CIT2016 – XII Congreso de Ingeniería del Transporte
València, Universitat Politècnica de València, 2016.
DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

Classification and prediction of port variables using Bayesian Networks

Beatriz Molina Serrano
Doctoral candidate and researcher. Civil Engineering Department. Transports; Polythecnic University of Madrid. Madrid-Spain
María Nicoletta González Cancelas
Civil Engineer Doctor. Assistant Professor. Civil Engineering Department. Transports; Polythecnic University of Madrid. Madrid-Spain
Francisco Soler Flores
Technology and Information System Doctor. Associate Professor. Mathematics and Computers applied to Civil and Naval Engineering Department; Polythecnic University of Madrid. Madrid-Spain
Alberto Camarero Orive
Civil Engineer Doctor. Principal Professor. Civil Engineering Department. Transports; Polythecnic University of Madrid. Madrid-Spain

ABSTRACT

Many variables are included in planning and management of port terminals. They can be economic, social, environmental and institutional. Agent needs to know relationship between these variables to modify planning conditions. Use of Bayesian Networks allows for classifying, predicting and diagnosing these variables. Bayesian Networks allow for estimating subsequent probability of unknown variables, basing on know variables.
In planning level, it means that it is not necessary to know all variables because their relationships are known. Agent can know interesting information about how port variables are connected. It can be interpreted as cause-effect relationship. Bayesian Networks can be used to make optimal decisions by introduction of possible actions and utility of their results.

In proposed methodology, a data base has been generated with more than 40 port variables. They have been classified in economic, social, environmental and institutional variables, in the same way that smart port studies in Spanish Port System make. From this data base, a network has been generated using a non-cyclic conducted grafo which allows for knowing port variable relationships - parents-children relationships-. Obtained network exhibits that economic variables are – in cause-effect terms- cause of rest of variable typologies. Economic variables represent parent role in the most of cases. Moreover, when environmental variables are known, obtained network allows for estimating subsequent probability of social variables.

It has been concluded that Bayesian Networks allow for modeling uncertainty in a probabilistic way, even when number of variables is high as occurs in planning and management of port terminals.

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

1. INTRODUCTION

Sustainable development is beeing applied emergently by transport authorities and in other activity sectors and industries all over the World. It is proponted by iniciatives which include environmental variables and social responsibility in strategic management of companies (Doerr, 2011). This is port case. Port sustainability is rooted in the proposals of the GRI (Global Reporting Initiative, 2000) and it preserves the four main ideas or dimensions which knock into shape sustainable development – institutional, economic, environmental and social dimensions- It is considered that sustainable management of a company or organism has as main target equilibrate keeping of its function and activity for a long time. So, it is necessary looking for a equilibrate development of economic, social, environmental and institutional dimensions (Serrano, 2015).

In this context, maritime transport requires an especial attention, because it transports over 80% foreign trade (tonelates-kilometres) (Sánchez, Jaimurzino, Willmsmeier, Pérez, Doerr y Pinto, 2015). So, sustainable management must be understood in port sector as "management which allows containers traffic, solid and liquid granels, general goods and number of passenger grow up at the same time that energy and natural resources purchase, rubbish volume and negative impacts over social systems and ecosystems decreases in port influence areas" (Crespo, Ripoll, Crespo y Giner, 2005). It is necessary that environmental and social variables gain importance to make port development travels to sustainability (Grupo de trabajo 23, 2004).

Spanish State Ports and Merchand Navy Law includes sustainability as one of the principles which must regulate planning model and port management. Article 55.4 foresees company planning of each Port Authority must be gone with a sustainability report- it is considered as an analysis and diagnostic tool-. This report use a methodology which is based on GRI´s one, but it does not determine Port Authority behaviour, it only describes results using performance quality indicators (Crespo, Giner, Morales, Pontet y Ripoll, 2007). Sustainability quality indicators are usually used because they allow an evaluation of sustainable development management. Widespread application to a port system is useful to perform accurate benchmarking in sustainability between ports in the same region or country (González, Guerra, Martín, Nóvoa, Otero y Penela, 2010).

