Enhancing Causal Discovery from Robot Sensor Data in Dynamic Scenarios

Author names withheld

Editor: Under Review for CLeaR 2023

Abstract

Identifying the main features and learning the causal relationships of a dynamic system from time-series of sensor data are key problems in many real-world robot applications. In this paper, we propose an extension of a state-of-the-art causal discovery method, PCMCI, embedding an additional feature-selection module based on transfer entropy. Starting from a prefixed set of variables, the new algorithm reconstructs the causal model of the observed system by considering only the its main features and neglecting those deemed unnecessary for understanding the evolution of the system. We first validate the method on a toy problem, for which the ground-truth model is available, and then on a real-world robotics scenario using a large-scale time-series dataset of human trajectories. The experiments demonstrate that our solution outperforms the previous state-of-the-art technique in terms of accuracy and computational efficiency, allowing better and faster causal discovery of meaningful models from robot sensor data.

Keywords: causal discovery, feature selection, time-series, transfer entropy, causal robotics.

1. Introduction

Knowing the main variables contributing to the evolution of a dynamic system and their causal relationship is important for many real-world applications. For this reason, feature selection and causal discovery methods have become a crucial problem in machine learning and related areas, including robotics (Schölkopf et al., 2021; Ahmed et al., 2021).

Recently, the concept of causal analysis has extended over two fronts. One the one hand, studying the causal relationships between features of complex and dynamical systems has become a turning point on understanding and forecast the evolution of the latter when their models are unknown. For this reasons, over the last few decades, many causal discovery algorithms for static and time-series data have been developed (Glymour et al., 2019; Runge, 2018). On the other hand, another branch of causality/machine learning deals with identifying the most meaningful set of variables that is responsible for the evolution of the system (Chandrashekar and Sahin, 2014).

To our knowledge though, none of these data-driven approaches allow to extract both the important features representing the system and the causal association between them. We therefore aim extend one of the state-of-the-art causal discovery methods, i.e. PCMCI (Runge, 2018), augmenting it with a feature-selection algorithm that is able to identify the correct subset of variables involved in the causal analysis, starting from a prefixed set of them. Therefore, we introduce an all-in-one approach that identifies the causal features representing the system and, based on them, builds a causal model directly from time-series data. As a consequence, the causal discovery process results faster and more accurate. For instance, in an automated warehouse scenario (see Fig. 2), where the working area is shared between objects, humans, and a mobile robot that observes the interactions among them (e.g. worker and shelf), it is important to know which features, among those detectable by the robot's on-board sensors, are important in order to describe the observed interaction (e.g.

© 2023 .

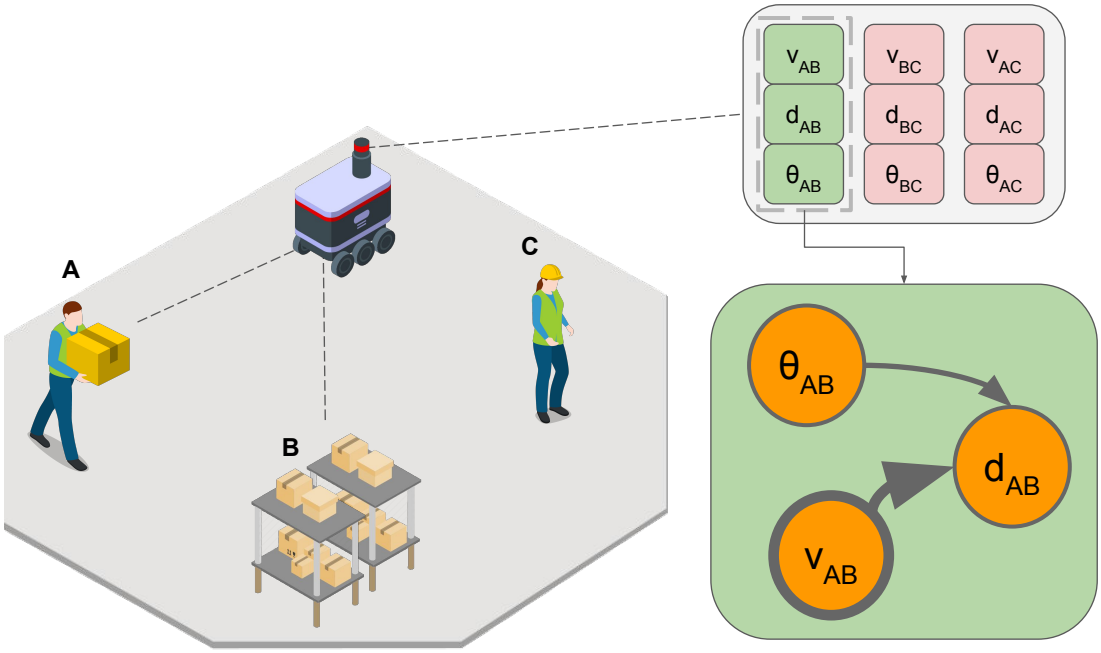

Figure 1: A warehouse-like environment where a mobile robot is observing the interaction between agents A and B. By applying our approach, the robot is able to neglect those variable not useful for the causal analysis. In this case, the variables of interaction AC and BC can be neglected since the agent B is a static object and the agent C is a standing human, both of them not involved in the interaction.

human-shelf distance/angle, human velocity, etc.), and which instead can be neglected (e.g. other humans not involved in the interaction). Our approach would allow the robot to discard unnecessary features and build a causal model of the interaction using only those actually involved in the process.

