

On Analysis of Active Querying for Recursive State Estimation

Aziz Koçanaoğulları¹, Murat Akçakaya², and Deniz Erdoğan³

Abstract—In stochastic linear/nonlinear active dynamic systems, states are estimated with the evidence through recursive measurements in response to queries of the system about the state to be estimated. Therefore, query selection is essential for such systems to improve state estimation accuracy and time. Query selection is conventionally achieved by minimization of the evidence variance or optimization of various information theoretic objectives. It was shown that optimization of mutual information-based objectives and variance-based objectives arrive at the same solution. However, existing approaches optimize approximations to the intended objectives rather than solving the exact optimization problems. To overcome these shortcomings, we propose an active querying procedure using mutual information maximization in recursive state estimation. First we show that mutual information generalizes variance based query selection methods and show the equivalence between objectives if the evidence likelihoods have unimodal distributions. We then solve the exact optimization problem for query selection and propose a query (measurement) selection algorithm. We specifically formulate the mutual information maximization for query selection as a combinatorial optimization problem and show that the objective is submodular, therefore can be solved efficiently with guaranteed convergence bounds through a greedy approach. Additionally, we analyze the performance of the query selection algorithm by testing it through a brain computer interface (BCI) typing system.

Index Terms—Active querying, mutual information, query selection, recursive state estimation.

I. INTRODUCTION

IN STOCHASTIC linear/nonlinear dynamic systems, a state vector describes one instance over a realization of the path of the stochastic propagation of a target identity that follows the specified dynamic system. State can be determined (estimated)

from relevant evidence collected by the dynamic system. For example, evidence can be obtained through the stimulation of the environment where the dynamic system is defined. In optical imaging and radar problems, propagating field from the active sensors is the query into the environment and scattered field is the evidence [1]. Noninvasive brain computer interfaces that rely on external stimuli to detect user intent is another field where state estimation (user intent) over dynamic systems (e.g., brain's time-varying response to external stimuli) with active queries for evidence extraction (external stimuli presentation to obtain brain response) is commonly used for inference of the time-varying user intent [2], [3]. To estimate state over the stochastic dynamic models, maximum a posteriori (MAP) estimation is generally used [4]. In intermediate steps of the state estimation, a common approach is to update the posterior distribution using the collected evidence until a certain confidence interval is achieved [5]. Achieving high confidence usually requires collecting more evidence; hence, usually decreases the rate of convergence to the true state. Therefore, we argue that informative query selection (selecting subset of sensors or stimuli to query the most informative evidence) in such processes is crucial to increase the confidence while keeping a faster convergence rate.

Seeking the most informative/useful evidence/measurement is sometimes referred as query (sensor/data/stimuli) selection in signal processing applications, such as computer-vision [6], [7], radar [1], robot-control [8], experiment design [9], and image retrieval [10], [11]. In recursive updates for MAP estimation, in one approach, the query is selected to maximize the expected posterior of a particular update [12]. This approach is then improved when Fisher information-based objectives are used for query selection [13], [14]. Such approaches mainly exploit measurement variance, assuming the measurement noise (which is usually assumed to be colored) should be minimized. Therefore, through the minimization of a function of noise variance, at each posterior update the ambiguity in the state estimation is expected to decrease. Additionally, information theoretic objectives are also used for query selection. For example, entropy is used for query selection for sensor allocation in a truss structure [15]. Entropy measures the amount of uncertainty in a stochastic process; thus minimizing entropy seeks to reduce expected uncertainty about the parameter of interest. Entropy can be used for sequential query selection, for instance as illustrated with the 20-questions approach [16]–[19]. Analogously, mutual information is used for query selection in dynamic systems [20]–[22]. Mutual information objective measures the dependence level between stochastic processes [23]. With maximum mutual information (MMI), which is commonly preferred over entropy minimization, a query that maximizes the information gain regarding the variable of interest is selected.

