



Towards Financial Valuation in Data-Driven Companies

M. EUGENIA PÉREZ-PONS¹, ALFONSO GONZÁLEZ-BRIONES^{1,2}, and JUAN M. CORCHADO^{1,2,3,4*}

¹BISITE Research Group, University of Salamanca. Edificio I+D+i, Calle Espejo2, 37007, Salamanca, Spain.

²Air Institute, IoT Digital Innovation Hub, Carbajosa de la Sagrada, 37188, Salamanca, Spain.

³Department of Electronics, Information and Communication, Faculty of Engineering, Osaka Institute of Technology, 535-8585 Osaka, Japan.

⁴Pusat Komputeran dan Informatik, Universiti Malaysia Kelantan, Karung Berkunci 36, Pengkaan Chepa, 16100 Kota Bharu, Kelantan, Malaysia.

Abstract

The following work presents a methodology of determining the economic value of the data owned by a company in a given time period. The ability to determine the value of data at any point of its lifecycle, would make it possible to study the added value that data gives to a company in the long term. Not only external data should be considered but also the impact that the internal data can have on company revenues. The project focuses on data-driven companies, which are different to the data-oriented ones, as explained below. Since some studies affirm that data-driven companies are more profitable, the indirect costs of using those data must be allocated somewhere to understand their financial value¹⁴ and to present a possible alternative for measuring the financial impact of data on the revenue of companies.



Article History

Received: 21 June 2019

Accepted: 03 July 2019

Keywords

Case-Based Reasoning;
Data-Driven Companies;
Financial Valuation;
Recommendation Systems.

Introduction


Nowadays, data-driven companies are widespread in all business contexts.^{19,17,4} However, the real implications of being a data-driven company are huge and not all the companies can easily transform into one.⁵ A data-driven approach is one in which companies organize and analyze their data carefully

in order to improve their customer service and predict future product demands. One of the most essential elements of data-driven companies are big data architectures.⁶ When we say Big Data we refer to large volumes of data, both structured and unstructured, that are generated and stored on a day-to-day basis. Although what is really important

CONTACT Juan M. Corchado ✉ corchad@usal.es 📍 BISITE Research Group, University of Salamanca. Edificio I+D+i, Calle Espejo2, 37007, Salamanca, Spain.



© 2019 The Author(s). Published by Oriental Scientific Publishing Company

This is an  Open Access article licensed under a Creative Commons license: Attribution 4.0 International (CC-BY).

Doi: <http://dx.doi.org/10.13005/ojcs12.02.01>

is not the amount of data a company has, but how it uses the data it collects and the knowledge it is capable of extracting, as well as how this knowledge helps the company improve and grow.

In the process of collecting and organizing the information, many big data projects fail because companies develop rigid architectures which do not have the flexibility to adapt to possible future changes.³ Many companies generate lots of data which they manage through business intelligence applications or dashboards, such companies can be called data-oriented but they are not data-driven because they only analyse a part of all the possible data that could be collected. Data-oriented companies collect real-time and historical data for monitoring, however, unlike data-driven companies, they do not look for the causes, nor do they make recommendations or decisions on the basis of this information.¹⁶

Most of the analyses of data-driven organizations are oriented at the future, they use predictive models to optimize spending or to respond to an element in the supply chain that is at a point of failure.^{18, 10, 8} They help prevent the loss of clients to competition through different strategies such as the collection

of external data or by offering a small discount to customers to persuade them not to leave. Certain tools and skills are necessary in order to perform these activities; However, above all, there must be a business culture that promotes the use of data as a basis for decision-making. It has not been demonstrated yet that data-driven companies benefit more from being oriented both to customers and to the internal organization, nevertheless it seems quite obvious that the decisions that are based on data will always have a better result. Therefore is suggested a CBR architecture to propose an alternative.

