International Journal of Advanced Computer Science, Vol. 3, No. 7, Pp. 322-329, Jul., **2013.**

Exploratory Analysis in Time-Varying Data Sets: a Healthcare Network Application

Narine Manukyan, Margaret J. Eppstein, Jeffrey D. Horbar, Kathleen A. Leahy, Michael J. Kenny, Shreya Mukherjee, and Donna M. Rizzo

Manuscript
Received: *10,Mar.,2013*
Revised: *30,Mar.,2013*
Accepted: *24,Apr.,2013*
Published: *15,Jun.,2013*

Keywords
Artificial intelligence, genetic algorithm, knowledge discovery, pattern recognition,

***Abstract*— We introduce a new method for exploratory analysis of large data sets with time-varying features, where the aim is to automatically discover novel relationships between features (over some time period) that are predictive of any of a number of time-varying outcomes (over some other time period). Using a genetic algorithm, we co-evolve (i) a subset of predictive features, (ii) which attribute will be predicted (iii) the time period over which to assess the predictive features, and (iv) the time period over which to assess the predicted attribute. After validating the method on 15 synthetic test problems, we used the approach for exploratory analysis of a large healthcare network data set. We discovered a strong association, with 100% sensitivity, between hospital participation in multi-institutional quality improvement collaboratives during or before 2002, and changes in the risk-adjusted rates of mortality and morbidity observed after a 1-2 year lag. The proposed approach is a potentially powerful and general tool for exploratory analysis of a wide range of time-series data sets.**

1. Introduction

The rapid growth of technology has facilitated widespread collection and storage of vast amounts of time-varying data (e.g. [1]). This data undoubtedly contains a wealth of potentially valuable information regarding relationships between various time-varying features and outcomes. However, the very size of these databases is an impediment to knowledge discovery, creating a need for automated exploratory analysis tools. Over recent decades the scientific community has expressed an increasing interest in knowledge discovery in large databases [4], and some exciting progress has been made in this area. For example, a new method for automated discovery of non-parametric associations between pairs of variables was recently proposed and was shown to discover a wide range of functional and non-functional associations [25]. However, it would be computationally prohibitive to extend this method for discovering multivariate associations.

In general, large data sets include many features, only a few of which may interact, potentially in very nonlinear ways, resulting in some association with other outcome features in the data. Thus, identifying the relevant features is a critical aspect of knowledge discovery in large data sets. Evolutionary algorithms provide a particularly attractive approach for feature selection, because they require no pre-determination of the number of features in the optimal feature subset. Genetic algorithms (GAs), in particular, have been widely and successfully applied for feature selection in a variety of problems (e.g., [3], [16], [19], [22], [24]).

However, identifying the correct set of features is only part of the challenge in exploratory data analysis. For example, one may also need to identify which outcome(s) those features are associated with. Indeed, many distinct complex relationships between different feature subsets and different predicted outcomes may be present in the same data set, waiting to be discovered. The problem is compounded with time-series data sets, where there may be time-dependent aspects to the association. There are domain specific solutions that can address this problem for specific tasks ([21], [29], [14]), but developing a general tool that can find novel multivariate associations between features in time varying data for arbitrary problems is a much bigger challenge.

Our motivation in addressing this problem stems from a particular application in the healthcare domain. The Vermont Oxford Network (VON) is a non-profit corporation dedicated to the mission of improving the quality and safety of medical care for newborn infants and their families through a coordinated program of research, education, and networking of neonatal intensive care units (NICUs) at hospitals around the world. Since its inception in 1990, the VON has maintained databases with detailed information about hospital characteristics, treatments, and outcomes for all of the very low birth weight (VLBW) infants (birth weight under 1500 grams) treated at member hospitals around the world (e.g., [2], [6]-[9], [18], [27]-[28], [30]). These data are used to quantify treatment practices and risk-adjusted morbidity and mortality for VLBW infants treated at NICUs in the VON. While they account for only one percent of births, VLBW infants account for half of infant deaths in the US each year [17]. A major and consistent finding of previous VON database analysis is the dramatic variation in outcomes among NICUs, even after adjusting for differences in case mix among units [6]-[7], [9], [18], [27]-[28], [30]. Differences in hospitals and unit characteristics such as teaching status, volume or NICU

This work was supported by the NIH Eunice Kennedy Shriver National Institute of Child Health & Human Development award 1R21HD068296.

level also fail to explain the large discrepancies in health outcomes [27]. We hypothesize that differences in VON-sponsored activities designed to improve healthcare practices may account for some of these unexplained discrepancies in patient outcomes in VON member hospitals. Of particular interest are VON-sponsored team quality improvement collaboratives, in which interdisciplinary teams from multiple institutions work together to identify, test, implement, and report on innovative evidence-based treatment strategies [10]-[13], [20], [23], [26]. In order to explore this hypothesis, we have assembled a large database of VON-sponsored interactions among member hospitals between 1995 and 2010. We seek to discover novel multivariate associations between time-varying VON-sponsored hospital interactions and patient outcomes. Discovering such relationships, if they exist, could potentially have widespread application to managing collaborative healthcare networks, such as the VON, that seeks to innovate and spread quality improvement practices between hospitals around the world.

