Trajectory-Based Behavior Analytics: Papers from the 2015 AAAI Workshop

Trajectory Analysis Based on Clustering and Casual Structures

Raymond K. Wong, Victor W. Chu, Mojgan Ghanavati, Asso Hamzehehi
School of Computer Science and Engineering
University of New South Wales
Sydney, Australia

Abstract
Causal structure discovery methods are investigated recently but none of them has taken possible time-varying structure into consideration. This paper uses a notion of causal time-varying dynamic Bayesian network (CTV-DBN) and define a causal boundary to govern cross-time information sharing. CTV-DBN is constructed by using asymmetric kernels to address sample scarcity and to adhere to causal principles; while maintaining good variance and bias trade-off. Upon satisfying causal Markov assumption, causal inference can be made based on manipulation rule. We explore trajectory data collected from taxis in Beijing which exhibit heterogeneous patterns, data sparseness and distribution skewness. Experiments show that by using casual structures and trajectory clustering, we can analyse the spatio-temporal behavior of the trajectory data.

Introduction
The massive amount of spatio-temporal data collected by tracking devices, e.g., sensor networks, GPS equipped smartphones, etc., and their improved storage capability have enabled the analysis of various real-world problems. In the domain of traffic management, the studies could cover space use, critical paths and their flows, movement relationship (Liu et al. 2011), traffic anomalies (Chawla, Zheng, and Hu 2012) (Pang et al. 2011), etc. Map animation and space-time cube are common approaches to represent and visualize spatio-temporal data (Andrienko, Andrienko, and Gatalisky 2003), but they all have their limitations, e.g., the effectiveness of map animation is in fact questionable (Tversky, Morrison, and Bétrancourt 2002). As a result, it leads us to explore alternative methods to model spatio-temporal data and their relationships.

Structure discovery for spatio-temporal data has been a cross disciplinary subject among databases (Ester et al. 1996) (Jeung et al. 2008) (Lu et al. 2011), data mining (Chawla, Zheng, and Hu 2012) (Wang, Wang, and Li 2006), artificial intelligence (Liu et al. 2010), etc., though they may have different focus and scalability requirements. While there is not much previous work done in analyzing the relationships among network conditions across time, Song et al. (Song, Kolar, and Xing 2009) outline a time-varying dynamic Bayesian networks (TV-DBN) to model gene-to-gene interaction networks. This paper summarizes our previous extension of TV-DBN (Chu et al. 2014) to model time-varying network causal relationships. In particular, in this paper, we demonstrate that by discovering their causal dependency network structure, we will be able to detect region-to-region interactions. As a result, by identifying interesting regions through their causal structures, we can further analyze the trajectory behavior through the trajectory clusters between these regions.

Most spatio-temporal observations from mobile devices exhibit unfavorable statistical properties: heterogeneous patterns, data sparseness and distribution skewness (Liu et al. 2011). By using a time-varying model, the network structure discovery process reveals the evolution of region connections to mitigate the issue of heterogeneous patterns. To address data sparseness and distribution skewness, we use density-based clustering methods to identify regions for a particular space-time interval from trajectories directly. Density-based clustering methods (Ester et al. 1996) (Kisilevich et al. 2010) (Rinaldo et al. 2012) are argued to be the best solution for spatio-temporal clustering (Nanni and Pedreschi 2006).

We introduce a notion of causal time-varying dynamic Bayesian networks (CTV-DBN) by extending the original TV-DBN. In CTV-DBN, causal Markov assumption (Pearl 2000) (Spirtes 2010) (Spirtes, Glymour, and Scheines 2000) (Spirtes, Meek, and Richardson 1999) is satisfied by considering causal boundary. It is achieved by asymmetric kernels (Mackenzie and Tieu 2004) to make the information sharing across time in better adherence to causal principles; but still, ensure sharing among suitable neighbors to address data scarcity while maintaining similar level of variance and bias trade-off. Asymmetric kernels also provide a solution to rectify boundary problem created by real-world data where they mirror a causal function (Mackenzie and Tieu 2004). Causal inference can be made based on manipulation rule once a Bayesian network is made isomorphic with a causal model (Lemmer 1996) (Pearl 2000) (Spirtes 2010) (Spirtes, Glymour, and Scheines 2000) (Spirtes, Meek, and Richardson 1995).

