Modeling Causal Relationships in Sociocultural Systems Using Ensemble Methods

Amy Sliva, Scott Neal Reilly, David Blumstein, Steve Hookway and John Chamberlain

Abstract Analyzing the political, military, economic, social, information, and infrastructure (PMESII) effects in a sociocultural system requires models that capture the causal and predictive dynamics. However, given the complexity of PMESII factors and the diversity of available data sources, accurately modeling causal relationships requires incorporating multiple domains of study and a variety of analytic methods. In this paper, we present an ensemble approach to modeling causal relationships of sociocultural systems, applying insights from machine learning where ensembles consistently outperform individual approaches. We describe three different types of ensemble models and combinations and explore the application of this approach in experiments using both synthetic and real-world datasets.

Keywords Dynamic PMESII modeling · Sociocultural systems · Prediction · Causality · Ensemble learning · Machine learning

1 Introduction

When planning operations or designing policy interventions, military decision-makers and policy experts must have an understanding of the world around them, particularly the political, military, economic, social, information, and infrastructure

A. Sliva (✉) · S.N. Reilly · D. Blumstein · S. Hookway · J. Chamberlain
Charles River Analytics, 625 Mount Auburn Street, Cambridge, MA 02138, USA
e-mail: asliva@cra.com

S.N. Reilly
e-mail: snealreilly@cra.com

D. Blumstein
e-mail: dblumstein@cra.com

S. Hookway
e-mail: shookway@cra.com

J. Chamberlain
e-mail: jchamberlain@cra.com

(PMESII) implications on and effects of these policies. Analyzing these effects requires models that accurately capture the causal and predictive dynamics of a sociocultural system, enabling policy makers to make informed decisions as to the likely results of proposed interventions. Unfortunately, it is notoriously difficult to analyze causality rigorously, more so in complex sociocultural systems. Controlled experiments are rarely possible and capturing rich and complex causal and predictive relationships from available data is extremely challenging.

Making practical use of causal concepts challenges researchers in a wide range of fields, including philosophy, economics, statistics, artificial intelligence, cognitive science, physics, and social sciences. As a result, researchers have developed an equally wide range of approaches to recognize, represent, and reason with causal relationships to understand the past and predict the future. For example, Granger causality [1, 2] and convergent cross mapping [3] can detect possible causality from time series data, probabilistic graphical models provide tools for describing and reasoning about causality and uncertainty [4, 5], and recent advancements in artificial intelligence provide approaches for determining the directionality of causal linkages from observational data [6, 7]. Consequently, the diversity of factors and data sources that comprise sociocultural systems strongly suggests that modeling causal relationships of PMESII factors cannot be successful without incorporating aspects from multiple domains of study and insights from a variety of analytic methods.

In this paper, we present an ensemble modeling approach to understanding causality for PMESII modeling, treating the variety of causal analysis methods and data sources as an asset rather than a challenge. To mitigate the impacts of PMESII modeling complexity, we leverage computational data mining and data analysis techniques to accelerate the analysis and modeling process, recommend causal hypotheses that might be unanticipated by human experts, and help humans validate their theories against available data. Our approach applies insights from machine learning, where similar types of ensembles have been consistently able to outperform individual learning approaches, to (1) employ different analytics as appropriate to disparate data sources; (2) apply distinct analytic approaches to address discrete aspects of the sociocultural system; and (3) combine the results into a single coherent analysis.

The remainder of this paper is organized as follows. In the next section, we discuss different types of automated and semi-automated methods for identifying and modeling causal relationships from various types of sociocultural data. Section 3 presents our approach to ensemble causal models. In Sect. 4, we illustrate our approach by applying ensemble causal models to both synthetic and real-world data sets. First, we developed an ensemble combination of additive noise models to determine the causality from observational, non-temporal data. Second, we constructed an ensemble of diverse causal analysis techniques for time-series data that exploited metadata to improve performance. In both of these cases, the ensemble outperformed the individual approaches and produced a more accurate model of causal relationships. Third, we used a human-in-the-loop process to develop an

ensemble model for extracting causal relationships from real-world data about Anbar province in Iraq to form a more complete picture of the causal dynamics and PMESII effects in the region. Finally, in Sect. 5 we discuss the implications of this research and directions for future work.

