No Data but Knowledge

Program DATA-INFO-KNOWLEDGE, Talk 2

Prepared by: Katerina Zalamova, CEO
26 February 2018

ABSTRACT

Today the economy is driven by data as the digitalisation of the daily life is going smoothly. However, what enterprises and governments are looking for is not data but knowledge, answers to their daily challenges. Thanks to the high availability of any kind of data in a digital format different industries emerged over new technology for the treatment of that data. The first market was called the Big Data market and it has emerged in early 2000s but it could not show a disruptive success for its players although it shows deficiency in technologies readiness and business models. The main deficiency of the Big Data Market is the cost of the treatment of the data and the poor ratio between effective data transformation vs. cost.

In this article, we are going to discuss the key factors for successful transformation of the data to information and knowledge. This work is an own study of CREA IDEA LAB and the analysis here presented is publish for first time.

INTRODUCTION

It has been said that our society is living in the 4th industrial revolution, The Society of Knowledge. In the last decades, there has been emerging different markets based on data in digital format as the Big Data, Open Data, software industry for data treatment and analysis, Blockchain and Artificial Intelligence. The main common point for all of them is their aim to improve the decision making process by objectivity that is base on real data while minimizing the uncertainty.

Another common point is the organization of the data in data bases under some criteria of classification. Then, the differences between them come from the use that is given to that classified data. Big Data market offers enormous volume of data for prediction. Open Data market aims to create new business opportunities for the individuals. Thick Data is oriented to identify trends in the consumer behaviour. Blockchain gives value to the data ownership and its sharing process. Artificial Intelligence aims to discover new knowledge beyond the establish data analysis methods.

In conclusion, all industries related to data need to organise the data in a certain manner and transform it for applications. In the continuation, we are going to discuss the transformational process and the key factors for its success.

DATA TRANSFORMATION PROCESS

For the purpose of this paper, we defined data as the documentation of the reality in numbers. For example, data is the quantity of 10 trees in a given park in the city but not their existence.

Then, we define information as the meaning of the data which in the previous example is the existence of exactly 10 trees in that given place. Another way to see the information is as classified data.

Then, we define knowledge as the application of the data or the understanding behind it. In the previous example, knowledge will be the purpose to have 10 trees at that given place.

Now, we can define the transformation process of the data to knowledge with the following:

- Through classification the data is converted to information.
- Through crossing different classified data (information), it is converted to knowledge.

In the previous example, we have the raw data of 10 trees at a given place in the city. The information is their existence on that place. To convert this information to knowledge, we need to cross the data with other data sets, for example the demographic situation around this place, or the noise level, or the traffic. Thus, we can find knowledge about the purpose to have those trees at that place. For example, they are providing a quite island in a intense busy urban area improving the quality of live of the residents by 10%.

KEY FACTORS FOR SUCCESSFUL DATA TRANSFORMATION

Transformation of raw data to information

Looking on the transformational flow of the data to knowledge, the first key factor is the definition of the criteria for classifying the data and organising it. Depending on this criteria the same data provides variety of informations. For the previos example, we can characterise 10 trees by their location, by their application, by their cost of maintenance or by their environmental or social impact. Through those criteria, we obtain information about the defined classes. For example:

- By application- 10 trees are **sufficient** to create a quite island in a busy urban area.
- By social impact- 10 trees **improved** the happiness of the employees in the offices around them by 10% because of the their green color that is seen from the windows.

Then, at this first step of the transformation from raw data (quantity) to classified data (information), there are two main approaches for the definition of the transformational flow. The first approach aims to characterise all possible classes for a given volume of raw data. The goal of this method is to get the largest list of characteristics and possible applications for the studied data set. Its benefits are a large amount of information that can be extract from a single data set and the indications for different ways to transform that information to knowledge.

The second approach is focused in a specific classification which is driven by external force. This external force is the desire for specific knowledge which is expressed under a specific question. This second methodology for data treatment recognises that external force as a base for the transformational process which will driven the transformation to predictable results. It can be said that this approach is conditioned by predictable results and the transformation depends on those predictions.

