Classification of operational risks in construction companies on the basis of big data

Tatyana Kostyunina^{1,*}

¹Saint Petersburg State University of Architecture and Civil Engineering, Vtoraya Krasnoarmeiskaya str. 4. St. Petersburg, 190005, Russia

Abstract. Nowadays, Big Data is commonly used in many business sectors. Its use is also relevant for the construction industry. One of the most promising areas of Big Data technologies application is their use for risk analysis and assessment. Big Data represents an efficient way to manage modern risks by analyzing the unlimited amount of structured and unstructured information. The study examines principles of operational risks classification in construction companies on the basis of Big Data technologies. The final goal of such classification is the creation of a solution pattern for subsequent use of Big Data. As an example, a solution pattern for such business problem as "Construction: Detection of Insurance Fraud" is created. Application of the Big Data analytics for fraud detection has a series of advantages as compared to traditional approaches. Insurance companies can build systems that include all relevant data sources. An analysis of operational risks by means of self-organizing Kohonen maps on the basis of the Deductor analytical platform is performed.

1 Introduction

In a recent decade, the term "Big Data" has become a symbol of a revolutionary breakthrough in the field of data processing. Initially it literally meant "data bulks". The meaning gradually expanded and currently it includes: huge data volumes, technologies of data processing and use, methods for information searching in data bulks.

Big Data have become commonly used in many business sectors. They are used in healthcare, telecommunications, trade, and logistics, in financial companies, as well as in the sphere of public administration. Several studies [1, 2] contain information regarding the state of the world and Russian markets of Big Data technologies. Forecasts for their implementation in the coming years are very optimistic for many fields of economics.

The construction industry has a plenty of examples of Big Data technologies application. First of all, prospects of "smart homes" should be mentioned. All kinds of integrated sensors collecting various data (pressure, temperature, number of visitors, environmental parameters) are meant here. The collected data are accumulated and analyzed. Finally, decision-making systems enabling complete automation of building operation are developed. Sensors can control the technical state of a facility in real-time, thus increasing its safety.

* Corresponding author: tnktn@yandex.ru

For a long time, the world has been looking closely at the analysis of Big Data as a tool for the creation of "smart cities". Programs that might consider and systematize information on citizens (their desires, needs, activities and ways of transportation) are very promising. It is expedient to use those data during city development planning and creation of "smart infrastructure".

Historical Big Data can be analyzed to reveal patterns and probabilities of construction risks in order to manage new projects. Currently, Big Data are widely used to optimize transport systems, find construction sites and allocate facilities, evaluate investment costs of facilities and resolve many other issues.

2 Operational risks in construction

One of the most promising areas of Big Data technologies application is their use for risk analysis and assessment. Big Data represents an efficient way to manage modern risks by analyzing the unlimited amount of structured and unstructured information.

Within the framework of this study, we will not address the whole variety of risks in the construction industry but rather will focus on operational risks. Such risks exist for all business processes including the construction industry. A problem of operational risks classification with consideration of the Big Data factor will be analyzed below.

At present, there is no clear definition of the "operational risk" concept in the economic literature. The following definition is most commonly used: **Operational risk** is a risk arising from execution of company's business functions. This concept includes fraud risks and external event risks [3].

In other words, it is a risk of loss due to the following factors:

- internal processes do not comply with the nature and range of business activities or with the requirements of the applicable legislation;
- disruption in business activities due to unintentional or intentional actions of employees;
- insufficient capabilities of corporate systems;
- failures of information, technological and other corporate systems;
- external events or fraud leading to disruptions in business activities;
- etc.

Operational risk is inherent in all business processes and systems, and the effective management of operational risk has always been a fundamental element of companies' risk management systems.

Classification of risks, including operational ones, is carried out on the basis of several distinguished criteria [4]. Such criteria include: risk sources (see Table 1), types of business activities, etc.

Table 1. Classification of operational risks by sources.

Sources of operational risks				
Management quality	System risks	Human factor	Processes	Force majeure
Decision error Insufficiency of data involved	Use of obsolete technologies Insufficient system reliability	Personnel incompetence Operating errors Fraud	Delivery risk Counterparty risks Insufficient process control	Floods Fires Terrorism

As seen from Table 1, operational risks can be divided by sources into those caused by poor management quality, defects in system operation, personnel actions, process structuring,

as well as force-majeure circumstances. It appeared from this classification that the majority of operational risks are in any case related to human activity. For example, direct and indirect losses (damages) arise due to personnel errors with regard to the compliance with internal regulations and procedures, due to decision errors, theft, misuse, as well as insufficient competence of personnel and low qualification level. Moreover, even in cases, when losses are caused by failures in telecommunication, computer or information systems, they mainly originate due to human errors.