Presumably, customer- or user-related data have a direct impact on the company sales and revenues. The data related to the main activities of a company, such as feedback from the users on its products, is important in order to be able to calculate the costs and profits of a company. The aim of this work is to propose an alternative methodology for the measurement of financial impact that the collection, processing and storage of data generated by users has on a company. In the last decade, news reports have stressed the effectiveness of data-driven companies, however a real financial demonstrative or conclusive valuation method has not been proposed yet. However, there are some

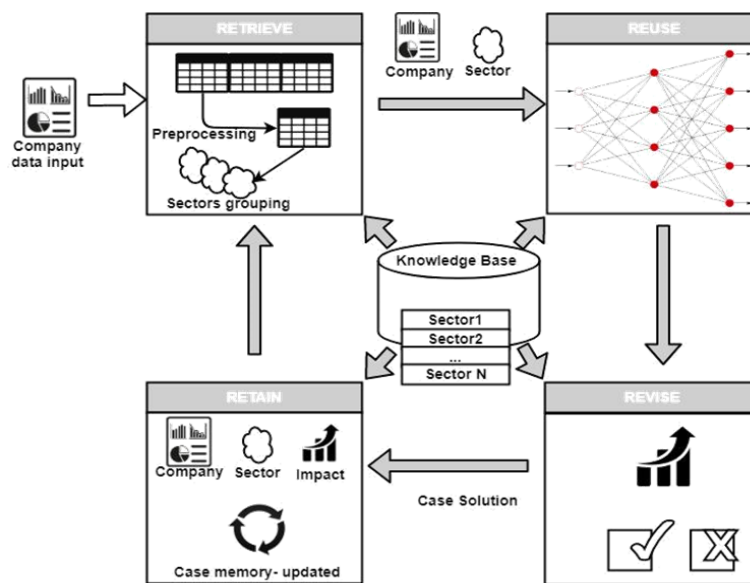


Fig.1: CBR cycle for the prediction of the financial impact of introducing a recommendation system

methodologies in the state of the art that make it possible to value the assets of a company, such as the Data - Information - Knowledge - Wisdom theory¹ which analyzes the value of data at a certain period in time (since it's raw Data to Wisdom) and valuing the data of a company as an asset,¹⁵ are some of the most known and representative methodologies. The actions carried out at each stage of the CBR architecture are detailed below.

Background

The costs that data capture, storage and processing entail are never the same for all companies. The amounts of data collected nowadays are big, however, with the growth of IoT and new technologies, those volumes will increase dramatically and all companies should be able to see where the break-even point is for their data. As mentioned before, since it is very hard to identify the point at which data has a certain value, when considering data as an asset, it can be evaluated with the DIKW methodology or by valuing the company as an asset.

That DIKW methodology is capable of calculating the value of the data at any point in its lifecycle. Also, the data have to be in the life-cycle like other organisational assets, information has a cost (collection, storage and maintenance) and a value (how it contributes to the revenue of a company). However, this is where the similarity ends. Information does not follow the same laws of economics as other assets do, because it has some unique properties which must be comprehended if their real value is to be measured correctly. (Glazer, 1993)⁹ State that data can be shared between multiple business areas and the cost will be the same as if a single party had exclusive use of the information. Another argument against valuing data as an asset is that we will miss the fact that duplicating data entails indirect costs as well as storage costs, these costs are not very noticeable nowadays but will be when the volume of data will increase.¹⁵ Due to those research gaps, the main objective of this work is to propose a methodology for analyzing the financial impact that data processing has on companies. The external data collected from users is going to be used in the development of a method for real conclusive economic assessment or prediction, that is, for the measurement of financial impact (ie: greater profit, lower costs,...).

It is sometimes argued that recommendation systems do not help consumers discover new services or products and only reinforce the popularity of the already popular ones.⁷ However, what is clear is that they are usually beneficial for the companies that use them.^{2,13} Therefore, recommendation systems can serve as a tool for the measurement of financial impact on company benefits.

