

DISEÑO DE UN MODELO DE PLANIFICACIÓN DE ZONAS DE ACTIVIDADES LOGÍSTICAS MEDIANTE EL EMPLEO DE REDES BAYESIANAS

AREAS OF ACTIVITIES LOGISTICS PLANIFICATION MODEL DESING USING BAYESIAN NETWORKS

Francisco Soler Flores^{1,*}, Nicoletta González Cancelas¹, Alberto Camarero Orive¹, José Luis Almazán Gárate², María del Carmen Palomino Monzón²

RESUMEN

La dificultad para el establecimiento de pautas de dimensionamiento de Plataformas Logísticas y, en especial, de las Zonas de Actividades Logísticas (ZAL), recae en la heterogeneidad del desarrollo de este tipo de nodos de transporte a nivel internacional. Bajo la denominación genérica de Plataformas Logísticas han surgido multitud de iniciativas en la escena internacional, que, respondiendo a diferentes motivos de implantación de un nodo de intercambio modal, ha producido la aparición de diferentes tipos de Plataformas con diversos objetivos que implican unidades funcionales específicas, con necesidades de localización, instalación y superficie necesaria diferentes. Este sector logístico tan importante, se encuentra sin metodologías, herramientas o programas que permitan establecer los parámetros de planificación y explotación óptimos para las diferentes zonas de actividades logísticas, si bien se han desarrollado tecnologías de trazabilidad de la carga y elementos basadas en la planificación logística, con el objetivo de determinar los parámetros óptimos de explotación y planificación portuaria, a través de la clasificación de las zonas de actividades logísticas, añadiendo la inferencia de escenarios virtuales. Como resultado principal se destaca que, mediante el empleo de herramientas de inteligencia artificial, modelos gráficos probabilísticos: Redes Bayesianas (BN), se han definido las principales variables de planificación de las zonas de actividades logísticas y se puede planificar con BN una ZAL seleccionando las variables conocidas y obteniendo las variables a predecir.

Palabras Claves: Zonas de Actividades Logísticas, Planificación, Minería de datos, Redes Bayesianas.

ABSTRACT

The difficulty of establishing guidelines for design of Logistics Platforms and specially Logistics Activity Zones is the heterogeneity of developing this type of transport nodes worldwide. Many initiatives under the generic term of Logistics Platforms have emerged to the international scene, responding to different reasons for implementing a modal interchange node, different types of platforms have been developed with different objectives. In this way specific functional units like location, installation and area are needed. All initiatives for Logistics Platforms have been developed in line with broad temporal changes in the priorities for planning

¹Departamento de Ingeniería Civil. Transportes, Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Madrid, España

²Departamento de Ingeniería Civil: Ordenación Del Territorio, Urbanismo y Medio Ambiente, Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid, Madrid, España

*Autor para correspondencia: f.soler@upm.es

Recibido: 22.07.2012 Aceptado: 09.03.2013

and design criteria that implies. The logistics sector is important and there isn't a methodology, tools or programs to set the parameters for optimal planning and operation for different logistic areas, although have developed technologies and cargo tracking and based on the planning logistics elements. The main planning variables of logistic zones are defined through classification of areas and with inference in virtual scenarios through the use of artificial intelligence tools, using Bayesian Networks.

Keywords: Logistics Platforms, Planning, Data Mining, Bayesian Networks

INTRODUCCIÓN

Los procesos de globalización han provocado que las cadenas de transporte sean cada vez más complejas, y sea necesario el empleo de recursos logísticos para optimizar el transporte. Paralelamente han surgido nodos de intercambio, infraestructuras logísticas de transporte, para mejorar la eficiencia de las cadenas de transporte (MUÑOZ et al. 2005).

Entre otras Plataformas Logísticas se han desarrollado las Zonas de Actividades Logísticas (ZALs), asociadas principalmente a los puertos. Los puertos son nodos de intercambio modal que han evolucionado atendiendo a las necesidades del mercado. Se encuentran en el sistema de transporte, de manera que actúan como un elemento más de la cadena de transporte, por lo que deben ser eficientes para no constituir un cuello de botella en dicha cadena. Para ello, y con el objetivo de mejorar los servicios del puerto y, por ende, la cadena de transporte que tiene un eslabón en el puerto, se ha favorecido la creación de Zonas de Actividades Logísticas (ZALs en adelante) adyacentes al puerto, en las que se llevan a cabo actividades de almacenaje, agrupamiento y distribución, e incorporación de valor añadido (Camarero & González, 2005), como consecuencia del rediseño de las redes de distribución europeas, que tienden a concentrarse en un número limitado de centros y rutas comerciales.

Lo que se pretende en este artículo es emplear las redes bayesianas, partir de la red construida y, al realizar inferencia, dadas ciertas variables conocidas de una ZAL (evidencia), calcular la probabilidad posterior de las demás variables de una ZAL (desconocidas)". Las redes bayesianas permiten modelar la incertidumbre de forma probabilística, incluso cuando el número de variables es elevado. Se basan en utilizar la noción de independencia para generar una representación compacta y modular de la distribución probabilística de las variables. La tarea fundamental en Redes Bayesianas es la inferencia: conocer la distribución de unas variables cuando se observan otras, que es lo que se pretende obtener en esta investigación, conocidas unas variables de partida para la planificación de una ZAL conocer la distribución del resto de las variables de planificación de dicha ZAL.

Los puertos están inmersos, a la vez que son impulsores, en los procesos de globalización de las sociedades, y es por ello que no son ajenos a la potencialidad de nuevas metodologías, como pueden ser los sistemas de inteligencia artificial. Así, este proceso de apertura constante impulsa el crecimiento económico global. Respecto al transporte marítimo mundial, más de 8.400 millones de toneladas se transportan por esta vía (año 2011) para abastecer a los 7.000 millones de personas que habitan el planeta. El tráfico de los contenedores supera los 500 millones de TEUs transportados (Development. Secretariat 2011). El sistema portuario español mueve alrededor de 450 millones de toneladas (año 2011), lo que representa aproximadamente un 3% del tráfico portuario total del mundo y más de un 10% del tráfico portuario de la Unión Europea, por lo que es uno de los puntales de la economía nacional (Rodríguez-Dapena A. 2009).

En el ámbito portuario, los servicios logísticos han dejado de ser una novedad, para convertirse en una necesidad, en una opción que los clientes de los puertos empiezan a exigir, por lo que las ZALS se están convirtiendo en un instrumento clave para el desarrollo integral del puerto y en un factor de competitividad. Ante esta situación, y con la necesidad de caracterizar las ZALS asociadas a puertos del sistema portuario español, se emplean modelos gráficos probabilísticos (redes bayesianas) con el objetivo de determinar los parámetros óptimos de explotación y planificación portuaria, a través de la clasificación de las zonas de actividades logísticas, añadiendo la inferencia de escenarios virtuales.

La tendencia general en el análisis de la explotación y planificación de nodos o Plataformas Logísticas es comparar los ratios y parámetros de la bibliografía internacional. Una de las bibliografías más empleadas es (Rodrigue *et al.* 2009), en la que se analiza que, según la teoría de polos de crecimiento, el desarrollo no es uniforme y se lleva a cabo en lugares específicos, alrededor del cual se desarrollan actividades de aglomeración. Por las infraestructuras que existen en torno a una actividad, se puede mejorar la accesibilidad a los proveedores y clientes. Las terminales de transporte, por lo tanto, son fuerzas económicas, generan vínculos con otros sectores de la economía y se convierten en focos de actividad económica, por lo que con frecuencia se consideran como polos de crecimiento. Por ejemplo, una estrategia de crecimiento para los terminales terrestres gira en torno a la formación de Zonas de Actividades Logísticas, donde los centros de distribución comparten instalaciones comunes, incluyendo un mejor acceso a una terminal de transporte.

Para abordar la planificación de una ZAL portuaria es necesario identificar la etapa de desarrollo logístico del puerto concreto, su inserción y visión a largo plazo. Una ZAL es una pieza importante de la estrategia logística global del puerto. Su funcionalidad y, en definitiva su potencial de contribución a dicha estrategia, estará en función del grado de desarrollo logístico general del propio puerto, que es preciso identificar de una forma realista.

