

Big Data en seguros

Montserrat Guillen

Universitat de Barcelona, Riskcenter

El seguro es una disciplina basada en los métodos estadísticos que siempre ha empleado las fuentes de datos disponibles para pronosticar los resultados de pérdidas y su probabilidad de ocurrencia. Por este motivo, la estadística no solo ha tenido una posición prominente en la ciencia actuarial sino que forma parte de la esencia de la gestión de riesgos.

Aun así, el reciente crecimiento digital ha sido tan vigoroso que ha desafiado la manera de entender el marco teórico asegurador y, por lo tanto, ha hecho necesario integrar la ciencia de los datos en los nuevos modelos que estudian los seguros, nutriéndose de grandes sistemas de información. En este sentido, los seguros están siendo precursores de la transformación que sufre la econometría como instrumento de medición de la economía. En los seguros toma protagonismo la información individualizada y longitudinal, se prima la versatilidad del uso intensivo de datos frente a la simplicidad computacional de un modelo elegante y se da mucha importancia a las decisiones individuales complejas e interrelacionadas, sin perder de vista que debe garantizarse la solvencia de la entidad. El *big data* en seguros es un ejemplo paradigmático de la nueva economía digital.

La Figura 1 muestra la complejidad del sistema de información en una entidad aseguradora. En dichas bases de datos los clientes pueden tener pólizas de distintos tipos e interactuar a través de los canales de mediación como agentes o corredores, de las redes sociales y aplicaciones en la web y además recibir servicios y prestaciones. Vinculados a cada contrato se encuentran los siniestros

sufridos y su seguimiento, que deben tener dotado suficiente capital para garantizar la cobertura. Además de todo lo anterior, los contratos atienden a sus respectivas fechas de renovación, lo que imprime un carácter longitudinal a los datos, puesto que deben mantenerse registros de lo ocurrido hasta el momento e información sobre las características del riesgo, así como la fecha futura que determina el límite en la vigencia de cada una de las pólizas.

Les presentaré tres ejemplos de *big data*, en el ámbito asegurador en los que se demuestra la complejidad del análisis debido al volumen y variedad de la información.

El primer ejemplo es un caso de tarificación y comercialización de seguros, en el que existen tensiones entre lograr retener a los buenos clientes, los de menor riesgo, a la vez que establecer para ellos unos precios suficientemente competitivos en el mercado, pero necesariamente elevados para poder cubrir las cuantías de indemnización. La solución consiste en una maximización de beneficios sujeta a un control de riesgos. En las compañías aseguradoras, el departamento actuarial dispone de especialistas en fijación de precios (equipos de *pricing* o tarificación) que calculan las tarifas y argumentan que incrementar los precios incrementa los beneficios por póliza, así como la reserva de seguridad para la solvencia. En contraposición, el departamento de comercialización consigue su éxito mediante la retención de los buenos clientes, la cual se ve amenazada si los precios son excesivamente altos. Se ha valorado en estudios empíricos que atraer a un nuevo cliente cuesta tres veces más que conseguir retener a un cliente existente.

El debate entre los equipos de tarificación y de comercialización es en realidad una lucha por lograr un equilibrio óptimo entre precios suficientemente bajos para mantener una situación interesante en el mercado y necesariamente altos para

El reciente crecimiento digital ha sido tan vigoroso que ha hecho necesario integrar la ciencia de los datos en los nuevos modelos que estudian los seguros nutriéndose de grandes sistemas de información

