V Simposio Nacional de Predicción. Editorial INM (2001).

MODELIZACIÓN Y PREDICCIÓN CON REDES PROBABILÍSTICAS

José Manuel Gutiérrez Llorente (1)
Rafael Cano Trueba (2)

(1) Dept. Matemática Aplicada, Universidad de Cantabria
(2) CMT CAS, INM

RESUMEN

En primer lugar se introducen las Redes Probabilísticas y más concretamente las Redes Bayesianas en el área de la Predicción Probabilística y el diseño de Modelos Conceptuales. A continuación, se da una breve descripción de los pasos para construir una Red Bayesiana, y dos ejemplos con la versión Bayesiana de métodos como el de los *Análogos de Lorenz* y el de los *Filtros de Kalman*. Finalmente se propone un modelo de Red Bayesiana que, combinando los métodos anteriores y eliminando algunas restricciones, garantiza la compatibilidad meteorológica y la coherencia espacial de las predicciones. Se verá a lo largo del trabajo cómo este tipo de sistemas asimila de forma natural salidas de *EPSs*. Además admite actualización *on line* con observaciones puntuales procedentes de la *vigilancia meteorológica*, entre otras ventajas.

1. Introducción

Tanto para diseñar un Modelo Conceptual, como para realizar un Pronóstico Meteorológico, todos los datos y observaciones disponibles son susceptibles de ser utilizados. La estadística elemental sugiere que a más datos, más precisión. Sin embargo demasiada información se hace redundante y puede ser inmanejable. Habitualmente se selecciona la información utilizando el concepto de correlación (Billet, 1997), ignorando a menudo que la falta de correlación no implica independencia ya que existen variables dependientes cuya correlación es insignificante. Las Redes Probabilísticas son herramientas más apropiadas para resolver problemas en este entorno.

1.1 Notación empleada

M: predictando meteorológico (Tmax, Tmin, Pcp, Vx, Ins, Tr, Gr, Nb, Nv, Es,...).
$\mathbf{m}^{\mathbf{l}}_{\mathbf{t}}$**:** valor concreto de predictando meteorológico en la localidad **l** en el instante **t**.
${}_{\mathbf{1}}\mathbf{m}^{\mathbf{l}}_{\mathbf{t}}$ **:** serie del predictando meteorológico hasta el instante **t**.
EA: conjunto de clases de escenarios atmosféricos.
$\mathbf{ea}_{\mathbf{t}}$**:** valor concreto de clase de escenarios atmosféricos (en el instante t).

2. Redes Probabilísticas

Están formadas por un conjunto de variables y su PDF conjunta. Se representan por grafos donde cada variable es un nodo y cada relación de dependencia es un segmento inter-nodal. Cuando el grafo es dirigido y acíclico las Redes son Bayesianas (Castillo y otros, 1997). Por ejemplo, el grafo de la Fig. 1 muestra una red Bayesiana formada por 6 nodos y las relaciones de dependencia que se establecen entre ellos. Los nodos circulares representan meteoros en una localidad dada en un cierto instante, y el nodo rectangular representa la salida del modelo numérico HIRLAM, que define el estado de la atmósfera en el mismo instante. El grafo resultante define las relaciones de dependencia que se establecen entre las variables. De esta forma, una Red Probabilística (RP) es una herramienta estadística capaz de obtener un modelo de dependencia para todas y cada una de las variables disponibles (mediante un grafo dirigido acíclico). Así mismo, una RP permite cuantificar estas relaciones definiendo una función de probabilidad conjunta de las mismas a través de un producto de funciones locales de probabilidad condicionada (la probabilidad de cada nodo condicionado a sus padres):

$$P(H,Tn,Tx,Pcp,Vx,Ins)=P(H)\ P(TN|H,Tx)\ P(Tx|H,Pcp)\ P(Pcp|H)\ P(Vx|Tx,Pcp)\ P(Ins|Pcp) \qquad [0]$$

La función de probabilidad resultante es compatible con las relaciones de dependencia e independencia codificadas en el grafo (ver Castillo y otros 1997 para más detalles). Por tanto, estos modelos proporcionan una alternativa más potente y general que los métodos estándar basados en correlación, y su utilidad meteorológica es muy versátil.

