A Multivariate Probabilistic Method for Comparing Two Clinical Datasets

Yuriy Sverchkov [*]
yus24@pitt.edu

Shyam Visweswaran [†*]
shv3@pitt.edu

Gilles Clermont [§]
cler@pitt.edu

Milos Hauskrecht [‡*]
milos@cs.pitt.edu

Gregory F. Cooper [†*]
gfc@pitt.edu

[*] Intelligent Systems Program
[†] Department of Biomedical Informatics
[‡] Department of Computer Science
[§] Departments of Critical Care Medicine, Mathematics and Industrial Engineering
University of Pittsburgh
Pittsburgh, PA 15260

ABSTRACT

We present a novel method for obtaining a concise and mathematically grounded description of multivariate differences between a pair of clinical datasets. Often data collected under similar circumstances reflect fundamentally different patterns. For example, information about patients undergoing similar treatments in different intensive care units (ICUs), or within the same ICU during different periods, may show systematically different outcomes. In such circumstances, the multivariate probability distributions induced by the datasets would differ in selected ways. To capture the probabilistic relationships, we learn a Bayesian network (BN) from the union of the two datasets. We include an indicator variable that represents the dataset from which a given patient record is obtained. We then extract the relevant conditional distributions from the network by finding the conditional probabilities that differ most when conditioning on the indicator variable. Our work is a form of explanation that bears some similarity to previous work on BN explanation; however, while previous work has mostly focused on justifying inference, our work is aimed at explaining multivariate differences between distributions.

Categories and Subject Descriptors

G.3 [**Probability and Statistics**]: Multivariate Statistics; I.2 [**Artificial Intelligence**]: Miscellaneous

General Terms

Algorithms, Theory

Keywords

Bayesian networks, data mining, explanation, multivariate differences, multivariate probability distributions, pattern recognition

1. INTRODUCTION

Identifying and explaining the similarities and differences between two clinical datasets can be very valuable. For example, consider two intensive care units (ICUs) in a healthcare system that collect similar electronic medical record (EMR) data, including patient history, symptoms, signs, therapies, and outcomes. If the two ICUs experience different outcomes (e.g., different patient re-admission rates), a clinical researcher may wish to compare these datasets to gain insight regarding as to how they otherwise differ. As another example, suppose an ICU has a marked decrease in the use of a given medication from one period of time to another. A quality-assurance officer may find it useful to gain insight into the details of this change. There are numerous other circumstances in which it is natural to compare the similarities and differences of two clinical datasets, such as clinical research trials and comparative effectiveness research.

This paper presents a novel method for examining the differences between two clinical datasets. We represent each of the datasets using a multivariate probability distribution, and then systematically examine the two distributions to gain insight into how they are different, and in what way. The explanation of why they differ is based on finding the subgroups that most contribute to the differences and describing how they combine to account for the differences. While there are many ways to implement these general ideas, the current paper describes an approach that seems promising as a launching point for this direction of research.

1.1 Overview of the Approach

This section is a brief overview of the basic technical methodology; details are provided in the methods section below. We use a Bayesian network (BN) to represent a multivariate, joint probability distribution over a set of variables $\mathbf{X} = \{X_1, \ldots, X_n\}$ in a clinical dataset $\mathbf{D}$. A BN consists of

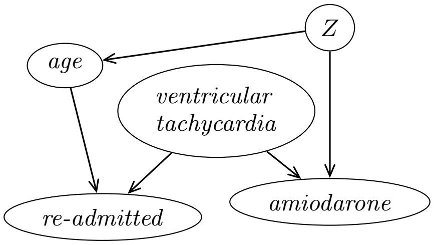

Figure 1: A Bayesian network structure for a simple example.