In this preliminary work, we demonstrate that by starting from a prefixed set of observable variables and applying a suitable feature selection method based on transfer entropy, our algorithm is able to extract those that better characterise the evolution of the system and find the causal relationships among them, significantly faster and more accurately than PCMCI. In summary, our contributions are the following:

- an all-in-one algorithm able to select the most meaningful features from a set of variables and build a causal model from such selection. To this end, we significantly enhance speed and accuracy of the causal discovery;
- experimental evaluation of the algorithm and the obtained causal models on a challenging robotics dataset to predict spatial interactions in dynamic environments.

The paper is structured as follows: related work about causal representation learning and causal discovery are presented in Section 2; Section 3 introduces the motivation for using a transfer entropy-based strategy in our approach; Section 4 explains the implementation details of our method;

Section 5 presents simulation and real-world results to validate, respectively, correctness and performance of our solution; finally, we conclude the paper in Section 6 discussing achievements and future applications.

2. Related Work

Feature selection: The increasing use of machine learning for high-dimensional data analysis led to the development of several feature selection methods. Indeed, feature selection helps to overcome the high dimensionality challenge by eliminating redundant and irrelevant data. Removing the irrelevant features improves learning accuracy and reduces the computation time. Feature selection methods can be categorized in three different categories, reviewed in (Chandrashekar and Sahin, 2014). The first, *filter methods*, use general statistics metrics (e.g. correlation, mutual information) to selected the most useful features; the second, *wrapper methods*, use a predictor as a black box and its performance as the objective function to evaluate the variable subset; finally, the third, *embedded methods*, integrate the feature selection process into a classification/regression model. For their capability to deal with high-dimensional data and low computational cost, filter methods are often used as feature selection ones. In particular, the Transfer Entropy (TE) (Schreiber, 2000; Bossomaier et al., 2016) is an extension of mutual information that measures the directed information transfer between time series of a source and a target variable. TE has become popular in many scientific disciplines to infer dependencies and whole networks from data. Recently, it has often been used as a causal discovery method (Zeng et al., 2022a; Chen et al., 2021; Zeng et al., 2022b), although Runge et al. (2012) demonstrated that TE does not satisfy two of the five causal measures principles that a causal measure must satisfy, as discussed later in Sec. 3. However, for its capability to find the parents for each target variable, in this paper TE has been adopted as feature selection method and moreover, to have a initial preview of a causal model.

Causal discovery: Structural causal models (SCMs) are at the core of causal discovery (Pearl, 2016), complemented by opportune Directed Acyclic Graphs (DAGs) with nodes and oriented edges to represent, respectively, system variables and dependencies between them. The knowledge of SCMs leads to the possibility to reason on the cause and effect relationship between variables. Several methods have been developed over the last few decades to derive causal relationships from observational data. They can be categorized into two main classes, reviewed in (Glymour et al., 2019). The first, *constraint-based methods*, such as Peter and Clark (PC) and Fast Causal Inference (FCI), rely on conditional independence tests as constraint-satisfaction to recover the causal graph, while the second, *score-based methods*, such as Greedy Equivalence Search (GES), assign a score to each DAG and perform a search in this score space. More recently, reinforcement learning-based methods have also been used to discover causal structure (Zhu et al., 2020). However, most of these algorithms work only with static data (i.e. no temporal information), which is a limitation in many robotics applications. Indeed, methods for time-dependent causal discovery are necessary to deal with time-series of sensor data.

To this end, time-dependent causal discovery methods come to the rescue when we are dealing with situations where the temporal order of the data is necessary to understand and recover the correct behaviour of the system. In (Runge, 2018), a variation of the PC algorithm adapted to time-series data was introduced. This method, called PCMCI, is designed to efficiently reconstruct causal graphs from high-dimensional time-series datasets with an additional optimisation for the

control of the false positive rate, performed by a momentary conditional independence (MCI) test. An extension of the PCMCI algorithm, called PCMCI$^+$, has been introduced in (Runge, 2020) to include discovery of contemporaneous dependencies. PCMCI applications can be found in climate and healthcare sectors (Runge et al., 2019; Saetia et al., 2021). An interesting causal discovery application with time-series data is presented in (Kobayashi et al., 2017), where the PC causal discovery has been used to build a mining algorithm to pinpoint failures and their causes in network syslog data.

In robotics, recent works include a method to build and learn a SCM through a mix of observation and self-supervised trials for tool affordance with a humanoid robot (Brawer et al., 2020). Two interesting applications include the use of PCMCI: one to derive the causal model of an underwater robot trying to reach a target position (Cao et al., 2021), and another for modeling and predicting human spatial interactions in a social robotics context (Castri et al., 2022). More causal approaches can also be found in the area of robot imitation learning (Katz et al., 2018; Angelov et al., 2019). However, all these solutions relies on a pre-determined set of variables for performing the causal analysis and do not extract the most meaningful ones for the reconstruction of the causal model. In real-world scenarios, including the robotics sector, such set of variables might be extremely large, leading to slow causal analysis due to the computational cost, which increases with the number of variable, and to inaccurate causal models with a high percentage of spurious links.

