

A Rank-Based Approach to Active Diagnosis

Gowtham Bellala, *Member, IEEE*, Jason Stanley, Suresh K. Bhavnani, and Clayton Scott

Abstract—The problem of active diagnosis arises in several applications such as disease diagnosis and fault diagnosis in computer networks, where the goal is to rapidly identify the binary states of a set of objects (e.g., faulty or working) by sequentially selecting, and observing, potentially noisy responses to binary valued queries. Previous work in this area chooses queries sequentially based on Information gain, and the object states are inferred by maximum a posteriori (MAP) estimation. In this work, rather than MAP estimation, we aim to rank objects according to their posterior fault probability. We propose a greedy algorithm to choose queries sequentially by maximizing the area under the ROC curve associated with the ranked list. The proposed algorithm overcomes limitations of existing work. When multiple faults may be present, the proposed algorithm does not rely on belief propagation, making it feasible for large scale networks with little loss in performance. When a single fault is present, the proposed algorithm can be implemented without knowledge of the underlying query noise distribution, making it robust to any misspecification of these noise parameters. We demonstrate the performance of the proposed algorithm through experiments on computer networks, a toxic chemical database, and synthetic datasets.

Index Terms—Active diagnosis, active learning, Bayesian network, persistent noise, area under the ROC curve



1 INTRODUCTION

THE problem of diagnosis appears in various applications such as medical diagnosis [1], fault diagnosis in nuclear plants [2], computer networks [3], [4], and power-delivery systems [5], and decoding of messages sent through a noisy channel. In these problems, the goal is to identify the binary states $\mathbf{X} = (X_1, \dots, X_M)$ of M different objects based on the binary outcomes $\mathbf{Z} = (Z_1, \dots, Z_N)$ of N distinct queries/tests, where the query responses are noisy. Moreover, the query noise is persistent in that repeated querying results in the same query response. For example, in the problem of medical diagnosis, the goal is to identify the presence/absence of a set of diseases based on the noisy outcomes of medical tests. Similarly, in a fault diagnosis problem, the goal is to identify the state (faulty/working) of each component based on noisy alarm/probe responses. For simplicity, we will refer to an object with state 1 as a fault in the rest of this paper.

In recent years, this problem has been formulated as an inference problem on a Bayesian network, with the goal of assigning most probable states to unobserved object nodes based on the outcome of the query nodes. An important issue in diagnosis is the tradeoff between the cost of

querying (uncovering the value of some Z_j) and the achieved accuracy of diagnosis. It is often too expensive or time consuming to obtain responses to all queries.

In this paper, we study the problem of *active diagnosis*, where the queries are selected sequentially to maximize the accuracy of diagnosis. We study active diagnosis in two different settings. First, we consider the case where multiple faults could be present, as mentioned in applications stated above. Then, we consider a special case where only one fault could be present (i.e., only one object can be in state 1). This special scenario arises in applications such as pool-based active learning [6], [7], toxic chemical identification [8], image processing [9], computer vision [10], job scheduling [11], and the adaptive traveling salesperson problem [12]. For example, in the context of pool-based active learning, there is an unknown hypothesis (belonging to a known, finite hypothesis class) that could accurately classify all of the unlabeled data points, and the goal is to identify this hypothesis by obtaining (noisy) labels for as few data points as necessary.

Though the problem of active diagnosis has been studied in the literature, there are still several limitations with existing methods. We will now briefly mention some of the popular approaches to active diagnosis in these two settings and state their limitations. These existing approaches will be described in more detail in the following sections.

In the multiple fault scenario, Zheng et al. [4] proposed the use of reduction in conditional entropy (equivalently, mutual information) as a measure to select the most informative subset of queries. They proposed an algorithm that uses the belief propagation (BP) framework to select queries sequentially based on the gain in mutual information, given the observed responses to past queries. This algorithm, which they refer to as BPEA, requires one run of BP for each query selection. Finally, the objects are assigned the most likely states based on the outcome of the selected queries, using a maximum a posteriori (MAP) inference algorithm. Refer to Section 4.1 for more details.

- G. Bellala is with Hewlett Packard Laboratories, 1501 Page Mill Road, Building 1U, Mail Stop #1143, Palo Alto, CA 94304. E-mail: gowtham.bellala@hp.com.
- J. Stanley is with the Citadel Investment Group, 131 S. Dearborn St., Chicago, IL 60603. E-mail: jasonsta@umich.edu.
- S.K. Bhavnani is with the Institute for Translational Sciences, University of Texas Medical Branch, 301 University Blvd, Galveston, TX 77555. E-mail: skbhavnani@gmail.com.
- C. Scott is with the Department of Electrical Engineering and Computer Science, University of Michigan, 1301 Beal Avenue, Room #4433, Ann Arbor, MI 48109. E-mail: clayscot@umich.edu.

Manuscript received 29 Dec. 2011; revised 18 Sept. 2012; accepted 14 Jan. 2013; published online 24 Jan. 2013.

Recommended for acceptance by E.P. Xing.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2011-12-0937.

Digital Object Identifier no. 10.1109/TPAMI.2013.30.

However, there are two limitations with this approach. First, the MAP estimate may not equal the true state vector \mathbf{X} , either due to noise in the observed query responses or due to suboptimal convergence of the MAP algorithm. This leads to false alarm and miss rates that may not be tolerable for a given application.

The second issue is that BPEA does not scale to large networks because the complexity of computing the approximate value of conditional entropy grows exponentially in the maximum degree of the underlying Bayesian network (see Section 4.1 for details). As we show in Section 6, it becomes intractable even in networks with a few thousand objects. In addition, because this approach relies on BP, it may suffer from the limitations of BP such as slow convergence or oscillation of the algorithm, especially when the prior fault probability is small [13]. As we discuss below, the prior fault probability is indeed very low in most real-world diagnosis problems. These factors render BPEA impractical in many large scale, real-world applications.

The problem of active diagnosis in the single fault scenario has been extensively studied in the literature. It has often been assumed that queries can be resampled such that repeated querying results in independent query responses [14], [15]. However, in most diagnosis problems, the query noise persists in that repeated querying results in the same query response. This is a more restricted noise model known as persistent query noise. Persistent noise has been considered earlier in the context of pool-based active learning [7], where it is assumed that the query set is large enough (possibly infinite) such that the unknown object (e.g., hypothesis) could be identified with great certainty. However, in most diagnosis problems, the query set is finite, making these approaches inapplicable.

Alternatively, Rish et al. [3] and Geman and Jedynek [10] considered the use of gain in mutual information as a criterion for active diagnosis in the single fault scenario. Unlike the multiple fault case, information gain can be computed efficiently in the single fault scenario, thus making it a tractable approach (refer to Section 4.2 for more details). However, one other limitation with the use of information gain as a query selection criterion is that it requires knowledge of the full data model. In particular, it requires knowledge of the probability of query errors, and can be sensitive to discrepancies in these values.

In this paper, we propose a novel rank-based approach to active diagnosis that addresses the above limitations.¹ First, to address the limitation of the MAP estimate, we propose to output a ranked list of objects rather than their most likely states, where the ranking is based on their posterior fault probability. Given such a ranked list, the object states can be estimated by choosing a threshold t , where the top t objects in the ranked list are declared as faults (i.e., state 1) and the remaining as 0. Varying the threshold t leads to a series of estimators with different false alarm and miss rates, giving rise to a receiver operating characteristic (ROC) curve. The quality of the obtained ranking is then measured in terms of the area under this ROC curve (AUC). We show how to choose

queries greedily such that the AUC, and thus the quality of diagnosis, is maximized.

The rank-based approach is motivated by the fact that in many applications there is a domain expert who makes the final decision on the objects' states. Such a ranking can be extremely useful to a domain expert who will use domain expertise and other sources of information to choose a threshold t that may lead to a permissible value of false alarm and miss rates for a given application. Note that even in a single fault scenario, the object with the highest posterior fault probability (i.e., the top-ranked object) need not be the true object, especially in situations of moderate to high noise. Hence, a ranked list of the objects can be useful to a domain expert who could use other sources of information to further determine the unknown object.

Second, to address the issue of scalability in the multiple fault scenario, we circumvent the use of BP in the query selection stage by making a single fault approximation. To be clear, we still intend to apply our algorithm when multiple faults are present; the single fault approximation is used in the design of the algorithm. This approximation is reasonable because the prior fault probability is quite low in many applications. For example, in the problem of fault diagnosis in computer networks, the prior probability of a router failing in any given hour is on the order of 10^{-6} [18]. Similarly, in the disease diagnosis problem of QMR-DT, the prior probability of a disease being "present" is typically on the order of 10^{-3} [13].

Under the single fault approximation, the computational complexity of both the proposed AUC criterion and the information gain-based criterion reduces significantly, making them both feasible on large datasets. However, as we show in Section 5.1, when multiple faults are present, the AUC criterion under a single fault approximation still makes a good choice of queries, and performs significantly better than the queries selected using the information gain-based criterion under the single fault approximation.

Further, in the scenario where only one fault is present, we note that both these methods perform equally well when the underlying query noise model (i.e., the values of the noise parameters) is completely known. However, as we show in Section 5.2, any minor discrepancies in the values of these noise parameters could significantly reduce the performance of the information gain-based criterion, while the proposed AUC criterion can be implemented without any knowledge of the underlying noise parameters by using a tight upper bound that is independent of the noise parameters.

Finally, we demonstrate through experiments that the proposed query selection criterion can achieve performance close to that of BPEA *in a multifault setting*, while having a computational complexity that is of orders less than BPEA (near quadratic versus the exponential complexity of BPEA). In addition, in the single fault scenario, we demonstrate its competitive performance to information gain-based query selection, despite not having knowledge of the underlying query noise. In summary, the proposed rank-based framework is a fast, robust, and reliable approach for active diagnosis in large-scale, real-world diagnosis problems.

1. Preliminary versions of this work appeared in [16] and [17].

1.1 Additional Related Work

The problem of active diagnosis in the single fault scenario has been widely studied in the literature. In the noise-free case, this problem has been referred to as binary testing or object/entity identification [19]. The goal of object identification is to construct an optimal binary decision tree, where each internal node is associated with a query and each leaf node corresponds to an object where optimality is often with respect to the average depth of the leaf nodes. This problem of finding an optimal decision tree is known to be NP-complete [20]. However, there exists an efficient, greedy algorithm known as the *splitting algorithm* or *generalized binary search* (GBS) that achieves a logarithmic approximation to the optimal solution [6].