In summary, our primary contributions of the work described in this paper are as follows:
1. A notion of causal time-varying dynamic Bayesian net-
works (CTV-DBN) is introduced that satisfies causal Markov assumption.

2. Asymmetric kernel is used to make the information sharing across time to better adhere to causal principles while maintaining a good level of variance and bias trade-off.

3. Upon satisfying causal Markov assumption, we base on manipulation rule to establish a structure for causal inference.

4. CTV-DBN network discovery is applied to road systems to reveal the evolution of their causal time-varying structures.

5. Trajectory clustering is used to further analyse the spatio-temporal behavior of drivers driving from one region to another region. These regions can be selected based on their causal structures.

Related Work

DBN (Murphy 2002) have been used to model sequences of variables and regarded as a method to overcome the expressive power limitation in Hidden Markov models and Kalman filter models, in which state-space is represented in factored form. Nevertheless, DBN is in fact a time-invariant model, in which the structure of the network is fixed but is capable to model dynamic systems (Song, Kolar, and Xing 2009). Non-stationary dynamic Bayesian networks (NS-DBN) are introduced by Robinson and Hartemink (Robinson and Hartemink 2008) (Robinson and Hartemink 2010) in recent time to model time-varying network structures. Markov chain Monte Carlo (MCMC) sampling method is used; however, Song et al. (Song, Kolar, and Xing 2009) point out that such an approach is unlikely to be scalable, and it is also prone to over-fitting.

In parallel, Grzegorczyk and Husmeier (Grzegorczyk and Husmeier 2009) (Grzegorczyk and Husmeier 2011) also develop an alternative approach. Their assumption of a fixed network structure is deemed to be too restrictive (Dondelinger, Lèbre, and Husmeier 2013), even though the interaction parameters of the model can vary with time to cater for non-stationary systems. Song et al. propose TV-DBN (Song, Kolar, and Xing 2009) to overcome those weaknesses. More recently, Dondelinger et al. (Dondelinger, Lèbre, and Husmeier 2013) propose a more complex non-homogeneous dynamic Bayesian networks for inferring gene regulatory networks with gradually time-varying structure. Although both Dondelinger and Song’s proposals can be traced back to a common root of Robinson and Hartemink (Robinson and Hartemink 2008) (Robinson and Hartemink 2010), Dondelinger et al. work with continuous time data whilst the method proposed by Song et al. (Song, Kolar, and Xing 2009) is suitable to discretization data. From a different direction, Liu et al. (Liu et al. 2010) recently attempt to address the problem of temporal causal graph discovery assuming relational time-series exist.

On the other hand, the theory of statistical causal inference developed by Perl (Pearl 2000) and Spirtes et al. (Spirtes, Glymour, and Scheines 2000) provides a platform for causal relationship to be detected based on observations. In recent time, Pellet and Elissseff (Pellet and Elissseff 2008) attempt to provide a causal structure discovery algorithm for causally insufficient data and show that their Markov blanket/collidet set (MBCS*) algorithm is in several orders of magnitude faster than the popular Fast Causal Inference (FCI) algorithm (Spirtes, Meek, and Richardson 1995). Another example is Inductive Causation* (IC*) (Pearl and Verma 1991), in which both algorithms IC* and FCI are formulated based on relaxing the causal sufficiency assumption. CTV-DBN is different from all of them as it takes the time-varying network structure into consideration and also adheres to causal Markov assumption to make a Bayesian network isomorphic with a causal model (Lemmer 1996).

