

A COLLABORATIVE 20 QUESTIONS MODEL FOR TARGET SEARCH WITH HUMAN-MACHINE INTERACTION

Theodoros Tsiligkaridis*, Brian M. Sadler[□] and Alfred O. Hero III^{*†}

University of Michigan, *EECS Dept. and [†] Dept. Statistics, Ann Arbor, MI, USA

[□] US Army Research Laboratory, Adelphi, MD, USA

ttsili@umich.edu, brian.m.sadler6.civ@mail.mil, hero@umich.edu

ABSTRACT

We consider the problem of 20 questions with noise for collaborative players under the minimum entropy criterion [1] in the setting of stochastic search, with application to target localization. First, assuming conditionally independent collaborators, we characterize the structure of the optimal policy for constructing the sequence of questions. This generalizes the single player probabilistic bisection method [1, 2] for stochastic search problems. Second, we prove a separation theorem showing that optimal joint queries achieve the same performance as a greedy sequential scheme. Third, we establish convergence rates of the mean-square error (MSE). Fourth, we derive upper bounds on the MSE of the sequential scheme. This framework provides a mathematical model for incorporating a human in the loop for active machine learning systems.

Index Terms— optimal query selection, human-machine interaction, target localization, convergence rate, minimum entropy.

1. INTRODUCTION

This paper addresses a problem related to maximizing the value of adding a human-in-the-loop to an autonomous learning machine, e.g., an automated target recognition (ATR) sensor. In the ATR setting the objective of the human-machine-interaction is to collaborate on estimating an unknown target location, where the human is repeatedly queried about target location in order to improve ATR performance. We propose a 20 questions framework for studying the value of including the human-in-the-loop and optimizing the sequence of queries.

Motivated by the approach of Jedynak et al [1], which was restricted to the single player case, we model the human-machine interaction as a noisy collaborative 20 questions game. In this framework a controller sequentially selects a pair of questions about target location and uses the noisy responses of the human and the machine to formulate the next pair of questions. Under the minimum expected entropy criterion, we show that even under independence between collaborative players, jointly optimal policies require overlapping non-identical queries. We prove that the expected entropy reduction for the optimal joint design is the same as that of a greedy sequential design. The greedy sequential design consists of a sequence of bisections. This yields a low complexity implementation that is guaranteed to have the same performance as the optimal query controller.

As in Jamieson et al [3], we use a simple noisy query-response model with different reliability functions for the machine and the

human (called derivative-free optimizers (DFO) in [3]). Under this model we specify the optimal query policy, establish a separation theorem, and obtain MSE bounds and convergence rates. Our model predicts that the value of including the human-in-the-loop, as measured by the MSE human gain ratio (HGR), initially increases when localization errors are large, and then slowly decreases over time as the location errors go below the human’s fine resolution capability.

2. NOISY 20 QUESTIONS WITH COLLABORATIVE PLAYERS: ENTROPY LOSS

Assume that there is a target with unknown state $X^* \in \mathcal{X} \subset \mathbb{R}^d$. Our focus in this paper is the case where the target state is spatial location, i.e., in $d = 2$ or 3 dimensions. However, our results are applicable to higher dimensions also, e.g., where X^* is a kinematic state or some other multi-dimensional target feature. Starting with a prior distribution $p_0(x)$ on X^* , the aim is to find an optimal policy for querying a machine (hereafter referred to as player 1), with the additional help of humans. The policy’s objective is to minimize the expected Shannon entropy of the posterior density $p_n(x)$ of target location after n questions.

There are M collaborating players that can be asked questions at each time instant. The objective of the players is to come up with the correct answer to a 20 questions game. Let the m th player’s query at time n be “does X^* lie in the region $A_n^{(m)} \subset \mathbb{R}^d$?”. We denote this query as the binary variable $Z_n^{(m)} = I(X^* \in A_n^{(m)}) \in \{0, 1\}$ to which the player yields provides a noisy response $Y_{n+1}^{(m)} \in \{0, 1\}$.