La heterogeneidad en el desarrollo de las Plataformas Logísticas en la escena internacional es un factor de dificultad para el establecimiento de pautas de dimensionamiento. Bajo la denominación genérica de Plataformas Logísticas se incluye toda una serie de iniciativas que, respondiendo a diferentes motivos de implantación, enmarcados dentro de esta denominación tan genérica, ha propiciado la aparición de diferentes nodos con objetivos diversos, que implican unidades funcionales particulares y unas necesidades de instalación y de superficie dispares. Las diferentes iniciativas han surgido, además, dentro de un proceso dilatado en el tiempo, con la codificación de las prioridades de los criterios de dimensionamiento que esto conlleva.

El proceso de desarrollo de las Plataformas Logísticas ha provocado una falta de homogeneización, sobre la que actúan muchos factores, algunos comunes a todas las actuaciones y otros impuestos por las condiciones de contorno de cada caso particular. La falta de estandarización de los criterios de dimensionamiento se debe principalmente a las cuestiones que a continuación se recogen:

La legislación sobre el suelo, diferente en cada país e incluso dentro de un mismo país, la titularidad del mismo y las condiciones de localización (suelo urbano & suelo no urbano), inciden en el precio final del mismo y en las posibilidades de desarrollo y dimensionamiento de la Plataforma Logística. Las diferentes condiciones que concurren en cada promoción provocan una importante dispersión de las opciones de la variable y la dificultad de su homogeneización.

Los flujos de tráfico que puede captar cada Plataforma Logística son muy variables; puede captarse de diversas zonas, ir a sitios diferentes, necesitar una cadena intermodal, etc. Este factor, junto con la variabilidad de la tipología del tráfico captado (graneles sólidos, graneles

líquidos, mercancía general, etc.), que requieren de diferentes equipos de manipulación, diferentes necesidades de superficie de almacenamiento y tipología de la instalación de almacenamiento, es determinante a la hora de configurar las dimensiones finales del centro logístico.

El período del proceso de creación de una Plataforma Logística provoca que las variaciones existentes en las demandas de mercado se traduzcan en una evolución de los criterios de dimensionamiento. Además, el propio concepto de Plataforma Logística ha sufrido modificaciones en su concepto, evolucionando desde los primeros centros de carretera hasta las actuales plataformas multimodales, modificaciones que han influido sustancialmente en el diseño de áreas funcionales.

La sensibilidad del sector sobre los espacios determina la ordenación del centro: si se concibe la Plataforma Logística como un espacio industrial de calidad, que significa principalmente que dicho espacio permite el desarrollo de centros muy extensos y poco densos (caso de Italia), ya que las empresas implicadas en el sector están dispuestas a pagar precios mayores por instalarse dentro de los centros a los que impone el mercado en localizaciones industriales de menor calidad, donde los centros son de menor espacio y más densos, como es el caso de Francia, donde los espacios son muy densos con un aprovechamiento mayor del suelo.

La promoción de Plataformas Logísticas ha recaído fundamentalmente en autoridades regionales y locales, y no en la iniciativa privada (exceptuando el caso de Italia), lo que ha supuesto una falta de integración en el ámbito nacional o supranacional de estos centros; la falta de iniciativas de estandarización por parte de las entidades supranacionales acentúa el problema, incluso de entidades naciones y regionales, dándoles un carácter localista a las iniciativas, por lo que la disparidad en la toma de decisiones de planeamiento se acentúa. Con la actual situación de crisis, el modelo a seguir debe incidir en la colaboración y participación público-privada. Pese a la relativización antes señalada de la validez de alcanzar un consenso sobre estándares de dimensionamiento de ZALs, es conveniente disponer de un marco de referencia de apoyo, tanto desde el punto de vista de promoción de iniciativas como de desarrollo de ZALs para poder abordar análisis concretos.

El diseño funcional y operativo de una ZAL portuaria exige un estudio sistemático del entorno portuario próximo. Las posibilidades de desarrollo de actividades logísticas de valor añadido, ligadas a la actividad portuaria, dependen de un número extenso de factores. La inserción del puerto en su entorno próximo condiciona sustancialmente, tanto su diseño funcional, como operativo y de negocio.

Una ZAL portuaria es, funcionalmente, una pieza singular del complejo logístico de un entorno portuario. El análisis pormenorizado de este entorno funcional no es, en consecuencia, una mera exigencia disciplinar o académica; de ese análisis pueden surgir múltiples alternativas de ubicación y ordenación, niveles de competencia o necesidades de actuación en proyectos complementarios, especialmente de infraestructuras complementarias de acceso a la ZAL, conexión con las terminales portuarias o de interrelación con otros polos logísticos y productivos del entorno.

El estudio y planificación de una zona de actividades logísticas lleva inherente el estudio del rendimiento de la misma. Tradicionalmente, la planificación de una infraestructura logística se realiza mediante ratios de eficiencia obtenidos del estudio de la explotación. Estos ratios son tomados como parámetros de referencia para proyectar nuevos nodos o planificar desarrollos futuros de otros ya existentes; por lo tanto, en el escenario de trabajo del rendimiento se emplean, habitualmente, con escasa precisión, diversos términos como tráfico, explotación óptima, capacidad, productividad, ocupación, eficiencia, etcétera.

En el análisis de eficiencia y rendimiento de infraestructuras de transporte existen dos grandes grupos de técnicas:

- La conocida como DEA (Data Envelopment Análisis), o Análisis Envoltante de Datos (Roll et al. 1993), (Tongzon et al. 2005), (Bonilla et al. 2004), ha sido tradicionalmente utilizada para la estimación de la eficiencia relativa de un conjunto de unidades productivas.
- La estimación econométrica de funciones y distancias (Liu et al. 2002), (Jara-Díaz et al. 1997), (Tovar et al. 2003), (Tovar et al. 2004), es una estimación empírica de funciones de costes; tiene su origen en la década de los años setenta con el trabajo de (Wanhill 1974), cuyo objetivo era diseñar un modelo que permitiera determinar el número de atraques óptimos que minimiza el coste total de uso del puerto, entendiendo como tal la suma de dos componentes: el coste de suministrar la infraestructura (el atraque) y el coste del tiempo del barco en el puerto. En el trabajo de (Wanhill 1974) se considera que la inversión y planificación futuras han de hacerse teniendo en cuenta que los servicios portuarios no se pueden almacenar y, por tanto, hay un trade-off entre el coste de la capacidad portuaria y el coste de permanencia de los barcos en el puerto (tiempo de servicio más tiempo de espera), que es determinante y ha de ser tenido en cuenta en el proceso de planificación (Bishop 2006).

La metodología de los modelos DEA supone la generalización del análisis tradicional de los ratios de actividad, permitiendo considerar de forma simultánea varios inputs y/o outputs. El DEA y la estimación de funciones de frontera son alternativas para calcular la frontera de producción y, por tanto, mediar la eficiencia en la producción y en los costes. A partir de estas dos técnicas se pueden obtener ratios de eficiencia relativa dentro de una muestra de unidades, que se comparan con la frontera de eficiencia. La estimación de funciones frontera supone el uso de métodos econométricos (métodos paramétricos), mientras que el DEA es un método no paramétrico, basado en el uso de la programación lineal.

En las últimas décadas se han desarrollado numerosas técnicas de análisis y modelización de datos en distintas áreas de la estadística y la inteligencia artificial (Todd et al. 1994), (Bishop 2006). La Minería de Datos (MD) es un área moderna interdisciplinar, que engloba a aquellas técnicas que operan de forma automática (requieren de la mínima intervención humana) y, además, son eficientes para trabajar con las grandes cantidades de información disponibles en las bases de datos de numerosos problemas prácticos (Lizaso Torres et al. 2011). Estas técnicas permiten extraer conocimiento útil (asociaciones entre variables, reglas, patrones, etc.) a partir de la información cruda almacenada, permitiendo así un mejor análisis y comprensión del problema. En algunos casos este conocimiento puede ser también post-procesado de forma automática, permitiendo obtener conclusiones e, incluso, tomar decisiones de forma casi automática, en situaciones prácticas concretas (sistemas inteligentes). La aplicación práctica de estas disciplinas se extiende a numerosos ámbitos comerciales y de investigación en problemas de predicción, clasificación o diagnóstico (Cios et al. 2007), (Witten et al. 2005). La aplicación de la minería de datos ha sido profusa en varias disciplinas, como la biología (Wong et al. 2006), economía (Vityaev et al. 2004), medicina (Cios et al. 2001).