a directed acyclic graph on a set of nodes that represent variables, and a conditional distribution, $P(X_i|parents(X_i))$, for each variable X_i given its parents [5]. The joint distribution of a BN can be factored as follows:

$$P(\mathbf{X}) = \prod_{i=1}^{n} P(X_i|parents(X_i)).$$

Let $\mathbf{D}_A$ and $\mathbf{D}_B$ be two datasets being compared that contain the same set of variables $\mathbf{X}$. We are interested in comparing various aspects of the joint distributions $P_A(\mathbf{X})$ and $P_B(\mathbf{X})$, which represent datasets $\mathbf{D}_A$ and $\mathbf{D}_B$, respectively. Rather than represent the two distributions with two different BNs, we instead use the conditional distribution $P(\mathbf{X}|Z = 0)$ to represent $P_A(\mathbf{X})$, and $P(\mathbf{X}|Z = 1)$ to represent $P_B(\mathbf{X})$. As we will see, doing so renders the comparisons more parsimonious and coherent.

Our method first learns a BN among the variables in $\mathbf{X}$. For convenience, we assume Z is a member of $\mathbf{X}$, and we place it first in the node ordering of the variables. If the BN is a causal model, then the explanations provided will be causal. Otherwise, the explanations will be probabilistic, but not necessarily causal. Those variables with an arc from Z have different conditional distributions, depending on the value of Z. These nodes are the key nodes in the sense that all differences between $P(\mathbf{X}|Z = 0)$ and $P(\mathbf{X}|Z = 1)$ are due to differences in the conditional distributions of the key nodes given their respective parents, including Z.

Next, the method identifies those variables X_i in $\mathbf{X}$ for which $P(X_i|Z = 0)$ is substantially different from $P(X_i|Z = 1)$; call each of these probabilities *marginal distributions*. It then uses the BN to explain the main reasons for those differences. One reason could be that $P(X_i|parents(X_i), Z = 0)$ may be different from $P(X_i|parents(X_i), Z = 1)$. Another reason could be that $P(parents(X_i)|Z = 0)$ is different than $P(parents(X_i)|Z = 1)$. Also, both reasons may hold. We conjecture that decomposing the differences into their component parts will provide insight into the source of the differences, which will be useful in understanding those differences more deeply.

We now describe a simple, fictitious example to convey the main ideas introduced above. Figure 1 shows a BN that has been learned from a fictitious ICU dataset with four domain variables, plus the dataset variable Z. *Age* is the patient's age in years. *Ventricular tachycardia* is a type of abnormally rapid heartbeat; this variable represents whether it ever occurred during the ICU stay. *Amiodarone* is a medication that is sometimes used to prevent or treat *ventricular tachycardia* and other heart conditions; this variable represents whether it was ever given during the ICU stay. *Re-admitted* denotes whether the patient was previously in the ICU within the past 30 days. A realistic BN would contain many more nodes and arcs.

Suppose that *age*, *re-admitted*, and *amiodarone* have different marginal distributions for different values of Z. Qualitatively, this difference is indicated by the arc from Z to *age*; it can also be quantified probabilistically. Since *age* has no parents, these two ICU populations simply have different *age* distributions. Upon further analysis, it turns out that the rate of being *re-admitted* is higher in ICU$_A$ ($Z = 0$) than in ICU$_B$ ($Z = 1$) due to the difference in *age* distributions; for a given *age* and *ventricular-tachycardia* status, the rate of being *re-admitted* is otherwise the same for the two ICUs. Regarding *amiodarone*, it is given to patients more commonly in ICU$_B$ when *ventricular tachycardia* is present; however, the prevalence of *ventricular tachycardia* is about the same in the two ICUs. These analyses convey a deeper understanding of why *age*, *re-admitted*, and *amiodarone* have different distributions in the two ICUs.

1.2 Related Work

We briefly discuss related methods for presenting information captured in BNs that have been explored in the literature in the context of BN explanation. The BN explanation literature largely focuses on the task of explaining inference. The task of BN inference is, given some evidence $\mathbf{e}$ in the form of an assignment of a subset of the variables in the network to values, to obtain a posterior probability $P(x|\mathbf{e})$ where x represents the assignment of one or more variables in the network (that are not a part of $\mathbf{e}$) to values. There are two major approaches to inference explanation: abduction and influence-tracing. Abduction consists of providing a most probable explanation (MPE) in terms of the assignment of unobserved variables to their most probable values. The methods of abduction themselves can be divided into methods of total abduction [1, 9, 4], where all unobserved variables are assigned values, and methods of partial abduction [4, 10], where only variables that are relevant to the particular inference task are assigned values.