Define the M -tuples $\mathbf{Y}_{n+1} = (Y_{n+1}^{(1)}, \dots, Y_{n+1}^{(M)})$ and $\mathbf{A}_n = \{A_n^{(1)}, \dots, A_n^{(M)}\}$.

Assumption 1. We assume that the players’ responses are conditionally independent:

$$P(\mathbf{Y}_{n+1} = \mathbf{y} | \mathbf{A}_n, X^*, p_n) = \prod_{m=1}^M P(Y_{n+1}^{(m)} = y^{(m)} | A_n^{(m)}, X^*, p_n) \quad (1)$$

where

$$P(Y_{n+1}^{(m)} = y^{(m)} | A_n^{(m)}, X^*, p_n) = f_1^{(m)}(y^{(m)} | A_n^{(m)}, p_n) I(X^* \in A_n^{(m)}) + f_0^{(m)}(y^{(m)} | A_n^{(m)}, p_n) I(X^* \notin A_n^{(m)}). \quad (2)$$

Assumption 2. We model the players’ responses as binary symmetric channels (BSC) [4] with crossover probabilities $\epsilon_m \in (0, 1/2)$. Therefore the conditional p.m.f. $f_j^{(m)} = P(Y_n^{(m)} = j | A_n^{(m)}, p_n)$ of the response of the M -th player can be written:

$$f_j^{(m)}(y^{(m)} | A_n^{(m)}, p_n) = (1 - \epsilon_m) I(y^{(m)} = j) + \epsilon_m I(y^{(m)} = 1 - j)$$

This research was partially supported by MURI grant W911NF-11-1-0391.

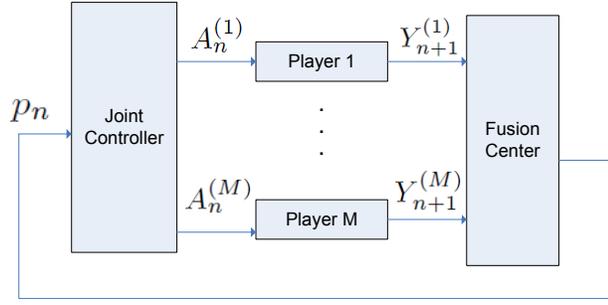


Fig. 1. Joint scheme for M collaborative players responding to binary valued queries about the location X^* of an unknown target. At time n , the controller chooses the queries $I(X^* \in A_n^{(m)})$ based on the posterior p_n . Then, the M players yield responses $Y_{n+1}^{(m)}$ that are fed into the fusion center, where the posterior is updated and fed back to the controller at the next time instant $n + 1$. The target location estimate X_n is the median of the posterior p_n .

where $m = 1, \dots, M, j = 0, 1$.

2.1. Optimal Joint Query Design

We consider a similar setting as in [1], which applied to the $M = 1$ player case, but now we have a joint controller that chooses M queries $A_n^{(m)}$ at time n . The system block diagram is shown in Fig. 1. Define the set of subsets of \mathbb{R}^d :

$$\tilde{\pi}(A^{(1)}, \dots, A^{(M)}) = \left\{ \bigcap_{m=1}^M (A^{(m)})^{i_m} : i_m \in \{0, 1\} \right\}$$

where $(A)^0 := A^c$ and $(A)^1 := A$. The cardinality of this set of subsets is 2^M and these subsets partition \mathbb{R}^d . The objective is to localize the target within a subset $A^{(m)}$.

Define the density parameterized by $\mathbf{A}_n, p_n, i_1, \dots, i_M$:

$$g_{i_1:i_M}(y^{(1)}, \dots, y^{(M)} | \mathbf{A}_n, p_n) := \prod_{m=1}^M f_{i_m}^{(m)}(y^{(m)} | A_n^{(m)}, p_n)$$

where $i_j \in \{0, 1\}$.

