REVIEW OF DATA-DRIVEN PROGNOSTICS AND HEALTH MANAGEMENT TECHNIQUES: LESSIONS LEARNED FROM PHM DATA CHALLENGE COMPETITIONS

Bin Huang, Yuan Di, Chao Jin and Jay Lee
NSF I/UCRC for Intelligent Maintenance Systems (IMS)
University of Cincinnati
560 Baldwin Hall
University of Cincinnati
Cincinnati, OH 45221
Telephone: (513) 556-4650
huangba@mail.uc.edu

Abstract: Machine learning and statistical algorithms are receiving considerable attention during the past decade in prognostics and health management (PHM). However, there is a lack of consensus and methodology on algorithm selection in different scenarios, which renders the random implementation of machine learning algorithms and inefficient development processes. PHM Data Challenge, an open data competition specialized in PHM, includes diverse issues in industrial data analytics and thus provides abundant resource for study and appropriate approach development. In this work, we first summarize the problems and datasets of PHM Data Challenge competitions. According to their objectives, the 9 problems can be classified into 3 categories, health assessment, fault classification and remaining useful life prediction. Second, common issues and unique challenges have been clearly pointed out for each problem and each category. Then, we analyze all solutions regarding what type of strategy a particular solution took, what algorithms it used and how it overcame the challenges. At last, insights in PHM solution strategies have been summarized to conclude the paper.

Key words: Diagnostics; Prognostics; Health Assessment; Data Challenge Competition; Fault Classification; Remaining Useful Life Prediction

1. Introduction: Prognostics and Health Management (PHM) is an interdisciplinary research area that aims at increasing productivity by managing operations based on assessment and prognosis of potential engineering system faults. The advancement of Industrial Internet of Things [1] and industrial big data analytics [2] has spurred huge promise of economic values through data-driven predictive modeling [3, 4] in a variety of industries. Due to its scalability, flexibility, and rapidity of deployment, data-driven PHM has clearly become a necessity for the next industrial revolution.

The PHM Society has been promoting research at the frontier of PHM by organizing the PHM Data Challenge competition nearly every year since 2008 [5]. The data in the competitions cover a wide spectrum of real-world industrial problems, from rotary machinery fault diagnosis [6] to jet engine remaining useful life prediction [7], and from time series based fault detection [8] to event-based failure prediction [9]. The proposed issues and winning algorithms each year serve as a diverse library of case studies from which we can learn about the current challenges in practice, the thinking flow of addressing these challenges, and the advantages and disadvantages of different methods.

This paper attempts to find the commonalities and insights of applying machine learning algorithms for PHM solutions based on the insights learned from the competitions. The resulting conclusions will serve as the foundation for the development of a systematic methodology in data-driven modeling and contributing to the advancement of the PHM research.

1.1. Review of Related Literature: With the rapid advancement in prognostics, quite a few research efforts provide a survey for remaining useful life (RUL) prediction as one of the most important tasks in prognostics. In [10], the author categorized the algorithms into physics-based methods, experienced-based methods and data-driven methods. In the data-driven section, four typical types of machine learning algorithms, including neural network, support vector machine, Bayesian network and hidden Markov model, were reviewed while other feature extraction techniques combined with statistical and machine learning algorithms were considered as hybrid methods. In [11], the author reviewed the data-driven methods and summarized them into two types based on whether it is possible to identify an indicator from the original raw data, direct condition monitoring approach and indirect condition monitoring data approach. Regression-based, Wiener process, Gamma process and Markovian-based methods can be used when an indicator is identified; stochastic filtering and covariate based hazard model and hidden Markov model are applicable if an indicator cannot be identified. The paper [12] reviewed the algorithms in both diagnosis and prognosis. It is noted that the author discussed prognostics based on different tasks. Other than that, it also reviewed the commonly used techniques in other steps during implementation, including data acquisition, preprocessing, feature extraction and selection and a special case when inconsistent types of data need to be processed. A systematic methodology on PHM design for rotary machine systems with a comprehensive review of algorithms was proposed in [13]. Although this paper proposed an algorithm selection method according to customer requirements and application conditions by quality function deployment (QFD) tables, its performance is largely limited by prior knowledge and subjective judgment. As we see from the above existing PHM reviews, most papers focus on discussing algorithms in terms of capability and functionality, but they rarely consider the application of these algorithms in real industrial situations. Hence this review paper will try to discuss the methods and algorithms in an application and problem-oriented fashion.

1.2. Motivation: In the past 9 years, several key issues that reflect the challenges in industrial data analytics in different industrial sectors have been generalized and released as PHM Data Challenge competitions organized by PHM Society and IEEE Reliability Society (in 2012 and 2014). The purpose of PHM Data Challenge is to gain more attention and efforts from academics and industry to address the real-world challenges. The topics of the competitions covers a wide range of areas, including aircraft engine, fuel cells, and gearbox. The topics also cover almost all the PHM tasks, which include remaining useful life (RUL) prediction, diagnosis, fault detection, health assessment, and condition monitoring. The competitions provide abundant open resource for study and research such that further improvement and extended research can be advanced. Many research works have been published based on popular datasets provided by PHM Data Challenge competitions, such as jet engine RUL prediction in 2008 [14] and gearbox diagnosis in 2009 [15, 16]. Therefore, reviewing how algorithms have been used and methods have been developed based on PHM Data Challenge datasets will surely benefit and facilitate the data-driven PHM methodology development in the future.

1.3. Objective: This study aims at providing a generalized problem-oriented PHM technique review by summarizing 9 PHM Data Challenge competitions and around thirty papers published as the winning solutions to these competitions.