In influence-tracing a description might include statements such as "X_i positively influences X_j" for example, meaning that an increase in X_i increases the probability of an increase in X_j [6]. Our method is closer to influence-tracing methods than to abduction methods. Influence-tracing methods aim to describe the influence that evidence has on unobserved variables in terms of the relationships between the variables. The literature contains methods for describing such influences both qualitatively [3] and quantitatively using differences [7], log-ratios [8], and other functions such as cross-entropy [11, 12] for comparing conditional probabilities. We follow a similar approach to quantitatively compare probabilities that are obtained by conditioning on Z. However, existing inference explanation methods focus on explaining the arrival at a posterior probability, hence they do not provide an explanation of the differences a particular variable induces. Our method is unique in that, unlike previous methods, it explicitly identifies the contribution of particular subgroups to the marginal differences. Additionally, by targeting a more focused task we are able to coherently compare probabilities in terms of differences where probabilities are decomposed additively, and in terms of ratios where probabilities are decomposed multiplicatively.

2. METHOD

We now describe in detail the method for providing a description of multivariate differences between two clinical datasets. The method identifies the terms responsible for the differences using a two-level report-generating procedure for comparison of probabilistic relationships (CPR). Figures 2 and 3 show an algorithmic representation of the process, while the following text explains the process and provides its mathematical foundation.

We first learn a BN from the union of the two datasets to which we have added a binary indicator variable $Z \in \{0, 1\}$ that indicates which dataset a record appears in. While the particular choice of the BN learning algorithm inherently influences all resulting analysis, the analysis of this influence is outside of the scope of this paper. Learning of the network is restricted to only consider networks where Z has no parents. Note that, in principle, this restriction of the structure does not reduce the space of possible joint probability distributions that can be represented by the resultant network.

The learned BN is taken to be a representation of the probability distribution of the data within the original datasets. The use of a BN model allows us to explore the conditional dependencies encoded in the model in an organized fashion while maintaining the frame of comparison between the two datasets. Suppose that the distribution of X_i is different between the two datasets. This would be reflected in the network by the fact that the marginal probability $P(X_i = x_i|Z = 0)$ is different from $P(X_i = x_i|Z = 1)$ for x_i, a particular value of X_i. The difference $P(x_i|Z = 1) - P(x_i|Z = 0)$ is used as a measure to quantify how different those two probabilities are.

When the node X_i has parents other than Z in the BN we explain the difference in the distribution of X_i between the two datasets in terms of its parents. Z may or may not be a parent of X_i in the general case. Let $\Pi_i := \mathit{parents}(X_i)\backslash\{Z\}$ denote the set of nodes that are parents of X_i, with Z excluded. Denoting possible assignments of Π_i to particular values by π_i, and a particular assignment of Z to 0 or 1 by z, we get the equality:

$$P(x_i|z) = \sum_{\pi_i} P(x_i, \pi_i|z). \quad (1)$$

Let $z_1 := \mathrm{argmax}_z\, P(x_i|z)$. Then

$$P(x_i|z_1) - P(x_i|1 - z_1) = \sum_{\pi_i} \left[P(x_i, \pi_i|z_1) - P(x_i, \pi_i|1 - z_1)\right]. \quad (2)$$

Hence for each assignment of the parents π_i, a positive difference $P(x_i, \pi_i|z_1) - P(x_i, \pi_i|1 - z_1)$ contributes towards the difference of interest $P(x_i|z_1) - P(x_i|1 - z_1)$, and a positive difference $P(x_i, \pi_i|1 - z_1) - P(x_i, \pi_i|z_1)$ contributes against it. Since the number of parent configurations π_i grows exponentially with the number of parents $|\Pi_i|$, the number of such additive terms can be quite large. For this reason we only report those terms that match or exceed a user-defined threshold t. The threshold controls the number of terms in the sum that are displayed: a smaller t displays more terms ($t = 0$ displays all terms) and a larger t displays less terms ($t > 1$ displays no terms).