The benefits of the second approach are a predictable and focused transformational process and a limited number of data classes which will be transform to knowledge in an established manner.

To resume, the transformation of raw data to information can be done by two approaches:

- natural transformation without any dependance on external conditions;
- or a conducted and focus transformation forced by external conditions.

The success of the transformational process of the data to knowledge greatly depends on the choice of the methodology for converting the raw data in information. This choice gives the framework for the definition of the criteria for the classification and organization of the raw data. Additionally, it establishes the base for the next transformational step of the obtained information to knowledge.

Transformation of classified data to knowledge

Once we transform the raw data to information, we can say that we have already obtained knowledge. However, for the purpose of our previous definition of knowledge, the results from the first stage of transformation can be called “raw knowledge”. The aim of the second stage transformation from “raw” to real knowledge is the understanding behind the data.

The key for the transformation of the “raw” knowledge to real one lies in the criteria for crossing sets of information. We get understanding of something when we take action over it. This is why, at this stage we force an action over the data, for example, to compare the studied data set to other ones.

The criteria for comparison between classified data sets dependes on the purpose and the needed level of understanding. For successful extraction of knowledge, it is needed a clear purpose and a pre-determined level of understanding that is aimed to be obtain.

For example, our purpose is to evaluate the level of importance of the obtained information for future applications. Here, the needed level of understanding of the information is basic as we will continue to classify it. However, we will need higher understanding if our purpose is to find a proof for undergoing phenomenon.

The complexity of the extraction of real knowledge from classified data comes from the need to define the base of this extraction which is build on a definition of the purpose, a level of understanding, a choice of classified data sets and a criteria for actioning over them.

In details, the purpose answers the question “What/Why do we want to know?” The level of understanding defines how much we want to know: improved classification, identification and characterising of relationships, validation of undergoing processes, etc.

Then, with a defined purpose and a level of understanding, we have insights about which data sets we should action together and what kind of actions we should perform. Possible actions over data sets are a comparison, a definition of interrelations, a classification of interrelations, an identification of relational flow “cause-effect”.

The key point here is the definition of criteria which will structure the action taken over the data sets. For example, if we want to compare to data sets, we need to do this comparison in relation to a reference. A reference can be a common characteristic of the data sets or an external factor and its impact over them.

Once we have all those definitions in place, we have settle the framework for the transformation of the classified data to knowledge. The success of the transformational process and the quality of its results strongly depend on the quality of the definitions of its framework.

In difference with the transformation of the raw data to information, its conversion to knowledge is a conditioned, predictable and focus process. There is no any natural process for the knowledge extraction as the transformation aim is the understanding behind the data which is related to a specific action.

DIGITALISATION OF THE DATA TRANSFORMATION

Here, we want to discuss how much of the process can be digitalise with the actual state-of-art of the technology. For the purpose of the discussion, we defined the data transformational flow by the following steps:

1. Generation/Collection of raw data.
2. Criteria definition for the classification of the raw data.
3. Classification and organization of the raw data in data sets.
4. Criteria definition for data treatment and data analysis.
5. Basic data analysis: classification of the data sets for further applications.
6. Historical data analysis: identification of interrelationships and processes.
7. Extraction of knowledge: validation of hypothesises, visualisation of evidences about undergoing phenomenon.

At the moment, the available technologies permit automation of steps 1, 3, 5 and 6 in the above process. The Step Nº2 can be automated in its first approach- natural process (see above). However, there is not any available digitalisation approach for the Step 4 to the moment. Sean McKenna has done very a detailed description of all classification needed for successful data transformation process and why it will be hard to automate [1]. As for Step 7, there is some methodologies that can semi-automate the process.

In conclusions, the existent technology cannot provide solutions for the key points for the success in the data transformation which is closely related to the human creativity and common sense. We suggest that the teams, working on data treatment technologies as software for data analytics, artificial intelligence and similar, improve their skills in criteria definition for classifying and crossing data. We believe those skills are of a high importance to asure a high quality knowledge extraction.

[1] https://track2analytics.com/2018/02/27/five-misconceptions-about-data-science-knowing-what-you-dont-know/