The basic ideas of this review consist of three steps: 1) summarize the common issues and analytical challenges from the data competitions; 2) compare and benchmark the relevant solutions; 3) discuss the lessons learned from these competitions and solutions. By following these steps, the objective is to understand the reason why the winning algorithm outperforms others and learn the rational strategy of algorithm selection in different scenarios.

1.4. Organization: Section 2 gives an overview of Data Challenge competition, and introduces the basic information about each year's problem. Then the details of three categories of problems, health assessment, fault detection and classification and RUL prediction, are analyzed individually in Section 3, Section 4 and Section 5, respectively. Section 6 concludes the findings in each type of problem and other common issues over all data challenge problems.

2. Overview: In general, an effective PHM system is expected to have the capabilities of providing early detection and isolation of incipient faults and monitoring and predicting the progression of faults, which will further support maintenance scheduling and asset management. Such expectation is a mixed requirement of diagnostics, prognostics, health assessment and health management. Conventional diagnostics is known as a fail-and-fix process to determine the root cause of machine failure. On the contrary, prognostics is a predict-and-prevent methodology aiming to reduce downtime and maintenance cost proactively. Health assessment is mostly regarded as a critical process for detecting incipient failures before an algorithm-centered prediction process. The last step in a complete PHM system is maintenance scheduling and operation management based on outputs from prognostic modeling process. It focuses on assessing the further impact of failures and minimizing the loss through optimization.

In PHM Data Challenge competitions, a variety of topics mentioned above in the general PHM scope are covered. For the convenience of discussing different competitions, we shorten the name of each year's PHM Data Challenge competition as the following format: "ORGANIZATION'YR" to represent the competition organizer and year. For example, "PHM'08" represents the competition held by PHM Society in 2008; "IEEE'12" represents the competition held by IEEE Reliability Society in 2012. We summarize the tasks in the general PHM framework as shown in Figure 1. Three competition problems involve more than one PHM tasks. In PHM’11, a sensor health monitoring problem, given two types of dataset with different number of sensors included, it requires the detection of faulty condition for shear data and identification of all the possible combinations of multiple faulty anemometer for pair data, which are considered as health assessment and diagnostics, respectively. IEEE’14 splits the competition into two tasks: RUL Prediction and State of Health (SoH) estimation, which involves all three tasks. PHM’15 requires prediction of the failure mode, failure start time and failure end time, which is a combined task of prognostics and diagnostics, while other prognostics problem only consider the RUL prediction without the identification of failure modes.

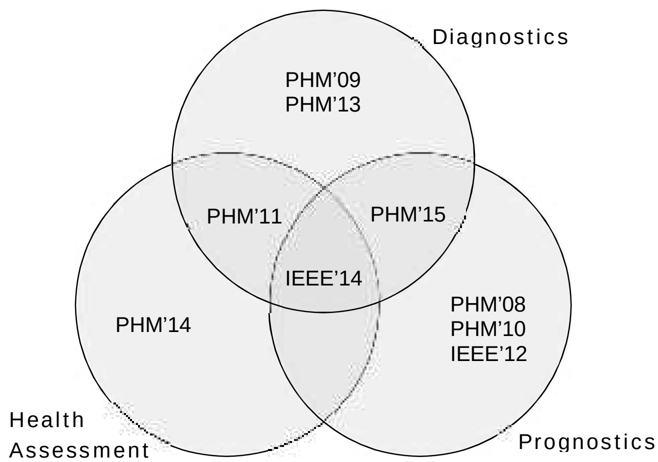

Figure 1: Tasks of PHM Data Challenge Competition Problems.

2.1. Basic Information of PHM Data Challenge Competitions: Since the tasks of competitions are sometimes multi-objective, we categorize all the competitions into three types according to the common objectives in PHM, namely, health assessment, fault classification and RUL prediction. Table 1 summarizes the basic information of PHM Data Challenge competitions and their datasets. The observations from this table are: 1) most time prediction tasks consider multi-asset and multi-regime issue without multiple failure modes, since the run-to-failure data provided is sufficient to represent the failure mode; 2) fault classification tasks focus more on the multi-failure mode issue than multi-asset and multi-regime scenarios; 3) health assessment tasks seem to focus less on regime complexity in the competitions.

3. Health Assessment: Health assessment task in the context of PHM Data Challenge competition can be defined as health condition estimation for a device or system at the time requested. Regarding its applications, health assessment can be applied either independently to real-time condition monitoring or to detecting the degradation trend with RUL prediction tasks as a pre-modeling step, as illustrated in Figure 1. Details of three health assessment problems have been listed in Table 2.

Table 1: Basic Information of PHM Data Challenge Competitions and Datasets

Task Objective		Organization & Year	Application	Asset	Failure	Regime	Data Source	Data Type	Variable Type
Health Assessment	NP	PHM 2011	Anemometer	NA	4	D	Monitoring	Short	Parameter
		PHM 2014	Unspecified Equipment	Fleet	14	NA	Maintenance & Usage	Long	Event & Parameter
	P	IEEE 2014	Fuel Cell	2	NA	S&D	Testbed	Short	Parameter
Fault Classification	NP	PHM 2009	Gearbox	NA	14	10	Testbed	Short	Waveform
		PHM 2013	Unspecified Equipment	NA	13	NA	Monitoring	Long	Event & Parameter
	P	PHM 2015	Plant Operation	Fleet	6	NA	Monitoring & Control	Long	Parameter
Remaining Useful Life Prediction		PHM 2008	Aircraft Engine	Fleet	NA	6	Simulation	RTF	Parameter
		PHM 2010	Milling Machine	6	NA	1	Monitoring & Usage	RTF	Waveform
		IEEE 2012	Bearing	17	NA	3	Testbed	RTF	Waveform
		IEEE 2014	Fuel Cell	2	NA	S&D	Testbed	RTF	Parameter

Notation:
P/NP – Prediction of failure time is required or not required.
NA – The criteria is neither clearly specified nor significant to problem.
Fleet – A large number of assets are considered.
S/D – Static or Dynamic regime is considered, respectively.
RTF – Run-to-Failure data is provided
Short/Long – Relatively short term or long term data is provided.

