Big Data in insurance

Veľké dáta v poisťovníctve

Zuzana Krátka

Abstrakt: Exponenciálne rastúce objemy dát a schopnosť spracovať ich v reálnom čase považujú mnohí experti a vrcholoví manažéri poisťovní za jeden z najvýznamnejších trendov najbližších rokov. Zhromažďujú a analyzujú veľký objem dát z množstva rôznorodých zdrojov a snažia sa z nich získať čo najviac hodnotných informácií, aby lepšie porozumeli potrebám svojich klientoch a maximálne ich zohľadnili v dizajne svojich produktov. Tento článok je zameraný na analýzu využívania veľkých dát v poistnom sektore.

Abstract: The exponentially growing volumes of data and ability of their processing in real time are considered by many experts and managers of insurance companies as one of the most important trend for next years. They collect and analyse big volume of data from many different sources and they are trying to achieve the most valuable information for better understanding of client´s claims and to use them maximally in new products. In this paper we will concentrate on the analysis of the use of Big Data in insurance.

Kľúčové slová: veľké dáta, techniky spracovávania veľkých dát, poisťovníctvo

Keywords: Big Data, technics of the big data processing, insurance

JEL classification: G22

1. Introduction

The value and importance of the information is steadily growing. Information – the knowledge and skills derived from them – are currently the engine of the world economy. With the development of a global network of the internet, the transmission of information over short and long distances became a simple activity. The development has caused a huge amount of information available today, with a relatively low-cost almost anywhere and to anyone.

As a result of the rapid development of information technology and digital communications, and also other relevant factors, in particular advances in the process of globalization, there is a fundamental change in society and its economy. The core of the new economy is made up of just the information and communication systems based on sophisticated information-communication technologies. Gain a competitive advantage increasingly depends on investments in modern electronic systems, because the speed, quantity and transaction interdependence are incomparably greater than ever before in the history of human society.

The intelligent data merging is still the one of main principle in insurance. We can say that insurance market worked with the big data from its beginning and always is working, even though there is a higher processing speed, big diversity of data and new possibilities nowadays. It is not important if we talk about insurance or reinsurance companies, both of

them have to perfectly understand the risks and rate them correctly. The Big Data can be very useful.

2. Characteristics of the Big Data

The Big Data is often characterized by the „3V": Volume, Velocity and Variety. In case the data meets at least two of three „V", we can talk about the big data.

Volume

The first feature of the Big Data is their big volume. We talk about such big volume of data which is not possible and analyze by common hardware or only with high expenses. The Big Data Technologies are used in case it is more effective to use massive parallel data processing on hundreds or thousands common servers then use big super computer or configuration of super computers. The big data have hundreds or thousands terabytes (1 terabyte = 1012 bytes) and more. The volume of created data raise in incredible tempo in consequence still growing worldwide using of internet and digitizing of day-to-day life.

The volume of canned data for analysis will grow very fast. It is connected with continuing digitizing, miniaturization and ambition for bigger mobility. The average smartphone contains approximately fifteen sensitive sensors which change this mobile phone on important multifunction equipment for communication, access to internet, navigation, purchases and other activities.

The big part on expanding volume of data will have also the devices of the so-called internet of things, different sensors or intelligent devices and domestic appliances. It is estimated that the internet of things will generate 400 zettabytes of data until year 2018. The more and more common products will be equipped with different sensors. Their producers will be able to collect data about using of their products by customers and to provide the service. It is more common today that the car factory can connect the car with the costumer remote via internet, initiate the diagnostic and collect data necessary for service control. It will have a big influence on the assistance services of insurance companies in future.

Velocity

The big volume is not only feature of the big data. There is specific for the big data that the volume raises very quickly and necessity to process them very quickly. More and more data should be updated and be at disposal very fast. The ability of processing data in real time brings new possibilities for the insurance companies, e.g. the possibility of sending information about the weather to all their insured customers in disaster area in real time to avoid the damages and loss on health and people´s life. Such information are useful only in case that they are provided quickly – their providing after few hours later (sometimes only few minutes later) are useless.

Variety

The big data are specific in their variety of data´s formats. They are not only the usual structured data, but also the semi-structured data, like e-mails, text documents, statuses on Facebook, logs, and unstructured data like video records, audio records and others, too. There is used the different methods for processing of unstructured data, especially Data mining. It allows the empirical correlations for identification among notes of internet forums, social medias and own databases of proper company. It is possibility to collect information about needs of clients for the insurance companies. These information can make basis for sale or service activities.