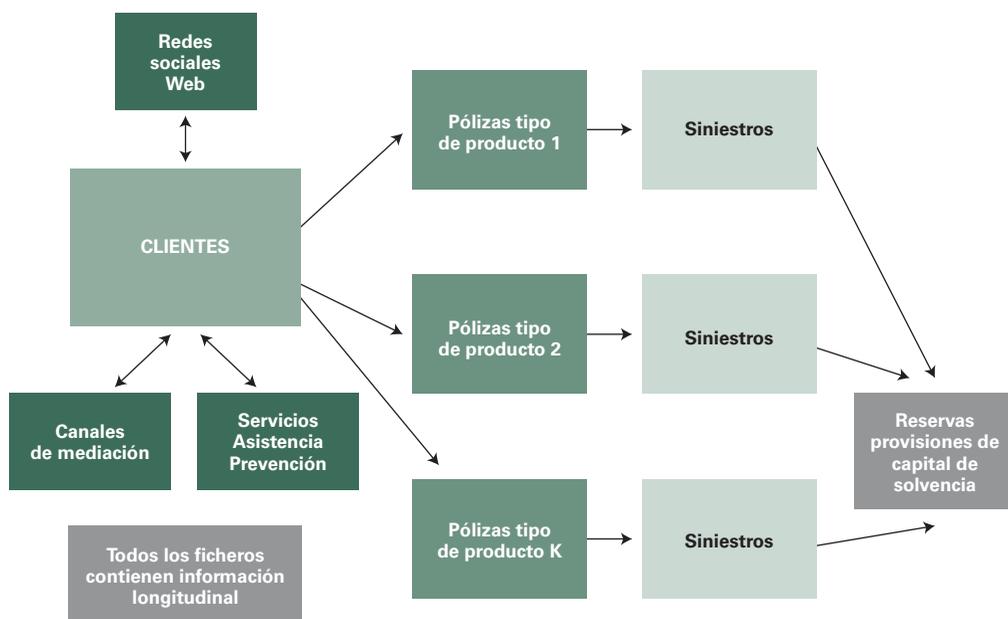
alcanzar los capitales mínimos exigidos por la regulación del sector. Pero existe un tercer elemento que incrementa la tensión. La presencia de un contagio entre los diferentes contratos dentro de una misma entidad aseguradora que a menudo no se tiene en cuenta como debería. Cuando un contrato no se renueva, es posible que haya existido algún hecho que haya influido en dicha decisión y que, como es lógico, también influya en el resto de contratos que posee el mismo tenedor (o sus allegados en sentido amplio) en la misma compañía. De esta forma, es posible que si se cancela una póliza de un asegurado, se produzcan cancelaciones encadenadas. Incluso se ha llegado a observar que este fenómeno viene alentado por competidores que pretenden realizar las denominadas ventas cruzadas y conseguir todos los productos contratados por un mismo asegurado una vez se han hecho con la primera de sus pólizas.

En línea con esta tradición, la aproximación estadística clásica ha trabajado durante mucho tiempo como si precios y ventas fueran problemas independientes. En contraposición con los métodos predictivos clásicos, las nuevas técnicas de modelización predictiva (ver Tabla 1) apuntan hacia una visión más integral, tendiendo un puente entre los tres objetivos: tarificar, comercializar y fidelizar, pero planteando un reto para el análisis de los datos y la modelización.

La segunda aplicación que les expongo tiene que ver con la generación de datos telemáticos en la conducción de vehículos mediante la instalación de un GPS que registra distancias recorridas y formas de conducir. Su utilización aporta elementos de prevención de accidentes, así como de identificación de segmentos de conductores con comportamientos potencialmente peligrosos, lo que obviamente redundará en las campañas de sensibilización y en la prima de seguros. Es, sin lugar a dudas, el paso previo y necesario para plantear la cobertura aseguradora de un vehículo que no requiera conductor.

En el seguro de automóvil basado en el uso, también conocido como el *pay-as-you-drive*, la prima se personaliza más allá de lo que venía siendo tradicional, y en definitiva los conductores ocasionales suelen pagar menos que los usuarios frecuentes. Sin embargo, se tienen en cuenta los hábitos generales de conducción al evaluar el riesgo y así, los excesos de velocidad, el tipo de carreteras por el que se circula habitualmente y el tiempo de día en que se conduce, pueden tenerse en cuenta en el nuevo sistema de tarificación. En realidad, dichos factores se demuestran que sirven para explicar la probabilidad de que el conductor se vea implicado en un accidente. Hoy en día, muchas compañías de seguros del mundo ya comercializan contratos basados en la telemetría, pero dada

Figura 1. Esquema de base de datos en una entidad aseguradora



Fuente: elaboración propia.