Further analysis of the terms of largest magnitude in the sum produces an explanation in terms of the probabilities that contribute most to (and against) the difference of distributions of X_i between the two datasets. Figure 2 shows an algorithmic representation of this first level of analysis described so far. The first level triggers the second level of analysis for each additive term that we now describe.

```
CPRL1( $X_i$, $t$ ):
for $x_i \in$ possible values of $X_i$, ordered in decreasing order
by $|P(x_i|Z = 1) - P(x_i|Z = 0)|$ do
  Let $z_1 := \mathrm{argmax}_z P(x_i|z)$
  Present the difference $P(x_i|z_1) - P(x_i|1 - z_1)$
  % List large positive contributions:
  for $\pi_i \in$ possible configurations of $\Pi_i$ such that
  $P(x_i, \pi_i|z_1) - P(x_i, \pi_i|1 - z_1) \geq t$ ordered in descending
  order by the difference do
    Call CPRL2($x_i, \pi_i, z_1$)
  end for
  % List large negative contributions:
  for $\pi_i \in$ possible configurations of $\Pi_i$ such that
  $P(x_i, \pi_i|1 - z_1) - P(x_i, \pi_i|z_1) \geq t$ ordered in descending
  order by the difference do
    Call CPRL2($x_i, \pi_i, 1 - z_1$)
  end for
end for
```

Figure 2: The first level of CPR for a given variable X_i and a threshold t.

Since it is difficult to assign intuitive meaning to the term $P(x_i, \pi_i|z)$, we take advantage of the natural decomposition of this term that is provided by the BN:

$$P(x_i, \pi_i|z) = P(x_i|\pi_i, z)P(\pi_i|z). \quad (3)$$

Let $z_2 := \mathrm{argmax}_z\, P(x_i, \pi_i|z)$. In order to meaningfully relate the multiplicative decomposition in Equation 3 to the additive nature of the differences discussed above, observe that:

$$\frac{P(x_i, \pi_i|z_2) - P(x_i, \pi_i|1 - z_2)}{P(x_i, \pi_i|1 - z_2)} = \frac{P(x_i, \pi_i|z_2)}{P(x_i, \pi_i|1 - z_2)} - 1. \quad (4)$$

The choice of z_2 allows for a natural interpretation of these terms: the left hand side of the equation is simply the fold increase in probability from the dataset indicated by $Z = 1 - z_2$ (the one with the smaller probability for that term) to the other dataset, indicated by $Z = z_2$. The right hand term is just a ratio of probabilities (guaranteed to be greater than or equal to 1) minus one, hence we have a way to relate the ratio to the difference. The decomposition of the ratio into multiplicative terms follows:

$$\frac{P(x_i, \pi_i|z_2)}{P(x_i, \pi_i|1 - z_2)} = \frac{P(x_i|\pi_i, z_2)}{P(x_i|\pi_i, 1 - z_2)} \frac{P(\pi_i|z_2)}{P(\pi_i|1 - z_2)}. \quad (5)$$

Here $P(x_i|\pi_i, z)$ is a conditional probability of an assignment of X_i given its parents.[1] $P(\pi_i|z)$ is the joint probability of the assignment of the parents within one of the datasets. We further decompose $P(\pi_i|z)$ into a product by representing the parents of X_i explicitly as $\Pi_i = \{Y_1, \ldots Y_m\}$:

