Decision-Driven Models with Probabilistic Soft Logic

Stephen H. Bach Matthias Broecheler Stanley Kok Lise Getoor
Department of Computer Science
University of Maryland, College Park
College Park, MD 20742, USA
{bach, matthias, skok, getoor}@cs.umd.edu

Abstract

We introduce the concept of a decision-driven model, a probabilistic model that reasons directly over the uncertain information of interest to a decision maker. We motivate the use of these models from the perspective of personalized medicine. Decision-driven models have a number of benefits that are of particular value in this domain, such as being easily interpretable and naturally quantifying confidences in both evidence and predictions. We show how decision-driven models can easily be constructed using probabilistic soft logic, a recently introduced framework for statistical relational learning and inference which allows the specification of medical domain knowledge in concise first-order-logic rules with assigned confidence values.

1 Introduction

Medical professionals have to make many important decisions in their daily work. Whether selecting appropriate tests, providing the patient with a diagnosis or prognosis, or devising an effective treatment plan, many of these decisions must be based on uncertain and sometimes contradictory information. Medical professionals must infer accurate conclusions from this information in order to make appropriate decisions. Complicating these inferences is the rapidly increasing amount of information available from advances in medical and information collection technologies. Such inferences are further complicated as we seek to personalize them. Health care providers know that they can improve outcomes by integrating information about a particular patient with more general information collected across patient populations. Therefore, identifying and integrating relevant information is becoming an ever more daunting task for the time-stressed health care provider. We propose *decision-driven modeling* as an approach to such complex inference processes. Starting from the decisions to be made, a decision-driven model (DDM) identifies those random variables on which the decisions primarily depend and provides a mathematical framework to relate the marginal probability distributions over these variables to the available uncertain information. Inference in DDMs computes these marginals in an easy-to-interpret way, and can also compute confidence values, which enable a deeper analysis of the conclusions reached.

To motivate decision-driven models we use the diagnosis and treatment of prostate cancer for a fictitious patient "Joe Black" as our running example. Figure 1 shows excerpts of the medical records of Joe Black and some of his relatives. Joe was recently screened for prostate cancer and had a digital rectal examination (DRE) and blood tests done. The DRE showed no signs of cancer, but the slightly high volume-correlated Prostate-Specific Antigen (PSA) level of 5.2 indicates a risk of prostate cancer. Joe's doctor must decide which of these options is best: (a) conduct an invasive biopsy to collect cell samples from the prostate, (b) monitor the PSA level more frequently to identify any trend, or (c) tell Joe not to worry and to return for his regular screening next year. This decision primarily depends on the probabilities of Joe having prostate cancer and whether it is aggressive, as well as the confidence in those probabilities. For the majority of men, prostate cancer is not aggressive and progresses too slowly to be health threatening. Hence, if the probabilities of either

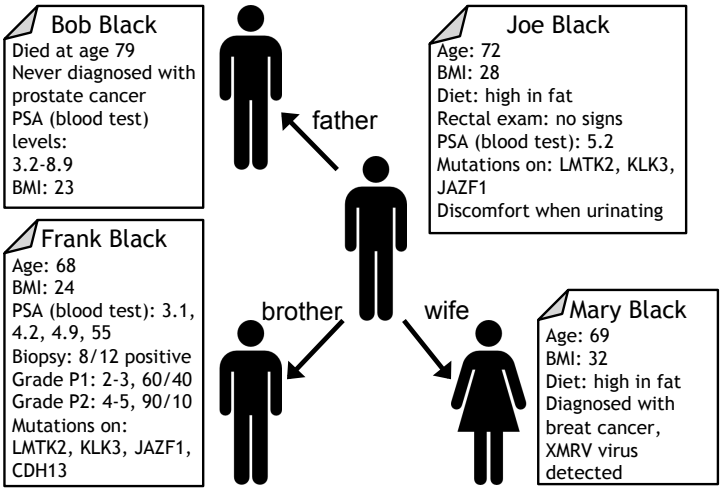

Figure 1: Example patient records and relationships

proposition $A \equiv$ "Joe has prostate cancer" or $B \equiv$ "Joe's prostate cancer is aggressive" (given that he has prostate cancer) is very low with high confidence, then the biopsy might cause more harm than good. We can help Joe's doctor make this decision by inferring the desired probabilities.