The term "Big Data" can also be summarized by the "5V" definition – Volume, Variety, Velocity, Veracity and Value, where the analytics applied to this mountain of data has the capacity to provide the insights for both risks and opportunities:

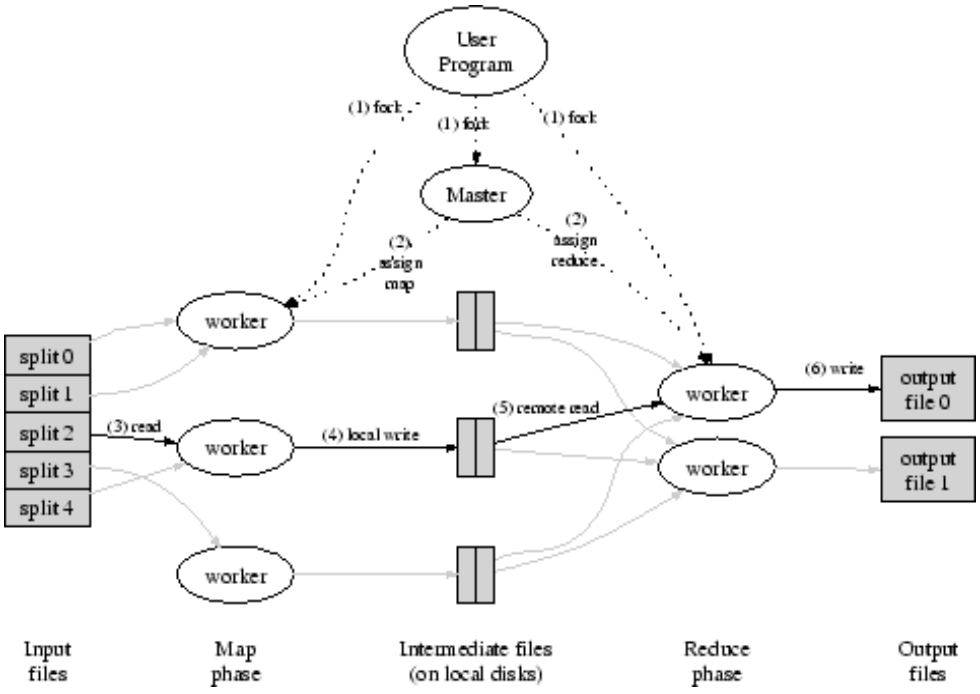
- Volume refers to the vast amount of data generated every second.
- Velocity refers to the speed at which new data is generated and the speed at which data moves around.
- Variety refers to the different types of data we can now use.
- Veracity refers to the messiness or trustworthiness of the data.
- Value refers to our ability to turn our data into value.

3. The technics of the Big Data processing

New technologies are generating a flood of data streams that far exceed the capabilities of traditional databases. Analytics are the tools to detect patterns that lie behind the tables, buried in mountains of data.

In 2004, the software engineers of Google launch the article about method known like MapReduce which simplified and reduced the price of the programming of parallel processing of data. MapReduce is now the most universal way of massive parallel processing of the big data.

Fig. 1: MapReduce: Simplified Data Processing on Large Clusters



Source: Google, Inc., 2004

Thanks to Google which allowed to use this method by other users, there was made several open space platforms based on the MapReduce (later also next technics of the big data processing), which enable the academic and commercial area to start processing the big data with minimal opening costs. The most used are Hadoop, Hive, MongoDB, CouchDB, Riak and other. The phenomenon of the open source software is the next important factor of the big data development. The cost reduction in processing of the big data is caused also by creation of the big clouds centers which allow to rent hundreds or thousands cheap virtual computers

with pre-installed applications, so the processing of the big data is in the financial possibilities of each insurance company today.

4. The Big Data have potential to significantly change the insurance sector

The intelligent data merging is still the one of main principle in insurance. We can say that insurance market worked with the big data from its beginning and always is working, even though there is a higher processing speed, big diversity of data and new possibilities nowadays. It is not important if we talk about insurance or reinsurance companies, both of them have to perfectly understand the risks and rate them correctly. The Big Data can be very useful. The Big Data is a phenomenon connected with a large growing volume of new data and their analysis. The phenomenon of the Dig Data is influenced by quick digitizing of industrial and commercial processes and by webs with bigger amount of linked websites, databases and social networks. The exponentially growing volumes of data and ability of their processing in real time are considered by many experts and managers of insurance companies as one of the most important trend for next years. The big data is used by insurance companies for marketing analysis, reducing the outflow of customers, optimization of stocks, prevention of technological failures and for many others purposes. They collect and analyze big volume of data from many different sources and they are trying to achieve the most valuable information for better understanding of client´s claims and to use them maximally in new products.