3. Transfer Entropy-based Filter

As already explained in the Sec. 2, the Transfer Entropy (TE) (Schreiber, 2000; Bossomaier et al., 2016) is an extension of mutual information that measures the directed information transfer between time-series of a source and a target variable. It is the information-theoretic analog of Granger causality (Barnett et al., 2009). TE has become popular in many scientific disciplines to infer dependencies and whole networks from data. Indeed, we can find some application of it for causal discovery (Zeng et al., 2022a; Chen et al., 2021; Zeng et al., 2022b).

Runge et al. (2012) established five key criteria that a causal measure should fulfill:

- *generality*: the measure should not be restricted to certain types of associations (e.g. linear association);
- *equitability*: the measure should reflect a certain heuristic notion of coupling strength, i.e., it should give similar scores to equally noisy dependencies;
- *causality*: the measure should give a nonzero value only to the dependency between lagged components of a multivariate process that are not independent conditional on the remaining process;
- *coupling-strength autonomy*: the measure is uniquely determined by the interaction of the two components alone and in a way autonomous of their interaction with the remaining process;
- *practically computable*.

Moreover, Runge et al. (2012) introduced a causal measure related to the TE, called *momentary information transfer* (MIT), which represents the base for the PCMCI causal discovery algorithm (Runge, 2018), highlighting that the proposed causal measure satisfies all the above-mentioned principles, unlike TE. Indeed, the latter has various limitations that make it not suitable for measuring the causal strength correctly, and it is therefore not suitable for causal discovery. In particular, its limitations are the following: *(i)* it is not lag-specific, which can lead to false interpretations, like

	TE	MIT
generality	X	X
equitability	X	X
causality	X	X
coupling-strength autonomy		X
practically computable	/	X

Table 1: Comparison between TE and MIT under the point of view of the key causal principles. The "X" corresponds to the satisfaction of the criteria, the "/" corresponds to the criteria satisfied with a workaround.

in the case of feedback in the analysed model; *(ii)* it is not uniquely determined by the interaction of the two components alone and depends on misleading effects of, e.g., autodependency and the interaction with other processes; finally *(iii)*, since TE measures the influence of a variable to another one considering all the past lags of the process, this makes it not practically computable. In essence, TE does not fulfill the criteria of coupling-strength autonomy and practically computable. A workaround for the second limitation though is the introduction of a truncation parameter to discard the infinite past lags of the process and maintain only the lags up to the maximal coupling delay of the variable involved in the measure. However, the introduction of this workaround has a strong influence on the value of the TE and affects the reliability of the measure. Table 3 summarises the comparison between TE and MIT.

Despite the above considerations and limitations, TE is still a valid choice for the feature selection problem. Due to the unmet coupling-strength autonomy criteria, it would not able to give a correct value of the measure but it can still indicate whether a relation between two variables exists (causality principle). For this reason, we adopted TE as feature selection method for our approach, as implemented in (Wollstadt et al., 2019).

4. Filtered-based Causal Discovery

Our approach, named Filtered PCMCI (FPCMCI), is based on the use of a feature selection method as "filter" in order to extrapolate the important features and possible associations between them from the whole set of variables before the causal analysis. As already explained in section 3, we used the Transfer Entropy (TE) (Schreiber, 2000; Bossomaier et al., 2016) to decide which variables and links can be excluded from the original set and instead, which variables and links need to be taken into account in the causal analysis. As output, the filter returns a set of variables and a set of possible causal associations between them, which then, need to be validated through a proper causal analysis. The latter is performed by using the PCMCI causal discovery algorithm. More in detail, PCMCI is a causal discovery algorithm (Runge, 2018) which consists of two main parts, both exploiting conditional independence tests (e.g. partial correlation, Gaussian processes and distance correlation) to measure the causal strength between variables. The first part is the well-known PC algorithm, which starts from a fully connected graph and outputs an initial causal model structure; the latter is then used by the second part, the MCI test, which validates the structure by estimating the test statistics values and p-values for all the links and outputs the final causal model.

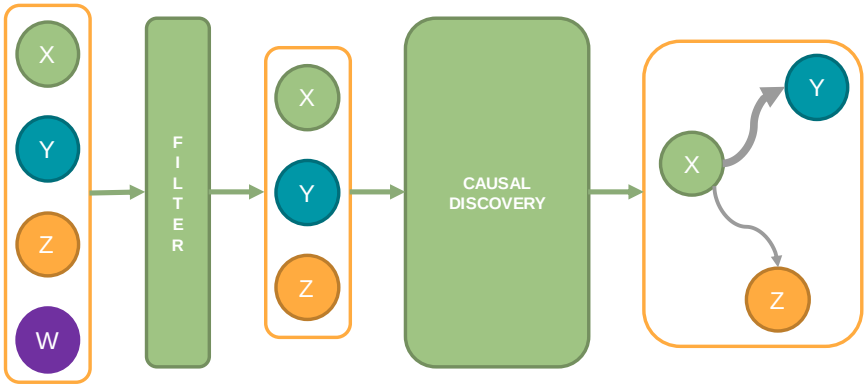

Figure 2: A block-scheme representation of our approach. Starting from a prefixed set of variable (X, Y, Z, W), apply the feature selection method to remove unneeded variables (W) and feed into the causal discovery algorithm only those variables (X, Y, Z) and the set of possible association chosen by the filter. As output, the causal model without unnecessary variables.