In this paper we propose a genetic algorithm for co-evolving four important aspects of exploratory multivariate time-series analysis: (i) a subset of features to be used as input into some sort of statistical predictor (such as a classifier or regression analysis), (ii) which attribute we can best predict from these features, (iii) a dividing year that partitions the time-series, and (iv) a time lag to be added to the dividing year. Fitness is determined by seeing how well the values of the selected features before the dividing year can be used to predict changes in the selected attribute after the dividing year + lag. In this proof-of-concept study, we first validate the approach using synthetic data, and then apply the method to a subset of the VON data.

2. Methods

We propose a new method that uses a Genetic Algorithm (GA) to co-evolve the inputs and output to a fitness function based on a statistical predictor, seeking causal associations in large time-varying data sets with multiple input features and potential prediction attributes. In this paper we focus on classification predictors, although one could easily employ other types of predictors (such as multiple regression). For brevity, we refer to this method as GAMET (Genetic Algorithm for Multivariate Exploration of Time-varying data).

In the general problem, the hypothesis is that there is some sort of causal relationship between a set of features that affect the value of some outcome attribute over some time period in the future. For example, we hypothesize that interactions between hospitals in the Vermont Oxford Network (e.g., as evidenced by participation in multi-institutional team quality improvement collaboratives, co-authored publications, case study presentations, and attendance at annual meetings) can influence future health outcomes at these hospitals (e.g., probability of patient death, infection, or other morbidity). However, even assuming this causal influence is true, there are doubtless a number of other (non-VON related) influences that affect the healthcare outcomes at these hospitals (see Fig. 1, top). Thus, it is not realistic to expect that we will be able to predict healthcare outcomes based on knowledge of the VON interactions alone. Furthermore, the number of hospitals that actively participate in the more intense types of VON interactions (such as team collaboratives and co-authorship on scientific studies) is much smaller than the number of member hospitals that don't actively participate, so these classes are very imbalanced. Consequently, for this application we seek to do the prediction in the opposite direction (see Fig. 1, bottom). That is, given knowledge of time-varying healthcare outcomes at various hospitals, can we predict which hospitals actively participated in VON-sponsored interactions (even if we cannot determine which hospitals did *not* actively participate)?

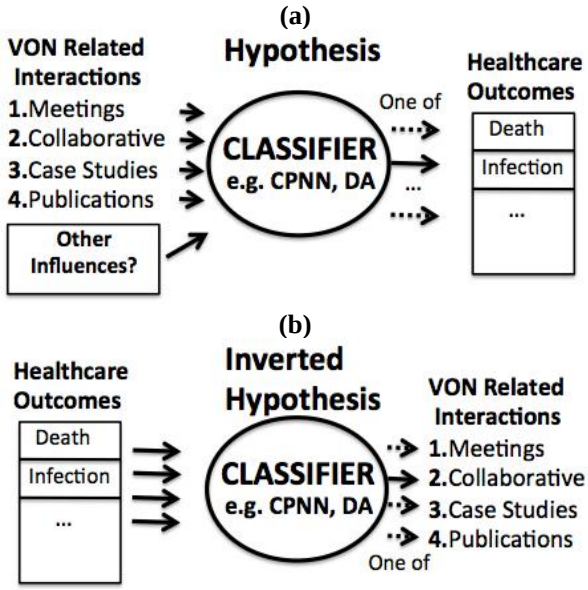


Fig. 1. a) Hypothesis of causality. b) Inverted hypothesis tested by the classifier.

In a problem like this where we hope to infer causal relationships, it is important to take the time-varying nature of the data into account. For example, if a hospital participates in a team collaborative designed to reduce infection rates, then one would hope to see infection rates decrease at that hospital at some time in the future, although there may be a time lag between when the collaborative activity took place and when measurable changes in infection rate can be detected. We handle this time component by looking at the change in health outcomes, averaged before and after a given points in time, relative to some "dividing year" and possibly with an intervening time lag, and see if we can use this to predict the presumed causal attribute (level of participation in VON-sponsored activities) before the dividing year (as illustrated in Fig. 2 for a dividing year of 2004 and a time lag of 2 years).

Thus, we desire to co-estimate three types of information simultaneously: which features to use as input to the classifier, what dividing year and lag to use in processing the time-series data, and which attribute to try to predict. The binary chromosomes used in GAMET thus include genes associated with each of these three parts (see

Fig. 3).

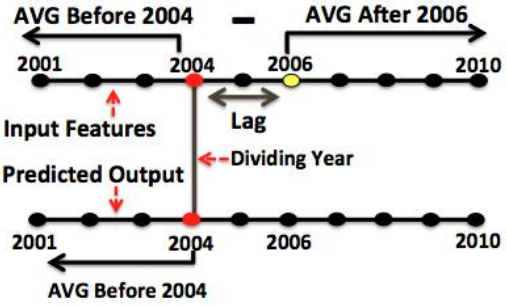
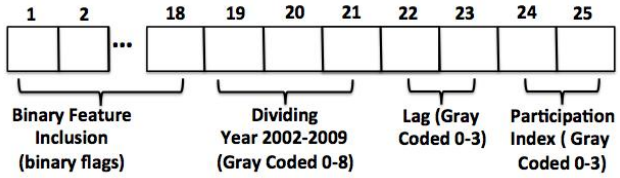


Fig. 2. Information is extracted and aggregated from the time-series data relative to a dividing year (2004, in this example) and lag (2 years, in this example).

