IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Issue-3, March 2019 E-ISSN: 2347-2693

Optimizing Analytics of Artificial Intelligence and Data Science

Mahesh Patidar[1*], V. B. Gupta[2], Seema Patidar[3]

[1]School of Data Science & Forecasting, Devi Ahilya Vishwavidyalaya, Indore, India
[2]Head School of Data Science & Forecasting, Devi Ahilya Vishwavidyalaya, Indore, India
[3]Department of Electronics & Communication, OIST, Bhopal, India

DOI: https://doi.org/10.26438/ijcse/v7i3.736740 | Available online at: www.ijcseonline.org

Accepted: 10/Mar/2019, Published: 31/Mar/2019

***Abstract*: -** Data science and machine learning are the key technologies when it comes to the processes and products with automatic learning and optimization to be used in the automotive industry of the future. This article defines the terms “data science” (also referred to as “data analytics”) and “machine learning” and how they are related. In addition, it defines the term “*optimizing analytics*” and illustrates the role of automatic optimization as a key technology in combination with data analytics. It also uses examples to explain the way that these technologies are currently being used in the automotive industry on the basis of the major sub-processes in the automotive value chain (development, procurement; logistics, production, marketing, sales and after-sales, connected customer). Since the industry is just starting to explore the broad range of potential uses for these technologies, visionary application examples are used to illustrate the revolutionary possibilities that they offer.

***Keywords*: -** Data science, big data, machine learning, automatic optimization, optimizing analytics, automotive industry

I. INTRODUTION

Data science and machine learning are now key technologies in our everyday lives, as we can see in a multitude of applications, such as voice recognition in vehicles and on cell phones, automatic facial and traffic sign recognition, as well as chess and, more recently, Go machine algorithms[1] which humans can no longer beat. The analysis of large data volumes based on search, pattern recognition, and learning algorithms provides insights into the behavior of processes, systems, nature, and ultimately people, opening the door to a world of fundamentally new possibilities. In fact, the now already implementable idea of autonomous driving is virtually a tangible reality for many drivers today with the help of lane keeping assistance and adaptive cruise control systems in the vehicle.

The fact that this is just the tip of the iceberg, even in the automotive industry, becomes readily apparent when one considers that, at the end of 2015, Toyota and Tesla's founder, Elon Musk, each announced investments amounting to one billion US dollars in artificial intelligence research and development almost at the same time. The trend towards connected, autonomous, and artificially intelligent systems that continuously learn from data and are able to make optimal decisions is advancing in ways that are simply revolutionary, not to mention fundamentally important to many industries. This includes the automotive industry, one of the key industries in Germany, in which international competitiveness will be influenced by a new factor in the near future – namely the new technical and service offerings that can be provided with the help of data science and machine learning.

II. ARTIFICIAL INTELLIGENCE

An early definition of artificial intelligence from the IEEE Neural Networks Council was “the study of how to make computers do things at which, at the moment, people are better.”[5] Although this still applies, current research is also focused on improving the way that software does things at which computers have always been better, such as analyzing large amounts of data. Data is also the basis for developing artificially intelligent software systems not only to collect information, but also to:

- Learn
- Understand and interpret information
- Behave adaptively
- Plan
- Make inferences
- Solve problems
- Think abstractly
- Understand and interpret ideas and language

2.1 Machine learning

At the most general level, machine learning (ML) algorithms can be subdivided into two categories:

supervised and unsupervised, depending on whether or not the respective algorithm requires a target variable to be specified.

1. Supervised learning algorithms

Apart from the input variables (predictors), supervised learning algorithms also require the known target values (labels) for a problem. In order to train an ML model to identify traffic signs using cameras, images of traffic signs – preferably with a variety of configurations – are required as input variables. In this case, light conditions, angles, soiling, etc. are compiled as noise or blurring in the data; nonetheless, it must be possible to recognize a traffic sign in rainy conditions with the same accuracy as when the sun is shining. The labels, i.e., the correct designations, for such data are normally assigned manually.

This correct set of input variables and their correct classification constitute a training data set. Although we only have one image per training data set in this case, we still speak of multiple input variables, since ML algorithms find relevant features in training data and learn how these features and the class assignment for the classification task indicated in the example are associated. Supervised learning is used primarily to predict numerical values (regression) and for classification purposes (predicting the appropriate class), and the corresponding data is not limited to a specific format – ML algorithms are more than capable of processing images, audio files, videos, numerical data, and text. Classification examples include object recognition (traffic signs, objects in front of a vehicle, etc.), face recognition, credit risk assessment, voice recognition, and customer churn, to name but a few.