The problem of active learning in the presence of query noise has been studied in [14], [15], where the noise is assumed to be independent in that posing the same query twice may yield different responses. This assumption suggests repeated selection of a query as a possible strategy to overcome query noise. The algorithms presented in [14], [15] are based on this principle. However, in applications such as fault diagnosis in computer networks, resampling or repeating a query does not change the query response, thereby confining an active diagnosis algorithm to non-repeatable queries.

This more stringent noise model where queries cannot be resampled is known as persistent noise [7], [21]. It has been studied earlier in the situation where the number of persistent errors is restricted such that unique identification of the faulty object is still possible [22], [23]. In particular, let each object i be associated with a bit string of length N corresponding to the responses of the N queries if object i is in state 1. Then, the number of query errors is restricted to being less than half of the minimum Hamming distance between any two object bit strings. However, this is often not reasonable as the minimum Hamming distance could be very small.

Finally, the problem of active diagnosis under persistent noise with no restriction on the number of query errors has been studied in the context of pool-based active learning in the Probably Approximately Correct (PAC) model [7]. However, here the query set is assumed to be large enough (possibly infinite) such that it is possible to get arbitrarily close to the optimal classifier for any given noise level.

1.2 Outline

The rest of the paper is structured as follows: In Section 2, we describe the data model, and formulate the problem of active diagnosis in Section 3. In Section 4, we briefly describe information gain-based active query selection in both the multiple fault and single fault settings, and discuss its limitations in each of these settings in more detail. In Section 5, we propose our rank-based active query selection algorithm, and describe how the AUC criterion overcomes the limitations of mutual information. Finally, in Section 6, we demonstrate the performance of the proposed rank-based approach through experiments on computer networks, a toxic chemical database, and synthetic networks.

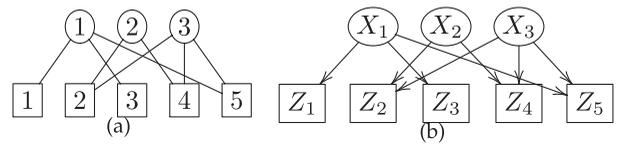


Fig. 1. (a) A toy BDG, where the circled nodes denote the objects and the square nodes denote queries. (b) A Bayesian network corresponding to the given BDG.

2 DATA MODEL

We will represent a diagnosis problem by a bipartite graph between a set of M different objects and a set of N distinct queries. This graph will be referred to as a bipartite diagnosis graph (BDG). The edges represent the relation or the interactions between the two entities. For example, in a fault diagnosis problem, the objects correspond to components and queries to alarms, where an edge indicates that a particular component-alarm pair is connected. Similarly, in a disease diagnosis problem, objects may correspond to diseases and queries to symptoms, where an edge indicates that a particular symptom is exhibited by a disease. Fig. 1 demonstrates a toy BDG.

We denote the state of each object (e.g., presence/absence of a disease) with a binary random variable X_i and the state of each query (i.e., the observed response to a query) by a binary random variable Z_j . Then, $\mathbf{X} = (X_1, \dots, X_M)$ is a binary random vector denoting the states of all the objects, and $\mathbf{Z} = (Z_1, \dots, Z_N)$ is a binary random vector denoting the responses to all the queries. We let $\mathbf{x} \in \{0, 1\}^M$ and $\mathbf{z} \in \{0, 1\}^N$ correspond to realizations of \mathbf{X} and \mathbf{Z} , respectively.

In addition, for any subset of queries $\mathcal{A} \subseteq \{1, \dots, N\}$, we denote by $\mathbf{Z}_{\mathcal{A}}$ the random variables associated with those queries, e.g., if $\mathcal{A} = \{1, 4, 7\}$, then $\mathbf{Z}_{\mathcal{A}} = (Z_1, Z_4, Z_7)$. Also, for any query j , let pa_j denote the objects that are connected to it in the BDG. Then, \mathbf{X}_{pa_j} denotes the states of all the objects connected to query j , e.g., for query 2 in Fig. 1, $\mathbf{X}_{\text{pa}_2} = (X_2, X_3)$.

We need to specify the joint distribution of (\mathbf{X}, \mathbf{Z}) and, more generally, $(\mathbf{X}, \mathbf{Z}_{\mathcal{A}})$ for any \mathcal{A} , which can be defined in terms of a prior probability distribution on \mathbf{X} and a conditional distribution on $\mathbf{Z}_{\mathcal{A}}$ given \mathbf{X} . To define the prior probability distribution on \mathbf{X} , we make the standard assumption that the object states are marginally independent, i.e., $\Pr(\mathbf{X} = \mathbf{x}) = \prod_{i=1}^M \Pr(X_i = x_i)$. Similarly, to define the conditional distribution on $\mathbf{Z}_{\mathcal{A}}$ given \mathbf{X} , we make the standard assumption that the observed responses to queries are conditionally independent given the states of the objects connected to them, i.e.,

$$\Pr(\mathbf{Z}_{\mathcal{A}} = \mathbf{z}_{\mathcal{A}} | \mathbf{X} = \mathbf{x}) = \prod_{j \in \mathcal{A}} \Pr(Z_j = z_j | \mathbf{x}_{\text{pa}_j}).$$

These assumptions hold reasonably well in many practical applications. For example, in a fault diagnosis problem, it can be reasonable to assume that the components fail independently and that the alarm responses are conditionally independent given the states of the components they are connected to, as justified in [3] and [18]. These dependencies can be encoded by a Bayesian network, as shown in Fig. 1.

In the ideal case when there is no noise, the observed response Z_j to query j is deterministic given the binary states of the objects in \mathbf{pa}_j . Specifically, it is given by the OR operation of the binary variables in $\mathbf{X}_{\mathbf{pa}_j}$, i.e., $Z_j = 1 \iff \exists i \in \mathbf{pa}_j$ s.t. $X_i = 1$. More generally, it is a noisy OR operation [24], where the conditional distribution of Z_j given $\mathbf{x}_{\mathbf{pa}_j}$ can be defined using standard noise models such as the Y-model [25] or the QMR-DT model [1].

We derive the AUC-based active diagnosis algorithm under this general probability model, and in Section 6, we demonstrate the performance of the proposed algorithm through experiments on both synthetic and real datasets under the QMR-DT noise model, where

$$\Pr(X_i = x) := (\alpha_i)^x (1 - \alpha_i)^{1-x}, \text{ and}$$

$$\Pr(Z_j = 0 | \mathbf{x}_{\mathbf{pa}_j}) := (1 - \rho_{0j}) \prod_{k \in \mathbf{pa}_j} \rho_{kj}^{x_k}.$$

Here, α_i is the prior fault probability, and ρ_{kj} and ρ_{0j} are the so-called inhibition and leak probabilities, respectively.

3 ACTIVE DIAGNOSIS

The goal of active diagnosis is to find a set of queries \mathcal{A} , subject to a constraint on the number of queries made, such that the expected value of a function $f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}})$ is maximized, i.e.,

$$\begin{aligned} \operatorname{argmax}_{\mathcal{A} \subseteq \{1, \dots, N\}} \mathbb{E}_{\mathbf{Z}_{\mathcal{A}}} [f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}})] \\ \text{s.t. } |\mathcal{A}| \leq k, \end{aligned} \quad (1)$$

where $f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}})$ corresponds to the quality of the estimate of \mathbf{X} based on the responses to queries in \mathcal{A} . The quality function $f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}} = \mathbf{z}_{\mathcal{A}})$ would implicitly involve an expectation over \mathbf{X} . Finding an optimal solution to this problem is typically not computationally feasible [3]. Instead, the queries can be chosen sequentially by greedily maximizing the quality function, i.e., given the observed responses to the past queries, the next best query is chosen to be

$$j^* := \operatorname{argmax}_{j \notin \mathcal{A}} \mathbb{E}_{Z_j} [f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}} \cup Z_j) - f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}}) | \mathbf{Z}_{\mathcal{A}} = \mathbf{z}_{\mathcal{A}}], \quad (2)$$

where $\mathbf{Z}_{\mathcal{A}} \cup Z_j$ denotes the random variables associated with queries in $\mathcal{A} \cup \{j\}$.

4 INFORMATION GAIN-BASED ACTIVE QUERY SELECTION

Mutual information has been traditionally chosen as a function to measure the quality of the estimate of the object states \mathbf{X} based on the responses to queries in \mathcal{A} . The expression for the quality function $f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}})$ is then given by $f(\mathbf{X}; \mathbf{Z}_{\mathcal{A}}) = I(\mathbf{X}; \mathbf{Z}_{\mathcal{A}}) := H(\mathbf{X}) - H(\mathbf{X} | \mathbf{Z}_{\mathcal{A}})$. However, the optimization problem in (1) with mutual information as the quality function is NP-hard [3]. Alternatively, the greedy approach can be used to choose queries sequentially where, given the observed responses $\mathbf{z}_{\mathcal{A}}$ to previously selected queries in \mathcal{A} , the next best query is chosen to be the one that maximizes the expected information gain, as shown below,

$$\begin{aligned} j^* &= \operatorname{argmax}_{j \notin \mathcal{A}} \mathbb{E}_{Z_j} [I(\mathbf{X}; \mathbf{Z}_{\mathcal{A}} \cup Z_j) - I(\mathbf{X}; \mathbf{Z}_{\mathcal{A}}) | \mathbf{Z}_{\mathcal{A}} = \mathbf{z}_{\mathcal{A}}] \\ &= \operatorname{argmin}_{j \notin \mathcal{A}} \sum_{z=0,1} \Pr(Z_j = z | \mathbf{z}_{\mathcal{A}}) H(\mathbf{X} | \mathbf{z}_{\mathcal{A}}, z). \end{aligned} \quad (3)$$

Note that information gain-based greedy query selection reduces to choosing a query that minimizes the expected conditional entropy of the object states \mathbf{X} . Hence, in the rest of this paper, we will refer to this approach as *entropy-based* active query selection. Below, we will describe how this greedy strategy has been implemented under the multiple fault and the single-fault settings, along with their limitations.