For feature selection, we are using binary flags that indicate whether the given feature is included in the final features subset or not. To evolve the time series component we evolve the dividing year and lag, both of which are represented as gray-coded integers in the chromosome. Finally, a gray-coded "participation index" specifying which single attribute (from a list of potentially predicted attributes) is to be predicted.



To calculate the fitness of an individual, we first process the data for the included features, using the dividing year and lag as described above (labeled as time series extraction and aggregation in Fig. 4). We then pass these time-processed features as inputs to the classifier, and compare the predicted classes to class outcomes of the attribute specified by the participation index, averaged prior to the dividing year. The data is divided into training and testing sets, using a parameter to control the percentage of the data used for training (80\% for our experiments). We use Latin hypercube sampling to ensure adequate distribution of samples in the training and testing sets for this highly unbalanced classification problem. After the training phase we evaluate the classifier performance using the confusion matrix, which shows the number of correctly and incorrectly classified samples in each class (see Fig. 4).

For our VON data set we are using two classes for all predicted outputs: a "positive" (*P*) classification means that we are predicting that a particular hospital participated in the specified activity before the dividing year, whereas a "negative" (*N*) classification means we are predicting the hospitals that didn't participate in the specified activity. The fitness is calculated using the following formula:

$$fitness = \frac{FP}{(FP+TN)} + \frac{FN}{(FN+TP)} + \frac{(FP+FN)}{2(TP+FP+TN+TP)}$$

where *FP* is the number of false positives, *TP* is the number of true positives, *FN* is the number of false negatives and *TN* is the number of true negatives. The first two terms represent the proportion of samples in each class that were classified incorrectly, whereas the last term is the proportion of the overall misclassified samples. This fitness function thus takes into consideration both the overall prediction rate and the individual class prediction rates (the latter is helpful for unbalanced classes). We would like to note that there is some stochasticity involved in the calculation of the fitness function (due to the Latin hypercube sampling and any stochasticity possibly associated with classifier), which can result in slightly different fitness values being evaluated for the same chromosome on different occasions.

We employ two different classifiers in this paper. For the synthetically generated data set, we were able to use a naïve Bayes quadratic discriminant analysis (DA) classifier. However, because the VON data set violated so many assumptions of the DA, for this application we used a non-parameter counter-propagation artificial neural network (CPNN) classifier [5]. The overall architecture of the approach, illustrated for the VON data set, is shown in Figure 4, where the co-evolved entities are indicated in red.

TABLE 1.
GA PARAMETERS USED IN THIS STUDY.

Parameter	Value
Population Type	bitstring
Population Size	500
Generations	100
Crossover Fcn	scattered
Mutation Fcn	{uniform, p = 0.04}
Crossover Fraction	0.8
Elite Count	1
Selection Fcn	{tournament, size = 4}

TABLE 2.
CPNN PARAMETERS USED IN THIS STUDY.

Parameter	Value
Learning rate	0.7
Bias	0.1
Mean Square Error to stop training	0.001

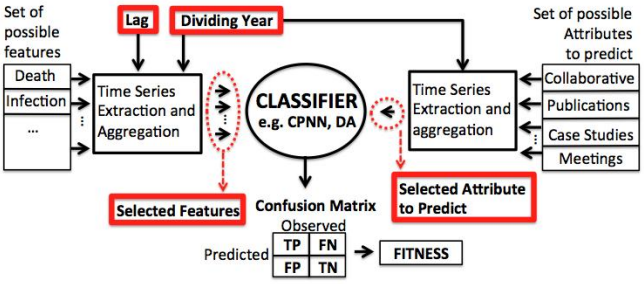


Fig. 3. Overall architecture of the approach, illustrated for use with the VON data set. Items outlined in red are co-evolved by GAMET.

3. Experiments

A. Synthetic data

In order to test the capability of GAMET for co-evolving correct feature sets of varying sizes, attribute to predict, year, and lag, we created synthetic data sets for 15 test problems, as follows.

We first generated 5 random "true" combinations of dividing year (selected uniformly from 2002..2009), lag (selected uniformly from 0..3 years), and index for the attribute to predict (selected uniformly from 0..3). We next generated 15 random multivariate expression trees in 3 sets of varying levels of difficulty; 5 expressions contained 2 variables, 5 contained 3 variables, and 5 contained 8 variables. For each of these 15 test problems, we generated a 300×100 matrix of uniformly distributed random real numbers in the range (0,1), representing synthetic data for 300 cases, each with 100 feature variables (e.g., synthetic values for 100 heath outcomes at 300 hospitals). The expression trees were generated using a function set of {+,-,*,*exp*, <, >, =} and were constructed so as to return binary class outcomes such that at most 2/3 of the outcomes had the same value.