We want our inference to be as accurate as possible, so we want to incorporate as much relevant information as we can. The first source of relevant information is Joe. His PSA level and DRE, as well as his lifestyle choices, such as diet, all influence the probabilities that A and B are true. Another source of information is Joe's brother, Frank. The chances of developing prostate cancer and whether that cancer is aggressive are believed to be dependent in part on heritable factors. Suppose that Frank had a prostate biopsy done in the past, resulting in 8 positive samples out of 12, after which he had the cancerous tissue removed surgically. This information increases our belief that Joe also has prostate cancer. Given that Joe has cancer, we additionally need to infer the probability that his cancer is also aggressive. Again, Frank's medical history provides information. A histopathological assessment of the cells from Frank's biopsy yielded a P1 grading of 2 or 3 (with 60% and 40% probability respectively). This indicates that, although there is no certainty how Frank's cancer would have progressed had it not been treated, his cancer did not appear to be particularly aggressive at the time. This affects our belief that "Joe's prostate cancer is aggressive" is true. DDMs can incorporate such probabilistic dependencies and correlations and effectively combine probabilities in a sound mathematical framework, which leads to accurate inferences and, therefore, better decisions.

2 Decision-Driven Models

Before discussing decision-driven models in full generality, we define *propositional decision-driven models* as a particular type of DDM to reason about propositional assertions like A and B from our running example.

Definition 1. *Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a set of propositions. We define a probability space $(\Omega, \mathbb{P})$ where $\Omega = 2^{\mathcal{A}}$ is the finite set of all possible worlds (where we treat non-existence as negation) and $\mathbb{P}$ is some discrete probability distribution over Ω. Furthermore, we define the random variables $X_i : \Omega \to \{0, 1\}$ as $X_i(\omega) = 1$ if $A_i \in \omega$ and 0 otherwise. We denote $\mathbb{P}_{X_i} = \mathbb{P}(X_i = 1) = \sum_{\omega \in \Omega, A_i \in \omega} \mathbb{P}(\omega)$.*

This standard definition of a probability space over propositions lies at the heart of many approaches to probabilistic modeling which aim to represent the probability function $\mathbb{P}$ compactly (see e.g. [1]). Compact representation is crucial since the number of possible worlds grows exponentially in the number of propositions and therefore it is intractable to enumerate their respective probabilities. However, in many applications one is not actually interested in the individual probability of any particular possible world ω but rather in the probability that some proposition A_i is true across

all possible worlds, i.e. the marginal probability $\mathbb{P}_{X_i}$. We call such applications *decision-driven* because making a single decision often requires only knowledge of the marginal probability of a single or few propositions. For instance, to devise a treatment plan, Joe's doctor needs to infer the probability of A and B but the probability of any single world such as "Joe has prostate cancer AND Frank has prostate cancer AND Bob has diabetes AND Mary has breast cancer AND ..." does not factor into his decision making. The decision the doctor needs to make determines the propositions he or she ought to consider. Reasoning probabilistically about these propositions might require considering additional ones. For instance, $\mathbb{P}_C$, where $C \equiv$ "Frank has prostate cancer," is helpful to infer $\mathbb{P}_A$. However, their joint probability distribution is often not relevant to the decision. For example, Joe's doctor will probably ignore the probability of A *and* B.

Prior approaches to probabilistic modeling infer these marginal distributions from the joint probability function $\mathbb{P}$ through the process of marginalization, which is often intractable. Therefore, approximations such as sampling or variational techniques are often used [2]. The key idea behind DDMs is to avoid this process by modeling the distribution of the random variables X_i directly, since, for decision-driven applications, the decision maker is only interested in those marginal probabilities $\mathbb{P}_{X_i}$.

