Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

HHS Public Access
Author manuscript
IEEE Signal Process Lett. Author manuscript; available in PMC 2019 December 23.

Published in final edited form as:
IEEE Signal Process Lett. 2018 June ; 25(6): 743–747. doi:10.1109/LSP.2018.2823271.

On Analysis of Active Querying for Recursive State Estimation

Aziz Koçanaoğulları[*], **Murat Akçakay**[†], **Deniz Erdoğmuş**[*]

Aziz Koçanaoğulları: akocanaogullari@ece.neu.edu; Murat Akçakay: akcakaya@pitt.edu; Deniz Erdoğmuş: erdogmus@ece.neu.edu

[*]Northeastern University

[†]University of Pittsburgh

Abstract

In stochastic linear/non-linear 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 sub-modular, 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 typing system.

Keywords

recursive state estimation; query selection; active querying; mutual information

I. Introduction

In stochastic linear/non-linear 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

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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; and 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 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], 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], [17], [18], [19]. Analogously, mutual information is used for query selection in dynamic systems [20], [21], [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.

In this paper, 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: (i) we solve the exact optimization problem without defining a lower/upper bound on the optimization objective to achieve a tractable solution, (ii) we analytically obtain the close to global solution of the exact problem, (iii) 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,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

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 $\boldsymbol{\varepsilon}^c$. Assume that we are just before the i^{th} recursion; that is, we have observed evidence $\boldsymbol{\varepsilon}^{i-1}$ for up to and including $(i-1)^{\text{th}}$ 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: $I\left(\varepsilon^i, \sigma \mid \Pi^i, \Pi^{i-1}, \boldsymbol{\varepsilon}^{i-1}\right)$. In all of the following, past queries $\boldsymbol{\Pi}^{i-1}$ and associated evidence $\boldsymbol{\varepsilon}^{i-1}$ are given together, therefore to simplify the notation, we will omit the past queries in the condition list, as in $I\left(\varepsilon^i, \sigma \mid \Pi^i, \boldsymbol{\varepsilon}^{i-1}\right) = H(\sigma \mid \Pi^i, \boldsymbol{\varepsilon}^{i-1}) - H(\sigma \mid \Pi^i, \boldsymbol{\varepsilon}^i)$. Observe that, in the absence of evidence $\boldsymbol{\varepsilon}^i$, state $\boldsymbol{\sigma}$ is independent of the prospective query Π^i. Consequently, the first entropy term is constant with respect to Π^i, and the optimization problem:

$$\underset{\Pi^i}{\arg\max}\, I\left(\varepsilon^i, \sigma \mid \Pi^i, \boldsymbol{\varepsilon}^{i-1}\right) = \underset{\Pi^i}{\arg\max} - H(\sigma \mid \Pi^i, \boldsymbol{\varepsilon}^i) \tag{1}$$

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

$$\widehat{\Pi}^i = \underset{\Pi^i}{\arg\max}\, E_{p(\sigma \mid \boldsymbol{\varepsilon}^{i-1})}\left(E_{p\left(\varepsilon^i \mid \sigma, \Pi^i\right)}\left(\log\left(p\left(\varepsilon^i \mid \sigma, \Pi^i\right) p(\sigma \mid \boldsymbol{\varepsilon}^{i-1}, \Pi^i) / p\left(\varepsilon^i \mid \Pi^i\right)\right)\right)\right) \tag{2}$$

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

$$\log E_{p(\sigma \mid \boldsymbol{\varepsilon}^{i-1})}\left[E_{p\left(\varepsilon^i \mid \sigma, \Pi^i\right)}(p(\sigma \mid \boldsymbol{\varepsilon}^i, \Pi^i))\right] \geq E_{p(\sigma \mid \boldsymbol{\varepsilon}^{i-1})}\left[E_{p\left(\varepsilon^i \mid \sigma, \Pi^i\right)}(\log p(\sigma \mid \boldsymbol{\varepsilon}^i, \Pi^i))\right] \tag{3}$$