Table 2: Overview of Health Assessment Problems and Datasets.

	PHM'11	PHM'14	IEEE'14
Equipment	Anemometer	Unknown	Fuel Cell
Data Type	Monitoring	Maintenance Log & Usage	Testbed (Run-to-Failure)
Asset (train-test)	NA	1913-2076	1-1
Regime	Dynamic	NA	Static – Dynamic
Failure Mode	NA	14	1
Sampling Interval	Every 10min	Event	Aging: Every 30s EIS, Polarization: Weekly
Variables	Single Value	Single Value	Single Value
Variables	Pair: 16 Shear: 20 or 24	Part Consumption: 5 Usage: 3	Aging: 25 Polarization: 8 EIS: 3
Note that PHM'11 only provide a series of sample data without knowing whether they come from the same group of sensors or various assets: For pair data, 12 25-day length samples for training; 420 5-day length samples for testing. For shear data, 7 25-day length samples for training; 255 5-day length sample for testing			

Three levels of health assessment can be classified in PHM Data Challenge competitions due to their different levels of requirement: 1) abnormal detection using a feature-based indicator, which represents a base level binary health condition assessment. For example, PHM'11 only differentiates normal/abnormal or healthy/unhealthy conditions; 2) quantitative estimation of health condition using a self-defined health index, such as the failure rate of system defined different from method to method in PHM'14; 3) quantitative estimation of health condition using a special indicator, such as the diagnostic tool of electrochemical impedance spectroscopy (EIS) test being used in fuel cell (FC) SoH estimation of IEEE'14.

3.1. Challenges: Health condition always changes randomly with noise. Although the randomness or other contribution from invisible factor exists, part of the unstable behaviors can be theoretically explained or physically understood. Challenges of each health assessment problem are summarized in Table 3. In PHM'11, slow running situations are not caused by mechanical issues but temporary weather conditions, such as icing. Since they are not permanent degradation and recoverable, such cases should not be considered as true failure, and need to be filtered out during the modelling process. In PHM'14, two major changing modes of failure rate have been discovered according to the preventive maintenance policy in which the failure rate follows a bathtub curve, a case of Weibull distribution [17]. Failure rate decreases after preventive maintenance (PM) determined when a large number of parts are replaced and dramatically increases for a short period of time known as infant mortality after corrective maintenance (CM) determined when a few parts are replaced [18]. Both winning methods published this year built their models upon similar maintenance policies. In IEEE'14, a self-healing effect on FC power has been found with a local reversible degradation behavior after weekly performance or characteristic test. Therefore, how to understand and explain such phenomenon becomes a common issue for health assessment. In order to address these issues, finding suitable algorithms with deep domain knowledge is critical to develop an effective PHM solution no matter what strategies being chosen. Thoroughly understanding the background, relevant theories, underlying mechanism of the problem may help handling the uncertainty and facilitate method development.

Table 3: Summary of Challenges of Health Assessment Problems

	PHM'11	PHM'14	IEEE'14
Common Issues	• Find health indicator • Health condition degrades with uncertainty • Call for domain knowledge		
Unique Challenges	• Icing • Pattern recognition	• Implementation of maintenance policy	• IES Estimation • Little Reference

3.2. Solutions: In the competitions, the strategies took in solutions are so diverse that we cannot directly compare the solutions. Each of them represents one type of health assessment problem. Fortunately, since each solution create a health index in modeling, we discuss the solutions by following what health index was chosen and what algorithm was used.

PHM'11 – Binary Abnormal Detection Using a Feature-Based Indicator: Three solutions for this competition are compared. Before pattern classification, Siegel [19] and Sun [20]'s methods uses wind speed difference as health indicator and both applies power law equation and unify the data from different heights, whereas Cassity [21]'s method uses percentage difference of wind speed as health indicator and applies same wind profile theory to create wind profile models and fit in each height level. The champion solution proposed by Siegel applies an advanced algorithm Auto-Associative Neural Network (AANN) further process the health indicator for shear data other than Sum of Squared Residual (SSR) and R^2 value used by Sun and Cassity, respectively. In all three methods, the icing issue which would potentially cause false alarm is addressed by eliminating outliers. For pattern classification part, the health indicators are used as critical features and inputs of pattern recognition models. Siegel's k-mean clustering based figure of merit (FOM) method is able to partition the dataset with a bias on the samples containing lagging anemometers. Sun conducted a Euclidean similarity measurement based pattern search with a hybrid decision making method which has a close performance to the 1st place. Cassity's technique performs not as well as other two top methods do. This is probably because of the additional step it takes for initial guessing some bad sensors, as an essential step of generating labeled bad/faulty data samples for discriminant analysis which is a supervised learning method. Due to the dynamic working regime and multiple possible failure mode which is not clear to the participants in this competition, such process of using initial guess as an untrusted baseline data to detect other faulty sensors would be more likely to generate a larger error. The expertise knowledge involved in solution procedure is also summarized in the last column.

