Causal Reasoning Engine: An Explanation-Based Approach to Syndromic Surveillance

Benjamin B. Perry ~ Tim Van Allen
Quantum Leap Innovations
bbp@quantumleap.us

Abstract

Quickly detecting an unexpected pathogen can save many lives. In cases of bioterrorism or naturally occurring epidemics, accurate diagnoses may not be made until much of the population has already been jeopardized. The goal of syndromic surveillance is to detect early anomalies that emerge from patient data in a given population area and to note disease patterns before more individuals begin to experience definitive symptoms. We developed a syndromic surveillance approach for generating advance warnings of potential wide-spread diseases as well as identifying demographic attributes that are predictive of the diseases. We describe the Causal Reasoning Engine (CRE), a multipurpose decision support system for diagnosing causes from observed symptoms and predictors. The CRE uses Bayesian inference and machine learning methods and deploys an intuitive explanation-based framework for causal modeling. We also present a diagnostic decision support tool based on the CRE that allows emergency responders to analyze and interrogate findings.

1. Introduction

In several possible infectious disease epidemics such as smallpox or SARS, the first few days of the unexpected epidemic are the most important when it comes to containing the number of casualties. Every hour that ticks by without the contact person(s) being contained can lead to many more potential source contacts, which can lead to widespread disaster. Unfortunately, in many instances, patients are discharged after their initial visit with incorrect diagnoses. It is not until these patients' symptoms advance that they are correctly diagnosed and a potential epidemic has been realized. The problem is compounded when various health facilities in the same region do not share data with a central data analyzer. If several patients present similar odd but seemingly harmless symptoms at the same time in different facilities who do not share data, the potential for quickly discovering an epidemic has been severely hampered. Because of this problem, we have developed a syndromic surveillance system that helps to detect early signs of epidemics by using an innovative explanation-based approach of analyzing patient symptom data from several health facilities' data warehouses. In this paper, we describe an overview of our syndromic surveillance system and the unique explanation-based approach we use for the diagnostic engine. We compare our approach to other diagnostic approaches. We finish up this paper with a discussion of healthcare data requirements, practicality, and privacy concerns.

2. System Overview

The overall design of our syndromic surveillance system is shown in figure 1. Several health facilities send anonymous patient symptom records to the Causal Reasoning Engine (CRE), which processes the records and sends results to the user-friendly front-end, which we call the Threat Assessment Module (TAM). The TAM has two main functions: it serves as an alert system that can run in the background, scanning for anomalies and trends, and it serves as a user-friendly tool that allows health officials to explore the patient data that the CRE has processed.
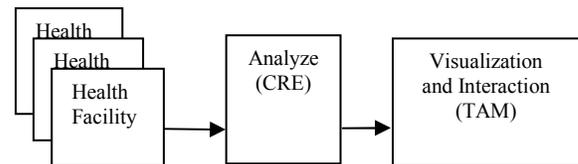

Figure 1 – Syndromic surveillance flow of information

The CRE examines a set of symptoms for each patient record it receives (from any number of health facilities via the HIPAA Electronic Data Interchange) and considers every likely disease(s) that can likely explain the symptoms. The CRE is capable of picking up multiple non-mutually exclusive diseases; for example, if someone registers symptoms of a broken arm, coughing, and headache, the CRE would likely report that the patient has both a broken arm and a mild case of influenza whereas many other surveillance systems would fail to report a probable disease since hardly any diseases list "broken arm" as a symptom.

For each disease, the CRE measures how likely the optimal subset of symptoms in the case match the symptomology of the disease and provides a probability

that accurately reflects the level of match. The CRE can be used as a general diagnosis tool or to monitor only those diseases that health officials are particularly interested in.

The CRE determines the likelihood of each disease for each patient record as well as the most probable explanations of the symptoms (which could be a conjunction of diseases). Once the set of possible diseases has been determined, the CRE can analyze how unusual each disease is given all of the circumstances. The CRE is capable of using environmental information to strengthen its analysis, such as pollen count, time of the year, weather, and etc. by adjusting disease priors according to environmental information. For example, a case of influenza in the middle of July might be considered to be unusual whereas the same case in March is not all that unexpected. The CRE uses these prior probabilities to return a set of diseases that most likely explains the symptoms. As more patient records are processed, the CRE adjusts the posterior probabilities of each disease in the overall population and sends those probabilities to the TAM. The CRE examines only the last predetermined number of cases in calculating the posterior probabilities as to avoid skewing the posterior probabilities towards a time-specific trend.