4.1 Multiple Fault Scenario

In the multiple fault scenario, the conditional entropy is given by

$$H(\mathbf{X} | \mathbf{z}_{\mathcal{A}}, z) = - \sum_{\mathbf{x} \in \{0,1\}^M} \Pr(\mathbf{x} | \mathbf{z}_{\mathcal{A}}, z) \log_2 \Pr(\mathbf{x} | \mathbf{z}_{\mathcal{A}}, z).$$

Note that direct computation of this expression is intractable. However, Zheng et al. [4] note that under the independence assumptions of Section 2, the conditional entropy can be simplified such that the query selection criterion in (3) is reduced to

$$\begin{aligned} \operatorname{argmin}_{j \notin \mathcal{A}} \left[- \sum_{\mathbf{x}_{\mathbf{pa}_j}, z} \Pr(\mathbf{x}_{\mathbf{pa}_j}, z | \mathbf{z}_{\mathcal{A}}) \log_2 \Pr(Z_j = z | \mathbf{x}_{\mathbf{pa}_j}) \right. \\ \left. + \sum_{z=0,1} \Pr(Z_j = z | \mathbf{z}_{\mathcal{A}}) \log_2 \Pr(Z_j = z | \mathbf{z}_{\mathcal{A}}) + \text{const} \right]. \end{aligned}$$

In addition, they propose an approximation algorithm that uses the loopy BP infrastructure to compute the above expression, which they refer to as belief propagation for entropy approximation (BPEA). Interestingly, this algorithm requires only one run of loopy BP for each query selection. After observing responses $\mathbf{z}_{\mathcal{A}}$ to a set of queries in \mathcal{A} , the object states are then estimated to be $\mathbf{x}^{\text{MAP}} := \operatorname{argmax}_{\mathbf{x} \in \{0,1\}^M} \Pr(\mathbf{X} = \mathbf{x} | \mathbf{z}_{\mathcal{A}})$, where the MAP estimator is obtained using a loopy version of the max-product algorithm. As far as we know, BPEA is the best known solution to the problem of active query selection in the multiple fault scenario.

However, this approach does not scale to large networks as BPEA involves a term whose computation grows exponentially in the number of parents to a query node. If m denotes the maximum number of parents to any query node, i.e., $m := \max_{j \in \{1, \dots, N\}} |\mathbf{pa}_j|$, then the computational complexity of choosing a query using BPEA is $O(N2^m)$, thus making it intractable in networks where m is greater than 25 or even less, especially when real-time query selection is desired.

Recently, Cheng et al. [26] proposed a speed up to query selection using BPEA by reducing the number of queries to be investigated at each stage. However, the exponential complexity still remains. Alternatively, we propose to assume a single fault in the query selection stage. As mentioned earlier, this approximation is motivated by the fact that in most diagnosis problems, the prior fault probability is very low. However, it is important for the

query selection criterion to be robust to violations of the single fault approximation, as multiple faults could be present in practice. As we show in Section 5.1, entropy-based query selection is not robust to such violations and can perform poorly when multiple faults are present.

4.2 Single Fault Scenario

In this special case where only one object can be in state 1 (i.e., only a single fault is present), the object state vector \mathbf{X} belongs to the set $\{\mathbb{I}_1, \dots, \mathbb{I}_M\}$, where \mathbb{I}_i is a binary vector whose i th element is 1 and the remaining elements are 0. This reduction in the state space of the object vector allows for efficient computation of the conditional entropy term in the query selection criterion of (3).

More specifically, the conditional entropy $H(\mathbf{X}|\mathbf{z}_A, z) = -\sum_{i=1}^M \Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A, z) \log_2 \Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A, z)$ can be computed in $O(M)$ complexity given the posterior probabilities. Moreover, using the conditional independence assumption of Section 2 whereby $\Pr(Z_j|\mathbf{X}, \mathbf{z}_A) = \Pr(Z_j|\mathbf{X})$, the posterior probabilities can be updated efficiently in $O(M)$ time as follows:

$$\begin{aligned} \Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A = \mathbf{z}_A, Z_j = z) &= \frac{\Pr(\mathbf{X} = \mathbb{I}_i, Z_j = z|\mathbf{z}_A = \mathbf{z}_A)}{\Pr(Z_j = z|\mathbf{z}_A = \mathbf{z}_A)} \\ &= \frac{\Pr(\mathbf{X} = \mathbb{I}_i, Z_j = z|\mathbf{z}_A = \mathbf{z}_A)}{\sum_{k=1}^M \Pr(\mathbf{X} = \mathbb{I}_k, Z_j = z|\mathbf{z}_A = \mathbf{z}_A)} \\ &= \frac{\Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A = \mathbf{z}_A)\Pr(Z_j = z|\mathbf{X} = \mathbb{I}_i)}{\sum_{k=1}^M \Pr(\mathbf{X} = \mathbb{I}_k|\mathbf{z}_A = \mathbf{z}_A)\Pr(Z_j = z|\mathbf{X} = \mathbb{I}_k)}. \end{aligned} \quad (4)$$

However, another critical limitation with the use of entropy-based active query selection is that it requires knowledge of the complete noise distribution or the parameters in the noise model. In particular, these parameters are required for the computation of posterior probabilities. As we show in Section 6, entropy-based active query selection can be sensitive to any discrepancies in the knowledge of these parameters.

In the next section, we derive a new query selection criterion that sequentially chooses queries such that the AUC of a rank-based output is maximized. We will show how this query selection criterion overcomes the limitations of entropy-based active query selection. In particular, we will show that the computational complexity of greedily choosing a query in the multiple fault scenario can be significantly reduced (from exponential to near-quadratic) using the proposed criterion, with little loss in performance. In addition, we will also show that the proposed algorithm can be implemented efficiently without knowledge of the underlying noise distribution in the single fault setting.

5 AUC-BASED ACTIVE QUERY SELECTION

AUC has been used earlier as a performance criterion in the classification setting with decision tree classifiers [27] and boosting [28], in the problem of ranking [29], and in an active learning setting [30]. In all the earlier settings, the AUC of a classifier is estimated using the training data whose binary labels are known. However, in our setting, the object states (binary labels) are neither known nor does

	X_1	X_2	X_3	X_4	X_5
$\Pr(X_i = 1 \mathbf{z}_A)$	0.3	0.15	0.35	0.15	0.05

Fig. 2. A rank order for the above example is $\mathbf{r} = (3, 1, 2, 4, 5)$.

there exist any training data. Hence, we propose a simple estimator for the AUC based on the posterior probabilities of the object states. Specifically, we propose three variants of this estimator, and discuss some interesting properties of each of these variants in the two settings.

Given the observed responses \mathbf{z}_A to queries in \mathcal{A} , let the objects be ranked based on their posterior fault probabilities, i.e., $\Pr(X_i = 1|\mathbf{z}_A)$, where ties involving objects with the same posterior probability are broken randomly. Then, let $\mathbf{r} = (r(1), \dots, r(M))$ denote the rank order of the objects, where $r(i)$ denotes the index of the i th ranked object. For example, a rank order corresponding to the toy example in Fig. 2 is $\mathbf{r} = (3, 1, 2, 4, 5)$. Also, note that \mathbf{r} depends on the queries chosen \mathcal{A} and their observed responses \mathbf{z}_A , though it is not explicitly shown in our notation.

Given this ranked list of objects, we get a series of estimators $\{\hat{\mathbf{x}}^t\}_{t=0}^M$ for the object state vector \mathbf{X} , where $\hat{\mathbf{x}}^t$ corresponds to the estimator which declares the states of the top t objects in the ranked list as 1 and the remainder as 0. For example, $\hat{\mathbf{x}}^2 = (1, 0, 1, 0, 0)$ for the toy example shown in Fig. 2.

These estimators have different false alarm and miss rates. The miss and false alarm rates associated with $\hat{\mathbf{x}}^t$ are given by

$$\begin{aligned} \text{MR}_t &= \frac{\sum_{\{i:\hat{x}_i=0\}} \mathbf{I}\{X_i = 1\}}{\sum_{i=1}^M \mathbf{I}\{X_i = 1\}} = \frac{\sum_{i=t+1}^M \mathbf{I}\{X_{r(i)} = 1\}}{\sum_{i=1}^M \mathbf{I}\{X_i = 1\}}, \\ \text{FAR}_t &= \frac{\sum_{\{i:\hat{x}_i=1\}} \mathbf{I}\{X_i = 0\}}{\sum_{i=1}^M \mathbf{I}\{X_i = 0\}} = \frac{\sum_{i=1}^t \mathbf{I}\{X_{r(i)} = 0\}}{\sum_{i=1}^M \mathbf{I}\{X_i = 0\}}, \end{aligned}$$

where $\mathbf{I}\{E\}$ is an indicator function which takes the value 1 when the event E is true, and 0 otherwise.

However, because the true states of the objects are not known, the false alarm and the miss rates need to be estimated. Given the responses \mathbf{z}_A to queries in \mathcal{A} , these two error rates can be approximated by using the expected value of the numerator and denominator conditioned on these responses as shown below:

$$\widehat{\text{MR}}_t(\mathbf{z}_A) = \frac{\sum_{i=t+1}^M \Pr(X_{r(i)} = 1|\mathbf{z}_A)}{\sum_{i=1}^M \Pr(X_i = 1|\mathbf{z}_A)}, \quad (5a)$$

$$\widehat{\text{FAR}}_t(\mathbf{z}_A) = \frac{\sum_{i=1}^t \Pr(X_{r(i)} = 0|\mathbf{z}_A)}{\sum_{i=1}^M \Pr(X_i = 0|\mathbf{z}_A)}. \quad (5b)$$

Note that these estimators are not the expected values for the error rates, but rather an approximation to the expected values. The complexity of computing the true expected values is exponential in M , while the above approximations can be computed in linear time. Moreover, as responses to more queries are obtained, the posterior fault probabilities tend to be close to either 0 or 1, thus making the above estimates close to their true values, respectively.

Using these estimates, the ROC curve can then be obtained by varying the threshold t from 0 to M leading

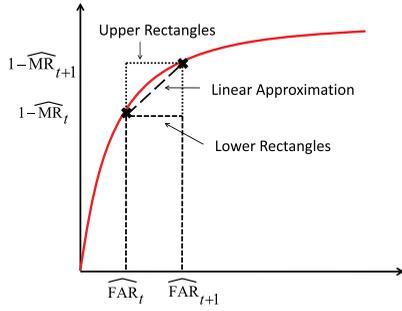


Fig. 3. The different approximations for AUC.

to different false alarm and miss rates. For example, $\hat{\mathbf{x}}^0$, which declares the states of all the objects to be equal to 0, has a false alarm rate of 0 and a miss rate of 1. On the other hand, $\hat{\mathbf{x}}^M$, which declares the states of all objects as 1, has a false alarm rate of 1 with a miss rate of 0. The other estimators have false alarm and miss rates that span the space between these two extremes.