Casual Structure Discovery

Causal Time-Varying Dynamic Bayesian Networks (CTV-DBN)

Although the structure of Bayesian networks (BN) may be directed, the directions of arrows do not define causal effects as the influence can flow both ways except a collider (v-structure) is hit. Therefore, all of the definitions in BN refer only to probabilistic properties, such as conditional independence (Koller and Friedman 2009). Song et al. (Song, Kolar, and Xing 2009) recognise this problem – BN does not necessarily imply causality, but suggest that dynamic Bayesian networks (DBN) bears a natural causal implication in which TV-DBN is part of this family. Each edge in a DBN only points from time \( t - 1 \) to \( t \) contributing to a natural causal implication. However, the network discovery method established in TV-DBN ignores causal relationships allowing the sharing of information across the whole time period.

The difference between causal models and probabilistic models arises when we care about interventions in a model (Koller and Friedman 2009). We aim to establish causal relationships between regions for causal inference by assigning manipulated probability density to a region of interest (Spirtes 2010). The condition of causal Markov assumption (CMA) is invoked to make a BN isomorphic with a causal model (Lemmer 1996), where the condition is defined as: given a causal graph \( G = (\mathcal{V}_G, \mathcal{E}_G, \mathcal{P}_G) \), where \( \mathcal{V}_G \) is a set of vertices and \( \mathcal{E}_G \) is a set of edges between vertices in \( \mathcal{V}_G \) and \( \mathcal{P}_G \) is a probability distribution over the vertices in \( \mathcal{V}_G \). \( \mathcal{G} \) satisfies CMA if and only if for every \( v \in \mathcal{V}_G \), \( v \) is independent of \( \{ \mathcal{V}_G \setminus (\text{Descendants}(v) \cup \text{Parents}(v)) \} \) given \( \text{Parents}(v) \), where \( \text{Parents}(v) \) is the set of parents of \( v \) in \( \mathcal{G} \) and \( \text{Descendants}(v) \) is the set of descendants of \( v \) in \( \mathcal{G} \) (Spirtes, Glymour, and Scheines 2000).

A probability density \( \mathcal{P}_G(\mathcal{V}_G) \) can be factorized according to \( \mathcal{G} \) if and only if

\[
\mathcal{P}_G(\mathcal{V}_G) = \prod_{v \in \mathcal{V}_G} \mathcal{P}_G(v|\text{Parents}(v)) \quad (\text{Spirtes 2010})
\]

Assuming \( n \subset \mathcal{V}_G \) with only non-descendants of \( m \), a manipulation of \( m \in \mathcal{V}_G \) to \( \mathcal{P}_G(m|n) \) can be achieved by replacing \( \mathcal{P}_G(m|\text{Parents}(m)) \) in Equation (1) by a manipu-
lated density $P_{\mathcal{G}}'(m|m)$ to form a manipulation rule:

$$P_{\mathcal{G}}(V_{\mathcal{G}}|P_{\mathcal{G}}'(m|m)) = P_{\mathcal{G}}(m|V_{\mathcal{G}}) \prod_{v \in V_{\mathcal{G}} \setminus \{m\}} P_{\mathcal{G}}(v|Parents(v)), \quad (2)$$

where the double bar indicates an assignment of probability and $P_{\mathcal{G}}'$ is a new probability density. Hence, the manipulation rule at time $t = \zeta$ (where \(\zeta \in \{2 \ldots T\}\)) and region $i$ is:

$$p(X^t, \ldots, X^{T}||p(X^\zeta_m|X^{\zeta-1}_m)) = \prod_{t=2,T} p(X^t) \prod_{i \neq m} p(X^t_{|X^{t-1}}_i). \quad (3)$$