The TAM compares the posterior probabilities against the prior probabilities and alerts the health officials if any diseases are having an unusually high posterior probability. In other words, if the number of occurrences of a specific disease is significantly higher than expected, the TAM can alert the health official. The TAM can also alert the health official if a single patient record presents symptoms that are best explained by a specific disease of interest; for example, health officials might want to be notified if even a single patient presents symptoms that closely match the symptomology of meningitis.

The CRE is also capable of learning associations between external information and diseases. As long as the use of the CRE is Health Insurance Portability and Accountability Act (HIPAA) compliant and adheres to the strictest of privacy controls, the external information might include gender, age, or any interview question such as any major places the patient might have visited or any foods they might have consumed. The CRE uses simple data mining techniques to associate the external information to diseases. For example, the CRE would be able to determine that a majority of people infected with E. coli in a specific time frame might have patronized a particular restaurant; that information could be presented to the health officials who could then act accordingly. This particular example might take much longer for health officials to determine the source without the cooperative sharing of patient records. When multiple health facilities send in their data to the CRE, the CRE is able to mediate between all facilities and conglomerate the information to have a more complete awareness of the region, which

allows for a vastly improved resolution time. One caveat is that the CRE's learning capabilities are only as good as the data recorded by the health facilities.

Once a health official has been alerted to an unusually high frequency of a specific disease in the population overall or in a specific patient, the TAM provides the health official with a front-end user interface to explore the data and make sense out of it. If any external associations have been linked to a disease, the user will be presented with that information.

3. Causal Reasoning Engine

In this section, we present the technical details for the Causal Reasoning Engine, its inference algorithm, and its learning algorithm.

3.1. Introduction

The Causal Reasoning Engine is the crux of Quantum Leap Innovations' syndromic surveillance system. The CRE is best described as a multipurpose engine for diagnosing probable causes from observed effects and linking the probable causes to observed predictors. The CRE provides posterior probabilities for each possible cause that is explained by the observed symptoms and provides combinations of causes that best explain the observations. The CRE is also able to dynamically learn what non-symptom attributes may be predictive of certain causes by using a simple data-mining technique.

One aspect of the CRE that sets it apart from other existing approaches to diagnosis is the form of local distributions of symptoms, which we later refer to as *Explanation-based*. There are two main benefits to this approach: (1) the CRE's ability to discover sets of non-mutually exclusive probable causes, and (2) the simplistic disease model, which allows users to quickly and accurately enter a very small amount of numeric parameters compared to other diagnosis approaches.

3.2. Model

The CRE uses a *Bayesian network* for its reasoning model framework. A Bayesian network is a directed, acyclic graph whose nodes represent random variables and whose arcs represent dependencies between variables. If an arc is pointing from node A to node B, then B is dependant upon A. Bayesian networks allow us to model probabilities of a domain; each non-root node has a conditional probability table which quantifies the effects that its parent nodes have on it. Root nodes have a prior probability table. Bayesian networks can model expert knowledge and are widely used in several aspects of computer science.

All symptom nodes in the CRE are leaf nodes that have two possible distributions: one distribution if it is *explained* and one distribution if it is *unexplained*. A symptom is *explained* if one or more of its causes are present; a symptom is *unexplained* otherwise. This model framework, which we call *Explanation-based*, allows for a very simple model representation as well as very efficient inference algorithms. For each symptom, only two numerical parameters have to be entered: p1, the probability of the symptom being present if at least one of its causes is present, and p2, the probability of the symptom being present if none of its causes are present. Table 1 shows a sample conditional probability table for a symptom node.

Table 1 - Symptom Conditional Probability Table

	Symptom explained (At least one cause present)	Symptom not explained
Symptom present	p1 = .9	p2 = .3
Symptom absent	1 – p1 = .1	1 – p2 = .7

Figure 2 shows a sample simplified model that the user sees, while figure 3 shows how the model is actually designed in a Bayesian network. The CRE uses a convention that all of the symptom nodes in the Bayesian network have at most one parent. This does not mean that symptoms can only have one cause; for example, the symptom node "Cough" might have hundreds of possible causes. In the model, all of the symptom's possible causes are combined into a disjunction node which acts as an OR truth table. The disjunction node is true if at least one of the causes is present and false if none of the causes are present. The disjunction node then becomes the symptom's parent (see the disjunction node "C1 or C2" in figure 3).
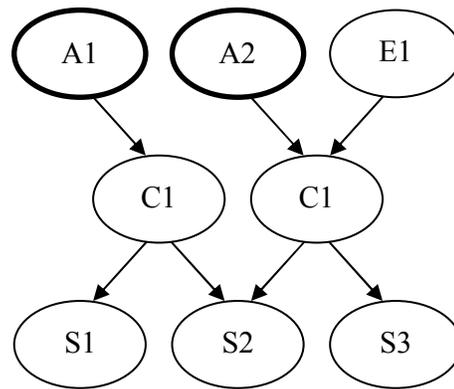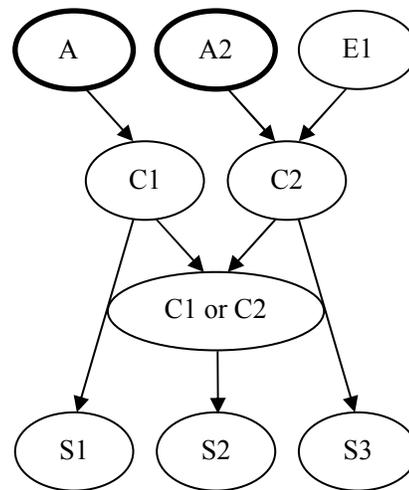