The expression trees were generated using a terminal set comprising 100 distinct real-valued variables (corresponding to the 100 feature columns in the synthetic data sets), as well as integer constants {1,2,3}. Each set of 5 expression trees with the same number of variables was associated with the set of 5 combinations of year, lag, and index of the attribute to be predicted, created as described above. The resulting specifications for these 15 test problems are outlined in Table 4, column 2.

For each of the 15 random problems, we then created a synthetic 300×128 outcomes matrix, where the 128 columns in this matrix correspond to all combinations of 4 possible attributes to predict (e.g., synthetic values for participation in 4 types of VON-related interactions), 8 possible dividing years, and 4 possible lags. All 96 columns in the outcomes matrix corresponding to the 3 incorrect attributes to predict (for all 8 dividing years and all 4 possible lags) were initialized to uniform random binary class outcomes. However, the remaining 32 columns associated with the correct attribute to predict (for all 8 dividing years and all 4 lags), were initialized to the "true" predicted binary class outcomes associated with the 15 random problems.

These "true" outcomes were calculated by evaluating the expression trees using the columns from the synthetic data matrix corresponding to the feature variables in the expression trees. Lastly, we added noise to 31 of these 32 columns, proportional to the Hamming distance (H) between the 5-bit gray-coded sequences representing their dividing years (3-bits) and lags (2-bits) and the 5-bit vector representing the "true" dividing year and lag. Specifically, we overwrote $30 \times H$ bits in each of these columns with random binary values.

This algorithm thus creates a synthetic data set that has known associations between a subset of feature vectors and one of the attributes to predict. By design this relationship has a perfect association when the dividing year and the lag exactly match the "true" target values, but the level of added random noise increases as the dividing year and lag get farther from the target values, as one might expect to see in real time series data.

B. VON data set

1) VON-related interactions: We assimilated a large database of VON-facilitated interactions between hospitals for the years 1995 through 2010. During this time, the VON network grew from around 100 hospitals to 850 hospitals. Here, we report on four specific types of VON-sponsored interactions: (i) participation in VON annual meetings; (ii) preparation of case studies that were presented at VON meetings; (iii) participation in VON-sponsored team collaboratives, which are 2-year long team projects where multidisciplinary quality improvement teams from participating hospitals work together to identify and implement potentially better health practices, and (iv) co-authorship on publications resulting from VON-related activities. It should be noted that the level of participation in these four types of interactions is quite variable, with many member hospitals not actively participating in any of these types of VON-sponsored interactions. On average, in any given year only {53.2%, 11.5%, 8.3%, and 21.5%} of all VON member hospitals participated in these four types of activities, respectively. Thus, although we have quantitative information on the amount of participation in each of these activities, for this preliminary study we have binarized the annual participation in these four types of interactions for each member hospital. Our initial goal is to see if changes in health outcomes are associated with any level of participation, in any of these types of VON-facilitated interactions. I.e., these four types of VON-sponsored interactions comprise four potential "attributes to predict", where the predicted values are the binary classes representing participation or non-participation. After the creation of this database, all identifying information was removed, to ensure member hospital privacy.

2) VON health outcomes: The VON maintains an extensive database of over 200 types of annual health outcomes at all member hospitals. In this preliminary study, we are focusing on only 18 risk-adjusted measures (see table 3) over the period 2001 through 2010, representing the health outcomes of over half a million VLBW infants. The risk adjusted outcome measures are recorded as observed divided by expected values of the outcome, where expected values vary with the number of patients at the hospital. These particular features were identified by VON staff as ones they thought had strong potential to have been impacted by VON-related interactions, based on collaborative studies they had sponsored during this time period. I.e., we want to see if subsets of these 18 real-valued features can be used to classify individual hospitals as participants or non-participants in any of the 4 types of VON-sponsored interactions described in Section 3.B.1. The distribution of health outcomes in the real VON data violates assumptions of normality and independence. Preliminary testing, using the real VON health outcome features described in Section 3.B.2 with synthetically generated known associations to class outcomes, confirmed that the parametric DA classifier was not able to correctly classify known outcomes associated with these data,

whereas the non-parametric CPNN was. Thus, as mentioned previously, we used CPNN-based fitness in the co-evolutionary method applied to the VON data. The data provided by VON for use in this study was de-identified to protect the confidentiality of patients and hospitals and did not include personal patient or hospital identifiers. The protocol for this research was submitted by the Committees on Human Research at the University of Vermont and determined to be exempt from formal Committee review and approval.

TABLE 3. HEALTH OUTCOMES USED AS POSSIBLE FEATURES IN OUR ANALYSIS OF THE VON DATA.

