Investigación

Estimación de sucesos raros mediante Redes Bayesianas

Rare events estimation using Bayesian Networks

Francisco Soler Flores y José Ángel Olivas Varela

Revista de Investigación

Volumen IV, Número 1, pp. 127–136, ISSN 2174-0410
Recepción: 30 Sep'13; Aceptación: 20 Mar'14

1 de abril de 2014

Resumen

La teoría y aplicaciones de sucesos de baja probabilidad de ocurrencia o sucesos raros está en boga en los últimos años debido a su importancia práctica en muy diferentes campos, como seguros, finanzas, ingeniería o ciencias del medio ambiente. Este trabajo presenta una metodología para predecir sucesos raros basada en Redes Bayesianas que permite a su vez estudiar situaciones alternativas.

Palabras Clave: Sucesos raros, redes bayesianas, poisson, Naive-Poisson .

Abstract

Theory and applications of rare events, very important in recent years due to its practical importance in very different fields such as insurance, finance, engineering or environmental science. This paper presents a methodology for predicting rare events based on Bayesian networks that in turn allows to study alternative scenarios.

Keywords: Rare events, Bayesian networks, poisson, Naive-Poisson.

1. Introducción

El tratamiento de sucesos raros, sucesos que ocurren con una probabilidad baja, es un problema complejo y amplio cuyo tratamiento se enmarca en el ámbito de modelización de la incertidumbre, teoría de la decisión y donde el estudio del riesgo, definido en términos de teoría de decisiones como las pérdidas promedio o las pérdidas que se pronostican cuando algo malo sucede, es importante. La 'Ley de Eventos Raros', demostrada por Poisson, fundamenta matemáticamente el concepto de suceso raro. Esta ley que lleva su nombre, la ley de los sucesos raros ([1]) es también llamada por Bortkiewicz 'ley de los pequeños números' ([2]).

En las últimas décadas se han desarrollado numerosas técnicas de análisis y modelización de datos en distintas áreas de la estadística ([3],[4]) y la inteligencia artificial ([5]) que se han aplicado al estudio de sucesos raros . La Minería de Datos (MD) ([6]) 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. 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 diagnosis.

Con el avance y la extensión de los modelos de inteligencia artificial, tales como las redes bayesianas, se plantea, en este trabajo, la utilización de estos últimos para estimar la probabilidad de ocurrencia de un suceso raro.

El objetivo de este trabajo es presentar una metodología que, a partir de los datos brutos de los que se dispone habitualmente a la hora de estudiar cualquier problema en el marco del análisis de datos, sistematiza el estudio de sucesos raros y es capaz de estudiar diferentes situaciones que modifican la probabilidad de ocurrencia de estos.

El trabajo se divide en tres secciones principales:

- Minería de datos y redes bayesianas: Se describen los conceptos básicos en los que se apoya la metodología.
- Metodología propuesta: Se describe en detalle el procedimiento propuesto.
- Conclusiones.

2. Minería de datos y Redes Bayesianas

Se denomina Minería de Datos ([7],[8]) al conjunto de técnicas y herramientas aplicadas al proceso no trivial de extraer y presentar conocimiento implícito, previamente desconocido, potencialmente útil y humanamente comprensible, a partir de grandes conjuntos de datos, con objeto de predecir de forma automatizada tendencias y comportamientos; y describir de forma automatizada modelos previamente desconocidos ([9]). El término Minería de Datos Inteligente ([8]) es referido específicamente a la aplicación de métodos de aprendizaje automático ([10]), para descubrir y enumerar patrones presentes en los datos, para estos, se desarrollaron un gran número de métodos de análisis de datos basados en la estadística. En la medida en que se incrementaba la cantidad de información almacenada ([11]) en bases de datos, estos métodos empezaron a utilizarse para problemas de eficiencia y escalabilidad y es entonces cuando aparece el concepto de minería de datos. Una de las diferencias entre al análisis de datos tradicional y la minería de datos es que el primero supone que las hipótesis ya están construidas y validadas contra los datos, mientras que el segundo supone que los patrones e hipótesis son automáticamente extraídos de los datos . Las tareas de la minería de datos se pueden clasificar en dos categorías: minería de datos descriptiva y minería de datos predicativa ([12]). Diferentes trabajos como [5] y [13] aplican la minería de datos al tratamiento de sucesos raros.