Finally, the AUC can be estimated using a piecewise approximation with either lower rectangles, upper rectangles, or a linear approximation, as shown in Fig. 3. As we discuss later, each of these variants have some interesting properties in different settings. The expressions related to each of these approximations are

$$\begin{aligned} \underline{\mathbf{A}}_{lr}(\mathbf{z}_A) &= \sum_{t=0}^{M-1} (1 - \widehat{\text{MR}}_t)(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t), \\ \underline{\mathbf{A}}_{ur}(\mathbf{z}_A) &= \sum_{t=0}^{M-1} (1 - \widehat{\text{MR}}_{t+1})(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t), \text{ and} \\ \underline{\mathbf{A}}_l(\mathbf{z}_A) &= \sum_{t=0}^{M-1} \left(1 - \frac{\widehat{\text{MR}}_t + \widehat{\text{MR}}_{t+1}}{2}\right)(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t), \end{aligned}$$

where we dropped the dependence of $\widehat{\text{MR}}_t$ and $\widehat{\text{FAR}}_t$ on \mathbf{z}_A to avoid cramping. Further, noting that $\widehat{\text{FAR}}_M = 1$ and $\widehat{\text{FAR}}_0 = 0$, $\underline{\mathbf{A}}_{lr}(\mathbf{z}_A)$ can be rewritten

$$\begin{aligned} \underline{\mathbf{A}}_{lr}(\mathbf{z}_A) &= \sum_{t=0}^{M-1} (1 - \widehat{\text{MR}}_t)(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t) \\ &= \widehat{\text{FAR}}_M - \widehat{\text{FAR}}_0 - \sum_{t=0}^{M-1} \widehat{\text{MR}}_t(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t) \\ &= 1 - \sum_{t=0}^{M-1} \widehat{\text{MR}}_t(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t), \end{aligned} \quad (6)$$

where $\sum_{t=0}^{M-1} \widehat{\text{MR}}_t(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t)$ corresponds to an estimate of the area *above* the ROC curve using lower rectangles, which we denote by $\underline{\mathbf{A}}_{lr}(\mathbf{z}_A)$. Similarly, the estimates of the area *above* the ROC curve using upper rectangles or a linear approximation are given by

$$\begin{aligned} \overline{\mathbf{A}}_{ur}(\mathbf{z}_A) &= \sum_{t=0}^{M-1} \widehat{\text{MR}}_{t+1}(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t) \text{ and} \\ \overline{\mathbf{A}}_l(\mathbf{z}_A) &= \sum_{t=0}^{M-1} \frac{\widehat{\text{MR}}_t + \widehat{\text{MR}}_{t+1}}{2}(\widehat{\text{FAR}}_{t+1} - \widehat{\text{FAR}}_t). \end{aligned}$$

Substituting the estimates for miss rate and false alarm rate from (5a) and (5b), the corresponding approximations for the area above the ROC curve are given by

$$\overline{\mathbf{A}}_{lr}(\mathbf{z}_A) = \frac{\sum_{i=1}^M \sum_{j=i}^M \mathbf{N}_{i,j}(\mathbf{z}_A)}{\mathbf{D}(\mathbf{z}_A)}, \quad (7a)$$

$$\overline{\mathbf{A}}_{ur}(\mathbf{z}_A) = \frac{\sum_{i=1}^{M-1} \sum_{j=i+1}^M \mathbf{N}_{i,j}(\mathbf{z}_A)}{\mathbf{D}(\mathbf{z}_A)}, \quad (7b)$$

$$\begin{aligned} \overline{\mathbf{A}}_l(\mathbf{z}_A) &= \frac{\sum_{i=1}^{M-1} \sum_{j=i+1}^M \mathbf{N}_{i,j}(\mathbf{z}_A)}{\mathbf{D}(\mathbf{z}_A)} \\ &+ \frac{\sum_{i=1}^M \Pr(X_i = 0|\mathbf{z}_A)\Pr(X_i = 1|\mathbf{z}_A)}{2\mathbf{D}(\mathbf{z}_A)}, \end{aligned} \quad (7c)$$

where $\mathbf{N}_{i,j}(\mathbf{z}_A) = \Pr(X_{r(i)} = 0|\mathbf{z}_A)\Pr(X_{r(j)} = 1|\mathbf{z}_A)$ and $\mathbf{D}(\mathbf{z}_A) = \sum_{i=1}^M \Pr(X_i = 1|\mathbf{z}_A) \sum_{i=1}^M \Pr(X_i = 0|\mathbf{z}_A)$.

Using the AUC as a quality function, the goal of active diagnosis is to maximize the accuracy of diagnosis given by the estimate of AUC, subject to a constraint on the number of queries made, i.e.,

$$\begin{aligned} \max_{\mathcal{A} \subseteq \{1, \dots, N\}} \underline{\mathbf{A}}(\mathbf{z}_A) \\ \text{s.t. } |\mathcal{A}| \leq k, \end{aligned}$$

where $\underline{\mathbf{A}}(\mathbf{z}_A)$ corresponds to an estimate of the AUC using any of the above approximations. More generically, in the rest of this paper, we will use the terms $\underline{\mathbf{A}}(\mathbf{z}_A)$ and $\overline{\mathbf{A}}(\mathbf{z}_A)$ to denote any of the above approximations for AUC and area above the ROC curve, respectively.

Once again the above optimization problem is NP-hard. Hence, we resort to the greedy strategy, where substituting this quality function in (2), we get the criterion for greedily choosing a query to be

$$\begin{aligned} j^* &= \operatorname{argmax}_{j \notin \mathcal{A}} \mathbb{E}_{Z_j} [\underline{\mathbf{A}}(\mathbf{Z}_A \cup Z_j) - \underline{\mathbf{A}}(\mathbf{Z}_A) | \mathbf{Z}_A = \mathbf{z}_A] \\ &= \operatorname{argmax}_{j \notin \mathcal{A}} \sum_{z=0,1} \Pr(Z_j = z | \mathbf{z}_A) \underline{\mathbf{A}}(\mathbf{z}_A \cup z) \\ &= \operatorname{argmin}_{j \notin \mathcal{A}} \sum_{z=0,1} \Pr(Z_j = z | \mathbf{z}_A) \overline{\mathbf{A}}(\mathbf{z}_A \cup z), \end{aligned} \quad (8)$$

where the second equality follows as $\underline{\mathbf{A}}(\mathbf{Z}_A)$ is independent of Z_j , and the last equality follows from (6), i.e., $\underline{\mathbf{A}}(\mathbf{z}_A) = 1 - \overline{\mathbf{A}}(\mathbf{z}_A)$.

5.1 Multiple Fault Scenario

Note that both the query selection criterion in (8) and the different approximations to the quality function $\overline{\mathbf{A}}(\mathbf{z}_A)$ in (7) depend only on the posterior probabilities of unobserved nodes given the states of the observed nodes. Since these probabilities can be approximated using loopy BP, the AUC-based active query selection can be performed using loopy BP similar to the entropy-based active query selection in BPEA.

However, a main focus of this paper is on active diagnosis for large scale networks, where query selection using loopy BP is slow and possibly intractable. Hence, we make a single fault approximation to compute the posterior probabilities during the query selection stage. Under

this approximation, both the AUC and the entropy criterion can be computed efficiently. However, as we argue below, when multiple faults are present, the AUC criterion under a single fault approximation still makes a good choice of queries, and performs significantly better than the queries selected using the entropy criterion under a single fault approximation.

Under this approximation, the object state vector \mathbf{X} is restricted to belong to the set $\{\mathbb{I}_1, \dots, \mathbb{I}_M\}$ in the query selection stage. This reduction in the state space of the object vector allows for query selection to be performed efficiently without the need for looped BP. More specifically, the posterior probabilities required to choose queries sequentially in (8) can be computed as described below.

Using the conditional independence assumption of Section 2,

$$\Pr(Z = z|\mathbf{z}_A) = \sum_{i=1}^M \Pr(Z = z|\mathbf{X} = \mathbb{I}_i)\Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A),$$

where the posterior probabilities $\Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A)$ can be updated efficiently in $O(M)$ time as shown in (4). Also, note that under a single fault approximation,

$$\sum_{i=1}^M \Pr(X_i = 1|\mathbf{z}_A) = \sum_{i=1}^M \Pr(\mathbf{X} = \mathbb{I}_i|\mathbf{z}_A) = 1, \quad (9a)$$

$$\sum_{i=1}^M \Pr(X_i = 0|\mathbf{z}_A) = M - 1. \quad (9b)$$

Using these constraints, the estimates for the area above the ROC curve in (7) can be equivalently expressed as shown in the following proposition.

Proposition 1. *Under the single fault approximation, the estimates for the area above the ROC curve, in (7), can be equivalently expressed as*

$$\bar{\mathbf{A}}_{lr}(\mathbf{z}_A) = \frac{\sum_{i=1}^M [2i + \Pr(X_{r(i)} = 0|\mathbf{z}_A)]\Pr(X_{r(i)} = 1|\mathbf{z}_A)}{2(M-1)}, \quad (10a)$$

$$\bar{\mathbf{A}}_l(\mathbf{z}_A) = \frac{\sum_{i=1}^M [2i]\Pr(X_{r(i)} = 1|\mathbf{z}_A)}{2(M-1)}, \quad (10b)$$

$$\bar{\mathbf{A}}_{ur}(\mathbf{z}_A) = \frac{\sum_{i=1}^M [2i - \Pr(X_{r(i)} = 0|\mathbf{z}_A)]\Pr(X_{r(i)} = 1|\mathbf{z}_A)}{2(M-1)}. \quad (10c)$$

Proof. Refer to the supplemental material, which can be found online in the Computer Society Digital Library at <http://doi.ieeecomputersociety.org/10.1109/TPAMI.2013.30> or in [31]. \square

Note from this result that given a ranked list of the objects along with their posterior probabilities, the complexity of estimating the area above the ROC curve $\bar{\mathbf{A}}(\mathbf{z}_A)$ under a single fault approximation is $O(M)$. Since the posterior probabilities can also be updated efficiently in $O(M)$ time,

the complexity of computing $\bar{\mathbf{A}}(\mathbf{z}_A)$ is dominated by the complexity of sorting, which is $O(M \log M)$. Hence, the computational complexity of choosing a query at each stage using the AUC-based criterion under a single fault approximation is $O(NM \log M)$. This allows active query selection to be tractable even in large networks.