Entre las diferentes técnicas disponibles en minería de datos, las redes probabilísticas o redes bayesianas permiten modelizar de forma conjunta toda la información relevante para un problema dado, utilizando posteriormente mecanismos de inferencia probabilística para obtener conclusiones en base a la evidencia disponible (Pearl 1988), (Castillo et al. 1997). Estos modelos se adaptan de forma natural a la concepción probabilística y local actual de la modelización en planificación.

Se puede decir que una Red Bayesiana es un conjunto de nodos y arcos. Cada nodo corresponde a una variable, que a su vez representa una entidad del mundo real, y los arcos que unen los nodos indican relaciones de influencia causal entre las variables. Una Red Bayesiana es un grafo acíclico dirigido, en el que cada nodo representa una variable y cada arco una dependencia probabilística, en la cual se especifica la probabilidad condicional de cada variable dados sus padres (Castillo et al. 1997). La variable a la que apunta el arco es dependiente (causa-efecto) de la que está en el origen de éste. La topología o estructura de la red ofrece información sobre las dependencias probabilísticas entre las variables y sus dependencias condicionales dada otra(s) variable(s). Dichas dependencias simplifican la representación del conocimiento (menos parámetros) y el razonamiento (propagación de las probabilidades). Una Red Bayesiana proporciona una forma compacta y modular de representar la distribución conjunta de varias variables aleatorias. Una Red Bayesiana consta de una parte cualitativa, que describe las relaciones entre las distintas variables, y una parte cuantitativa, que describe la fuerza de dichas relaciones mediante probabilidades condicionadas

La potencia de las Redes Bayesianas radica en que, una vez especificada la estructura de la red, es posible realizar cualquier tipo de inferencia dada la información disponible (Figura 1). Se plantea como la obtención de la distribución a posteriori de ciertas variables de la red, dado que se ha observado el valor que toman otras variables. Esto es lo que se conoce como "propagación de probabilidades". Es decir, se puede hacer inferencias predictivas (si la terminal de transporte tiene una superficie X , ¿cuál es la probabilidad de que tenga Z grúas de patio en el almacenamiento?) o abductivas (si la ZAL tiene menos de X grúas de patio en el almacenamiento, ¿cuál es la probabilidad de que conozca el concepto superficie de almacenamiento necesario?), el modelo gráfico especifica la distribución conjunta de las variables involucradas; por eso, para contestar cualquier pregunta de inferencia, se puede sumar sobre todas las posibles combinaciones de eventos no relevantes para la pregunta. De esta forma, un mismo nodo puede ser tanto fuente de información como objeto de predicción. Dichas inferencias se realizan aplicando algoritmos de propagación de probabilidades, que se han desarrollado específicamente para tal fin, lo que permite las red bayesina, una vez conocida la estructura de la red, o lo que es lo mismo cada nodo representa una variable de la ZAL y cada arco una dependencia probabilística entre variables de la ZAL, en la cual se especifica la probabilidad condicional de cada variable de la ZAL dados sus padres. A partir de la red construida, y al realizar la inferencia, la inferencia probabilística consiste en: "dadas ciertas variables conocidas (evidencia), calcular la probabilidad posterior de las demás variables (desconocidas)" lo que permite alcanzar el objetivo buscado de que dadas ciertas variables "conocidas" a la hora de planificar una ZAL (como puede ser la superficie disponible, el modelo de gestión esperado, etc) se puede calcular la probabilidad posterior del resto de variables "no conocidas" (como la zona destinada a área logística o el tráfico esperado asociado a la ZAL), es un esquema general de la propagación de la incertidumbre.

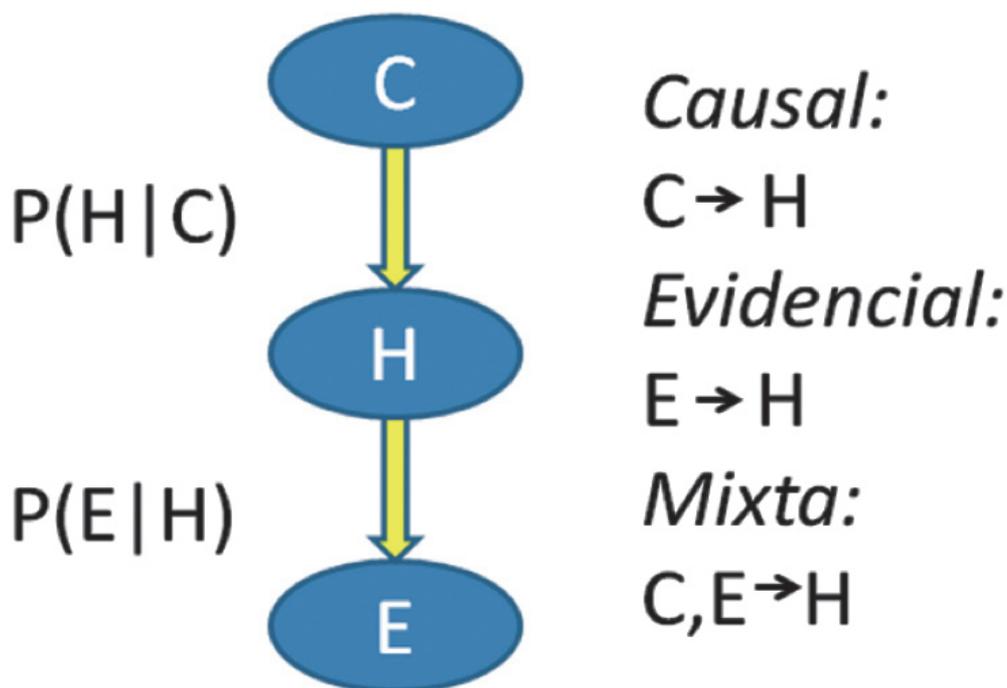


Figura 1. Inferencia Bayesiana
 (Elaboración propia)

Para utilizar una Red Bayesiana se deben identificar las variables y las relaciones de influencia causal entre ellas, y cuantificar estas relaciones de influencia causal asignando las probabilidades condicionadas y a priori. Una vez definida la red, se puede utilizar alguno de los paquetes de software comerciales o de libre disposición como (HUGIN EXPERT A/S 2012), (Norsys 2012), (Sotolongo et al. 2011), (Chávez et al. 1999) para realizar las inferencias que interesen. En el modelo que se propone se utiliza el software (Elvira Consortium 2002).

Las Redes Bayesianas son una representación compacta de una distribución de probabilidad multivariante. Formalmente, una Red Bayesiana es un grafo dirigido acíclico, donde cada nodo representa una variable aleatoria y las dependencias entre las variables quedan codificadas en la propia estructura del grafo según el criterio de d-separación. Asociada a cada nodo de la red hay una distribución de probabilidad condicionada a los padres de ese nodo, de manera que la distribución conjunta factoriza como el producto de las distribuciones condicionadas asociadas a los nodos de la red. Es decir, para una red con n variables x_1, x_2, \dots, x_n , como se aprecia en la ecuación:

$$\prod_{i=1}^n p(x_i | pa(x_i))$$

Las redes probabilísticas automatizan el proceso de modelización probabilística, utilizando la expresividad de los grafos. Los modelos resultantes combinan resultados de la teoría de grafos (para representar las relaciones de dependencia e independencia del conjunto de variables) y de la probabilidad (para cuantificar estas relaciones). Esta unión permite realizar de forma eficiente, tanto el aprendizaje automático del modelo, como la inferencia a partir de la evidencia disponible. La base de conocimiento de estos sistemas es una estimación de la

función de probabilidad conjunta de todas las variables del modelo, mientras que el módulo de razonamiento es donde se hace el cálculo de probabilidades condicionadas. El estudio de esta técnica proporciona una buena perspectiva global del problema del aprendizaje estadístico y la minería de datos, y permite entender mejor otras técnicas alternativas (Hastie et al. 2009), (Duda et al.).