PHM'14 – Quantitative Estimation of Health Condition Using a Self-Defined Health Index: Similar to PHM'11, the dataset of PHM'14 also has two parts, maintenance log based part consumption data and usage measurement data. In both solutions, a deep expertise on maintenance policy has been adopted firstly, especially when defining and detecting Predictive Maintenance (PM) or Scheduled Maintenance (SM), and Corrective maintenance (CM). In spite of other statistical analysis, histogram chart is seen as the most frequently used statistical tool due to its capability of accumulating discrete data samples as frequency information that can be further converted as risk indicator. The champion's solution proposed by Rezvani [18] is quite complicated in statistic based analysis, in which two high risk time interval detection models have been individually developed for PM and CM based on Weibull Bathtub curve by only using the part consumption dataset, then used to test each asset sequentially at the requested times to determine its risk level. Instead of calculating high risk time interval sample by sample, the 2nd placed solution [22] created a high/low risk histogram-based spectrum for any type of failure mode (reason code) in time domain and in usage domain as well during the training stage, then simply test each sample whether it located in the high/low risk bins by a consensus ensemble strategy. Furthermore, unlike the champion method, this method utilizes the usage dataset revealing a linear relation of failure rate with accumulated usage, also makes a further inference from reason code to part number instead of summarizing parts information in repeated number of part numbers replaced (RNPNR). Despite of some of these part number being found to be uniquely associated to SM and CM, respectively, this solution treated them in one strategy with no difference instead of using separate strategies for PM and CM in champion's method. Overall, these two solutions are similar in methodology, technically compensate with each other and can be possibly fused into one more effective method.

PHM'15 - Quantitative Estimation of Health Condition Using a Special Indicator: As shown in the literature review in both papers, we see that the majority of research on FC or battery focuses on the prognostics and diagnosis, but less effort on EIS estimation which is related to the dynamic performance of FC. Basic background knowledge on FC and some expertise on EIS would be essential to solve this problem effectively. The champion of this task [23] developed a solution targeting to each issues mentioned above. The solution merged the power degradation information with Equivalent Circuit Model (ECM) for EIS estimation by considering an exponential local reversible degradation model for the degradation uncertainty issue. This requires a deep understanding of physical system of FC stacks. In the 2nd placed solution [24], not too much expertise was used in modeling process. Several different regression methods are applied to fit the real part and imaginary part independently, though much more background and domain knowledge is mentioned at the beginning. This method is very straightforward and only considered the EIS data without any degradation process. Data with current information and polarization curve hasn't been used in both solutions, but may lead to better results. Another solution published to RUL prediction of FC, anther split task of this competition using the same dataset, will be discussed later in Section 5.2.

4. Fault Classification: Finding root cause of machine failure is generally considered as the major task of diagnostics. In the content of PHM Data Challenge competition, the task of root cause finding is mostly expressed as fault classification problem that is required to determine a certain sample's fault type from a series of possible failure modes. Details of three fault classification problems have been summarized in Table 4.

Table 4: Overview of Fault Detection Problems and Datasets

	PHM'09	PHM'13	PHM'15
Equipment	Gearbox	Unknown	Plant
Data Type	Short-time Testbed	Maintenance Log & Monitoring	Monitoring
Asset (train-test)	NA	1	33-15
Regime	10	NA	Multiple
Failure Mode	14	13	6
Sampling Interval	Test	Event	Every 15min
Variables	Waveform	Single value	Single Value
Variables	3	2, 2	Monitoring: 14 Fault Event: 3

Two classification strategies can be grouped according to the data type that needs to be processed:

1) Feature-based pattern classification, in which the performance of the algorithm highly depends on the quality of features extracted from the dataset, e.g. PHM'09 and PHM'15 give a no training data situation and a no background situation for classification task, respectively, thus the former requires a powerful feature extraction methods, and the latter requires thorough data exploration process. In both situations, deep expert knowledge would be necessary.

2) Event-based inference and classification, in which major effort would be taken in revealing the complex relationship among multiple character type variables so as to classify the data samples, e.g. PHM'13 asks participants to develop a maintenance action recommender which is capable of infer the problem ID based on event type data generated from the onboard condition monitoring system.

4.1. Challenges: Despite of the types of classification technique, the challenges that each problem faces are different, mainly due to their diverse backgrounds. Challenges of each health assessment problem are summarized in Table 5. For PHM’09, because of no training data (labeled data) provided, we would only be able to mine useful information or extract features from unlabeled data, which further implies the necessity of developing more powerful methods with more expertise in signal processing for vibration data and gearbox domain knowledge. For PHM’13, three character type variables, “event”, “case” and “problem type”, formalizes a complex triangular relationship in which each two of them are somehow related. In addition, once each event occurs, 30 numerical parameters would be generated by device, as a snapshot of system condition. Hence the unique challenges in this competition are 1) how to reveal the triangular relationship and 2) how to appropriately combine the two sources of data and systematically use them. For PHM’15, a task consisting of both fault classification and failure time prediction, the first challenge would be determine a strategy to deal with such multi-objective task. Besides, finding high quality features without given details about the physical system is another challenge.

Table 5: Summary of Challenges of Fault Classification Problems

	PHM’09	PHM’13	PHM’15
Common Issues	- Find health indicator - Health condition degrades with uncertainty - Call for domain knowledge		
Unique Challenges	- No training data - Signal processing	- Complex Relationship among “event”, “case” and “problem type” - Data Fusion	- Fault classification and time prediction - Feature extraction without knowing the mechanism of system

4.2. Solutions: The solutions to fault classification problems are discussed by following the two types of strategies mentioned above. PHM’09 and PHM’14 belong to feature based pattern classification problem. PHM’13 belongs to event based inference and classification problems.