2.1. Redes Bayesianas

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 y utilizando posteriormente mecanismos de inferencia probabilística para obtener conclusiones en base a la evidencia disponible. Las redes bayesianas han sido utilizados en el contexto de la estimación de sucesos sucesos raros en algunos trabajos([14] y [15]) sin llegar a señalar un método de estimación general.

Las redes bayesianas ([16], [17]) 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 ([18]). 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 factorizada como el producto de las distribuciones condicionadas asociadas a los nodos de la red. Es decir, para una red con n variables $X_1, X_2, \ldots, X_n$ (ecuacion 1).

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

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: no supervisados y supervisados ([19]).

El concepto de causalidad ([20]) en una red bayesiana se traduce en un caso particular de estas denominado red causal ([21]). Las redes bayesianas pueden tener una interpretación causal y aunque se utilizan a menudo para representar relaciones causales, el modelo no tiene por que representarlas de esta forma, Naive-Bayes (apartado 2.2) es un ejemplo de esto, sus relaciones no son causales.

Las redes probabilísticas automatizan el proceso de modelización probabilística ([22]) 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, a través del cálculo de parámetros ([23]) que para el caso de variables binarias se modela con una distribución Beta y para variables multivaluadas mediante su extensión, que es la distrubución Dirichlet (tabla 1),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.

Tabla 1. Estimación paramétrica

estimador	**expresión**
Máxima verosimilitud. Multinomial	$\theta_k^* = \frac{N_k}{N}$
Estimación bayesiana. Dirichlet	$\theta_k^* = \frac{N_k + a_k}{N + \sum_{i=1}^{r} a_i}$

2.2. Modelo Naive Bayes

Una de las formas más sencillas que se pueden idear al considerar la estructura de una red Bayesiana con objetivos clasificatorios, es el denominado *Naive-Bayes* o *Bayes ingenuo* ([24]).

Su denominación proviene de la hipótesis ingenua sobre la que se construye, que consiste en considerar que todas las variables predictoras son condicionalmente independientes dada la variable a clasificar (figura 1)

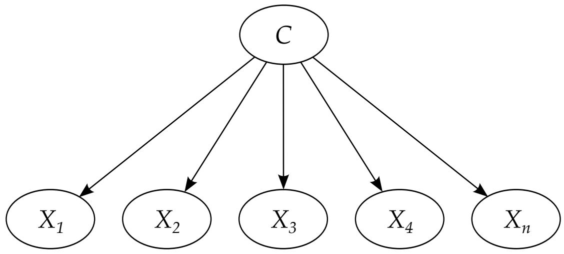


Figura 1. Naive Bayes

El modelo Naive Bayes es muy utilizados debido a que presentan, entre otras, ciertas ventajas ([6]):

- Generalmente, es sencillo de construir y de entender.
- El proceso de inducción es rápido
- Es muy robusto considerando atributos irrelevantes.
- Toma evidencia de muchos atributos para realizar la predicción final.

En el sentido de que su capacidad predictiva es competitiva con el resto de clasificadores existentes, el llamado *Naive-Bayes*, descrito, por ejemplo, por [24] y por [25] es uno de los más efectivos clasificadores. Este clasificador aprende de un conjunto de entrenamiento la probabilidad condicional de cada atributo X_i dada la clase C. La clasificación se hace entonces aplicando la regla de Bayes para calcular la probabilidad de C dadas las instancias de $X_1, X_2, \ldots, X_n$ tomando como clase predicha la de mayor probabilidad a posteriori. Estos cálculos se basan en una fuerte suposición de independencia: todos los atributos X_i son condicionalmente independientes dado el valor de la clase C.