Figure 2 - CRE model that the user sees

Figure 3 – Figure 2 CRE model implemented in Bayesian network

In figures 2 and 3, there are four sets of variables: *A* nodes, which represent predictor variables, *E* nodes, which represent environmental nodes, *C* nodes, which represent cause variables, and *S* nodes, which represent symptom variables. In the health domain, example predictor variables might be "ate at Joe's" or "attended football game", environmental variables might be "time of year" or "pollen count", cause variables might be "has flu" or "has SARS", and symptoms might be "coughing" or "diarrhea."

Values for environment nodes are provided at runtime as evidence. As each case is submitted to the CRE, the CRE's runtime environment can populate the environment nodes by querying various services, such as a weather service tool. Neither predictor nodes nor environment nodes require any numeric parameters from the user since they are instantiated as evidence at runtime.
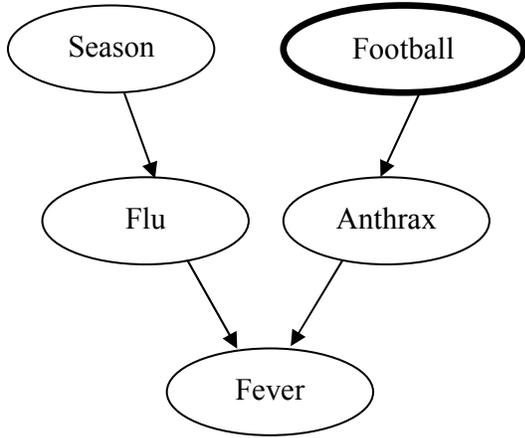
Figure 4 - Very simple CRE model

The simplest network consists of only cause nodes and symptom nodes. To improve the accuracy of cause priors, environmental nodes can be used. Such nodes might be time of year, pollen count, etc. Figure 4 shows a very simple actual network. This simple network demonstrates how influenza might be chosen as the cause if a patient presents the "fever" symptom in the winter, but anthrax might have a higher chance of being chosen if the season is summer. In this network, "Season" is an environmental node that is given as evidence with each record. "Football" is a predictor node that is learned at runtime and not provided by the modeler. "Influenza" and "Anthrax" are cause nodes, and "Fever" is a symptom node. Among the patients whose cases support anthrax, a majority might have gone to a particular football game.

3.3. Inference Algorithm

In order to infer the posterior probabilities of (a) each individual cause as well as (b) the most probable explanations for the evidence, an inference algorithm has been developed. The two types of inference are similar; for inference type (a), the causes are not necessarily independent in the posterior, so (b) may provide information about the posterior dependencies that (a) does not. The inference algorithm first generates a set of probable explanations and then marginalizes over the set to compute the approximate marginal cause posteriors.

The inference algorithm uses a greedy local search algorithm to find the sets of causes that have the best posterior probability given the observed symptoms. In most cases, only one cause will be present. Figure 6 shows a chart of the inference algorithm.

3.4. Learning Algorithm

The predictor nodes and their links to cause nodes are determined at run-time by a decision-tree learning algorithm that uses a pseudo marginal log likelihood scoring function and a simple greedy branch-and-bound search algorithm. Our goal is to learn the distribution $P(C|A)$. The learning algorithm learns a decision tree T_i for each cause distribution $P(C_i|A)$. Each leaf of the tree is a distribution $P(C_i|A=a)$, where a is the assignment to the predictor variables determined by the path from the root of the tree to the leaf. The final Bayesian model has links from A_j to C_i if and only if A_j appears in T_i. T_i corresponds to a partitioning of the A-space and a local distribution of C_j for each partition.