Definition 2 (Propositional Decision-Driven Model)**.** *Let $\mathcal{A} = \{A_1, \ldots, A_n\}$ be a set of propositions. A* decision-driven model *for $\mathcal{A}$ is defined by the sample space $\Omega_{\mathbb{P}} = [0,1]^n$, the set of marginal distribution denotations $\mathbf{M} = \{\mathbb{P}_{X_1}, \ldots, \mathbb{P}_{X_n}\}$, a probability density function $f(\mathbf{x} = \langle x_1, \ldots, x_n \rangle) : \Omega_{\mathbb{P}} \to \mathbb{R}_0^+$ such that $\int_{\mathbf{x} \in [0,1]^n} f(\mathbf{x}) d\mathbf{x} = 1$, and a mapping $g : \Omega_{\mathbb{P}} \to (\mathbf{M} \to [0,1])$.*

We are reusing the notation $\mathbb{P}_{X_i}$ in the definition to illustrate the relationship to the probability space over all possible worlds and highlight the fact that propositional DDMs model a probability density function f over the space of distributions of the random variables X_i. The mapping g defines this connection between each x_i and $\mathbb{P}_{X_i}$ explicitly. That is, instead of computing $\mathbb{P}_{X_i}$ bottom-up through the process of marginalization, we reason about those distributions directly as first-class citizens in the model. In this sense, DDMs are second-order probability distributions over the marginal distributions of interest to the decision-maker. We make no assumption about the density function f other than they produce a valid continuous probability distribution.

In propositional DDMs each random variable X_i is binary and therefore we restrict the marginal distributions $\mathbb{P}_{X_i}$ to be Bernoulli distributions. In theory, we can define decision-driven models for arbitrary random variables, such as real-valued random variables. However, the range of real-valued random variables is no longer enumerable. Hence, we would have to introduce a $\sigma - algebra$ for the n-fold product of the space of arbitrary probability distributions over the real line. In practice, this full level of generality is intractable and we have to either fix the parameterization of the marginal distributions (e.g. Gaussian) or approximate the distribution using kernel densities.

Definition 3 (General Decision-Driven Model)**.** *A decision-driven model is a probability distribution over a set of random variables, each of which represents a probability distribution.*

A decision-driven model (DDM) places a joint probability distribution *over* the marginal distributions of interest to a decision maker, as well as the other distributions which influence them.

To construct a DDM, we start from the decision(s) to be made. We then recursively identify probability distributions that are needed to make the decisions, with each distribution providing information about others. Finally, we place a second-order probability distribution over these distributions. In our example, the decision is whether to conduct a biopsy, for which we need to infer the probabilities over the truth values of the propositions $A \equiv$ "Joe has prostate cancer" or $B \equiv$ "Joe's prostate cancer is aggressive." Those probabilities in turn are influenced by other probabilities as discussed above. Such a relationship encoded in a DDM is a statement that the probability of a proposition provides (uncertain) information about the probability of another. These influences might be believed to be causal, or just correlational. For some propositions, we assign fixed probabilities. For instance, the proposition "Frank's prostate cells have a P1 grading of 2" would have a fixed probability of 60% stemming from the pathological assessment. These fixed probabilities serve as the evidence in the inference process. The goal of inference is to determine the probabilities of the other propositions.

We now consider the form of the probability density function f in a propositional DDM. A natural choice is a constrained continuous Markov random field (CCMRF) [3]. We define a set of *compatibility functions* $\{\phi_i\}_{i=1}^m$, which implicitly expresss dependencies among probability assignments to

propositions. These dependencies might be based on domain or case knowledge, or learned from data. The compatibility functions induce a probability density function over $\mathbf{x}$ as follows:

$$f(\mathbf{x}) = \frac{1}{Z(\Lambda)} \exp[-\sum_{i=1}^{m} \lambda_i \phi_i(\mathbf{x})] \quad ; \quad Z(\Lambda) = \int_{\mathbf{x} \in [0,1]^n} \exp[-\sum_{i=1}^{m} \lambda_i \phi_i(\mathbf{x})] \, d\mathbf{x} \tag{1}$$

where $\Lambda = \{\lambda_i\}$ are parameters associated with the compatibility functions. In addition, a CCMRF has a set of constraints which must be satisfied for a setting of $\mathbf{x}$ to have non-zero density. This lets us constrain the probabilities of propositions to be between zero and one. We can also specify other constraints of use in DDMs. For example, sets of propositions are often known to be mutually exclusive. We can easily express this by adding the constraint that the sum of the probabilities of those propositions must be less than or equal to one.