Theorem 1. (Joint Optimality Conditions) Under Assumption 1, an optimal joint policy that minimizes the Shannon entropy of the posterior distribution p_n achieving the following entropy loss:

$$G^* = \sup_{A^{(1)}, \dots, A^{(M)}} \left\{ H \left(\sum_{i_1:i_M=0}^1 g_{i_1:i_M} P_n \left(\bigcap_{m=1}^M (A_n^{(m)})^{i_m} \right) \right) - \sum_{i_1:i_M=0}^1 H(g_{i_1:i_M}) P_n \left(\bigcap_{m=1}^M (A_n^{(m)})^{i_m} \right) \right\}, \quad (3)$$

where $H(f)$ is the Shannon entropy of the p.m.f. f .

Proof. Using (1) and (2), we have:

$$P(\mathbf{Y}_{n+1} = \mathbf{y} | \mathbf{A}_n, X^* = x, p_n) = \sum_{i_1:i_M=0}^1 g_{i_1:i_M}(\mathbf{y} | \mathbf{A}_n, p_n) I \left(x \in \bigcap_{m=1}^M (A_n^{(m)})^{i_m} \right). \quad (4)$$

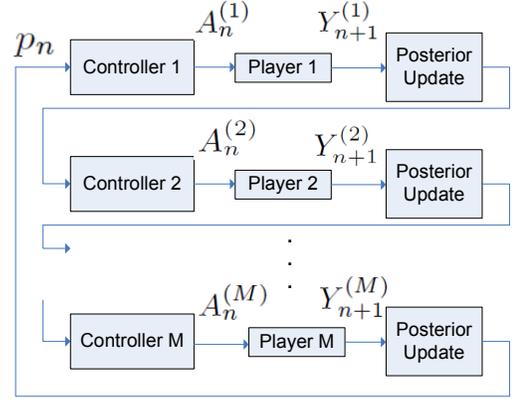


Fig. 2. Sequential scheme for M collaborative players responding to binary valued queries about the location X^* of an unknown target. At time n , the first controller chooses the query $I(X^* \in A_n^{(1)})$ based on the posterior p_n . Then, player 1 yields the response $Y_{n+1}^{(1)}$ that is used to update the posterior, and the second controller chooses the next query $I(X^* \in A_n^{(2)})$ for player 2 based on the updated posterior, etc.

By integrating over $x \in \mathbb{R}^d$, we have:

$$P(\mathbf{Y}_{n+1} = \mathbf{y} | \mathbf{A}_n, p_n) = \mathbb{E}[P(\mathbf{Y}_{n+1} = \mathbf{y} | \mathbf{A}_n, X^*, p_n)] = \sum_{i_1:i_M=0}^1 g_{i_1:i_M}(\mathbf{y} | \mathbf{A}_n, p_n) P_n \left(\bigcap_{m=1}^M (A_n^{(m)})^{i_m} \right). \quad (5)$$

The difference between the entropy at time n and the predicted entropy at time $n + 1$ is the mutual information ($I(X; Y | A, p)$ denotes the conditional mutual information of random variables X and Y given random variables A and p):

$$H(p_n) - \mathbb{E}[H(p_{n+1}) | \mathbf{A}_n, p_n] = I(X^*; \mathbf{Y}_{n+1} | \mathbf{A}_n, p_n) = H(\mathbf{Y}_{n+1} | \mathbf{A}_n, p_n) - \mathbb{E}[H(\mathbf{Y}_{n+1}) | X^*, \mathbf{A}_n, p_n],$$

which is equal to the argument of the “sup” on the right hand side of (3). Here we used (5) and (4). Using dynamic programming similar to Thm. 2 in [1], it follows that optimal queries satisfy (3). \square

Thm. 1 generalizes the bisection policy [1, 2] to multiple players. The fusion rule is a posterior update and by Bayes rule:

$$p_{n+1}(x) \propto P(\mathbf{Y}_{n+1} = \mathbf{y}_{n+1} | \mathbf{A}_n, X^* = x, p_n) \times p_n(x)$$

where $\mathbf{y}_{n+1} \in \{0, 1\}^M$ are the observations at time n .

