Chapter 12
Big Data Analytics in the Maintenance of Off-Shore Wind Turbines: A Study on Data Characteristics

Elaheh Gholamzadeh Nabati and Klaus Dieter Thoben

Abstract The aim of this study is to discuss the characteristics of the input data to the data analytical algorithms of a predictive maintenance system, from the viewpoint of big data technology. The discussed application is for the maintenance of off-shore wind turbines. The maintenance of off-shore wind turbines is an expensive and sensitive task. Therefore, making decision for planning and scheduling of maintenance in a wind farm (which is made by the operating company of a wind farm) is important and plays a critical role in the cost of maintenance. In this paper, the current state of the art for big data technology in the maintenance of off-shore wind turbines is presented. The dimensions of big data analytics and the technical requirements of data for the use of this technology in the maintenance of off-shore wind turbines are described. A contribution of this paper is to study the technical requirements of suitable data for decision-making. The outcomes of presented study are identifying the characteristics of input data to predictive maintenance in the era of big data and the discussion of these characteristics in the condition monitoring for off-shore wind energy.

Keywords Predictive maintenance · Big data · Off-shore wind energy · Data analytics

E.G. Nabati (✉)
International Graduate School of Dynamics in Logistics, University of Bremen, Bremen, Germany
e-mail: nab@biba.uni-bremen.de

K.D. Thoben
BIBA- Bremer Institut für Produktion und Logistik GmbH, Bremen, Germany
e-mail: tho@biba.uni-bremen.de

12.1 Introduction

Maintenance process is a critical element of system's operation. The maintenance of off-shore wind turbines is usually done by the operating companies. The operating company is responsible for operating a wind farm and making the decision on the service in order to prevent the equipment stopping from producing the electricity. In the first five years of the operation, the maintenance and service tasks are done in cooperation with the wind turbine manufacturer.

The maintenance strategies of the off-shore wind mills which the operating companies perform have three categories: the corrective maintenance, preventive maintenance and the predictive maintenance. In the corrective maintenance the turbine runs until the failure. Failure causes the system to stop. At this time, the fault diagnosis is done and then it is decided whether to repair or replace (Puig 2011). The preventive maintenance is the interval visits of the wind turbines and service all the parts based on fixed time intervals. This type of maintenance usually requires the part to be changed before they reach the end of their life time. Therefore, this maintenance strategy is costly. The third strategy for maintenance is the predictive maintenance where the status of system is monitored with the condition monitoring system and the parts are replaced when the monitoring system shows, there is need for replacement. In other words as stated by (Karyotakis 2011) the equipment are *monitored for their condition and statistical reliability tools are used to preventively maintain or exchange critical items before any failure occurs.*

The off-shore wind equipment has several characteristics which calls for more control and monitoring of the equipment from remote. First, the wind farms are normally located far from the land and the accessibility to these equipments is only with ship and helicopter possible. The dispatching of ship or helicopter depends on the weather condition, whether if the sea is calm and safe enough to travel. Second, the wind turbines are exposed to moisture, salt and sever marine condition as well as constant vibration because of the sea waves. These factors cause the early erosion and consequently need for more service in comparison of on the land turbines.

Therefore the maintenance strategies such as predictive maintenance are the most preferable solution among the above maintenance strategies for the off-shore. It enables the control from the remote and if properly implemented causes the least expenses during the services.

From one hand, the predictive maintenance uses the data sources such as sensor measurements and health and environmental data for performing the predictions. From the other hand, currently the advances in the technology, caused a big data to appear, new types of sensors, new measurement methods are being generated. These new data sources cause challenges and need for modifications in the current predictive maintenance and reliability estimation approaches. It is necessary to clarify the characteristics of new data to enable faster adoption of maintenance system to the new situation. Moreover, to enable innovation in the data modelling approaches for the maintenance. Although recently several papers are published regarding the issue of big data dimensions in maintenance (Meeker and Hong 2014;

Göb 2014), still more work and applications are needed. In this paper the characteristics of input data to the data analytical algorithms of a predictive maintenance system is discussed from the viewpoint of big data technology. The application is for the maintenance of off-shore wind turbines.

This paper is organized as follows. Section 12.2 introduces the big data characteristics and the big data technology. Section 12.3 describes the relation of predictive maintenance with the big data analytics. Section 12.4 provides a review of literature on big data technology in the maintenance of wind energy. Section 12.5 presents the research methodology. Section 12.6 discusses the technical requirements of maintenance data to be used in the big data system. Section 12.6 is dedicated to the conclusion.