4.1. Toy Problems

To evaluate the correctness of our approach, an application with a known ground-truth causal model was needed. Indeed, we implemented two different toy problems (i.e. nonlinear system of equations) in order to assess the result of our approach in two different situations: derive the causal model of the nonlinear system with 1 as maximum dependencies time lag, described in (1) hereinafter defined as S_1; derive the causal model of a nonlinear system of equation with 2 as maximum dependencies time lag, described in (2) hereinafter defined as S_2. In both cases, the structure of the nonlinear systems were defined in advanced and they are defined as follows:

$$S_1 = \begin{cases} x_0(t) = c_{00}x_0(t-1) - c_{01}x_1(t-1) * c_{02}x_2(t-1) + \eta_0(t) \\ x_1(t) = \eta_1(t) \\ x_2(t) = \frac{c_{21}x_1^2(t-1)}{1+c_{22}x_2(t-1)} + \eta_2(t) \\ x_3(t) = c_{33} + \sqrt{x_3(t-1)} + \eta_3(t) \\ x_4(t) = \frac{c_{41}x_1(t-1)c_{42}x_2(t-1)}{1+c_{43}x_3(t-1)} + \eta_4(t) \\ x_5(t) = \eta_5(t) \\ x_6(t) = \frac{c_{60}x_6(t-1)}{1+c_{65}x_5(t-1)} + \eta_6(t) \end{cases} \tag{1}$$

$$S_2 = \begin{cases} x_0(t) = c_{01}x_1(t-2)c_{02}x_2(t-1) + \eta_0(t) \\ x_1(t) = \eta_1(t) \\ x_2(t) = c_{21}x_1^2(t-2) + \eta_2(t) \\ x_3(t) = c_{33} + x_3(t-1) + \eta_3(t) \\ x_4(t) = c_{42}x_2(t-2) - c_{43} * x_3(t-1) + \eta_4(t) \\ x_5(t) = \frac{c_{50}x_0(t-1)}{1+c_{55}x_5(t-2)} + \eta_5(t) \\ x_6(t) = \eta_6(t) \end{cases} \tag{2}$$

where x represents the system variables, c the coefficients and η a uniform noise with value in the range $[0, 1)$.

4.2. Modeling Real-world Human Spatial Interactions

Our approach has been applied for modeling and predicting spatial interactions (Fig. 2). It can be decomposed in three main steps: (*i*) extract all time-series sensor data from a human spatial interaction scenarios; (*ii*) use them for the causal discovery performed by our approach; (*iii*) finally, embed the obtained causal model in a LSTM-based prediction system.

In more detail, for the LSTM-based prediction we used the network described in (Yin et al., 2021). It proposes an encoder-decoder network with three different attention layers: two of them are in the encoder block and they are, respectively, an input attention mechanism which selects the most useful time steps of the considered observation window of the target variable for predicting the latter, a self-attention mechanism which, instead, selects the more meaningful drivers for predicting the target variable. Finally, the third attention mechanism is positioned in the decoder network. For our purpose and for having a simpler network, the latter attention layer has been removed.

Moreover, since the self-attention mechanism is the one that selects the drivers for each target variable, we decided to embed our causal model coming from our approach in this layer. In particular, the attention mechanism in this layer is computed as follows:

$$g_t = \tanh(W_g x_t + b_g) \tag{3}$$

$$\alpha_t^k = \sigma(W_\alpha g_t + b_\alpha) \tag{4}$$

where σ denotes the sigmoid function, W_g and W_α are the learnable parameters, b_g and b_α are the bias vectors, x_t represents the input driving series at time step t for the target variable x^k. We used the bias vector b_α to embed the causal inference vector extracted from our causal model corresponding to the target variable x^k and we set it as a non-trainable parameter for the network.

5. Experiments

To evaluate the correctness of our approach, we first implemented various toy problems described in Sec. 4.1 in order to verify that our approach brings advantages with respect to the current state-of-the-art method PCMCI under the point of view of the computational cost and the causal models' accuracy. Once, established that our approach works correctly, we applied it for modeling and predicting human (using the network described in Sec. 4.2) spatial behaviours on a challenging dataset, i.e. THÖR (Rudenko et al., 2020). The latter contains data of people moving in indoor environments, a workshop/warehouse. Our strategy is first to extract the real sensor time-series data

from the dataset, as explained in Sec. 4, and then use it for causal discovery. In order to prove the usefulness of our approach, a comparison between causal and a non-causal predictions is finally shown.

