Enhanced K-Nearest Neighbors Method Application in Case of Draglines Reliability Analysis

A. Taghizadeh Vahed[1], B. Ghodrati[1], S.H. Hosseinie[2]

[1] Division of Operation and Maintenance Engineering, Lulea University of Technology, Sweden
[2]Faculty of Mining Engineering, Isfahan University of Technology, Iran

Email of Corresponding author: Behzad@ltu.se

Abstract. Dragline's availability plays a major role in sustaining economic feasibility and operation of opencast coal mine. Thus, its reliability is essential for the production availability of mine. The dragline's reliability and maintenance optimization are key issues, which should seriously be considered. Draglines' unexpected failures and consequently unavailability result in delayed productions and increased maintenance and operating costs. The applications of methodologies which can predict the failure mode of dragline based on the historical dataset of failure are not only useful to reduce the maintenance and operating costs but also increase the availability and the production rate of mining machineries. In this research a historical failure dataset of a dragline has been utilized in order to analyze and conduct predictive maintenance. Authors have already utilized the K-Nearest Neighbors (KNN) algorithm in order to predict the failure mode; however, there was a chance of getting into local optimum by utilization of the mentioned methodology. In this case, combination of Genetic Algorithm and K-Nearest Neighbor algorithm (i.e. called enhanced K-Nearest Neighbors) was applied for the failure dataset, so the probability of local optimum has been decreased by application of Genetic Algorithm. In previous studies, the Artificial Neural Network methods and conventional method of K-Nearest Neighbor has been applied to the same dataset, yet the result from enhanced K-Nearest Neighbor reveals better regression analysis.

Keywords: Draglines Maintenance, Machine Learning, K-Nearest Neighbor, Genetic Algorithm, Reliability.

1 Introduction

Draglines are major production equipment in surface strip mining operation. High productivity with minimum operating cost are the major goals with a reliable dragline. Thanks to new technologies and innovations, new mining machineries are more reliable than the previous one; however, the harsh mining environment do not allow the mining machineries to be as available as it is expected. Lack of proper services, maintenance and investigation may increase the rate of failures and unavailability [1]. In this condition, the harsh environment causes to unexpected breakdowns for a dragline and failure of its components as well. A failure takes a place in a subsystem of a dragline stops the whole machine operation. Therefore, prediction of failure time and failure mode in advance based on the historical data of machine assists the maintenance crews and managers to carry out the proactive actions, which eventually enhance the reliability and increase the availability of the machine.

Reliability of mining machineries have to be seriously considered, so mines can have predictive/preventive maintenance plans which carry out the maintenance program on time, which assure the availability of machineries. In order to conduct the preventive maintenance, it is vital to make an appropriate maintenance plan. In this case, predictive maintenance reveals its functionality. In the predictive maintenance, there is need to get more information and knowledge from historical data, which represents the previous status of machine such as: i) Time to Failure (TTF), ii) Time Between Failures (TBF), iii) Time to Repair (TTR), iv) Time Between Repairs (TBF), and v) maintenance actions. Considering the historical data, it is possible to extract a pattern(s), which is called Data Mining procedure. Thus, extracted pattern that includes more information and knowledge will be utilized to better proactive maintenance. In conventional method, mines conduct periodic maintenance that is based on a specific period of time or amount of machine use [2, 3]. In the mentioned method, regular inspection, repair, and replacement of parts is done in specific time which is defined by the manufacturer. Therefore, the benefit of preventive maintenance is the reduced probability of equipment breakdowns and extension of equipment life [4]. On the other hand, the disadvantage of preventive maintenance is the need to interrupt production at scheduled intervals to perform the maintenance action [4]. Therefore, based on the machine condition monitoring data and its historical data, there is possibility to plan appropriate proactive maintenance. In order to make a better plan for maintenance, different types of methodologies have been utilized such as i) Failure Mode, Effect and Criticality Analysis (FMECA), Fault Tree Analysis (FTA), Markovian Analysis (MA), Reliability Block Diagram (RBD). Since all the named methodologies are based on the conventional linear algebra and statistical algorithms, the application of them in a domain with high volume and variety of data is impractical [5]. However, to the best of the authors' knowledge, machine learning methodologies can be applied in the historical dataset of the dragline in order to make a predictive maintenance model. The machine learning techniques are widely used in order to not only figure out the pattern(s) but also develop a predictive model. Machine learning algorithms can be utilized for fault diagnostics, fault prediction, and prognostics actions. Machine learning gives an extra hand to System Health Monitoring which contains (i) a set of activities preformed on a system to maintain it in operable condition, (ii) maintenance data management systems, and (iii) online reliability estimation by components' degradation signal processing [6]. Therefore, machine learning methodologies can be applied in the mining industry in order to enhance the proactive maintenance. The best practice in the mining industry is based on the application of classical methodologies of reliability analysis, which measure the Key Performance Indicators [7-12]. In the previous studies, TBF and TBR have been applied; however, more data give more chance for making better predictive model and represents more information about status of a machine.

