Big Data is a powerful tool for environmental improvements in the construction business

Aleksandr Konikov¹ and Gregory Konikov

¹Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, 129337, Russia
E-mail: a.konikov@gmail.com

Abstract. The work investigates the possibility of applying the Big Data method as a tool to implement environmental improvements in the construction business. The method is recognized as effective in analyzing big volumes of heterogeneous data. It is noted that all preconditions exist for this method to be successfully used for resolution of environmental issues in the construction business. It is proven that the principal Big Data techniques (cluster analysis, crowdsourcing, data mixing and integration) can be applied in the sphere in question. It is concluded that Big Data is a truly powerful tool to implement environmental improvements in the construction business.

1. Introduction
Before proceeding to the main theme of the article, i.e. using the Big Data technology for environmental improvements in the construction business, let us make a comment. Environmental characteristics of buildings are becoming increasingly important. Therefore we believe that in addition to ABC–XYZ–FMR–VEN ranging [1-5] another way of ranging should be implemented which we call here KOL for convenience (letters are set randomly, not to repeat already used ones). We recall that in ABC ranging the construction sites are divided into categories depending on the profitability, while XYZ ranges construction sites depending on the stability of income inflow. In FMR, sites are ranged depending on the visiting frequency. VEN ranging divides sites into absolutely necessary “V”, desirable “E” and the remainder “E”. In the construction sphere, VEN ranging is advisable when designing micro-districts where “N” category covers, for example, schools. In KOL ranging, “K” category may include environmentally safe sites, “O” category — sites with certain environmental issues (for example, some elevated noise levels), “L” category — sites of environmental concern.

2. Materials and Methods
Recently, Big Data has become one of the most demanded trends in the IT field. The Big Data concept is based on the principles which can be called “three V” for convenience:

Volume: the technology allows processing of very big volumes of data;
Velocity: both in the sense of increment rate and the necessity for high-speed processing and obtaining results;
Variety: the possibility to simultaneously process data of different types (including semistructured).

In a certain sense, Big Data continues and improves the concept of Data Mining — smart analysis of hidden regularities contained in heterogeneous data. It should be noted that so called “Soft Computing” is used which is oriented to solving problems with semistructured objects. They include fuzzy sets, fuzzy
logic, fuzzy controller, precedent-based reasoning systems, evolutionary modeling, methods based on the application of decision tree, associative memory and other. At the same time the Big Data technology contains its own solutions making it possible to single out the Big Data as a new development path.

I. Cluster analysis — multivariate statistical procedure performing acquisition of data containing information on items sampling. Dividing the sample into groups of similar items allows simplification of further data processing and decision making while applying an individual analysis method to each cluster [6]. Data compression is an important technique of cluster analysis. If the initial sample is excessively large, it can be reduced by leaving one most typical representative of each cluster. Another technique consists in detection of novelties. Non-typical items are singled out which may not be added to any cluster. Clusterization is a complicated task: there are used probabilistic approach, approaches based on artificial intelligence systems, logical approach (based on decision tree), graph-theoretic approach, hierarchical approach.

In the cluster analysis, each item is described with a set of its characteristics called features. The matrix of distances between the items and the matrix of similarity between the items are used for the analysis [7]. For the sake of justice it should be noted that the cluster analysis method has been known for some time (and applied, in particular, in Data Mining.). But the Big Data technology essentially improved the method.

II. Crowdsourcing — categorization and enrichment of data by the resources of general public engaged in the project based on a public offer without entering into employment relationships [8]. As an example, Wikipedia may serve — an electronic encyclopedia created mainly by volunteers. Another example: Muji, Japanese furniture company, which gathers ideas for its furniture items through its corporate website and makes decisions on production start-up based on contest results.

III. Data mixing and integration — a set of methods and techniques allowing integration of heterogeneous data from various sources to enable in-depth analysis.

It is important to understand that Big Data is not limited to methods and mathematic approaches. There exist concrete software and hardware solutions which are successfully used in various business applications.

For example, in Russia Big Data techniques are used in mobile communications. All companies from the “big three” mobile operators in the RF have division specializing in the big data and they bring additional profit. Besides, the Big Data technology is successfully used in the retail business, banking and other sectors. As for the construction sector, the Big Data technology is not yet widely used in Russia (as far as we know). Yet it is clearly only a matter of time.

