis of a multiagent system. This is similar to economists' "revealed preference" models, where a preference theory is developed by observing consumer choices. If the observed choices are consistent in a particular way we can then represent them as if consumers maximized utility functions that satisfy the axioms of preferences. The benefit of a more bottom up, data-driven inductive modeling is that no prior assumptions need to be made about the rationality of the players, unlike deductive model-based assumptions of Mechanism Design. However, the cost of this methodology is the difficulty in constructing a theory of system equilibria since no generalization can be established given many sources of conflicting evidence. Nonetheless, this is also a problem for deductive models of MD in domains with large scaled dynamicity, such as wireless/ad-hoc networks, where standard equilibrium solutions no longer apply given the uninformed nature of the players. We have therefore had to trade-off analytical expressiveness for an engineered solution.

Our future goal is to continue to develop better learning models and evaluate whether these estimated models derived from observed consumer data does indeed (weakly or strongly) satisfy the consistency conditions required by the model-based deductive equilibrium models. If so we can then make valid inferences using deductive models. The preliminary empirical work reported here shows that primitive conclusions can be made, even in dynamic environments as wireless networks.

Acknowledgment

This work is supported by the NSF grant ANI-0082503 on Protocols for Open Access Wireless Deployment.

References

[1] G. Boutilier. A POMDP formulation of preference elicitation problems. In *Proceedings of American Association of Artificial Intelligence*, pages 239–246, Edmonton, Alberta, Canada, 2002.

[2] D. Clark. A model for cost allocation and pricing in the internet. *MIT Technical Report*, 2000.

[3] W. Conen and T. Sandholm. Minimal preference elicitation in combinatorial auctions. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCA I), Workshop on Economic Agents, Models, and Mechanisms, August 6th*, Seattle, WA, 2001.

[4] K. Deb. *MultiObjective Optimization using Evolutionary Algorithms*. Wiley, New York, 2001.

[5] P. Faratin, G. Lee, J. Wroclawski, and S. Parsons. Social user agents for dynamic access to wireless networks. In *Proceedings of AAAI Spring Symposium on Human Interaction with Autonomous Systems in Complex Environments*, pages 52–59, Stanford, CA, US, 2003.

[6] M. Jackson. Mechanism theory. *Encyclopedia of Life Support Systems*, 2000.

[7] R. L. Keeney and H. Raiffa. *Decisions with Multiple Objectives*. John Wiley and Sons, New York, 1976.

[8] D. Kreps. *Notes on the Theory of Choice*. Westview, 1988.

[9] R. J. S. L. Nania. *Flat–The Minimalist Price: Internet Economics*. MIT Press, Cambridge, MA, 1997.

[10] G. Lee. Automatic service selection in dynamic wireless networks. Master's thesis, Massachusetts Institute of Technology, September 2003.

[11] G. Lee, P. Faratin, S. Bauer, and J. Wroclawski. A user-guided cognitive agent for network service selection in pervasive computing environments. In *proceedings of The IEEE Second International Conference on Pervasive Computing and Communication (Percom-04), March 14–17*, Orlando, FL, USA, 2004.

[12] J. K. MacKie-Mason and H. R. Varian. Some FAQs about usage-based pricing. *Computer Networks and ISDN Systems*, 28(1–2):257–265, 1995.

[13] A. Mas-Colell, M.D.Whinston, and J.Green. *Microeconomic Theory*. Oxford Press, 1995.

[14] K. Miettinen. *Nonlinear MultiObjective Optimization*. Kluwer International Series, Boston, 1999.

[15] A. M. Odlyzko. Internet pricing and the history of communications. *Computer Networks*, 36:493–517, 2001.

[16] A. M. Odlyzko. The many paradoxes of broadband. *First Monday*, 8(9), 2003.

[17] A. M. Odlyzko. Privacy, economics, and price discrimination on the internet. *Proceedings of ICEC2003: The Fifth International Conference on Electronic Commerce*, pages 355–266, 2003.

[18] Personal Router Research Project. `http://pr.lcs.mit.edu`.

[19] S. Shenker, D. D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: Reshaping the research agenda. *Telecommunications Policy*, 20(3), 1996.

[20] H. R. Varian. Estimating the demand for bandwidth. In *Internet Services: The Economics of Quality of Service in Networked Markes*, Cambridge, MA, 2001. MIT Press.

[21] X. Yang. Designing traffic profiles for bursty internet traffic. In *Proceedings of IEEE Global Internet*, Taipei, Taiwan, 2002.