5.1. Data Processing

From the dataset, we extracted the x-y positions of each agent and derived all the quantities from them (i.e. orientation θ, velocity v, etc.). In more details, the THÖR dataset provides a wide variety of interactions between humans, robot, and static objects (Fig. 3). Helmets and infrared cameras were used to track the motion of the agents at $100\ Hz$. We used this dataset to analyse human spatial interactions as described in Sec. 4.2. Moreover, to reduce the computational cost of causal discovery on this dataset, due to the large amount of samples in the time-series, we selected only a subset of them using an entropy-based adaptive-sampling strategy (Aldana-Bobadilla and Alfaro-Pérez, 2015), with an additional variable size windowing approach to reduce the number of samples. In order to represent the human spatial interactions, from the agents' trajectories extracted from the THÖR dataset we derived, for each of them, a set of 8 variables reasonable for the application. These are the following:

- d_g distance between the current position and the current goal position of the selected agent;
- v velocity of the selected agent;
- $risk$ a variable which models the risk due to the presence of other agents in the same environment, inspired by (Castri et al., 2022). It depends on v and d_{obs};
- θ orientation of the selected agent;
- θ_g angle between the current position and the current goal position of the selected agent;
- ω angular velocity of the selected agent;
- g_{seq} sequence of goal positions reached by the selected agent;
- d_{obs} distance between the current position of the selected agent and the closest obstacle in the scenario.

5.2. Network Parameter Settings

As already explained in Sec. 4.2, the network is a LSTM-based encode-decoder model for Multi-Output Multi-Step Forecast. During the training phase, the learning rate is 0.0001 and the batch size is 32. In our model, we set the length of the observation window size equal to 32 and the length of the forecasting window size equal to 48 which correspond to 3.2sec and 4.8sec. For simplicity, we use the same number of LSTM cells per hidden layer for the encoder (2 hidden layers) and the decoder (2 hidden layers), and conduct a grid search over $\{64, 128, 256, 512\}$. With a value equal to 256, our approach achieves the best performance over the validation set.

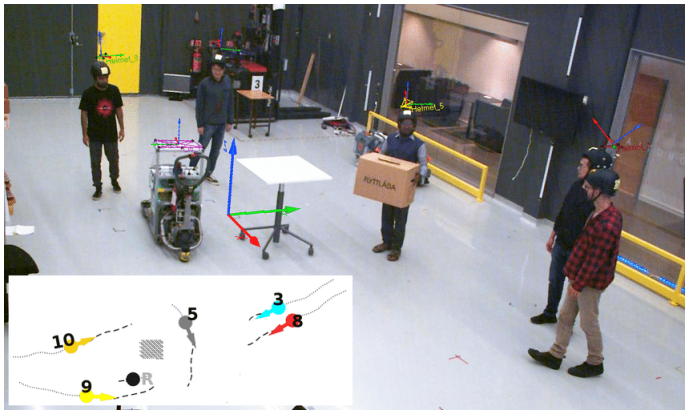

Figure 3: Image from THÖR dataset (Rudenko et al., 2020)

5.3. Evaluation on Toy Problems

Here, we reports the results of the evaluation carried out by using the system of equation S_1 and S_2 defined in Sec. 4.1. In both cases, the number of variables considered in the system varies from 3 to 7 and moreover, for each configuration of system, we performed 10 run tests with different random coefficients, taking as evaluation metric the mean F1-score and the mean execution time (in seconds) over all the 10 tests for each system configuration. Figure 4 shows the comparison between the results of the causal discovery analysis performed by using the state-of-the-art PCMCI algorithm and our approach FPCMCI. In particular, the graphs in the top line correspond to the F1-score and execution time related to the system S_1. Instead, the bottom line reports the F1-score and the execution time related to the system S_2.

Fig. 4 clearly shows that our approach outperforms the PCMCI in terms of accuracy of the causal models and of execution time, highlighting how the the PCMCI algorithm suffers when the dimension of the system increases and how instead, the FPCMCI is able to maintain a reasonable execution time. This difference becomes important in the case of the system S_2, where the dependency time lag equal to 2 increases a lot the number of possible links to check for both algorithms.

5.4. Real-world Experiment

In order to have a complete evaluation of the proposed approach, we applied the FPCMCI to a real-world scenario of human spatial interaction. In this case though, since we do not know the ground-truth causal model we evaluate the correctness and accuracy of the obtained causal model using as metric the accuracy of the prediction approach explained in Sec. 4.2. First of all, we extracted the data from the THÖR dataset and derived a reasonable set of variable for a human spatial interaction application. Then, the latter is filtered and used to derive the causal model by using our approach FPCMCI. After that, we used the causal model in the prediction approach to verify that, the network with the knowledge of the causal model underlying the process that has to be predicted, is able to improve the accuracy of the prediction compared to the non-causal prediction approach. Moreover, in order to have a comparison with the state-of-the-art causal discovery algorithm, we repeated exactly the same analysis but in this case using the causal model derived through the PCMCI.

