Technologies and Big data in Motor Insurance as instrument to combat frauds

Sara Landini

University of Florence

sara.landini@unifi.it

It's a matter of fact that Motor Insurance is a compensation mechanism having a relevant impact on the free movement of persons and vehicles. Therefore it should have reasonable costs .

We can put at the top of the list of the reasons of increased cost of motor insurance: the increase in the propensity to claim, the increase in the amounts awarded with specific regards to physical damages not objectively assessable (like in case of whiplash claims) and to non pecuniary losses. Non pecuniary losses can be characterized as losses or harm not assessable in money on the basis of any standard financial yardstick. We can consider for example pain and suffering as a result of physical injury, or even the death of a person.

Another reason of insurance increased cost could be frauds. Typical motor insurance frauds are: the so called crash for cash (for example a person causes crashes with innocent motorists to make fraudulent insurance claims) and the cases of exaggerated claims (for example a real accident may occur, but the dishonest victim may take the opportunity to ask compensation also for a whole range of previous minor damages of the vehicle).

Instruments to combat frauds can be: Black Boxes ; databases to detect frauds ; the use of Big data for frauds prevention and frauds detection.

As well known the term big data indicates a complex of big data size that can be used to form new knowledge through relationships between data knowable . The big data have a heuristic value. In fact they represent the starting point for identifying correlations that can be relevant for future developments [1] .

There are various techniques used:

1- The "data mining" is the process of analyzing data from different perspectives, summarizing it into information and finding correlations [2]

2- The "data fusion" is the process of integration of multiple data and knowledge expecting that fused data are more informative than the original inputs [3].

---

[1] Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., et al. (2011). Big data: The next frontier for innovation, competition, and productivity. The McKinsey Global Institute; McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. Harvard Business Review.

[2] Cabena; P. Hadjinian; R. Stadler; J. Verhees; A. Zanasi. *Discovering data mining from concept to implementation*, Prentice Hall PTR 1997; Dulli Susi; Furini Sara; Peron Edmondo. *Data Mining, Springer Verlag, 2009;*
[3]*Hall, David L.; McMullen, Sonya A. H. (2004). Mathematical Techniques in Multisensor Data Fusion, Second Edition. Norwood, MA: Artech House, Inc.; Mitchell, H. B. (2007). Multi-sensor Data Fusion – An Introduction. Berlin: Springer-Verlag.;Das, S. (2008). High-Level Data Fusion. Norwood, MA: Artech House Publishers.*

3- The Clustering procedure. The goal of data clustering is to organize a set of data into clusters so data contained in the same cluster are more similar in respect to objects in different clusters.

4- *The Regression analysis* is used to estimate the strength and direction of the relationship between variables that are *linearly* related to each other.

There is no single definition of big data . Sometimes the adjective "Big" refers to the amount of data , other times it is reported to their extent and granularity analysis we can carry on them . In general, they are characterized by large volume : this means that the amount of data can not be handled with traditional methods such as spreadsheets or the data base . Another feature is the large velocity relating to the rapid transfer of information . Eventually there is the aspect of variety of data. Big data come from different and have different formats funds.

The Big Data may be used for purposes of frauds detection and frauds prevention of particular interest to companies not only in the field of Motor Insurance Claims. Through Big Data technologies companies are able to manage to improve their fraud detection and pattern identification capability. An efficient metodoly is to apply a Bayesian model in fraud recognition combined with Big Data analysis techniques. This approach, includes data discovery through all the available internal and external structured and unstructured data sources, integrated with the computational capabilities of a Big Data infrastructure in order to support the claims during all phases of the investigation.

We can indicate the following phases:

1-   network analysis that will identify any historical relationship between the actors in a specific claim, in the past that could suggest a propensity to commit a fraud.

2-   a clusterization of the actors and related behaviors based on a self-learning statistical model. It permits a better representation of relations and attitudes to plausible fraud existence.

From a viewpoint of a similar legal use of big data it must still be governed to avoid similar correlations allow to attribute liability without assessing the action of the people .

In short, the risk is that people are held responsible and punished before they have acted. A such perspective seems to be contrary to the basic values of all juridical systems : fairness and justice.