3. Results
Application of Big Data in the construction sector seems very promising — both from purely “builder’s” point of view and the environmental perspective. All preconditions exist for successful application of this technology:

The volume of data which should be considered during the construction process (especially in relation to large construction sites) is very big. If the environmental factor is properly considered, this volume will increase even more. Indeed, account should be taken not only of the potential adverse health effects of the construction and finishing materials of the building but also of the fact that two materials may be harmless separately but cause adverse effect in combination.

When considering detrimental effects from nearby industrial enterprises, account should be taken of, among other factors (the amount of emission released into the atmosphere, their composition, etc.), the prevailing wind direction. As an example, a story from the Soviet Union times. One of the authors, then a student, was sent to complete his practical training program to a small city. In its vicinity there was an industrial enterprise which released noxious emissions to the atmosphere. In the building where students lived it was sometimes hard to breathe. But in the city which was approximately equally distant from the enterprise, but on the other side, the inhabitants did not experience any inconveniences since the wind almost always carried the emissions away.
Account should also be taken of the existing and designed highways. Their harmful effects are related to the noise and automobile emissions. Then appropriate measures can be taken, from installing improved window frames and other merely building solutions to negotiations with the municipal authorities on relocating the highway.

The factors which affect the ecology of a building comprise vibrations and noises from industrial enterprises, railways. It should be noted that noise is propagated over various distances depending on the frequency and requires different protection methods at construction sites. It is quite clear that account should be taken of the local landscape and the objects situated between the designed building and sources of noise — digital map or GIS (geo information system) data should be analyzed.

Here we only discussed a small part of the wide problem which can be called “Environmental matters in the construction business”. These include issues related to the influence of construction on the atmosphere, environmental safety of construction materials and items, ecological security of foundations being built, application of noise protection materials and structures and many others. Their resolution is related to processing large volumes of data.

It follows from the above examples that huge volumes of data needs to be taken into consideration in development projects, thus the first precondition (first “V”, Volume) is fulfilled for the Big Data analysis. The same examples demonstrate that the second and third preconditions are also fulfilled (third “V” is Variety), since data of various types need to be processed. The necessity of the Velocity condition (second “V”) is evident, since there is a need for high speed processing and obtaining results with the initial data continuously increasing.

4. Discussion
As for the methods used within the Big Data technology in terms of the environment in the construction industry, here are the most useful ones:
- Cluster analysis. In our case this method is advisable, for example, when evaluating a micro-district: obviously, environmental requirements for kindergartens should differ from those for trading enterprises.
- Crowdsourcing. This method is extremely effective when constructing residential houses: potential inhabitants of such buildings will, voluntarily and free of charge, obtain information related to the environmental situation, i.e. whether that are any plans for construction of roads or railways close to the building, or industrial enterprises which can negatively affect the ecology of the building, etc.

Data mixing and integration. This approach in part corresponds to the above discussed. For example, people planning to become inhabitants of a constructed building can, within their networks, obtain information which will enrich the knowledge about the environmental conditions of the future building — not necessarily negative (see crowdsourcing examples), it can be positive — that a park is planned to be laid out and so forth.

5. Conclusions
In view of the above and the fact that software and hardware solutions are at hand which allow transforming the available knowledge into concrete recommendation and decisions, it may be safely assumed that Big Data is a powerful tool for environmental improvements in the construction business.

References
[1] Konikov A and Konikov G 2016 Procedia Engineering 165 pp. 1052–1056
[2] Konikov A and Konikov G 2016 Logistics and supply chain management 5 (64) pp. 70-73
[3] Borboni A, Mor M, Faglia R 2016 Journal of Dynamic Systems, Measurement and Control, Transactions of the ASME, 138 (11), art. no. 111003. DOI: 10.1115/1.4033831
[4] Aggogeri F, Borboni A, Merlo A, Pellegrini N, Ricatto R 2016 Sensors 16 (10), art. no. 1577 DOI: 10.3390/s16101577
[5] Borboni A, Lancini M 2015 ) Journal of Vibration and Acoustics, Transactions of the ASME, 137 (3), art. no. A1 DOI: 10.1115/1.4029575