Using Domain Knowledge to Constrain Structure Learning in a Bayesian Bioagent Detector

Anshu Saksena, Dennis Lucarelli, I-Jeng Wang
The Johns Hopkins University
Applied Physics Laboratory
Laurel, MD 20723
E-mail: anshu.saksena@jhuapl.edu

Abstract— A novel procedure for learning a probabilistic model from mass spectrometry data that accounts for domain specific noise and mitigates the complexity of Bayesian structure learning is presented. We evaluate the algorithm by applying the learned probabilistic model to microorganism detection from mass spectrometry data.

I. INTRODUCTION

Recent advances in high-throughput laboratory procedures in the life sciences are beginning to produce data sets that are amenable to investigation with methods from data mining and machine learning. Biological data sets, however, are often noisy and very sparse, thus prompting researchers to craft new learning algorithms to address these limitations. Of particular interest is the prospect of using mass spectrometry as a tool for identification, prediction and diagnosis [1], [2]. Some of the enthusiasm surrounding mass spectrometry can be attributed to its possible impact on *biomarker profiling* — the process of determining the discriminative or functional molecular species, or *biomarkers*, in biological processes. This is a topic of considerable interest in this post-genomic era of the life sciences.

In this paper, we describe an algorithm to learn a joint probability distribution over a set of correlated random variables modeling the underlying physical processes in time-of-flight mass spectrometry. A procedure for incorporating domain knowledge as a second source of data into the learning algorithm is presented as a way to mitigate the complexity of the learning algorithm and to counter a significant source of noise in biomarker identification. After training on a particular microorganism, the algorithm produces a probabilistic model of the ensemble of mass spectrometry experiments.

As an initial application, we assessed its effectiveness in detecting the organism *Bacillus globigii* from mass spectrometry data. Exploiting the qualitative features of the learned probabilistic model for possible use in biomarker profiling is touched upon.

II. TARGET APPLICATION

Mass spectrometry (MS) is a highly sensitive method of empirical chemical identification. One of the many variants of MS is *time-of-flight* (TOF) MS in which a sample is ionized and accelerated in an electric field. The accelerated ions then travel through a drift region of constant potential, at the end of which they strike a detector that measures the flux of ions as a function of time. Since the force F applied to each ion in this field is proportional to its charge q by $F = qE$, the more charge an ion has, the faster it will be moving through the drift region and the earlier its arrival time. Also, since acceleration a is inversely proportional to mass m by Newton's second law of motion $F = ma$, the heavier an ion is, the slower it will travel and the later it will arrive at the detector. Assuming (incorrectly) that all the molecules are ionized simultaneously at the same position and that at that time they have zero velocity, it follows by integrating the acceleration twice with respect to time that their arrival times t at the detector d go as the square root of their mass-to-charge ratios: $t = \sqrt{\frac{2d}{E}\frac{m}{q}}$. After calibration, the arrival times can be converted into mass-to-charge ratio and the resulting mass spectrum consisting of the flux time series provides data from which the masses of the molecules of interest can be inferred.

Ionization of the sample is usually done by laser. However, the laser often causes the molecules of interest to break down. The spectra usually consist of a combination of original molecules and their various fragments. For small and irregularly structured molecules, the fragments can shed light on the structure of the original molecules. For proteins, which are long chains of the same 23 basic amino acids, the fragments are usually smaller chains of amino acids and the masses of the fragments do not uniquely determine the mass of the original protein. Additionally, microorganism identification by MS is further complicated by the presence of several proteins in the sample. Therefore, microorganism identification by MS is typically simpler when the proteins are intact.

Matrix-assisted laser desorption and ionization (MALDI) TOF MS does this by embedding the sample among smaller molecules of a compound called a *matrix*. These molecules absorb the laser energy very efficiently and are lifted off the surface of the sample, carrying the intact proteins, or *monomers*, with them. In that plume of ions and proteins, charges are exchanged and the intact proteins are ionized.

