



La nueva era de la información

Las Universidad de Almería, la danesa de Aalborg y la noruega Norwegian University of Science and Technology participan en un proyecto europeo que romperá barreras en el ámbito de la gestión de datos. En el proyecto colaboran Cajamar, que diseñará una herramienta matemática y un software capaz de reducir el riesgo de las operaciones de crédito, y las multinacionales Daimler-Mercedes Benz, Hugin Expert y Verdande Technology. Por Alberto F. Cerdera

La morosidad bancaria subió hasta el 13,6 por ciento en 2013, según el Banco de España. Esta cifra, que supone un nuevo récord, se corresponde con los más de 197.000 millones de euros en créditos dudosos concedidos por las entidades bancarias de todo el país. Estas cifras reflejan una tendencia al alza de la morosidad, que se ha convertido en uno de los lastres más visibles de la economía de este país.

Viendo los datos se puede hablar de una mala gestión realizada por parte de las entidades de crédito, que han prestado más dinero de la cuenta, que no han realizado estudios de riesgo lo suficientemente exhaustivos como para valorar la idoneidad de la operación. Sin embargo también se puede hablar de unas herramientas de

valoración insuficientes, cuyas estimaciones no alcanzan la realidad de los clientes que se acercan a pedir un crédito.

Un grupo de investigación de la Universidad de Almería ha comenzado a trabajar en el desarrollo de una serie de algoritmos y aplicaciones informáticas que mejoran la gestión de altos volúmenes de datos, y con la que las entidades bancarias reducirán los riesgos de impago en las operaciones de crédito, gracias a que se conseguirán estudios de riesgo mucho más fiables y basados variables cien veces superiores a las que actualmente se emplean.

Esta nueva solución, a medio camino entre la estadística y la informática, se enmarca en el proyecto europeo AMIDST, financiado con fondos procedentes del Séptimo

Programa Marco, y que nace fruto de la colaboración de las universidades de Almería, de la danesa Aalborg University y de la noruega Norwegian University of Science and Technology. Dotado con un presupuesto de 2,762 millones de euros, de los que 323.000 se quedarán en el Campus de La Cañada, este proyecto servirá para el desarrollo de aplicaciones para la gestión de grandes volúmenes de datos, cuyos resultados se aplicarán en campos tan dispares como la banca, el coche autónomo y la búsqueda de yacimientos de petróleo.

AMIDST nace de la colaboración entre las instituciones académicas y empresas privadas, que participarán en el desarrollo de las diferentes herramientas y, al tiempo, serán unas de las principales beneficiarias de los resultados obtenidos con estos futuros entes de gestión de datos. Aunque no las únicas, ya que las soluciones desarrolladas se pondrán de forma gratuita al servicio de la comunidad para que puedan ser usadas por cualquier empresa que las necesite, y además estarán disponibles con código abierto, para que puedan ser implementadas por otros desarrolladores, interesados en mejorar su potencial.

Las tres universidades participan con grandes empresas como Cajamar, en el sector de la banca; Daimler, que es la encargada de fabricar vehículos Mercedes, y la empresa tecnológica Hugin Expert, que se dedica al ámbito de la automoción; y Verdande Technology, dedicada al desarrollo de soluciones para la toma de decisiones en el sector petrolero.

El proyecto comenzó el pasado mes de enero y se prolongará hasta finales de 2016, para cuando está previsto que los tres integrantes de esta iniciativa internacional tengan listas sendas herramientas para la gestión de los datos en el ámbito financiero, en el desarrollo del vehículo autónomo y en la detección de fallos en prospecciones petrolíferas.

Cientes sin secretos.

Antonio Salmerón es el investigador que lidera el grupo de la Universidad de Almería que, en colaboración con Cajamar, se centrará en el desarrollo de una herramienta para la gestión de un volumen elevado de datos, y cuya finalidad será la de determinar el riesgo en operaciones de crédito. Esta solución matemática hará posible incorporar un volumen muy superior de datos que los modelos disponibles actualmente, con lo que los resultados y análisis de riesgos serán más acertados.

Según Antonio Salmerón "hablamos de pasar de modelos que actualmente tienen en cuenta unas 20 ó 30 variables, a operar con unas 2.000". En la actualidad las herra-



Trabajo con supercomputadores

El desarrollo del sistema matemático y las diferentes pruebas para testar el funcionamiento de la aplicación informática para la gestión de datos requiere un elevadísimo coste computacional y los investigadores de la Universidad de Almería tendrán que realizar estos trabajos en el supercomputador de la Norwegian University of Science and Technology (en la imagen de arriba), cuya capacidad equivale a la de varios miles de ordenadores personales.



Antonio Fernández Álvarez, Antonio Salmerón y Rafael Rumí integran el equipo almeriense participante en el proyecto AMIDST.

mientas para la gestión de datos que emplea la banca tienen unos límites operativos para incorporar todo este nuevo volumen de información sobre los clientes que está disponible en las bases de datos de las entidades de crédito.