Proposed CBR Architecture

This work presents a hypothetical CBR architecture for the prediction of the financial impact that recommendation systems have on a company. The methodology will be based on a set of variables that are considered key to understanding the performance and the evolution of a company in a given sector. We have chosen several Key Performance Indicators as our variables which will make it possible to extract company performance patterns and compare them with different companies. Those variables will indicate the financial evolution of the company, the human capital and market concentration. Companies that use recommendation systems and base their company model on it, need to collect large volumes of information to train the artificial intelligence models which help offer products or services adapted to the client's profile.

Some business models view the real-time data generated by customers as the added value of companies. Nowadays, we are capable of extracting value from all types of data as they all contribute to improve business. Nevertheless, it is true that the collection of customer data and customer flow is essential, without those elements companies would not be able to become as competitive. For instance, any social network based on a localization service requires constant information updates and algorithms that process information in real time.

Everyday more data is created and stored around the world, i.e: Google processes over 40,000 search queries every second on average, which is 3.5 billion searches per day and 1.2 trillion searches per year worldwide. As Google's director of research Peter Norvig, puts it: "We don't have better algorithms. We just have more data". The storage, processing and analysis of data entail costs for the companies that collect them. A small mistake in data mining can have serious implications, leading, for example, to the poor

performance of the algorithms that are going to be trained with incorrect data.

To measure the financial impact that recommender-system-facilitated data processing has on enterprises, it is first necessary to study the variables that allow for the development of cases in the knowledge base of the proposed architecture. The following set of variables must be used in the description of cases: (i) Company Sales, (ii) Company Earnings, (iii) Staff Costs, (iv) Cash Flows, (v) Company Size, (vi) Company Sector, (vii) EBITDA and (viii) Company Location.

Figure 1 shows the flow of the proposed theoretical architecture which makes it possible to obtain a solution on the basis of information contained in past cases. The proposed architecture must be capable of obtaining case solutions for various companies. In this respect, it is necessary that the CBR architecture can be suitable for any company, regardless of the sector in which it develops its economic activity.

To best adapt the solutions of the architecture to the input data, the system's case memory separates the information of each company into sectors. In addition, to calculate the financial impact, the architecture must have an ANN for each sector, which must be trained with the input variables (the ones listed above).

In the Retrieve step, the system recovers the neural network which has been trained with data from companies in the same sector. Having obtained the cases for a concrete sector, data is pre-processed to eliminate the information that does not provide relevant cases for study.

In the Reuse step, the neural network will estimate the economic impact of a recommendation system on a company. Once the objective solution has been reached and the impact of using a recommendation system is known, the ANN will be trained at regular intervals automatically as the amount of input data increases through the addition of new cases to the database. To validate whether the training of the database has been successful, the ANN will train the database with 70% of its data and will use the remaining 30% to test its functionality.^{12, 11}

The learning process of the CBR architecture is performed in the Revision step. This review process must be carried out by a human in order to determine if the obtained economic impact corresponds to the company's input data (This process will be carried out according to personal criteria and experience, nevertheless different adjustments can be made). Otherwise, the case is discarded and the CBR cycle ends.

In the Retain step, regardless of whether the case had to be modified or not, if in the end it has been accepted it is stored in the Knowledge Database as a possible solution to a future problem.

Conclusions

The conclusion drawn from our research is that we still don't have a real valuation method and it is something that companies will need shortly as the amount of data being stored by companies will double. It is necessary for them, therefore, to have the tools necessary to filter the data or to be able to measure their real cost at any level.

A real database will be used to test the CBR methodology on Spanish companies that do not operate in the stock market. The financial information of the companies will be anonymized and then used in the CBR to train the system and find possible patterns that will demonstrate the economic profit that comes from having a recommender system or correct data processing in a company. In the first step, the data of Spanish companies is going to be used, however, at a later stage it would be preferable to use the information of different European companies to identify tendencies in different countries and sectors.