If $V_{\mathcal{G}}$ represents region variables from all time points, the network structure estimated by optimization does not satisfy CMA. It is because the weighting function $w^t_i(t)$ considers the time points in \(\{V_{\mathcal{G}} \setminus \{Descendants(v) \cup Parents(v)\}\}\), e.g., \(\hat{A}_i^t\) at time $t^*$ is not only determined by $X_i^{t-1}$ but also $X_i^{t+1}$ where $z > 1$. In order to comply with CMA, the weighting function should only gather evidence from $S = \{\{Descendants(v) \cup Parents(v)\}\}$. We define causal boundary $B$ as a set of points in the closure of $S$ but not belonging to the interior of $S$. We therefore propose to adopt a causal weighting function $w^c_i(t)$ to fulfil the requirement, such that $w^c_i(t) = \frac{K^c_{\beta}(t-t^*)}{\sum_{t=1}^{T} K^c_{\beta}(t-t)}$ where $K^c_{\beta}()$ is an asymmetric and non-negative kernel function satisfying CMA. Hence, we rewrite the network structure estimation to:

$$\hat{A}_i^t = \arg\min_{A_i^t \in \mathbb{R}^{1 \times T}} \frac{1}{T} \sum_{t=1}^{T} w^c_i(t)(\hat{A}_i^{t-1} - x_i^t)^2 + \lambda ||A_i^t||_1. \quad (4)$$

The use of asymmetric kernel for non-parametric regression can be found in economic literature (Gospodinov and Hirukawa 2008) (Gospodinov and Hirukawa 2012) but rarely discussed in structure discovery. Mackenzie and Tieu (Mackenzie and Tieu 2004) discuss an application of asymmetric kernel regression to radial-basis neural networks with an opinion that the available real-life data reproduce a causal function; and therefore, are naturally bounded by an interval. Hence, a truncation of a symmetric kernel at the boundary makes the model to suffer from material bias error. Although there are several attempts to resolve boundary problem (Hall and Wehrly 1991) (Zhang, Karunamuni, and Jones 1999), most of them cannot correct bias without increasing noise level and/or variance error (Mackenzie and Tieu 2004). Mackenzie and Tieu propose to correct boundary error by replacing a symmetric kernel with an asymmetric one. Apart from the favorable boundary property (Mackenzie and Tieu 2004) – maintaining similar level of variance and bias trade-off, asymmetric kernel can also provide a weighting function which is within causal boundary $B$. We define $K^c_{\beta}()$ by using gamma function (Bain and Engelhardt 1992).

For the case of using kernel regression to estimate functional relationship, e.g., $y_i = y(t_i) + \epsilon$, where $i \in \{1 \ldots N\}$, $0 \leq t_i \leq T$ and $\epsilon$ is random noise (Mackenzie and Tieu 2004). By using symmetric Gaussian kernel $(K_{\beta})$ with boundary, we obtain significant bias at the boundary as the odd moments of $K_{\beta}$ are no longer zero due to truncation (Mackenzie and Tieu 2004):

$$bias[\hat{y}(\eta)] = \left\{ y(\eta) \int_{0}^{\infty} K_{\beta}(t-\eta)dt + y'(\eta) \int_{0}^{\infty} (t-\eta)K_{\beta}(t-\eta)dt \right\} + \frac{1}{2} y''(\eta) \int_{0}^{\infty} (t-\eta)^2K_{\beta}(t-\eta)dt + \ldots - y(\eta), \quad (5)$$

where $\hat{y}(\eta)$ is a Priestley-Chao estimator (Mackenzie and Tieu 2004) of $y(\eta)$; versus the scenario of no boundary:

$$bias[\hat{y}(\eta)] = \int_{-\infty}^{\infty} y(t_i)K_{\beta}(\eta - t_i)dt_i - y(\eta) \quad (6)$$

$$\simeq \frac{\sigma^2}{2} y''(\eta) + \sigma^4 \frac{y^{'''}(\eta)}{8}. \quad (7)$$