For PHM’09, fault classification without training data, the champion Wu proposes a compact and integral solution which works in a sequential process with the best performance in this competition [25]. The idea of using many bandpass filters to decompose the whole frequency spectrum to achieve the global frequency analysis as well as local analysis of frequency component of interest simultaneously requires a deep insight of vibration signal processing skills. Al-Atat’s solution is relatively complicated and module-based, but follows a more general diagnosis approach [26]. It consists of many sub-methods regarding to regime segmentation, health assessment and fault classification. These sub-methods are highly independent with each other, and thus have more flexibility being used in other applications. In this competition, all methods are used with prior expertise or for exploring the hidden relations.

For PHM’15, fault classification with downtime prediction, in order to address this multi-objective task, both two top ranked solutions take the same strategy that machine learning based fault classifiers are built in each time interval for each plant, and features are extracted based on the comprehensive data exploration. In champion’s solution [27], several algorithms are used for correlation analysis for variables and components, and periodicity of failure time is also found. For 2nd place [28], features are extracted based on inference from physical interpretation, basically guessing the possible physical meaning of each correlated sensors, seasonality analysis, health condition detection by a visualization tool named empirical Probability Density Function (PDF), and a “one hour” rule being discovered by time series analysis. Both of the above two methods performs very close to each other, with score of 21015.55, 20639.71, respectively, nearly double as the 3rd place solution [29], 10221.54. The 3rd place solution also uses ensemble technique to combine Random Forest (FR) and Gradient Boosting Decision Tree (GBDT) to form a decision tree based classifier. Besides, the 3rd place solution focused on an issue of fault event overlapping. It is observed that building a multiclass classifier in each time step seems a good strategy for multi-objective task of fault classification and prediction, and preparing high quality features which requires thorough exploration is critical to the performance of a method.

Table 6 Summary of Solutions to Event Based Inference and Classification Problems

PHM’13	Score	Dataset Used	Basic Strategy	Algorithms
1st - Das	86	Event data	Matrix decomposition	ALS-NMF
2nd - Katsouros	60	Event data	Bayesian Classification	Naïve Bayesian Classifier
3rd - Kimotho Method 1	51	Event data and parametric data	Classification for event	Event-based DT + Ensemble of classifiers of BT, RF and SVM
3rd - Kimotho Method 2	48	Event data and parametric data	Classification for event	Event-based DT + SVM
2nd - Katsouros Method 2	35	Parametric data	Classification	SVM with radial based kernel
6th - Siegel Method 1	35	Event data and parametric data	Regression for problem	Event based random forest
6th - Siegel Method 2	30	Parametric data	Regression for problem	Random forest

For PHM’13, a typical fault classification problem with a dataset mixed with event data and parametric data, a benchmark of solutions to this event based inference and classification problem has been illustrated in Table 6. Parametric data is an extra dataset with 30 parameters generated by control system when a specific condition corresponding to an event code in event data is met onboard. Based on the papers and presentation slides [9] published by PHM Society, 7 solutions have been benchmarked and ranked according to their scores shown in Table 6. High performance (score>60): Both of the top two methods in this range only use event data and take their main efforts in relationship inference. The champion Das [30] introduces a collaborative filtering based method typically used for 'Recommender System' development in Information Technology (IT) field, e.g. e-commerce item recommendation [31]. The 2nd placed Bayesian classifier proposed by Katsouros [32] has much lower accuracy than the 1st placed Non-Negative Matrix Factorization (NMF) methods, but it has a straightforward inference with a big advantage in its simple structure and computationally effectiveness. Medium performance (60>score>40): The two second level methods are all based on classification for each event. Solutions in this range are generated by the 3rd place participant Kimotho, using both event data and parametric data to classify different event codes. Low performance (score<40): Other three methods given by Katsouros and Siegel either choose regression instead of classification or completely ignore the event data. Therefore we may conclude that parametric data may not be very useful as the Event ID/Code itself summaries the parameters settings for a specific condition [30]. Beyond the performance, due to the time constraint of competition, most participants in PHM'13 follows a trail-and-error strategy and gives more than one solutions to the problem, however the champion who gains more scores than others seizes the essence of the problem and introduces a highly effective method from the IT field.

5. Remaining Useful Life Prediction: RUL prediction tasks occupy nearly half of all PHM Data Challenge competitions, because of its significance of assisting maintenance scheduling and health management. It is very common to see the health assessment works as a critical step to support RUL prediction in real applications. In competitions, the objective of RUL prediction is to estimate the remaining working time of a system before its health condition deteriorates below a threshold indicating a faulty condition. In competitions, RUL prediction doesn't explicitly require to describe the health condition during degradation as long as the solution accurately estimates the time before actual failure, which make it literally differentiable from health assessment tasks in this competition. Details of four RUL prediction problems have been summarized in Table 7.