The challenge in our decision tree algorithm is that the C nodes are uncertain. In order to get around that, we use a modified form of the marginal log-likelihood of T_i for the scoring function, which we call pseudo marginal log-likelihood (psMML). Instead of inputting pairs <A, C> to the scoring function, we pass <A, P>, where P is a vector of the posterior probabilities of the causes, based on the original prior probabilities, the cause-symptom model, and the observed symptoms. The actual psMLL equation is shown below in figure 5

$$psMLL \ (C_i \mid network \ ) = \sum_{j=1}^{m} \left(p_j \log \ P(C_i = present \ \mid p_{j-1}, p_{j-2}, \cdots p_1, network \ ) + (1 - p_j) \log \ P(C_i = absent \ \mid p_{j-1}, p_{j-2}, \cdots p_1, network \ ) \right)$$
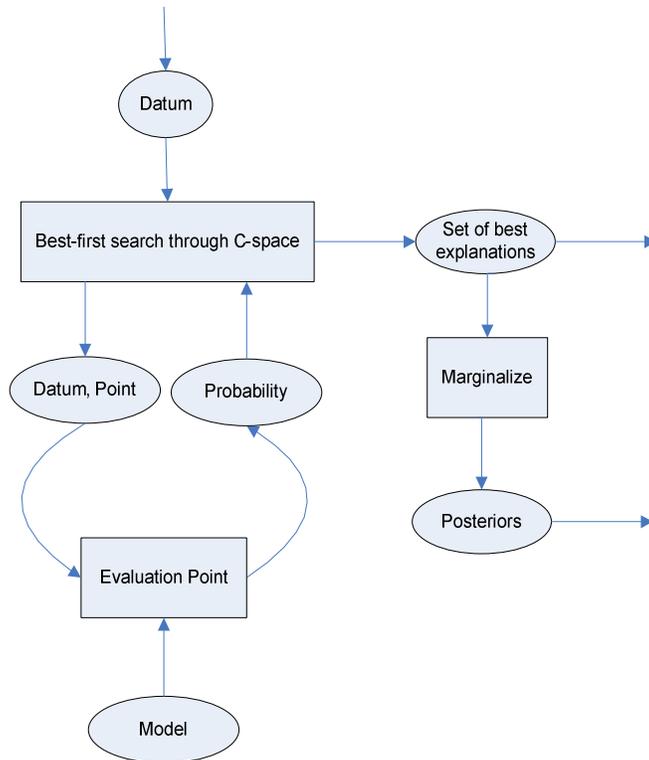
Figure 5 - psMLL equation

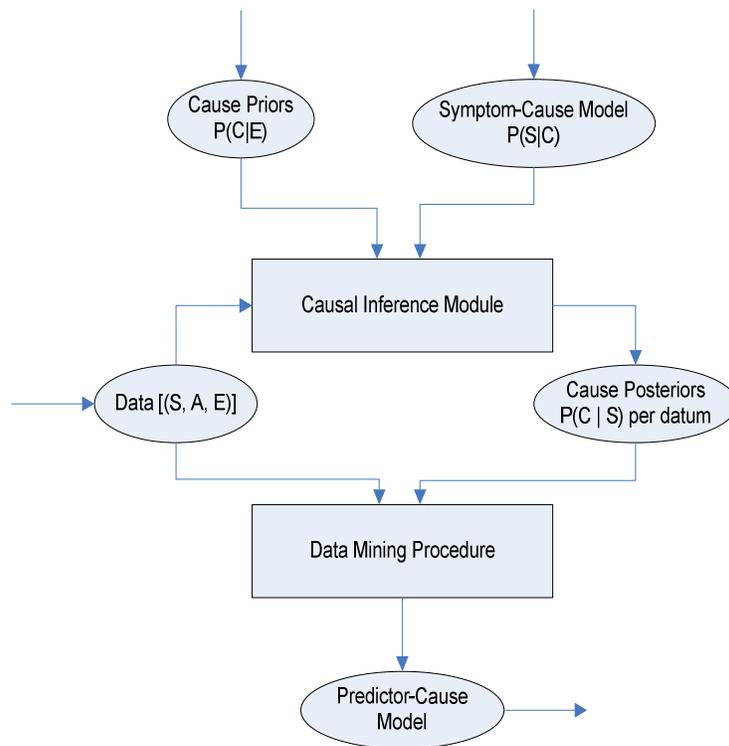**Figure 6 - CRE inference algorithm**

Figure 7 – CRE Learning algorithm

4. Results

We have tested the syndromic surveillance system with synthetic data generated by an advanced data generator; this data generator takes as input a Bayesian network and allows the user to draw a Bezier curve to represent a probability distribution function for cases per date as well as a Bezier curve to represent the probability enhancement curve (evidence instantiation or increased evidence likelihood) for every state of every node. The generator then outputs several thousand cases, which we submitted to the CRE in a time-controlled manner. The CRE monitored the cases and correctly sent warnings during various onsets of diseases. The CRE also correctly identified predictor nodes. Quantum Leap Innovations is in the process of securing authorization to receive non-synthetic patient data from the state of Delaware and we plan to examine various onsets that happened in the past. We are confident that the CRE would have picked up on them.