Different microorganisms have both common and unique proteins. The proteins that are unique to a species, called *biomarkers*, can be used as the basis for discriminative identification of that species. An appropriately processed sample of microorganism may have a characteristic signature in its mass
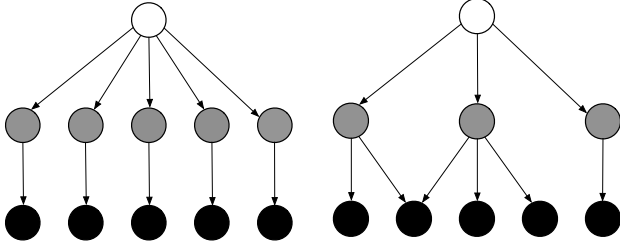
Fig. 1. BBN Structures

is posed as an optimization problem of the likelihood of the data $\mathcal{D}$ given the underlying graph,

$$max_{\mathcal{G}} P(\mathcal{D} \,|\, \mathcal{G}). \tag{2}$$

Under certain assumptions [6] (the most problematic of which we will describe below), the likelihood admits a local decomposition given by

$$P(\mathcal{D} \,|\, \mathcal{G}) \propto \prod_{i}^{n} score(X_i, \Pi(X_i), \mathcal{D}) \tag{3}$$

where $score(X_i, \Pi(X_i), \mathcal{D})$ is a function of the counts occurring in the data. Namely, under the assumption that the data is sampled from a multinomial distribution and the prior distribution on the parameters is the Dirichlet distribution, the local score can be expressed as

$$score(X_i, \Pi(X_i), \mathcal{D}) =$$
$$\rho(X_i, \Pi(X_i)) \cdot \prod_{j=1}^{q_i} \frac{\Gamma(N'_{ij})}{\Gamma(N_{ij} + N'_{ij})} \prod_{k=1}^{r_i} \frac{\Gamma(N_{ijk} + N'_{ijk})}{\Gamma(N'_{ijk})} \tag{4}$$

where $\Gamma(\cdot)$ is the gamma function, q_i is the number of configurations of the parents of X_i, r_i is the number of values of X_i, N_{ijk} is the number of occurrences of kth state of X_i with the jth configuration of its parents, N_{ij} is the number of occurrences of the jth configuration of the parents of X_i, and N'_{ijk}, N'_{ij} denote the parameters of the Dirichlet distribution. Finally, $\rho(X_i, \Pi(X_i))$ is a modular prior over structures, usually chosen either uniformly or to penalize network complexity. This score is the known as the *Bayesian Direchlet* (BD) score and is used in our experiments with a strong prior over structures that we describe in the following.

Thus under these assumptions, given a graph, it is relatively efficient to evaluate the function to be maximized; however, the size of graph space prevents a global optimization without imposing further constraints. Therefore it is standard to pursue a hill climbing procedure by making *local* changes to the network and evaluating the affected factors in the expression (3) until settling into a local maximum.

The derivation leading to the above expressions relies on the assumption that the data has no hidden variables or missing values. This rather severe restriction can be circumvented by a procedure known as the *structural expectation maximization* (SEM) algorithm [7]. SEM is an iterative method that uses the current fully specified BBN to perform inference over the hidden (or missing) variables to complete the data. With a fully instantiated data set, a structure learning routine is performed inside the loop. Parameters are then learned for the highest scoring structure and the process repeats until convergence. While this algorithm helps to alleviate the full database assumption, it is a computationally demanding procedure that puts the globally intractable problem of learning the structure *inside* the loop. The algorithm described in the following was designed to limit this search through graph space by allowing only those graphs that conform to our understanding of the problem domain. The fusion of domain knowledge with data to

spectrum. However, the spectra are extremely sensitive to the environment in which the organism grew and its specific intra-species genetic and expression variation, resulting in highly noisy and broadly varying spectra among different samples of the same species, making organism identification by mass spectrometry very difficult. However, a detector based on a probabilistic representation of this variation may have hope. We use a Bayesian belief network as our detector, which we train on data taking advantage of our knowledge of the properties of this application.