#	Description
1	Any Late Infection
2	Chronic Lung Disease
3	Chronic Lung Disease before 33 Weeks
4	Coagulase Negative Staph
5	Mortality
6	Mortality or Morbidity
7	Fungal Infection
8	Intraventricular Hemorrhage
9	Mortality Excluding Early Deaths
10	Bacterial Pathogen after Day 3
11	Necrotizing Enterocolitis
12	Necrotizing Enterocolitis, where occurred
13	Nosocomial Infection
14	Pneumothorax
15	Cystic Periventricular Leukomalacia
16	Retinopathy of Prematurity
17	Severe Intraventricular Hemorrhage
18	Severe Retinopathy of Prematurity

C. Experimental design

For the 15 synthetic problems described in Section 3.A we ran 10 replicates of the GA, using the DA-based fitness function. For the actual VON data described in Section 3.B, we ran 10 replicates of the GA, using the CPNN-based fitness function. Because both the DA and the CPNN can still classify well even with a certain number of excess features given as inputs, we subsequently intersected the feature sets of the best individuals resulting from each of the 10 replicates. The results of these experiments are described in the following section.

4. Results

In all 10 replications of each of the fifteen 100-feature synthetic problems GAMET was able to correctly identify the dividing year, lag, which attribute to predict (labeled "output"), and all of the 2, 3, or 8 true features (see table 4, compare columns 2 and 3), using the DA-based fitness function. As the number of true features increased, the tendency of GAMET to return excess features also increased (see table 4, column 4), since the DA can accommodate excess features (but simply not give them much weight).

TABLE 4.
EXPERIMENTAL RESULTS ON THE 15 SYNTHETIC TEST PROBLEMS.

#	**True** year, lag, output, #features	**Found** year, lag, output, #true feat.	**#Feat.** mean± std	**Excess** #features found in ∩
1	2002, 2, 3, 2	2002, 2, 3, 2	46±4	0
2	2003, 1, 1, 2	2003, 1, 1, 2	45±5	0
3	2007, 0, 2, 2	2007, 0, 2, 2	46±6	0
4	2005, 2, 4, 2	2005, 2, 4, 2	44±4	0
5	2004, 1, 3, 2	2004, 1, 3, 2	45±5	1
6	2002, 2, 3, 3	2002, 2, 3, 3	48±5	0
7	2003, 1, 1, 3	2003, 1, 1, 3	47±6	1
8	2007, 0, 2, 3	2007, 0, 2, 3	47±6	1
9	2005, 2, 4, 3	2005, 2, 4, 3	48±6	1
10	2004, 1, 3, 3	2004, 1, 3, 3	47±5	1
11	2002, 2, 3, 8	2002, 2, 3, 8	49±7	3
12	2003, 1, 1, 8	2003, 1, 1, 8	48±6	1
13	2007, 0, 2, 8	2007, 0, 2, 8	52±8	3
14	2005, 2, 4, 8	2005, 2, 4, 8	50±7	2
15	2004, 1, 3, 8	2004, 1, 3, 8	51±8	1

However, the intersections of the feature sets in the 10 replications contained relatively few excess features (see table 4, column 5). These results demonstrate that the system is able to co-evolve the correct feature subsets, correct attribute to classify, correct dividing year, and correct lag in time-series data with known relationships between input features and attribute to classify.

On the VON data set, all 10 runs consistently returned a dividing year of 2002, and discovered that participation in VON-sponsored team collaboratives was the attribute that could most accurately be classified. In 7 of the 10 runs, the lag was determined to be 2 years, whereas in the remaining 3 runs the lag was determined to be 1 years. The health outcome features selected as input to the CPNN-based fitness function were also relatively consistent between the 10 runs (see Figure 5). However, since the CPNN can do robust predictions even when given a few excess inputs, we then searched for consensus in the selected features between the different runs.

In all cases, the CPNN was able to predict the "true positives" in the smaller class (participants) with 100% accuracy (i.e., based on the selected health outcomes, the CPNN could correctly predict which hospitals *had* participated in a VON-sponsored team collaborative during or before the dividing year) (see table 5, column 1). However, the classifier was not able to use the selected health outcomes to accurately predict the "true negatives" (hospitals that didn't participate in any VON-sponsored team collaboratives during or before the dividing year) (see table 5, column 2). In other words, the identified classifier has high sensitivity, but low specificity. We assessed the overall classification accuracy in prediction participation in a VON-sponsored team collaborative, using health outcome feature sets that included the top *n* in {4,8,9,11,12,13,16} most consistently selected features, based on the consensus features selected in {100%, 90%, 80%, 70%, 30%, 20%, 10%} of the replicates, respectively (i.e., those features whose frequency bars are at or above the horizontal dotted

lines in Figure 5), using a dividing year of 2002 and a lag of 2 years. We report the resulting percentage accuracies to the right of Figure 5. Four features (5, 6, 8, and 14) occurred in the selected features of all 10 replicates, but the highest prediction accuracy (33%) was obtained when using the 9 features (2, 5, 6, 7, 8, 9, 10, 11, 12, 14, and 17) that were found in at least 7 of the 10 replicates; this 9-feature set also coincides with the best single individual found in the 10 runs (see Figure 5, red asterisks). The confusion matrix for this individual is shown in Table 5. Note that differences in these percent accuracies only reflect the differences in the specificity of the classifier, since all had perfect sensitivity. Conversely, we also found that the overall classification accuracy dropped dramatically to only 16%-18% when predicting from any 3 of the top 4 features, indicating that all four of these are important predictive features.