2 Modeling Causality in Sociocultural Systems

Causality is extremely difficult to analyze, in part because there is no agreed upon definition of what makes a relationship "causal" or "predictive." Goldthorpe [8] identifies three possible definitions of causality that have been prominent in social science research: (1) causality as robust dependence, where causation implies association or correlation with strong predictive power; (2) causation as consequential manipulation where different consequences are observed when the causal variable is manipulated or varied; and (3) causality as a generative process, where the relationship is determined by some underlying social process or mechanism, which itself may be unobservable. Each of these definitions can provide insight into causal relationships in sociocultural models of PMESII factors, and their usefulness depends on the current decision-making context and the available information. For example, if a Commander must make an immediate decision for a transportation route, a causal model of the first type that predicts IED placement from the relative locations of roads, infrastructure, known weapons caches, and previous attacks, may be sufficient for route planning. However, for larger operations, such as reconstruction or humanitarian aid, a more detailed understanding of the causal relationship between local ethnic demographics and how this influences which water sources a household uses and how road closures or construction may impact their ability to acquire water, may be necessary. In this section, we review several automated methods for analyzing causal/predictive relationships and extracting them from data (or validating hypotheses against observational data). These techniques are drawn from a variety of social, natural, and computational domains and include a range of interpretations on the nature and definition of causality.

Many models adhere to the first definition of causality and describe relationships that are highly correlated with a strong predictive component. For example, because causality has an inherently temporal component—the effect does not precede the cause—techniques that exploit this can help analyze and extract causal models from time-series data. The most basic model of predictive power is simple correlation, such as the classic Pearson correlation measure; by adding a temporal offset to the correlation to mimic the sequencing of cause and effect, Pearson can successfully capture many causal/predictive relationships in time-series data. Granger causality [1, 2], a technique familiar to many social science researchers, takes strong predictive correlation further. Granger causality was originally introduced for time series analysis in economics and assumes a temporal order of cause and effect and that the causal variables provides information about the effect that can otherwise be unavailable.

Dynamic time warping is a similar approach that was developed by the authors for reasoning over arbitrary time series data [9, 10]. It was originally created for gait recognition where temporal offsets may be inconsistent over time (similar to how the distance between steps may change depending on whether a person is walking or running, an increase in crime might happen anywhere from six to 12 months after an uptick in unemployment). Convergent cross mapping (CCM) [3], a recent advance in biological studies, can identify even deeper, feedback relationships among time series, enabling scientists to model cyclic causal relationships, such as the feedback relationship between poverty and conflict known as the conflict trap [9, 11, 12]. These approaches can be very useful for identifying causal/predictive relationships in time-series datasets, for example validating models relating drops in unemployment with successive increases in criminal activity. However, we have found that this temporal data only represents a small proportion of the available sociocultural data sets.

Capturing Goldthorpe's third definition, relatively recent advances in the uncertainty in artificial intelligence (UAI) community seek to model the structure of some underlying process that is responsible for generating observed outcomes, capturing the direct and indirect causal relationships that exist between PMESII factors. For example, rather than simply indicating there is a statistically strong relationship between unemployment and crime data, these approaches will explicitly model the causal chain from lack of jobs in the legitimate economy to unemployment to economic hardship to criminal behavior, capturing the entire process that links the observed rise in unemployment with the subsequent rise in crime. Unlike the causal/predictive methods described above, the observational data used to construct and validate these models can be non-temporal, such as survey data (e.g., household development indicators surveys) or cross-sectional records (e.g., election outcomes for different regions) often encountered when trying to understand various PMESII characteristics.