2.2. Greedy Sequential Query Design

As an alternative, we consider the following greedy sequential coordinate-by-coordinate design: ask an optimal query to the first player, then update posterior density and ask an optimal query to the second player, and so on (see Fig. 2). In [1], the optimal query of a single player was given as a bisection rule. We show that this greedy sequential scheme achieves the same expected entropy loss as the optimal joint design of Thm. 1.

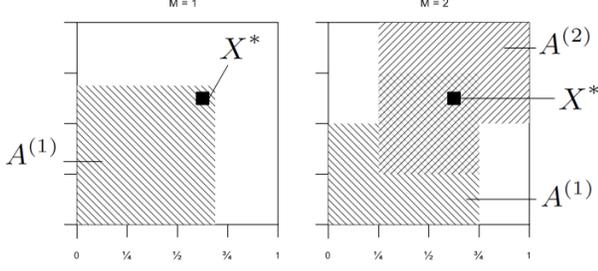


Fig. 3. Jointly optimal queries under uniform prior for two dimensional target search. The target X^* is indicated by a black square. The one-player bisection rule (left) satisfies the optimality condition (7) with optimal query $A^{(1)} = [0, \frac{1}{\sqrt{2}}] \times [0, \frac{1}{\sqrt{2}}]$. The two-player bisection rule (right) satisfies (7) with optimal queries $A^{(1)} = [0, \frac{3}{4}] \times [0, \frac{1}{2}] \cup [\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{2}, \frac{3}{4}]$, $A^{(2)} = [\frac{1}{4}, 1] \times [\frac{1}{2}, 1] \cup [\frac{1}{4}, \frac{3}{4}] \times [\frac{1}{4}, \frac{1}{2}]$. We note that using the policy on the left, if player 1 responds that $X^* \in [0, \frac{1}{\sqrt{2}}] \times [0, \frac{1}{\sqrt{2}}]$, with high probability, then the posterior will concentrate on that region. When using the policy on the right, if player 1 and 2 respond that $X^* \in A^{(1)} \cap A^{(2)}$ with high probability, then the posterior will concentrate more on the intersection of the queries, thus better localizing the target as compared with the single player policy.

Theorem 2. (Separation) Under Assumptions 1 and 2:

1. The expected entropy loss under an optimal joint query design is the same as the greedy sequential query design. This loss is given by:

$$C = \sum_{m=1}^M C(\epsilon_m) = \sum_{m=1}^M (1 - h_b(\epsilon_m)) \quad (6)$$

where $h_b(\cdot)$ is the binary entropy function [4].

2. All jointly optimal control laws satisfy:

$$\int_{\mathcal{R}} p_n(x) dx = 2^{-M}, \forall R \in \tilde{\pi}(\mathbf{A}_n). \quad (7)$$

Proof. Let G_{seq} denote the expected entropy loss after querying M players sequentially. The bisection policy yields an expected entropy loss of $C(\epsilon_m) = 1 - h_b(\epsilon_m)$ ¹ after querying the m th player [1]. Thus, $G_{seq} = \sum_{m=1}^M C(\epsilon_m)$. Since the joint controller is the optimal controller, we have $G_{seq} \leq G^*$. To finish the proof, we show $G_{seq} \geq G^*$. From Thm. 1,

$$\begin{aligned} G^* &= \sup_{A^{(1)}, \dots, A^{(M)}} \left\{ H \left(\sum_{i_1:i_M=0}^1 g_{i_1:i_M} P_n \left(\prod_{m=1}^M (A_n^{(m)})^{i_m} \right) \right) \right. \\ &\quad \left. - \sum_{i_1:i_M=0}^1 H(g_{i_1:i_M}) P_n \left(\prod_{m=1}^M (A_n^{(m)})^{i_m} \right) \right\} \\ &\leq \sup_{\mathbf{p}} \{ H(\mathbf{p}^T \mathbf{g}) - \mathbf{p}^T H(\mathbf{g}) : \mathbf{p} \geq 0, \mathbf{1}^T \mathbf{p} = 1 \} \\ &= G_{seq} \end{aligned} \quad (8)$$

where the last equality follows by the symmetry of BSC. The supremum in the strictly concave problem (8) is achieved by the uniform distribution. \square

¹This is the channel capacity of the m th BSC [1, 4].