La probabilidad de que el j-ésimo ejemplo pertenezca a la clase i-ésima de la variable C, puede aplicarse sin más que aplicar el teorema de Bayes, de la siguiente manera (ecuación 2)

$$P(C = \theta_i | X_1 = x_{1j}, \ldots, X_n = x_{nj}) \propto P(C = \theta_i) \cdot P(X_1 = x_{1j}, \ldots, X_n = x_{nj} | C = \theta_i) \qquad (2)$$

Dado que suponemos que las variables predictoras son condicionalmente independeientes dada la variable C, se obtiene que (ecuación 3)

$$P(C = \theta_i | X_1 = x_{1j}, \ldots, X_n = x_{nj}) \propto P(C = \theta_i) \cdot \prod_{r=1}^{n} P(X_r = x_{rj} | C = \theta_i) \qquad (3)$$

El modelo *Naive-Bayes* presenta un comportamiento muy dependiente del tipo de dominio ([26]) . Así por ejemplo, en dominios médicos, donde el conocimiento sobre el problema es elevado y por tanto tan sólo se recoge información relativa a variables que podríamos decir que se complementan, el Naive-Bayes, proporciona resultados aceptables, mientras que en dominios poco estructurados, en los que las variables del sistema se encuentran altamente correlacionadas, el comportamiento del Naive-Bayes puede ser más pobre.

El modelo Naive-Bayes combinado con la distribución de Poisson es utilizado para la clasificación de texto en el trabajo de [27] con buenos resultados. En este trabajo se propone el añadido de la minería de datos utilizando redes bayesianas y aplicando la distribución de probabilidad conocida para el estudio de sucesos raros. De esta forma y conocidos los valores de las variables que se utilizan como predictoras se pueden estudiar las diferentes situaciones posibles y observar cuándo es más probable la ocurrencia de un suceso raro.

3. Metodología propuesta

A continuación se detallará la metodología presentada en este trabajo. El objetivo de la misma es sistematizar el estudio de sucesos raros, proporcionar una herramienta para su estudio partiendo de los datos brutos, a priori, de cualquier problema.

Así las fases de este metodología son las siguientes:

- Preprocesamiento de los datos
  1. Selección de variables y obtención de sus valores.
  2. Discretización de las variables
- Construcción de la red bayesiana, estructura Naive-Bayes
- Aplicación modelo Naive-Poisson

3.1. Preprocesamiento de los datos

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: no supervisados y supervisados ([19]).

En este apartado de preparación de los datos, éstos son formateados de forma que las herramientas informáticas que se uilicen puedan manipularlos. A su vez, este apartado consiste también en la selección de variables y discretización de los datos en el caso de variables continuas. En este paso se identifica también la variable de la cual se quieren estimar los sucesos raros.

3.2. Construcción de la red bayesiana

A partir de la fase anterior, la construcción de la red consta de los siguientes puntos:

- Identificación de la Estructura: identificar las relaciones causales; analizar las variables en cuanto a dependencias e independencias
- Cálculo de Parámetros (probabilidades): cuantificar relaciones e interacciones

En este caso se propone un modelo concreto, el Naive-Bayes (apartado 2.2), adecuado para el proceso de clasificación y para el que la variable de la cual se quieren estimar los sucesos raros será la variable denominada padre(([18])).

Se va a optar para el desarrollo de la metodología por el modelo Naive Bayes. El modelo Naive Bayes es muy utilizados debido a que presentan, entre otras, ciertas ventajas:

- Generalmente, es sencillo de construir y de entender.

- Las inducciones de son extremadamente rápidas, requiriendo sólo un paso para hacerlo.
- Es muy robusto considerando atributos irrelevantes.
- Toma evidencia de muchos atributos para realizar la predicción final.

Una vez determinada la estructura de la red se calculan las tablas de probabilidades que permiten realizar la descomposición de la distribución de probabilidad y posteriormente realizar inferencia basándose en las evidencias obtenidas y la distribución de probabilidad asociada a la red.