However, multiple faults could be present in practice, and hence, it is important for a query selection criterion under a single fault approximation to be robust to violations of that approximation. We will now provide an intuitive explanation as to why the proposed AUC criterion makes a robust choice of queries under a single fault approximation, while the entropy-based criterion fails to do so. We will demonstrate the same through extensive experiments in Section 6.

The following result helps to explain why entropy-based query selection under a single fault approximation performs poorly in a multifault setting.

Proposition 2. *Under the single fault approximation and the conditional independence assumption of Section 2, the entropy-based query selection criterion in (2) reduces to*

$$j^* := \operatorname{argmin}_{j \notin A} \sum_{i=1}^M \Pr(X_i = 1|\mathbf{z}_A)H(\Pr(Z_j = 0|X_i = 1)) - H(\Pr(Z_j = 0|\mathbf{z}_A)), \quad (11)$$

where $H(p) := -p \log_2 p - (1-p) \log_2 (1-p)$ denotes the binary entropy function.

Proof. Refer to the supplemental material, available online in the CSDL or at [31]. \square

As noted in (9a), under a single fault approximation, the posterior fault probabilities are constrained to sum to 1. Hence, objects with high posterior fault probability decrease the posterior fault probabilities of the remaining objects. Given this scenario, note from (11) in Proposition 2 that both the terms in this query selection criterion are highly dominated by the object(s) with high posterior fault probabilities (even the second term, since $\Pr(Z_j = 0|\mathbf{z}_A) = \sum_{i=1}^M \Pr(X_i = 1|\mathbf{z}_A)\Pr(Z_j = 0|X_i = 1)$). Hence, at any given stage, the query chosen according to this criterion is highly biased toward objects that already have a high posterior fault probability. This could lead to a poor choice of queries as the objects with high posterior fault probability need not have their true states as 1, especially in the initial stages.

On the other hand, the AUC-based criterion under a single fault approximation chooses queries at each stage by taking into account its effect on all the objects, leading to a more balanced and informative choice of queries. This can be observed from the expressions of the estimators for area above the ROC curve in (10), where the object with the highest posterior fault probability $X_{r(1)}$ is assigned the least weight, with monotonically increasing weights as the posterior fault probability of the objects decreases. This forces choosing a query that takes into consideration the effect on all the objects.

Alternatively, the key difference between these two approaches can also be explained using an exploitation versus exploration terminology, where the entropy criterion chooses queries by exploiting the objects with high posterior

fault probabilities, whereas the AUC criterion chooses queries that explore in the early stages, but exploit in the later stages.

Though all three approximations for AUC are robust to violations of the single fault approximation, for reasons similar to the above and explained in detail in the supplemental material [31], AUC approximated using upper rectangles turns out to be a better choice for active diagnosis of multiple faults under a single fault approximation.

5.2 Single Fault Scenario

In this section, we will consider a special case of the active diagnosis problem where only one object can be in state 1. Note that this is different from the above scenario, where multiple faults could be present even though a single fault approximation is made. Here, the object state vector \mathbf{X} can only take values from the set $\{\mathbb{I}_1, \dots, \mathbb{I}_M\}$. This scenario arises in several applications, such as pool-based active learning, job scheduling, image processing, and computer vision. In applications where there can be no object in state 1, the above set can be modified to include an all zero vector.

Here, we will use some additional notation besides the ones defined in Section 2. We will use an $M \times N$ binary matrix \mathbf{B} to denote the bipartite relationship between objects and queries, where an entry b_{ij} in this matrix is 1 if there is an edge between object i and query j , and 0 otherwise. Also, we will use the vector $\Pi = (\pi_1, \dots, \pi_M)$ to denote the prior probability distribution of \mathbf{X} , where $\pi_i := \Pr(\mathbf{X} = \mathbb{I}_i)$ and $\sum_i \pi_i = 1$. Finally, in the single fault scenario, we will denote the event $\mathbf{X} = \mathbb{I}_i$ by $X_i = 1$.

We begin by noting that in this special scenario, the ROC curve corresponding to the rank-based estimators reduces to a step function. In particular, note that the miss rate of an estimator can only take two values in this case, either 0 or 1, as there is only one object whose true state is equal to 1. Hence, the ROC curve corresponding to a ranked list of objects is a step function, where the step corresponds to the location of the faulty object (object with state 1) in the ranked list. Thus, in this scenario, maximizing the AUC (or minimizing the area above the ROC curve) corresponding to a ranked list of objects is equivalent to minimizing the rank of the faulty object.

In fact, note from (10b) that in a single fault scenario, the estimate of the area above the ROC curve using a linear approximation corresponds to the expected rank of the faulty object in the ranked list. We will now introduce a slight variation on this criterion that offers a key advantage over entropy-based query selection in terms of not requiring knowledge of the underlying noise parameters.

Specifically, we replace the ranking $r(i)$ defined previously with a worst-case ranking. Given the observed responses \mathbf{z}_A to a set of queries in \mathcal{A} , the worst-case rank of object i is defined to be

$$\begin{aligned} r_{wc}(i|\mathbf{z}_A) &= \sum_{k=1}^M \mathbf{I}\{\Pr(X_k = 1|\mathbf{z}_A) \geq \Pr(X_i = 1|\mathbf{z}_A)\} \\ &= \sum_{k=1}^M \mathbf{I}\{\pi_k \Pr(\mathbf{z}_A|X_k = 1) \geq \pi_i \Pr(\mathbf{z}_A|X_i = 1)\}, \end{aligned}$$

	X_1	X_2	X_3	X_4	X_5
$\Pr(X_i = 1 \mathbf{z}_A)$	0.2	0.2	0.3	0.2	0.1
$r_{wc}(i \mathbf{z}_A)$	4	4	1	4	5

Fig. 4. Worst-case ranking.

$\mathbf{I}\{E\}$ being the indicator function. If the posterior probabilities are distinct, then $r_{wc}(i|\mathbf{z}_A)$ coincides with the previous definition. However, when multiple objects have the same posterior fault probabilities, each of those objects is assigned the worst-case rank, as shown in Fig. 4.

Given these rankings, the area above the ROC curve estimated using a linear approximation in (10b) can be upper bounded as shown below:

$$\begin{aligned} \bar{\mathbf{A}}_l(\mathbf{z}_A) &= \frac{1}{M-1} \sum_{i=1}^M i \cdot \Pr(X_{r(i)} = 1|\mathbf{z}_A) \\ &= \frac{1}{M-1} \sum_{i=1}^M r(i) \cdot \Pr(X_i = 1|\mathbf{z}_A) \\ &\leq \frac{1}{M-1} \sum_{i=1}^M r_{wc}(i|\mathbf{z}_A) \cdot \Pr(X_i = 1|\mathbf{z}_A) =: \bar{\bar{\mathbf{A}}}_l(\mathbf{z}_A). \end{aligned}$$

Substituting this in (8), we get the criterion for greedily choosing the next query to be

$$\begin{aligned} j^* &= \operatorname{argmin}_{j \notin \mathcal{A}} \sum_{z=0,1} \Pr(Z_j = z|\mathbf{z}_A) \bar{\bar{\mathbf{A}}}_l(\mathbf{z}_A \cup z) \\ &= \operatorname{argmin}_{j \notin \mathcal{A}} \frac{1}{M-1} \sum_{z=0,1} \sum_{i=1}^M [\Pr(Z_j = z|\mathbf{z}_A) \\ &\quad \Pr(X_i = 1|\mathbf{z}_A \cup z) r_{wc}(i|\mathbf{z}_A \cup z)] \\ &= \operatorname{argmin}_{j \notin \mathcal{A}} \frac{1}{M-1} \sum_{z=0,1} \sum_{i=1}^M \left[\frac{\pi_i \Pr(\mathbf{z}_A, z|X_i = 1)}{\Pr(\mathbf{z}_A)} r_{wc}(i|\mathbf{z}_A \cup z) \right] \\ &= \operatorname{argmin}_{j \notin \mathcal{A}} \sum_{z=0,1} \sum_{i=1}^M \pi_i \Pr(\mathbf{z}_A, z|X_i = 1) r_{wc}(i|\mathbf{z}_A \cup z), \end{aligned} \tag{12}$$

where (12) follows as $1/(M-1)$ and $\Pr(\mathbf{z}_A)$ are constants that do not depend on query j . In the noise-free case with uniform prior, this greedy strategy reduces to GBS [6], as shown in the supplemental material [31].

In the noisy case, given the knowledge of the prior distribution Π and the noise parameters such as the leak and the inhibition probabilities in the QMR-DT noise model, the greedy algorithm in (12) can be implemented efficiently. However, these noise parameters are often not known, and hence it is desirable for a greedy query selection criterion to be robust to any discrepancies in the knowledge of these parameters. As we show in Section 6, entropy-based active query selection can be sensitive to discrepancies in the noise parameters.

In the next two sections, we consider two special cases of the noise model discussed in Section 2 that appear in many applications, and present a noise independent estimate of the query selection criterion in (12). Specifically, we take advantage of the fact that this query selection criterion depends on the noise parameters only through the likelihood function, and provide a good upper bound on the

likelihood function that is independent of noise parameters. This enables accurate prediction of the worst-case rank of the objects without requiring knowledge of the true noise parameters. Furthermore, we show that in some cases it is possible to estimate the true ranks exactly with limited knowledge on the query noise. The bound on the likelihood function is based on the following lemma.

Lemma 1. *Let h, k be integers with $0 \leq h \leq k$ and $k \geq 1$. Then, for any $0 < p < 1$,*

$$p^h(1-p)^{k-h} \leq \varepsilon_h^h(1-\varepsilon_h)^{k-h}, \quad (13)$$

where $\varepsilon_h = \frac{h}{k}$. If it is known that $p \leq p_2 < 1$, then (13) holds with $\varepsilon_h = \min\{p_2, \frac{h}{k}\}$. If it is known that $p \geq p_1 > 0$, then (13) holds with $\varepsilon_h = \max\{p_1, \frac{h}{k}\}$. If it is known that $0 < p_1 \leq p \leq p_2 < 1$, then (13) holds with $\varepsilon_h = \min\{p_2, \max\{p_1, \frac{h}{k}\}\}$.