12.2 Big Data Technology and the Place of Data Analytics

Big data is a term which refers to the data which has a huge size, high speed of production and comes from several sources with different formats. In the context of decision-making, big data is considered as a new source of competitive advantage which has the potential to derive new insight from it, and can help in making more realistic decisions. Form the technical point of view, the definition of big data is the data which is more than the size of current relational data bases and so complex in format and dimension, which is hard or impossible to use and apply traditional analytical techniques on it. (Fosso Wamba et al. 2015) provided the different definitions of big data in the literature. The basic definition contains three characteristics (3 V).Volume, which refers to the size of data. Variety, that refers to the several data sources and data formats and Velocity, which is the increasing speed of data generation and transmission. Some authors add more dimensions to the 3 V. For example (Opresnik and Taisch 2015; White 2012), added the two dimensions of Veracity and Value to the big data. Veracity means the accuracy of the data and Value describes the insight which these information can provide for us. These authors applied the concept of big data on a certain business oriented domain such as IT system, product service system, business and organizational data sources. Therefore, taking into account the business value of the big data. From another perspective, (Göb 2014) named the fourth dimension of big data as the complexity. Indeed he looked at the big data problem from a more technical view. Complexity describes the data management difficulties, statistical analysis and algorithmic complexity aspects of big data (Göb 2014).

Big data analytics is a part of big data technology, where the analysis of data is performed. Data analysis basically is an interdisciplinary field between the sciences of probability theory, statistics, machine learning and data mining. From the application purpose, data analytics can be used for describing the characteristics of a certain entity such as the mean value of a measurement, exploring the correlation
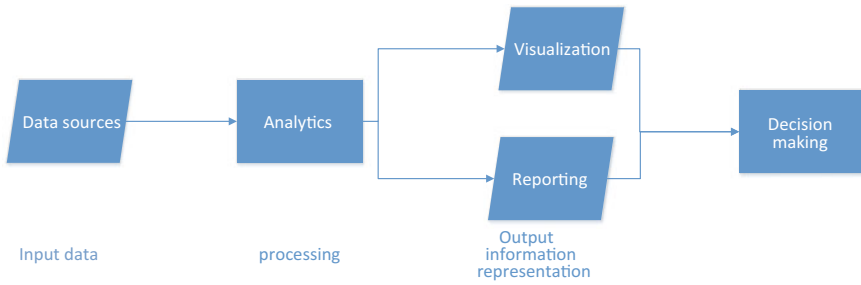
Fig. 12.1 Simple systemic view of data analytics

and relations between different variables, e.g. the relationship between the increase in the speed and increase of temperature, monitoring the current situation (diagnosis), e.g. fault detection, predicting the future condition (prognosis), e.g. predicting the life of a part. A good representation of different approaches of data analysis is illustrated in the research done by (Freitag et al. 2015).

In the next section the location of data analytics is discussed in the predictive maintenance.

Figure 12.1 shows a simple systematic view to the data analytics. First, different data from the different data sources are being preprocessed. Importantly, the missing values and errors are handled. Afterwards, the data analytical process is being done. It usually contains the feature selection, model fitting or the algorithm application and assessing the appropriateness of the output information. In the next step, the gained information are represented to the decision-maker for further interpretation and use in the decision-making process.

12.3 Data Analytics and Predictive Maintenance

Predictive maintenance uses a part of data analytical techniques. These techniques are diagnosis and prognosis models. With them, it is possible to monitor the condition of the products (Puig 2011). Therefore, the analytical module of the predictive maintenance is called Condition Monitoring System (CMS). It is always beneficial to distinguish between prognosis and diagnosis. While diagnosis is used to monitor and finding the faults, abnormalities and control if the process is functioning within certain standards, prognosis is done to understand the problem (fault) before it has occurred. The example of prognosing is estimating the remaining life of the parts in a wind turbine. Several algorithms exist for prognosing and diagnosing. Good references about the classification of these algorithms in the predictive maintenance and wind industry are available at (Hameed et al. 2009; García Márquez et al. 2012; Takoutsing et al. 2014; Sheng 2015; Lau et al. 2012).

12.4 Big Data Analytics in Wind Maintenance: Review of the Literature

Wind energy has used the data analytics from long ago in the maintenance practices for analyzing the system health monitoring data (SHM) and the environmental data (Meeker and Hong 2014). Nevertheless, still no comprehensive work about the application of big data technology in the off-shore wind energy maintenance exists. Only, some authors addressed different parts of this technology. In this section an overview of existing literature is presented.