Regression examples include determining continuous numerical values on the basis of multiple (sometimes hundreds or thousands) input variables, such as a self-driving car calculating its ideal speed on the basis of road and ambient conditions, determining a financial indicator such as gross domestic product based on a changing number of input variables (use of arable land, population education levels, industrial production, etc.), and determining potential market shares with the introduction of new models. Each of these problems is highly complex and cannot be represented by simple, linear relationships in simple equations. Or, to put it another way that more accurately represents the enormous challenge involved: the necessary expertise does not even exist.

2. Unsupervised learning algorithms

Unsupervised learning algorithms do not focus on individual target variables, but instead have the goal of characterizing a data set in general. Unsupervised ML algorithms are often used to group (cluster) data sets, i.e., to identify relationships between individual data points (that can consist of any number of attributes) and group them into clusters. In certain cases, the output from unsupervised ML algorithms can in turn be used as an input for supervised methods. Examples of unsupervised learning include forming customer groups based on their buying behavior or demographic data, or clustering time series in order to group millions of time series from sensors into groups that were previously not obvious.

In other words, machine learning is the area of artificial intelligence (AI) that enables computers to learn without being programmed explicitly. Machine learning focuses on developing programs that grow and change by themselves as soon as new data is provided. Accordingly, processes that can be represented in a flowchart are not suitable candidates for machine learning – in contrast, everything that requires dynamic and changing solution strategies and cannot be constrained to static rules is potentially suitable for solution with ML. For example, ML is used when:

- No relevant human expertise exists
- People are unable to express their expertise
- The solution changes over time
- The solution needs to be adapted to specific cases

In contrast to statistics, which follows the approach of making inferences based on samples, computer science is interested in developing efficient algorithms for solving optimization problems, as well as in developing a representation of the model for evaluating inferences. Methods frequently used for optimization in this context include so-called “evolutionary algorithms” (genetic algorithms, evolution strategies), the basic principles of which emulate natural evolution[6]. These methods are very efficient when applied to complex, nonlinear optimization problems.

Even though ML is used in certain data mining applications, and both look for patterns in data, ML and data mining are not the same thing. Instead of extracting data that people can understand, as is the case with data mining, ML methods are used by programs to improve their own understanding of the data provided. Software that implements ML methods recognizes patterns in data and can dynamically adjust the behavior based on them. If, for example, a self-driving car (or the software that interprets the visual signal from the corresponding camera) has been trained to initiate a braking maneuver if a pedestrian appears in front it, this must work with all pedestrians regardless of whether they are short, tall, fat, thin, clothed, coming from the left, coming from the right, etc. In turn, the vehicle must not brake if there is a stationary garbage bin on the side of the road.

2.2 Computer vision

Computer vision (CV) is a very wide field of research that merges scientific theories from various fields (as is often the case with AI), starting from biology, neuroscience, and psychology and extending all the way to computer science, mathematics, and physics. First, it is important to know how an image is produced physically. Before light hits sensors in a two-dimensional array, it is refracted, absorbed, scattered, or reflected, and an image is produced by measuring the intensity of the light beams through each element in the image (pixel). The three primary focuses of CV are:

- Reconstructing a scene and the point from which the scene is observed based on an image, an image sequence, or a video.
- Emulating biological visual perception in order to better understand which physical and biological processes are involved, how the wetware works, and how the corresponding interpretation and understanding work.
- Technical research and development focuses on efficient, algorithmic solutions – when it comes to CV software, problem-specific solutions that only have limited commonalities with the visual perception of biological organisms are often developed.

All three areas overlap and influence each other. If, for example, the focus in an application is on obstacle recognition in order to initiate an automated braking maneuver in the event of a pedestrian appearing in front of the vehicle, the most important thing is to identify the pedestrian as an obstacle. Interpreting the entire scene – e.g., understanding that the vehicle is moving towards a family having a picnic in a field – is not necessary in this case.

III. DATA MINING AND ARTIFICIAL INTELLIGENCE

At a high level of abstraction, the value chain in the automotive industry can broadly be described with the following sub-processes:

1. Development
2. Procurement
3. Logistics
4. Production
5. Marketing
6. Sales, after-sales, and retail
7. Connected customer

Each of these areas already features a significant level of complexity, so the following description of data mining and artificial intelligence applications has necessarily been restricted to an overview.

Development

Vehicle development has become a largely virtual process that is now the accepted state of the art for all manufacturers. CAD models and simulations (typically of physical processes, such as mechanics, flow, acoustics, vibration, etc., on the basis of finite element models) are used extensively in all stages of the development process.