Graphical probabilistic models, such as Bayesian networks [5], have provided the foundation for work in causal analysis since Pearl's *Causality: Models, Reasoning, and Inference* [4]. These models have proven useful as formal tools for describing and reasoning about causality for many fields of inquiry (e.g., epidemiology, policy, and social sciences). Graphical models mimic naturalistic ways of thinking about causality by explicitly representing the *structure* of causal relationships. Figure 1 shows an example of a graphical model illustrating the causal links between economic factors and increases in criminal activity. With their basis

Fig. 1 Graphical causal model illustrating the causal chain from economic recession to increases in crime
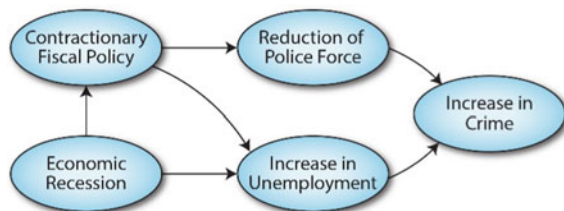
in formal probability theory, these models also facilitate rigorous mathematical reasoning and inference; however, the intuitive graphical structure means that it is often straightforward to reason visually over these graphs and make complex assessments about causal interactions in real-world situations quite easily and naturally, often without needing to resort to algebraic analysis. The graphical model structure can either be provided by human experts, or through automated structure learning techniques, such as heuristic search or constraint satisfaction (e.g., inductive causation (IC) [13] or the PC algorithm [14]). However, these automated structure learning methods depend on large quantities of data from all variables in the system, which in many cases may be unavailable or researchers have not identified all of the necessary variables present in the system. Mooij et al. [6] have developed a technique for determining the directionality of a causal relationship between only two variables that does not rely on any additional structural knowledge, enabling structured causal analysis from more limited observational data. For example, if we know that there is some causal relationship between crime and unemployment, we can determine the directionality of this link simply from observational data about these two variables, rather than requiring knowledge of other factors in the graph. This approach assumes that there is noise in any causal relationship and that by evaluating the effects of some standard noise models on causal relationships, it is possible to tease apart cause and effect in some situations. This has proven effective in many practical cases (see Sect. 4).

The wide range of methods described above—by no means an exhaustive list— make clear that is no single causality analysis technique that addresses all of the challenges and complexities of modeling PMESII factors in sociocultural systems. Rather, social scientists have at their disposal a suite of causal/predictive modeling approaches when dealing with observational data, each with strengths and weaknesses. For example, time-series analysis techniques can create useful models, but may draw superficial or over-generalized conclusions; probabilistic graphical models provide intuitive views of causal relationships, but require a priori identification of the important variables; and additive noise models can recognize causal structure between two variables, but cannot account for confounding relationships. Having these analytic techniques available allows scientists to develop more robust models of causal relationships in complex systems. However, we believe the real strength of this analysis suite lies not in finding the perfect technique for a particular system, but in combining these approaches in unique ways to produce a more comprehensive model [9, 10].

3   Ensemble Methods for Causal Analysis

Ensemble reasoning has proven extremely useful in the computational modeling and the machine learning communities to provide capabilities beyond those provide by any individual technique. For instance, the DARPA ICEWS program was successful at predicting nation-state instability by combining a variety of

independent predictive models and combing the results into a single prediction that was more accurate than any individual model provided [15]. Also, when using machine learning to, for instance, learn a classifier, it is possible to create ensembles that combine various learning techniques (e.g., support vector machines, neural networks, naïve Bayes classifiers, combinations of decision trees)—each with its own strengths and weaknesses—to provide better results than any individual approach [16]. We extend this concept to the domain of causal reasoning to create multiple causality-analysis techniques, each with its own strengths and weaknesses, that more thoroughly and accurately model causality in real-world systems.

3.1 Constructing Ensembles

There are many possible structures for combining multiple causal reasoning methods into ensembles and for joining multiple ensembles into larger ensembles. We call the building blocks of our approach data ensembles, chain ensembles, technique ensembles, and nested ensembles. Figure 2 shows these graphically.

Data ensembles. There are qualitatively different types of causal relationships to reason about (e.g., causality when there are hidden variables, non-deterministic causality) and qualitatively different types of available data to learn from (e.g., continuous, categorical, noisy). Just as different machine learning techniques can learn different kinds of things from different kinds of data, we need different causal analysis techniques to extract causal models based on the nature of the causal relationship present and the types of available data. For instance, Granger causality works on temporal data; PC works on non-temporal data.