By reviewing the literature on the applications of big data for the maintenance of wind energy, four main categories are recognized:

1. **SHM and environmental data Storage**: (Viharos et al. 2013) studied collecting and storing large volumes of data from the wind park site and store it for further fault prognosis. The study aims at providing guidelines for practitioners. This research groups the challenges of big data storage and processing as the following: "

   - *The volume of the data puts very high load on both the local industrial computers and the data centers.*
   - *End user applications must balance between flexibility of querying and quick query response times as distributed systems pose limitations to certain elements of SQL including the join operation*".

Markovic et al. (2013) studied the cloud computing for real-time high volume data capturing and analysis. Nguyen et al. (2014) discussed the data management and metadata generation aspects.

2. **Data processing and analysis**: The current literature can be divided to two subgroups, first prediction and reliability analysis and second system health monitoring and controlling. For the first subgroup, (Xiang et al. 2009) studied the utilizing the sensor data for fault prediction. They applied fusion technique which is a data-driven approach to take advantage of the available measurements without installing extra sensors explicitly for fault diagnosis. Xiang et al. (2009 and Spahić et al. (2009) presented a reliability model for several wind mills in a large off-shore wind farm. They computed the reliability model with high volume of data and also they included new data characteristic such as wind farm location, single generator power and power of the entire wind farm, wind farm grid type (radial/meshed), switchgear and protection type, and automation technology type. They applied statistical methods on the data.

   For the second subgroup which is monitoring and controlling applications, Lau et al. (2012) provided a review of failure diagnosis types. Wan (2004) analyzed the long time data to see the behaviour of wind mills at the farm.

3. **Smart grinds**: The management and efficiency of produced electricity in the era of big data has been addressed by some authors. SunGard solutions (2013) named the advantages of big data for energy trading. IBM (2012) discussed the

characteristics and relevant system architectures for big data in smart grinds (IBM 2012). Diamantoulakis et al. (2015) addressed the energy management in smart grids with the big data analytics.

4. **Social media and unstructured data analysis**: Anninni (2014) reported the use of new sources of unstructured data in the wind energy. For example, the use of social media about the opinions of public towards the wind energy. He tested the sentiment analysis on the opinions of Vastas employee towards the finical benefits of big data business intelligent platform in this wind energy company (Anninni 2014).

12.5 Research Methodology

The models and algorithms described in the previous section need the input data to produce the output information, upon which the decision-makers can make decisions. In this section with a systemic view, we look at the characteristics of input data to the data analysis module of predictive maintenance.

Through a study on different characteristics of input data for the big data analytics, the following main categories were found.

Taking a closer look at these dimensions, several common characteristics are recognized. Considering their relevancy to the predictive maintenance these characteristics are to be seen: Volume, Variety, Veracity (quality structure), Complexity and Velocity.

In the next section, these dimensions of data will be discussed in the context of off-shore wind energy maintenance.

Some of the dimensioned mentioned by the authors in Table 12.1 are not relevant to the input data to a data analysis system. Two of them are "Sensitivity" and "Value" (Table 12.1). Géczy (2014) described sensitivity as one of the characteristics of big data. When we distinguish between the input data to and the output information from an analytical system, these characteristics belong to the output information and they are not relevant for the input data. As stated by Geczy,

Table 12.1 Different characteristics of big data

Big data characteristics	Reference
3 V: Volume, Variety, Velocity	(Wamba et al. 2015; Gandomi and Haider 2015)
4 V: Volume, Variety, Velocity, Value	(Wamba et al. 2015)
5 V: Volume, Variety, Velocity, Value, Veracity	(Opresnik and Taisch 2015; White 2012; Wamba et al. 2015)
Interdisciplinary characteristics framework: Sensitivity, Diversity (Variety), Quality, Structure, Volume, Speed (Velocity)	(Géczy 2014)

sensitivity means the information which contains know-how and the personal or confidential information. The raw data such as sensors and machine logs does not reveal any of such qualities before they are processed.

The same is true for the "Value". Value is the knowledge or insight which is being extracted. From the raw data and before analyzing the data, it is not possible to get the value.