Our first test used a simple network which consists of only five causes and five symptoms. Three predictor variables were used in this simple network. Figures 8 and 9 show the Bayesian network used to generate the cases and the Bayesian network that the CRE used to monitor the cases.

When we did not introduce any unexpected onsets, the posterior probabilities of each cause closely matched the prior probabilities of each cause; for example, the root mean square error between the posterior probabilities and the original network when evaluating 10,000 cases for this simple network was 0.0117. This indicates that the CRE correctly identified a strong majority of the causes and correctly learned which predictor nodes were associated with their respective causes. Table 2 shows a matrix of various trials.

When we did introduce onsets, the threat assessment module correctly indicated to the user that an onset was approaching. It should be noted, however, that we are not health experts. The probabilities we used for the cause priors and the symptom probabilities were based on limited research. A qualified health domain expert would be the ideal person to create the actual networks.

More advanced networks in other domains yielded similar results. As is expected, though, as the network becomes more and more complicated, and the probabilities become less and less distinguishable, the accuracy of the network decreases. This is partially why the CRE was not designed to be a general diagnostic replacement. We don't intend to put doctors out of jobs, or pretend to be able to diagnose any possible ailment. The CRE is mainly used to diagnose a limited number of diseases that the healthcare officials are interested in monitoring. Thus, the fewer diseases that are being monitored, the more accurate the model will be.
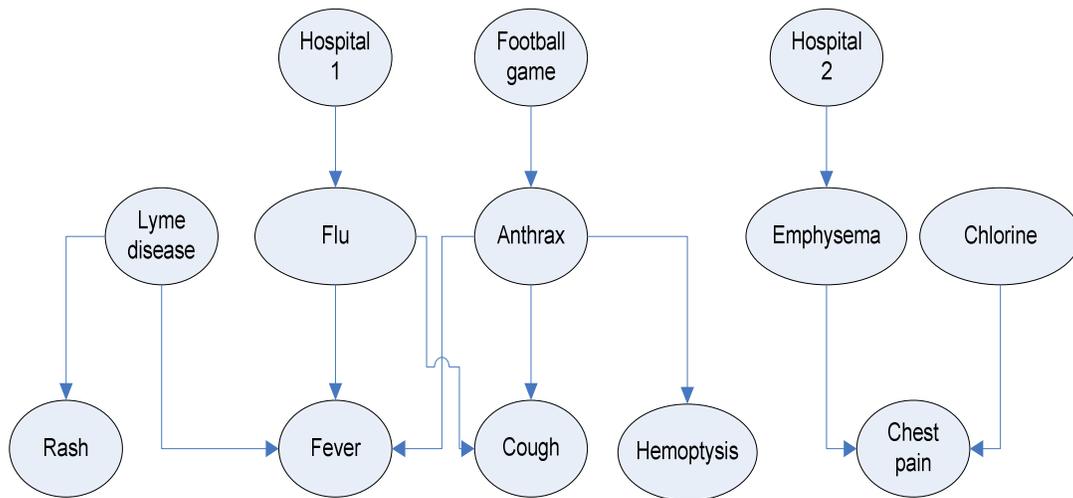

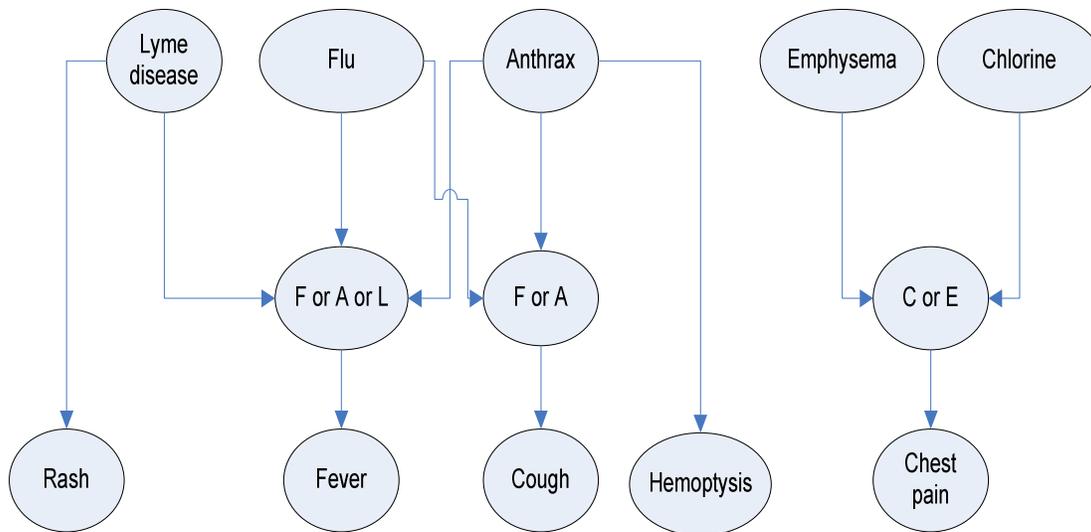**Figure 8 - Simple Bayesian network used for synthesizing data**