Manuscript received February 19, 2018; revised March 27, 2018; accepted March 27, 2018. Date of publication April 5, 2018; date of current version April 23, 2018. The work of A. Koçanaoğulları and D. Erdoğan was supported in part by National Science Foundation under Grant IIS-1149570, Grant CNS-1544895, and Grant IIS-1715858; in part by Department of Health and Human Services under Grant 90RE5017-02-01; and in part by National Institutes of Health under Grant R01DC009834. The work of M. Akçakaya was supported in part by NSF IIS-1717654, and in part by the Air Force Office of Scientific Research, the Dynamic Data-Driven Applications Systems Program, under Grant FA9550-16-1-0386. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Samuel Davey. (Corresponding author: Aziz Koçanaoğulları.)

A. Koçanaoğulları and D. Erdoğan are with the Northeastern University, Boston, MA 02115 USA (e-mail: akocanaoğullari@ece.neu.edu; erdogmus@ece.neu.edu).

M. Akçakaya is with the University of Pittsburgh, Pittsburgh, PA 15260, Pennsylvania (e-mail: akcakaya@pitt.edu).

Color versions of one or more of the figures in this letter are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/LSP.2018.2823271

In this letter, we focus on using an information theoretic approach based on mutual information for active query selection from a countable set of queries in recursive state estimation. First, we show that information theoretic approaches generalize the variance based query selection methods. Second, we extend the current literature on information theoretic query selection in the following ways:

- 1) we solve the exact optimization problem without defining a lower/upper bound on the optimization objective to achieve a tractable solution;
- 2) we analytically obtain the close to global solution of the exact problem;
- 3) we introduce a computationally efficient combinatorial optimization algorithm that achieves the global solution. We demonstrate the proposed approach at work in non-invasive brain computer interface design, specifically addressing the issue of optimal selection of stimuli.

II. PROBLEM FORMULATION

In this section, we formulate query selection based on MMI for recursive state estimation. We denote the state, the evidence, and the query for the i th recursion as σ , ε^i , Π^i , respectively. In order to represent a sequence of variables such as $\varepsilon^{0:c}$ from index 0 up to c , we use the bold notation ε^c . Assume that we are just before the i th recursion; that is, we have observed evidence ε^{i-1} for up to and including $(i-1)$ th query Π^{i-1} . In order to obtain the i th query Π^i , we will maximize the mutual information between prospective evidence and state given all queries and past evidence: $I(\varepsilon^i, \sigma | \Pi^i, \Pi^{i-1}, \varepsilon^{i-1})$. In all of the following, past queries Π^{i-1} and associated evidence ε^{i-1} are given together, therefore to simplify the notation, we will omit the past queries in the condition list, as in $I(\varepsilon^i, \sigma | \Pi^i, \varepsilon^{i-1}) = H(\sigma | \Pi^i, \varepsilon^{i-1}) - H(\sigma | \Pi^i, \varepsilon^i)$. Observe that, in the absence of evidence ε^i , state σ is independent of the prospective query Π^i . Consequently, the first entropy term is constant with respect to Π^i and the optimization problem

$$\arg \max_{\Pi^i} I(\varepsilon^i, \sigma | \Pi^i, \varepsilon^{i-1}) = \arg \max_{\Pi^i} -H(\sigma | \Pi^i, \varepsilon^i). \quad (1)$$

Since ε^{i-1} has already been observed and Π^i is a particular subset of all potential queries, (1) can also be written as

$$\hat{\Pi}^i = \arg \max_{\Pi^i} E_{p(\sigma | \varepsilon^{i-1})} (E_{p(\varepsilon^i | \sigma, \Pi^i)} (\log (p(\varepsilon^i | \sigma, \Pi^i) p(\sigma | \varepsilon^{i-1}, \Pi^i) / p(\varepsilon^i | \Pi^i))))). \quad (2)$$

We can relate (2) with evidence variance minimization techniques as a natural lower bound

$$\begin{aligned} & \log E_{p(\sigma | \varepsilon^{i-1})} [E_{p(\varepsilon^i | \sigma, \Pi^i)} (p(\sigma | \varepsilon^i, \Pi^i))] \\ & \geq E_{p(\sigma | \varepsilon^{i-1})} [E_{p(\varepsilon^i | \sigma, \Pi^i)} (\log p(\sigma | \varepsilon^i, \Pi^i))]. \end{aligned} \quad (3)$$

This can be trivially observed using Jensen's inequality. Therefore, the simplified MMI objective is a lower bound to the expected posterior maximization (EPM) method [12].