The subject of optimization (often with the use of evolution strategies [31] or genetic algorithms and related methods) is usually less well covered, even though it is precisely here in the development process that it can frequently yield impressive results.

Procurement

The procurement process uses a wide variety of data concerning suppliers, purchase prices, discounts, delivery reliability, hourly rates, raw material specifications, and other variables. Consequently, computing KPIs for the purpose of evaluating and ranking suppliers poses no problem whatsoever today. Data mining methods allow the available data to be used, for example, to generate forecasts, to identify important supplier characteristics with the greatest impact on performance criteria, or to predict delivery reliability. In terms of *optimizing analytics*, the specific parameters that an automotive manufacturer can influence in order to achieve optimum conditions are also important.

Logistics

In the field of logistics, a distinction can be made between procurement logistics, production logistics, distribution logistics, and spare parts logistics.

Procurement logistics considers the process chain extending from the purchasing of goods through to shipment of the material to the receiving warehouse. When it comes to the purchasing of goods, a large amount of historical price information is available for data mining purposes, which can be used to generate price forecasts and, in combination with delivery reliability data, to analyze supplier performance. As for shipment, *optimizing analytics* can be used to identify and optimize the key cost factors.

A similar situation applies to production logistics, which deals with planning, controlling, and monitoring internal transportation, handling, and storage processes. Depending on the granularity of the available data, it is possible to identify bottlenecks, optimize stock levels, and minimize the time required, for example here.

Production

Every sub-step of the production process will benefit from the consistent use of data mining. It is therefore essential for all manufacturing process parameters to be continuously recorded and stored. Since the main goal of optimization is usually to improve quality or reduce the incidence of defects, data concerning the defects that occur and the type of defect is required, and it must be possible to clearly assign this data to the process parameters. This approach can be used to achieve significant improvements, particularly in new types of production process – one example being CFRP[36]. Other potential optimization areas include energy consumption and the throughput of a production process per time unit. *Optimizing analytics* can be applied both offline and online in this context.

When used in offline applications, the analysis identifies variables that have a significant influence on the process. Furthermore, correlations are derived between these influencing variables and their targets (quality, etc.) and, if applicable, actions are also derived from this, which can improve the targets. Frequently, such analyses focus on a specific problem or an urgent issue with the process and can deliver a solution very efficiently – however, they are not geared towards continuous process optimization. Conducting the analyses and interpreting and implementing the results consistently requires manual sub-steps that can be carried out by data scientists or statisticians – usually in consultation with the respective process experts.

Marketing

The focus in marketing is to reach the end customer as efficiently as possible and to convince people either to become customers of the company or to remain customers. The success of marketing activities can be measured in sales figures, whereby it is important to differentiate marketing effects from other effects, such as the general financial situation of customers. Measuring the success of marketing activities can therefore be a complex endeavor, since multivariate influencing factors can be involved.

It would also be ideal if *optimizing analytics* could always be used in marketing, because optimization goals, such as maximizing return business from a marketing activity, maximizing sales figures while minimizing the budget employed, optimizing the marketing mix, and optimizing the order in which things are done, are all vital concerns. Forecast models, such as those for predicting additional sales figures over time as a result of a specific marketing campaign, are only one part of the required data mining results – multi-criteria decision-making support also plays a decisive role in this context.

Sales, after-sales, and retail

The diversity of potential applications and existing applications in this area is significant. Since the “human factor,” embodied by the end customer, plays a crucial role within this context, it is not only necessary to take into account objective data such as sales figures, individual price discounts, and dealer campaigns; subjective customer data such as customer satisfaction analyses based on surveys or third-party market studies covering such subjects as brand image, breakdown rates, brand loyalty, and many others may also be required. At the same time, it is often necessary to procure and integrate a variety of data sources, make them accessible for analysis, and finally analyze them correctly in terms of the potential subjectivity of the evaluations[37] – a process that currently depends to a large extent on the expertise of the data scientists conducting the analysis.

Connected customer

While this term is not yet established as such at present, it does describe a future in which both the customer and their vehicle are fully integrated with state-of-the-art information technology. This aspect is closely linked to marketing and sales issues, such as customer loyalty, personalized user interfaces, vehicle behavior in general, and other visionary aspects (see also section 5). With a connection to the Internet and by using intelligent algorithms, a vehicle can react to spoken commands and search for answers that, for example, can communicate directly with the navigation system and change the destination. Communication between vehicles makes it possible to collect and exchange information on road and traffic conditions, which is much more precise and up-to-date than that which can be obtained via centralized systems. One example is the formation of black ice, which is often very localized and temporary, and which can be detected and communicated in the form of a warning to other vehicles very easily today.