12.6 Technical Characteristics of Input Data for Predictive Maintenance in the Era of Big Data

This section describes the technical characteristics/requirements of input data to the predictive maintenance in the era of big data. These characteristics as stated in Sect. 12.2 are volume, variety and veracity and complexity. We discuss these four main characteristics in the field of off-shore maintenance.

Volumn

The data are growing in the wind energy. The sensor data are growing in number and frequency of information broad casting. For a wind farm, we have the data of around 100 turbines, which each turbine has around 20–30 sensors installed on it (Viharos et al. 2013). So, the amount of information is huge when all needs to be processed. During the maintenance process usually not one but several turbines are being serviced together. Therefore, having the data of the status of all those turbines and making the maintenance in a coordinated way based on the real condition of turbines would be more effective. This needs processing the data of several turbines together. In this case, if three turbines are being maintained together and the results from the analysis shows the probability of failure in one turbine is higher than the rest two it is more practical to start the maintenance process not randomly but from that turbine which has the higher probability of failure.

The other source of big data (increase in data volume) comes from digitalization of the business processes. Traditionally the data on the paper such as maintenance reports from the practitioners, fault descriptions, were not included in the condition monitoring system. Currently by the advances in automation; emergence of mobile devices and smart tags, it is possible to integrate more of these data sources. For example, now the maintenance check lists can be filled by the technicians on the web-based applications installed on the tablets. Therefore, it facilitates the use of them in the analytics (Liebstückel 2012).

Velocity

Velocity of data generation shows the speed in which the data are produced and transmitted. The sensor measurements from the wind farms are broad casted every few minutes. The new sensor technologies such as wireless sensor and acoustic sensors are increasing this pace of data exchange event more. Improved speed of data availability calls for real-time analytics and data aggregation which in a large

scale is a challenge. That is because, the transmitted data should be checked for the data quality quicker than before and the feature selection and the analytics should be done near real time. It means the predictive maintenance system should respond faster to the input data. As an example, feature selection for fault prognostic in gearbox uses the vibration sensor data. The axel gear vibration is monitored for this aim. The selection of important features which signal the fault can be recognized by the neural network algorithm (Lau et al. 2012). Incorporating the techniques such as sliding window to deal with the input data pace can help to convert this application to a real-time monitoring.

Variety

The traditional data for the predictive maintenance and reliability engineering are in the form of failure data (status code logs), service and maintenance activity list, and system health management logs. The latter is a combination of sensor measurement and the records of wind parameters such as stress and acceleration (Sheng 2015). Currently the SHM sources are expanding. For example in the wind energy not only the speed of the wind and temperature of environment are being measured and used in the analytics but also the weather condition such as the see state, the wave heights based on the season of the year, the salt in the water are being taken into account.

Audio, video multimedia data and unstructured text data are other new sources of data. For example the non-structured data of critical failures which is provided by the wind turbine technicians, in form of event annotations or the controllers at the control room are also very promising sources of information. They can provide many good insights to the real conditions which caused the failure to happen.

Veracity

Veracity refers to the correct structure and good quality of input data. The measured data should be of high quality. It means the missing values, the errors in the measures should be as less as possible. The format of input data to the algorithm should be correct. For example if the algorithm models the integer values the use of numeric values (with fraction) causes the error in the prediction results. One good advantage of big data sources as stated by (Meeker and Hong 2014) is, classically the estimation of life time of a part usually is done based on the data from laboratory tests during the product development. Those tests used the historical data of similar products (the failures of previous products). But today with the installation of sensors and chips on the products it is possible to increase the veracity.

Complexity

The modern filed data of wind energy, such as the sensor measures and health and environmental data, are captured more in form of vector (a batch of data at every measurement) rather than a single value. This property makes it possible to find more relations between the different variable and learn how a change in one parameter affects the other, also how change in both parameter together signal the degradation and abnormality in the function of a part.

An example of complex aspect of big data is "*sensor data collected in short time intervals are inevitably serially correlated*" (Göb 2014). So, this data cannot be used directly in the algorithms which use the assumption that the input data are normally distributed. As most of analytical algorithms has the assumption of normality, dealing with the autocorrelated data increases the complexity of analysis.

The other opportunity is we can monitor the degradation of the part over time rather than just wait/work with the failure data. The degradation shows the decrease in the performance of the part and with that we can recognize a failure before it happens. For example, performance degradation of bearing, or the colour coating (Meeker and Hong 2014).