III. BAYESIAN BELIEF NETWORKS

Bayesian belief networks (BBN) provide a compact representation of the joint distribution over high dimensional spaces. Formally, a BBN structure is defined by a directed acyclic graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, representing the space of random variables $(\mathcal{V})$ and the conditional independencies among them $(\mathcal{E})$. Two example structures are shown in Figure 1. For a random variable, X, the set of *parents* of X, denoted $\Pi(X)$, is defined by the set of variables joined via directed edges terminating in X. A directed edge joining two random variables in the network implies a possible conditional dependency between the two variables.[1] The joint distribution over the space of random variables is given by

$$P(\mathcal{V}) = \prod_{i}^{n} P(X_i | \Pi(X_i)). \tag{1}$$

Each element in the product (1) is specified by a conditional probability table (CPT) defined at each variable in the network. In the following, we will refer to the CPT's generically as the *parameters* of the model. The innovation of the BBN formalism is that the additional information provided by the structural graph lends itself to inferencing and learning algorithms. For example, the *clique tree* algorithm [3] is an exact method for computing marginal probabilities, which is efficient for small to medium size networks. Efficient variational methods of inferencing adapted from statistical physics [4], [5] can be employed for larger sized networks.

In this paper, we are interested in learning the dependency structure (i.e., $\mathcal{E}$) of the network given data [6]. The problem

---

[1] More precisely, the structure encodes conditional *independence* statements by asserting that a random variable is independent of its non-descendants given its parents.

enhance machine learning is a topic that has recently received attention in the literature [10].

IV. BAYESIAN DETECTOR

A Bayesian detector contains the relevant random variables — specifically, variables for the observables and variables for the queries of interest — and their conditional dependencies [9]. The dependencies in the MALDI process that we capture in our Bayesian detector are as follows. Obviously, the presence or absence of the organism in question affects the likelihood of presence of the biomarker proteins. By definition, the likelihood of their presence is low if the organism is not present. Similarly, the presence or absence of the different biomarkers affects the likelihood of presence of the different ions of the protein. The presence or absence of each ion affects the likelihood of seeing a peak at the corresponding mass-to-charge ratio in the mass spectrum.

These fundamental physical dependencies translate directly into generic conditional dependencies in the Bayesian network. The root node of our BBN (i.e., the unique node that has no parents), which we call the *detection node*, represents the presence or absence of the *bioagent*, or organism, we are trying to detect. The state of this node is visible during training, and hidden when the BBN is being used to perform detection. We have hidden nodes representing the presence or absence of biomarker proteins of the organism in question. We will also call these *protein* nodes. The fact that they are hidden means that we do not need specific knowledge *a priori* of what these biomarkers should be. Finally, we have nodes that represent the presence of peaks at particular masses that are determined by inspection of the training spectra. To simplify learning, we chose not to have random variables in our BBN represent the presence of the actual ion variant molecules in the MALDI plume, since these would be hidden as well, increasing the computational complexity of the learning process by many orders of magnitude. All nodes in our BBN are binary. The peak presence nodes, which we will interchangably call *mass nodes* and *peak nodes*, are conditionally dependent only on some subset of the hidden protein nodes. The protein nodes are conditionally dependent on the detection node. The result is a 3-layer Bayesian network, an example of which is illustrated in Figure 1 and whose joint probability distribution factors as follows:

$$P(\mathcal{V}) = P(\mathcal{B}) \cdot \prod_i P(\mathcal{M}_i | \vec{\mathcal{P}}) \prod_j P(\mathcal{P}_j | \mathcal{B}) \qquad (5)$$

where $\mathcal{B}$ is the detection node, the $\mathcal{M}_i$ are the mass nodes, $\vec{\mathcal{P}}$ is the set of all the protein nodes, and $\mathcal{P}_j$ is one particular protein node. To perform detection, each spectrum is processed to get absent/present values for each mass node. The nodes are set accordingly, inference is performed, and the probability that the detection node is in its bioagent-present state is thresholded to give the detector result.