Chain ensembles. Most causal analysis techniques are very specific about what kind of knowledge they extract, so they need to be composed with algorithms that perform other types of reasoning to provide the necessary analysis. For instance, correlation analysis (with appropriate human supervision) can be used to find pairs of likely causal variables, but not the direction of the causation. Then additive-noise methods (e.g., those described in [6, 7]) can identify the direction of the causation given two causal parameters, but those parameters are not identified. By combining the two, we provide the ability to find pairs of cause-effect parameters in data.

Technique ensembles. There are different ways to extract similar kinds of results from similar kinds of data, combining the results in ways that produce better results than any individual method can provide. For instance, Granger causality, dynamic time warping, Pearson correlation, and offset-correlation are all mechanisms to find possible causal links in temporal data [9, 10], though all are known to have limitations. By combining the results using mechanisms like those used in ensemble machine learning (e.g., various weighted voting schemes), we can provide an ensemble approach that is more accurate than any of the individual methods.
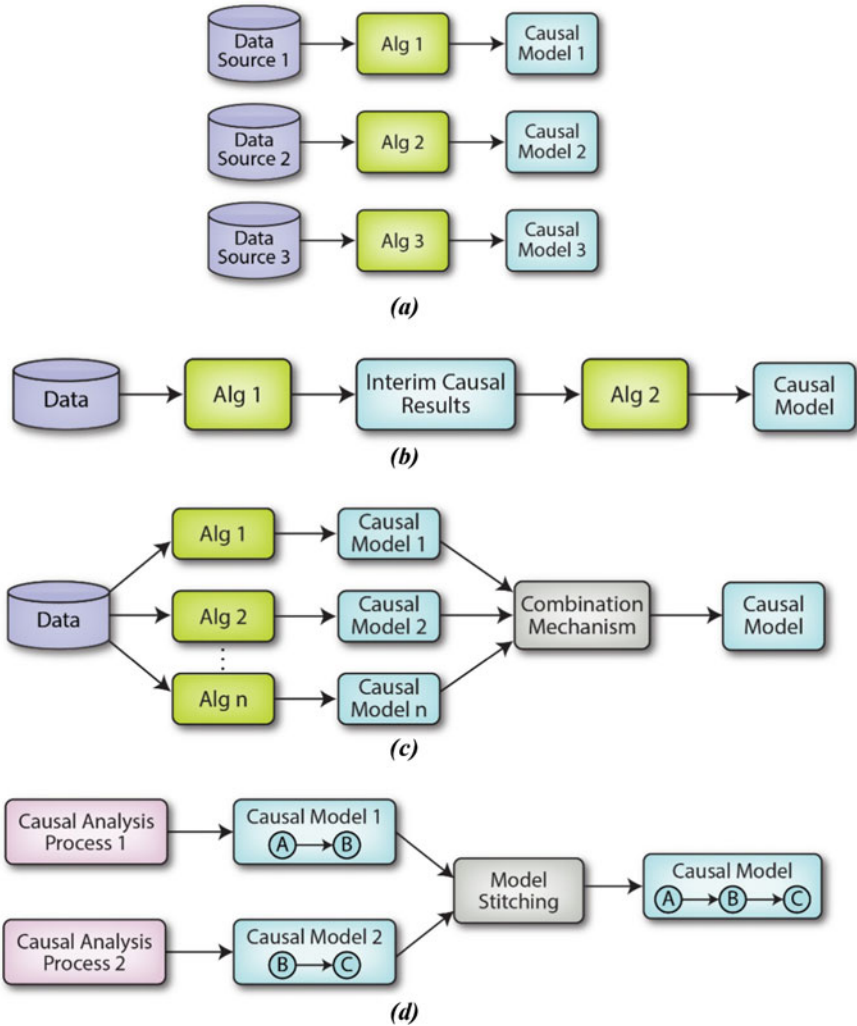
Fig. 2 The building blocks of ensemble causal reasoning: **a** data ensembles; **b** chain ensembles; **c** technique ensembles; **d** nested ensembles