Las redes bayesianas están siendo empleadas dentro de los sistemas de transporte, preferentemente, para desarrollar actuaciones en carreteras, como (Sun et al. 2006) que emplea las redes bayesianas para predecir flujos de vehículos, así como (Tebaldi et al. 1998). Dentro de la planificación se puede encontrar (Bromley et al. 2005), donde se desarrolla una gestión integrada de recursos hídricos o (Cain 2001) mejoras en la planificación, en la gestión de los recursos naturales.

La escasez de modelos de planificación bayesianos estriba en que aplicar un modelo de Redes Bayesianas exige mucho más esfuerzo que aplicar otros modelos de razonamiento aproximado, o desarrollar una heurística para la definición y actualización del modelo. Este esfuerzo adicional viene provocado principalmente por dos causas:

a) Especificación de la red (estructura y parámetros). Especificar una Red Bayesiana exige estudiar cuidadosamente cuáles son las variables que intervienen en el sistema y las relaciones de influencia causal entre ellas. Además, una vez definida la red se deben estimar las probabilidades condicionadas asociadas, que normalmente son un número bastante grande de parámetros difíciles de estimar.

b) Dificultad de implementar los algoritmos de propagación de probabilidades, que además de ser más o menos complejos, son computacionalmente muy costosos.

MATERIAL Y MÉTODOS

Para la consecución del objetivo de caracterizar los parámetros físicos de las zonas de actividades logísticas mediante redes bayesianas, de manera que una vez construida la y al realizar inferencia probabilística dadas ciertas variables conocidas de la ZAL a planificar (evidencia), calcular la probabilidad posterior de las demás variables (desconocidas) de dicha ZAL. Para la consecución del objetivo se ha desarrollado la siguiente metodología, que se divide en dos tareas: una, para determinar el escenario de trabajo y, la otra, para desarrollar el modelo de inteligencia artificial.

Determinación del escenario de trabajo

Diagnóstico y estado del arte

Consiste en la revisión del estado del arte para identificar el conjunto de variables de medida de la explotación en ZALs, mediante el empleo de buscadores especializados y gestores de aplicaciones. Se desarrolla en dos etapas:

a) Determinación de las variables de la ZAL y selección de las mismas: se realiza un estudio de todas las variables de la ZAL susceptibles de investigación.

b) Obtención del valor de las variables para las distintas ZALs de estudio: una vez seleccionadas las ZALs de estudio que se quieren obtener se utilizan diferentes fuentes de información para obtener, los valores de las mismas.

Las variables (rasgos) que intervienen en el cálculo de las probabilidades, tanto a priori como las condicionales son las variables de planificación de la ZAL. El número de variables (rasgos) conforma una tabla de 2^n combinaciones, donde n es un número natural que puede ser considerablemente grande; lo cual constituye una dificultad, cuya solución pudiera consistir en reducir el espacio de representación inicial, de forma tal que si existen variables superfluas, se analice si se mantienen o no, según su importancia desde el punto de vista metodológico.

Una alternativa de solución al problema de la selección de variables es el uso del conjunto de testores típicos (Ruiz-Shulcloper et al. 1990). Los testores típicos son el número de variables (atributos o rasgos), con los cuales se deben describir los objetos (cuestionarios) que inciden de manera determinante en un problema. En el modelo propuesto se aplica el algoritmo LEX (Alganza et al. 2003) para obtener los testores típicos.

La información proporcionada por una o más variables que se observan (evidencia), se propaga por la red y actualiza nuestra creencia acerca de las variables no observadas; por ello es muy importante la selección de las variables que darán información para la planificación. Tras el desarrollo del escenario de trabajo se concluye que para la planificación de ZALs, las variables empleadas para la investigación son las que se recogen en la Tabla 1.

Tabla 1. Variables de estudio

NOMBRE DE LA VARIABLE	DESCRIPCIÓN
Ubicación de la ZAL	Las ZAL pueden identificarse según su grado de integración con el área del Puerto
Importancia estratégica de la ZAL	La ZAL se puede definir según su rol o grado de centralidad logística respecto a su "hinterland": es decir, su rango como nodo logístico y la amplitud de su ámbito de influencia
Intermodalidad ferroviaria	La ZAL puede identificarse según la ordenación de la intermodalidad ferroviaria, es decir, en función de la relación general funcional de los Centros Logísticos de la ZAL y la intermodalidad ferroviaria
Grado de especialización logística	La ZAL se caracteriza según su grado de especialización en la logística exclusivamente portuaria o en actividades de logística mixta
Nivel de la ZAL	La ZAL se caracteriza según su nivel de multifuncionalidad (demanda no sectorizada) o especialización sectorial en mercancías concretas
Ordenación interna de la ZAL	La ZAL se caracteriza en función del consumo de espacio y al edificabilidad media bruta
Área logística de la ZAL	En el caso de las parcelas, su ordenación puede responder a distintos criterios de promoción de la ZAL, que dependerán de la orientación y el potencial logístico detectado. Según la ordenación de las parcelas se clasifica la ZAL
Área intermodal de la ZAL	El área intermodal es un ámbito delimitado dentro de la ZAL en el que se realizan todas las actividades necesarias para transferir carga del ferrocarril a la carretera y viceversa, si se produce este intercambio modal
Área de servicios de la ZAL	El área de servicios es el ámbito o ámbitos dentro de la ZAL en el que se prestan servicios de todo tipo a los vehículos y equipos, a las personas y a las empresas. Es el ámbito más representativo de la ZAL, y el nivel de calidad de la ZAL se expresa, en buena medida, a través de los estándares de dimensionamiento y calidad de esta área de servicios
Tráfico total del puerto ligado a la ZAL	Corresponde al tráfico total anual movido en el puerto asociado a la ZAL
Tráfico de contenedores del puerto asociado a la ZAL	La variable comprende el tráfico de contenedores anuales de medido en TEUs del puerto asociado a la ZAL
Tráfico de mercancía general del puerto asociado a la ZAL	Corresponde al tráfico en toneladas de mercancía general anual del puerto asociado a la ZAL
Tráfico de contenedores anuales de entrada/salida	Corresponde al tráfico de contenedores anuales de entrada/salida (TEUs) del puerto asociado a la ZAL

Construcción del modelo de inteligencia artificial

Obtener una red bayesiana a partir de datos, es un proceso de aprendizaje que se divide en dos etapas: el aprendizaje estructural y el aprendizaje paramétrico (Pearl 1988). La primera de ellas consiste en obtener la estructura de la red bayesiana, es decir, las relaciones de dependencia e independencia entre las variables involucradas. La segunda etapa, tiene como finalidad obtener las probabilidades a priori y condicionales requeridas, a partir de una estructura dada. A continuación se describe la discretización de las variables, la construcción del modelo y la inferencia y clasificación.

Discretización de variables

Una vez seleccionadas las variables de estudio, es necesario, para el proceso de construcción de los modelos, la discretización de las variables. Normalmente las redes bayesianas consideran variables discretas o nominales, por lo que si no lo son hay que discretizarlas antes de construir el modelo. Aunque existen modelos de redes bayesianas con variables continuas, éstos están limitados a variables gaussianas y relaciones lineales. Los métodos de discretización se dividen en dos tipos principales: (i) no supervisados y (ii) supervisados. Así se estudian diferentes tipos de discretizaciones y será una opción del software desarrollado.

La discretización consiste en dividir el rango de las variables continuas en un número finito de intervalos exhaustivos y exclusivos. En el caso de las redes Bayesianas, las herramientas generalmente necesitan los atributos discretizados. Para discretizar los atributos es necesario seleccionar el número de intervalos. Los intervalos se dividen teniendo en cuenta la frecuencia, es decir, el número de instancias en cada intervalo o utilizando el mismo rango, es decir, la misma distancia. Naturalmente, siempre se pierde 'información' al discretizar. Una vez seleccionadas las variables de estudio en las tareas anteriores, es necesario, para el proceso de construcción de los modelos, la discretización de las variables. Para este estudio se ha considerado emplear variables discretas, para lo cual se han discretizado las variables continuas. Dicha discretización se realiza atendiendo a criterio experto para la selección de estratos.