The above description of conditional dependencies is overly generic in that each mass node in the BBN should depend only
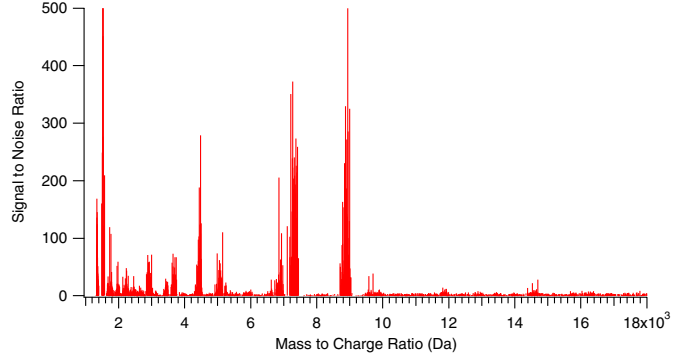

Fig. 2. Aggregated Processed Spectra of Bg

on some particular subset of the protein nodes. The specific dependencies must be learned from the data. This structure learning problem does not scale well with the number of mass nodes in the BBN. Moreover, the training data typically available does not substantiate the volume of performance comparisons that have to be done — the statistical significance of the result of the learning would be questionable. The following two sections describe our approach to help alleviate this problem.

V. DOMAIN KNOWLEDGE

In the MALDI plume, ions can transfer charge to one another, creating the possibility of multiply charged species. Also, ions can merge together to form a larger ion. It has been shown that there is a correlation among the frequency of peaks of these *ion variants* that are rational multiples of the mass of a protein [8]. Feldman *et al.* showed that when the equine cytochrome C protein is analyzed with sinapinic acid as the matrix, some of the strongest correlations occur among peaks of the monomer ion, the doubly-charged monomer ion, and a singly-charged ion consisting of two proteins, called a dimer. Figure 2 shows a sample mass spectrum in which candidates for each of these species can be found (for example, there is a peak at mass 9000 Da as well as at half that mass). Since the appearance of these rationally related masses are correlated and since correlation implies dependence, the structure of the Bayesian detector must not impose conditional independence among these variables. We assume that all other pairs of variables representing peak presence are independent given the state of the detection node. Since the Bayesian detector described in the previous section assumes that all protein nodes are hidden, it follows that the children of each protein node are dependent as long as the protein node itself is not deterministic given the detection node. Therefore, the desired dependency structure is obtained by having each protein node only have its monomer, dimer, and doubly-charged monomer as potential children.

We say potential children because there will be instances in which a peak could be the dimer of another or be the doubly-charged of a third or both, in addition to itself being an independent monomer related or unrelated to any other

peaks. We allow structures that have any combination of these protein-peak mappings and we rely on model selection via SEM [7] limited to those structures to determine the most likely mapping given the data. Note that our approach does not take advantage of genetic or proteomic information of the organism being detected. When that information is available, it would reduce the search space of structures, but disambiguation of protein-peak mappings could still be necessary.

Assuming there are the same number of protein nodes as there are mass nodes, the total number of possible structures before constraining the search space as above is of order 2^{n^2}, super exponential in n, the number of mass nodes. Our domain constraint limits the number of candidate children of a protein node to 3, reducing the size of the search space to exponential in n.

This physically motivated constraint not only significantly reduces the complexity of selecting a model, but it also biases the learning to models that are more likely to represent the distribution of the data based on this domain knowledge. We believe this will make the detector less likely to overfit the training data. Additionally, because of the dependency between peaks, the learned structure should be more robust to data representing a mixture of different microorganisms than a *naive Bayes* structure, which does not capture any inter-peak dependencies. See the left-hand structure in Figure 1 for an example of a naive Bayes network.

In addition to the domain-driven constraint, we impose additional constraints to avoid including multiple equivalent BBNs in the model selection search and to constrain the parameters of the selected model as much as possible. Equivalent BBNs in this context are those whose structure is the same up to permutation of the hidden nodes. Since hidden nodes are only identifiable by their topological neighbors, different protein nodes parenting the same mass nodes are equivalent. Therefore, we identify protein nodes by imposing the constraint that if a protein node has any children, one of them must be a monomer, and we disallow the possibility of multiple monomers sharing the same mass.