TABLE 5. CONFUSION MATRIX FOR THE BEST INDIVIDUAL FOUND BY GAMET ON THE VON DATA. HERE, "PARTICIPANTS" REFERS THOSE HOSPITALS WHO PARTICIPATED IN VON-SPONSORED QUALITY IMPROVEMENT COLLABORATIVES DURING OR BEFORE TO 2002.

	Observed participants	Observed non-participants
Predicted participants	54	263
Predicted non-participants	0	91

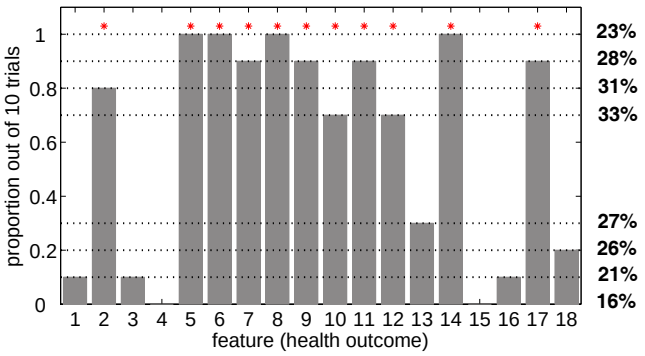


Fig. 4. Experimental results on the VON data set. The bars indicate the frequency with which each of the individual features was selected in 10 GAMET trials. The red asterisks near the top indicate the features selected in the single best individual.

5. Discussion and Conclusions

In this paper, we introduce a method for exploratory analysis of large data sets with time-varying features. Such data sets may contain information about many different potential relationships between features and outcomes. The aim is to automatically discover novel relationships between features (over some time period) that are predictive of any of a number of time-varying outcomes (over a different time period), but where the specific features, outcomes, and time periods are not known in advance. The application that motivated this study concerns exploratory analysis of a large healthcare network data set, comprising various types of time-varying interactions between subsets of hospitals in the Vermont Oxford Network (VON) and a variety of annual health outcomes at those hospitals.

The approach we take uses a Genetic Algorithm for Multivariate Exploration of Time-varying data (GAMET), in which we co-evolve (i) a subset of health outcomes, (ii) one of four types of VON-sponsored interactions to consider, (iii) the maximum ``dividing" year up to which we consider these VON-sponsored interactions, and (iv) how many years' time lag after the dividing year before which we assess changes in the health outcomes.

We first validated that GAMET was able to select the correct features, outcomes, dividing year, and lag in 15 synthetically designed problems with 2, 3, and 8 non-linearly interacting features with known associations to a specific binary-valued attribute. For these synthetic problems we assessed fitness based on the classification accuracy of a naive Bayes quadratic discriminant analysis classifier.

We then conducted preliminary exploration of the actual VON data set with 18 potential health outcome features, 4 types of VON-sponsored interactions, 8 possible dividing years, and 4 possible lags, representing a search space of over 33 million possible combinations of solutions. Due to the non-parametric nature of this actual data set, we assessed fitness based on the classification accuracy of non-parametric counterpropagation artificial neural network classifier. In addition, because the participation classes were highly unbalanced, we used Latin hypercube sampling to determine how to subdivide the data into appropriate training and testing sets.

The strongest association so far discovered by GAMET in the VON data set was between participation in VON-sponsored team quality improvement collaboratives during or before 2002, and changes in the risk-adjusted rates of mortality and morbidities including intraventrical hemorrhage and pneumothorax (collapsed lung) that were observed after 2003 or 2004, relative to these rates during or before 2002. Using changes in only 4 health outcomes selected by GAMET, we achieved 100% sensitivity in predicting which hospitals had participated in these collaboratives in 2002 or earlier. The identified lag of 1-2 years is a reasonable amount of time one would expect such changes in health practices to be implemented, and the health impacts of these changes observed, in the annually-updated health outcome records.

Our results on the VON data had relatively low specificity, however. The best individual returned by GAMET was still only able to achieve an overall classification accuracy of 33%, because the classifier was not able to accurately predict which hospitals had *not* participated in VON-sponsored interactions during or before 2002, based on the changes in health outcomes after 2003 or 2004. This result is actually to be expected, because there are many changes in healthcare practices at VON member hospitals that were independent of participation in VON-sponsored activities (and are consequently not in our database) that are expected to contribute to changes in health outcomes.

Having established proof-of-concept for the method, we now plan to apply GAMET to a more complete set of health outcome features and VON-sponsored interactions aimed at

stimulating improvements in healthcare practices. We will then more closely examine the specific nature of the relationships embedded in the associations discovered by GAMET. For example, we intend to use genetic programming (GP) for symbolic regression, using GAMET-selected features as variables in the GP terminal set (much as in [3]).