Los atributos continuos se transforman en intervalos que pueden ser usados como etiquetas discretas; por ejemplo, la variable tamaño de la ZAL se ha dividido en un número de intervalos etiquetados como: menos de 100 hectáreas, entre 100 y 200 hectáreas, y más de 200 hectáreas. El experto del dominio decidirá el número de intervalos, rechazar o corregir instancias con outliers, etc.

Una vez que hemos preprocesado los datos, empieza el proceso de aprendizaje (el proceso de minería de datos). En el caso de las redes Bayesianas, consiste en aprender la estructura de la red y generar las tablas de probabilidades para cada nodo. En la mayoría de los casos, en este paso también serán necesarios expertos del dominio para definir total o parcialmente la estructura de la red, decidiendo la dirección de los arcos (relaciones causa-efecto), nodos raíz/ hoja y/o parámetros como umbrales, algoritmos, etc. Una vez obtenida la red, es posible que sea necesario editarla para añadir, borrar o invertir los arcos.

Construcción de los modelos

En esta parte del proceso, el aprendizaje estructural consiste en encontrar las relaciones de dependencia entre las variables, de forma que se pueda determinar la topología o estructura de la red bayesiana. De acuerdo al tipo de estructura, se aplican diferentes métodos de aprendizaje estructural: aprendizaje de árboles, aprendizaje de poli-árboles, aprendizaje de redes multiconectadas, métodos basados en medidas y búsqueda, métodos basados en relaciones de dependencia.

Para la construcción del modelo, en esta investigación se ha empleado el algoritmo K2. El algoritmo K2 está basado en la optimización de una medida, que es lo que se pretende en la planificación, optimizar los ratios de explotación. Esa medida se usa para explorar, mediante un algoritmo de ascensión de colines, el espacio de búsqueda formado por todas las redes que contienen las variables de la base de datos. Se parte de un red inicial y ésta se va modificando (añadiendo arcos, borrándolos o cambiándolos de dirección), obteniendo una nueva red con mejor medida. En concreto, la medida K2 (Cooper et al. 1992) para una red G y una base de datos D es la siguiente:

$$f(G:D) = \log P(G) + \sum_{i=1}^n \left[\sum_{k=1}^{s_i} \left[\frac{\log(\Gamma(\eta_{ik}))}{\Gamma(\eta N_{ik} + \eta_{ik})} + \sum_{j=1}^{r_i} \frac{\log \left(\frac{\Gamma(\eta N_{ijk} + \eta_{ijk})}{\Gamma(\eta_{ijk})} \right)}{\Gamma(\eta_{ijk})} \right] \right]$$

donde N_{ijk} es la frecuencia de las configuraciones encontradas en la base de datos D de las variables x_i , donde n es el número de variables, tomando su j-ésimo valor y sus padres en G tomando su k-ésima configuración, donde s_i es el número de configuraciones posibles del conjunto de padres y r_i es el número de valores que puede tomar la variable x_i .

Las redes Bayesianas llevan tiempo incorporándose a tareas de clasificación supervisada, pero no al caso de planificación en lo relacionado con los puertos. En base a ideas expuestas por (Acid et al. 1995), (Friedman et al. 1996), y ampliadas en (Sierra et al. 1998), se pueden utilizar las factorizaciones de probabilidad representadas por las redes Bayesianas para realizar clasificaciones, considerando para ello la existencia de una variable especial, la variable a clasificar, que viene a ser predicha por un grupo de variables, el resto, y de forma que la estructura de la red obtenida puede ser utilizada para la predicción del valor de la clase de esta variable a clasificar, mediante la asignación de valores a las predictoras, y la posterior propagación de la evidencia introducida en la red; esto es, mediante el cálculo de la probabilidad a posteriori del nodo asociado a la variable especial, dados los valores del resto.

En la construcción de la red se ha utilizado el software Elvira específico para trabajar con redes bayesianas (Elvira Consortium, 2002).

El programa Elvira cuenta con un formato propio para la codificación de los modelos, un lector-intérprete para los modelos codificados, una interfaz gráfica para la construcción de redes, con opciones específicas para modelos canónicos, algoritmos exactos y aproximados (estocásticos) de razonamiento, tanto para variables discretas como continuas, métodos de explicación del razonamiento, algoritmos de toma de decisiones, aprendizaje de modelos a partir de bases de datos, fusión de redes, etc.

Inferencia y clasificación

Una vez construidos los modelos, en esta fase se estudia la capacidad de inferencia. Una red bayesiana proporciona un sistema de inferencia, donde una vez encontradas nuevas evidencias sobre el estado de ciertos nodos, se modifican sus tablas de probabilidad; y, a su vez, las nuevas probabilidades son propagadas al resto de los nodos. La propagación de probabilidades, o inferencia probabilística, es la probabilidad de algunas variables de ser calculadas, dadas las evidencias en otras variables. Las probabilidades, antes de introducir evidencias, se conocen como probabilidades a priori; una vez introducidas las evidencias, las nuevas evidencias propagadas se llaman probabilidades a posteriori.

En esta fase se aprovechan las características que poseen los métodos bayesianos en tareas de aprendizaje. Cada ejemplo observado va a modificar la probabilidad de que la hipótesis formulada sea correcta (aumentándola o disminuyéndola). Es decir, una hipótesis que no concuerda con un conjunto de ejemplos más o menos grande, no es desechada por completo sino que lo que harán será disminuir esa probabilidad estimada para la hipótesis. Estos métodos son robustos al posible ruido presente en los ejemplos de entrenamiento y a la posibilidad de tener entre esos ejemplos de entrenamiento datos incompletos o posiblemente erróneos. Los métodos bayesianos nos permitirán tener en cuenta en la predicción de la hipótesis el conocimiento a prior o conocimiento del dominio en forma de probabilidades.

RESULTADOS

Con la aplicación de la Red Bayesiana mediante el algoritmo K2, la red construida es la que aparece en la figura 2.

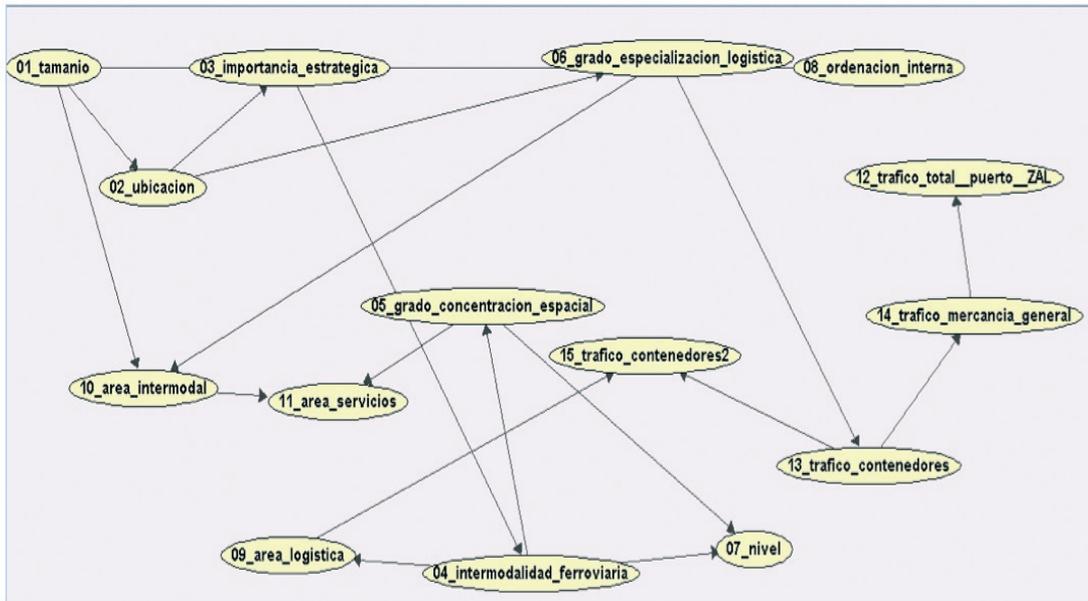


Figura 2. Red Bayesiana, algoritmo K2 para la estimación de parámetros para Zona de Actividades Logísticas

La probabilidad a priori es la que aparece en la fFigura 3.