Acknowledgments

This work has been developed thanks to the project "Virtual-Ledgers-Tecnologías DLT/Blockchain y Cripto-IOT sobre organizaciones virtuales de agentes ligeros y su aplicación en la eficiencia en el transporte de última milla" (ID SA267P18), financed by Junta Castilla y León, Consejería de Educación and FEDER funds.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Terje Aven. A conceptual framework for linking risk and the elements of the data–information–knowledge–wisdom (dikw) hierarchy. *Reliability Engineering & System Safety*, 111:30–36, 2013.
2. Amos Azaria, Avinatan Hassidim, Sarit Kraus, Adi Eshkol, Ofer Weintraub, and Irit Netanel. Movie recommender system for profit maximization. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 121–128. ACM, 2013.
3. Aniello Castiglione, Marco Gribaudo, Mauro Iacono, and Francesco Palmieri. Exploiting mean field analysis to model performances of big data architectures. *Future Generation Computer Systems*, 37:203–211, 2014.
4. Pablo Chamoso, Alfonso González-Briones, and Francisco José García-Peñalvo. Data analysis platform for the optimization of employability in technological profiles. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 322–325. Springer, 2019.
5. Hsinchun Chen, Roger HL Chiang, and Veda C Storey. Business intelligence and analytics: From big data to big impact. *MIS quarterly*, 36(4), 2012.
6. Yuri Demchenko, Cees De Laat, and Peter Membrey. Defining architecture components of the big data ecosystem. In *2014 International Conference on Collaboration Technologies and Systems (CTS)*, pages 104–112. IEEE, 2014.
7. Daniel M Fleder and Kartik Hosanagar. Recommender systems and their impact on sales diversity. In *Proceedings of the 8th ACM conference on Electronic commerce*, pages 192–199. ACM, 2007.
8. David García-Retuerta, Álvaro Bartolomé, Pablo Chamoso, Juan M Corchado, and Alfonso González-Briones. Original content verification using hash-based video analysis. In *International Symposium on Ambient Intelligence*, pages 120–127. Springer, 2019.
9. Rashi Glazer. Measuring the value of information: The information-intensive organization. *IBM Systems Journal*, 32(1):99–110, 1993.
10. Alfonso González-Briones, Pablo Chamoso, Roberto Casado-Vara, Alberto Rivas, Sigeru Omatu, and Juan M Corchado. Internet of things platform to encourage recycling in a smart city. 2019.
11. Alfonso González-Briones, Javier Prieto, Fernando De La Prieta, Enrique Herrera-Viedma, and Juan Corchado. Energy optimization using a case-based reasoning strategy. *Sensors*, 18(3):865, 2018.
12. Alfonso González-Briones, Alberto Rivas, Pablo Chamoso, Roberto Casado-Vara, and Juan Manuel Corchado. Case-based reasoning and agent based job offer recommender system. In *The 13th International Conference on Soft Computing Models in Industrial and Environmental Applications*, pages 21–33. Springer, 2018.
13. Jo Yong Ju, Il Young Choi, Hyun Sil Moon, and Jae Kyeong Kim. Reinforcement learning for profit maximization of recommender systems. 2017.
14. Steve Lohr. The age of big data. *New York Times*, 11(2012), 2012.
15. Daniel L Moody and Peter Walsh. Measuring the value of information-an asset valuation approach. In *ECIS*, pages 496–512, 1999.
16. Paul Resnick and Hal R Varian. Recommender systems. *Communications of the ACM*, 40(3):56–59, 1997.
17. Alberto Rivas, Pablo Chamoso, Alfonso González-Briones, Roberto Casado-Vara, and Juan Manuel Corchado. Hybrid job offer recommender system in a social network. *Expert Systems*, page e12416.
18. Alberto Rivas, Jesús M Fraile, Pablo Chamoso, Alfonso González-Briones, Sara Rodríguez, and Juan M Corchado. Students performance analysis based on machine learning techniques. In *International Workshop on Learning Technology for Education in Cloud*, pages 428–438. Springer,

- 2019.
19. Alberto Rivas, Jesús M Fraile, Pablo Chamoso, Alfonso González-Briones, Inés Sittón, and Juan M Corchado. A predictive maintenance model using recurrent neural networks. In *International Workshop on Soft Computing Models in Industrial and Environmental Applications*, pages 261–270. Springer, 2019.