Whenever there are hidden nodes, the parameters have more degrees of freedom than any training set can constrain. Parameter learning algorithms must implicitly or explicitly constrain the parameters in some way in order to converge to a solution. We perform parameter learning using the expectation-maximization algorithm [12], which iterates through the following to converge on a local maximum:

$$\theta_{n+1} = argmax_\theta \sum_{\mathcal{H}} \log P(\mathcal{O}, \mathcal{H}|\theta) P(\mathcal{H}|\mathcal{O}, \theta_n) \quad (6)$$

where θ_n is the value of the parameters at the n-th iteration of the algorithm, $\mathcal{H}$ is the set of hidden nodes, $\mathcal{O}$ is the set of visible or observed nodes, and the sum is taken over all combinations of values of the hidden nodes.

We also disallow structures in which all the children of a protein node have other proteins as parents as well to prevent the parameters from being further underconstrained. In the following section we detail the implementation of these constraints in learning the BBN.

VI. INCORPORATION OF DOMAIN KNOWLEDGE INTO CONSTRAINED STRUCTURE LEARNING

As mentioned above, MALDI-TOF mass spectra have many sources of noise that serve to broaden peaks and shift their times of arrival significantly enough to require that we consider some window width within which any energy will be considered contributory towards a peak of that mass. This means that multiple mass windows can be candidates for a double- or half-mass window of any one mass. A mass window is a candidate for being a double-mass ion of another if when the endpoint masses of the latter are doubled, the resulting doubled window overlaps with the former. A mass window is a candidate for being a half-mass ion of another if the latter is a double-mass ion of the former. We want to consider all BBN structures that are consistent with any of the double- or half-mass assignments that conform with these overlaps. Note that the simplest such structure maps each protein to exactly 1 peak with no dependencies on other peaks, making every peak represent a monomer. In this case, the hidden protein nodes are superfluous and the structure is equivalent to the naive Bayes structure.

In order to do this, we first create a data structure that encapsulates all of these candidate double- and half-mass pairings. We create a directed acyclic graph (DAG) whose nodes are the mass nodes of interest. The DAG has a directed edge from mass node A to mass node B if mass node B is a candidate for being a double-mass ion of mass node A, as shown in Figure 3. It is easy to see that there cannot be any cycles in this construction. Now all candidate sets of mass nodes that can be the child sets of protein nodes by our constraint are exactly all combinations of distinct paths of length 0 (consisting of a single node), 1, or 2 in the DAG.

We call the process of constructing the DAG clustering, because when this is done for all the potential mass windows of interest, the result may be several disjoint DAGs, or *clusters*. The choice of protein parent in the BBN for mass nodes in one cluster is independent of that for mass nodes in a different cluster. Therefore, the joint probability density function of the target BBNs factor, and a globally maximal *a posteriori* structure is the union of the maximum likelihood structures of the nodes within each cluster determined independently of each other. Within each iteration of our SEM applied to each cluster, we exhaustively evaluate every structure consistent with that cluster. Therefore, the complexity of the search is dominated by the size of the largest cluster, not by the number of peak nodes or the number of clusters, and on average the size of the largest cluster cannot grow faster than linearly with respect to the number of peak nodes.

VII. RESULTS AND CONCLUSION

We have encouraging results on both simulated and real data. We focus on only one cluster in our results for simulated
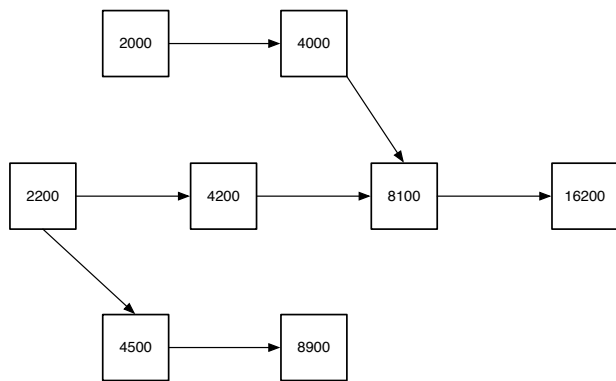
Fig. 3.   Example DAG