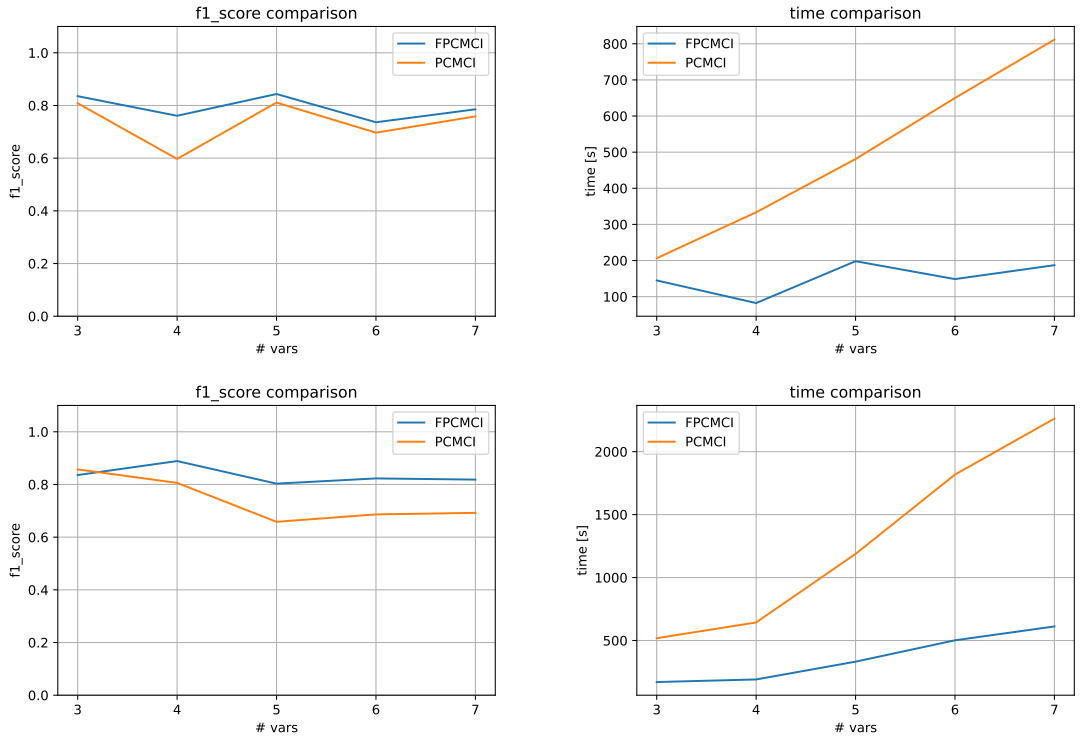

Figure 4: F1-score and execution time comparison between PCMCI (orange) and FPCMCI (blue). The graphs show the two metrics for each particular configuration of the system (e.g. 3 variables, 4 variables, etc.). (Top line) comparison related to the system S_1. (Bottom line) comparison related to the system S_2.

Figure 5 shows the two causal models derived through the PCMCI (left) and the FPCMCI (right) corresponding to the agent 11 of the THÖR dataset. A expected, due to the large number of variables and links, the PCMCI suffers of the presence of spurious links which, is not able to filter out. On the other side, the FPCMCI is able to provide a cleaner causal model, considering the full set of variable (as PCMCI) but only the most meaningful links between them thanks to the introduction of the transfer entropy-based filter. The execution time reflects what just said, indeed the PCMCI had an execution time of 01hr-19min-45sec, instead the FPCMCI's execution lasted only 17min-33sec. These execution times are related to the machine used for the experiment, i.e. Lenovo Legion i7. However, taking into account the PC used for deriving the causal model, the FPCMCI method results 4 time faster than the PCMCI.

For the accuracy, we trained our network for each agent involved in the scenario and, for each training, we use the other agents for the testing. As evaluation metrics for the prediction we used the *Normalised Mean Absolute Error* (NMAE) and the *Normalised Root Mean Square Error* (NRMSE), defined as follows:

$$NMAE(y, \hat{y}) = \frac{\sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{n}}{\sigma(y)} \tag{5}$$

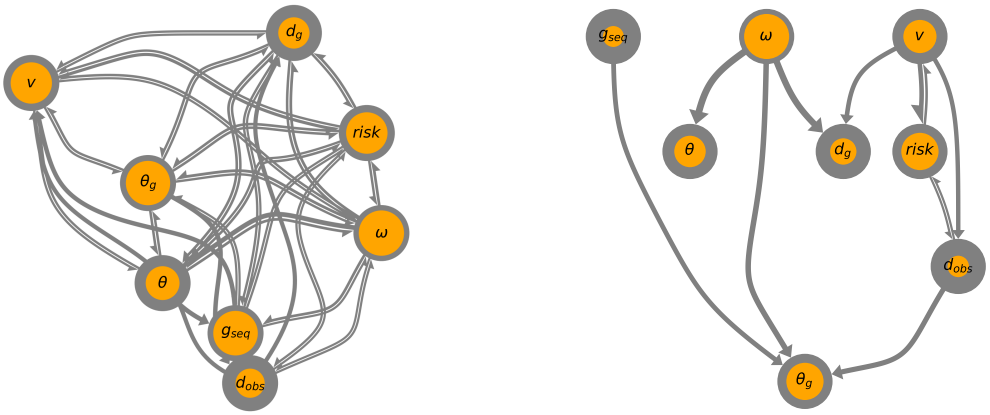

Figure 5: Comparison between the causal models coming from the time-series data extracted from the THÖR dataset (agent 11). (Left) Causal model derived through the PCMCI. (Right) Causal model derived through the FPCMCI. The thickness of the arrows and of the nodes' border represents, respectively, the strength of the cross and auto-causal dependency (the stronger the dependency, the thicker the line/border). All the dependencies correspond to a 1-step lag time.

