FutureTDM
Explore . Analyse . Improve

CHALLENGES

IMPROVING UPTAKE OF TEXT AND DATA MINING IN THE EU

Facts

Project No: 665940

Program: H2020 | CSA | GARRI-3-2014

Duration: 09/2015 - 08/2017

TEXT AND DATA MINING AS AN ECONOMIC ASSET

What is TDM?

TDM facilitates the extraction of useful and instrumental pieces of information from typically large corpora of essentially unstructured text and other types of data; it also allows for the translation of this information into actionable intelligence for advancing a specific process – be it public policy intervention, market actions or actions performed by other entities for various reasons.

TDM and its Business Connotation

In the business (or organizational in general) context we equate "useful knowledge" with "actionable intelligence". In other words, useful knowledge can be used for action. Therefore, just like specific conditions have to be met in order to be able to secure the finding of useful knowledge, the same applies to TDM understood as a tool for developing actionable intelligence. First, miners must understand not only the data they use but, even more importantly, the context in which they operate. In business, this usually means familiarity with the company and its operations: the more in-depth this knowledge is, the greater the chances of discovering useful patterns and making useful predictions. The lack of mutual understanding between data miners and a given company's strategic decision-makers is one of the most prominent obstacles for developing Big Data projects.

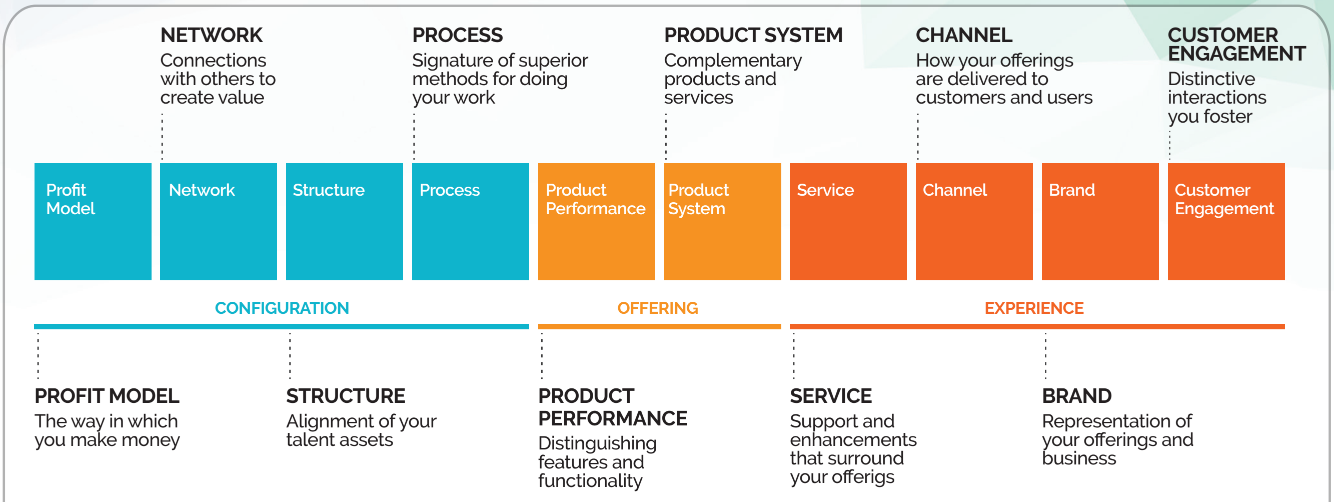
After the context is understood miners can start making decisions about the data they should take into focus and start cleaning and preprocessing it.

Step tag	Explanation of the proces
1 Understanding	First is developing an understanding of the application domain and the relevant prior knowledge and identifying the goal of the KDD process from the practical problem viewpoint.
2 Selecting	Second is creating a target data set: selecting a data set, or focusing on a subset of variables or data samples, on which discovery is to be performed.
3 Preprocessing	Third is data cleaning and preprocessing. Basic operations include removing noise if appropriate, collecting the necessary information to model or account for noise, deciding on strategies for handling missing data fields and accounting for time-sequence information and known changes.
4 Transformation	Fourth is data reduction and projection: finding useful features to represent the data depending on the goal of the task. With dimensionality reduction or transformation methods, the effective number of variables under consideration can be reduced, or invariant representations for the data can be found.
5 Choosing the method	Fifth is matching the goals of the KDD (Knowledge Discovery in Databases)process (step 1) to a particular data - mining method. For example, summarization, classification, regression, clustering and so on.
6 Exploring	Sixth is exploratory analysis and model and hypothesis selection: choosing the datamining algorithm(s) and selecting method(s) to be used for searching for data patterns. This process includes deciding which models and parameters might be appropriate (for example, models of categorical data are different than models of vectors over the reals) and matching a particular data-mining method with the overall criteria of the KDD process (for example, the end user might be more interested in understanding the model than its predictive capabilities).
7 Discovering patterns	Seventh is data mining: searching for patterns of interest in a particular representational form or a set of such representations, including classification rules or trees, regression and clustering. The user can significantly aid the data-mining method by correctly performing the preceding steps.
8 Interpreting	Eighth is interpreting mined patterns, possibly returning to any of steps 1 through 7 for further iteration. This step can also involve visualization of the extracted patterns and models or visualization of the data given the extracted models.
9 Acting	Ninth is acting on the discovered knowledge: using the knowledge directly, incorporating the knowledge into another system for further action, or simply documenting it and reporting it to interested parties. This process also includes checking for and resolving potential conflicts with previously believed (or extracted) knowledge.