data. This data was created by first randomly constructing a DAG that is consistent with intersections of intervals with their doubles. From this DAG we randomly chose one of the BBN structures that can be derived from it as described above. Parameters for this BBN are drawn from a uniform distribution. This BBN is then sampled to create a data set of sufficient size to provide a reasonable estimate of the distribution represented. For many different randomly generated DAGs of sizes ranging from 4 mass nodes to 8, our structure learning technique learned the generating structure consistently for small DAGs and about 50% of the time for the larger DAGs. Although the simulated data was probably sufficient to learn the parameters for each structure, it may not have been sufficient for the multiple comparisons required to determine the maximum likelihood structure to be significant.

We also applied the learning algorithm to a data set consisting of over 50,000 positive spectra of *B. globigii* (Bg) and the same number of negative spectra taken from outdoor air samples. 420 individual, unaveraged spectra were collected from different locations on each of 240 samples and processed to highlight the relevant peaks and deemphasize typical MALDI matrix noise [11]. Mass windows were determined from the positive spectra based on regions of visible concentrations of peaks. Clustering is done as described above with these mass windows. The data were shuffled and half used for training and half for testing. This was repeated many times to mitigate sensitivity to different partitions of this data set. In testing, detection was inferred from the ensemble of all 420 spectra from each sample to take advantage of the correlations in the learned structure.

We compared the performance of the learned structure as a detector with a structure that did not benefit from the incorporation of domain knowledge, but was trained on the same mass spectrometry data. The resulting structure resembles a naive Bayes detector with a layer of hidden nodes as depicted in the left panel of Figure 1. Given that the learned structure represents a dependency structure inspired by both the mass spectrometry data and domain knowledge, we expected it to have better separation in output probability

between when the bioagent is present and when it is not. Indeed, the learned structure, maximizing the BD score [6] during training, resembled the naive Bayes structure in only about 10% of cases, indicating that in 90% of the cases, the learned structure scored better than the naive Bayes.

However, since test data is available, perhaps a more pertinent score would be based on the ROC curve of the detector on the test data. To compare the two structures, we used a weighted area under the ROC curve, which we call the ROC score, that places more importance on regions of the ROC curve that correspond to low false alarm rates. Each of the 50 reshuffles of the data we performed generated a new ROC score for each cluster for the naive Bayes structure and for the structure learned from that shuffle of the data. For each cluster, we wanted to evaluate whether there was a difference between these two sets of 50 observations that was not likely to be caused by random noise. We performed the Mann-Whitney test, a non-parametric statistical hypothesis test designed to decide whether the null hypothesis that the two sets of observations are drawn from the same unspecified distribution is true versus the alternative that their distributions are different. In 2 of the 5 clusters, we can conclude with 95% confidence that the learned structure outperformed the naive Bayes structure as a detector, whereas the null could not be rejected for the other 3 clusters. In one of the 3 indeterminate clusters, while it is not statistically significant, the learned structure was the better performing structure more often than the naive structure. We conclude that the learned structure performed as well as or better than the naive Bayes structure as a detector, with improved performance observed in 3 out of 5 clusters.

We have a couple of possible explanations for why performance was not improved for all clusters. First of all, the naive Bayes structure performed rather well with this data, and with the exception of a couple of possibly mislabelled spectra, there was not much room for improvement. The raw spectra were heavily processed as described in [11] to remove most of the MALDI noise and make peaks discernable by inspection. It is very possible that this processing itself makes the problem easy enough to not require the second-order features being pursued here. In fact, the results presented here are after one step, the averaging together of about 420 spectra at a time, was removed from the processing. Before removing this step, none of the clusters showed improved performance with our detector. In that case, the process of averaging removed the very correlations we were trying to leverage, making the problem a perfect fit for the naive Bayes detector. Additionally, our mass nodes are binary indicators of peak presence – the mass spectral peaks had to be thresholded in some way to generate the data for the detectors. Making the data continuous would improve the resolution and possibly increase the complexity of the decision boundary and allow more than a handful of possible inputs to the detector, and therefore only a handful of outputs from which performance differences needed to be found.