In our model for Joe's condition, the mapping g assigns each $x_i \in \mathbf{x}$ to the marginal probability $\mathbb{P}_{X_i}$ that proposition A_i is true. We can use compatibility kernels to encode dependencies among these probabilities. As an example, recall $A \equiv$ "Joe has prostate cancer" and $C \equiv$ "Frank has prostate cancer." Statistical data show that 2 out of 5 men whose brothers have prostate cancer develop cancer as well. Based on this medical knowledge we can devise the following compatibility function:

$$\phi_1(< \mathbb{P}_A = x_1, \mathbb{P}_B = x_2, \mathbb{P}_C = x_3 >) = \quad \max(0, x_3 * 0.4 - x_1)$$

If $\mathbb{P}_A = 0.05$ and $\mathbb{P}_C = 0.8$, then $\phi_1(\mathbf{x}) = 0.27$. This means, intuitively, that the probability assignments are incompatible with degree 0.27 according to ϕ_1, since we would have expected a higher probability that Joe has cancer given that Frank had cancer. In this example, the value of λ_1 reflects our confidence in this piece of medical knowledge. In Section 3, we describe a convenient language for defining such compatibility functions.

We now discuss some of the properties of DDMs and why they are of particular value in the medical community.

2.1 Quantification of Confidence

The nature of these second-order distributions enables the quantification of confidence, both in evidence and in inferences. This is of particular value in cost-sensitive domains, such as personalized medicine. Users can express relative confidence in different sources of evidence by encoding different strengths in the interactions among a marginal probability and various sources of evidence. For example, one could express the relative reliability of different medical tests. Then, a DDM will incorporate these beliefs when learning and predicting. Further, since the marginals are a MAP state in a DDM, one can quantify how sharply peaked the distribution is around that state. In other words, one can infer confidence intervals around the MAP value of one or more variables. Broecheler and Getoor [3] showed how to efficiently infer such intervals in continuous CCMRFs. This is of particular value in the medical community because health care providers must regularly make probabilistic predictions and express their confidences in them.

2.2 Interpretability

The MAP states which DDMs infer are inherently straightforward to understand and interpret, since users can easily see how each variable was affected by its relationships to other variables. When human decision makers, such as health care providers, use predictive models, it is important that they have confidence in the model's predictions and that they can recognize any errors in the model. To do this, they need to be able to understand the process by which a model arrives at its predictions. Since DDMs directly encode the relationships among the marginals, users can gain insight into why the model considers a set of marginal distributions to be optimal. Also, the confidence quantification discussed above further aids in the interpretation of the model.

2.3 Second-Order Distributions

As mentioned above, a DDM is a second-order probability distribution, meaning that it is a probability distribution defined over probability distributions. In addition to enabling the benefits mentioned above, this is semantically most appropriate for medical predictions. Medical information, e.g., test results, clinical trials, etc., is often uncertain. For example, in our hypothetical biopsy decision,

Joe's brother, Frank, had the cells obtained from his biopsy sent to a laboratory for histo-pathological assessment. The results from an expert analysis are a probability distribution over the five possible grades. DDMs can incorporate such subjective probabilistic information naturally in their inferences. Sources of subjective probability can be human, such as specialized medical examiners, or software systems optimized for particular recognition tasks, such as computer vision system for cancer classification [4]. DDMs reason using probabilistic evidence, and so are a natural fit for personalized medicine. Using probabilities as evidence to infer other probabilities is exactly what medical professionals routinely do.

3 Probabilistic Soft Logic

We implemented a variant of DDMs using probabilistic soft logic[1] (PSL), a framework for statistical relational learning and inference [5]. Instead of the propositional formulation of DDMs presented above, PSL uses a first-order-logic syntax to concisely represent CCMRFs. PSL reasons probabilistically about the values of *atoms*, which are first-order predicates grounded with arguments. Each atom corresponds to a (vector valued) continuous random variable. We use each atom to represent the marginal probability that a proposition is true. A PSL program is a set of logical rules. Each rule contains first-order predicates combined with logic operators corresponding to atom-combination functions. The sequence of predicates and operators in each rule defines a template for a compatibility function. Each assignment of the variables in a rule to constants defines a compatibility function over those atoms. A compatibility function assigns a non-negative value to how compatible the possible values of the atoms are, with zero being perfect compatibility. Finally, each rule can be a hard rule, meaning it must have a value of zero, i.e., perfect compatibility, or a soft rule, which has an associated, non-negative weight, corresponding to the confidence in the rule.