Sólo Cajamar podría ahorrar 7,5 millones de euros al año aplicando estos sistemas para la estimación del riesgo de las operaciones de crédito.

Cada vez que interactuamos con una entidad bancaria generamos información que puede ser útil para una estimación de riesgo. Nuestros pagos con tarjeta, las consultas realizadas a través de banca electrónica, desde qué lugar geográfico las realizamos, nuestros movimientos en la cuenta corriente... "toda esa información puede tener un valor a la hora de predecir la probabilidad de que un hipotético crédito puede acabar en un impago o no" y ahora no se emplea en su totalidad, debido a la falta de unos sistemas de gestión potentes.

El equipo almeriense que lidera Antonio Salmerón está integrado por el también matemático y docente de la Universidad de

Almería, Rafael Rumí, y por el doctorado en Informática, Antonio Fernández Álvarez. Ellos serán los encargados de desarrollar los algoritmos necesarios para la interpretación de toda la información, así como del desarrollo de una aplicación de gestión,

que será la que finalmente se ponga a disposición de Cajamar y del resto de la comunidad de usuarios, y que probablemente se ponga en uso comercial para la estimación del riesgo de las operaciones de crédito.

La clave se basa en estudios probabilísticos, en los que se tienen en cuenta una cantidad de variables cien veces mayor que las aplicaciones actuales. Y que, en el caso de Cajamar, puede suponer un ahorro de unos 7,5 millones de euros al año, según se explica en la página web del proyecto AMIDST. Los métodos con los que van a trabajar los investigadores almerienses tratan de descubrir "a ciegas" qué información puede haber en los datos que sea de utilidad para

la evaluación del riesgo.

Precisamente, la gran ventaja de la minería de datos es que pueden descubrir relaciones que pasan desapercibidas a un experto en la materia, explica Antonio Salmerón; lo que intentarán conseguir los investigadores almerienses es establecer relaciones entre las diferentes baterías de datos, gracias a los registros históricos que las entidades bancarias conservan de cada uno de sus usuarios. De esta manera, "vamos a poder decir qué características suelen tener las personas pagadoras y con cuáles se pueden identificar a posibles morosos".

La minería de datos, es decir, el establecimiento de relaciones que a simple vista pasan desapercibidas entre las diferentes variables, es la baza más significativa de este proyecto, y también la que lo hace único hasta el momento.

Las dificultades teóricas han impedido que, hasta el momento, una entidad como Cajamar ponga en uso toda la información que dispone de sus clientes, que es mucho mayor de lo que muchos podrían imaginar. Lo que Antonio Salmerón llama "manejabilidad de la información", y que será potenciada de forma espectacular gracias al trabajo de estos tres investigadores almerienses, que mantienen una relación estrecha con el Departamento de Metodología para la Medición del Riesgo de Cajamar, integrado, en gran parte, por antiguos alumnos de la carrera de Matemáticas de la Universidad de Almería.

La primera parte del trabajo en el que ya están inmersos los investigadores almerienses consiste en el desarrollo de los modelos desde el punto de vista teórico y algoritmos capaces de procesar esas cantidades de información, asegura Antonio Salmerón; y a partir de ese trabajo "vamos a desarrollar un modelo de predicción, de manera que conocida cierta información de un cliente determine la probabilidad de producirse un impago en un periodo de tiempo determinado", añade.

Evidentemente, esta nueva herramienta matemática necesita de una plataforma desde la que ponerse en uso y ahí entra en juego el informático del grupo, que será el encargado de construir un prototipo de software, capaz de captar los datos desde las fuentes de información de la propia entidad bancaria y realizar las predicciones y evaluaciones de riesgo sobre las operaciones de crédito.

Con el proyecto AMIDST, tanto los investigadores españoles como los noruegos y daneses van a dar solución a la gestión masiva de datos, un campo en el que todavía hay mucho por hacer. "Lo que vamos a desarrollar es una solución de carácter genérico, que

se va a instanciar a los problemas concretos que vamos a abordar en este proyecto”, aclara Antonio Salmerón. Los métodos actuales no llegan a cubrir las necesidades planteadas en los tres casos que se van a abordar en este proyecto europeo. Uno de los principales retos a los que los investigadores deben dar solución es al manejo y cruce de magnitudes cualitativas y cuantitativas simultáneamente. “Magnitudes numéricas y magnitudes cualitativas del tipo alto o bajo, hombre o mujer, profesión... manejar la interacción entre esas dos variables, y establecer la relación entre variables sociodemográficas, por ejemplo, y variables numéricas, como ingresos o cuantías de impago, requiere un alto

Otra de los grandes avances del trabajo de estos grupos de investigación es que van a utilizar “modelos gráficos probabilísticos” que, a diferencia de otros modelos conocidos como de caja negra, explican cómo han hecho las predicciones, cómo toman las decisiones. “Éste es un valor añadido importante porque no solo hace la predicción, sino que te permite aprender de ella”, añade Antonio Salmerón. El modelo ofrecerá información de cuáles son las relaciones encontradas. Por ejemplo, dirá que a partir de esta base de datos, la variable *a* se relaciona con la variable *b* de una manera determinada.