Application of these tools emerge objectives (economics, environmental, socials and institutionals) whose must be achieved by a port authority or port company to assure its sustainable development and its port growth (Autoridad Portuaria de A Coruña, Autoridad Portuaria de Valencia, Organismo Público Puertos del Estado, 2008). Therefore, the objective pursued is that through these four main sustainability ideas, ports conform as a system and not be seen as isolated entities and subject to a specific business situation, but as elements that interact with an environment physical, social and environmental, which are to be integrated effectively, that is, being able to adapt to a changing situation and in

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

turn, pointing to a renewal that will help achieve the best possible future scenarios (Puertos del Estado, 2011).

2. METHODOLOGY AND RESULTS

Probabilistic graphical models are graphs in which nodes represent random variables, and the (lack of) arcs represent conditional independence assumptions. Hence they provide a compact representation of joint probability distributions. Undirected graphical models, also called Markov Random Fields (MRFs) or Markov Networks, have a simple definition of independence: two (sets of) nodes A and B are conditionally independent given a third set, C, if all paths between nodes A and B are separated by a node in C. By contrast, directed graphical models, also called Bayesian Networks or Belief Networks (BNs), have a more complicated notion of independence, which takes into account the directionality of the arcs, as we explain below (Almazán-Gárate, Palomino-Monzón, González-Cancelas and Soler-Flores, 2014). Undirected graphical model are more popular with physics and vision communities and directed models are more popular with AI and statistic communities (it is possible to have a model in which both directed and undirected arcs are included. This model is called chain graph). For a careful study of relationship between directed and undirected graphical models, review references (Castillo, Gutiérrez and Hadi, 1997), (Duda, Hart and Stork, 2001) and (Pearl, 1982).

Selected variables in the study are listed in Table 1, which is included below:

ID	Description
herramgestion_dimins	Management systems to support decision-making
geninfraortuaria_dimins	Port Authority role as infrastructure dealer
mercservidos_dimins	Structure and evolution of main good traffics
dinamact_dimins	Main sector or activities which are relevant in local economic development
servicios concauto_dimins	Types, delivery framework and regulation
inicprivada_dimins	Private iniciative presence
transconcu_dimins	Initiatives to ensure that any operator could provide services in port
calidserv_dimins	Initiatives promoted by Port Authority to improve efficiency
intetrans_dimins	Port integration in transport system
sitecofin_dimecon	Economic and finantial situation (EBITDA/tonne, etc.)
inv_dimecon	Level and structure of investments
negserv_dimecon	Business and service (income occupancy rates and activity, etc.)
vgenprod_dimecon	Generated value and productivity
caphum_dimsoci	Employment, internal communication, training, etc.

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

ID	Description
empl_dimsoci	Employment and job security in the port community.
gestamb_dimma	Degree of implementation of environmental management systems (EMAS, ISO 14001 and PERLS) and expenses invested financial resources
calaire_dimma	Air quality: main sources involving significant emissions
calagua_dimma	Water quality : main source discharges located at port to have a significant impact on water quality
calacust_dimma	Sound quality : major emission sources (point and diffuse) port involving significant noise pollution
residuos_dimma	Rubbish management
ecoef_dimma	Efficiency in land use, water and energy consumption
comport_dimma	Conditions or requirements on environmental issues

Table 1 – Selected factors

Although directed models have a more complicated notion of independence than undirected ones, but they have several advantages. The main advantage is that everyone can regard an arc from A to B to indicate A "causes" B. This can be uses as a guide to build graph structure. Moreover, directed models can encode deterministic relationships, and they are easier to learn (fit to data). In addition, it is necessary to specify parameters of the model to define graph structure. In a directed model Conditional Probability Distribution (CPD) must be specify for each node. If variables are discrete, it can be represented as a table (CPT) in which there is listed probability of a child node takes on all its different values for each combination of its parent´s values.