Proof. Refer to the supplemental material, which can be found online in the CSDL or at [31]. \square

5.2.1 Constant Noise Level

We begin with the following special case of the noise model described in Section 2, where $0 < \rho_{0j} = p < 1$, $\forall j$, and $\rho_{kj} = p/(1-p)$, $\forall k, j$. This noise model has been used in the context of pool-based active learning with a faulty oracle [7], [15], experimental design [21], computer vision, and image processing [9], where the responses to some queries are assumed to be randomly flipped.

In this setting, $\Pr(Z_j = z_j | X_i = 1) = p^{|b_{ij} - z_j|} (1-p)^{1-|b_{ij} - z_j|}$. More generally, the likelihood function can be expressed as shown below:

$$\Pr(\mathbf{Z}_A = \mathbf{z}_A | X_i = 1) = p^{\delta_{i,A}} (1-p)^{|\mathcal{A}| - \delta_{i,A}},$$

where $\delta_{i,A} = \sum_{j \in \mathcal{A}} |b_{ij} - z_j|$ is the local Hamming distance between the true responses of object i to queries in \mathcal{A} and the observed responses \mathbf{z}_A . Note that the bipartite relationship between the objects and queries or the binary matrix \mathbf{B} should be known to compute $\delta_{i,A}$.

Using the result in Lemma 1, the above likelihood function can then be upper bounded by

$$\overline{\Pr}(\mathbf{z}_A | X_i = 1) := \left(\frac{\delta_{i,A}}{|\mathcal{A}|} \right)^{\delta_{i,A}} \left(1 - \frac{\delta_{i,A}}{|\mathcal{A}|} \right)^{|\mathcal{A}| - \delta_{i,A}}.$$

The lemma also states that given an upper or lower bound on the noise parameter p , this bound can be further improved.

Finally, let $\bar{r}_{wc}(i | \mathbf{z}_A)$ denote the estimated worst-case rank of object i based on the upper bound on the likelihood function:

$$\bar{r}_{wc}(i | \mathbf{z}_A) := \sum_{j=1}^M \mathbf{I}\{\pi_j \overline{\Pr}(\mathbf{z}_A | X_j = 1) \geq \pi_i \overline{\Pr}(\mathbf{z}_A | X_i = 1)\}. \quad (14)$$

Then, the query selection criterion in (12) can be replaced by the following noise-independent criterion:

$$\operatorname{argmin}_{j \notin \mathcal{A}} \sum_{z=0,1} \sum_{i=1}^M \pi_i \overline{\Pr}(\mathbf{z}_A, z | X_i = 1) \bar{r}_{wc}(i | \mathbf{z}_A \cup z). \quad (15)$$

The result in Proposition 3 presents conditions under which the true rank can be estimated accurately. It states that, under uniform prior on the objects, it suffices to know whether $p < 0.5$ or $p > 0.5$, for the estimated ranks to be exactly equal to the true ranks.

More generally, for any given prior Π with $\rho := \min_i \pi_i / \max_i \pi_i$, it suffices to know whether $p < \frac{\rho}{1+\rho}$ or $p > \frac{1}{1+\rho}$ for the estimated ranks to be equal to the true ranks. Even in the case where $\frac{\rho}{1+\rho} \leq p < 0.5$ or $0.5 < p \leq \frac{1}{1+\rho}$, we observe through experiments that the estimated ranks are equal to the true ranks for most objects.

Proposition 3. *Let \mathbf{z}_A be the observed responses to a sequence of queries in \mathcal{A} , under some unknown noise parameter p . Let $\rho := \min_i \pi_i / \max_i \pi_i$. Given a $\bar{p} \in (0, \frac{\rho}{1+\rho})$ such that $0 < p \leq \bar{p}$ or a $\underline{p} \in (\frac{1}{1+\rho}, 1)$ such that $1 > p \geq \underline{p}$, the estimated ranks $\bar{r}_{wc}(i | \mathbf{z}_A)$ computed only with the knowledge of \bar{p} or \underline{p} are equal to the true ranks $r_{wc}(i | \mathbf{z}_A)$, $\forall 1 \leq i \leq M$.*

Proof. Refer to the supplemental material, which can be found online in the CSDL or at [31]. \square

5.2.2 Response Dependent Noise

We now consider the noise model where the probability of error depends on the true response to a query. When the true response is 0, the probability of observing a noisy response is given by ν_0 , and by ν_1 when the true response is 1, i.e.,

$$\Pr(Z_j = 0 | X_i = 1) = 1 - \nu_0, \text{ if } b_{ij} = 0, \\ \text{and } \Pr(Z_j = 0 | X_i = 1) = \nu_1, \text{ if } b_{ij} = 1.$$

For example, consider the following special case of the QMR-DT noise model described in Section 2, where $\rho_{0j} = \rho_0$, $\forall j$, and $\rho_{kj} = \rho$, $\forall k \neq 0, j$. This case reduces to the above setting with $\nu_0 = \rho_0$ and $\nu_1 = (1 - \rho_0)\rho$, where $0 < \rho_0, \rho < 1$ are the leak and inhibition probabilities, respectively.

For any subset of indices $\mathcal{A} \subseteq \{1, \dots, N\}$, let $\mathcal{A}_0^i = \{j \in \mathcal{A} : b_{ij} = 0\}$ and $\mathcal{A}_1^i = \{j \in \mathcal{A} : b_{ij} = 1\}$ be partitions of \mathcal{A} for each $i = 1, \dots, M$ such that the true response b_{ij} of object i to queries in \mathcal{A}_0^i is 0 and that in \mathcal{A}_1^i is 1. The likelihood function is then given by

$$\Pr(\mathbf{Z}_A = \mathbf{z}_A | X_i = 1) = \nu_0^{\delta_{i,\mathcal{A}_0^i}} (1 - \nu_0)^{|\mathcal{A}_0^i| - \delta_{i,\mathcal{A}_0^i}} \\ \times \nu_1^{\delta_{i,\mathcal{A}_1^i}} (1 - \nu_1)^{|\mathcal{A}_1^i| - \delta_{i,\mathcal{A}_1^i}},$$

where $\delta_{i,\mathcal{A}_0^i} = \sum_{j \in \mathcal{A}_0^i} |0 - z_j|$ and $\delta_{i,\mathcal{A}_1^i} = \sum_{j \in \mathcal{A}_1^i} |1 - z_j|$ are the local Hamming distances between the true responses of object i to queries in \mathcal{A}_0^i and \mathcal{A}_1^i and that of their observed responses.

Once again, using Lemma 1, this likelihood function can be upper bounded by

$$\overline{\Pr}(\mathbf{z}_A = \mathbf{z}_A | X_i = 1) = \left(1 - \frac{\delta_{i,\mathcal{A}_0^i}}{|\mathcal{A}_0^i|} \right)^{|\mathcal{A}_0^i| - \delta_{i,\mathcal{A}_0^i}} \\ \left(\frac{\delta_{i,\mathcal{A}_0^i}}{|\mathcal{A}_0^i|} \right)^{\delta_{i,\mathcal{A}_0^i}} \times \left(1 - \frac{\delta_{i,\mathcal{A}_1^i}}{|\mathcal{A}_1^i|} \right)^{|\mathcal{A}_1^i| - \delta_{i,\mathcal{A}_1^i}} \left(\frac{\delta_{i,\mathcal{A}_1^i}}{|\mathcal{A}_1^i|} \right)^{\delta_{i,\mathcal{A}_1^i}}.$$

Hence, the ranks of the objects can be estimated using (14) and the rank-based query selection can be performed

using (15), without requiring any knowledge of the query noise parameters.

Unfortunately, it is not possible to extend the result of Proposition 3 to this case. Yet, the experimental results in Section 6 demonstrate that the noise-independent rank-based algorithm performs comparably to the entropy-based algorithm, which requires knowledge of ν_0 and ν_1 .

6 EXPERIMENTS

We compare the performance of the proposed rank-based algorithm with that of entropy-based query selection and random search in both the multiple fault and the single fault scenarios. Random search serves as a baseline and is not expected to perform well as it selects queries at random. We compare their performance on two real diagnosis applications and several synthetic datasets. In particular, we consider the problem of fault diagnosis in computer networks, which is explained in more detail in Section 6.1. In the single fault scenario, we also consider the emergency response problem of toxic chemical identification, which is described in more detail in Section 6.2. In both cases, we also show experimental results on synthetic datasets generated using random network models such as the Erdős-Rényi (ER) and Preferential Attachment (PA) random network models. For more details on how the random networks and the computer networks were generated, refer to the supplemental material [31].

6.1 Multiple-Fault Scenario

We consider the problem of fault diagnosis in computer networks. In this application, the goal is to monitor a system of networked computers for faults, where each computer can be associated with a binary random variable X_i (0 for working and 1 for faulty). It is not possible to test each individual computer directly in a large network. Hence, a common solution is to test a subset of computers with a single test probe Z_j , where a probe can be as simple as a ping request or more sophisticated such as an e-mail message or a webpage access request. Thus, there is a BDG with each query (probe) connected to all the objects (computers) it passes through. In these networks, certain computers are designated as probe stations, which are instrumented in sending out probes to test the response of the networked elements. However, the available set of probes \mathbf{Z} is often very large, and hence, it is desirable to minimize the number of probes required to identify the faulty computers. Refer to Rish et al. [3] for further details.

We compare the performance of the proposed AUC criterion under a single fault approximation (AUC+SF) with that of the entropy criterion computed using BP (BPEA) and under a single fault approximation (Entropy+SF), along with random search. We have also implemented the AUC criterion using BP. In all our experiments, we observed its performance to be at least as good as BPEA, but not significantly better. Hence, we elected not to include them in the paper.