$$P(\pi_i|z) = \prod_{j=1}^{m} P(y_j|y_1, \ldots, y_{j-1}, z) \quad (6)$$

[1] When Z is not a parent of X_i, the fact that X_i cannot be an ancestor of Z implies that conditioning on X_i's parents and not conditioning on any descendants of X_i d-separates X_i from Z, making the probability equal to $P(x_i|\pi_i)$.

```
CPRL2($x_i, \pi_i, z_2$):
Present $P(x_i, \pi_i|Z = z_2)$ and $P(x_i, \pi_i|Z = 1 - z_2)$, their difference, and ratio.
Present the conditional probabilities $P(x_i|\pi_i, z_2)$, $P(x_i|\pi_i, 1 - z_2)$ and their ratio.
Let $\{Y_1, \ldots, Y_m\} = \Pi_i$ be the parents of $X_i$ other than $Z$.
Let $\{y_1, \ldots, y_m\}$ be the assignments of $\{Y_1, \ldots, Y_m\}$ in $\pi_i$.
% List the ratios contributing towards $\frac{P(x_i,\pi_i|z_2)}{P(x_i,\pi_i|1-z_2)}$:
for $Y_j \in \Pi_i$ ordered in descending order by $\frac{P(y_j|y_1,\ldots,y_{j-1},z_2)}{P(y_j|y_1,\ldots,y_{j-1},1-z_2)}$ do
  Present $P(y_j|y_1, \ldots, y_{j-1}, z_2)$, $P(y_j|y_1, \ldots, y_{j-1}, 1 - z_2)$ and their ratio.
end for
% List the ratios contributing against $\frac{P(x_i,\pi_i|z_2)}{P(x_i,\pi_i|1-z_2)}$:
for $Y_j \in \Pi_i$ ordered in descending order by $\frac{P(y_j|y_1,\ldots,y_{j-1},1-z_2)}{P(y_j|y_1,\ldots,y_{j-1},z_2)}$ do
  Present $P(y_j|y_1, \ldots, y_{j-1}, 1 - z_2)$, $P(y_j|y_1, \ldots, y_{j-1}, z_2)$ and their ratio.
end for
```

Figure 3: The second level of CPR for a given variable assignment x_i, a parent assignment π_i, and a Z-assignment z_2.

and obtain the following decomposition of the ratio

$$\frac{P(\pi_i|z_2)}{P(\pi_i|1-z_2)} = \prod_{j=1}^{m} \frac{P(y_j|y_1, \ldots, y_{j-1}, z_2)}{P(y_j|y_1, \ldots, y_{j-1}, 1 - z_2)}. \qquad (7)$$

Examining each term in the resultant product leads to two observations: testing whether a term is greater or less than 1 shows whether it contributes towards or against the ratio $\frac{P(x_i,\pi_i|z_2)}{P(x_i,\pi_i|1-z_2)}$. Moreover, we obtain a measure of the magnitude of the contribution as a multiplicative factor by looking at the value of the ratio for terms that contribute towards $\frac{P(x_i,\pi_i|z_2)}{P(x_i,\pi_i|1-z_2)}$ and at the inverse of the value for terms that contribute against the ratio. Figure 3 shows an algorithmic representation of this phase of the analysis.

We can expand the explanation capabilities by repeating the process with y_j as x_i; however, we leave such analysis for future work as stronger theoretical grounding is required in order to make such a recursive explanation process coherent.

3. APPLICATION TO A DATASET ON PATIENTS WITH PNEUMONIA

We applied our method to a clinical dataset that has been previously used to study community acquired pneumonia. The data used were collected in a prospective cohort study of hospitalized and ambulatory care patients conducted from October 1991 to March 1994 at five medical institutions [2]. Patients included in the study had to have one or more symptoms suggestive of pneumonia, as well as radiographic evidence of pneumonia within 24 hours of presentation. The variables available in the dataset include categorical variables, continuous variables, and discretized versions of continuous variables. We restricted ourselves only to categorical variables and one discretization of each continuous variable, yielding 165 variables. The available variables included demographic information, history and physical examination information, laboratory results, chest X-ray findings, and outcomes.

To demonstrate our method we selected two of the five medical institutions that participated in the study as the two data sources to compare. We will refer to these institutions as *subsite A* and *subsite B*. Our merged dataset then consists of patient records from *subsite A* and *subsite B*, and the indicator variable Z corresponds to the *subsite* variable.