Thm. 2 shows that the optimal policy can be determined and implemented using the simpler greedy sequential query design. Note that, despite the fact that all players are conditionally independent, the joint policy does not decouple into separate single player optimal policies. This is analogous to the non-separability of the optimal vector-quantizer in source coding even for independent sources [5]. In addition, the optimal queries must be overlapping—i.e., $\bigcap_{m=1}^M A_n^{(m)} \neq \emptyset$, but not identical. Finally, we remark that the optimal query \mathbf{A}_n is not unique, so it is possible that there exists an even simpler control law than the sequential greedy policy.

3. LOWER BOUNDS ON MSE VIA ENTROPY LOSS

Thm. 2 yields the value of the 20 questions game in terms of expected entropy reduction, which is the sum of the “capacities”² of all the players. This value function is used next to provide a lower bound on the MSE of the sequential Bayesian estimator.

Theorem 3. Let Assumptions 1,2 hold. Assume $H(p_0)$ is finite. Then, the MSE of the joint or sequential query policies in Thm 1 and 2 satisfies:

$$\frac{K}{2\pi e} de^{-2nC/d} \leq \mathbb{E}[\|X^* - X_n\|_2^2] \quad (9)$$

where $K = e^{2H(p_0)}$, C is the entropy loss given in (6) and X_n is the posterior median.

Observe that the bound in (9) is uniform over all policies π . The bound is met with equality if the optimal policy is used, the estimation error is Gaussian with covariance K_n , the target estimate is taken as the conditional mean, $\mathbb{E}^\pi[H(p_n)] = \log((2\pi e)^d \det(\mathbb{E}^\pi[K_n]))$ and $\det(\mathbb{E}^\pi[K_n]) = (\frac{\mathbb{E}^\pi[\text{tr}(K_n)]}{d})^d$. We finally note the MSE bound behaves exponentially as a function of the number of queries n . The proof is given in [6].

4. UPPER BOUNDS ON MSE

The performance analysis of the bisection method is difficult primarily due to the continuous nature of the posterior [2]. A discretized version of the probabilistic bisection method was proposed in [7], using the Burnashev-Zingagirov (BZ) algorithm, which imposes a piecewise constant structure on the posterior. A description of the BZ algorithm and its convergence rate is given in [2] (also see App. A in [8]). For simplicity of discussion, we assume the target location is constrained to the unit interval $\mathcal{X} = [0, 1]$. A step size $\Delta > 0$ is defined such that $\Delta^{-1} \in \mathbb{N}$ and the posterior after j iterations is $p_j : \mathcal{X} \rightarrow \mathbb{R}$, given by

$$p_j(x) = \frac{1}{\Delta} \sum_{i=1}^{\Delta^{-1}} a_i(j) I(x \in I_i)$$

where $I_1 = [0, \Delta]$, $I_i = ((i-1)\Delta, i\Delta]$ for $i = 2, \dots, \Delta^{-1}$. The initial posterior is $a_i(0) = \Delta$. The posterior is characterized completely by the pseudo-posterior $\mathbf{a}(j) = [a_1(j), \dots, a_{\Delta^{-1}}(j)]$ which is updated at each iteration via Bayes rule [8].

Convergence rates were derived for the one-dimensional case in [2] for the bounded noise case (i.e., constant error probability) and

²The “capacity” of each player is the channel capacity of each BSC [4].

for the unbounded noise case (i.e., error probability depends on distance from target X^* and converges to $1/2$ as the estimate reaches the target) in [9]. A modified version of this algorithm that is proven to handle unbounded noise was shown in [9]. Thm. 4 derives upper bounds on MSE using ideas from [9]. The proof is given in [6].