We compare the algorithms on one synthetic dataset and two computer networks. Unlike Zheng et al. [4] and Cheng et al. [26] who only considered networks of size up to 500 components and 580 probes, here we also consider a large

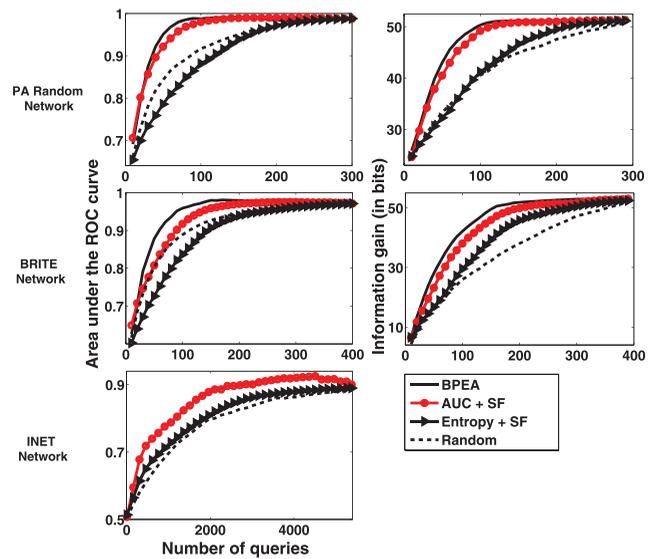


Fig. 5. The competitive performance of the AUC-based query selection under single fault approximation compared to that of BPEA while having a computational complexity that is orders less (near quadratic versus exponential complexity of BPEA). On INET, we compare AUC+SF only with Entropy+SF and random search as BPEA becomes slow and intractable.

scale network. The first dataset is a random BDG generated using the standard PA random graph model [32]. The second and the third datasets are network topologies built using the BRITE [33] and the INET [34] generators, which simulate an Internet-like topology at the Autonomous Systems level. To generate a BDG of components and probes from these topologies, we used the approach described by Rish et al. [3] and Zheng et al. [4].

For the random graph model considered, we generated a random BDG consisting of 300 objects and 300 queries. We generated a BRITE network consisting of 300 components and around 400 probes, and an INET network consisting of 4,000 components and 5,380 probes. We consider the QMR-DT noise model described in Section 2; parameters are given below. We compare the four query selection criteria under two performance measures, AUC and information gain.

Fig. 5 compares their performance as a function of the number of queries inputted. Information gain is computed using BPEA. To compute the AUC, we rank the objects based on their posterior fault probabilities that are computed using a single-fault approximation. Alternatively, note that these posterior probabilities could be computed using BP for the PA and BRITE networks (BP is intractable on the INET). Using this alternate approach does not change our conclusions.

We used the inference engines in the libDAI [35] package for implementing BPEA and BP. However, BPEA (and BP) became slow and intractable on the INET, with BP often not converging and resulting in oscillations. Hence, on this network, we only compare the performance of AUC+SF and Entropy+SF based on the AUC criterion, which is computed based on rankings obtained from posterior probabilities under a single-fault approximation.

The results in this figure correspond to a prior fault probability value of 0.03, with the leak and inhibition

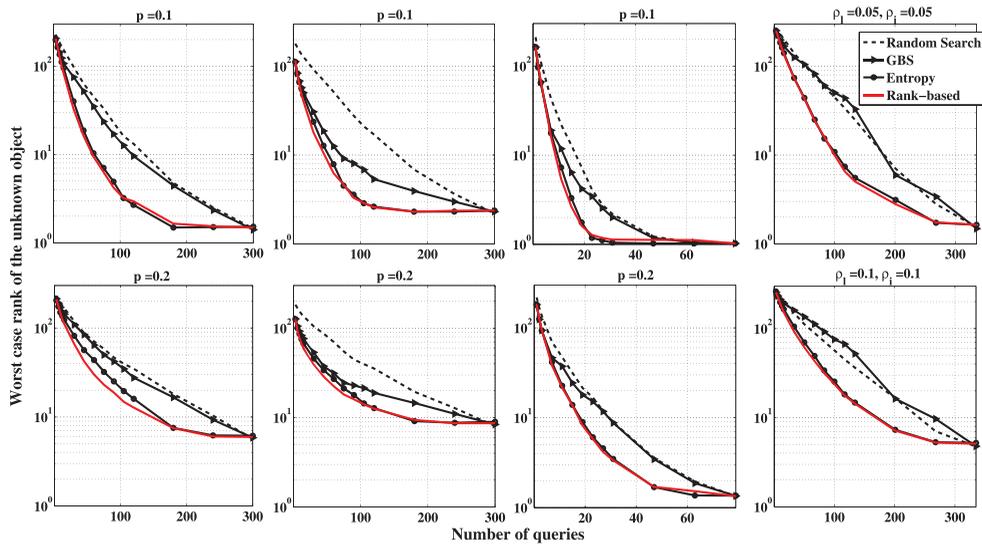


Fig. 6. The first column corresponds to a dataset generated using the ER model, the second column corresponds to a dataset generated using the PA model, the third column corresponds to the WISER database, and the last column corresponds to a BRITE network. In all the experiments, the rank-based algorithm has no knowledge of the noise parameters.

probabilities at 0.05. Each curve in this figure is averaged over 200 random realizations, where each random realization corresponds to a random state of \mathbf{X} and random generation of the noisy query responses. For the PA and BRITE models, the results were observed to be consistent across different realizations of the underlying bipartite network. For INET, we considered only one network with 25 probe stations.

Note from this figure that while BPEA performs the best, Entropy+SF performs as badly as a random search. On the other hand, AUC+SF performs significantly better than Entropy+SF, and comparable to BPEA while having a computational complexity that is orders less (near-quadratic versus exponential complexity of BPEA). We compared the performance of the methods for different values of prior fault, inhibition, and leak probabilities, as shown in the supplemental material [31]. We also observed that as the prior fault probability is increased beyond 0.1, the performance of AUC+SF starts to deteriorate and slowly converges to that of Entropy+SF, and this was observed irrespective of the size of the network. However, this should not be a problem as most real-world applications have a fairly low ($\ll 0.1$) prior fault probability.

6.2 Single-Fault Scenario

For the single fault scenario, we examine the fault diagnosis application discussed above, as well as the emergency response application of toxic chemical identification. In this context, the objects are toxic chemicals and the queries are symptoms. In the event of a toxic chemical accident, a first responder is often faced with the task of rapidly identifying the toxic chemical by posing symptom-based queries to a victim. Unfortunately, many symptoms tend to be non-specific (i.e., vomiting can be caused by many different chemicals), and it is therefore critical for the first responder to pose these questions in a sequence that leads to chemical identification in as few questions as possible.

We compare the performance of the proposed rank-based noise independent algorithm with that of entropy-

based query selection and random search on two synthetic datasets, a computer network, and a toxic chemical database. In addition, we also compare the proposed algorithm with that of GBS, as the proposed rank-based approach can be considered as an extension of GBS that is designed to handle query noise. Once again, GBS serves as a baseline, as it does not account for noise.

The first two datasets are random bipartite networks generated using the standard ER random network model and the PA random network model. The third dataset is a network topology built using the BRITE generator. The last dataset is the WISER database (<http://wiser.nlm.nih.gov>), which is a toxic chemical database describing the binary relation between 298 toxic chemicals and 79 acute symptoms.

We generated a random network for each of the random network models considered, where each network consisted of around 200 objects and 300 queries. We generated a BRITE network consisting of 300 objects (components/computers) and around 350 queries (probes). For the synthetic datasets and WISER, we assumed the noise model to be that of Section 5.2.1, and for the BRITE network, we considered the noise model in Section 5.2.2. Here, we present the results under uniform prior, where $\pi_i = 1/M$. We observed similar performance under nonuniform prior.

Fig. 6 shows the worst-case rank of the unknown object (i.e., the faulty computer or the leaked toxic chemical). Each curve in this figure is averaged over 500 random realizations, where each random realization corresponds to a random selection of the unknown object from the set of M objects and random generation of the noisy query responses. The resulting confidence intervals were very small and, hence, are not shown in the figure. For the two random network models and BRITE, the results were observed to be consistent across different realizations of the underlying bipartite network.

For the ER, PA, and the WISER datasets, we consider two different values for the probability of error, $p = 0.1, 0.2$. The entropy-based query selection is performed assuming the knowledge of p , whereas the rank-based query selection is

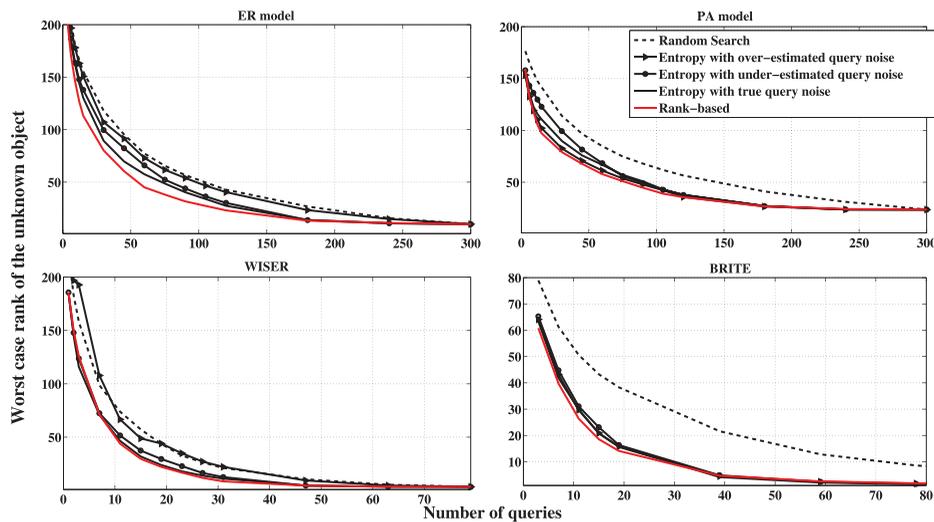


Fig. 7. The sensitivity of entropy-based query selection to misspecification of noise parameters.

performed using only the fact that $p < \bar{p} = 0.5$. The BRITE networks are simulated using the QMR-DT noise model, where we considered the leak and the inhibition probabilities to be $(\rho, \rho_0) = (0.05, 0.05)$ and $(0.1, 0.1)$. This noise model reduces to that in Section 5.2.2 with $\nu_0 = \rho_0$ and $\nu_1 = (1 - \rho_0)\rho$. Once again, the entropy-based query selection is performed assuming the knowledge of ν_0 and ν_1 , whereas the rank-based query selection is performed using only the fact that $\nu_0, \nu_1 \leq \bar{p} = 0.25$.

Finally, Fig. 7 demonstrates the sensitivity of entropy-based query selection to misspecification of the value of noise parameters. For the ER, PA, and WISER datasets, the true noise parameter is $p = 0.25$, while the underestimated and the overestimated curves are obtained using $p = 0.15$ and 0.4 , respectively. For the BRITE network, while the true noise parameters are $(0.1, 0.1)$, the other two curves are obtained using $(0.05, 0.05)$ and $(0.15, 0.15)$. Once again, the rank-based algorithm is performed without knowledge of the noise parameters. This demonstrates that the entropy-based query selection can perform poorly when the noise parameters are misspecified.

