

AMISSIMA, L'IMPORTANZA DEI DATI NELLA GESTIONE SINISTRI

di PAOLO LANDI, direttore area statistico-attuariale di Amissima

ATTRAVERSO L'UTILIZZO DI MODELLI PIÙ AVANZATI (ANCHE DI MACHINE LEARNING), LA COMPAGNIA PUNTA A PERSONALIZZARE LA STIMA DEI CLAIMS TARDIVI E DEL COSTO ULTIMO, COSÌ DA SUPPORTARE IN MODO EFFICACE LA DINAMICA GESTIONALE DEL BUSINESS E L'OTTIMIZZAZIONE TARIFFARIA



In **Amissima**, a partire dal 2017, il nuovo amministratore delegato **Alessandro Santoliquido** ha dato un impulso decisivo alle attività finalizzate alla creazione di valore attraverso l'analisi dei dati e la sintesi di informazioni a supporto del business management.

In ambito Rca, in una prima fase, l'attuarialo della compagnia si è concentrato nell'ottimizzazione delle tariffe e nella creazione di un efficace sistema di monitoraggio delle performance tecniche e i risultati di

tali attività hanno contribuito al progressivo recupero della redditività avvenuto negli ultimi tre anni.

La seconda fase, in corso ormai da più di un anno, si è concentrata nell'analisi multivariata dei dati relativi alle caratteristiche dei sinistri (circostanze, complessità, zona, mezzi coinvolti...) con l'obiettivo di arrivare a una *best estimate* personalizzata del costo ultimo dei sinistri e dei sinistri tardivi.

Tale azione ha l'ambizione di contribuire alla soluzio-

ne di alcune problematiche che caratterizzano il business management assicurativo.

- La stima del costo ultimo dei sinistri e dei tardivi avviene normalmente attraverso consolidate tecniche di analisi e proiezione su base temporale dei dati contabili (*chain ladder*, ecc.). L'ipotesi chiave alla base di tali metodologie consiste nella omogeneità e adeguata numerosità dei gruppi di sinistri analizzati. Tanto più tali ipotesi non si verificano (tipico il caso di discontinuità nella gestione sinistri o nel cambiamento di profilo del portafoglio sinistri), tanto più i risultati forniti dalle varie metodologie tendono a divergere indebolendo sia la precisione, sia l'oggettività dei risultati. Un indirizzo di soluzione di tale problema consiste nell'ampliare l'analisi dalla sola dimensione temporale ad altre dimensioni significative in grado di spiegare gli effetti dei cambiamenti gestionali depurando i trend temporali dai relativi effetti distorsivi.
- La qualità delle analisi tariffarie, fortemente personalizzate, viene degradata dalla limitata differenziazione dei coefficienti correttivi per tardivi e costo ultimo. Tali negatività impattano maggiormente le stime relative agli anni/mesi più recenti che sono quelli maggiormente rilevanti nel determinare sul campo l'efficacia predittiva della tariffa.
- La qualità degli indicatori tecnici a supporto delle azioni di business management attuate su singoli cluster di rischi (territoriali e distributivi), viene degradata dal fatto che la mutualizzazione dei sinistri di punta viene normalmente effettuata sulla base di criteri esclusivamente contabili. Un indirizzo di soluzione consiste nell'adozione di metodologie multivariate che consentano di differenziare il caricamento per sinistri di punta in funzione delle caratteristiche dei sinistri.
- L'utilizzo di differenti metodologie per la stima del costo ultimo e dei tardivi tra bilancio/Solvency II, analisi tariffarie e valutazioni di redditività genera

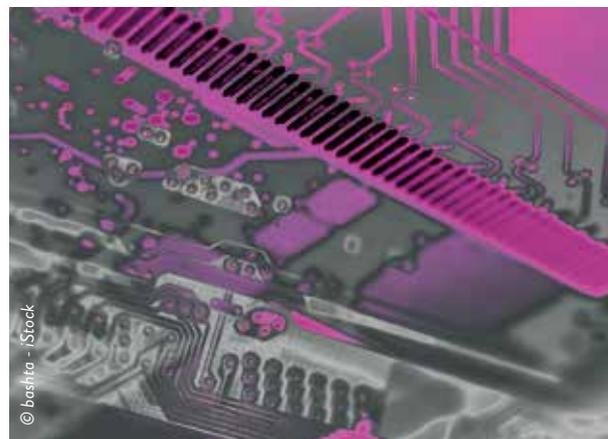
distorsioni e incoerenze informative a vari livelli che impattano l'efficacia dell'azione di business management. L'adozione generalizzata dei risultati di una metodologia multivariata che consenta sia la scalabilità dell'informazione che la sua personalizzazione, rappresenta un possibile indirizzo di soluzione del problema.