Using first-order logic one can succinctly represent the propositions about Joe and Frank having prostate cancer, which we would write as follows: `hasCancer(joe,prostate)`, `hasCancer(frank,prostate)`. The compatibility function ϕ_1 would be represented as the following ground rule in PSL:

$$\texttt{hasCancer(frank,prostate)} \times 0.4 \Rightarrow \texttt{hasCancer(joe,prostate)}$$

The implication operator $A \Rightarrow B$ corresponds to the atom-combination function $\max(0, A - B)$, meaning that the greater A is, the greater B must be to maintain a given compatibility. The notation "$\times 0.4$" indicates that the value of the body of the rule should be multiplied by 0.4. PSL allows us to compactly formulate such rules using variable symbols instead of the constants for Joe and Frank. Hence, we can write one rule which applies to all brothers and therefore encode the general knowledge that brothers of individuals with prostate cancer have a 40% chance of developing prostate cancer themselves:

$$\texttt{hasCancer(P,prostate)} \wedge \texttt{brother(P,Q)} \times 0.4 \Rightarrow \texttt{hasCancer(Q,prostate)}$$

As before, the weight assigned to this rule would reflect our confidence in the rule *across all groundings*. The weights of PSL rules can either be specified by a domain expert or learned from available data.

To construct a probability density function, a PSL program is applied to a set of (possibly typed) entities. For each possible atom in rules of the program, a variable is added to a CCMRF. Then, for each grounding of each soft rule, a compatibility function among the atoms is added to the CCMRF. Each compatibility function has the weight λ of the rule that created it. Finally, each grounding of each hard rule is added as a constraint. It is easy to see that this induces a probability density of the same form as Equation 1. PSL programs can therefore be seen as templates for CCMRFs. The advantage of using a template is that one can define a density function in terms of classes of variables, instead of individual variables. This leads to more compact definitions in large, relational domains. Personalized medicine is obviously one such domain, and PSL rules allow medical professionals to compactly express domain knowledge. Also, the same template can produce different density functions when reasoning over different entities, such as different patients with different structures

[1] Probabilistic soft logic was originally presented as probabilistic *similarity* logic [5]. The former is now used to refer to the modeling framework, to express its ability to model continuous values in general, not just similarities.

of genetic relatives, different sets of available test results, etc. In this way, the same general pieces of knowledge can be applied to different tasks with minimal user effort, as shown in the example rule above.

This also facilitates the adoption of novel medical knowledge into medical practice. For the example of prostate cancer, it has recently been suggested that the XMRV retrovirus is linked to prostate cancer [6]. Assume furthermore that XMRV can be sexually transmitted. Hence, we might devise the following rules which would apply to Joe and Mary from our example:

$$\texttt{hasVirus(P,xmrv)} \wedge \texttt{sexualPartner(P,Q)} \times 0.5 \Rightarrow \texttt{hasVirus(Q,xmrv)}$$
$$\texttt{hasVirus(P,xmrv)} \times 0.4 \Rightarrow \texttt{hasCancer(P,prostate)}$$

Now, as those rules encode novel and to some extent speculative medical knowledge, we would assign those rules low weights or choose to learn the weights from medical records as it becomes available. In any case, the ease with which such rules can be devised and deployed across decision systems allows the rapid adoption and verification of novel medical knowledge.

PSL is well suited for constructing DDMs. First, PSL can model marginal distributions accurately. Broecheler et. al. [5] used PSL to predict categorical labels for Wikipedia articles. They combined the probabilistic predictions of a Naive Bayes classifier with the observed relationships among documents to improve performance over Naive Bayes alone. PSL is also an efficient framework. Since the MAP-state problem is over continuous variables, we can formulate it as a numerical optimization problem. Therefore, one can leverage many of the efficient algorithms that exist for this well-understood class of problems. If the compatibility functions are linear or conic in their arguments, then the MAP-state problem can be solved as a second-order conic program (SOCP) in $O(R^{3.5})$ time, where R is the number of compatibility functions plus the number of random variables in the CCMRF. Additionally, the optimizer can incorporate any hard, linear equality or inequality rules. The exact derivation of the optimization problem is presented in [5]. Finally, PSL can scale to applications with millions of uncertain propositions and 8 million relationships between them [7]. Scalability to problems of such size is important in order to achieve the full potential of personalized medicine since the amount of data is rapidly increasing due to advances in medical and information technologies. For these reasons, constructing DDMs with PSL is a promising approach.