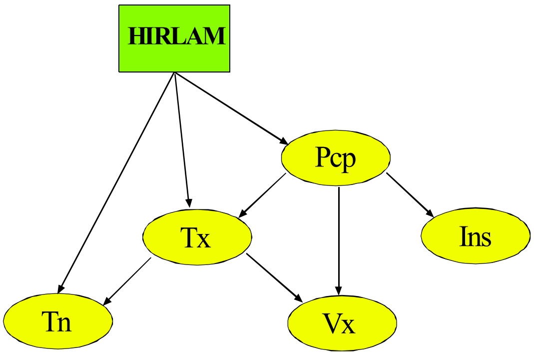

Figura 1: Ejemplo de Red Probabilística.

2.1 Especificación

El primer paso para construir una RP consiste en especificar las variables a incluir en el modelo, sus características: si son discretas o contínuas, numéricas o alfanuméricas, gaussianas o no gaussianas, predictores o predictandos, si la red será estática, dinámica, acoplada etc. Resumiendo, se decide qué series y en qué formato se incorporan al modelo. Para más información, entrar en:
http://www.cs.berkeley.edu/~murphyk/Bayes/usage_dbn.html#spec

2.2 Aprendizaje

Una vez especificados los datos que forman la **base de conocimiento** del modelo, hay que encontrar la estructura (el grafo) y los parámetros (las probabilidades condicionadas) que mejor se ajustan a la base de datos utilizada como muestra; tanto la estructura como los parámetros pueden estimarse a partir de los datos utilizando algoritmos eficientes.

2.2.1 Aprendizaje de la estructura

Para cada número ***n*** de variables, existe un número limitado de posibilidades de construir una Red Bayesiana (o lo que es lo mismo un Grafo Dirigido Acíclico (DAG)). En el ejemplo de la figura 1, tenemos un DAG con 6 variables. Por desgracia el número de DAGs posibles para cada ***n*** variables es combinatorio en ***n*.** Con ***n*** = 4, hay 543 DAG posibles, con ***n*** = 5, hay 29 281 DAG posibles, para ***n*** = 6, hay 3 781 503 DAG posibles, etc. Esto significa que para un número de nodos no demasiado grande, el aprendizaje estructural no puede ser exhaustivo. Una solución es predefinir el tipo de DAG por hipótesis, como se hace en los 2 primeros ejemplos que se mostrarán después. Pero si nuestro desconocimiento de las relaciones de dependencia del sistema nos obliga a realizar un aprendizaje estructural, hay varios métodos de búsqueda no exhaustiva, casi todos ellos basados en la maximización de una función de verosimilitud de la Red sobre la base de datos. Por ejemplo, uno de estos métodos aplicado sobre datos de precipitación de distintas localidades da como resultado el grafo mostrado en la figura 2. Una inspección visual de este grafo revela una dependencia espacial de las localidades consistente con la idea a priori que pueda tener un experto de los distintos modos de precipitación. Lo más interesante de estos modelos es que son obtenidos de forma automática a partir de los datos disponibles. Más detalles en:
http://www.cs.berkeley.edu/~murphyk/Bayes/usage.html#structure_learning

2.2.2 Aprendizaje de los parámetros

Un modelo consiste en la estructura gráfica (definida por el DAG) y los parámetros. Los parámetros vienen dados por las PDF condicionadas de cada nodo dados sus padres (los padres de un nodo son los nodos del grafo cuya flecha apunta al nodo), y son fáciles de calcular con una buena base de conocimiento. Más información en: http://www.cs.berkeley.edu/~murphyk/Bayes/usage.html#param_learning