Table 7: Overview of RUL Prediction Problems and Datasets

	PHM'08	PHM'10	IEEE'12	IEEE'14
Equipment	Engine	Milling Cutter	Bearing	Fuel Cell
Data	Simulation (Run-to-Failure)	Monitoring & Usage (Run-to-Failure)	Testbed (Run-to-Failure)	Testbed (Run-to-Failure)
Asset (train-test)	218-218	3-3	6-11	1-1
Regime	6	1	3	2(Static-Dynamic)
Failure Mode	NA	NA	NA	NA
Sampling Interval	Operation Cycle - Flight	Operation Cycle - Cut	Every 10s	Every 30s
Variables	Single Value	Single Value	Waveform	Single Value
Variables	30	8	3	25

5.1. Challenges: As shown in Table 8, the common challenges of RUL prediction is finding an appropriate degradation trend which can be identified as an indicator for modeling process. However, in different systems and situations, degradation detection would meet different challenges related to number of assets and number of regimes. PHM'08 aircraft engine dataset shows a clear degradation trend with multiple regimes and a large number of units. PHM'10 milling cutter machine dataset also shows a clear degradation trend but with few units available to analysis, which causes data scarcity issue. PHM'12 bearing dataset shows a very unclear end-of-life degradation signature without an ideal gradually monotonic degradation trend. We also notice the difficulty of dealing with bear's uncertain behavior which is caused by its unbalanced dataset and multiple regimes. PHM'14 fuel cell dataset has a unique training/testing mode that training data in static regime is used as reference to estimate the test sample's RUL in dynamic regime and a special degradation mode, self-healing effect or local reversible degradation, caused by its electrochemical features.

Table 8: Summary of Challenges of Remaining Useful Life Prediction Problems

	PHM'08	PHM'10	PHM'12	IEEE'14
Common Issues	• Degradation trend detection			
Unique Challenges	• A fleet of asset • Multiple regimes	• Data scarcity	• Unclear end-of-life signature • Unbalanced dataset	• Static/dynamic regime • Reversible degradation

5.2. Solutions: A comprehensive review [14] on the papers using the C-MAPSS dataset in which PHM'08 competition dataset is included has been published. In that review, the solutions of 2nd place [33] and 3rd place [34] are classified as "Category 1 - using functional mappings between set of inputs and RUL", and champion's solution [35] are classified as "Category 2 - functional mapping between health index and RUL" as well as "Category 3 - similarity-based matching". Our review of methods provided in all competitions related to RUL prediction is basically conducted in a similar way. Over all solutions to RUL prediction problems, two strategies have been observed so far for data-driven RUL prediction methods: 1) finding a degradation indicator and predicting the trend until a threshold; 2) direct mapping from extracted features to RUL index. Since we follow a problem-centered manor of review, the solutions are discussed with their unique difficulties.

In PHM'08, RUL prediction with a fleet of assets and multiple regimes, the champion's solution proposed by Wang [35] takes the first strategy and clearly applies three techniques targeting on each of the single challenges, degradation trend fitting, asset variety and multi-regime. The other two solutions take the 2nd strategy. The 2nd place Heimes' solution [33] brings a mature tool that is specially developed for nonlinear dynamic system with capability of dealing with uncertainty and complexity. The 3rd place Peel's solution takes advantages of Kalman Filter (KF) to treat variables as time series data integrating past information and reduce noise in raw data and error in asset variety.

For PHM'10, RUL prediction with data scarcity which means only 3 assets of training samples are available for predicting another 3 testing samples whose data is only partially given, the champion's solution [36] proposed by Das follows a classic and general PHM solution workflow. The key point for his high accuracy prediction is attributed to successfully finding a critical feature, harmonic of tooth pass frequency, and employing a batch of 100 neural networks to deal with the data scarcity issue. The 2nd place solution given by Chen [37] is developed in the framework of Bayesian statistic, but not using any Bayesian inference methodology. Both of them are second category regression based approach, however, the difference is Chen proposes an individual step of estimating the initial wear state of the cutter instead of using the wear state related feature from last cut as input of regression model in Das' solution.

For PHM'12, RUL prediction with big challenges in handling an unclear degradation end-of-life signature and unbalanced dataset, the champion Sutrisno developed three independent solutions [38], two of them take the first strategy and another one take the second strategy. For the solutions with a degradation indicator, Sutrisno's Method 1 detects an ideal degradation indicator, moving-averaged spectral (MAS) kutosis with a gradual and monotonic trend, then predicts the trend by Bayesian Monte Carlo; Sutrisno's Method 3 extracts a "prognostic feature" by taking averaging of 5 maximum values of variables, eventually calculate the end-of-life time using the ratio of multiple degradation stages which can be easily observed. For the regression-based solution in Sutrisno's Method 3, also the only one in this competition, least squares-support vector regressor (LS-SVR) with features in time domain and wavelet analysis is applied for exploring the underlying relationship between features and RUL directly by the regression method, and achieved the best performance in this competition. The 2nd place solution given by Wang [39] is actually similar to Sutrisno's Method 1 in terms of its workflow, but using an health index extracted by Principal Component Analysis (PCA) and T^2 statistic. Though it applies some bearing diagnosis techniques, such as bearing characteristic features and natural frequency analysis, this method proposed does not have a strong capability of detecting the degradation trend to overcome this challenge. By comparing the strategies took in different solutions, we conclude that multi-stage degradation identification should be the key to RUL prediction problem with unclear end-of-life trend.

For IEEE'14, RUL prediction with variety of static/dynamic regime, due to its unusual dataset with only one sample in either training or testing dataset, it assumes that the variety issues between different FC stacks can be neglected, and uncertainty observed in data only comes from dynamic regime and testbed. The only solution published for this split task in IEEE'14 by Kimotho [40] is developed based on Particle Filter method with a self-healing factor adapted in each iteration. It tried 5 different state modes in the adaptive particle filtering process and combine them using ensemble technique.