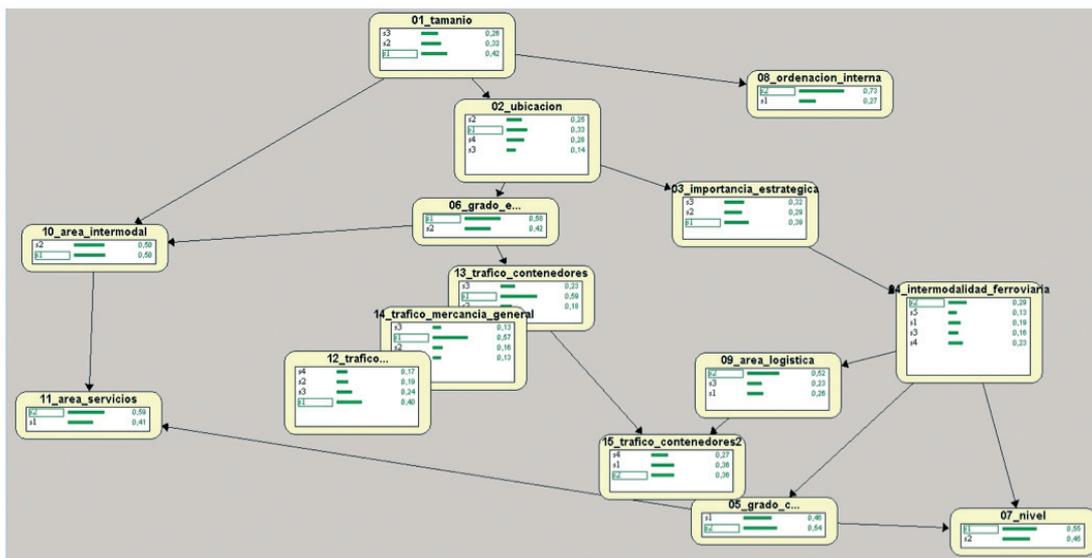


Figura 3. Probabilidad a priori de la Red Bayesiana, algoritmo K2 para la estimación de parámetros para Zona de Actividades Logísticas

A partir de la red construida mediante inferencia, se pueden obtener las probabilidades que pueden adoptar las discretizaciones de las diferentes variables. A partir de la red construida, y al realizar la inferencia (Figura 4), la inferencia probabilística permite: dadas ciertas variables conocidas (evidencia): que la superficie de la ZAL esté en el estrato de menos de 100 hectáreas, que la ubicación de la ZAL es ubicación intraportuaria dentro de la zona de servicio y que el área intermodal esté en desarrollo (lo que corresponde a un área intermodal ferroviaria poco madura, con ausencia de conexión con la red general del puerto), calcular la probabilidad posterior de las demás variables (desconocidas) como la importancia estratégica de la ZAL, su ordenación interna, el área de servicios. El ejemplo se representa en la Figura 4.

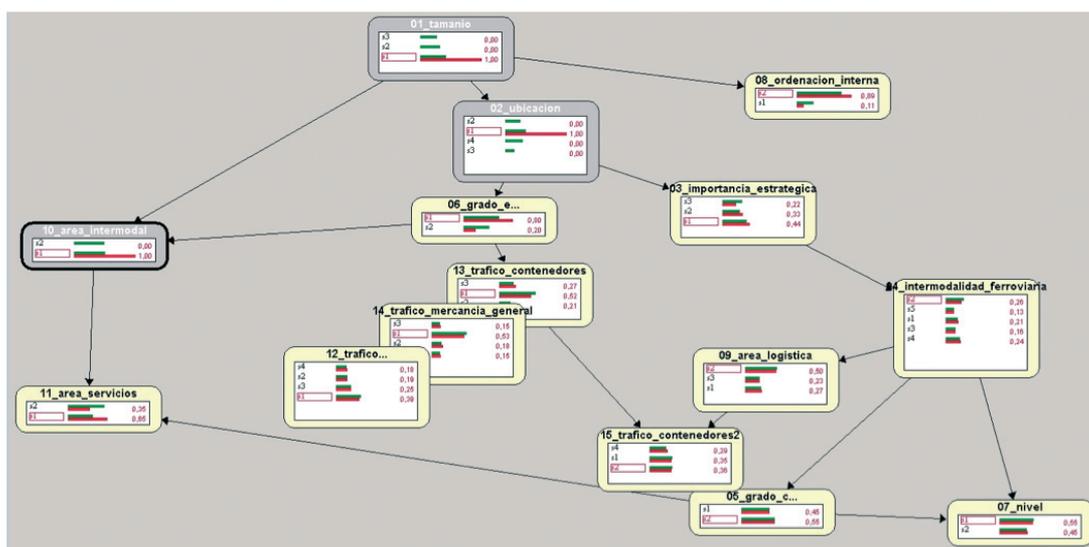


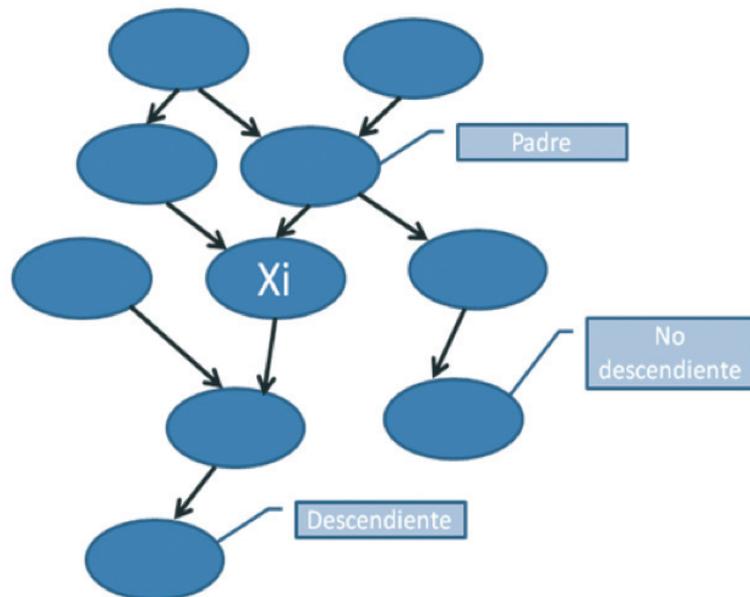
Figura 4. Ejemplo de inferencia de la Red Bayesiana, algoritmo K2 para la estimación de parámetros para zona de actividades logísticas

DISCUSIÓN

Si una vez construida la red (Figura 2) y obtenida la probabilidad a priori (Figura 3), respecto al tráfico asociado a la ZAL se puede concluir, a partir del modelo de Red Bayesiana desarrollado empleando el algoritmo k2 y cuya discretización corresponde a mover: menos de 10 millones de toneladas anuales, entre 10 y 20 millones de toneladas anuales, entre 20 y 40 millones de toneladas anuales y más de 40 millones de toneladas anuales, de la tabla de probabilidad de cada nodo, (las herramientas no dan un único valor, sino una probabilidad por cada uno de los estados más probable), la ZAL que presenta mayor probabilidad es aquella cuyo tráfico total anual movido en el puerto asociado a la misma es menor de 10 millones de toneladas anuales, y corresponde al tráfico de mercancía general anual de dicho puerto asociado a la ZAL menos de 5 millones de toneladas. Respecto al tráfico de contenedores, la ZAL tendrá asociado un tráfico de menos de 500.000 TEUs anuales. De éste, el tráfico de entrada/salida es menor de 50.000 TEUs anuales.