Esta información que nos proporciona la red construida se completa aplicando el modelo Naive-Poisson descrito a continuación (apartado 3.3).

3.3. Modelo Naive-Poisson

El modelo o procedimiento por el que, a partir de la red bayesiana construida mediante el modelo Naive-Bayes, obtenemos la distribución de probabilidad asociada que permite estimar la probabilidad de ocurrencia de un suceso raro es denominado Naive-Poisson ([28]).

El procedimiento de asignación de la distribución de probabilidad consta de los siguientes apartados:

- Suposición Poisson
- Construcción de la distribución de probabilidad

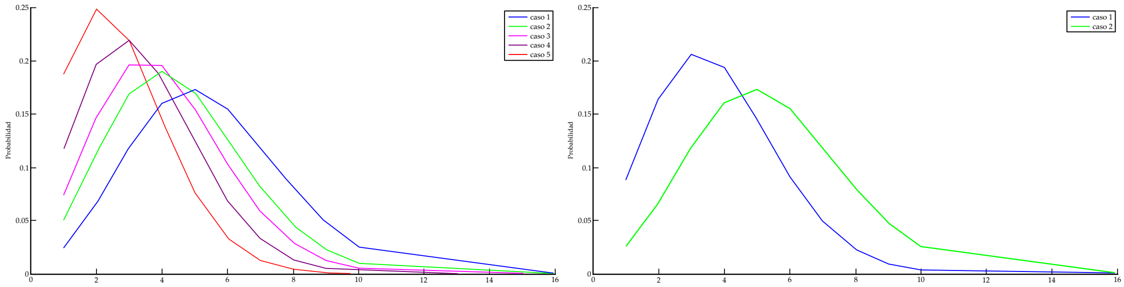
Es admitido en la comunidad científica que la distribución de la frecuencia de sucesos raros se ajusta a una distribución de Poisson ([1]). Así, se supondrá para el modelo a construir, la frecuencia de de sucesos raros sigue una distribución de Poisson. Una vez obtenidos los valores de la distribución real se ajustarán los datos por este tipo de distribución ([28],[29]). La distribución de probabilidad que proporciona la red bayesiana construida permite estimar los diferentes valores de probabilidad para cada uno de los valores de las variables que proporciona la discretización. A partir de los resultados obtenidos, la red se ajusta con una distribución de Poisson, que como es sabido viene determinada por su media (ecuacion 4).

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!} \tag{4}$$

Para el cálculo del parámetro que determina la distribución e Poisson, la media, tomaremos su estimador máximo verosímil que viene dado por la ecuación 5, donde x_i son los valores discretos de la accidentalidad y $p(x_i)$ los valores de probabilidad que proporciona la red construida mediante el algoritmo Naive-Bayes , siendo la variable discreta a clasificar C, la variable de la que queremos estudiar los valores que consideramos sucesos raros y el resto, las variables utilizadas para tal fin el resto de las variables.

$$\lambda = \sum_{i=1}^{n} x_i \cdot p(x_i) \tag{5}$$

Para cada uno de los valores de los estratos $a_1, a_2, \ldots, a_n$ que proporciona la discretización de la variable a estudiar los sucesos raros, se tomarán los valores reales de la frecuencia del suceso raro a estimar excepto para el a_n para el que se asigna un valor dado por la media de los valores x_i superiores al mayor estrato a_{n-1} (ecuación 6).