La gestión masiva de datos es un asunto en el que han avanzado mucho empresas

explícito” de los usuarios para el manejo de esos datos, no como ocurre mientras navegamos por Internet, que constantemente estamos enviando información de nuestros intereses, prácticamente sin darnos cuenta. “En este proyecto no hay cabida a vulneración de datos personales. La Comisión Europea así nos lo exige y se ha mostrado inflexible con este tema. Hay una garantía absoluta de que no se va a vulnerar la intimidad”, aclara Antonio Salmerón.

Este proyecto será una realidad en 2016, cuando se espera que estén listos los tres sistemas de gestión de datos previstos y serán las entidades privadas participantes las que tengan la última palabra para

Gestión de datos aplicada al desarrollo del vehículo

La danesa Aalborg University, en colaboración con la empresa tecnológica Hugin Expert y el gigante de la automoción Daimler, fabricante de vehículos Mercedes, trabajarán en el desarrollo de un sistema de gestión de datos aplicado al coche autónomo. Gracias a la incorporación de un número mayor de variables y de una gestión de la información más eficiente, los técnicos participantes en el proyecto conseguirán un sistema de toma de decisiones que aplicarán al vehículo. Gracias a él, el coche autónomo será capaz de reconocer maniobras en tráfico por autopista, de manera que se mejora la seguridad en la circulación de estos vehículos, que serán una realidad en no muchos años. Según explica la empresa Daimler, con AMIDST se optimizará la eficiencia de las técnicas de monitorización y análisis de un alto volumen de datos tomados en tiempo real, que son percibidos por los sistemas de control que incorporarán este tipo de vehículos. Gracias a AMIDST se conseguirá el objetivo de reducir en un 50 por ciento el número de accidentes de tráfico mortales en el año 2020.

El coche autónomo será capaz de reconocer maniobras en tráfico por autopista con el objetivo de reducir un 50% los accidentes mortales para 2020



Más eficiencia buscando petróleo

La toma de decisiones en la búsqueda de nuevos yacimientos de petróleo se verá mejorada una vez que esté en marcha los desarrollos previstos en el proyecto AMIDST. En este apartado trabajan la Norwegian University of Science and Technology, en colaboración con la empresa Verdande Technology, convertida en uno de los referentes mundiales en sistemas de gestión de datos y toma de decisiones en tiempo real, aplicados al sector petrolero.

El trabajo conjunto se traducirá en la creación de una serie de herramientas matemáticas y un sistema de gestión que permitirán reducir el margen de error en las prospecciones petrolíferas, gracias a la incorporación de un volumen de datos mucho mayor que el empleado en la actualidad. El número de mediciones en los registros de perforación en tiempo real es enorme. Por ejemplo, la ubicación precisa del agujero, la presión de los poros de las formaciones, su estabilidad. El conocimiento de estos y otros factores son importantes para optimizar el proceso de perforación de manera que es posible una perforación eficaz y el tiempo no productivo bajo, y con este sistema se mejorará.



costo computacional.

Hasta ahora, los modelos que hay para gestionar grandes volúmenes de datos, llamados bigdata, principalmente se centran en magnitudes cuantitativas, y si son numéricas lo que hace es discretizarlas, es decir, agruparlas por rangos de valor”, explica el investigador almeriense. La novedad es que somos capaces de manejar los datos en su “estado original, sin necesidad de discretizarlos ni procesarlos, con lo cual no perdemos ninguna información”, argumenta.

como Google o Amazon, que manejan cantidades ingentes de información sobre la vida de los usuarios, gracias al registro que realizan de los movimientos a través de Internet. Últimamente se ha puesto en cuestión el uso que se realiza de esta información y si, en cierto modo, este seguimiento atenta contra la privacidad de las personas. En el proyecto liderado por los investigadores almerienses no hay ningún conflicto con la intimidad, en la medida en que las entidades bancarias cuentan con “permiso

ponerlos en uso comercial.

Lo que sí es cierto es que suponen un hito en la forma en la que se gestionan los grandes volúmenes de datos. Un antes y un después en la manera en que se tomarán decisiones en las grandes empresas, y que ponen en uso una cantidad de datos que hasta ahora solo ocupan espacio en los discos duros gigantes de las entidades bancarias, las empresas petroleras y la industria de la automoción interesada en el desarrollo del coche autónomo. ■