2.3 Inferencia

Con un DAG y sus parámetros, la RP está creada y puede ser usada para realizar inferencia. Esto es, calcular las PDF marginales dadas ciertas evidencias. Normalmente, en PPSs, las evidencias las proporcionan las salidas de los modelos numéricos de predicción, estas evidencias pueden ser asimiladas por la RP en forma categórica o en forma de PDF (caso de EPSs). Una de las grandes ventajas que tienen las RPs es que admiten observaciones parciales, pueden incluso ser puntuales. Observaciones parciales recibidas en tiempo real pudeden ser inmediatamente asimiladas por la RP como evidencias. Esto permite la actualización de los pronósticos en tiempo real a medida que se van incorporando nuevas observaciones (**vigilancia meteorológica)**. Incluso el predictor puede perturbar cualquier dato a su criterio para mejorar su perspectiva.

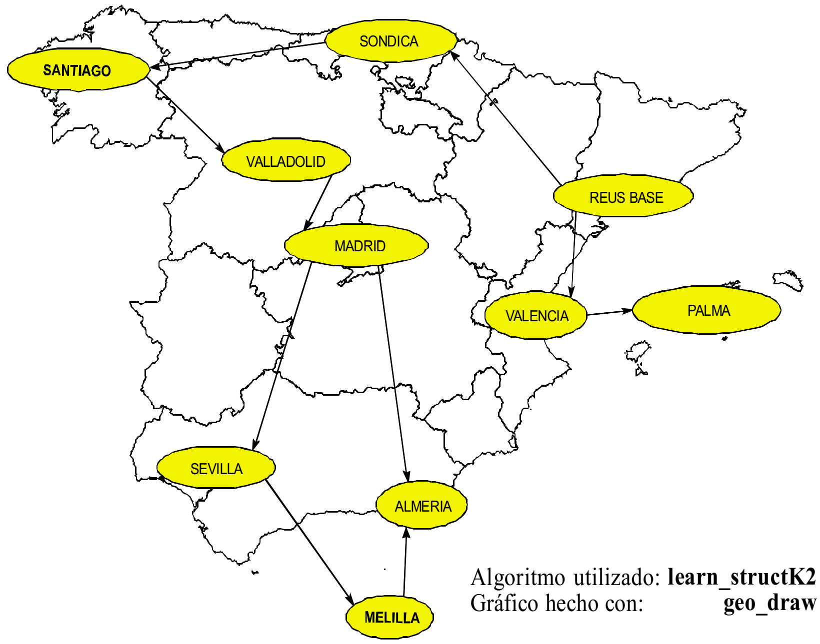

Figura 2. Ejemplo de aprendizaje estructural para 10 variables, que ilustra la dependencia espacial.

3 Predicción Probabilística

Dado que el objetivo principal es entrar en el área de la predicción probabilística conviene, antes de continuar, definir lo que se entiende por **Predicción Probabilística**: *Consiste en estimar la distribución de probabilidad (PDF) de los fenómenos meteorológicos una vez asimilada la información disponible.*

La climatología y la persistencia son Sistemas de Predicción Probabilística (PPSs) elementales; de hecho son empleados como referencia en los sistemas de validación de PPSs (Brier, 1950).

Desde el punto de vista de las Redes Probabilísticas, es sencillo expresar el problema general de la predicción probabilística: consiste en estimar las probabilidades $P(m^l_t)$ para cada posibilidad m de M, en cada localidad y para todos los predictandos, dada la información *on line* disponible, es decir:

$$P(m^1_t) \_ P(m^1_t \mid ea_t, {}_1m^1_{t-1}, m^j_t, z^1_t) \qquad [1]$$

Donde:

ea_t: valor de EA correspondiente al escenario previsto por el modelo numérico (con EPS, se toma la PDF) .
${}_1m^1_{t-1}$: valores previos de M (Auto-Regresión).
m^j_t : valores de M en otras localidades (Dependencia espacial).
z^1_t: valores de otros predictandos (Compatibilidad Física).