These experiments demonstrate the competitive performance of the rank-based algorithm to entropy-based query selection, despite not having knowledge of the underlying noise parameters.

7 CONCLUSIONS

We study active diagnosis in two different settings—multiple fault and single fault. In the multiple-fault scenario, active query selection algorithms such as BPEA rely on belief propagation, making them intractable in large networks. Thus, we propose to make the simplifying approximation of a single fault in the query selection stage. Under this approximation, several query selection criterion can be implemented efficiently. However, we note that traditional approaches such as entropy-based query selection under a single fault approximation perform poorly in a multiple fault setting. Hence, we propose a new query selection criterion where the queries are selected sequentially such that the AUC of a rank-based output is maximized. We demonstrate

the competitive performance of the proposed algorithm to BPEA in the context of fault diagnosis in computer networks. The competitive performance of the proposed algorithm, while having a computational complexity that is orders less than that of BPEA (near quadratic versus the exponential complexity of BPEA), makes it a fast and a reliable substitute for BPEA in large scale diagnosis problems. Furthermore, we show that the proposed criterion has another interesting feature in the single fault scenario in that it does not require knowledge of the underlying noise distribution. On the other hand, entropy-based query selection requires knowledge of these noise parameters, and can be sensitive to misspecification of these values.

ACKNOWLEDGMENTS

This work was supported in part by US National Science Foundation (NSF) Awards No. 0830490 and 0953135, and CDC/NIOSH Grant No. R21 OH009441-01A.

REFERENCES

- [1] T.S. Jaakkola and M.I. Jordan, "Variational Methods and the QMR-DT Databases," *J. Artificial Intelligence Research*, vol. 10, pp. 291-322, 1999.
- [2] N.I. Santoso, C. Darken, G. Povh, and J. Erdmann, "Nuclear Plant Fault Diagnosis Using Probabilistic Reasoning," *Proc. IEEE Power Eng. Soc. Meeting*, vol. 2, pp. 714-719, 1999.
- [3] I. Rish, M. Brodie, S. Ma, N. Odintsova, A. Beygelzimer, G. Grabarnik, and K. Hernandez, "Adaptive Diagnosis in Distributed Systems," *IEEE Trans. Neural Networks*, vol. 16, no. 5, pp. 1088-1109, Sept. 2005.
- [4] A.X. Zheng, I. Rish, and A. Beygelzimer, "Efficient Test Selection in Active Diagnosis via Entropy Approximation," *Proc. Int'l Conf. Uncertainty in Artificial Intelligence*, 2005.
- [5] Z. Yongli, H. Limin, and L. Jinling, "Bayesian Networks Based Approach for Power Systems Fault Diagnosis," *IEEE Trans. Power Delivery*, vol. 21, no. 2, pp. 634-639, Apr. 2006.
- [6] S. Dasgupta, "Analysis of a Greedy Active Learning Strategy," *Proc. Advances in Neural Information Processing Systems*, 2004.
- [7] S. Hanneke, "Teaching Dimension and the Complexity of Active Learning," *Proc. 20th Conf. Learning Theory*, 2007.
- [8] S.K. Bhavnani, A. Abraham, C. Demeniuk, M. Gebrekristos, A. Gong, S. Nainwal, G. Vallabha, and R. Richardson, "Network Analysis of Toxic Chemicals and Symptoms: Implications for Designing First-Responder Systems," *Proc. Am. Medical Informatics Assoc. Ann. Symp.*, 2007.

- [9] A.P. Korostelev and J.C. Kim, "Rates of Convergence of the Sup-Norm Risk in Image Models under Sequential Designs," *Statistics and Probability Letters*, vol. 46, pp. 391-399, 2000.
- [10] D. Geman and B. Jedynak, "An Active Testing Model for Tracking Roads in Satellite Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 1, pp. 1-14, Jan. 1996.
- [11] S.R. Kosaraju, T.M. Przytycka, and R.S. Borgstrom, "On an Optimal Split Tree Problem," *Proc. Sixth Int'l Workshop Algorithms and Data Structures*, pp. 11-14, 1999.
- [12] A. Gupta, R. Krishnaswamy, V. Nagarajan, and R. Ravi, "Approximation Algorithms for Optimal Decision Trees and Adaptive TSP Problems," *Proc. 37th Int'l Colloquium Conf. Automata, Languages and Programming*, 2010.
- [13] K.P. Murphy, Y. Weiss, and M. Jordan, "Loopy Belief Propagation for Approximate Inference: An Empirical Study," *Proc. Int'l Conf. Uncertainty in Artificial Intelligence*, pp. 467-475, 1999.
- [14] M. Kärräinen, "Active Learning in the Non-Realizable Case," *Proc. 17th Int'l Conf. Algorithmic Learning Theory*, pp. 63-77, 2006.
- [15] R. Nowak, "Noisy Generalized Binary Search," *Proc. Advances in Neural Information Processing Systems 21*, 2009.
- [16] G. Bellala, S.K. Bhavnani, and C. Scott, "Active Diagnosis under Persistent Noise with Unknown Noise Distribution: A Rank-Based Approach," *Proc. 14th Int'l Conf. Artificial Intelligence and Statistics*, 2011.
- [17] G. Bellala, J. Stanley, C. Scott, and S.K. Bhavnani, "Active Diagnosis via AUC Maximization: An Efficient Approach for Multiple Fault Identification in Large Scale, Noisy Networks," *Proc. 27th Int'l Conf. Uncertainty in Artificial Intelligence*, 2011.
- [18] S. Kandula, D. Katabi, and J.P. Vasseur, "Shrink: A Tool for Failure Diagnosis in IP Networks," *Proc. ACM SIGCOMM MineNet Workshop*, Aug. 2005.
- [19] D.W. Loveland, "Performance Bounds for Binary Testing with Arbitrary Weights," *Acta Informatica*, vol. 22, pp. 101-114, 1985.
- [20] L. Hyafil and R. Rivest, "Constructing Optimal Binary Decision Trees Is NP-Complete," *Information Processing Letters*, vol. 5, no. 1, pp. 15-17, 1976.
- [21] A. Rényi, "On a Problem of Information Theory," *MTA Mat. Kut. Int. Kozl.*, vol. 6B, pp. 505-516, 1961.
- [22] G. Bellala, S.K. Bhavnani, and C. Scott, "Group-Based Active Query Selection for Rapid Diagnosis in Time-Critical Situations," *IEEE Trans. Information Theory*, vol. 58, no. 1, pp. 459-478, Jan. 2012.
- [23] D. Golovin, D. Ray, and A. Krause, "Near-Optimal Bayesian Active Learning with Noisy Observations," *Proc. Advances in Neural Information Processing Systems 23*, 2010.
- [24] J. Pearl, *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.
- [25] T. Le and C.N. Hadjicostis, "Max-Product Algorithms for the Generalized Multiple-Fault Diagnosis Problem," *IEEE Trans. Systems, Man, and Cybernetics*, vol. 37, no. 6, pp. 1607-1621, Dec. 2007.
- [26] L. Cheng, X. Qui, L. Meng, Y. Qiao, and R. Boutaba, "Efficient Active Probing for Fault Diagnosis in Large Scale and Noisy Networks," *Proc. IEEE INFOCOM*, 2010.
- [27] C. Cortes and M. Mohri, "AUC Optimization versus Error Rate Minimization," *Proc. Advances in Neural Information Processing Systems 15*, 2003.
- [28] P.M. Long and R.A. Servedio, "Boosting the Area under the ROC Curve," *Proc. Advances in Neural Information Processing Systems 19*, 2007.
- [29] K. Ataman, W.N. Street, and Y. Zhang, "Learning to Rank by Maximizing AUC with Linear Programming," *Proc. IEEE Int'l Joint Conf. Neural Networks*, pp. 123-129, 2006.
- [30] M. Culver, K. Deng, and S. Scott, "Active Learning to Maximize Area under the ROC Curve," *Proc. Sixth Int'l Conf. Data Mining*, 2006.
- [31] "Supplementary Material," <http://web.eecs.umich.edu/~cscott/pubs.html>, 2012.
- [32] J. Guillaume and M. Latapy, *Bipartite Graphs as Models of Complex Networks*. Springer, 2004.
- [33] A. Medina, A. Lakhina, I. Matta, and J. Byers, "BRIT: An Approach to Universal Topology Generation," *Proc. Ninth Int'l Symp. Modeling, Analysis and Simulation of Computer and Telecomm. Systems*, 2001.
- [34] J. Winick and S. Jamin, "INET-3.0: Internet Topology Generator," Technical Report CSE-TR-456-02, Univ. of Michigan, 2002.

- [35] J.M. Mooij, "libDAI: A Free and Open Source C++ Library for Discrete Approximate Inference in Graphical Models," *J. Machine Learning Research*, vol. 11, pp. 2169-2173, 2010.



Gowtham Bellala received the BTech degree in electrical engineering from the Indian Institute of Technology, Madras, in 2006, and the MS and PhD degrees in electrical engineering and computer science from the University of Michigan, Ann Arbor, in 2008 and 2012, respectively. He is currently a research scientist at Hewlett Packard Labs in Palo Alto, California. His research interests include pattern recognition, data mining, machine learning, and their applications. He is a member of the IEEE.



Jason Stanley received the BSE degree in computer science from the University of Michigan, Ann Arbor, in 2012. As an undergraduate, he worked as a research assistant in the Electrical Engineering and Computer Science Department. He is currently with the Citadel Investment Group. His research interests include computer science theory and machine learning.



Suresh K. Bhavnani received the PhD degree in computational design and human-computer interaction from Carnegie Mellon University, Pittsburgh, Pennsylvania, in 1998. He is currently an associate professor of biomedical informatics in the Institute for Translational Sciences (ITS), University of Texas Medical Branch, and holds a secondary appointment with the Department of Preventive Medicine and Community Health, and an adjunct associate professor appointment with the School of Biomedical Informatics, University of Texas Houston. His research interests include network visualization and analysis of biomedical data, with translation to the design of decision-support systems.



Clayton Scott received the AB degree in mathematics from Harvard University in 1998, and the MS and PhD degrees in electrical engineering from Rice University in 2000 and 2004, respectively. He is currently an associate professor in the Departments of Electrical Engineering and Computer Science and of Statistics at the University of Michigan, Ann Arbor. His research interests include statistical machine learning theory, algorithms, and applications.

► For more information on this or any other computing topic, please visit our Digital Library at www.computer.org/publications/dlib.