4 Related Work

Applications of machine learning techniques and probabilistic modeling in general to clinical decision support have been widely studied in the literature. What primarily distinguishes DDMs from most other approaches to inferring uncertain information is their use of second order probabilities, that is, reasoning probabilistically *about* probabilities. In contrast, most previous work on probabilistic modeling has focused on defining joint probability distributions directly over data. In case of complex dependencies and relationships between pieces of information, inference requires marginalization of the distribution. Such approaches often (a) have difficulty incorporating prior subjective probabilistic knowledge about propositions, such as a pathologist's grading of prostate cells, (b) cannot compute the confidence in the inference results, and (c) are difficult to understand and trace due to the complex process of marginalization – in particular for undirected models.

Our implementation of DDMs uses PSL, a framework for statistical relational learning (SRL), an area of research which generally seeks to combine logical formalisms and probabilistic models to reason in relational domains [8]. A related SRL technique is the CLP($\beta\mathcal{N}$) language [9], which uses inductive logic programming (ILP) to describe how the outputs of probabilistic models should be used to infer uncertain information. A CLP($\beta\mathcal{N}$) program could be viewed as a deterministic counterpart to a DDM. The probabilistic nature of DDMs allows multiples sources of uncertain evidence to be combined in different ways depending on their values. Instead of a probability distribution, CLP($\beta\mathcal{N}$) programs use only hard rules to constrain the inferred marginals in terms of the outputs of probabilistic models. CLP($\beta\mathcal{N}$) programs have been used for a number of biomedical problems. Srinivasan et. al. [10] created quantitative pharmacophore models, which describe Euclidean constraints on the location of small molecules and can be used in drug design. Costa et. al. [9] implemented a bioinformatics model for detecting a homology between a sequence of amino acids and a family of such sequences. They observed that the use of logical formalisms led to a model that was easy to interpret and study further.

Other approaches to personalized medicine have also used ILP techniques. For example, Davis et. al. [11] predicted which patients would be at substantial risk for heart attacks if they were to take Cox-2 inhibitors. They learned ILP rules to use as binary features in a probabilistic model. This approach is related to DDMs in that DDMs also can use predictions as features. However, binary features are limited in expressivity. They cannot capture the uncertainty usually inherent in evidence or predictions. DDMs can integrate the degrees of uncertainty in different predictions to make an overall prediction. Davis et. al. concluded three things: SRL techniques were best for their personalized medicine problem, a user's ability to interpret their model was of paramount importance, and that more work needs to be done to improve personalized medicine. These conclusions support the position of this paper.

A number of approaches to personalized medicine have sought to identify good predictors from a larger pool of features and/or models. Moon et. al. [12] used an ensemble approach to make diagnostic and treatment decisions. Predictors were trained on subsets of the feature space, and a cross-validation technique was used to select the optimal number to include in the ensemble. The International Warfarin Pharmacogenetics Consortium [13] used probabilistic models to predict good dosages of the anticoagulant warfarin. They started with a wide range of predictors, performed cross-validation on training data, eventually selecting linear regression. Farinas et. al. [14] used discriminant analysis to identify genes useful in predicting responses to the medicine alefacept in psoriasis. They selected 23 genes from a pool of thousands of candidates. In all of these approaches, absolute decisions had to be made about which features to use in the model. DDMs can help improve and automate this selection problem. Standard statistical learning techniques can determine how to weigh the relative influence of features, and even identify different subsets of features that are best for different regions of the feature space. For example, using PSL, weight learning would assign greater weights to the rules that related the marginals to the outputs of models which are better predictors.