Si se plantea el problema en estos términos, de la ecuación [1] se vé que hay que calcular una probabilidad conjunta que involucra a todas las variables en diferentes puntos y en diferentes instantes de tiempo, lo cual es inabordable computacionalmente, por lo que conviene plantear algunas hipótesis restrictivas. A continuación se muestran 3 tipos de PPSs que aproximan el problema de manera eficiente.

3.1 Modelo de Análogos de Lorenz

Según Lorenz (Lorenz, 1969), toda la información sobre cualquier predictando meteorológico se puede obtener a partir de un perfecto conocimiento del Escenario Atmosférico. Esto significa que cada **M** sólo depende de **EA**, por lo que se ignoran la dependencia espacial, la secuencia temporal y los restantes predictandos. Con estas hipótesis la ecuación [1] se convierte en:

$$P(m^{1}_{t}) \_ P(m^{1}_{t} \mid ea_{t}) \qquad [2]$$

Ignorar la secuencia temporal implica que no se tiene en cuenta la evolución del sistema, de aquí que este método sea más adecuado para predictandos con muy poca autocorrelación (tipo precipitación). La independencia espacial añade el problema de no garantizar la coherencia de la prediccion entre las diferentes localidades, ya que aunque cada uno de los análogos es obviamente coherente, su combinación puede que no lo sea. Además, al ser de tipo climatológico suaviza mucho los extremos.

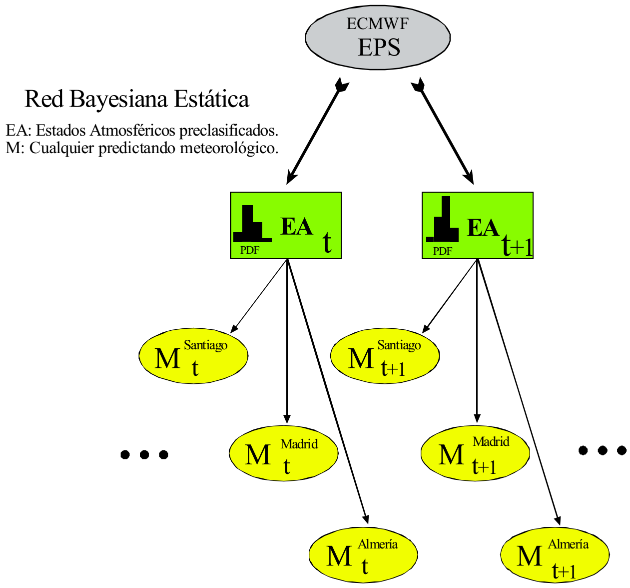

Figura 3: Modelo gráfico que expresa el PPS de Análogos de Lorenz.

3.2 Filtros de Kalman

Los filtros de Kalman son modelos autoregresivos, es decir, tienen en cuenta la evolución temporal del sistema, de tal manera que cada variable **Mg** (ver figura 4) no sólo depende de la salida **S** (ver figura 4) sino que también depende de su evolución. Con estas hipótesis la ecuación [1] queda:

$$P(mg^{1}_{t}) \_ P(mg^{1}_{t} \mid s_{t}, {}_{1}mg^{1}_{t-1}) \qquad [3]$$

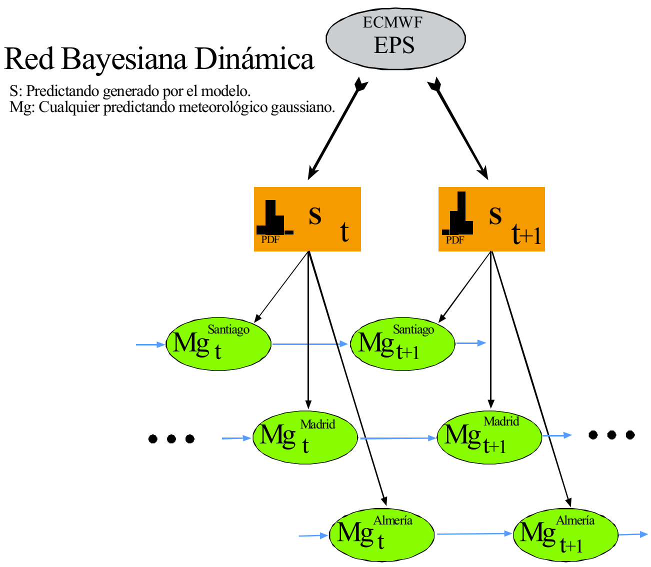

Figura 4: Modelo gráfico que expresa el PPS de Filtros de Kalman.

Por su estructura este método es óptimo para predictandos con mucha autocorrelación (como temperatura), aunque por esto mismo, tiene tendencia a la persistencia y no se adapta bien a los cámbios bruscos. Al igual que el método de análogos ignora la correlación espacial por lo que no garantiza la coherencia espacial de las predicciones. Detecta aceptablemente los eventos extremos, pero sólo admite variables de tipo gaussiano (no es el más adecuado para variables tipo precipitación). Más detalles en:
http://www.cs.berkeley.edu/~murphyk/Bayes/usage_dbn.html#lds

3.3 Modelo de Análogos Acoplado

Hasta hoy, la hipótesis de Lorenz es inaccesible, no se alcanza un perfecto conocimiento del Escenario Atmosférico. Por ello tiene sentido, y es necesario, añadir información procedente de otras vías. Lo más sencillo es recurrir a la información residente en las propias series climatológicas: por una parte está la dependencia espacial, es decir información mútua que pueden intercambiar parejas de localidades con características climatológicas parecidas (ver figura 2) y por otra parte está la dependencia temporal, es decir información de instantes anteriores o auto-regresión. El modelo de análogos acoplado que aquí se propone surge al tener en cuenta estas consideraciones, quedando la ecuación [1]:

$$P(m^{1}_{t}) \_ P(m^{1}_{t} \mid ea_{t}, m^{1}_{t-1}, m^{j}_{t}) \qquad [4]$$

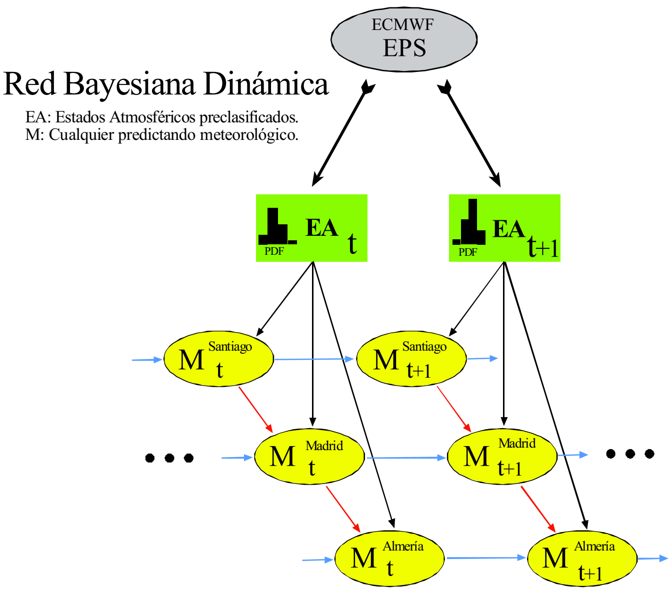

Figura 5: Modelo gráfico que expresa el PPS de Análogos Acoplado.