Source: based on Fayyad, Usma, Gregory Piatetsky-Shapiro and Padhraic Smyth. 1994. From Data Mining to Knowledge Discovery in Databases, „AI Magazine", nr 17/3: 37-54.

This project has received funding from the European Union's Horizon 2020 (H2020) Research and Innovation Programme.

CHALLENGES


NETWORK
Connections with others to create value
PROCESS
Signature of superior methods for doing your work
PRODUCT SYSTEM
Complementary products and services
CHANNEL
How your offerings are delivered to customers and users
CUSTOMER ENGAGEMENT
Distinctive interactions you foster
Profit Model
Network
Structure
Process
Product Performance
Product System
Service
Channel
Brand
Customer Engagement
CONFIGURATION
OFFERING
EXPERIENCE
PROFIT MODEL
The way in which you make money
STRUCTURE
Alignment of your talent assets
PRODUCT PERFORMANCE
Distinguishing features and functionality
SERVICE
Support and enhancements that surround your offerigs
BRAND
Representation of your offerings and business

Source: Larry Keeley, Ryan Pikkel, Brian Quinn, and Helen Walters, Ten Types of Innovation: The discipline of building Breakthroughs (Hoboken, New Jersey: John Wiley & Sons, 2013).

Then, they may proceed to shed unimportant data, choose the appropriate data mining method, start exploratory analysis, search for patterns and start interpreting them. The final interpretation should be actionable enough to enable making a decision aimed at solving a given problem. What is important is that this process may have many iterations and loops between any two steps as the process itself is highly creative and requires the miner to possess a high level of complex and interdisciplinary sensitivity.

All these steps in the process point to necessary conditions for TDM to offer value in business projects. Pattern recognition and prediction have to take place within a deep understanding of the complexity of a specific business model. This necessarily has to mean that the business knowledge of data miners has to be updated and aligned to strategic decisions constantly. When they are properly linked, solutions may be mined for thousands of specific business problems to hone greater competitive advantage.

However, the application of TDM may go far beyond what we are observing currently. From a theoretical and a more abstract, perspective, TDM may lead to improvements (or innovations) in (1) business configuration, (2) offering and/or (3) customer experiences.

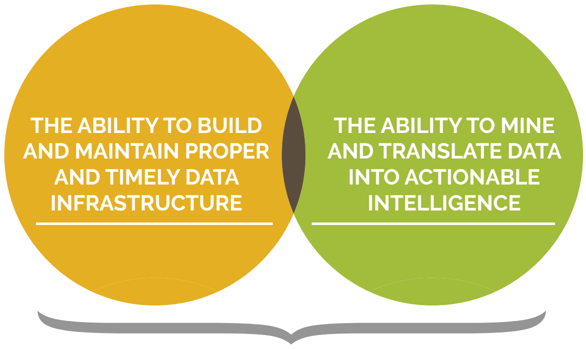
It has been estimated that Big (and Open) Data will give an incremental boost of 1.9% to European economic growth by 2020 (Buchholtz et al. 2014). This growth will transpire mainly through three types of economic gains to organizations.

1. Resource efficiency improvements through better use of information concerning resource waste in production, distribution and marketing activities.

2. Product and process improvements through innovation based on R&D activities, day-to-day process monitoring and consumer feedback.

3. Management improvements through evidence-based, data-driven decision making (Buchholtz et al. 2010: 11).

TDM accessibility and quality will, to a large extent, determine its uptake and therefore also the economic impact exerted by Big Data. While Big Data discussions have so far pointed more to aspects concerned with data infrastructure development and maintenance, TDM presents an opportunity to address the issue of human skills and capacities to extract value from data leveraging data infrastructure.


THE ABILITY TO BUILD AND MAINTAIN PROPER AND TIMELY DATA INFRASTRUCTURE
THE ABILITY TO MINE AND TRANSLATE DATA INTO ACTIONABLE INTELLIGENCE

Incremental 1.9% boost in GDP by 2020
if full TDM potential is used