2.1 Parameter learning

In order to specify Bayesian Network and thus fully represent the joint probability distribution, it is necessary to specify probability distribution for X conditional upon X´s parents in each node X. Distribution of X conditional upon its parents may have any form. It is very common using discrete or Gaussian distributions when this simplifies calculations. Sometimes, only constraints on a distribution are known; in this case, one can use maximum entropy principle to define a single distribution, The one which has greatest entropy gives constraints. Analogously, in a specific context of a dynamic Bayesian Network, it is usual that one specifies conditional distribution of hidden state´s temporal evolution to maximize entropy rate of implied stochastic process.

These conditional distributions often include parameters which are unknown and must be estimated from data. Sometimes, parameters are estimated using maximum likelihood approach. Maximization of likelihood (or posterior probability) in a direct way is often complex when there are unobserved variables. A typical approach of this problem is the expectation-maximization algorithm. This algorithm alternates computing expected values of unobserved variables which are determined by observed data, and maximizing full likelihood (or posterior) considering previous computed expected values are correct. Under

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

mild regularity conditions, this process converges on parameter values of maximum likelihood (or maximum posterior).

A deeper Bayesian parameter approach is treating parameters as additional unobserved variables and computing a full posterior distribution over all nodes conditional upon observed data to integrate out parameters. This approach can be expensive and lead to large dimension models, so in practice it is more common use of classical parameter-setting approaches.

2.2 Learning

It is necessary to specify two things to describe a Bayesian Network: graph topology (structure) and parameters of each Conditional Probability Distribution (CPD). It makes possible to learn both of them using data. However, learning structure is harder than learning parameters. Also, when some of nodes are hidden or there are missing data learning is harder than learning when everything is observed.

Learning Bayesian Network structure is considered a harder problem than learning Bayesian Network parameters. Moreover, another obstacle arises in situation of partial observably in case of nodes are hidden or when dais missing. Simple case is a Bayesian Network which is specified by an expert. Then, it is used to perform inference. In other applications, task of network definition is too complex for human people, so network structure and local distribution parameters must be learned from data. Automatically learning of a Bayesian Network graph structure is a challenge pursued within machine learning. In this case, obtained network is displayed by K2 algorithm. It can be observed that variable has been pulled away (Fig. 1)

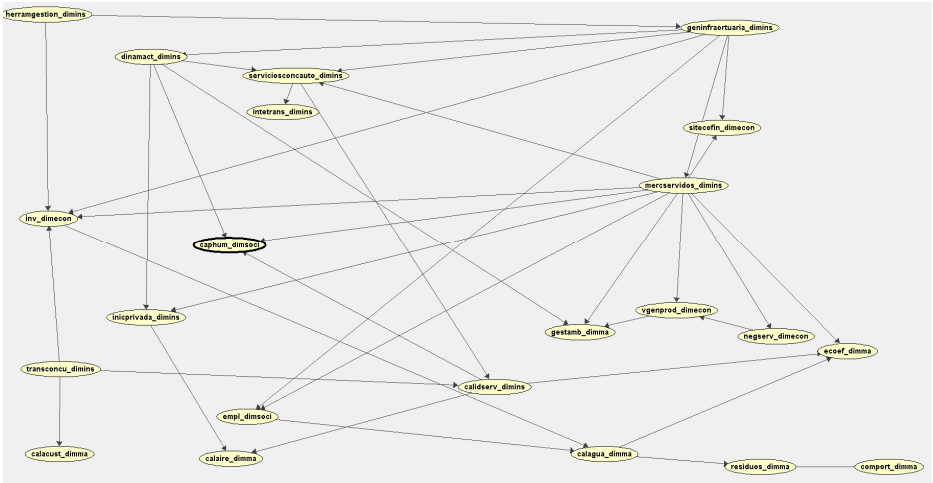

Fig. 1 – Bayesian Network. Algorithm K2

Topology needs to identify factors that are relevant, to determine how those factors are causally related to each other. The arc cause-effect does mean that cause is a factor involved in causing effect. In this case, for example, parent of the variables inv_dimecon and geoinfraortuaria_dimins is the node herramgestion_dimins. Then, when herramgestion_dimins is known, inv_dimecon and geoinfraortuaria_dimins, are

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

conditionally independent (Fig 2).