Figure 9 - CRE Model

Table 2 - RMSE values of learned networks compared to original networks

Number of nodes	Number of cases	RMSE
16	100	0.0573
16	1,000	0.0197
16	10,000	0.0117
42	10,000	0.118
97	10,000	0.279

5. Comparison to Other Diagnostic Approaches

In this section, we present a comparison of the CRE's diagnostic capabilities to those of various other existing diagnosis approaches.

Most of the following diagnostic reasoning approaches assume a static network and cannot handle unexpected information. However, the CRE makes use of unexpected information and attempts to note any relevancy of the unexpected information to any causes that are likely. This behavior alone makes the CRE a formidable diagnostic tool. Although the learning algorithm might not be novel, the use of a dynamic learning algorithm makes the CRE a unique diagnostic tool.

Perhaps the most powerful feature of the CRE is its simplicity. Each symptom requires only two numerical parameters; either the symptom is *explained* or it is not. This simplifies the amount of data a domain expert must enter and makes the models much simpler than in other approaches.

5.1. Noisy-Or Networks

In a noisy-or network, variable dependencies are modeled as binary valued noisy OR logic gates. Each child has a probability matrix that models its parent's probability of activating the child. Noisy-Or networks can supply the local distributions of symptoms that the CRE provides. Like the CRE, each symptom can be presented as a disjunction of its causes; however, in noisy-or networks, disjunctions are probabilistic as opposed to truth-functional. Each symptom has some probability, usually high, of not occurring when none of its causes are present. Every cause that is present decreases the probability of the symptom not occurring by a multiplicative factor.

A CRE model represents relationships between symptoms and *sufficient* causes -- any parent of a symptom is sufficient to explain that symptom, and additional causes do not increase its probability. (Such a model can also represent *necessary* causes by simply negating all variables: "X is necessary for Y" is equivalent to "not Y is sufficient for not X". Thus, if "Has Gas" and "Live Battery" are necessary causes for "Car Starts", then "Out of Gas" and "Dead Battery" are sufficient causes for "Car Won't Start".)

Since the CRE's approach assumes that each cause is sufficient to explain a symptom, the presence of additional causes has no effect on the distribution of the symptom, unlike a noisy-or network. The CRE's approach assumes that a symptom has two different distributions given the causes; it does not assume anything about what those distributions are.

Noisy-Or networks cannot be used to model the more complex causal relationships. They can only model a "noisy" form of causal sufficiency.

Furthermore, the number of parameters required for a noisy-or network is $O(M + M*N)$ whereas the CRE's model requires only $O(M + 2N)$ parameters.

5.2. Naïve-Bayes Networks

Many diagnostic reasoning tools use a Naïve-Bayes network for the disease model. A Naïve-Bayes network is a Bayesian network under the assumption of conditional-independence. Naïve-Bayes classifiers assume that the effect of a variable value on a given node is independent of the values of all other nodes. This assumption simplifies the computation and is in this sense considered to be "Naïve". The Naïve-Bayes model's root node ranges over all possible causes.

This assumption is fairly strong and is often not applicable. It is an effective approach when causes are mutually exclusive; however, it cannot model multiple causes being present in a single case. As mentioned in section 2, if a person has both a broken arm and flu-like symptoms, a Naïve-Bayes approach would fail to classify a cause since no cause has both a broken arm and flu-like symptoms. The CRE would be able to correctly identify two distinct causes: broken arm and flu.

A Naïve-Bayes model also has $O(M + M*N)$ numerical parameters compared to the CRE's $O(M + 2*N)$. Thus, Naïve-Bayes networks are much more complicated to enter than CRE models.