Secondly, 50 reshuffles of the data may not be a sufficient

sample size from which to record a significant difference, especially for a nonparametric test. More reshuffles, or better yet, the use of cross-validation may make the results more conclusive.

Finally, it is possible that the 2 clusters that did not show improvement do not involve proteins that either display the ion variations our source domain knowledge describes, in which case the naive structure can represent the distribution, or involve biomarkers for Bg. In addition to improved detector performance, we believe that given sufficient data for structure performance comparisons within the SEM procedure to be statistically significant, the learned structure has the added benefit of shedding light on the MALDI processes that occur on the molecular level. To the extent that the domain knowledge on which we base this approach is accurate and complete, it may be possible to glean an appropriate set of biomarkers from the children sets of each protein node that would be effective in discrimination. When a protein has an odd number of children, our domain constraint uniquely determines which peak node corresponds to the monomer. If a protein has two children then either mass represents a potential monomer provided neither is already a monomer of another protein. Interestingly, Ref. [13] identifies three proteins, whose masses are about 7000 Da, 7300 Da, and 8900 Da as biomarkers for Bg. The three clusters that showed improved performance each contain a mass node corresponding to each of these known protein masses. The presence of these proteins may have been what made our detector outperform the naive one in those clusters, and that improved performance indicated the possibility of biomarker identification. The unimproved performance of the other two clusters may simply be an indication that no biomarkers exist in them.

ACKNOWLEDGEMENT

REFERENCES

[1] M. Mann, Mass tool for diagnosis, *Nature*, vol. 418, pp. 731-732, Aug. 2002.

[2] M. Campa et al. eds., Special issue: Mining MALDI-TOF data, *Proteomics* vol. 3, 2003.

[3] S.L. Lauritzen and D.J. Spiegelhalter, "Local computations with probabilities on graphical structures and their application to expert systems," *Journal of the Royal Statistical Society B* vol. 50 pp. 157-224, 1988.

[4] J.S. Yedida, W. Freeman, and Y. Weiss, Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms, *NIPS*, 2000.

[5] D. Lucarelli, C. Resch, I-J. Wang, and F. Pineda, Field-theoretic Methods for Intractable Probabilistic Models, *SIAM Conference on Data Mining*, 2003.

[6] D. Heckerman, A Tutorial on Learning with Bayesian Networks in *Learning in Graphical Models*, M. Jordan, ed. MIT Press, Cambridge, MA, 1999.

[7] N. Friedman, The Bayesian Structural EM algorithm, *Conference on Uncertainty in Artificial Intelligence*, 1998.

[8] A.B. Feldman et al., Covariance mapping in matrix-assisted laser desorption/ionization time-of-flight mass spectrometry, *Rapid Communications in Mass Spectrometry*, vol. 17 pp. 991-995, 2003.

[9] N. Friedman, D. Geiger, and M. Goldszmidt, Bayesian network classifiers, *Machine Learning*, vol. 29 pp. 131-163, 1997.

[10] R. Dybowski et al., Introduction to the special issue on the fusion of domain knowledge with data for decision support, *Journal of Machine Learning Research* vol. 4, 2003.

[11] J.S. Lin, W.A. Bryden, F.J. Pineda, C.L. Resch, A. Saksena, A.B. Feldman, A Fully-Automated Peak Extraction and Baseline Estimation Method for High-Throughput MALDI-TOF-MS-Based Detection of Biological Agents, *Proceedings of the 50th ASMS Conference*, Orlando, FL, 2002.

[12] A.P. Dempster, N.M. Laird, and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society B*, vol. 39, pp.1-39, 1977.

[13] Y. Hathout et al., Small, Acid-Soluble Proteins as Biomarkers in Mass Spectrometry Analysis of *Bacillus* Spores, *Applied and Environmental Microbiology*, vol. 69, pp. 1100-1107, 2003.