(a) Variable ejemplo 1 (b) Variable ejemplo 2

Figura 2. Ejemplos de estimación

$$a_n = \bar{x}_i \text{ con } x_i > a_{n-1} \tag{6}$$

De esta forma descrita obtenemos las distribución de poisson asociada a cualquiera de las situaciones ($j = 1 \ldots, m$) que estudiemos (ecuación 7) con cualquier conjunto de valores de las diferentes variables seleccionadas (en la figura 2 se observan dos ejemplos distintos), con lo cual es posible determinar en que situación es más alta la probabilidad de ocurrencia de los sucesos de baja probabilidad una vez detectado cuales son los valores de la variable estudiada que dada su distribución son de probabilidadmás baja (figura 3).

$$P_j(X = x) = e^{-\lambda_j} \frac{\lambda_j^x}{x!} \tag{7}$$

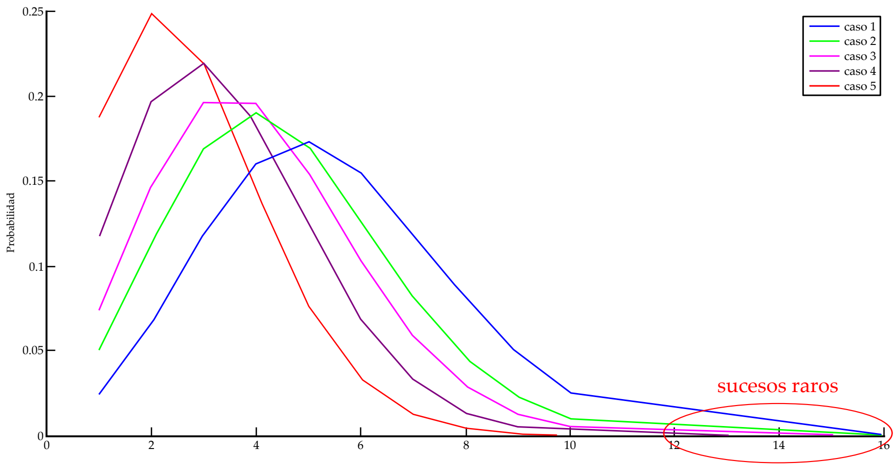


Figura 3. Sucesos raros y sus distribuciones según los valores de las variables predictoras

4. Conclusiones

En este trabajo se propone una metodología para que, de forma sencilla, a partir cualquier conjunto de datos y una vez seleccionada la variable objeto de estudio, se analicen los sucesos de baja probabilidad o sucesos raros, además de comparar diferentes situaciones para poder controlar así la probabilidad con la que se producen. La metodología presentada permite de manera sistemática el estudio de la ocurrencia de los sucesos objeto del estudio y plantear alternativas para controlar su ocurrencia.

Las líneas que abre este trabajo son entre otras la del desarrollo de software que permita su utilización de manera intuitiva, el estudio de las medidas de bondad de ajuste que completen su aplicación y el contraste con otros modelos de inteligencia artificial.

Referencias

[1] Simon Denis Poisson. *Recherches sur la probabilité des jugements en matière criminelle et en matière civile, précédées des règles générales du calcul des probabilités*. Bachelier, 1837.

[2] Ladislaus Bortkiewicz. Das gesetz der klienen zahlen (the law of small numbers.). *Leipzig, Germany: Teubner*, 1898.

[3] Micahael Tomz, Gary King, and Langche Zeng. Relogit: Rare events logistic regression. *Journal of statistical software*, 8(i02), 2003.

[4] F. Soler-Flores, J. M. Pardillo Mayora, and R. Jurado Piña. Tratamiento de outliers en los modelos de predicción de accidentes de tráfico. *VIII CIT, 02/07/2008-04/07/2008, La Coruña, España.*, 2008.

[5] Chris Seiffert, Taghi M. Khoshgoftaar, Jason Van Hulse, and Amri Napolitano. Mining data with rare events: a case study. In *Tools with Artificial Intelligence, 2007. ICTAI 2007. 19th IEEE International Conference on*, volume 2, pages 132–139. IEEE, 2007.

[6] D. E. Holmes, J. Tweedale, and L. C. Jain. Data mining techniques in clustering, association and classification. *Data Mining: Foundations and Intelligent Paradigms*, pages 1–6, 2012.

[7] D. J. Hand. Data mining: statistics and more? *The American Statistician*, 52(2):112–118, 1998.