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: (i) the query set is countable; and (ii) 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 tracktable 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 $\boldsymbol{\varepsilon}^i$ which is not observed. Therefore, it requires Monte Carlo Simulations based on a generative evidence model. In this paper, 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), $p(\sigma \mid \boldsymbol{\varepsilon}^{i-1}, \Pi^i) = p(\sigma \mid \boldsymbol{\varepsilon}^{i-1})$. With these two simplifications, the problem reduces to:

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

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

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

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, $p\left(\varepsilon_k^i|\sigma,\Pi_k^i\right) = p\left(\varepsilon_k^i|c_k^i(\sigma)\right)$. With this notation we can rewrite (4) as the following equation,

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

Probability distributions of evidence given state-query dependent class label can be different for each state-query combination. Let $z(\sigma,\Pi) = \inf_{\sigma' \neq \sigma} E_{p(\varepsilon|\sigma,\Pi^i)}(p(\varepsilon|\sigma,\Pi)/p(\varepsilon|\sigma',\Pi))$. Queries for which $z(\sigma,\Pi) > > 1$ we designate $c_k^i(\sigma) = c_\sigma(\sigma)$, and for which $z(\sigma,\Pi) < < 1$ we designate $c_k^i(\sigma) = c_{\not\sigma}(\sigma)$. After making this substitution, adding/subtracting $\sum_k \log\ p\left(\varepsilon_k^i|c_{\not\sigma}\right)$ (6) becomes;

$$\widehat{\Pi}^i = E_{p(\sigma|\boldsymbol{\varepsilon}^{i-1})}\left(E_{p\left(\varepsilon^i|\sigma,\Pi^i\right)}\left(\sum_k \log\frac{p\left(\varepsilon_k^i|c_\sigma(\sigma)\right)}{p\left(\varepsilon_k^i|c_{\not\sigma}(\sigma)\right)} - \log\left(\sum_v\prod_k\frac{p\left(\varepsilon_k^i|c_k^i(v)\right)}{p\left(\varepsilon_k^i|c_{\not\sigma}(\sigma)\right)}p(v)\right)\right)\right) \tag{7}$$

With $c_k^i \in \{c_\sigma, c_{\not\sigma}\}$, first summation can be separated into $\bar{k}: = \{k|c_k^i = c_\sigma\}$ and $\{k|c_k^i = c_{\not\sigma}\}$. Additionally, we assume resulting evidence of $c_\sigma(\sigma)$ class has close to uniform likelihood for mismatched classes $c_{\not v}(v)\forall v$. Thus, $\varepsilon \sim p(\varepsilon|c_\sigma(\sigma))\ v \neq \sigma,\ p(\varepsilon|c_{\not v}(v))/p(\varepsilon|c_{\not\sigma}(\sigma)) \approx 1$ Thus, in the

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

second term we only require $c_k^i = c_\upsilon$, we define the set of indices $\tilde{k}_\upsilon := \left\{ k \mid c_k^i = c_\upsilon \right\}$. With simple algebraic manipulations (7) can be rewritten as,

$$\begin{aligned}\widehat{\Pi}^i = \arg\max_{\Pi^i} E_{p(\sigma \mid \boldsymbol{\varepsilon}^{i-1})}\Bigg(E_{p\left(\varepsilon^i \mid \sigma, \Pi^i\right)}\Bigg(\sum_k \log \frac{p\left(\varepsilon_k^i \mid c_\sigma(\sigma)\right)}{p\left(\varepsilon_k^i \mid c_{\not\sigma}(\sigma)\right)} \\ - \log\left(\sum_\upsilon \prod_{\tilde{k}_\upsilon} \frac{p\left(\varepsilon_{k_\upsilon}^i \mid c_\upsilon(\upsilon)\right)}{p\left(\varepsilon_{k_\upsilon}^i \mid c_{\not\sigma}(\sigma)\right)} p(\upsilon)\right)\Bigg)\Bigg)\end{aligned} \quad (8)$$

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 $\varepsilon_k^i \big| c_k^i$ 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.