We have two main assumptions that enable the results in Section III: 1) the query set is countable; and 2) the response to the query, given a particular state, generates a noisy evidence [17] and evidence likelihood conditioned on state-query tuples have concentrated-unimodal distribution. In the next section, we derive a tractable solution to objective (2) by showing the equivalence between EPM and MMI using these assumptions.

III. SOLUTION

It can be observed that the MMI objective in (2) requires evidence ε^i , which is not observed. Therefore, it requires Monte Carlo simulations based on a generative evidence model. In this letter, we assume that such a method to evaluate/estimate the necessary likelihoods and expectations is available with satisfactory accuracy. Due to the fact that the state is independent of the upcoming query, in (2), $p(\sigma | \varepsilon^{i-1}, \Pi^i) = p(\sigma | \varepsilon^{i-1})$. With these two simplifications, the problem reduces to

$$\arg \max_{\Pi^i} E_{p(\sigma | \varepsilon^{i-1})} \left(E_{p(\varepsilon^i | \sigma, \Pi^i)} \log \frac{p(\varepsilon^i | \sigma, \Pi^i)}{E_{p(v)} p(\varepsilon^i | v, \Pi^i)} \right) \quad (4)$$

where $p(v) \sim p(\sigma | \varepsilon^{i-1})$. With v we can simplify the normalization term using the facts noted before. Furthermore, assuming that evidence ε^i acquired after query Π^i consists of statistically independent measurements for each element in the query subset, $p(\varepsilon^i | \sigma, \Pi^i) = \prod_k p(\varepsilon_k^i | \sigma, \Pi_k^i)$, rewrite (4)

$$\begin{aligned} \hat{\Pi}^i = \arg \max_{\Pi^i} E_{p(\sigma | \varepsilon^{i-1})} & \left(E_{p(\varepsilon^i | \sigma, \Pi^i)} \left(\sum_k \log p(\varepsilon_k^i | \sigma, \Pi_k^i) \right. \right. \\ & \left. \left. - \log \left(\sum_v \prod_k p(\varepsilon_k^i | v, \Pi_k^i) p(v) \right) \right) \right). \end{aligned} \quad (5)$$

If the number of discrete values that a state can take is finite, we define a class assignment for state and there exists a specific question for each class, where the evidence obtained using this query is highly likely. We design a sequence of questions to identify/infer the state. Therefore, we can relate the correspondence of queries to the states with a class indicating notation. We can denote the condition on the query and the state using the class notation as, $p(\varepsilon_k^i | \sigma, \Pi_k^i) = p(\varepsilon_k^i | c_k^i(\sigma))$. With this notation, we can rewrite (4) as the following equation:

$$\begin{aligned} \hat{\Pi}^i = \arg \max_{\Pi^i} E_{p(\sigma | \varepsilon^{i-1})} & \left(E_{p(\varepsilon^i | \sigma, \Pi^i)} \left(\sum_k \log p(\varepsilon_k^i | c_k^i(\sigma)) \right. \right. \\ & \left. \left. - \log \left(\sum_v \prod_k p(\varepsilon_k^i | c_k^i(v)) p(v) \right) \right) \right). \end{aligned} \quad (6)$$

Probability distributions of evidence, given a state-query dependent class label, can be different for each state-query combination. Let $z(\sigma, \Pi) = \inf_{\sigma' \neq \sigma} E_{p(\varepsilon | \sigma, \Pi)} (p(\varepsilon | \sigma, \Pi) / p(\varepsilon | \sigma', \Pi))$. Queries for which $z(\sigma, \Pi) \gg 1$ we designate $c_k^i(\sigma) = c_\sigma(\sigma)$, and for which $z(\sigma, \Pi) \ll 1$ we designate $c_k^i(\sigma) = c_\sigma(\sigma)$. After making this substitution, adding/subtracting $\sum_k \log p(\varepsilon_k^i | c_\sigma(\sigma))$, (6) becomes

$$\begin{aligned} \hat{\Pi}^i = E_{p(\sigma | \varepsilon^{i-1})} & \left(E_{p(\varepsilon^i | \sigma, \Pi^i)} \left(\sum_k \log \frac{p(\varepsilon_k^i | c_\sigma(\sigma))}{p(\varepsilon_k^i | c_\sigma(\sigma))} \right. \right. \\ & \left. \left. - \log \left(\sum_v \prod_k \frac{p(\varepsilon_k^i | c_k^i(v))}{p(\varepsilon_k^i | c_\sigma(\sigma))} p(v) \right) \right) \right). \end{aligned} \quad (7)$$