Para interpretar las relaciones entre variables, se ha de conocer que, los arcos en una red bayesiana proporcionan una forma de codificar relaciones de independencia. Estas relaciones se pueden especificar como (Figura 5):

- Dada una red bayesiana con nodos X_1, X_2, \dots, X_n . Si $\text{Padres}(X_i)$ son los padres de X_i y $\text{NoDescendientes}(X_i)$ los nodos que no son descendientes de X_i .
- Entonces, para cada variable X_i se tiene que X_i es independiente de sus No Descendientes dados sus Padres. Esto se expresa como $\text{Ind}(X_i; \text{NoDescendientes}(X_i) | \text{Pa}(X_i))$



De la red construida (Figura 2), las relaciones de dependencia resultantes destacan que, en este caso, el tráfico de mercancía general depende del tráfico de contenedores, y el tráfico total depende del tráfico de mercancía general. En términos causales se diría que el tráfico de contenedores es causa del tráfico de mercancía general, y que el tráfico de mercancía general

es causa del tráfico total del puerto asociado a la ZAL. En este caso, dada la dependencia entre las variables, cuando se conoce información sobre el tráfico de contenedores se puede modificar la certeza sobre el estado del tráfico total del puerto; y, a la inversa, cuando se sabe algo sobre el estado del tráfico total asociado al puerto, la creencia sobre el estado del tráfico de contenedores se altera. Sin embargo, si el estado del tráfico de mercancía general es conocido, conocer alguna información sobre el tráfico de contenedores o el tráfico total del puerto asociado a la ZAL no modificará la creencia sobre el estado de estas variables. Se puede decir que la propagación de la información se bloquea, y se dice que el tráfico de contenedores y el tráfico total se tornan condicionalmente independientes dado el tráfico de mercancía general.

Entre las variables ubicación, importancia estratégica e intermodalidad ferroviaria, se produce una conexión serial (Figura 2). Cuando se conoce información sobre la ubicación de la ZAL, se puede modificar la certeza sobre el estado de la intermodalidad ferroviaria; y, a la inversa, cuando se sabe algo sobre el estado de la intermodalidad ferroviaria, la creencia sobre el estado de la ubicación se altera. Sin embargo, si el estado de la variable de la importancia estratégica es conocido, conocer alguna información sobre la ubicación o intermodalidad ferroviaria asociada a la ZAL no modificará la creencia sobre el estado de estas variables, la propagación de la información se bloquea y la ubicación y la intermodalidad ferroviaria se tornan condicionalmente independientes dado la importancia estratégica.

La variable tamaño de la ZAL va a ser la variable decisora principal del proceso de planificación. Aparece en la red como una variable: nodo de la que salen los arcos, por lo que se genera una conexión divergente, se tiene un nodo padre que proyecta sus arcos hacia varios hijos. O, lo que es lo mismo, las flechas que salen de él y divergen hacia sus hijos.

Cuando no se conoce el estado de la variable padre, existe una dependencia entre las variables, análoga a la que se da en la conexión de serial (la anterior). Sin embargo, cuando el estado de la variable padre se conoce, las variables hijo se tornan independientes y la información no se propaga si se añaden evidencias sobre los nodos hijo. En el estudio, las variables ubicación, ordenación interna y área intermodal de la ZAL, son independientes dado el tamaño de la ZAL.

Si se estudian las relaciones entre las variables grado de concentración de la ZAL, área de servicios de la ZAL y área intermodal (Figura 2), se produce una conexión convergente, dado que varias variables apuntan con sus arcos hacia una variable de convergencia (área de servicios de la ZAL), esto es, convergen en un mismo nodo hijo. En este tipo de conexiones, las variables padre son independientes entre sí, pero una variable de convergencia puede influir sobre sus nodos padre. La propiedad importante de este tipo de conexión a la hora de propagar la información es que, cuando se tiene evidencia sobre la variable de convergencia, los nodos padre se vuelven dependientes y la evidencia del estado de uno de ellos se propaga por los demás; así, el grado de concentración y el área intermodal de la ZAL son dependientes, dada la variable zona de servicio.

Respecto a la concentración espacial, nivel de la ZAL y área logística, se produce una conexión divergente, se tiene un nodo padre: intermodalidad ferroviaria, que proyecta sus arcos hacia varios hijos, por lo que las variables anteriores son independientes, conocida la intermodalidad ferroviaria.

Las Redes Bayesianas son herramientas estadísticas, que permiten modelar óptimamente situaciones causales gracias a sus bases teóricas que permiten representar el aspecto cualitativo y cuantitativo de la realidad. Si el objetivo de la Inteligencia Artificial (IA) es construir modelos artificiales fidedignos de fenómenos naturales, es conveniente adaptar este tipo de

herramientas para que sean consistentes con el funcionamiento que se observa en situaciones naturales, como en el caso del estudio.

AGRADECIMIENTOS

La investigación que ha dado lugar al artículo se enmarca dentro del Proyecto: Diseño de una red de transporte marítimo de corta distancia para productos perecederos con centro en el sureste de España y destino en la costa mediterránea y atlántica (REDYMAR) Ref.: P10/08, Programa Nacional de Cooperación Público-Privada. Subprograma de Transporte e Infraestructura de Ministerio de Fomento y del Ministerio de Ciencia e Innovación de España.

Nuestro agradecimiento a la Universidad Politécnica de Madrid por el apoyo prestado para la realización de dicho proyecto, así como a la Fundación Agustín de Betancourt.

BIBLIOGRAFÍA

ACID, S., and DE CAMPOS, L. Approximations of Causal Networks by Polytrees: An Empirical Study. *Advances in Intelligent Computing—IPMU'94*, 1995, pp. 149-158.

ALGANZA, Y. S., and PORRATA, A. P. LEX: Un Nuevo Algoritmo Para El Cálculo De Los Testores Típicos. *REVISTA CIENCIAS MATEMATICAS*, Santiago De Cuba, CUBA, 2003, vol. 21, no. 1, pp. 2-3.

BISHOP, C. M. *Pattern Recognition and Machine Learning*. Springer New York, SpringerLink (Online service), 2006.

BONILLA, M., et al. An Efficiency Analysis with Tolerances of the Spanish Ports System. *International Journal of Transport Economics*, 2004, vol. 32, no. 3, pp. 380-400.

BROMLEY, J., et al. The use of Hugin® to Develop Bayesian Networks as an Aid to Integrated Water Resource Planning* 1. *Environmental Modelling & Software*, 2005, vol. 20, no. 2, pp. 231-242.

CAIN, J. *Planning Improvements in Natural Resource Management. Guidelines for using Bayesian Networks to Support the Planning and Management of Development Programmes in the Water Sector and Beyond*, 2001.

CAMARERO, A., and GONZÁLEZ, N. *Cadenas Integradas De Transporte*. Madrid, España: Fundación Agustín de Betancourt. Ministerio de Fomento, 2005. ISBN 84-609-8349-8.

CASTILLO, E., GUTIÉRREZ, J. M. and HADI, A. S. *Expert Systems and Probabilistic Network Models*. Springer Verlag, 1997.

CHÁVEZ, MC., GRAU, R., and GARCÍA, M. *Un Método Para Construir Redes Bayesianas*. *Revista De Ingeniería De La Universidad De Antioquia*, 1999.

CIOS, K. J., and KACPRZYK, J. *Medical Data Mining and Knowledge Discovery*. Physica-Verlag, 2001.

CIOS, K. J.; PEDRYCZ, W. and SWINIARSKI, R. W. *Data Mining: A Knowledge Discovery Approach*. Springer Verlag, 2007.

COOPER, G. F.; and HERSKOVITS, E. A Bayesian Method for the Induction of Probabilistic Networks from Data. *Machine Learning*, 1992, vol. 9, no. 4, pp. 309-347.

Development. Secretariat. *Review of Maritime Transport*. United Nations Pubns, 2011.

DUDA, R. O.; HART, P. E. and STORK, D. G. *Pattern Classification*. 2000. NY Wiley.

Elvira Consortium. *Elvira: An Environment for Probabilistic Graphical Models*. Proceedings of the First European Workshop on Probabilistic Graphical Models. J. Gaamez and A. Salmerson ed. 2002.

FRIEDMAN, N.; and GOLDSZMIDT, M. *Building Classifiers using Bayesian Networks*. , 1996.

HASTIE, T.; TIBSHIRANI, R. and FRIEDMAN, J. H. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Verlag, 2009.

HUGIN EXPERT A/S. HUGIN EXPERT A/S. Gasværksvej 5 • DK-9000 Aalborg • DENMARK. 2012. Available from: <www.hugin.com>.

JARA-DÍAZ, S., et al. *Marginal Costs and Scale Economies in Spanish Ports*. 25th European Transport Forum. , 1997.

LIU, C. I.; JULA, H. and IOANNOU, P. A. Design, Simulation, and Evaluation of Automated Container Terminals. *Intelligent Transportation Systems*, IEEE Transactions on, 2002, vol. 3, no. 1, pp. 12-26.