$$\begin{aligned}E_{\varepsilon \sim p(\varepsilon \mid c_\sigma(\sigma))}\left(\frac{p(\varepsilon \mid c_\sigma(\sigma))}{p(\varepsilon \mid c_{\not\sigma}(\sigma))}\right) &= s_\sigma^1 \\ E_{\varepsilon \sim p(\varepsilon \mid c_{\not\sigma}(\sigma))}\left(\frac{p(\varepsilon \mid c_{\not\sigma}(\upsilon))}{p(\varepsilon \mid c_{\not\sigma}(\sigma))}\right) &= s_{\sigma,\upsilon}^0\end{aligned} \quad (9)$$

By design $\forall \sigma, \upsilon \neq \sigma \quad s_\sigma^1 \geq s_{\sigma,\upsilon}^0$ and $\sup_\upsilon s_{\sigma,\upsilon}^0 = 1$ 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 $s_\sigma^1 \to \infty, s_{\sigma,\upsilon}^0 \to 0$ with improved class separability. Notice that these scores are the expected values of the evidence likelihood ratios that appear in (8). Let $k_\sigma\left(\Pi^i\right) = \sum_k \delta_{c_k^i, c_\sigma}$ 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 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 $s_{\sigma,\upsilon}^0$. Thus the normalization term has both score values. The following equation is trivial to obtain from (8);

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

$$
\begin{aligned}
\widehat{\Pi}^i &= \arg\max_{\Pi^i} E_{p(\sigma|\boldsymbol{\varepsilon}^{i-1})}\left(k_\sigma \log s_\sigma^1 - \log\left(p(\sigma|\boldsymbol{\varepsilon}^{i-1})\left(s_\sigma^1\right)^{k_\sigma} + \sum_{\upsilon \neq \sigma}\left(s_{\sigma,\upsilon}^0\right)^{k_\upsilon} p(\upsilon)\right)\right) \\
&= \arg\min_{\Pi^i} E_p(\sigma|\boldsymbol{\varepsilon}^{i-1})\left(\log\left(p(\sigma|\boldsymbol{\varepsilon}^{i-1}) + \sum_{\upsilon \neq \sigma}\frac{\left(s_{\sigma,\upsilon}^0\right)^{k_\upsilon}}{\left(s_\sigma^1\right)^{k_\sigma}} p(\upsilon)\right)\right)
\end{aligned} \tag{10}
$$

Here we reduce to objective to a form which 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).

IV. Method

Here we propose a solution to Eq.(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 sub-modular 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 $s_{\sigma,\upsilon}^0$, we have the following observations.

Observation 1.

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

Observation 2.

Following Obs.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 $\Sigma\left(\Pi^i\right) = \left\{\sigma | c_k^i = c_\sigma \forall k\right\}$ we define the objective,

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

Observation 3.

Observe that the objective in (11) monotonically decreases wrt. 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 $\{x | x \notin \Pi_2\} \subset \{x | x \notin \Pi_1\}$ Define $\Sigma_1 = \Sigma(\Pi_1)$, $\Sigma_2 = \Sigma(\Pi_2)$ and observe that $\Sigma_1 \subseteq \Sigma_2$

$$O(\sigma,\Pi_2) - O(\sigma,\Pi_1) = p(a|\boldsymbol{\varepsilon}^{i-1})\left(\log\frac{p(a|\boldsymbol{\varepsilon}^{i-1})}{\sum_{\sigma\notin\Sigma_1} p(\sigma|\boldsymbol{\varepsilon}^{i-1})}\right) + \left(\sum_{\sigma\notin\Sigma_2} p(\sigma|\boldsymbol{\varepsilon}^{i-1})\right)\left(\log\frac{\sum_{\sigma\notin\Sigma_2} p(\sigma|\boldsymbol{\varepsilon}^{i-1})}{\sum_{\sigma\notin\Sigma_1} p(\sigma|\boldsymbol{\varepsilon}^{i-1})}\right) \le 0 \tag{12}$$