5.3. Non-restricted Bayesian networks

Non-restricted Bayesian network models differ from the CRE's Bayesian network models in that the symptom nodes are not restricted to having just one parent. (In the CRE's model, each symptom may have more than one cause, but the symptom's parent node is a disjunction node of all of the possible causes.)

Non-restricted Bayesian networks can support more complex causal patterns than the CRE's models. One such causal pattern is multiple causes having an additive effect on the probability of a symptom. Suppose that the symptom "Violent Behavior" depends on the causes "Childhood Abuse" and "Brain Injury". Neither cause alone is sufficient or necessary to explain "Violent Behavior," but when both causes are present, "Violent Behavior" is explained. Regular Bayesian networks can support this pattern while the CRE cannot. Another causal pattern that regular Bayesian networks can support while the CRE's models cannot is one cause being blocked by another cause. For example, suppose that "Death" depends on "Smallpox contraction" and "Smallpox immunization", where "Smallpox immunization" takes away most or all of the explanatory power of "Smallpox contraction". A regular Bayesian network can model this example whereas the CRE's model cannot.

Although regular non-restricted Bayesian network models are more expressive, the extra expressiveness comes at an unfortunate cost: numerical parameter complexity. The number of priors grows with the number of causes linked to a symptom, which can quickly balloon to unmanageable sizes. The number of numerical parameters for the network can be as bad as $O(M + M^2 * N)$ whereas the CRE's model only requires $O(M + 2N)$. The complexity of non-restricted Bayesian network models makes them potentially unusable for most modelers, as the modelers simply cannot accurately enter such massive amounts of values. Furthermore, as the complexity of the model increases, the likelihood of a domain expert making faulty assumptions or mistakes in the specification increases.

5.4. Recognition Approaches

Recognition approaches are those where one tries to simply recognize the causes by the symptoms rather than doing probabilistic inference to infer the posterior distribution over all causes. One might have a cause recognizer for each individual cause and then call each recognizer on each datum to return a degree of match. The main issue with doing this is how to interpret the degree of match for several causes on a single datum. For example, if the similarity function is the inner product, then the degree of match is simply the number of symptoms of the cause that are present (optionally normalized by the total number of symptoms of the cause). If Influenza had symptoms {"cough", "fever", "runny nose"} and Anthrax had symptoms {"cough", "fever", "hemoptysis"} then both would match the evidence with a degree of 2 (or 2/3), corresponding with the evidence {"cough", "fever"}. Of course, suppose we know a priori that Influenza is much more probable; then the evidence hardly supports a conclusion of "Anthrax" because it equally well supports a more probable diagnosis of Influenza. This effect of one cause on the probability of another is called "explaining away". However, if we observed {"fever", "hemoptysis"}, we would have a much better justification for believing Anthrax was present. In a pattern-matching system, there is no principled way of resolving this issue and thus the CRE's probabilistic approach is much more accurate.

6. Health Domain Application Discussion

In section 2, we presented an overview of the entire syndromic surveillance system as applied to epidemiology. This section discusses some of the requirements of the patient case monitoring process, privacy issues, and practicality.

6.1. Requirements

For Quantum Leap's syndromic surveillance system to maximally benefit health officials, several health facilities should connect to the system and provide as much information as legally able. The HIPAA allows this sort of collaboration under tight restrictions and controls. Each health facility can send batches of anonymous patient records from their data warehouse, or stream the cases in real time as the patients' records are recorded.

The CRE uses a very simplistic patient data input language, but it can easily be extended to accept any type of file format, such as XML, HIPAA Electronic Data Interchange, and etc. Algorithms exist to interpret free-form text, such as latent semantic indexing or Naïve-Bayes classification, so even free-form text patient records could be submitted to the CRE.

The actual data record that the CRE expects is basically a list of present symptoms, the date of the patient record, and any other non-symptom data, such as "attended football game" or "ate at Joe's" that the health facility records.

An informed healthcare domain expert would be the ideal person who would enter the probabilities for the CRE monitor model; they would be familiar with symptom occurrences in the population and could distinguish between anomalies and false positives.

6.2. Practicality

To take full advantage of the learning capabilities of the CRE, Quantum Leap Innovations envisions having access to a wide array of interview questions and other miscellaneous information about individual patients. We realize that much of this information is simply not recorded in today's healthcare facilities. For example, it's not very likely that many patients are going to have been asked where he or she has eaten in the past three or four days, or which major events he or she has attended. This kind of information is usually not recorded until after a major outbreak is imminent (and thus much damage already caused). However, the CRE can still use predefined environmental information, such as pollen count and weather, as well as any other attribute that the CRE is given which doesn't necessarily come from a health facility when making its diagnoses. For example, the CRE could be told that a concert attended by 9,000 people took place in a particular zip code. If many patient records have the same zip code as the concert's venue, and these particular patients share a common illness, then the CRE might positively associate the concert with the illness.