IV. CONCLUSION

Artificial intelligence has already found its way into our daily lives, and is no longer solely the subject of science fiction novels. At present, AI is used primarily in the following areas:

- Analytical data processing
- Domains in which qualified decisions need to be made quickly on the basis of a large amount of (often heterogeneous) data
- Monotonous activities that still require constant alertness

In the field of analytical data processing, the next few years will see us transition from exclusive use of decision-support systems to additional use of systems that make decisions on our behalf. Particularly in the field of data analysis, we are currently developing individual analytical solutions for specific problems, although these solutions cannot be used across different contexts – for example, a solution developed to detect anomalies in stock price movements cannot be used to understand the contents of images. This will remain the case in the future, although AI systems will integrate individual interacting components and consequently be able to take care of increasingly complex tasks that are currently reserved exclusively for humans – a clear trend that we can already observe today.

REFERENCES

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five Disruptive Technology Directions for 5G,” *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 74-80, Feb. 2014.

[2] M. Paolini, “Mastering Analytics: How to Benefit From Big Data and Network Complexity,” [online]. http://content.rcrwireless.com/20170620 Mastering Analytics Report.

[3] S. Bi, R. Zhang, Z. Ding, and S. Cui, “Wireless Communications in the Era of Big Data,” *IEEE Commun. Mag.*, vol. 53, no. 10, pp. 190-199, Oct. 2015.

[4] 3GPP TR 23.793, “Study on Access Traffic Steering, Switching and Splitting support in the 5G system architecture,” V0.1.0 Aug. 2017.

[5] S. Han, C.-L I, G. Li, and S. Wang, “Big Data Enabled Mobile Network Design for 5G and Beyond,” *IEEE Commun. Mag.*, vol. 55, no. 9, pp. 150- 157, Sep. 2017.

[6] X. Cheng, L. Fang, L. Yang, and Shuguang Cui, “Mobile Big Data: The Fuel for Data-Driven Wireless,” *IEEE Intenet Things J.*, vol. 4, no. 5, pp. 1489-1516, Oct. 2017.

[7] X. Cheng, L. Fang, X. Hong, and L. Yang, “Exploiting mobile big data: Sources, Features, and Application,” *IEEE Netw.*, vol. 31, no. 1, pp. 72- 79, Jan./Feb. 2017.

[8] A. Engelbrecht, “Computational Intelligence: An Introduction,” 2nd ed. NY, USA: John Wiley & Sons, 2007.

[9] O. Acker, A. Blockus, and F. Potscher, “Benefiting From Big Data: A New Approach for the Telecom Industry,” [online]. https://www.strategy nd.pwc.com/reports/benefiting-big-data.

[10] C. Jiang, H. Zhang, Y. Ren, Z. Han, K.-C. Chen, and L. Hanzo, “Machine Learning Paradigms for Next-Generation Wireless Networks,” *IEEE Wireless Commun. Mag.*, vol. 24, no. 2, pp. 98-105, Apr. 2017.

[11] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, “Machine Learn- ing for Wireless Networks with Artificial Intelligence: A Tutorial on Neu- ral Networks,” [Online]. Available: https://arxiv.org/pdf/1710.02913.pdf, accessed on Feb 1, 2018.

[12] S. A. Kyriazakos and G. T. Karetsos “Practical Radio Resource Man- agement in Wireless Systems,” Boston, USA: Artech House. 2004.

[13] X. Lu, E. Wetter, N. Bharti, A. J. Tatem, and L. Bengtsson, “Approaching the Limit of Predictability in Human Mobility,” *Sci. Rep.*, vol. 3, Nov. 2013, p. 324.

[14] R. Atawia, H. S. Hassanein, and A. Noureldin “Fair Robust Predictive Resource Allocation for Video Streaming under Rate Uncertainties,” in *Proc. IEEE Globecom*, pp. 1-6, Dec. 2016.

[15] E. Oh, B. Krishnamachari, X. Liu, and Z. Niu, “Toward Dynamic Energy-Efficient Operation of Cellular Network Infrastructure,” *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 56-61, Jun. 2011.

[16] A. Banerjee, “Advanced Predictive Network Analytics: Optimize Your Network Investments & Transform Customer Experience,” Heavy Read- ing, Feb. 2014.

[17] W. Hong *et al.*, “Multibeam Antenna Technologies for 5G Wireless Communications,” *IEEE Trans. Ant. Prop.*, vol. 65, no. 12, Dec. 2017.

[18] E. J. Black *et al.*, “Broadband Surface Scattering Antennas,” *US2017/0187123A1*, Jun. 2017.