12.7   Conclusion

This paper reviewed the current literature on the big data analytics technology in the predictive maintenance, with the focus on the off-shore wind energy maintenance. It is possible to control a few wind turbines from remote with the current data analytical techniques (currently available prognosis and diagnosis models) but for a wind farm with several turbines and sensors, it is hardly possible to do the efficient analysis of the data, either the offline analysis for prognosis of next failures or the online analysis for monitoring the health condition of the equipment without considering the big data technology. However, there is a need for more research which clarifies the characteristics of big data in the maintenance of wind power. This can enhance the development of new maintenance strategies. Further research in this area can be discussing the requirements of data analytics algorithms as well as the output information, also the other aspects of big data technology such as storage, parallel processing in the condition monitoring off-shore wind turbines.

References

Anninni A (2014) Big data in the wind power industry. Master thesis, Grenoble Graduate School of Business

Diamantoulakis PD, Kapinas VM, Karagiannidis GK (2015) Big data analytics for dynamic energy management in smart grids. Big Data Anal High-Perform Comput 2(3):94–101

Wamba FS, Akter S, Edwards A, Chopin G, Gnanzou D (2015) How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. Int J Prod Econ (in press)

Freitag M, Kück M, Ait Alla A, Lütjen M (2015) Potentiazale von Data Science in Produktion und Logistik. Industrie 4.0:22–26

Gandomi A, Haider M (2015) Beyond the hype: big data concepts, methods, and analytics. Int J Inf Manag 35:137–144

García Márquez FP, Tobias AM, Pinar Pérez J, Papaelias M (2012) Condition monitoring of wind turbines: techniques and methods. Renew Energy 46:169–178

Géczy P (2014) Big data characteristics. Macrotheme Rev 3(6):94–105

Göb R (2014) Discussion of reliability meets big data: opportunities. Qual Eng 26(1):121–126

Hameed Z, Hong YS, Cho YM, Ahn SH, Song CK (2009) Condition monitoring and fault detection of wind turbines and related algorithms: a review. Renew Sustain Energy Rev 13 (1):1–39

IBM (2012) Managing big data for smart grids and smart meters

Karyotakis A (2011) On the optimisation of operation and maintenance strategies for offshore wind farms. Ph.D. dissertation. University Collage London

Lau BCP, Ma EWM, Pecht M (2012) Review of offshore wind turbine failures and fault prognostic methods. In: IEEE prognostics & system health management conference, pp 1–5

Liebstückel K (2012) Plant maintenance with SAP contents at a glance. Galileo Press, Bonn, Boston

Markovic D, Zivkovic D, Branovic I, Popovic R, Cvetkovic D (2013) Smart power grid and cloud computing. Renew Sustain Energy Rev 24:566–577

Meeker W, Hong Y (2014) Reliability meets big data: opportunities and challenges. Qual Eng 26:102–116

Nguyen TH, Nunavath V, Prinz A (2014) big data metadata management in smart grids. Stud Comput Intell 546:189–214

Opresnik D, Taisch M (2015) The value of big data in servitization. Int J Prod Econ 165:174–184

Puig J (2011) Asset optimization and predictive maintenance in discrete manufacturing industry. Master thesis, École Polytechnique Fédérale De Lausanne, Switzerland

Sheng S (2015) Improving component reliability through performance and condition monitoring data analysis. National Revenewable Energy Labratory (NREL), Houston, Texas

Spahić E, Underbrink A, Buchert V, Hanson J, Jeromin I, Balzer G (2009) Reliability model of large offshore wind farms. In: Proceedings of the IEEE Power Tech conference, pp 2–7

SunGard Solutions (2013) White paper big data challenges and opportunities for the energy industry

Takoutsing P, Wamkeue R, Ouhrouche M, Slaoui-Hasnaoui F, Tameghe T, Ekemb G (2014) Wind turbine condition monitoring: state-of-the-art review, new trends, and future challenges. Energies 7(4):2595–2630

Viharos ZJ, Sidló CI, Benczúr A et al (2013) "Big data" initiative as an IT solution for improved operation and maintenance of wind turbines. In: EWEA, Austria, pp 184–188

Wan Y (2004) Wind power plant behaviors: analyses of long-term wind power data. NREL

White M (2012) Digital workplaces, vision and reality. Int J Bus Rev 29(4):205–214

Xiang Y, Veeramachaneni K, Yanjun Y, Osadciw L (2009) Unsupervised learning and fusion for failure detection in wind turbines. In: 12th international conference on information fusion, Seatel, USA, pp 1497–1503