Using these observations we state that the optimization presented in (10) is a sub-modular 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 $\Pi^i_{0:k}$, and we would like to select a specific item $\boldsymbol{\pi}$ as the $(k+1)^{\text{th}}$ query with $\Sigma(\boldsymbol{\pi}) = a$, then we use the following greedy update,

$$\hat{\pi}^{k+1} = \arg\min_{\pi\notin\Pi^i_{0:k}} p(a|\boldsymbol{\varepsilon}^{i-1})\log\left(p(a|\boldsymbol{\varepsilon}^{i-1})\right) + \left(\sum_{\sigma\notin\Sigma\left(\Pi^i_k\cup\pi\right)} p(\sigma|\boldsymbol{\varepsilon}^{i-1})\right)\log\left(\sum_{\sigma\notin\Sigma\left(\Pi^i_s\cup\pi\right)} p(\sigma|\boldsymbol{\varepsilon}^{i-1})\right) \tag{13}$$

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

$$\hat{\pi}^{k+1} = \arg\min_{\pi\notin\Pi^i_k} p(a|\boldsymbol{\varepsilon}^{i-1})\log\left(p(a|\boldsymbol{\varepsilon}^{i-1})\right) + \left(\mathscr{P}_k - p(a|\boldsymbol{\varepsilon}^{i-1})\right)\log\left(\mathscr{P}_k - p(a|\boldsymbol{\varepsilon}^{i-1})\right) \quad \text{where } \mathscr{P}_k \in [0,1] \tag{14}$$

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

Proposition 1. Alg.1 can be realized with "N-best selection".

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

- $p(a|\boldsymbol{\varepsilon}^i) \ge \mathscr{P}_s/2$, $\exists\ p(\tilde{a}|\boldsymbol{\varepsilon}^i) = 1 - p(a|\boldsymbol{\varepsilon}^i)$ where $\tilde{\pi} \ne \pi$ and $p(\bar{a}|\boldsymbol{\varepsilon}^i) = 0\ \forall\bar{\pi} \ne \tilde{\pi}$. $O(\sigma,\pi) = O(\sigma,\tilde{\pi})$ therefore we can pick $\boldsymbol{\pi}$.
- $p(a|\boldsymbol{\varepsilon}^i) \ge \mathscr{P}_s/2$, $p(\tilde{a}|\boldsymbol{\varepsilon}^i) \ne 0\ \forall\tilde{\pi} \ne \pi$. It is trivial that the objective is minimized with $\boldsymbol{\pi}$.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

- Where $p(\tilde{a}|\varepsilon^i) < \mathscr{P}_s/2\ \forall\tilde{a}$. As the objective is strictly monotonically decreasing in $[0, \mathscr{P}_s]$, pick $\boldsymbol{\pi}$ with $\Sigma(\boldsymbol{\pi})$ satisfying the highest likelihood which minimizes the objective.

Algorithm 1 Mutual Information Based Query Selection

1: $\Pi^i \leftarrow \{\}, N \in \mathbb{Z}^+$

2: $x \leftarrow \{x_c | p(\sigma_c | \boldsymbol{\varepsilon}^{i-1}) \forall c\}$

3: **procedure** GREEDY SUBSET SELECTION(Π^i, σ, x)

4: $p \leftarrow 1$

5: **for** $i \in [1, 2, \cdots, |\Pi^i|]$ **do**

6: $\hat{c} = \arg\min_c x_c \log x_c + (p - x_c)\log(p - x_c)_c$

7: $\Pi^i \leftarrow \Pi^i \cup \{\Sigma^{-1}(\sigma_{\hat{c}})\}$

8: $p \leftarrow p - x_{\hat{c}}$

9: $x \leftarrow x \backslash \{x_{\hat{c}}\}$

10: **return:** Π^i

Observe that in any case $p(a|\varepsilon^i) \geq p(\tilde{a}|\boldsymbol{\varepsilon}^i) \forall \tilde{a} \neq \pi$. Other cases can be created by switching $\boldsymbol{\pi}$ without loss of generality. Since selection of $\mathscr{P}_s$ is arbitrary, one can conclude Alg.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 10 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 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/non-target 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 AUC ∈ [75, 85] and 6 users

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Author Manuscript

with AUC ∈ (85, 95]. We compare MMI with min/max querying [26], max only querying (sub-optimal variance minimization 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 Sec. 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 paper 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.