We learned a BN from the merged dataset using a two-phase greedy algorithm. The algorithm attempts to maximize the K2 score of the structure by adding arcs to an initially empty network in the first phase, and removing arcs from the result in the second phase [5]. The structure was constrained to at most five parents per node. The order of the variables in the network was, ordered from ancestors to descendants, constrained as follows: the *subsite* variable is first, followed by demographic variables such as *age* and *sex*, followed by variables that describe the patient's history and state at admission such as *smoke* (whether the patient smokes) and *flu* (whether the patient had influenza six weeks prior to presentation), followed by other variables which represent findings such as test results and other information about the patient's state while in the hospital, and outcome variables such as *dead30pr* and *dead90pr* (whether the patient has died within 30 or 90 days after presentation) are last. While the order constraints have a loosely causal and temporal justification, the resultant network is not guaranteed to be causal, and the results must be interpreted probabilistically rather than causally.

The threshold we selected to use with our method for the purposes of these evaluations is $t = 0.01$, as it was found to provide an informative yet manageable level of detail. In a practical application, the threshold would be selected based on the user's (e.g. clinical researcher's) goals and preference for level of detail. In the learned BN, there were 15 variables that were children of the *subsite* variable, and their marginal differences exceeded the threshold. We selected two of these variables, *flu* and *aspevent* (aspiration event), to illustrate the features of the method.

For the *flu* variable, the marginal probability for the value $flu = yes$, $P(flu = yes|subsite)$, is 0.130 for *subsite A* and 0.326 for *subsite B*, yielding a difference of 0.196. In the BN, the variable has only one parent besides *subsite*, namely, *age*. The additive terms that contribute to this difference take the form

$$P(flu = yes, age|subsite = B) - P(flu = yes, age|subsite = A).$$

The values of *age* (as discretized ranges) corresponding to terms that have a positive difference that exceeds the 0.01 threshold are: 30-44 years old, 0.082; 18-29 years old, 0.056; 75-90 years old, 0.037; 60-74 years old, 0.011. No terms exceeded the threshold and contributed negatively to the difference.

Proceeding to the second level of analysis for the first of these terms, we compute the ratio

$$\frac{P(flu = yes, age = \text{30-44}|subsite = B)}{P(flu = yes, age = \text{30-44}|subsite = A)} = 3.307,$$

which is further decomposed into the conditional part

$$\frac{P(flu = yes|age = \text{30-44}, subsite = B)}{P(flu = yes|age = \text{30-44}, subsite = A)} = 2.713$$

and the parent part

$$\frac{P(age = \text{30-44}|subsite = B)}{P(age = \text{30-44}|subsite = A)} = 1.415.$$

Thus, both parts contribute to the main ratio, with the conditional part contributing more, meaning that the difference for the subgroup of patients between 30 and 44 years of age is mostly explained by a higher proportion of 30-44 year-olds that had influenza recently, but the fact that there are proportionally more patients that are 30-44 also contributed to the difference. We observe similar numbers for the terms corresponding to an *age* value of 18-29 and 75-90, with the notable exception that the parent part of the 75-90 ratio contributes slightly against the additive term's ratio:

$$\frac{P(age = \text{75-90}|subsite = B)}{P(age = \text{75-90}|subsite = A)} = 0.987.$$

These results show that the higher proportion of patients with flu at *subsite B* is explained by the observation that *subsite B* patients who were 18-29, 40-44, and 75-90 had more flu recently than *subsite A* patients, and additionally, there were proportionally more patients with *age* 18-29 and 30-44 at *subsite B* than *subsite A*.

The analysis for the *aspevent* variable is both more interesting and more complex. The marginal distribution difference is 0.083 for *aspevent* = *yes*, with *subsite A* seeing proportionally more aspiration events. Unlike *flu*, *aspevent* has *alert* (patient alertness level), *swalldia* (presence of swallowing disorders), *nutrstaa* (malnutrition or poor nutritional status), and *sza* (seizures) as parents. Only two additive terms exceed the difference threshold of $t = 0.01$, both correspond to absence of *swalldia*, *nutrstaa* and *sza*, and both contribute positively to the marginal difference. One term corresponds to *alert* = *alert* (patient is alert) and the other to *alert* = *lethargic* (patient is lethargic). The decomposed ratios of both terms show that the contribution of the conditional term (with a factor between 8 and 9) outweighs the contribution of the parent terms (all with factors close to 1).