$$NRMSE(y, \hat{y}) = \frac{\sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}}{\sigma(y)} \tag{6}$$

where y and $\hat{y}$ are, respectively, the actual and the predicted values, and $\sigma(y)$ is the standard deviation of the actual value.

Figure 6 reports the comparison between prediction accuracy in the cases of: non-causal prediction, PCMCI-based prediction, FPCMCI-based prediction. The metrics NMAE and NRMSE are computed for each training and then averaged. The figure clearly shows that the knowledge of the causal model helps to have a more accurate prediction and moreover, the causal model built by the FPCMCI results more accurate than the one derived by the PCMCI, since both the metrics have a lower value in the FPCMCI case compared to the PCMCI.

6. Conclusion

In this work, we proposed an extension of a state-of-the-art causal discovery method, PCMCI, embedding an additional feature-selection module based on transfer entropy. The approach was initially evaluated on a toy problems, for which the ground-truth model was known, to verify the correctness of the approach, and then on a real-world robotics dataset with large-scale time-series of human trajectories. We showed that the our approach significantly improves the PCMCI causal discovery method in terms of accuracy and computational efficiency. This leads to better and faster causal discovery of dynamic models from real-world sensor data. Future work will be devoted to causal discovery and reasoning on both observational and interventional data, exploiting the

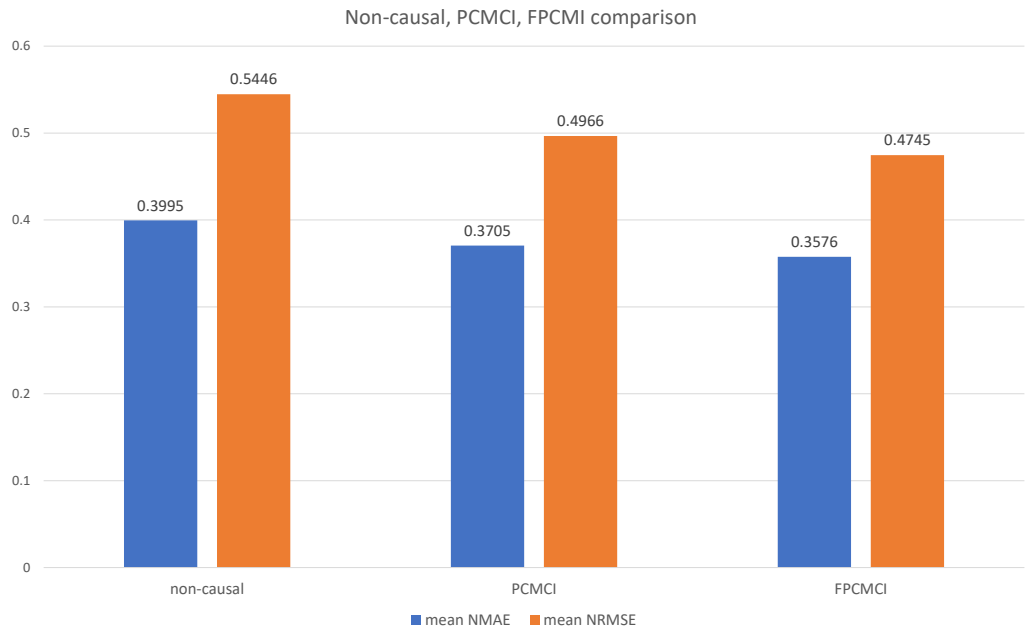

Figure 6: Comparison between non-causal and causal prediction (using PCMCI and FPCMCI) using as metrics the mean NMAE and the mean NRMSE over all the agents in the scenario.

influence that the robot's presence can have on nearby people, with a special interest on industrial and intralogistics applications.

References

O. Ahmed, F. Träuble, A. Goyal, A. Neitz, Y. Bengio, B. Schölkopf, M. Wüthrich, and S. Bauer. Causalworld: A robotic manipulation benchmark for causal structure and transfer learning. In *International Conference on Learning Representations (ICLR)*, 2021.

Edwin Aldana-Bobadilla and Carlos Alfaro-Pérez. Finding the optimal sample based on shannon's entropy and genetic algorithms. In *Mexican Int. Conf. on Artificial Intelligence*, pages 353–363. Springer, 2015.

Daniel Angelov, Yordan Hristov, and Subramanian Ramamoorthy. Using causal analysis to learn specifications from task demonstrations. In *Proc. of the Int. Joint Conf. on Autonomous Agents and Multiagent Systems, AAMAS*, volume 3, 2019.