4 Conclusiones

Cuando se consigue mejorar la perspectiva en el planteamiento de un problema, todo son ventajas. Hemos visto cómo varios sistemas de predicción probabilística operativos en el INM, cuando son planteados como Redes Bayesianas, son más compactos, más sencillos y más potentes. Además admiten las salidas del EPS como una distribución de probabilidad (PDF) de Escenarios Atmosféricos (ver figuras 3, 4 y 5), mientras que en su forma original hay que ejecutar el PPS una vez para cada elemento del EPS y después refundir toda la información resultante. Otra importante ventaja del formato en Red Bayesiana es la facultad de actualizar en tiempo real las inferencias con datos aislados. También se ponen claramente de manifiesto las fuertes restricciones que implícitamente se hacen en cada caso. Con el formato en RB es más sencillo descubrir el origen de sus puntos débiles y aplicar soluciones. Precisamente siguiendo esta línea de razonamiento se ha concebido el Modelo de Análogos Acoplado (**MAA**). Y estas son sus principales ventajas:

· Probabilidad: no correlación.
· Dependencia: utiliza el subconjunto de predictores realmente relevante sin criterios de correlación.
· Coherencia: no genera pronósticos incompatibles.
· Causalidad: permite el estudio *off line* de fenómenos adversos y modelos conceptuales.

· Genérico: es apto para predecir cualquier predictando.
· Actualización: admite datos aislados, o conjeturas del predictor y actualiza la predicción.
· Escalable: puede ser operativo desde la vigilancia hasta la predicción climática.
· EPS: es el *input* ideal por su naturaleza probabilística.
· Compatibilidad: tiene el mismo formato de salida y de validación de los PPS operativos.

Agradecimientos

Los autores agradecemos a la Universidad de Cantabria, al Instituto Nacional de Meteorología y a la Comisión Interministerial de Ciencia y Tecnología (CICYT Proyecto REN2000-1572) su apoyo en este proyecto. Más información en: http://etsiso2.macc.unican.es/~meteo/

Tambien agradecemos a Kevin Patrick Murphy, autor de los excelentes programas de la Bayes Net Toolbox for Matlab. http://www.cs.berkeley.edu/~murphyk/

Referencias

Billet J., M. Delisi, B.G. Smith, and C. Gates (1997): Use of regression techniques to predict Hail size and the probability of large Hail. Weather and Forecasting, 12, 154-164.

Bellone E., Hughes J.P. and Guttorp P. (2000): A Hidden Markov model for downscaling synoptic atmospheric patterns to precipitation amounts. Climate Research, 15(1), 1-15.

Brier G.W. (1950):Verification of Forecasts expressed in terms of probability. Monthly Weather Review, 78, 1-3.

Castillo E., Gutiérrez J.M. y Hadi A.S. (1997): Expert Systems and Probabilistic Network Models. Springer-Verlag.

Cano R., Gutiérrez J.M. (2001): Dynamic Bayesian Networks for Rainfall Forecasting.

Gutiérrez J.M., Cano R., Rodríguez M.A., and Cofiño A.S. (1999c) : Redes Neuronales y Patrones de Analogías Aplicados al Downscaling en Modelos Climáticos, en Proceedings del II Congreso Nacional de Climatología, 234--241, Instituto Nacional de Meteorología, Madrid.

Hughes J.P.,Guttorp P. and Charles S.P. (1998): A Nonhomogeneus Hidden Markov Model for Precipitacion Occurence, Journal of Educational Psichology, 24.

Lorenz E. N. (1969): Atmospheric Predictability as Revealed by Naturally Occurring Analogues. Journal of the Atmospheric Sciences, 26, 636-646.

Stanski, Wilson, Burrows: Survey of Common Verification Methods in Meteorology. WWW No. 8. WMO/TD. No. 358; 1989.

Toth, Z. (1990): Estimation of Atmospheric Predictability by Circulation Analogs. Monthly Weather Review, Vol 119, Nº 1, 65-119, January 1991.