We can also envision many ways in which to improve the GAMET algorithm itself. For example, since the two types of classifiers employed here (the DA and the CPNN) can be trained to ignore excess features, the features selected by GAMET also contained excess features. Consequently, we applied a post-processing step to further reduce the final feature sets, by looking for features common to the selected feature sets from different GAMET replicates. Others have reported promising results in GA-based feature selection by actually embedding set intersection directly into the crossover operator [3], [15]. Although we found that strict set intersection was too aggressive in reducing features in the VON application, we plan to explore whether a probabilistic application of a "softer" form of multiset intersection (i.e., including all elements that occur in a certain percentage of parents) in multiparent crossover could help improve feature selection in GAMET, and therefore preclude the need for the post-processing of multiple replicates, as done here. In addition, the current version of GAMET only allows for the evolution of a single dividing year. We plan to explore whether it may prove more powerful to apply the evolved lag directly to the hospital-specific years of participation for selected types of VON interactions.

Although the proposed method was originally developed for analysis of the VON healthcare network data set described here, the GAMET approach is a potentially powerful and general tool for exploratory analysis of a wide range of time-series data sets. Future work will include the application of GAMET to time-vary problems in a variety of other domains (such as those in [1]).

Acknowledgment

This work was funded in part by the NIH Eunice Kennedy Shriver National Institute of Child Health and Human Development award 1R21HD068296. We thank the staff of the Vermont Oxford Network for their help in assembling the data sets.

References

[1] A. Banks. Cross-national time-series data archive 1815-2007. Databanks International, Jerusalem, Israel, 2008.

[2] I. Bernstein, J. Horbar, G. Badger, A. Ohlsson, A. Golan, et al. Morbidity and mortality among very-low-birth-weight neonates with intrauterine growth restriction. American journal of obstetrics and gynecology, 182(1):198–206, 2000.

[3] D. DeHaas, J. Craig, C. Rickert, P. Haake, K. Stor, and M. Eppstein. Feature selection and classification in noisy epistatic problems using a hybrid evolutionary approach. poster and published extended abstract accepted for Genetic and Evolutionary Computation Conference (GECCO), 2007.

[4] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth. From data mining to knowledge discovery in databases. AI magazine, 17(3):37, 1996.

[5] R. Hecht-Nielsen. Counterpropagation networks. Applied optics, 26(23):4979–4983, 1987.

[6] J. Horbar, G. Badger, J. Carpenter, A. Fanaroff, S. Kilpatrick, M. LaCorte, R. Phibbs, R. Soll, et al. Trends in mortality and morbidity for very low birth weight infants, 1991–1999. Pediatrics, 110(1):143, 2002.

[7] J. Horbar, G. Badger, E. Lewit, J. Rogowski, P. Shiono, et al. Hospital and patient characteristics associated with variation in 28-day mortality rates for very low birth weight infants. Pediatrics, 99(2):149, 1997.

[8] J. Horbar et al. The Vermont Oxford Neonatal Network: integrating research and clinical practice to improve the quality of medical care. In Seminars in perinatology, volume 19, page 124, 1995.

[9] J. Horbar and J. Lucey. Evaluation of neonatal intensive care technologies. The Future of Children, pages 139–161, 1995.

[10] J. Horbar, P. Plsek, and K. Leahy. Nic/q 2000: establishing habits for improvement in neonatal intensive care units. Pediatrics, 111(Supplement):e397, 2003.

[11] J. Horbar, P. Plsek, J. Schriefer, and K. Leahy. Evidence-based quality improvement in neonatal and perinatal medicine: the neonatal intensive care quality improvement collaborative experience. Pediatrics, 118(Supplement):S57, 2006.

[12] J. Horbar, J. Rogowski, P. Plsek, P. Delmore, W. Edwards, J. Hocker, A. Kantak, P. Lewallen, W. Lewis, E. Lewit, et al. Collaborative quality improvement for neonatal intensive care. Pediatrics, 107(1):14–22, 2001.

[13] J. Horbar, R. Soll, and W. Edwards. The vermont oxford network: a community of practice. Clin Perinatol, 37(1):29–47, 2010.

[14] T. Jenssen, W. Kuo, T. Stokke, and E. Hovig. Associations between gene expressions in breast cancer and patient survival. Human genetics, 111(4):411–420, 2002.

[15] Submitted to the 21st International GECCO Confference, 2012.

[16] H. Liu, J. Li, and L. Wong. A comparative study on feature selection and classification methods using gene expression profiles and proteomic patterns. Genome Informatics Series, pages 51–60, 2002.

[17] J. Martin, K. Kochanek, D. Strobino, B. Guyer, and M. MacDorman. Annual summary of vital statistics—2003. Pediatrics, 115(3):619, 2005.

[18] L. Morales, D. Staiger, J. Horbar, J. Carpenter, M. Kenny, J. Geppert, and J. Rogowski. Mortality among very low birthweight infants in hospitals serving minority populations. American journal of public health, 95(12):2206, 2005.

[19] I. Oh, J. Lee, and B. Moon. Hybrid genetic algorithms for feature selection. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 26(11):1424–1437, 2004.