Direct insurers were pioneers in collecting and exploiting new data to develop more detailed pricing systems using predictive models. Many insurers now use customer data to estimate not only short-term risk behaviour, but also renewal patterns and conversion rates of potential new customers. All these have a direct impact on the pricing strategy, since acquiring new business is expensive and profiles with poor retention rates are a bad investment. Conversion rates of new business are monitored to design more effective marketing campaigns. Some companies even use automatic price adjustments based on conversion rates. In liberalised markets, direct insurers will experiment with changes in the number and type of questions applied to different customer profiles, policy forms can evolve and adapt to changing circumstances. But this also means that the data collected is not a consistent set of data, and this can complicate a lot of standard analysis.

Customer sentiment and feedback can be monitored in online forums, feedback comments, e-mails, etc. Insurance customers can and do communicate with one other. Comments left on public online forums remain as a record that can be read by any search engine. There are applications for collecting and analysing information provided through social networks. Other unstructured data such as claim reports, medical records and court decisions all contains information that is critical to the company but difficult to extract. Text-mining tools are used to extract information from unstructured notes, and are even applied to converting handwritten records to detect trends and patterns. The rise of big data and transparency of the internet has also raised concerns about the right to privacy. Some regulators have published guidelines to protect citizens' privacy and insurance companies will have to monitor how data is collected and used.

As with most technological developments, big data will serve to differentiate the winners from the losers. Companies willing to adapt and learn from experience will survive when traditional methods are no longer good enough. The key to survival is adaptability. In the more developed markets, it is no longer possible to simply copy the premium rate from leading competitors. Premium rates are constantly changing and depend on too many

variables. Knowledge about the client is a result of a lot of proprietary information and analysis.

Fig. 2: Potential applications Big Data in the insurance industry

Source: own

The main advantages of the use of Big Data in insurance are following:

- The work with structural and nonstructural data in combination with innovative analytics methods opens the new possibilities for insurance companies. The insurance products can be more focused on the individual needs of clients using the big data and can help preciously estimate the prices and risks. The qualitative changes in structure of existing risks wear markedly nowadays and the new global risks appear with hardly estimated effects. The risks are more global, very complicated and extremely systemic. The big data can be very useful in such situation.
- Big data is a term applied to data sets whose size or type is beyond the ability of traditional relational databases. Big data comes from sensors, devices, video/audio, networks, log files, transactional applications, web, and social media - much of it generated in real time and in a very large scale. The examples of datasets which peoples create every day - the smart phone checking the blood

pressure, the car's navigation system, the websites you visit daily - they make the daily life easier but they also do much more: they model the health and motor liability risk profile.

- In underwriting and claims management, business analytics offers the possibility of optimisation. Employed in the past in what was really a more qualitative basis, techniques such as Six Sigma can make a quantitative contribution to process improvements by using more complex statistical modelling approaches. As a result of the high level of manpower required in underwriting and claims management, business analytics offers opportunities for major improvements, particularly in this field. In claims settlement, for example, the trade-off in each individual loss between operational and medical costs or possible savings in treatment, which is determined by the accuracy of the inspection, can be calculated. While the operating costs based on the inspection period are fairly simple to calculate, a complex, data-based approach is required for the correlation with medical savings. Irrespective of its complexity and the probabilities of fraud and abuse already discussed, the optimal inspection effort can be defined for each claim and the overall process therefore made more efficient.
- In product development the combination of internal and external data plays a key role in designing customised products. Intelligent use of social media data through content analytics, i.e. the analysis of unstructured text and filtering of core data, increases knowledge of consumer behaviour. These data, which can be extracted free of charge, are then combined with internal data on the basis of a statistical similarity term. This provides information on demand behaviour and the insurability of existing and potential clients. Policyholders are divided into homogeneous population groups that are, however, as heterogeneous as possible with regard to their interest in insurance products. Possible demand for new insurance products and the preferred sales channels are then determined for each group/cluster. This paves the way for perfectly customised offers.
- Using statistical methods to systematically observe global data streams can also help in the timely detection of potentially relevant losses. This information can then be checked against a company's own portfolio, serving as a kind of early-warning system and enabling loss minimisation measures to be taken at an early stage.
- Very important for insurance industry is improving the detection of fraud. The most prevalent fraud is misrepresentation either at the point of underwriting (omitting information that affects risk assessment) or at the time of the claims declaration. Many insurers use some scoring system to determine which claims are worth investigating for fraud, and there are systems for investigating the links between the claimants and other interested parties (workshops, adjusters, etc.) to determine whether there is a pattern.
- In sales the cross-selling probability of an existing client can be predicted. This is then consolidated into an individual business case for that client, together with the forecast of the best possible sales channel, the acquisition costs involved and the expected revenues for the new product. The optimum procedure in economic terms is then determined on this basis for each client.
- The enormous increase in additional data can prove to be a decisive competitive advantage. Pricing, underwriting, claims management and client service can in

all cases be optimised. A purposeful combination of big data and business/predictive analytics helps to identify and predict individual risks, client behaviour and clients' needs more precisely, so that customised insurance products can subsequently be offered. It is therefore imperative for insurance companies of every size to define their own individual big data strategy.