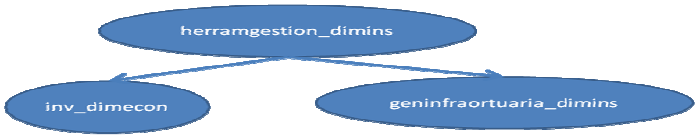

Fig 2. Relationship 1

Herramgestion_dimins is a resolution variable which appears in network as a "node". Some arcs started on it, so this variable generates a divergent connection. This way, herramgestion_dimins is a parent node which projects arcs to several sons, that is to say, arrows start in this variable and diverges to its sons (Fig 2).

When parent variable state is known, there is a dependence relationship between variables. However, when a parent state is unknown, son variables are taken in an independent way and information will not spread along network if some evidences are included over son nodes (Fig 2). An effect that has two or more ingoing arcs from other vertices is a common effect of those causes. A cause that has two or more outgoing arcs to other vertices is a common cause (factor) of those effects. The effects of a common cause are usually observables.

Following the Bayesian Network independence assumption, several independence statements can be observed in this case, in respect to each of the factors. When mercservidos_dimins is known, sitecofin_dimecon, servicios concauto_dimins, inicprivada_dimins, ecoef_dimma, vgenprod_dimecon, empl-dimsoci, inv_dimecon, negserv_dimecon, gestamb_dimma, and caphum_dimensoci are conditionally independent of its ancestor's geoinfraortuaria_dimins. (Fig 3)

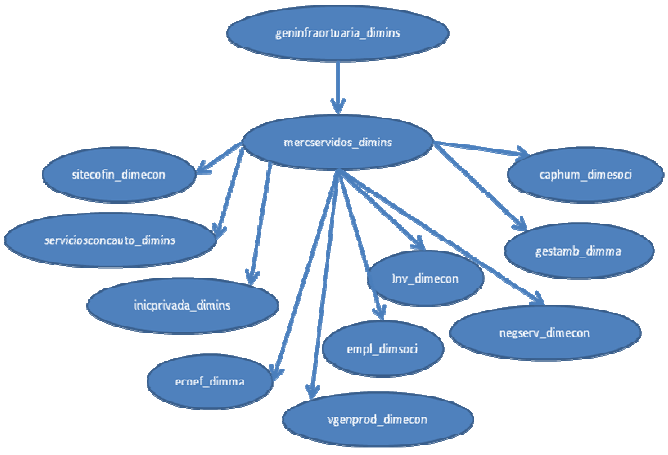

Fig 3. Relationship 3

Casual graph: The variable mercservidos_dimins has ten common effects: sitecofin_dimecon, servicios concauto_dimins, inicprivada_dimins, ecoef_dimma, vgenprod_dimecon, empl-dimsoci, inv_dimecon, negserv_dimecon, gestamb_dimma, and caphum_dimensoci (Fig. 5)

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

3. CONCLUSIONS

The most decision-making category, as network obtained by using the algorithm K2 shows, is institutional category, then economic and social at the same height, and finally environmental category. Management systems supporting decision-making includes quality management systems, scorecards, market characterization campaigns, etc, and they are represented by herramgestion_dimins. It is considered as a parent variable in network. The same goes for transconcu_dimins which represents initiatives to ensure that any operator could provide services in port. It is a parent node in network too. Furthermore, transcocu_dimins is a decision-making variable, so it appears in network as a "node", so a divergent connection is created and this parent node throws its arcs toward several of its sons.

Other essential variable in network structure is mercservidos_dimins. 10 arrows star on it and go to 10 different nodes. These are effects of structure and main good traffic evolution, so they are social, economic, institutional and environmental effects. That is to say, served markets have effects on rates, delivery framework and regulation of port services the number of companies operating in the port (institutional category). It has effects on EBITDA, EBITDA/tonne, public investment relative to cash flow and: income from employment and activity rates among others too (economic category). Even, about social status, it has effects on variables representing port community employment, job security and training services and health work, among others. Finally, in environmental category, served markets causes different grades of environmental management systems implementation (EMAS, ISO 14001 y PERLS), economic resource investment and investments associated to implementation, certification and maintenance of environmental management system. Therefore, served markets are a very important variable in planning from a sustainable perspective.