Nested ensembles. Ensemble causal analysis can enable us to combine causal models that derive from different data sources. This approach abstracts away from the raw data, which may come in a variety of forms (e.g., qualitative observations, surveys, time series) that are very challenging to analyze jointly, to identify linkages that can combine causal models from diverse data sources. For example, a time series causal analysis method, such as Granger causality, may indicate that increasing un-employment leads to increased crime. A human survey might indicate that a higher level of crime results in increased preference for more police spending.

Using an ensemble approach, we can combine these two causal models from datasets that cannot be linked using standard quantitative or qualitative methods to achieve a more complete picture of the causal process that leads from unemployment to police spending.

3.2   *Ensemble Combination and Model Stitching*

Automating model combination and stitching to create ensembles is a core problem area. In particular, in real-world application where the size of the model space may grow large and reconfiguration is required to apply existing tools to a new problem set or domain, purely manual approaches are likely to become infeasible. In the experiments described in this paper, we extended the approach described in [11] to create graphical causal models out of smaller models discovered through a variety of analysis techniques. When these probabilistic graphical models (e.g., Bayesian Networks) have identical nodes, the models can be stitched through these nodes reliably into a new, larger network. Similarly, approaches like structural equation models [17] allow models to be stitched as long as they are consistent (i.e., non-contradictory).

More challenging cases arise when stitching nodes that are not identical, such as when estimating different properties of the same phenomenon, which is a quite common misalignment when analyses span multiple domains, data sets, and techniques. For example, one model might characterize the expected crime count where another estimates the likelihood of an increase in crime rate. Human expertise is one mechanism to resolve this kind of mismatch, but achieving practical performance is likely to require automated semantic reasoning, such as through multi-formalism methods [18].

One of the primary concerns with using purely automated causal-analysis techniques is that they often overfit available data and find spurious relationships. We have found that, instead of having an automated tool that finds the "right" answer and informs the user of that answer, it is more effective to use techniques, such as iterative modeling, where the user proposes some concepts and the automation attempts to validate (or, more accurately, invalidate) the hypothesis and explain problems with the proposal given the available data. It can also suggest a number of possible alternatives for the human modeler to review. Once approved, the human or the automation can suggest additional extensions to the causal model (e.g., now that we believe A causes B, can we determine what might cause A?). This sort of human-machine (or top-down/bottom-up, or mixed-methods) approach is described in more detail in [9].

In particular, user insights and hypotheses can be incorporated as an explicit component of the causality analysis ensembles, combining interactive human input with various automated methods according to various voting or heuristic

combination method. For example, one ensemble might first use methods to look for direct causal relationships between observational variables by validating a model provided by the user. Then, automated methods will look for latent variables to supplement the model and determine if there may be some confounding variables the user missed. The system can then solicit input from the user to provide insight into what this latent variable might be and to identify supporting data sources that might capture this variable.

4 Experimentation and Evaluation

To evaluate the ensemble approach described in the previous section, we performed several experiments using a variety of ensemble types and data sources, including both synthetic and real-world data. In this section, we describe the results of these experiments and provide a case study applying our ensemble modeling approach to understanding violence in Iraq's Anbar Province in 2006.

4.1 Ensemble Model Using Time Series Data

The first experiment was designed to evaluate a technique ensemble over time-series data. We developed a synthetic causal data generator to create consistent ground truth temporal data consisting of two time series, A and B, of which one was designated the cause and the other the effect. The generator enables us to vary the mathematical function that relates the cause and effect (i.e., linear, sinusoidal, quadratic) as well as the temporal offset between the cause and effect. We generated datasets for each type of function at each offset ranging from 0 (i.e., the cause and effect occur at exactly the same time) to 10 (i.e., the cause precedes the effect by 10 time steps). We used implementations of Granger causality and offset Pearson correlation to determine the strength of the causal relationship between each A and B pair and identify which is the cause. In this baseline evaluation, Granger causality had a 46 % accuracy and Pearson had a 73 % accuracy in identifying the causal factor.