With $c_k^i \in \{c_\sigma, c_\sigma\}$, first summation can be separated into $\bar{k} := \{k | c_k^i = c_\sigma\}$ and $\{k | c_k^i = c_\sigma\}$. Additionally, we assume resulting evidence of $c_\sigma(\sigma)$ class has close to uniform likelihood for mismatched classes $c_\sigma(v) \forall v$. Thus, $\varepsilon \sim p(\varepsilon | c_\sigma(\sigma))$ $v \neq \sigma$, $p(\varepsilon | c_\sigma(v)) / p(\varepsilon | c_\sigma(\sigma)) \approx 1$. Thus, in the second term we only require $c_k^i = c_v$, and we define the set of indices

$\tilde{k}_v := \{k | c_k^i = c_v\}$. With simple algebraic manipulations (7) can be rewritten as

$$\hat{\Pi}^i = \arg \max_{\Pi^i} E_{p(\sigma|\epsilon^{i-1})} \left(E_{p(\epsilon^i|\sigma, \Pi^i)} \left(\sum_{\tilde{k}} \log \frac{p(\epsilon_k^i | c_\sigma(\sigma))}{p(\epsilon_k^i | c_\sigma(\sigma))} \right) - \log \left(\sum_v \prod_{\tilde{k}_v} \frac{p(\epsilon_{k_v}^i | c_v(v))}{p(\epsilon_{k_v}^i | c_\sigma(\sigma))} p(v) \right) \right). \quad (8)$$

Note that the exhaustive search for the optimal solution is exponentially growing with respect to query set size $|\Pi^i|$. In order to propose a tractable solution we introduce an assumption for the objective in (8) that allows us to show equivalence between EPM and MMI. Specifically, for a particular class c_k^i the corresponding evidence $\epsilon_k^i | c_k^i$ has a unimodal distribution with a variance that is negligible compared to its mean-value-squared. Furthermore, with that assumption, we introduce the following notation to indicate scalar-valued conditional expected likelihood ratios of the evidence indicating model performance:

$$E_{\epsilon \sim p(\epsilon | c_\sigma(\sigma))} \left(\frac{p(\epsilon | c_\sigma(\sigma))}{p(\epsilon | c_\sigma(\sigma))} \right) = s_\sigma^1$$

$$E_{\epsilon \sim p(\epsilon | c_\sigma(v))} \left(\frac{p(\epsilon | c_\sigma(v))}{p(\epsilon | c_\sigma(\sigma))} \right) = s_{\sigma,v}^0. \quad (9)$$

By design $\forall \sigma, v \neq \sigma \ s_\sigma^1 \geq s_{\sigma,v}^0$ and $\sup_v s_{\sigma,v}^0 = 1$, and this is satisfied when $v = \sigma$ (or model is insufficient to discriminate); note that in the event this is not satisfied, by swapping the class-conditional distributions between the two labels, this condition can be satisfied. Equality arises when the evidence is not class-specific on average. Additionally, we note that $s_\sigma^1 \rightarrow \infty, s_{\sigma,v}^0 \rightarrow 0$ with improved class separability. Notice that these scores are the expected values of the evidence likelihood ratios that appear in (8). Let $k_\sigma(\Pi^i) = \sum_k \delta_{c_k^i, c_\sigma}$ denote the number of times a particular state is queried in a query set. Observe that in (8) since the likelihood ratio in the first term is only calculated for the expected state σ , this expectation yields only s_σ^1 . However, in the normalization term, v takes every possible state value including σ . If v is equal to sigma, we have s_σ^1 in the normalization. In the cases where v takes different state values than σ , we observe $s_{\sigma,v}^0$. Thus, the normalization term has both score values. The following equation is trivial to obtain from (8):