These results show that the higher proportion of aspiration events at *subsite A* is primarily explained by the observation that among patients without malnutrition, swallowing disorders, or seizures, who are either alert or lethargic, more experience aspiration events at *subsite A* than at *subsite B*.

4. CONCLUSIONS AND FUTURE WORK

We have presented a method for exploring the probabilistic relationships between variables that are relevant to examining the differences between two datasets. In order to do so we first learn a BN that captures the probabilistic relationships in the datasets. We then present a decomposition of the marginal probabilities according to the learned BN in a way that identifies particular subgroups of records that contribute most to the difference between datasets. The analysis of the pneumonia patient dataset yielded good results: isolated subgroups that were primarily responsible for the differences of probabilities in the dataset were successfully identified. The results exemplified the explanatory power of the method and showed reasonable findings. This encourages further development and extensions of the method.

While we have focused the application to clinical datasets, the method presented here can be applied to other kinds of data as well. A possible extension of the method mentioned above, involves in expanding the analysis produced beyond a given node and its parents in the BN. The idea is to recursively examine each parent Y_j as we examined the initial variable X_i in the analysis here. Another direction for extending the method is the incorporation of model averaging to enable the estimation of the variance of probabilities in the analysis. Yet another improvement on the method that is more in the spirit of explanation would add the capability to consolidate the explanations provided, in order to automatically generate the sort of summaries that were compiled by hand in this paper.

5. ACKNOWLEDGMENTS

This research was supported by grant IIS-0911032 from the National Science Foundation.

6. REFERENCES

[1] E. Charniak and S. E. Shimony. Cost-based abduction and map explanation. *Artificial Intelligence*, 66:345–374, 1994.

[2] G. F. Cooper, V. Abraham, C. F. Aliferis, J. M. Aronis, B. G. Buchanan, R. Caruana, M. J. Fine, J. E. Janosky, G. Livingston, T. Mitchell, S. Monti, and P. Spirtes. Predicting dire outcomes of patients with community acquired pneumonia. *Journal of Biomedical Informatics*, 38(5):347 – 366, 2005.

[3] M. J. Druzdzel. *Probabilistic reasoning in decision support systems: from computation to common sense*. PhD thesis, Pittsburgh, PA, USA, 1993. UMI Order No. GAX93-22863.

[4] J. A. Gámez. Abductive inference in bayesian networks: A review. In *Advances in Bayesian Networks*, pages 101–117. Springer, 2004.

[5] D. Heckerman. A tutorial on learning with bayesian networks. In M. I. Jordan, editor, *Learning in graphical models*, pages 301–354. MIT Press, Cambridge, MA, USA, 1999.

[6] C. Lacave and F. J. Diez. A review of explanation methods for bayesian networks. *Knowledge Engineering Review*, 17:2002, 2000.

[7] C. Lacave, M. Luque, and F. J. Díez. Explanation of bayesian networks and influence diagrams in elvira. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 37(4):952–965, 2007.

[8] D. Madigan, K. Mosurski, and R. G. Almond. Graphical explanation in belief networks. *In Journal of Computational and Graphical Statistics*, 6:160–181, 1997.

[9] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[10] S. E. Shimony. A probabilistic framework for explanation. Technical report, Providence, RI, USA, 1991.

[11] H. J. Suermondt. *Explanation in Bayesian belief networks*. PhD thesis, Stanford, CA, USA, 1992. UMI Order No. GAX92-21673.

[12] H. J. Suermondt and G. F. Cooper. An evaluation of explanations of probabilistic inference. *Computers and Biomedical Research*, 26(3):242 – 254, 1993.