Tabla 1. Modelización estadística en seguros

Fundamentos	Modelos predictivos	Modelos Bayesianos, series temporales y longitudinales
• Linear models	• Longitudinal and panel data models	• Bayesian regression models
• Regression with categorical dependent variables	• Linear mixed models	• Generalized additive models and nonparametric regression
• Regression with count-dependent variables	• Credibility and regression modeling	• Non-linear mixed models
• Generalized linear models	• Fat-tailed regression models	• Claims triangles/loss reserves
• Frequency and severity models	• Spatial modeling	• Survival models
	• Unsupervised learning	• Transition modeling

Fuente: Fuente: recogidos en el libro Frees, E. W., Derrig, R. A., & Meyers, G. (Eds.). (2014). *Predictive Modeling Applications in Actuarial Science* (Vol. 1). Cambridge University Press.

la introducción tan reciente de este sistema, este es todavía escasamente conocido y su fundamento técnico no ha sido investigado como merece. Preocupa la privacidad de la información, aunque la futura vinculación de dicho sistema a los dispositivos móviles y las enormes ventajas en caso de tener que localizar el vehículo, sin duda eliminarán dicho recelo.

En definitiva, el avance de la telemetría constituye una verdadera revolución en la manera de abordar el precio de los seguros, de forma que en el futuro en lugar de valorar el tiempo como factor de riesgo, se considerará la distancia recorrida como la verdadera variable de exposición al riesgo y, además, se tendrán en cuenta los hábitos de conducción, mantenimiento del vehículo y prevenciones particulares de cada asegurado.

Finalmente, la tercera de las aportaciones que les presento se circunscribe en la tarificación dinámica en el contexto de los seguros de salud. En este caso se trata de lograr una evaluación de la longevidad conjuntamente con la predicción del número de servicios médicos demandados por el paciente, en función de las características perso-

nales y del histórico de uso de los servicios registrado. En este caso, el objetivo es mejorar los precios de las coberturas privadas de salud en personas mayores, usando el principio de la diversificación. Cuando una persona se siente enferma y visita al médico, se produce una reclamación en términos de cobertura de su póliza. Simultáneamente, la probabilidad de supervivencia disminuye, debido a que la presencia de una enfermedad se asocia a una reducción repentina de las condiciones de salud, sin embargo, casi de inmediato, la aplicación de un tratamiento médico implica aumentar las posibilidades de sobrevivir. Por lo tanto, se produce un efecto de disminución de las condiciones de salud y de aumento de las perspectivas de supervivencia sin que se sepa cuál de los dos efectos es mayor.

Si la evaluación de riesgo para calcular la prima es proporcional a la demanda de servicios médicos multiplicado por la probabilidad de sobrevivir todo el año, la aproximación de modelo conjunto puede proporcionar una herramienta dinámica para generar una evaluación de riesgo dinámica y personalizada, es decir continua e individualizada, con la que fijar precios competitivos.

Está claro que la transformación de la era digital que se plantea en los tres ejemplos que se acaban de mencionar, con el aumento de recursos informáticos no convencionales y grandes bases de datos no estructuradas, exige un cambio de enfoque en los métodos. En síntesis, la abundancia de datos y la modelización predictiva constituyen una oportunidad para canalizar la teoría del riesgo hasta la predicción de las pérdidas. Los modelos del futuro serán más personalizados, incluso bajo la presencia de restricciones legales y reglamentarias y deberán abordar cómo mantener el carácter mutuo y la exigencia de solidaridad que significa el compartir un riesgo. La modelización predictiva y el *big data* está suponiendo un cambio radical en la forma de plantear los seguros.

Para saber más...

- Ayuso, M., Guillen, M., Pérez-Marín, A.M. (2014) "Time and distance to first accident and driving patterns of young drivers with pay-as-you-drive insurance". *Accident Analysis and Prevention*, 73, 125-131.
- Guelman, L., Guillen, M., Pérez-Marín, A.M. (2015) "A decision support framework to implement optimal personalized marketing interventions". *Decision Support Systems*, 72, 24-32. [7]
- Piulachs, X., Alemany, R., Guillen, M., Serrat, C. (2015) "Joint modeling of health care usage and longevity uncertainty for an insurance portfolio". *Advances in Intelligent Systems and Computing*, 377, 289-297.