I risultati incoraggianti ottenuti da Amissima verso la soluzione di tali problematiche sono stati raggiunti attraverso un oneroso lavoro di progettazione, sintesi e ottimizzazione delle strutture multivariate di analisi. Nel seguito verranno forniti alcuni concetti e indicazioni sulla struttura generale del sistema di modellizzazione.

LA STIMA PERSONALIZZATA DEL COSTO ULTIMO

I punti chiave sono i seguenti:

- un sinistro denunciato ha una determinata probabilità di essere chiuso, con pagamento, nel mese successivo alla denuncia; qualora ciò non avvenga esso ha una probabilità di essere chiuso nel secondo mese, nel terzo e così via. Tali probabilità possono essere stimate attraverso analisi multivariate basa-





te su variabili che sono significative nel determinare la *durata del sinistro* (es. numero di danneggiati coinvolti, tardività della denuncia, territorio, ecc.)

- Alla stima multivariata della probabilità di chiusura di un sinistro in un certo mese/periodo è stata affiancata una stima multivariata del relativo costo di chiusura, personalizzata sulla base delle variabili che sono significative nel determinare il costo del sinistro (es. anzianità del veicolo, età del danneggiato, numero di partite coinvolte, territorio, tipologia veicoli coinvolti, esiti IP perizia medico legale, contenzioso, ecc.).

LA CATENA DELLE PROBABILITÀ

Per ogni sinistro, attraverso l'applicazione dei vari modelli di frequenza e costo, viene costruita la *catena delle probabilità* di chiusura in tutti i periodi e le corrispondenti stime del costo. Il costo ultimo statistico del sinistro è dato dalla somma dei prodotti delle probabilità e costi relativi a tutti i periodi analizzati.

La necessità di una duplice dimensione dell'analisi si può comprendere intuitivamente considerando che il costo ultimo del sinistro è determinato dalle caratteristiche del sinistro (complessità, gravità lesioni ecc.) ma

anche che i costi medi dei sinistri liquidati dopo molto tempo dalla denuncia sono in generale più elevati rispetto a quelli liquidati in tempi brevi.

Applicando tali modellizzazioni, con opportuni accorgimenti, ai sinistri, chiusi e ai sinistri aperti di un portafoglio sinistri si arriva a disporre di una stima degli stessi coerente e personalizzata in funzione delle loro caratteristiche. Inoltre tali valori stimati presentano un *range* di variabilità molto più ridotto rispetto a quello del costo dei sinistri effettivi (per esempio, per i sinistri *Nocard Lesioni* l'intervallo di variabilità è circa 0-80mila euro) determinando quindi una mutualizzazione personalizzata, e quindi ottimale, sia nella valutazione tecnica dei singoli business cluster (monitoraggio tecnico) sia per l'utilizzo nelle analisi multivariate a fini tariffari.

Esempio: backtest di confronto tra dati interni (costo totale sinistri aperti+sinistri chiusi) e dati stimati dalla modellizzazione:

SETTORE 1 AUTOVETTURE -CARDG- DANNI A VEICOLI

| | Costo totale stimato | Costo totale reale | Costo totale % diff. |
|------|----------------------|--------------------|----------------------|
| 2012 | 56.455.672 | 56.477.950 | 0% |
| 2013 | 43.412.327 | 43.098.427 | 1% |
| 2014 | 28.380.148 | 28.298.421 | 0% |
| 2015 | 26.876.939 | 27.312.582 | -2% |
| 2016 | 28.436.371 | 28.274.420 | 1% |
| 2017 | 38.529.253 | 38.168.170 | 1% |
| 2018 | 30.921.619 | 31.083.652 | -1% |
| 2019 | 20.684.774 | 20.201.723 | 2% |
| | 273.697.102 | 272.915.346 | 0% |