Acknowledgments

Our work is supported by NSF (IIS-1149570, CNS-1544895, IIS-1715858), DHHS (90RE5017-02-01), and NIH (R01DC009834)..

Our work is supported by NSF IIS-1717654, and by the Air Force Office of Scientific Research (AFOSR), the DDDAS Program, under Grant No. FA9550-16-1-0386.

Thanks to Paula Gonzalez Navarro for the efforts in data collection.

References

[1]. Gupta V, Chung TH, Hassibi B, and Murray RM, “On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage,” Automatica, vol. 42, no. 2, pp. 251–260, 2006.

[2]. Muller-Putz GR and Pfurtscheller G, “Control of an electrical prosthesis with an ssvep-based bci,” IEEE Transactions on Biomedical Engineering, vol. 55, no. 1, pp. 361–364, 2008. [PubMed: 18232384]

[3]. Orhan U, Hild KE, Erdogmus D, Roark B, Oken B, and Fried-Oken M, “Rsvp keyboard: An eeg based typing interface,” pp. 645–648, 2012.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[4]. Gauvain J-L and Lee C-H, "Maximum a posteriori estimation for multivariate gaussian mixture observations of markov chains," IEEE transactions on speech and audio processing, vol. 2, no. 2, pp. 291–298, 1994.

[5]. Arulampalam MS, Maskell S, Gordon N, and Clapp T, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," IEEE Transactions on signal processing, vol. 50, no. 2, pp. 174–188, 2002.

[6]. Denzler J and Brown CM, "Information theoretic sensor data selection for active object recognition and state estimation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 2, pp. 145–157, 2002.

[7]. Maes F, Collignon A, Vandermeulen D, Marchal G, and Suetens P, "Multimodality image registration by maximization of mutual information," IEEE transactions on medical imaging, vol. 16, no. 2, pp. 187–198, 1997. [PubMed: 9101328]

[8]. Meyer F, Hlinka O, Wymeersch H, Riegler E, and Hlawatsch F, "Distributed localization and tracking of mobile networks including noncooperative objects," IEEE Transactions on Signal and Information Processing over Networks, vol. 2, no. 1, pp. 57–71, 2016.

[9]. Chaloner K and Verdinelli I, "Bayesian experimental design: A review," Statistical Science, pp. 273–304, 1995.

[10]. Gosselin PH and Cord M, "Active learning methods for interactive image retrieval," IEEE Transactions on Image Processing, vol. 17, no. 7, pp. 1200–1211, 2008. [PubMed: 18586627]

[11]. Tong S and Chang E, "Support vector machine active learning for image retrieval," in Proceedings of the ninth ACM international conference on Multimedia. ACM, 2001, pp. 107–118.

[12]. Cohn DA, Ghahramani Z, and Jordan MI, "Active learning with statistical models," Journal of artificial intelligence research, 1996.

[13]. Chepuri SP and Leus G, "Sparsity-promoting sensor selection for nonlinear measurement models," IEEE Transactions on Signal Processing, vol. 63, no. 3, pp. 684–698, 2015.

[14]. Liu S, Chepuri SP, Fardad M, Maşazade E, Leus G, and Varshney PK, "Sensor selection for estimation with correlated measurement noise," IEEE Transactions on Signal Processing, vol. 64, no. 13, pp. 3509–3522, 2016.

[15]. Papadimitriou C, Beck JL, and Au S-K, "Entropy-based optimal sensor location for structural model updating," Journal of Vibration and Control, vol. 6, no. 5, pp. 781–800, 2000.

[16]. Tsiligkaridis T, Sadler BM, and Hero AO, "Collaborative 20 questions for target localization," IEEE Transactions on Information Theory, vol. 60, no. 4, pp. 2233–2252, 2014.