$$\begin{aligned} \hat{\Pi}^i &= \arg \max_{\Pi^i} E_{p(\sigma|\epsilon^{i-1})} \left(k_\sigma \log s_\sigma^1 \right. \\ &\quad \left. - \log(p(\sigma|\epsilon^{i-1})(s_\sigma^1)^{k_\sigma} + \sum_{v \neq \sigma} (s_{\sigma,v}^0)^{k_v} p(v)) \right) \\ &= \arg \min_{\Pi^i} E_{p(\sigma|\epsilon^{i-1})} \left(\log(p(\sigma|\epsilon^{i-1})) \right. \\ &\quad \left. + \sum_{v \neq \sigma} \frac{(s_{\sigma,v}^0)^{k_v}}{(s_\sigma^1)^{k_\sigma}} p(v) \right). \end{aligned} \quad (10)$$

Here, we reduce the objective to a form that is described by scores only, which can be solved combinatorially. In the next section, we introduce an algorithm to solve the optimization problem presented in (10).

Algorithm 1: Mutual Information Based Query Selection.

```

1:  $\Pi^i \leftarrow \{\}, N \in \mathbb{Z}^+$ 
2:  $x \leftarrow \{x_c | p(\sigma_c | \epsilon^{i-1}) \forall c\}$ 
3: procedure GREEDY SUBSET SELECTION ( $\Pi^i, \sigma, x$ )
4:    $p \leftarrow 1$ 
5:   for  $i \in [1, 2, \dots, |\Pi^i|]$  do
6:      $\hat{c} = \arg \min_c x_c \log x_c + (p - x_c) \log(p - x_c)$ 
7:      $\Pi^i \leftarrow \Pi^i \cup \{\Sigma^{-1}(\sigma_{\hat{c}})\}$ 
8:      $p \leftarrow p - x_{\hat{c}}$ 
9:      $x \leftarrow x \setminus \{x_{\hat{c}}\}$ 
10:  end for
11: return:  $\Pi^i$ 
12: end procedure

```

IV. METHOD

Here, we propose a solution to (10) without discarding the normalization factor, which is also a function of the query. We will show that the optimization in (10) can be represented as a submodular subset selection problem; therefore, we will develop a greedy algorithm that can find a solution close to global optimum solution [24].

Based on our assumptions on s_σ^1 and $s_{\sigma,v}^0$, we have the following observations.

Observation 1: Previously we stated that $s_\sigma^1 \rightarrow \infty$ and $s_{\sigma,v}^0 \rightarrow 0$. Therefore, $\forall k_v, k_\sigma \in \mathbb{Z}^+, (s_{\sigma,v}^0)^{k_v} / (s_\sigma^1)^{k_\sigma} \rightarrow 0$. Therefore, the query set Π^i should contain unique elements.

Observation 2: Following Observation 1, we can reduce the objective in (10) to a form without scores. We require a notation to represent the states addressed by a particular query Π^i , which we denote as $\Sigma(\Pi^i) = \{\sigma | c_k^i = c_\sigma \ \forall k\}$, and we define the objective

$$\begin{aligned} O(\sigma, \Pi^i) &= \sum_{\sigma \in \Sigma} p(\sigma | \epsilon^{i-1}) \log(p(\sigma | \epsilon^{i-1})) \\ &\quad + \left(\sum_{\sigma \notin \Sigma} p(\sigma | \epsilon^{i-1}) \right) \log \left(\sum_{\sigma \notin \Sigma} p(\sigma | \epsilon^{i-1}) \right). \end{aligned} \quad (11)$$

Observation 3: Observe that the objective in (11) monotonically decreases w.r.t. Π^i . Pick two random sets $\Pi_2 = \Pi_1 \cup \{\pi\}$ where $\pi \notin \Pi_1$ and observe that $\{\pi\} \subseteq \{x | x \notin \Pi_1\}$ and $\{x | x \notin \Pi_2\} \subseteq \{x | x \notin \Pi_1\}$. Define $\Sigma_1 = \Sigma(\Pi_1)$, $\Sigma_2 = \Sigma(\Pi_2)$ and observe that $\Sigma_1 \subseteq \Sigma_2$

$$\begin{aligned} O(\sigma, \Pi_2) - O(\sigma, \Pi_1) &= p(a | \epsilon^{i-1}) \left(\log \frac{p(a | \epsilon^{i-1})}{\sum_{\sigma \notin \Sigma_1} p(\sigma | \epsilon^{i-1})} \right) \\ &\quad + \left(\sum_{\sigma \notin \Sigma_2} p(\sigma | \epsilon^{i-1}) \right) \left(\log \frac{\sum_{\sigma \notin \Sigma_2} p(\sigma | \epsilon^{i-1})}{\sum_{\sigma \notin \Sigma_1} p(\sigma | \epsilon^{i-1})} \right) \leq 0. \end{aligned} \quad (12)$$