Lionel Barnett, Adam B Barrett, and Anil K Seth. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical review letters*, 103(23):238701, 2009.

T. R. J. Bossomaier, Lionel C. Barnett, Michael S. Harré, and Joseph T. Lizier. An introduction to transfer entropy. In *Springer International Publishing*, 2016.

Jake Brawer, Meiying Qin, and Brian Scassellati. A causal approach to tool affordance learning. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 8394–8399. IEEE, 2020.

Yu Cao, Boyang Li, Qian Li, Adam Stokes, David Ingram, and Aristides Kiprakis. Reasoning Operational Decisions for Robots via Time Series Causal Inference. In *2021 IEEE Int. Conf. on Robotics and Automation (ICRA)*, pages 6124–6131. IEEE, 2021. ISBN 978-1-72819-077-8.

Luca Castri, Sariah Mghames, Marc Hanheide, and Nicola Bellotto. Causal discovery of dynamic models for predicting human spatial interactions. In *International Conference on Social Robotics (ICSR)*, 2022.

Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014.

Siya Chen, G Jin, and Xinyu Ma. Satellite on-orbit anomaly detection method based on a dynamic threshold and causality pruning. *IEEE Access*, 9:86751–86758, 2021.

Clark Glymour, Kun Zhang, and Peter Spirtes. Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in Genetics*, 2019.

Garrett Katz, Di Wei Huang, Theresa Hauge, Rodolphe Gentili, and James Reggia. A novel parsimonious cause-effect reasoning algorithm for robot imitation and plan recognition. *IEEE Transactions on Cognitive and Developmental Systems*, 10, 2018. ISSN 23798939.

Satoru Kobayashi, Kazuki Otomo, Kensuke Fukuda, and Hiroshi Esaki. Mining causality of network events in log data. *IEEE Transactions on Network and Service Management*, 15(1):53–67, 2017.

Judea Pearl. *Causal inference in statistics : a primer*. Wiley, 2016.

Andrey Rudenko, Tomasz P Kucner, Chittaranjan S Swaminathan, Ravi T Chadalavada, Kai O Arras, and Achim J Lilienthal. Thör: Human-robot navigation data collection and accurate motion trajectories dataset. *IEEE Robotics and Automation Letters*, 5:676–682, 2020.

J. Runge. Causal network reconstruction from time series: From theoretical assumptions to practical estimation. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28:075310, 2018.

Jakob Runge. Discovering contemporaneous and lagged causal relations in autocorrelated nonlinear time series datasets. In *Conference on Uncertainty in Artificial Intelligence*, pages 1388–1397. PMLR, 2020.

Jakob Runge, Jobst Heitzig, Norbert Marwan, and Jürgen Kurths. Quantifying causal coupling strength: A lag-specific measure for multivariate time series related to transfer entropy. *Physical Review E*, 86(6):061121, 2012.

Jakob Runge, Peer Nowack, Marlene Kretschmer, Seth Flaxman, and Dino Sejdinovic. Detecting and quantifying causal associations in large nonlinear time series datasets. *Science Advances*, 5, 2019.

Supat Saetia, Natsue Yoshimura, and Yasuharu Koike. Constructing Brain Connectivity Model Using Causal Network Reconstruction Approach. *Frontiers in Neuroinformatics*, 15:5, 2021. ISSN 1662-5196.

Thomas Schreiber. Measuring information transfer. *Phys. Rev. Lett.*, 85:461–464, Jul 2000. doi: 10.1103/PhysRevLett.85.461. URL https://link.aps.org/doi/10.1103/PhysRevLett.85.461.

Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634, 2021. doi: 10.1109/JPROC.2021.3058954.

Patricia Wollstadt, Joseph T. Lizier, Raul Vicente, Conor Finn, Mario Martinez-Zarzuela, Pedro Mediano, Leonardo Novelli, and Michael Wibral. Idtxl: The information dynamics toolkit xl: a python package for the efficient analysis of multivariate information dynamics in networks. *Journal of Open Source Software*, 4(34):1081, 2019. doi: 10.21105/joss.01081. URL https://doi.org/10.21105/joss.01081.

Xiang Yin, Yanni Han, Hongyu Sun, Zhen Xu, Haibo Yu, and Xiaoyu Duan. Multi-attention generative adversarial network for multivariate time series prediction. *IEEE Access*, 9:57351–57363, 2021.

Zefan Zeng, Guang Jin, Chi Xu, Siya Chen, Zhelong Zeng, and Lu Zhang. Satellite telemetry data anomaly detection using causal network and feature-attention-based lstm. *IEEE Transactions on Instrumentation and Measurement*, 71:1–21, 2022a.

Zefan Zeng, Guang Jin, Chi Xu, Siya Chen, and Lu Zhang. Spacecraft telemetry anomaly detection based on parametric causality and double-criteria drift streaming peaks over threshold. *Applied Sciences*, 12(4):1803, 2022b.

Shengyu Zhu, Ignavier Ng, and Zhitang Chen. Causal discovery with reinforcement learning. *ICLR*, 2020.