[20] J. Øvretveit, P. Bate, P. Cleary, S. Cretin, D. Gustafson, K. McInnes, H. McLeod, T. Molfenter, P. Plsek, G. Robert, et al. Quality collaboratives: lessons from research. Quality and safety in health care, 11(4):345–351, 2002.

[21] N. Payne, M. Finkelstein, M. Liu, J. Kaempf,P. Sharek, and S. Olsen. Nicu practices and outcomes associated with 9 years of quality improvement collaboratives. Pediatrics, 125(3):437–446, 2010.

[22] F. Pernkopf and P. O'Leary. Feature selection for classification using genetic algorithms with a novel encoding. In Computer Analysis of Images and Patterns, pages 161–168. Springer, 2001.

[23] P. Plsek. Collaborating across organizational boundaries to improve the quality of care. American journal of infection control, 25(2):85–95, 1997.

[24] M. Raymer, W. Punch, E. Goodman, L. Kuhn, and A. Jain. Dimensionality reduction using genetic algorithms. Evolutionary Computation, IEEE Transactions on, 4(2):164–171, 2000.

[25] D. Reshef, Y. Reshef, H. Finucane, S. Grossman, G. McVean, P. Turnbaugh, E. Lander, M. Mitzenmacher, and P. Sabeti. Detecting novel associations in large data sets. science, 334(6062):1518–1524, 2011.

[26] J. Rogowski, J. Horbar, P. Plsek, L. Baker, J. Deterding, W. Edwards, J. Hocker, A. Kantak, P. Lewallen, W. Lewis, et al. Economic implications of neonatal intensive care unit collaborative quality improvement. Pediatrics, 107(1):23, 2001.

[27] J. Rogowski, J. Horbar, D. Staiger, M. Kenny, J. Carpenter, and J. Geppert. Indirect vs direct hospital quality indicators for very low-birth-weight infants. JAMA: the journal of the American Medical Association, 291(2):202, 2004.

[28] J. Rogowski, D. Staiger, and J. Horbar. Variations in the quality of care for very low birthweight infants: implications for policy. Health Affairs, 23(5):88–97, 2004.

[29] Z. Yu, F. Haghighat, B. Fung, and L. Zhou. A novel methodology for knowledge discovery through mining associations between building operational data. Energy and Buildings, 2011.

[30] J. Zupancic, D. Richardson, J. Horbar, J. Carpenter, S. Lee, G. Escobar, et al. Revalidation of the score for neonatal acute physiology in the vermont oxford network. Pediatrics, 119(1):e156–e163, 2007.

Narine Manukyan received a BS in Applied Mathematics and Informatics at Yerevan State University in Armenia in 2008, and her MS in Computer Science at the University of Vermont in 2011. She is currently working on her PhD in Computer Science at the University of Vermont. Her research interests include artificial intelligence, evolutionary computation, complex networks, data mining, machine learning and modeling complex systems.

Margaret J. Eppstein is Associate Professor and Chair of the Department of Computer Science at the University of Vermont, Burlington, VT, USA. She received the B.S. degree in zoology from Michigan State University in 1978, and the M.S. degree in computer science and the Ph.D. degree in environmental engineering from the University of Vermont, Burlington, in 1983 and 1997, respectively. She has been on the Computer Science Faculty at the University of Vermont since 1983 (Lecturer from 1983–2001; Research Assistant Professor from 1997–2002; Assistant Professor from 2002–2008; Associate Professor since 2008) and was the founding Director of the Vermont Complex Systems Center from 2006–2010. Dr. Eppstein's current research interests involve complex systems analysis and modeling in a wide variety of application domains, including biological, environmental, technological, and social systems.

Jeffrey D. Horbar, MD is a board-certified neonatologist and clinical scientist with extensive experience in clinical research and its application to the improvement of neonatal care. He is currently the Jerold F. Lucey Professor of Neonatal Medicine at the University of Vermont College of Medicine, Chief Executive and Scientific Officer of the Vermont Oxford Network, an Associate Editor of *Pediatrics*, and Co-Editor of the Neonatal Review Group of the Cochrane Collaboration. Dr. Horbar leads the Vermont Oxford Network Database which is used by over 950 NICUs to monitor and improve outcomes for high risk newborn infants. He is the Director of the Vermont Oxford Network NICQ and iNICQ Improvement Collaboratives in which multidisciplinary teams of health professionals from around the world work together under the guidance of expert faculty to improve the quality and safety of medical care for newborn infants and their families.

Michael J. Kenny, MS is a statistical consultant in the University of Vermont's Department of Medical Biostatistics. He received a biostatistics degree from the University of Vermont in 1995 and has collaborated with the researchers at the Vermont Oxford Network since 1998, with research interests including statistical reporting, risk-adjustment, and logistic regression methods.

Donna M. Rizzo, Associate professor in the School of Engineering, holds undergraduate degrees in Civil Engineering from the University of Connecticut, Fine Arts from the University of Florence, Italy and a M.S. and Ph.D. from the University of California, Irvine, and University of Vermont, respectively. Her research focuses on the development of new computational tools to improve the understanding of human-induced changes on natural systems and the way we make decisions about natural resources.