Using these observations, we state that the optimization presented in (10) is a submodular subset selection problem. It was shown that such problems can be efficiently solved using greedy algorithms [24]. Accordingly, we can write a greedy update to select query elements based on the objective presented in (11). Assume that at the k th step, we have the query set $\Pi_{0:k}^i$, and we would like to select a specific item π as the $(k+1)$ th query with $\Sigma(\pi) = a$, then we use the following greedy update:

$$\begin{aligned} \hat{\pi}^{k+1} = & \arg \min_{\pi \notin \Pi_{0:k}^i} p(a|\epsilon^{i-1}) \log(p(a|\epsilon^{i-1})) \\ & + \left(\sum_{\sigma \notin \Sigma(\Pi_k^i \cup \pi)} p(\sigma|\epsilon^{i-1}) \log \left(\sum_{\sigma \notin \Sigma(\Pi_k^i \cup \pi)} p(\sigma|\epsilon^{i-1}) \right) \right). \end{aligned} \quad (13)$$

We denote the remaining probability mass at k th query selection by $\mathcal{P}_k = 1 - \sum_{\sigma \in \Sigma(\Pi_k^i)} p(\sigma|\epsilon^{i-1})$. Inserting this equation into (13) and with algebraic manipulations, we have

$$\begin{aligned} \hat{\pi}^{k+1} = & \arg \min_{\pi \notin \Pi_k^i} p(a|\epsilon^{i-1}) \log(p(a|\epsilon^{i-1})) \\ & + (\mathcal{P}_k - p(a|\epsilon^{i-1})) \log(\mathcal{P}_k - p(a|\epsilon^{i-1})) \\ & \text{where } \mathcal{P}_k \in [0, 1]. \end{aligned} \quad (14)$$

Using the relation in (14), we propose Algorithm 1.

Proposition 1: Algorithm 1 can be realized with “N-best selection”.

Proof: $\forall \mathcal{P}_s \in [0, 1] \exists \Pi_s^i, p(\sigma|\epsilon^i) \forall \sigma$. Pick $\pi, \tilde{\pi}, \bar{\pi}$ with corresponding states $a, \tilde{a}, \bar{a}$.

- 1) $p(a|\epsilon^i) \geq \mathcal{P}_s/2, \exists p(\tilde{a}|\epsilon^i) = 1 - p(a|\epsilon^i)$ where $\tilde{\pi} \neq \pi$ and $p(\bar{a}|\epsilon^i) = 0 \forall \bar{\pi} \neq \tilde{\pi}$. $O(\sigma, \pi) = O(\sigma, \tilde{\pi})$, and therefore, we can pick π .
- 2) $p(a|\epsilon^i) \geq \mathcal{P}_s/2, p(\tilde{a}|\epsilon^i) \neq 0 \forall \tilde{\pi} \neq \pi$. It is trivial that the objective is minimized with π .
- 3) Where $p(\tilde{a}|\epsilon^i) < \mathcal{P}_s/2 \forall \tilde{a}$. As the objective is strictly monotonically decreasing in $[0, \mathcal{P}_s]$, pick π with $\Sigma(\pi)$ satisfying the highest likelihood which minimizes the objective.