Another group of operational risk sources associated with personnel includes misuse and fraud due to dishonesty of employees or insufficient quality of procedures and actions developed to eliminate the risk of misuse. Fraud can be committed when such possibility exists. And such possibility is a result of internal problems and errors. Engagement of a company in commercial relations with shadow economy, intentional conduction of transactions and non-disclosure of their results causing damage to the company may serve as an example of fraud. However, not all possible risks are listed in Table 1. The list may be continued, and each of the mentioned risks may be presented in more detail (divided into components). This points to the complexity of the problem of operational risks classification. The complexity of this problem is indirectly confirmed by Figure 1, where relations between construction companies are shown. Obviously, a business risk analysis should be carried out within the context of business operation overview. Figure 1 shows a scheme of functional relations between construction companies in the context of their cooperation with main economic counterparties [5].

Fig. 1. Cooperation between construction companies.

In the center of Figure 1, relations between companies upon the implementation of an investment and construction project are shown. Financial, material, labor and information flows (indicated by arrows), reflecting movement and consumption of investments, commodities, materials, equipment, projects and technologies (both applied and being developed), money (cash and non-cash) payments, manufactured products, etc., are seen between companies. Tax authorities, local administration and public authorities are not

related to construction companies directly. However, the listed entities significantly influence the progress of investment and construction project implementation thus increasing or decreasing construction company risks. The complexity of relations between construction companies, numerous material and information flows further complicate problems of the classification and analysis of operational risks in the construction industry. Thus, the use of capabilities provided by Big Data technologies becomes especially relevant [6].

3 Operational risks classification in the construction industry with the consideration of Big Data

An approach described in [7, 8, 9, 10, 11] is used in the present study for operational risks classification with the consideration of Big Data. The final goal of this problem is the selection (or creation) of a solution pattern for subsequent use of Big Data. The first step is to classify business problems according to Big Data types. Further, we will use this type to determine the appropriate classification pattern (atomic or composite) and the appropriate Big Data solution.

Out of the variety of business problems in the construction industry, let us choose such problem as "Construction: Detection of Insurance Fraud" to use Big Data. It is necessary to map the business problem to its Big Data type. To do that, we can use the following list:

- Transaction data
- Machine-generated data
- Human-generated data
- Web and social media data
- Biometrics

Let us choose the "Human-generated data" type for our problem.

Categorizing Big Data problems by type makes it easier to see the characteristics of each kind of data [7]. These characteristics can help us understand how the data is acquired, how it is processed into the appropriate format, and how frequently new data becomes available. Data from different sources has different characteristics; for example, social media data can have video, images, and unstructured text such as blog posts, coming in continuously.

Then data shall be assessed according to the common characteristics listed below:

- Analysis type.
- Processing methodology.
- Data frequency and size.
- Data type.
- Content format.
- Data source.
- Data consumers.
- Hardware.

Table 2. ForData characteristics for the "Detection of Insurance Fraud" problem.

Characteristics	Meaning
Analysis type.	An analysis must be done in real time
Processing methodology	Text analytics
Data frequency and size	On demand feeds
Data type	Historical data
Content format	Documents
Data source	Human-generated data
Data consumers	Enterprise applications
Hardware	Commodity hardware

The next step is to identify the components required for defining a Big Data solution for the project. A Big Data solution typically comprises the following logical layers:

1. Big Data sources.

2. Data messaging and store layer.

3. Analysis layer.

4. Consumption layer.

• Big Data sources. All of the data available for analysis, coming in from all channels, are considered.

• Data messaging and store layer. This layer is responsible for acquiring data from the data sources and, if necessary, converting it to a format that suits how the data is to be analyzed.

• Analysis layer. The analysis layer reads the data digested by the data messaging and store layer. It is required to make decisions with regard to how to manage the subsequent problems.

• Consumption layer. This layer consumes the output provided by the analysis layer. The consumers can be visualization applications, human beings, business processes, or services.