5. Conclusion

The intelligent merging is still the one of main principle in insurance. We can say that insurance market worked with the big data from its beginning and always is working, even though there is a higher processing speed, big diversity of data and new possibilities nowadays. It is not important if we talk about insurance or reinsurance companies, both of them have to perfectly understand the risks and rate them correctly. The Big Data can be very useful.

The Big Data is a phenomenon connected with a large growing volume of new data and their analysis. The phenomenon of the Dig Data is influenced by quick digitizing of industrial and commercial processes and by webs with bigger amount of linked websites, databases and social networks. The intense development of the Big Data is caused also by falling prices of processors and data stores with hardware visualization and the formation of algorithm and technics allow massively parallel processing of data on common, relatively cheap hardware. The analysis of the big volume on-line and also off-line data is still quicker and cheaper thanks to the modern technologies, globalization and digitizing.

The exponentially growing volumes of data and ability of their processing in real time are considered by many experts and managers of insurance companies as one of the most important trend for next years. The big data is used by insurance companies for marketing analysis, reducing the outflow of customers, optimization of stocks, prevention of technological failures and for many others purposes. They collect and analyze big volume of data from many different sources and they are trying to achieve the most valuable information for better understanding of client´s claims and to use them maximally in new products.

Of course, the Big Data and its models cannot solve all of problems of the insurance companies. There is no perfect model and the big dependence on data and processing and analyzing models cannot lead always to the optimal result. There is a risk that reality will be replaced by data. This phenomenon could be noticed before financial crisis in 2008 when many people thought that the financial models and financial data was reflected the reality, but it was no true. It is necessary to register that the big data are made by different devices, go through different systems and they are locked in different databases and these reasons caused the big differences in their credibility and redemption value. The next problem is the simplification of reality in processing model of the big data, in spite of their complexity. There is possibility that they will not cover all of relevant areas and factors and finally, they will not have a sufficient redemption value. The scientist, managers and experts in insurance industry should not rely upon the big data uncritically, but the calculations of risk, price and others significant decisions should be always the combination of science and personal deduction.

This paper presents partial results of the research conducted within the VEGA project number 1/0221/17.

Literature

AUER, D. J. – KROENKE, D. M. 2015. *Databáze*. Brno: Computer Press.

CRO FORUM. 2015. *Big Data & Analytics: the algorithm of modern business*. http://www.thecroforum.org/wp-content/uploads/2016/01/cro-forum-big-data-201602011.pdf

DEAN, J. – GHEMAWAT, S. (GOOGLE, INC.). 2004. *MapReduce: Simplified Data Processing on Large Clusters*. OSDI'04: Sixth Symposium on Operating System Design and Implementation, San Francisco, CA, December, 2004.

GEMIGNANI, Z. – CHRIS GEMIGNANI, C. – GALENTINO, R. – SCHUERMANN, P. 2015. *Efektivní analýza a využití dat*. Brno: Computer Press.

HOLUBOVÁ, I. – KOSEK, J. – MINAŘÍK, K. – NOVÁK, D. 2015. *Big Data a NoSQL databáze.* Praha: Grada.

KOSEČEK, M. 2015. *Big Data*. http://aadaptive.biz/freelancer/big-data/

LIENHARDT, F. 2014. *Big Data for Insurance Industry*. Swiss RE.

MUNICH RE. 2015. *Topics Magazine* 1/2015. https://www.munichre.com/site/corporate/get/documents_E-56262598/mr/assetpool.shared/Documents/5_Touch/_Publications/302-08466_en.pdf

MAYER-SCHÖNBERGER, V. – CUKIER, K. 2013. *Big Data: A Revolution That Will Transform How We Live, Work, and Think*. London: John Murray Publisher.

SCHMARZO, B. 2013. *Big Data: Understanding How Data Powers Big Business*. Wiley-Blackwell.

SUJA, R. 2014. *Big Data*. Bratislava: Inštitút informatiky a štatistiky. http://www.infostat.sk/web2015/sk/_publikacie/Big_Data.pdf

Author's address:

Zuzana Krátka, Mgr., Ing., PhD.
University of Economics in Bratislava
Department of Mathematics and Actuarial Science
Dolnozemská 1, 852 35 Bratislava, Slovak Republic
e-mail: zuzana.kratka@euba.sk