Observe that in any case $p(a|\epsilon^i) \geq p(\tilde{a}|\epsilon^i) \forall \tilde{a} \neq \pi$. Other cases can be created by switching π without loss of generality. Since selection of $\mathcal{P}_s$ is arbitrary, one can conclude that Algorithm 1 can be realized by using “N-best selection algorithm”. ■

V. EXPERIMENTS

We use electroencephalography (EEG) data recorded when human users performed typing tasks using RSVP Keyboard [3], which is a noninvasive EEG-based brain computer interface for typing. Specifically, calibration data of 12 healthy participants (collected with Northeastern University IRB-130107). During calibration, participants are presented with 100 sequences of symbols. A sequence contains randomly ordered ten symbols, one of which is the target symbol. EEG is acquired from 16 channels using the International 10–20 configuration (Fp1, Fp2, F3, F4, Fz, Fc1, Fc2, Cz, P1, P2, C1, C2, Cp3, Cp4, P5, P6). These signals are used to learn class (target and nontarget) conditional EEG feature distributions. A copy-letter task where simulated users (modeled with the said conditional EEG feature distributions) copy particular letters, where the state (intent of the user) is to be estimated through EEG measurements induced by stimulus symbol sequences (queries). During simulations when a target/nontarget letter is presented to the user, EEG features from appropriate class conditional distributions are sampled. These samples are used to estimate the state. For the details of this simulation framework, the reader is referred to Orhan’s work [25].

We categorize the users based on their calibration performance (based on area under the receiver operating characteristics curve (AUC)): 6 users with $\text{AUC} \in [75, 85]$ and 6 users with $\text{AUC} \in (85, 95]$. We compare MMI with min/max querying [26], max only querying (suboptimal variance minimization

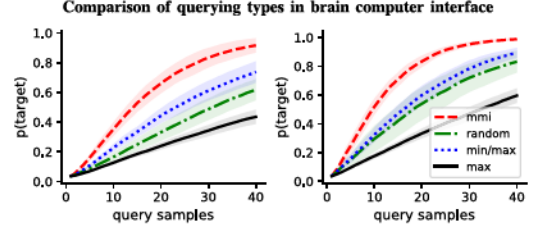


Fig. 1. $\text{AUC} \in [75, 85]$ (left) and $\in (85, 95]$ (right).

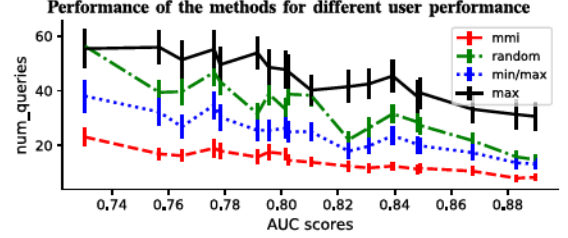


Fig. 2. Using the same 12 users, we are using different sessions to simulate the task.

for query scheduling) and random querying (baseline). We perform 1000 Monte Carlo simulations using the same prior for each method. In Fig. 1, we observe that the proposed querying method converge to the confidence interval faster than the other methods.

We also stated in Section III, proposed querying method also relies on the confidence of the evidence model. Observe that AUC, as a performance metric in the framework, can be related to the scores in (9). As AUC increases scores diverge from one. Therefore, estimation requires less number of queries for evidences that are more descriptive. We visualize this concept in Fig. 2. As performance of the users increases, number of query samples required for the estimation reduces.

VI. CONCLUSION

In this letter, we provided an efficient algorithm with guaranteed convergence properties of the optimization of MMI for query selection during recursive Bayesian state estimation. We analyzed the performance of the algorithm through BCI typing interface. Our results showed that the proposed method outperforms other commonly used query techniques for BCI typing. Specifically, as illustrated, the method always chooses N-best queries if N is the maximum number of available query. This technique is useful if we can always trust the currently available posterior distribution defined over the state. However, if the statistical properties of the evidence, which enables the estimation of the posterior distribution changes (e.g., nonstationarity in the EEG measurements) in a way to decrease our belief in the estimation accuracy, then a method purely relying on choosing the N -best based on the current posterior may not always be the optimum query selection. Therefore, in our future work, we will extend our approach to include also exploration beyond our current belief on the estimated posterior, which is fully exploited. We aim to formulate the query selection as a multi-objective optimization, which will balance between exploration and exploitation.

ACKNOWLEDGMENT

The authors would like to thank P. G. Navarro for the efforts in data collection.