For developers, layers offer a way to categorize the functions that must be performed by a Big Data solution, and suggest an organization for the code that must address these functions. In fact, mechanisms for acquiring, processing, storing, and consuming Big Data are meant here. Further, such mechanisms will be implemented using atomic patterns.

Atomic patterns help identify how the data is acquired, processed, stored, and consumed for recurring problems in a Big Data context [8,10]. Moreover, these patterns help to identify the required components. Acquiring, storing, and processing a variety of data from different data sources require different approaches. Basic atomic patterns are given in Table 3.

Table 3. Basic atomic patterns.

Pattern	Purpose
Data consumption patterns This type of pattern addresses the various ways in which the outcome of data analysis is consumed.	Visualization Ad-hoc discovery Augment traditional data stores Notifications Initiate automated response
Processing patterns Big Data can be processed when data is at rest or in motion. This pattern addresses how the Big Data is processed: in real time, near real time, or batch.	Historical data analysis Advanced analytics Pre-process raw data Ad-hoc analysis
Access patterns Although there are many data sources and ways data can be accessed in a Big Data solution, this section covers the most common.	Web and social media data Device-generated data Transactional, operational and warehouse data
Storage patterns Storage patterns help to determine the appropriate storage for various data types and formats. Data can be stored as is, as key–value pairs, or in predefined formats.	Distributed and unstructured data Distributed and structured data Traditional data stores Cloud storage

Atomic patterns focus on providing capabilities required to perform individual functions. Composite patterns, however, are classified based on corresponding end-to-end solutions.

Each composite pattern has one or more dimensions to consider. There are many variations in the cases that apply to each pattern. Composite patterns map to one or more atomic patterns to solve a given business problem.

Solution patterns are developed on the basis of the corresponding composite patterns. Let us address the problem using our example — "Detection of Insurance Fraud". At least two solution patterns shall be created:

- "Getting started" solution pattern (based on the "Store & explore" composite pattern)
- "Gaining advanced business insight" solution pattern (based on the "Purposeful and predictive analysis" composite pattern)

Insurance fraud is an action (or inaction) to gain dishonest or unlawful advantage, either for the party committing the fraud or for other related parties. Types of insurance fraud can be divided into the following categories:

- Policy holder fraud and insurance claims fraud — fraud against the insurer in the purchase and execution of an insurance product, including fraud at the time of making an insurance claim.
- Intermediary fraud — fraud committed by an insurance agent, corporate agent, intermediary, or third party-agent against the insurer or the policy holder.
- Internal fraud — fraud against the insurer committed by its director, manager, or any other staff member.

Issues with the current fraud-detection process lie in the fact that the current fraud-detection methods rely on what is known about the existing fraud cases. Thus, every time a new type of fraud occurs, insurance companies have to bear the consequences for the first time. Most traditional methods work within a particular data source and cannot accommodate the ever-growing variety of data from different sources. A Big Data solution can help to address these challenges and play an important role in fraud detection for insurance companies.

The "Getting started" solution pattern focuses on acquiring and storing the relevant data from various sources inside or outside the enterprise [9, 11]. Those data include unstructured data from such sources as blogs, social media, news agencies, reports from various agencies and regulatory authorities. With Big Data analytics, the information from those various sources can be correlated and combined and — with the help of defined rules— analyzed to determine the possibility of fraud.

Structured data can be easily converted into the format most appropriate for analysis and directly stored in Big Data structured storages. Ad-hoc queries can be performed on this data to get the information like:

- Overall fraud risk profile for a given customer, region, insurance product, agent, or approving staff in the given period.
- Inspection of past claims related to certain agents, employees or clients across several insurers.

The "Gaining advanced business insight" solution pattern predicts fraud at three stages of claim processing: the claim request is just received; the claim processing is underway; the claim is already settled.

In the first case, the claims can be processed in near-real time. For cases 2 and 3, the claims can be processed in batch, and the fraud-detection process can be initiated as part of the regular reporting process or as requested by the business.

The pattern allows using the following indicators to detect fraud, as well as using corresponding technologies to implement systems to combat fraud. Let us review a list of common fraud indicators [12, 13, 14]:

- The authenticity of documents is doubtful.
- Claims are made shortly after the policy inception.
- Serious underwriting lapses occur while processing a claim.