Using machine learning, we created an ensemble classifier that combines these two approaches. We found that the ensemble learning did not provide much improvement over the offset Pearson correlation until additional metadata was also included as features, such as the types of functions defining the causal relationship. Using this, the ensemble approach was able to achieve over 90 % accuracy in determining the correct causal model for each pair of time series. Interestingly, we found that adding the temporal offset as a feature in the ensemble learning did not seem to improve the scores.

4.2 Ensemble Model Using Non-temporal Data

In our second experiment, we created another same-type ensemble, combining several parallel additive noise models to determine the directionality of causal relationships in non-temporal observational data. For this experiment, we used the CauseEffectPairs (CEP) benchmark data set used in previous studies [6, 7] of additive noise models of causality. This data consists of 99 different "cause-effect pairs" taken from data sets from a variety of different domains (e.g., meteorology, biology, medicine, engineering, economy, etc.). For example the data contains pairs of altitude and temperature measurements from weather stations across Germany (here, altitude is clearly the cause, and temperature the effect), and pairs of CO_2 emissions and energy use compiled by the United Nations for 152 countries between 1960 and 2005 (energy use is the cause of CO_2 emissions). Each cause-effect pair consists of samples of a pair of statistically dependent random variables, where one variable is known to cause the other one, providing us with ground truth for evaluating our ensemble approach.

The ensemble approach proved very effective when dealing with this non-temporal data. We implemented a variety of state-of-the-art additive noise models from [6]. Using machine learning, we created an ensemble of additive noise models that was much more effective at identifying causal relationships than any single model. Individually, the best additive noise models were only capable of 60 % accuracy when determining the direction of the cause-effect pairs in the dataset, while most models had an approximate 50 % accuracy rate. Using our learned ensemble, this improved dramatically to approximately 90 % accuracy. This result indicates that an ensemble can identify important features of a dataset, such as the amount and type of noise, and determine which combination of additive noise models will be most successful. The wide range of observational data used in this experiment indicates the general applicability of this approach to a variety of domains, including PMESII modeling.

4.3 Case Study: Applying Ensemble Causal Models to Conflict in Iraq

To demonstrate how ensemble models can help identify causal relationships to characterize PMESII features in sociocultural systems, we applied this approach to modeling violence in Iraq. Currently, Iraq is plagued by continuing and increasing violence from the terror organization that calls itself the Islamic State of Iraq and the Levant (ISIL), and militant activity is growing throughout Iraq, Syria, and parts of Lebanon. World leaders have been debating how best to address this increasing threat and stabilize the region. Many experts have turned to the recent past to better understand and model the present PMESII effects, looking at parallels to Iraq's Anbar Province in 2006 when the region was under increasing control by Al Qaeda

militants. Using a variety of causal analysis approaches and diverse data, we developed a rich nested ensemble that employs data, chain, and technique ensembles to capture the dynamics of the situation, illustrating how decision-makers might use these approaches to understand the possible impacts of different policy options.

For our analysis, we used Empirical Studies of Conflict (ESOC) data from Princeton University [9]. ESOC contains a variety of different types and sources of data regarding Iraq that we used for this evaluation: (1) time series of violent events in Iraq yearly, monthly, and weekly at the province and district levels from February 2004–February 2008 as compiled from the significant activity (SIGACT) reports by Coalition Forces; (2) time-series reconstruction data on all projects executed in the country from 2003 to 2008, including the type of project and the amount and source of funding; and (3) non-temporal demographic data based on a national census in 2008. Each of these data sources can provide insight into a different aspect of the situation in Anbar province in 2006, capturing the militant activity, sociocultural, and sectarian issues. In addition, the choice of data for this experiment was itself motivated by a causal theory presented by human experts, indicating that there is some relationship between socioeconomic conditions and the propensity for violence and sectarian conflict in the region. To make use of all this data, including expert social science theories, we employed a variety of the analysis methods discussed in Sect. 2. Using a chain ensemble, we combined human expertise in identifying potential causes and effects with both additive-noise and time-series analyses to determine the actual direction and nature of the causation. Because we have different types of data sources, we used a technique ensemble approach to combine additive noise results and time-series analysis into a single coherent causal model. For intuitive analysis by policy-makers, the resulting model was represented as a graphical model with a structure similar to Pearl's [4, 5]. Given all of these factors, we constructed a complex nested ensemble, enabling us to combine human expertise with the additive noise and time-series ensembles developed for the prior two experiments in a graphical model representation.