[17]. Jedynak B, Frazier PI, and Sznitman R, "Twenty questions with noise: Bayes optimal policies for entropy loss," Journal of Applied Probability, vol. 49, no. 1, pp. 114–136, 2012.

[18]. Tsiligkaridis T, Sadler BM, and Hero AO, "On decentralized estimation with active queries," IEEE Transactions on Signal Processing, vol. 63, no. 10, pp. 2610–2622, 2015.

[19]. Rajan P, Han W, Sznitman R, Frazier P, and Jedynak B, "Bayesian multiple target localization," Journal of Machine Learning Research, vol. 37, pp. 1945–1953, 2015.

[20]. Ertin E, Fisher JW, and Potter LC, "Maximum mutual information principle for dynamic sensor query problems," in IPSN, vol. 2003 Springer, 2003, pp. 405–416.

[21]. Wang H, Yao K, Pottie G, and Estrin D, "Entropy-based sensor selection heuristic for target localization," in Proceedings of the 3rd international symposium on Information processing in sensor networks. ACM, 2004, pp. 36–45.

[22]. Higger M, Quivira F, Akcakaya M, Moghadamfalahi M, Nezamfar H, Cetin M, and Erdogmus D, "Recursive bayesian coding for bcis," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 25, no. 6, pp. 704–714, 2017. [PubMed: 27416602]

[23]. Peng H, Long F, and Ding C, "Feature selection based on mutual information criteria of max-dependency, max-relevance, and minredundancy," IEEE Transactions on pattern analysis and machine intelligence, vol. 27, no. 8, pp. 1226–1238, 2005. [PubMed: 16119262]

[24]. Nemhauser GL, Wolsey LA, and Fisher ML, "An analysis of approximations for maximizing submodular set functionsi," Mathematical Programming, vol. 14, no. 1, pp. 265–294, 1978.

Author Manuscript
Author Manuscript
Author Manuscript
Author Manuscript

[25]. Orhan U, Nezamfar H, Akcakaya M, Erdogmus D, Higger M, Moghadamfalahi M, Fowler A, Roark B, Oken B, and Fried-Oken M, "Probabilistic simulation framework for eeg-based bci design," Brain-Computer Interfaces, vol. 3, no. 4, pp. 171–185, 2016. [PubMed: 29250562]

[26]. Mallapragada PK, Jin R, and Jain AK, "Active query selection for semi-supervised clustering," in Pattern Recognition, 2008. ICPR 2008. 19th International Conference on. IEEE, 2008, pp. 1–4.