Theorem 4. *Consider the sequential bisection algorithm for M players in one-dimension, where each bisection is implemented using the BZ algorithm. Then, we have:*

$$\begin{aligned} P(|X^* - \hat{X}_n| > \Delta) &\leq \left(\frac{1}{\Delta} - 1\right) \exp(-n\bar{C}) \\ \mathbb{E}[(X^* - \hat{X}_n)^2] &\leq (2^{-2/3} + 2^{1/3}) \exp\left(-\frac{2}{3}n\bar{C}\right) \end{aligned} \quad (10)$$

where $\bar{C} = \sum_{m=1}^M \bar{C}(\epsilon_m)$, $\bar{C}(\epsilon) = 1/2 - \sqrt{\epsilon(1-\epsilon)}$.

The combination of the lower bound (Thm. 3) and the upper bound (Thm. 4) imply that the MSE of the BZ algorithm goes to zero at an exponential rate with rate constant between $2C$ and $2/3\bar{C}$.

5. HUMAN-IN-THE-LOOP

In this section, we consider a particular 2-player case where player 1 (the machine) has a constant error probability $\epsilon_1 \in (0, 1/2)$ and player 2 (the human) has error probability depending on the target localization error after the most recent query:

$$P(Y_{n+1}^{(2)} = y^{(2)} | Z_n^{(2)} = 1 - y^{(2)}) = \frac{1}{2} - \min(\delta_0, \mu |X^* - X_n|^{\kappa-1}) \quad (11)$$

where $\kappa > 1$, $0 < \delta_0 < \mu < 1/2$ is a reliability parameter to parameterize the human³. This is a popular model used for human-based optimization [3] and has also appeared in the unbounded noise case [9] for binary classification. From the nature of the error probability (11), we expect that the answers provided by the human will be helpful in the beginning iterations but their value will go to zero as the number of iterations grows to infinity. This is because the human propensity for error becomes larger as the questions become more highly resolved.

Using a similar technique as in the proof of Thm. 4, and using the modified BZ algorithm [9], from Lemma 1 in [9], we have the following. For $\kappa \geq 2$ with $\alpha_1 = \frac{\sqrt{\epsilon_1}}{\sqrt{\epsilon_1 + \sqrt{1-\epsilon_1}}}$, $\alpha_2 = 0.09\mu(3\Delta/4)^{\kappa-1}$:

$$P(|X^* - \hat{X}_n| > \Delta) \leq \Delta^{-1} \exp\left(-n \left[\bar{C}(\epsilon_1) + \frac{\mu^2}{50} \left(\frac{3\Delta}{4}\right)^{2\kappa-2} \right]\right).$$

This leads to the MSE upper bound dependent on Δ :

$$\mathbb{E}[(X^* - \hat{X}_n)^2] \leq \Delta^2 + \Delta^{-1} \exp\left(-n \left[\bar{C}(\epsilon_1) + \frac{\mu^2}{50} \left(\frac{3\Delta}{4}\right)^{2\kappa-2} \right]\right). \quad (12)$$

With the choice $\Delta = 2^{-1/3} e^{-n\bar{C}(\epsilon_1)/3}$,

$$\begin{aligned} \mathbb{E}[(X^* - \hat{X}_n)^2] &\leq \exp\left(-\frac{2}{3}n\bar{C}(\epsilon_1)\right) \\ &\times \left[2^{-2/3} + 2^{1/3} \exp\left(-\frac{\mu^2}{50} \left(\frac{3 \cdot 2^{-1/3}}{4}\right)^{2\kappa-2} n e^{-n\bar{C}(\epsilon_1) \frac{2\kappa-2}{3}} \right) \right] \end{aligned} \quad (13)$$

³The parameter κ controls the “resolution” of the human. It becomes increasingly difficult for the human to decide between close hypotheses as κ goes to infinity.

which is no greater than the “player 1” MSE bound (compare (13) with (10)). Asymptotically as $n \rightarrow \infty$, the two bounds both converge to zero at the same rate.