Figure 3 shows the final model constructed using these nested ensembles over the ESOC data. Looking at the time series in the reconstruction and SIGACTS data, our temporal ensemble of Granger causality, offset Pearson correlation, and dynamic time warping identified several relationships where specific types of reconstruction projects (e.g., Military Facilities, Democracy Building, Infrastructure Improvements, etc.) influence (either increase or decrease) the number of violent acts, or vice versa. From the non-temporal demographic data, the additive noise ensemble discovered causal links from the size of the Shia, Sunni, and mixed populations to the number of SIGACTS. In addition, the causal ensemble was also able to identify specific human-created hypotheses that were not supported by the data. In the structure proposed by human experts, reconstruction efforts targeting courts and schools were considered important factors in reducing the violence; however, subsequent analysis in the chain ensemble by automated time series methods did not support this assumption and with concurrence from the human, these factors were removed from the final model.
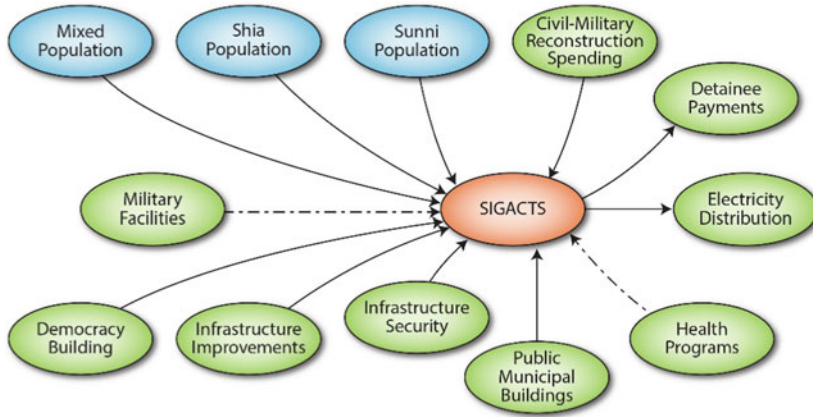
Fig. 3 Causal model of the impact of socioeconomic and demographic factors on militant activities in Anbar province, extracted from real-world data using an ensemble of ensembles

This case study demonstrates the power of ensemble combinations for analyzing causality in sociocultural systems, enabling us to combine various different types of data and models into a single analysis that provides richer and more accurate results than any single modeling approach. By combining causal analysis techniques, policy makers weighing different approaches to a situation, such as stabilizing Anbar in 2006 or dealing with ISIL in 2016, can employ the variety and richness of information available in different types of data sets, ultimately producing a model of the sociocultural system that considers factors across the full PMESII spectrum. For example, using the model in Fig. 3, a decision maker can determine the types of reconstruction projects that might be most influential for stabilizing the region, which demographic features contribute most to unrest, and identify new or unexpected causal linkages.

5   Discussion and Conclusions

The example presented in the previous section provides an illustration of the utility of our ensemble approach for PMESII analysis and modeling. While this experiment did not capture all of the intricacies of a real-world policy investigation, even this relatively small dataset with a few thousand measurements clearly illustrates the multiplicity of domains of study, potential stakeholders, and policy implications that can feature in such an analysis. In this instance, as well as the other two experiments described above, our approach demonstratively achieved the essential ambition promised by the field of ensemble machine learning: combinations of multiple, diverse techniques outperforming individual approaches. It achieved this by constructing data ensembles with a variety of data and types of causality, chain

ensembles incorporating human perspectives, technique ensembles that crowd-source the application of different algorithms, and a case study that used a nested ensemble of all the previous types.