Author Manuscript

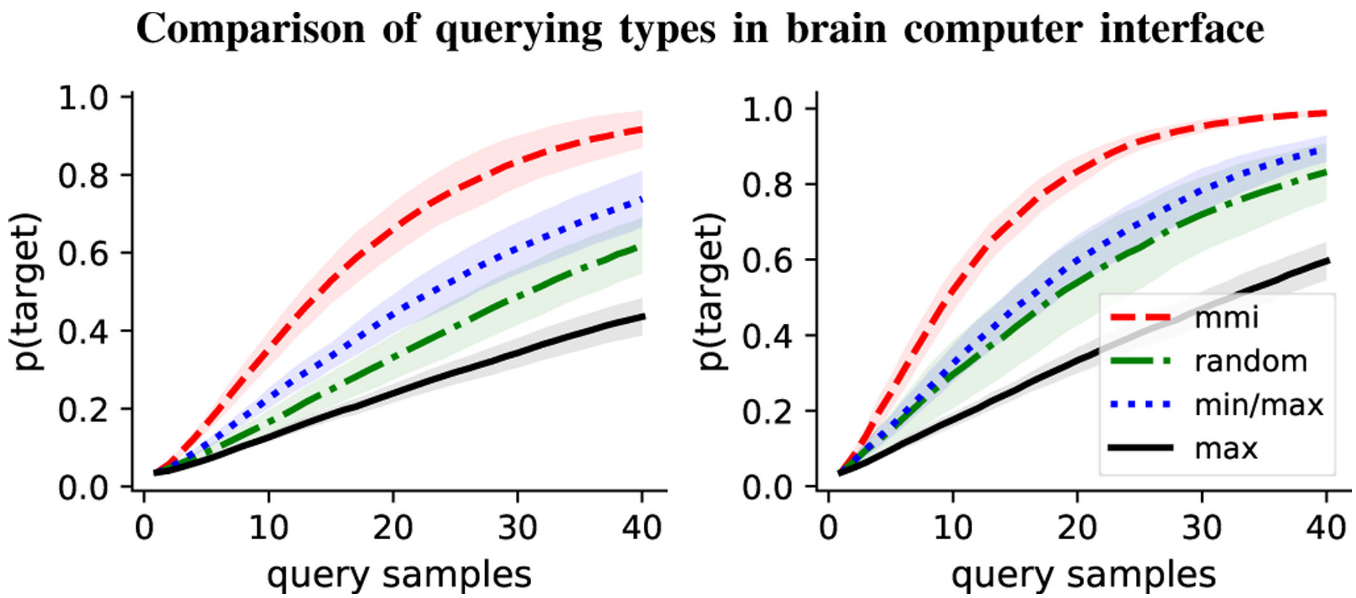

Fig. 1:
AUC ∈ [75, 85] (left) and ∈ (85, 95] (right).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

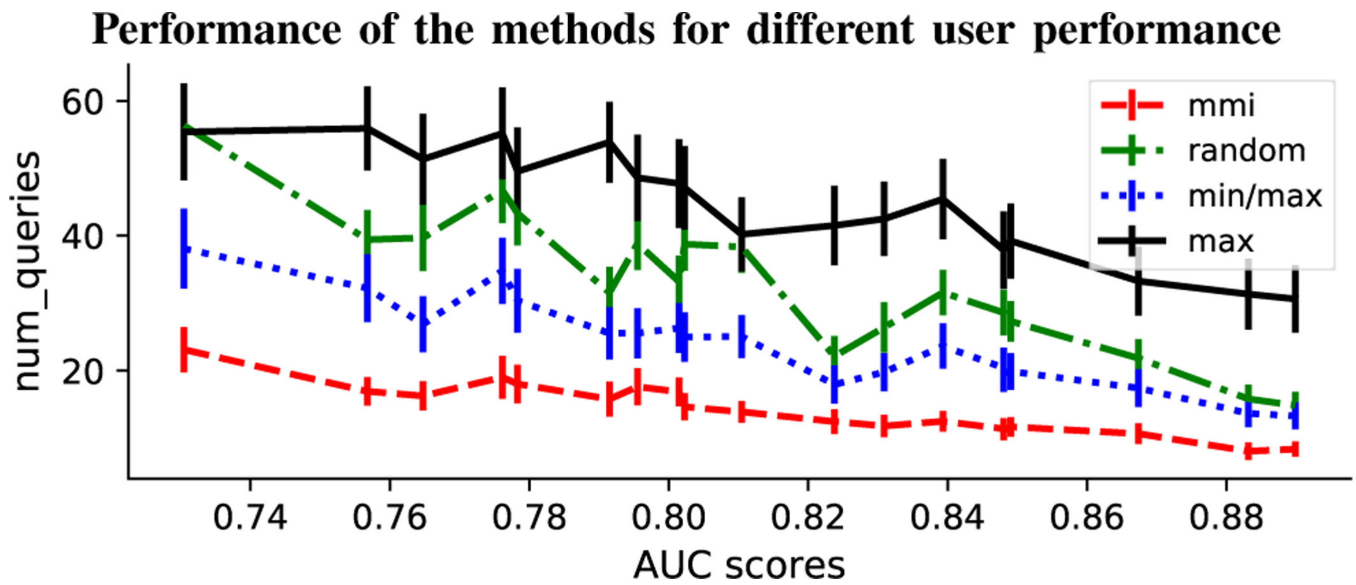

Fig. 2:

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