LA QUESTIONE DEI SINISTRI TARDIVI

I punti chiave sono i seguenti:

- in generale, fissato un certo periodo di avvenimento, il numero delle denunce tardive che si aggiungono a quelle già pervenute diminuisce progressivamente con il trascorrere del tempo. Inoltre il numero dei tardivi cambia significativamente in funzione sia della zona ove sono avvenuti i sinistri ma anche in funzione della loro tipologia e complessità.
- Assumendo come obiettivo la stima della percentuale di sinistri tardivi attesi, in funzione del periodo e delle caratteristiche dei sinistri denunciati, sono state effettuate delle analisi multivariate che hanno generato dei modelli che, applicati a gruppo di sinistri con certe caratteristiche, stimano la percentuale di tardivi attesi.
- Applicando i modelli multivariati a un portafoglio sinistri è possibile ottenere la stima del numero dei sinistri tardivi, attesi per ogni singolo sinistro denunciato, in funzione delle sue caratteristiche e della sua *anzianità*.
- Alcuni fattori correttivi vengono poi stimati per tenere conto del diverso costo medio dei tardivi rispetto ai sinistri denunciati.

Esempio: backtest di confronto tra dati interni e dati stimati dalla modellizzazione

SETTORE 1 AUTOVETTURE -CARDG-

| totale | avvenuti denunciati | simulati | reali | simulati/denunciati |
|---------|---------------------|----------|---------|---------------------|
| 1/12/14 | 15.015 | 15.015 | 15.018 | 0% |
| 1/12/15 | 26.284 | 26.287 | 26.295 | 0% |
| 1/12/16 | 28.352 | 28.409 | 28.433 | 0% |
| 1/12/17 | 34.052 | 37.422 | 36.916 | 1% |
| TOT | 103.703 | 107.132 | 106.662 | 0% |

COME OTTENERE UN PORTAFOGLIO SINISTRI ADJUSTED

Applicando a un portafoglio sinistri prima i modelli di stima del costo ultimo, sia ai sinistri chiusi sia a quelli aperti, e successivamente i modelli stimatori dei tardivi si ottiene un portafoglio sinistri adjusted che presenta le seguenti caratteristiche:

- coerenza e omogeneità tra i valori dei sinistri aperti e quelli chiusi e allineamento dei valori al costo ultimo sia a livello complessivo, sia sui singoli cluster di variabili significative (territoriali, ecc.);
- valori di costo *ultimo&tardivi* personalizzati per ogni sinistro e legati alle caratteristiche del sinistro;
- range di variazione del costo dei sinistri significativamente ridotto. La mutualizzazione del costo dei sinistri di punta è personalizzata e ottimizzata per il suo utilizzo nelle valutazioni di redditività e nelle analisi tariffarie.

Va infine sottolineato che l'utilizzo di analisi multivariate non sostituisce ma integra le tradizionali metodologie di analisi e stima del costo ultimo (chain ladder ecc.) in quanto la proiezione della variabile anno/generazione, sempre presente nelle analisi multivariate, resta comunque assolutamente fondamentale, ed è facilitata dal fatto che il suo andamento è depurato dall'effetto delle altre variabili.

IL CUORE DEI MODELLI DI MACHINE LEARNING

Nel progetto Amissima, i test e i backtest effettuati sia sulle singole generazioni sia sulle altre variabili di personalizzazione mostrano uno scostamento di pochi punti percentuali rispetto ai dati reali, e una bassa sensibilità a diverse selezioni di variabili e di ottimizzazione modelli. Sulla base di ciò si ritiene che attraverso l'utilizzo di modelli lineari generalizzati e prospetticamente anche di modelli più avanzati di *machine learning* sia possibile arrivare progressivamente a delle modellizzazioni personalizzate della stima dei tardivi e del costo ultimo dei sinistri in grado di supportare in modo efficace la sempre più accelerata dinamica gestionale del business e di ottimizzazione tariffaria. 