REFERENCES

- [1] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, 2006.
- [2] G. R. Muller-Putz and G. Pfurtscheller, "Control of an electrical prosthesis with an SSVEP-based BCI," *IEEE Trans. Biomed. Eng.*, vol. 55, no. 1, pp. 361–364, Jan. 2008.
- [3] U. Orhan, K. E. Hild, D. Erdogmus, B. Roark, B. Oken, and M. Fried-Oken, "RSVP keyboard: An EEG based typing interface," in *Proc. 2012 IEEE Int. Conf. Acoust., Speech Sig. Process.*, Kyoto, Japan, 2012, pp. 645–648.
- [4] J.-L. Gauvain and C.-H. Lee, "Maximum a posteriori estimation for multivariate Gaussian mixture observations of Markov chains," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 2, pp. 291–298, Apr. 1994.
- [5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [6] J. Denzler and C. M. Brown, "Information theoretic sensor data selection for active object recognition and state estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 145–157, Feb. 2002.
- [7] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens, "Multimodality image registration by maximization of mutual information," *IEEE Trans. Med. Imag.*, vol. 16, no. 2, pp. 187–198, Apr. 1997.
- [8] F. Meyer, O. Hlinka, H. Wymeersch, E. Riegler, and F. Hlawatsch, "Distributed localization and tracking of mobile networks including noncooperative objects," *IEEE Trans. Signal Inf. Process. Netw.*, vol. 2, no. 1, pp. 57–71, Mar. 2016.
- [9] K. Chaloner and I. Verdinelli, "Bayesian experimental design: A review," *Statistical Sci.*, vol. 10, pp. 273–304, 1995.
- [10] P. H. Gosselin and M. Cord, "Active learning methods for interactive image retrieval," *IEEE Trans. Image Process.*, vol. 17, no. 7, pp. 1200–1211, Jul. 2008.
- [11] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. 9th ACM Int. Conf. Multimedia*, 2001, pp. 107–118.
- [12] D. A. Cohn, Z. Ghahramani, and M. I. Jordan, "Active learning with statistical models," *J. Artif. Intell. Res.*, vol. 4, pp. 129–145, 1996.
- [13] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for nonlinear measurement models," *IEEE Transactions on Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.
- [14] S. Liu, S. P. Chepuri, M. Fardad, E. Maşazade, G. Leus, and P. K. Varshney, "Sensor selection for estimation with correlated measurement noise," *IEEE Trans. Signal Proc.*, vol. 64, no. 13, pp. 3509–3522, Jul. 2016.
- [15] C. Papadimitriou, J. L. Beck, and S.-K. Au, "Entropy-based optimal sensor location for structural model updating," *J. Vibration Control*, vol. 6, no. 5, pp. 781–800, 2000.
- [16] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "Collaborative 20 questions for target localization," *IEEE Trans. Inf. Theory*, vol. 60, no. 4, pp. 2233–2252, Apr. 2014.
- [17] B. Jedynak, P. I. Frazier, and R. Sznitman, "Twenty questions with noise: Bayes optimal policies for entropy loss," *J. Appl. Probability*, vol. 49, no. 1, pp. 114–136, 2012.
- [18] T. Tsiligkaridis, B. M. Sadler, and A. O. Hero, "On decentralized estimation with active queries," *IEEE Trans. Signal Process.*, vol. 63, no. 10, pp. 2610–2622, May 2015.
- [19] P. Rajan, W. Han, R. Sznitman, P. Frazier, and B. Jedynak, "Bayesian multiple target localization," *J. Mach. Learn. Res.*, vol. 37, pp. 1945–1953, 2015.
- [20] E. Erten, J. W. Fisher, and L. C. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *Proc. Inf. Process. Sensor Netw.*, 2003, vol. 2003, pp. 405–416.
- [21] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *Proc. 3rd Int. Symp. Inf. Process. Sensor Netw.*, 2004, pp. 36–45.
- [22] M. Higger *et al.*, "Recursive Bayesian coding for BCIs," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 6, pp. 704–714, Jun. 2017.
- [23] H. Peng, F. Long, and C. Ding, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 8, pp. 1226–1238, Aug. 2005.
- [24] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions-I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.
- [25] U. Orhan *et al.*, "Probabilistic simulation framework for EEG-based BCI design," *Brain-Comput. Interfaces*, vol. 3, no. 4, pp. 171–185, 2016.
- [26] P. K. Mallapragada, R. Jin, and A. K. Jain, "Active query selection for semi-supervised clustering," in *Proc. 19th Int. Conf. Pattern Recog.*, 2008, pp. 1–4.