[8] E. Castillo, J. M. Gutiérrez, and A. S. Hadi. *Expert systems and probabilistic network models*. Springer Verlag, 1997.

[9] J. Bromley, NA Jackson, OJ Clymer, AM Giacomello, and F. V. Jensen. The use of hugin to develop bayesian networks as an aid to integrated water resource planning. *Environmental Modelling and Software*, 20(2):231–242, 2005.

[10] K. J. Cios, W. Pedrycz, and R. W. Swiniarski. *Data mining: A knowledge discovery approach*. Springer Verlag, 2007.

[11] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth. From data mining to knowledge discovery in databases. *AI magazine*, 17(3):37, 1996.

[12] S. Acid and L. de Campos. Approximations of causal networks by polytrees: an empirical study. *Advances in Intelligent ComputingâĂŤIPMU'94*, pages 149–158, 1995. ID: 149.

[13] Gary M. Weiss. Mining with rarity: a unifying framework. *ACM SIGKDD Explorations Newsletter*, 6(1):7–19, 2004.

[14] Seong-Pyo Cheon, Sungshin Kim, So-Young Lee, and Chong-Bum Lee. Bayesian networks based rare event prediction with sensor data. *Knowledge-Based Systems*, 22(5):336–343, 2009.

[15] A. Ebrahimi and T. Daemi. Considering the rare events in construction of the bayesian network associated with power systems. In *Probabilistic Methods Applied to Power Systems (PMAPS), 2010 IEEE 11th International Conference on*, pages 659–663. IEEE, 2010.

[16] J. H. Kim and J. Pearl. A computational model for causal and diagnostic reasoning in inference systems. In *Proceedings of the 8th International Joint Conference on Artificial Intelligence*, pages 190–193. Citeseer, 1983.

[17] J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

[18] E. Castillo, J. M. Gutierrez, and A. S. Hadi. *Expert systems and probabilistic network models*. Springer Verlag, 1997.

[19] James Dougherty, Ron Kohavi, and Mehran Sahami. Supervised and unsupervised discretization of continuous features. In *ICML*, pages 194–202, 1995.

[20] Cristina Puente Agueda. Causality in sciencie. *Pensamiento Matemático*, (1):12, 2011.

[21] Judea Pearl. *Causality: models, reasoning and inference*, volume 29. Cambridge Univ Press, 2000.

[22] Laura Uusitalo. Advantages and challenges of bayesian networks in environmental modelling. *Ecological Modelling*, 203(3):312–318, 2007.

[23] Luis Enrique Sucar. Redes bayesianas.

[24] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern classification. 2000. *NY Wiley*. ID: 129.

[25] P. Langley, W. Iba, and K. Thompson. An analysis of bayesian classifiers. In *Proceedings of the National Conference on Artificial Intelligence*, pages 223–223. JOHN WILEY and SONS LTD, 1992.

[26] José Antonio Gámez and José Miguel Puerta. *Sistemas expertos probabilísticos*, volume 20. Univ de Castilla La Mancha, 1998.

[27] Sang-Bum Kim, Hee-Cheol Seo, and Hae-Chang Rim. Poisson naive bayes for text classification with feature weighting. In *Proceedings of the sixth international workshop on Information retrieval with Asian languages-Volume 11*, pages 33–40. Association for Computational Linguistics, 2003.

[28] F. Soler-Flores. Naive-poisson, a mathematical model for road accidents frequency estimation. *Conference of Informatics and Management Sciences*, pages 384–391, 2013.

[29] F. Soler-Flores. Expert system for road accidents frequency estimation based in naive-poisson. *Global Virtual Conference*, page 646, 2013.

Sobre el autor:

Nombre: Francisco Soler Flores
Correo electrónico: f.soler@upm.es
Institución: Universidad Politécnica de Madrid, España.

Nombre: José Ángel Olivas Varela
Correo electrónico: joseangel.olivas@uclm.es
Institución: Departamento de Tecnologías y Sistemas de Información, Escuela Superior de Informática, Universidad de Castilla La Mancha, España.