In other hand, institutional variables are interconnected. Economics ones are important as cause-effect because they are effects of served markets which belong to institutional dimension. Generated value and productivity depend on kind of business and service. Moreover, social variables are effects of institutional variables but they have not a direct relationship with their same dimension, social one. Finally, environmental variables are closely interconnected in Bayesian Network and they are principally effects of institutional category. Therefore, economic, social and environmental variables are effects of institutional ones.

As a conclusion, key issue is that Port Authorities start to incorporate sustainable elements-included in Port Law- in their tools, used to regulate port services and public possession management.

DOI: http://dx.doi.org/10.4995/CIT2016.2016.3226

REFERENCES

ALMAZÁN-GÁRATE, J.L.; PALOMINO-MONZÓN, M.C.; GONZÁLEZ-CANCELAS, N.; SOLER-FLORES, F. (2014). Relationship between air pollution and natural gas with respect to maritime transport. Methodology based on Bayesian Networks. *Global Virtual Conference. 7-11 April 2014.* Transport and Logistics Section.

AUTORIDAD PORTUARIA DE A CORUÑA, AUTORIDAD PORTUARIA DE VALENCIA, ORGANISMO PÚBLICO PUERTOS DEL ESTADO (2008). Guía para la elaboración de memorias de sostenibilidad en el sistema portuario español. FEPORTS

CASTILLO, E.; GUTIÉRREZ, J.M. and HADI, A.S. (1997). Expert Systems and Probabilistic Network Models. Springer Verlag, New York

CRESPO SOLER, C.; RIPOLL FELIU; V.M., CRESPO TRUJILLO, A.M.; GINER FILLOL, A.; (2005). La sostenibilidad ambiental en el sistema portuario de titularidad estatal. *XIII Congreso AECA. Armonización y Gobierno de la Diversidad. 22-24 Septiembre 2005*

CRESPO SOLER, C.; GINER FILLOL, A.; MORALES BARAZA, J.A.; PONTE TUBAL, N; RIPOLL FELIU; V.M. (2007). La información de sostenibilidad en el marco de las cuentas anuales: análisis aplicado al caso de la Autoridad Portuaria de Valencia. *Revista do Contabilidado de Maestrado em Ciências Contábeis da UERJ, Rio Janeiro, v-12, n.3, p-11 set./dez, 2007*

DOERR, O. (2011). Políticas portuarias sostenibles. *Boletín FAL. CEPAL. Edición nº 299, número 7 de 2011*

DUDA, R.O.; HART, P. E. and STORK, D.G. (2001). Pattern Classification. Wiley, New York

GLOBAL REPORTING INITIATIVE (2000). Guía para la elaboración de Memorias de Sostenibilidad. Versión 3.1. GRI

GONZÁLEZ, F.; GUERRA, A.; MARTÍN, F.; NÓVOA, J.J.; OTERO, C. y PENELA, J. (2010). Medición de la sostenibilidad en el sistema portuario Español: propuesta metodológica a través de indicadores sintéticos de desarrollo sostenible. *XII Reunión de economía mundial, mayo de 2010*. Santiago de Compostela

GRUPO DE TRABAJO 23 (2004). La sostenibilidad en los puertos. *CONAMA VII Cumbre del desarrollo sostenible, 24 de noviembre de 2004*

PEARL, J. (1982). The Solution for the Branching Factor of the Alpha-Beta Pruning Algorithm and its Optimality. *Communications of the ACM. 1982*. Vol 25, no.8

PUERTOS DEL ESTADO (2011). Memoria de sostenibilidad del sistema portuario de interés general.

SÁNCHEZ, R.J.; JAUIMURZONO, A.; WILLMSMEIER, G.; PÉREZ SALAS, G.; DOERR, O.; PINTO, F. (2015). Transporte marítimo y Puertos. Desafío y oportunidades en busca de un desarrollo sostenible de América Latina y El Caribe. *CEPAL. Serie Recursos Naturales e Infraestructura.*

SERRANO, O. (2015). Operativa portuaria y sostenibilidad. *CONAMA LOCAL 2015, 7 octubre de 2015*. Málaga