LIZAZO TORRES, D.; DELFOR MEYER, R. and TORRES CÁRDENAS, V. *Minería De Datos En La Encuesta Permanente De Hogares 2009*, Universidad Nacional Del Litoral, Argentina. *Revista Ingeniería Industrial*, 2011, no. 1, pp. 19-28.

MUÑOZ, A.; and ALARCON, H. *Diseño De Una Herramienta Para La Evaluación De La Calidad De Servicio De Operadores Logísticos*. *Revista Ingeniería Industrial*, 2005, vol. 4, no. 1, pp. 13-27.

Norsys. *Netica*. , 2012. Available from: <www.norsys.com/netica.html>.

PEARL, J. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

RODRIGUE, J. P.; COMTOIS, C. and SLACK, B. *The Geography of Transport Systems*. Taylor & Francis, 2009.

Rodríguez-Dapena A. *Prospectiva Económica De Interés Portuario*. . Puertos del Estado ed., , 2009. ISBN 978-84-88975-70-6.

ROLL, Y.; and HAYUTH, Y. *Port Performance Comparison Applying Data Envelopment Analysis (DEA)*. *Maritime Policy & Management*, 1993, vol. 20, no. 2, pp. 153-161.

RUIZ-SHULCLOPER, J.; and LAZO, M. *Modelos Matemáticos Para El Reconocimiento De Patrones*. Editorial UCLV, Santa Clara, Cuba, 1990.

SIERRA, B.; and LARRANAGA, P. *Predicting Survival in Malignant Skin Melanoma using Bayesian Networks Automatically Induced by Genetic Algorithms*. an Empirical Comparison

between Different Approaches. *Artificial Intelligence in Medicine*, 1998, vol. 14, no. 1-2, pp. 215-230.

SOTOLONGO, D. M., and VALDIVIA, Z. Z. G. RED BAYESIANA CON ASPECTO DE MAPA CONCEPTUAL: UNA PERSPECTIVA PARA LOS SISTEMAS DE ENSEÑANZA/APRENDIZAJE INTELIGENTES, 2011.

SUN, S., ZHANG, C., and YU, G. A Bayesian Network Approach to Traffic Flow Forecasting. *Intelligent Transportation Systems*, IEEE Transactions on, 2006, vol. 7, no. 1, pp. 124-132.

TEBALDI, C., and WEST, M. Bayesian Inference on Network Traffic using Link Count Data. *Journal of the American Statistical Association*, 1998, vol. 93, no. 442, pp. 557-573.

TODD, B. S., STAMPER, R., and MACPHERSON, P. The Design and Construction of a Medical Simulation Model. *Computer Methods and Programs in Biomedicine*, 1994, vol. 42, no. 2, pp. 77-91.

TONGZON, J.; and HENG, W. Port Privatization, Efficiency and Competitiveness: Some Empirical Evidence from Container Ports (Terminals). *Transportation Research Part A: Policy and Practice*, 2005, vol. 39, no. 5, pp. 405-424.

TOVAR, B.; JARA-DIAZ, S. R. and TRUJILLO, L. A Multioutput Cost Function for Port Terminals: Some Guidelines for Regulation. *World*, 2003.

TOVAR, B.; TRUJILLO, L. and JARA-DÍAZ, S. Organization and Regulation of the Port Industry: Europe and Spain. *Essays on Microeconomics and Industrial Organization*. 2nd Edn. Springer-Verlag, Heidelberg, 2004, pp. 189-207.

VITYAEV, E.; and KOVALERCHUK, B. Empirical Theories Discovery Based on the Measurement Theory. *Mind and Machine*, 2004, vol. 14, no. 4, pp. 551-573.

WANHILL, S. R. C. A Study in Port Planning: The Example of Mina Zayed. *Maritime Policy & Management*, 1974, vol. 2, no. 1, pp. 48-55.

WITTEN, I. H.; and FRANK, E. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann Pub, 2005.

WONG, R. C. W., et al. (α , k)-Anonymity: An Enhanced k-Anonymity Model for Privacy Preserving Data Publishing. *ACM*, 2006.

Biografía Autores

Francisco Soler Flores

Licenciado en Matemáticas por la Universidad de Almería.

Doctorando del Departamento de Ingeniería Civil y Transportes en la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos de la Universidad Politécnica de Madrid. Pertenece al Grupo de Investigación reconocido de la Universidad Politécnica de Madrid Grupo de Investigación Logística y Explotación del Transporte y Gestión Operativa. Let & Go. Especialista en matemáticas aplicadas a la Ingeniería Civil.

Nicoletta González Cancelas

Doctor Ingeniero de Caminos, Canales y Puertos por la Universidad Politécnica de Madrid.

Profesor Ayudante, Doctor del Departamento de Ingeniería Civil y Transportes en la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos de la Universidad Politécnica de Madrid.

Es profesora del Máster Universitario en Sistemas de Ingeniería Civil (MUSIC) de la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos de la Universidad Politécnica de Madrid y del Master en Gestión y Planificación Portuaria e Intermodalidad; Master del Campus Portuario de Puertos del Estado. Es especialista en Logística y Transporte.

Pertenece al Grupo de Investigación reconocido de la Universidad Politécnica de Madrid, Grupo de Investigación Logística y Explotación del Transporte y Gestión Operativa. Let & Go, es responsable de la líneas de investigación en Logística y Data Mining aplicado al transporte.

Alberto Camarero-Orive

Doctor Ingeniero de Caminos, Canales y Puertos por la Universidad Politécnica de Madrid.

Universidad Politécnica de Madrid. Profesor titular del Departamento de Ingeniería Civil y Transportes en la Unidad Docente de Explotación de Puertos. Además, es Profesor del Programa de Doctorado "Ingeniería e Infraestructura de los transportes". Doctor Ingeniero de Caminos, Canales y Puertos (E.T.S. de Ingenieros de Caminos, Canales y Puertos, Universidad Politécnica de Madrid), Licenciado en Economía y en Administración, y Dirección de Empresas (Facultad de Ciencias Sociales y Jurídicas de la Universidad Carlos III de Madrid).

M. Carmen Palomino Monzón

Licenciada CC Físicas. Profesora Contratada de la Universidad Politécnica de Madrid.

Doctora en Ciencias Físicas por la Universidad Complutense de Madrid. Diplomada en Humanidades por la Universidad Autónoma de Madrid. Profesora Titular de Universidad de Puertos y Costas e Ingeniería Portuaria, en la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos, de la Universidad Politécnica de Madrid. Dirige el Curso de Postgrado de Energías Renovables, y es especialista en energías renovables marinas. Ha participado como conferenciante o chairperson en varios congresos y seminarios nacionales e internacionales. Es autora de varias publicaciones, libros y artículos en materia de Medio Ambiente. Pertenece al Grupo de Investigación reconocido de la Universidad Politécnica de Madrid, Grupo de Investigación en Ingeniería Marítima y Portuaria.

José Luis Almazán Gárate

Doctor Ingeniero de Caminos, Canales y Puertos por la Universidad Politécnica de Madrid. Licenciado en Ciencias Económicas y Empresariales. Euro ingeniero.

Funcionario por oposición del Cuerpo de Ingenieros de Caminos, del antiguo Ministerio de Obras Públicas, habiendo estado destinado en la Dirección General de Puertos y Costas y el Centro de Estudios de Ordenación del Territorio, Urbanismo y Medio Ambiente. Ha sido Director de Cooperación Técnica y Secretario General de la Empresa Pública, SECEG, S.A., encargada de realizar los estudios de factibilidad del enlace fijo Europa - África a través del Estrecho de Gibraltar. Ha sido Vicepresidente del Consorcio Europeo para el Programa Internacional de

Perforación de los Océanos (Ocean Drilling Program). Ha sido Director Facultativo de las obras del tramo submarino del Gasoducto del Magreb, bajo las aguas del Estrecho de Gibraltar. Es Profesor Titular de Universidad de Puertos y Costas e Ingeniería Portuaria, en la Escuela Técnica Superior de Ingenieros de Caminos, Canales y Puertos de la Universidad Politécnica de Madrid. Dirige el Grupo de Investigación reconocido de la Universidad Politécnica de Madrid, Grupo de Investigación en Ingeniería Marítima y Portuaria.