As we continue to refine and mature these ideas, we are cognizant of a number of significant challenges that lay ahead. For example, the experiments above required a great deal of human guidance and intervention in selecting data, assembling models, configuring algorithms, and choosing the most useful and interesting features. To this end, the commoditization of machine learning in the open source community suggests a number of promising avenues of improvement. In addition, it is clear that human knowledge, context, and perspective is a necessary component to any successful PMESII analysis. In the current implementation of our approach provides near infinite control and visibility to the human, who can, for example, see and modify every element of the model. However, the basic interaction patterns described here scale very poorly as the size of the data and model space grow very large. For these problems, we gain inspiration from research in human systems, especially in managing Big Data, such as recommendation systems that facilitate access to large knowledge bases.

In conclusion, these experiments have demonstrated that our ensemble approach has the potential to take advantage of the great diversity in the conceivable inputs, dynamics, and objectives of PMESII modeling. As researchers, we look forward to the many challenges that remain to operationalize such an approach.

References

1. Granger, C.W.: Investigating causal relations by econometric models and cross-spectral methods. Economet.: J. Economet. Soc. 424–438 (1969)
2. Granger, C.W.: Testing for causality: a personal viewpoint. J. Econ. Dyn. Control **2**, 329–352 (1980)
3. Sugihara, G., May, R., Ye, H., Hsieh, C., Deyle, E., Fogarty, M., Munch, S.: Detecting causality in complex ecosystems. Science **338**(6106), 496–500 (2012)
4. Pearl, J.: Causality: Models, Reasoning, and Inference. Cambridge University Press (2009)
5. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann, San Mateo, CA (1988)
6. Mooij, J.M., Peters, J., Janzing, D., Zscheischler, J., Schölkopf, B.: Distinguishing cause from effect using observational data: methods and benchmarks. J. Mach. Learn. Res. (To Appear)
7. Peters, J., Mooij, J.M., Janzing, D., Schölkopf, B.: Causal discovery with continuous additive noise models. J. Mach. Learn. Res. **15**(1), 2009–2053 (2014)
8. Goldthorpe, J.H.: Causation, statistics, and sociology. Eur. Sociol. Rev. **17**(1), 1–20 (2001)
9. Sliva, A., Reilly, S.N., Casstevens, R., Chamberlain, J.: Tools for validating causal and predictive claims in social science models. Proceedings of the 4th international conference on cross cultural decision-making. Proc. Manuf. **3**, 3925–3932 (2015)

10. Sliva, A., Neal Reilly, S.: A big data methodology for bridging quantitative and qualitative political science research. In: American Political Science Association Annual Meeting (2014)
11. Collier, P., Elliott, V.L., Hegre, H., Hoeffler, A., Reynal-Querol, M., Sambanis, N.: Breaking the Conflict Trap: Civil War and Development Policy. World Bank and Oxford University Press, Washington, DC (2003)
12. Collier, P., Hoeffler, A.: Greed and grievances in civil wars. Oxford Econ. Papers **56** (2004)
13. Verma, T., Pearl, J.: Equivalence and synthesis of causal models. In: Proceedings of Sixth Conference on Uncertainty in Artificial Intelligence, pp. 220–227 (1990)
14. Spirtes, P., Glymour, C., Scheines, R.: Causation, Prediction, and Search. MIT Press (2001)
15. Ruvinsky, A., Lustick, I., Levey, B., Siedlecki, T.: Leveraging big data and social science theory for sociocultural modeling (Special presentation). In: Proceedings of the 4th International Conference on Cross-Cultural Decision Making, AHFE International (2015)
16. Opitz, D., Maclin, R.: Popular ensemble methods: an empirical study. J. Artif. Intell. Res. 169–198 (1999)
17. Hancock, Gregory R.: Fortune cookies, measurement error, and experimental design. J. Modern Appl. Stat. Methods **2**(2), Article 3 (2003)
18. Levis, A.: Multi-formalism modeling of human organizations. In: 29th European Conference on Modelling and Simulation (2015)