5 Conclusions and Future Work

In this paper we introduced the concept of a decision-driven model, a probabilistic model that reasons directly over the marginals of interest to a decision maker. We explained why such models would be of particular use in personalized medicine. Finally, we explained how DDMs can easily be constructed using probabilistic soft logic. This approach to predictive models has a number of benefits, including the ability to express confidences in evidence and inferences, and ease of interpretability. These properties are especially important in personalized medicine, since medical professionals, often making life-and-death decisions, need to know not just what the model predicts, but also why it predicts that and how confident it is in that prediction. DDMs will enable the creation of decision-support tools which can give medical professionals all of this information.

There has been much interest in pooling available medical information to create large resources for personalized medicine. For example, Cancer Commons [15] is working to pool information on molecular subtypes of tumors and provide a network for medical professionals to share information as it is discovered. The goal is to provide the best care possible for each individual patient, while maximizing the information learned to help future patients. Using computational techniques to aid this process has already been identified as a priority. We believe this is a promising application of DDMs to improve decision making in the diagnosis, prognosis, and treatment of cancer patients.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. 0937094. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the NSF.

References

[1] Daphne Koller and Nir Friedman. *Probabilistic Graphical Models: Principles and Techniques (Adaptive Computation and Machine Learning)*. The MIT Press, 2009.

[2] Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer, 1st ed. 2006. corr. 2nd printing edition, 2007.

[3] Matthias Broecheler and Lise Getoor. Computing marginal distributions over continuous markov networks for statistical relational learning. In *Advances in Neural Information Processing Systems (NIPS)*, 2010.

[4] H. D Cheng, X. Cai, X. Chen, L. Hu, and X. Lou. Computer-aided detection and classification of microcalcifications in mammograms: a survey. *Pattern Recognition*, 36(12):29672991, 2003.

[5] Matthias Broecheler, Lilyana Mihalkova, and Lise Getoor. Probabilistic similarity logic. In *Conference on Uncertainty in Artificial Intelligence*, 2010.

[6] Robert Schlaberg, Daniel J. Choe, Kristy R. Brown, Harshwardhan M. Thaker, and Ila R. Singh. XMRV is present in malignant prostatic epithelium and is associated with prostate cancer, especially high-grade tumors. *Proceedings of the National Academy of Sciences*, 106(38):16351 –16356, 2009.

[7] M. Broecheler, P. Shakarian, and V. S. Subrahmanian. A scalable framework for modeling competitive diffusion in social networks. In *Proceedings of the 2010 IEEE International Conference on Social Computing*, 2010.

[8] Lise Getoor and Benjamin Taskar. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.

[9] Vítor Santos Costa, David Page, and James Cussens. Clp(bn): Constraint logic programming for probabilistic knowledge. In Luc De Raedt, Paolo Frasconi, Kristian Kersting, and Stephen Muggleton, editors, *Probabilistic Inductive Logic Programming*, volume 4911 of *Lecture Notes in Computer Science*, chapter 6, pages 156–188. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2008.

[10] Ashwin Srinivasan, David Page, Rui Camacho, and Ross King. Quantitative pharmacophore models with inductive logic programming. *Machine Learning*, 64(1):65–90, 2006.

[11] Jesse Davis, Eric Lantz, David Page, Jan Struyf, Peggy Peissig, Humberto Vidaillet, and Michael Caldwell. Machine learning for personalized medicine: Will this drug give me a heart attack? In *Machine Learning in Health Care Applications Workshop*, 2008.

[12] Hojin Moon, Hongshik Ahn, Ralph L. Kodell, Songjoon Baek, Chien-Ju J. Lin, and James J. Chen. Ensemble methods for classification of patients for personalized medicine with high-dimensional data. *Artificial intelligence in medicine*, 41(3):197–207, 2007.

[13] The International Warfarin Pharmacogenetics Consortium. Estimation of the warfarin dose with clinical and pharmacogenetic data. *New England Journal of Medicine*, 360(8):753–764, 2009.

[14] Mayte S. Farinas, Kejal Shah, Asifa Haider, James Krueger, and Michelle Lowes. Personalized medicine in psoriasis: developing a genomic classifier to predict histological response to alefacept. *BMC Dermatology*, 10(1):1+, 2010.

[15] Jay M. Tenenbaum. Cancer commons: Defeating cancer: One patient at a time. Technical report, CollabRx Inc., 2010.