6. Conclusion: To conclude, all competitions in the past 9 years mainly involve the three basic PHM tasks, health assessment estimating the machine or system health condition, diagnostics finding the machine or system failure root cause, and prognostics predicting the machine or system failure time before failure actually occurs. We found that health assessment always supports diagnostics and prognostics. Based on the objectives of PHM Data Challenge competitions, PHM'11, PHM'14 and IEEE'14 are categorized into health assessment problems; PHM'09, PHM'13 and PHM'15 are categorized into fault classification problems; PHM'08, PHM'10, IEEE'12 and IEEE'14 are categorized into RUL prediction problems.

Most RUL prediction problems and fault prediction problem in PHM'15 consider multi-asset and multi-regime issues. A fleet-based analysis, such as similarity-based approach, is effective for RUL prediction with a large number of assets in PHM'08. Initial wear estimation and ensemble technique are necessary in RUL prediction with data scarcity issue. Multi-stage degradation identification is observed to be very effective in RUL prediction problem with unclear end-of-life signature. Domain knowledge would contribute the RUL modeling process only when some special effects have to be considered. A good example of applying domain knowledge in RUL prediction is that a factor which represents the self-healing behavior of fuel cell stacks is specially adapted into the particle filter technique in the solution to RUL prediction in IEEE'14 [40]. A negative example is that common bearing diagnosis technique such as characteristic features does not assist in degradation trend detection in IEEE'12 [39]. After reviewing all solutions, two basic strategies can be summarized in RUL prediction: 1) finding a degradation indicator and predicting the trend until a threshold; 2) direct mapping from extracted features to RUL by regression. According to our review, both two strategies are capable of solving RUL prediction as long as the algorithm of the selected strategy is used for a clear purpose and targeting to the challenges.

Fault classification tasks focus less on multi-asset and multi-regime issues in the PHM Data Challenge competitions. High quality features which require thorough exploration directly impact the method's performance. Prior expertise knowledge and data exploration is critical to classification without training data in PHM'09. Building a multiclass classifier in each time step has been seen as a general strategy for multi-objective task of fault classification and prediction in PHM'15. Data selection and filtering is significant in the situation with heterogeneous and complex dataset especially in inference based classification problems. The benchmark in PHM'13 gives us a lesson that choosing a right tool would be much more effective and efficient than trail-and-error based algorithm selection.

Health assessment tasks in competitions have less concerns on regimes. Three problems including PHM'11, PHM'14 and IEEE'14 in this category represents three levels of health assessment according to their requirements, binary health/unhealthy detection, quantitative health condition estimation using a self-defined health indicator and quantitative health condition estimation using special indicator, respectively. According to our review, all health assessment solutions rely on domain knowledge, e.g. wind profile equation used for handling shear data in PHM'11, expertise related to maintenance policy used to define PM/SM and CM in PHM'14, fuel cell property related domain knowledge used for modeling fuel cell dynamic behavior in EIS spectrum. Since we have already known that health assessment can possibly collaborate with any PHM related tasks, it implies that any other problems, either diagnostics or prognostics, would heavily rely on domain knowledge if health assessment is included.

6.1. Further Work: In this work, we only benchmark the algorithm performance for PHM'13. In the future, similar algorithm benchmark needs to be conducted for other competitions. In addition, other issues such as tradeoff between accuracy and efficiency of PHM solutions and use of ensemble techniques also deserve further analysis. The problem-centered review proposed in this paper could be extended to a wider range of resources, such as publications based on the same datasets, publication with the same objectives or using the same class of techniques, etc.

References:

[1] Accenture and GE. (2015). *Industrial Internet Insights Report. Industrial Insights Report*. Available: https://www.accenture.com/us-en/_acnmedia/Accenture/next-gen/reassembling-industry/pdf/Accenture-Industrial-Internet-Changing-Competitive-Landscape-Industries.pdf
[2] J. Lee, *Industrial Big Data*: Mechanical Industry Press, China, 2015.
[3] N. Millman, " Big data to unlock value from the industrial internet of things.," ed, 2015.