We define the human gain ratio (HGR) as the ratio of MSE upper bounds associated with “player 1” alone and “player 1 + human”, respectively, given by

$$R_n(\kappa) = \frac{2^{-2/3} + 2^{1/3}}{2^{-2/3} + 2^{1/3} \exp\left(-\frac{\mu^2}{50} \left(\frac{3 \cdot 2^{-1/3}}{4}\right)^{2\kappa-2} n e^{-n\bar{C}(\epsilon_1) \frac{2\kappa-2}{3}}\right)} \quad (14)$$

The HGR is plotted in Fig. 4 as a function of κ . This analysis quantifies the value of including the human-in-the-loop for a sequential target localization task. We note that the larger ϵ_1 is, the larger is the HGR. Also, as κ decreases to 1, the ratio increases, meaning that the value of including the human in the loop increases.

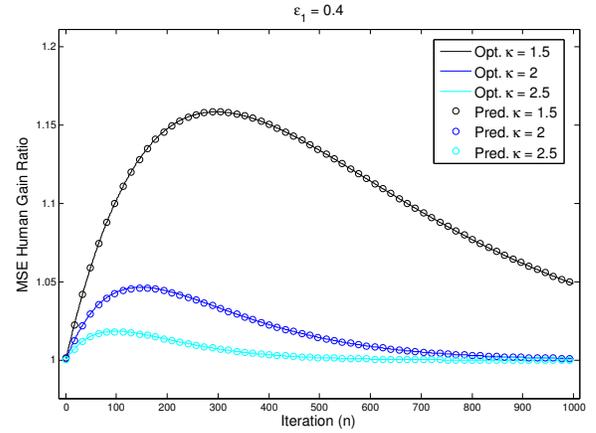


Fig. 4. Human gain ratio (see Eq. (14)) as a function of κ . The human provides the largest gain in the beginning few question iterations and the additional contribution of the human decreases as $n \rightarrow \infty$. The circles are the predicted curves according to (13), while the solid lines are the optimized versions of the bound (12) as a function of Δ for each n . The predictions match well the optimized bounds predicted by the theory in Section 5.

Monte Carlo simulation experiments are given in [6], where it is observed that employing a human in the loop reduces the MSE relative to only having player 1. It is also noted that the gap between the MSE curves associated with “player 1” and “player 1 + human” initially increases to a maximum and then diminishes to zero. This could be used to motivate a stopping rule for including the human when the cost of using the human is increasing over time or human fatigue prevents repeated querying; a worthwhile subject of future work.

6. CONCLUSION

We studied the problem of collaborative 20 questions with noise for the multiplayer case. We derived a separation theorem that shows the jointly optimal design is equivalent to a greedy sequential design that can be more easily implemented. Using this framework, we obtained bounds for the performance of human-in-the-loop target localization systems. Future work includes integration of a noisy continuous valued sensor and controlling the human query rate.

7. REFERENCES

- [1] B. Jedynek, P. I. Frazier, and R. Sznitman, “Twenty questions with noise: Bayes optimal policies for entropy loss,” *Journal of Applied Probability*, vol. 49, pp. 114–136, 2012.
- [2] R. Castro and R. Nowak, “Active learning and sampling,” in *Foundations and Applications of Sensor Management*. Springer, 2007.
- [3] K. G. Jamieson, R. D. Nowak, and B. Recht, “Query complexity of derivative-free optimization,” *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- [4] T. D. Cover and J. A. Thomas, *Elements of Information Theory*, Wiley, 2006.
- [5] A. Gersho and R. M. Gray, *Vector quantization and Signal Compression*, Kluwer Academic Press/Springer, 1992.
- [6] T. Tsiligkaridis, B. Sadler, and A. O. Hero III, “A collaborative 20 questions model for human-machine interaction,” Tech. Rep., University of Michigan - Ann Arbor, 2012.
- [7] M. V. Burnashev and K. Sh. Zigangirov, “An interval estimation problem for controlled observations,” *Problems in Information Transmission*, vol. 10, pp. 223231, 1974.
- [8] R. Castro, *Active Learning and Adaptive Sampling for Non-parametric Inference*, Ph.D. thesis, Rice University, August 2007.
- [9] R. Castro and R. D. Nowak, “Upper and lower bounds for active learning,” in *44th Annual Allerton Conference on Communication, Control and Computing*, 2006.