- The insured person is behind in loan payments.
- The damage incurred is not visible.
- A high-value claim has no known casualty.

During analysis, the search for all of these indicators can occur simultaneously on a huge volume of data [15]. Every indicator is weighted. The total weight across all indicators indicates the accuracy and severity of the assumed fraud.

When the analysis is complete, alerts and notifications can be sent to relevant stakeholders, and reports can be generated to show the outcome of analysis.

4 Possibilities of software implementation of the considered business problems

By now, complex software platforms to resolve similar problems have already been developed, e.g. IBM Big Data appliances such as IBM PureData™ System for Hadoop and IBM PureData System for Analytics, cutting across layers.

However, for training purposes, individual layers can be implemented within the framework of Russian analytical platforms, in particular, those based on the Deductor program. An example of the use of this program for analysis of test data by fraud indicators mentioned above is provided below. In particular, self-organizing Kohonen maps are constructed (Figure 2).

Fig. 2. Self-organizing map for classification of fraud indicators.

Analysis of the obtained results shows that two out of four examined parameters ("The authenticity of documents is doubtful" and "Serious underwriting lapses occur while processing a claim") are in the risk zone. They are indicated by red dots on the two upper maps. Thus, some of the considered problems may be resolved using this program.

The next stage of work is the development of an information system for monitoring operational risks. The proposed development methodology [6] allows to automate the processes of collecting information about security incidents, which is the basis of an integrated approach to the creation of operational risk management system.

5 Conclusion

Using business scenarios based on the use case of identifying fraud in the insurance industry, this article describes solution patterns that vary in complexity. The simplest pattern addresses storing data from various sources and doing some initial exploration. The most complex pattern covers how to gain insight from the data and take action based on the analysis.

Application of the Big Data analytics for fraud detection has a series of advantages as compared to traditional approaches. Insurance companies can build systems that include all relevant data sources. Such system helps to detect uncommon cases of fraud.

Analytics technologies enable an organization to extract important information from unstructured data. Although volumes of structured information are stored in data warehouses, most of the crucial information about fraud is in unstructured data, such as third-party reports, which are rarely analyzed.

Moreover, a Big Data solution can also help to build a global perspective of the anti-fraud efforts throughout the enterprise. Such a perspective often leads to better fraud detection by linking associated information within the organization. Combined data from various sources enables better predictions.

References

1. URL : <http://www.ipoboard.ru/files/cms/5e3af134b9942559eb802ea93a1c9050>
2. URL : <https://habr.com/company/moex/blog/256747/>
3. Operational risk, <https://dic.academic.ru/dic.nsf/ruwiki/666706>
4. A.A. Lobanov, A.V. Chugunov, *Encyclopedia of financial risk management* (Alpina.Business.Books, Moscow, 2005)
5. B.S. Seitzhanov. Risks in the construction industry — classification and analysis peculiarities, <http://be5.biz/ekonomika1/r2011/00585.htm>
6. T.N. Kostyunina, *Journal of Physics: Conference Series* **1015** (2018)
7. URL : https://www.ibm.com/developerworks/library/bd-archpatterns1/index.html?S_TACT=05AGX99&S_CMP=CP.
8. URL : https://www.ibm.com/developerworks/library/bd-archpatterns2/index.html?S_TACT=105AGX99&S_CMP=CP
9. URL : https://www.ibm.com/developerworks/library/bd-archpatterns1/index.html?S_TACT=105AGX99&S_CMP=CP
10. URL : https://www.ibm.com/developerworks/library/bd-archpatterns4/index.html?S_TACT=105AGX99&S_CMP=CP
11. URL : https://www.ibm.com/developerworks/library/bd-archpatterns5/index.html?S_TACT=105AGX99&S_CMP=CP
12. L.N. Doronkina, *Development of a risk management system in the investment and construction industry. Monograph* (KnoRus, Moscow, 2006)
13. M.V. Petrovskaya, *Risks in the investment and construction sector: management theory and methods* (KnoRus, Moscow, 2006)
14. E.A. Soboleva, *Specifics of design activity development in the investment and construction sector: detailed elaboration and prospects: monograph* (National Research University, Moscow State University of Civil Engineering, Moscow, 2016)
15. G.A. Makhovikova, T.G. Kasyanenko, *Business risk analysis and assessment. Second edition, updated and revised* (URAIT, Moscow, 2016)