[4] A. Alter, P. Banerjee, P. E.Daugherty, and W. Negm, "Driving Unconventional Growth through the Industrial Internet of Things.," ed, 2014.
[5] *PHM Competitions*. Available: https://www.phmsociety.org/competition
[6] *2009 PHM Challenge Competition Data Set*. Available: https://www.phmsociety.org/references/datasets
[7] K. G. A. Saxena, ""PHM08 Challenge Data Set", NASA Ames Prognostics Data Repository," 2008.
[8] *2011 phm society conference data challenge*. Available: https://www.phmsociety.org/competition/phm/11
[9] (11 March). *2013 PHM Society Conference Data Challenge*. Available: https://www.phmsociety.org/events/conference/phm/13/challenge
[10] F. Ahmadzadeh and J. Lundberg, "Remaining useful life estimation: review," *International Journal of System Assurance Engineering and Management,* vol. 5, pp. 461-474, 2014// 2014.
[11] X.-S. Si, W. Wang, C.-H. Hu, and D.-H. Zhou, "Remaining useful life estimation – A review on the statistical data driven approaches," *European Journal of Operational Research,* vol. 213, pp. 1-14, 8/16/ 2011.
[12] A. K. S. Jardine, D. Lin, and D. Banjevic, "A review on machinery diagnostics and prognostics implementing condition-based maintenance," *Mechanical Systems and Signal Processing,* vol. 20, pp. 1483-1510, 10// 2006.
[13] J. Lee, F. Wu, W. Zhao, M. Ghaffari, L. Liao, and D. Siegel, "Prognostics and health management design for rotary machinery systems—Reviews, methodology and applications," *Mechanical systems and signal processing,* vol. 42, pp. 314-334, 2014.
[14] E. Ramasso and A. Saxena, "Performance Benchmarking and Analysis of Prognostic Methods for CMAPSS Datasets," *International Journal of Prognostics and Health Management,* vol. 5, pp. 1-15, 2014.
[15] G. Zhang and J. Isom, "Gearbox vibration source separation by integration of time synchronous averaged signals," 2011.
[16] P. Boškoski and A. Urevc, "Bearing fault detection with application to PHM Data Challenge," *International Journal of Prognostics and Health Management Volume 2 (color),* p. 32, 2011.
[17] T. Nakagawa, "Periodic and sequential preventive maintenance policies," *Journal of Applied Probability,* vol. 23, pp. 536-542, 1986/006/001 1986.
[18] S. M. Rezvanizaniani, J. Dempsey, and J. Lee, "An effective predictive maintenance approach based on historical maintenance data using a probabilistic risk assessment: PHM14 data challenge," *International Journal of Prognostics and Health Management,* vol. 5, 2014.
[19] D. Siegel and J. Lee, "An auto-associative residual processing and K-means clustering approach for anemometer health assessment," *International Journal of Prognostics and Health Management Volume 2 (color),* p. 117, 2011.
[20] L. Sun, C. Chen, and Q. Cheng, "Feature extraction and pattern identification for anemometer condition diagnosis," *Int. J. Progn. Heal. Manag,* vol. 3, pp. 8-18, 2012.
[21] J. Cassity, C. Aven, and D. Parker, "Applying weibull distribution and discriminant function techniques to predict damaged cup anemometers in the 2011 PHM competition," *Int. J. Progn. Heal. Manag,* vol. 3, pp. 1-7, 2012.

[22] H. Kim, T. Hwang, J. Park, H. Oh, and B. D. Youn, "Risk Prediction of Engineering Assets: An Ensemble of Part Lifespan Calculation and Usage Classification Methods," 2014.
[23] T. Kim, H. Kim, J. Ha, K. Kim, J. Youn, J. Jung, *et al.*, "A degenerated equivalent circuit model and hybrid prediction for state-of-health (SOH) of PEM fuel cell," 2014, pp. 1-7.
[24] W. O. L. Vianna, I. P. d. Medeiros, B. S. Aflalo, L. R. Rodrigues, J. P. P. Mal, x00E, *et al.*, "Proton Exchange Membrane Fuel Cells (PEMFC) impedance estimation using regression analysis," in *Prognostics and Health Management (PHM), 2014 IEEE Conference on*, 2014, pp. 1-8.
[25] F. Wu and J. Lee, "Information Reconstruction Method for Improved Clustering and Diagnosis of Generic Gearbox Signals," *International Journal of Prognostics and Health Management Volume 2 (color)*, p. 42, 2011.
[26] H. Al-Atat, D. Siegel, and J. Lee, "A systematic methodology for gearbox health assessment and fault classification," *Int J Prognostics Health Manage Soc*, vol. 2, p. 16, 2011.
[27] W. Xiao, "A Probabilistic Machine Learning Approach to Detect Industrial Plant Faults," *arXiv preprint arXiv:1603.05770*, 2016.
[28] H. Kim, J. M. Ha, J. Park, S. Kim, K. Kim, B. C. Jang, *et al.*, "Fault Log Recovery Using an Incomplete-data-trained FDA Classifier for Failure Diagnosis of Engineered Systems," 2016.
[29] C. Xie, D. Yang, Y. Huang, and D. Sun, "Feature Extraction and Ensemble Decision Tree Classifier in Plant Failure Detection," presented at the Annual Conference of the Prognostics and Health Management Society, 2015.
[30] S. Das, "Maintenance action recommendation using collaborative filtering," *Int. J. Progn. Health Manag*, vol. 4, pp. 7-12, 2013.
[31] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, pp. 76-80, 2003.
[32] V. Katsouros, V. Papavassiliou, and C. Emmanouilidis, "A Bayesian Approach for Maintenance Action Recommendation," *International Journal of Prognostics and Health Management, International Journal of Prognostics and Health Management*, vol. 4, p. 034, 2013.
[33] F. O. Heimes, "Recurrent neural networks for remaining useful life estimation," 2008, pp. 1-6.
[34] L. Peel, "Data driven prognostics using a Kalman filter ensemble of neural network models," in *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, 2008, pp. 1-6.
[35] T. Wang, Y. Jianbo, D. Siegel, and J. Lee, "A similarity-based prognostics approach for Remaining Useful Life estimation of engineered systems," in *Prognostics and Health Management, 2008. PHM 2008. International Conference on*, 2008, pp. 1-6.
[36] S. Das, R. Hall, S. Herzog, G. Harrison, M. Bodkin, and L. Martin, "Essential steps in prognostic health management," 2011, pp. 1-9.
[37] H. Chen, "A multiple model prediction algorithm for CNC machine wear PHM," *International Journal of Prognostics and Health Management Volume 2 (color)*, p. 129, 2011.

[38] E. Sutrisno, H. Oh, A. S. S. Vasan, and M. Pecht, "Estimation of remaining useful life of ball bearings using data driven methodologies," 2012, pp. 1-7.
[39] T. Wang, "Bearing life prediction based on vibration signals: A case study and lessons learned," 2012, pp. 1-7.
[40] J. K. Kimotho, T. Meyer, and W. Sextro, "PEM fuel cell prognostics using particle filter with model parameter adaptation," in *Prognostics and Health Management (PHM), 2014 IEEE Conference on*, 2014, pp. 1-6.