However, the boundary error term is vanished by replacing symmetric kernel $K_{\beta}$ by an asymmetric Gamma kernel $(K_{\beta})$ in the case of kernel regression with boundary:

$$bias[\hat{y}(\eta)] = \frac{\sigma^2}{2} y''(\eta) + \frac{\sigma^4}{3\eta} y^{'''}(\eta) + \frac{\sigma^4}{8} \left(1 + 2 \left(\frac{\eta}{\sigma}\right)\right) y^{''''}(\eta) + \ldots. \quad (7)$$

As a result, we transform TV-DBN to CTV-DBN by adopting Equation (4) to be our new objective function. The CTV-DBN structure discovery algorithm is summarized in Algorithm 1. In this implementation, the weighting function $w^c_i(t)$ is pushed into the square loss function by rescaling $x_i^t$ and $x_i^{t-1}$ such that it becomes a standard $\ell_1$-regularized least-squares problem.

**Experiments**

**Structure Discovery for Ring Road System**

We apply CTV-DBN on Beijing taxi trajectories from Complex Engineered Systems Lab, Tsinghua University, China, where Beijing is a densely populated city with special ring topology road structure. The dataset consists of one month of trajectories of 28,000 taxis in Beijing captured in May 2009, where each record includes the following information: 1) a taxi identifier, 2) a time-stamp in UTC of the time when the location was taken, and 3) latitude and 4) longitude specifying the position of the taxi.

The trajectories are firstly passed through a spatial filter with a boundary of Beijing city centre at the Forbidden City (city centre) and extended to its three international airport terminals (top right hand corner). We then apply density-based clustering method DBSCAN on one week (Monday-Friday) of trajectories at 8am to obtain a driver's

http://sensor.ee.tsinghua.edu.cn/datasets.php

\(\text{a rectangle formed by latitude and longitude pairs (40.08200, 116.16054) and (39.75030, 116.62000)}\)
Algorithm 1 CTV-DBN Structure Discovery

1: Initialize $\hat{A}^0$ randomly
2: for $i \in \{1 \ldots r\}$ do
3:   for $t^* \in \{1 \ldots T\}$ do
4:     $\hat{A}^t_{i} \leftarrow \hat{A}^{t-1}_{i}$ \# Warm-start initialization
5:     $\hat{x}^t_i \leftarrow \sqrt{w^t_i(t)} \hat{x}^{t-1}_{i} \hat{x}^t_i \leftarrow \sqrt{w^t_i(t)} \hat{x}^{t-1}_{i}, \forall t \in \{1 \ldots T\}$ \# Rescaling
6:     while $\hat{A}^t_{i}$ not converges do
7:       \# Search for the argument of the minimum based on Equation (4)
8:       for $j \in \{1 \ldots r\}$ do
9:       $S_j \leftarrow 2T \sum_{t=1}^{T} \left( \sum_{k \neq j} A^t_{ik} \hat{x}^{t-1}_{k} - \hat{x}^t_i \right) \hat{x}^{t-1}_{j} \# \beta_j \leftarrow 2T \sum_{t=1}^{T} \hat{x}^{t-1}_{j} \hat{x}^{t-1}_{j}$
10:      $A^t_{ij} \leftarrow \frac{\text{sign}(S_j - \lambda) \lambda - S_j}{\beta_j}$, if $|S_j| > \lambda; 0$ otherwise
11:     end for
12:   end while
13: end for
14: end for
15: return $\{\hat{A}^1, \ldots, \hat{A}^T\}$

view of regions. The trajectory average speeds within clusters are calculated.

We obtain estimated network structures by using the structure discovery process of relaxed TV-DBN and CTV-DBN. The structures at 8:20am and 8:30am are shown in Figure 1 as examples, where the cells filled with black colour represent connections and blank otherwise. Because of the truncation at causal boundary $B$, the method of CTV-DBN with truncated $K_{\Gamma}$ is expected to suffer from higher bias as well as information loss (Figures 1(a) and 1(b)). So far, we can identify regions 3, 6, 11, 16, 22, 23, 24 and 28 (let’s call them X) are the ones heavily depending on nearly all regions in the city. Apart from region 16, all the other regions are between the city centre and the airports.