Quantum Leap Innovations envisions a much more complete interview process for patients in the relatively near future. Homeland security improvements, a general desire for more complete information, and other such demands might warrant more thorough standardized data collection throughout the healthcare industry. When those standards come, the CRE will already be prepared to analyze such data and provide more awareness for health officials.

Furthermore, the accuracy of the CRE is only as good as the network that it uses. A healthcare domain expert would be the ideal person to edit the models. Alternatively, network structure learning techniques could be applied to learn networks from data. However, accuracy is often sacrificed in these approaches.

6.3. Privacy Concerns

Privacy is a great concern in healthcare. Strict HIPAA regulations dictate what can and cannot be shared, what requirements data systems must pass, which protocols may be used, who can see the shared data, etc.

Quantum Leap Innovations envisions the user of our syndromic surveillance system to be a top government health official who can only access the system from a secure location after passing strict identification checks. This top official is the ideal person to use the system in order to get an overview of the types of symptoms that each health facility in the region is presenting, the list of most probable illnesses in the region, and to be alerted when probabilities of particular illnesses are abnormally high. The CRE is not incapable of false-positives; it takes a well-informed health official to interpret the findings of the CRE.

Because the intended user of the system is a top health official, and because the use of the system is to monitor overall awareness of disease in a region, use of the system falls under the general health oversight activities authorized by law to prevent or control disease. Therefore, patient information can legally be sent to the CRE without fear of breach of privacy.

7. Summary

In this paper, we presented an overview of Quantum Leap Innovations' syndromic surveillance system. The system uses an innovative *explanation-based* approach for modeling diseases and their symptoms. The system is also capable of learning non-symptom predictors of causes, such as "a majority of the patients who exhibit

anthrax attended a football game." We explained how the Causal Reasoning Engine (CRE) operates; in particular, we outlined the explanation-based framework for the Bayesian network data model that the CRE uses, the inference algorithm, and the learning algorithm. We have shown that the CRE can be very accurate with synthetic patient data. We then compared the CRE against several other diagnostic approaches and found that the main benefits of the CRE are its data model simplicity and its dynamic learning capability. We concluded the paper with a brief discussion on the requirements, practicality, and privacy concerns of using the CRE in the healthcare domain.

8. References

[1] E. Charniak. "Bayesian Networks Without Tears." *Proceedings of the 13th National Conference on Artificial Intelligence*, Portland, 1991, pp. 50-63.

[2] G. F. Cooper and E. Herskovits. "A Bayesian method for the induction of probabilistic networks from data." *Machine Learning 9*(4), 1995, pp. 309-347.

[3] H'ajek, P., Havr'anek, T., and Jirousek, R., *Uncertain Information Processing in Expert Systems*, CRC Press, Inc., Boca Raton, 1992.

[4] David Heckerman. "A Tutorial on Learning with Bayesian Networks." *Technical Report MSR-TR-95-06*, Microsoft Research, Redmond, Washington, 1995.

[5] E. Horvitz. "Automated reasoning for biology and medicine." In *R. Fortuner*, Johns Hopkins University Press, Baltimore, 1993.

[6] W. H. Hsu, H. Guo, B. B. Perry, and J. A. Stilson. "A Permutation Genetic Algorithm for Variable Ordering in Learning Bayesian Networks from Data." In *Proceedings of the Genetic and Evolutionary Computation Conference* (GECCO-2002), New York, NY, 2002.

[7] Joachims, T. "Text Categorization with Support Vector Machines: Learning with Many Relevant Features." In *Machine Learning: ECML-98*, 1998, pp. 137--142.

[8] S. L. Lauritzen and D. J. Spiegelhalter. "Local computations with probabilities on graphical structures and their application to expert systems." *Royal Statistic Society 50*, 1988, pp. 154-227.

[9] Peter McCullagh. "Quotient spaces and statistical models." *Canadian J. Statist 27*, 1999, pp. 447-456.

[10] Y. Peng and J. A. Reggia. "A probabilistic causal model for diagnostic problem solving, part one: integrating symbolic causal inference with numeric probabilistic inference." *IEEE Transactions on Systems, Man and Cybernetics, v.17 n.2*, 1987, pp.146-162.

[11] M.A. Shwe *et. al*. "Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base". In *Medical Informatics*, Stanford Press, 1991.

[12] P. Sykacek. "Bayesian Inference for Reliable Biomedical Signal Processing." PhD thesis, *Institute for Medical Cybernetics and Artificial Intelligence*, 2000.