Finally, the method of CTV-DBN with $K_{\Gamma}$ (Figures 1(c) and 1(d)) does not only come with a theoretical strength of low bias at the causal boundary $B$ and satisfying CMA, it also reveals more details of causal relationships among regions. Based on the same level of regularization, regions with insufficient causal connections are eliminated in CTV-DBN (versus relaxed TV-DBN) and additional connections are added based on the evidence within causal boundary $B$. The structural differences between CTV-DBN with $K_{\Gamma}$ and relaxed TV-DBN with $K_{G}$ are mainly from the enforced causal relationship in the former.

Out of all the regions in X, only regions 3, 24 and 28 are the top 3 regions causally impacted by most of the regions and they are all located along the Beijing Airport Expressway (S12)\(^3\) between the city and the airports, in which traffic

\[^3\text{http://en.wikipedia.org/wiki/Airport_Expressway_(Beijing)}\]
jam is common\textsuperscript{4}. Since the three regions are located just at or before ring roads\textsuperscript{6} which diverge traffic to all major districts in the city, any congestion in the other regions would have ultimate impact to these three regions (major artery between the city and the airports). These findings have also been confirmed by (Zhao, Lu, and de Roo 2011) and from numerous published facts, such as a report by China Central Television\textsuperscript{7} about the relationship between ring roads and other districts in the Beijing city.

**Trajectory Clustering Between Two Busy Regions**

We integrate the trajectory clustering algorithm, TRACLUS(Lee, Han, and Whang 2007), to mine trajectory patterns among regions of interest. Interesting regions are first identified by their time-varying traffic relationship as mentioned previously. The trajectory clustering algorithm consists of two phases of partitioning and grouping. It partitions trajectories into set of lines and similar line segments create sub-trajectory clusters. Unlike Gaffney (Gaffney et al. 2007; Gaffney and Smyth 1999) that cluster trajectories as a whole, TRACLUS discover similar moving object behaviours in sub-trajectories. Therefore, it is applicable for common behaviour mining of moving objects among regions of interests.

Due to space limitation, here we show an example of two snapshots, Figures 2 and 3, of the trajectory clusters between two busy regions (near CBD of Beijing). We pick these two regions as example as the casual relationship from CTV-DBN has shown that these two regions are very connected. The centers of these two regions are 10 km apart. Each region has a radius of 2.5 km. A trajectory is of our interest if it passes both regions. During our experiment, we found that many stationary points during peak hours especially during morning peak hours. After consulting the locals, this is due to the fact that many taxis are stuck in heavy congested roads with almost no movement, and also many drivers refuse to take passengers in the heavy congested regions during peak hours.

In Figures 2 and 3, each black line represents a trajectory cluster. In general, on-peak black lines are shorter and discrete, while off-peak black lines are usually longer and continuous. This is due to the fact that drivers make instantaneous decision to detour in order to avoid congestions. By studying the trajectory clusters (frequent routes), we can investigate the driver behavior that has shown to be spatio-temporal (i.e., depends on time and location, e.g., busy region, peak hours etc).

---

\textsuperscript{4}http://www.bjtgl.gov.cn/publish/portal1/
\textsuperscript{5}http://wikitravel.org/en/Beijing
\textsuperscript{6}http://en.wikipedia.org/wiki/Ring_roads_of_Beijing
\textsuperscript{7}http://www.cctv.com/lm/124/41/90128.html

**Conclusion**

This paper presents a notion of causal time-varying dynamic Bayesian networks (CTV-DBN), as well as defining causal boundary for cross time information sharing. In CTV-DBN, causal Markov assumption is satisfied by using asymmetric Gamma kernel ($K_T$). We apply CTV-DBN to spatio-temporal data by combining the model with density-based clustering for region discovery. By revealing the time-varying region structures using moving objects’ view of territories, causal relationships among regions are captured and available for causal inference. Regions of interest are further analyzed by considering trajectory clusters between them.

**References**