In this study, the machine learning methodology, which is K-Nearest Neighbor (KNN) in order to make a predictor model (i.e., called predictor) for failure type prediction is used. However, based on our previous studies [13-14], it has been recognized that the machine learning methodologies require some more practice in order to make an appropriate predictor in the case of tuned parameters and hyper-parameters as well. Therefore, for enhancing the previous predictor, which was created by KNN, Genetic Algorithms (GAs) was combined with and better result was obtained. KNN prone to fall in the local minimum which decrease the predictor's efficiency; therefore, GAs algorithm handles the mentioned issue, and direct the parameter tuning in the way which is close to global minimum. The remaining part of this paper is divided as follow: Section II represents the background information of machine learning and the utilized methodologies; Section III includes general information about the dragline and how Enhanced-KK was implemented; and finally, in section IV, conclusion and some comments are represented.

2 Methods

In this paper, Machine learning techniques and algorithms have been applied regards to make a predictor/classifier. A classifier assists to make a prediction for an unlabeled vector of input. Indeed, in the topic of machine learning, there are two major subjects which are widely used: i) supervised learning, and ii) unsupervised learning. In fact, supervised learning is called classification or regression, but unsupervised learning is called clustering. In unsupervised learning, the predictive model is going to make an algorithm based on a dataset, which does not have any label in its output data. On the other hand, supervised learning methodology uses the datasets which output data has label. Clustering and classification are also called un-labelled and labelled respectively. The modeling procedure starts by dividing a dataset into two groups: training dataset and testing dataset. In this study, based on the best of the authors' knowledge, 70 percent of the dataset is used for training and remaining for testing one.

In this study, due to labeled output a supervised learning methodology is used. KNN was utilized that is known as supervised learning when data analyzer is faced with labeled output data (the historical dataset of the dragline has labeled output, which is classifying problem).

2.1 Supervised Learning Methodology: KNN

The used methodology in this study, KNN, gets a set of n data point in d-dimensional space R^d and an integer k, and the problem is to determine a set of k points in R^d, called centers, so as to minimize the squared distance from each data point to its nearest center [15]. K-Nearest Neighbors algorithm finds a partition such that squared error between the empirical mean of a cluster and the points in the cluster is minimized. If μ_x be the mean of the cluster c_k, then the squared error between μ_k and the points n cluster c_k is defined as:

$$J(c_k) = \sum_{x_i} \|x_i - \mu_k\|^2 \ . \tag{1}$$

The main goal of KNN is to minimize the sum of the squared error over all *k* cluster, which is defined in this study. One way to overcome the local minima is to run the KNN algorithm, for a given K, with multiple different initial partitions and choose the partition with the smallest squared error.

KNN is one of the most popular algorithms for pattern recognition. However, KNN has some limitations such as i) computationally expensive due to utilizing all the training instants for classification, ii) KNN's performance is dependent to the training dataset, and iii) the data in any training data set does not have any difference with each other. In this case, for handling the limitation of KNN and improve it, Genetic Algorithms (GAs) can be used.

2.2 Genetic Algorithms

Evolutionary computing started by adapting ideas from biological theory into computer science. Genetic algorithms are most popular technique in evolutionary computing system. Evolutionary algorithms are used in the problems for optimization such as i) machine intelligence, ii) traveling sales person problem, iii) expert system, iv) medicine, v) engineering application, and vi) wired and wireless communication systems. Genetic Algorithms are implemented for searching in complex, large and multidimensional landscapes, which represents near-optimal solutions for objective or fitness function for the optimization issues.

GA is encoded the parameters in the search space to form a strings (i.e., called chromosomes). As a sequence, the collection of chromosomes creates population. Firstly, a random population is created, which represents different points in the search space. An objective and fitness function is associated with each string that shows the degree of goodness of string. Regarding to the survival of the fittest, a few of the strings are selected and each is assigned a number of copies that go into the mating pool. Some operation take places on the population, for instance, cross-over and mutation, which yields new generation of strings. Selection process based on cross-over or mutation are continued until the termination threshold is satisfied.

Based on applied genetic operators, the local maximum fitness value is calculated and is compared with the global maximum. In this stage, if the local maximum is bigger than global one, then the global maximum is replaced with the local maximum, and the next iteration takes a place. Algorithm is shown as follow:

---

1. Number of samples are selected which the training set is going to generate an initial population.
2. Distance between training instance in each chromosome and testing instance are evaluated.
3. Highest fitness value is selected and assigned as global maximum.
    a. For i = 1 to L
        i. Reproduction
        ii. Cross-over operator
        iii. Mutation and new population
        iv. Calculate the local Maximum
        v. If local maximum > global maximum

1. Global maximum = local maximum

b. Repeat

4. Final result: the chromosome which contains global maximum has the optimum KNN, and it is labeled as classification result.

The mentioned methodology was utilized in this study.

3 Case study

In this study, enhanced KNN implemented on a dataset, which was formed based on captured data from a dragline that currently is operating in a coal mine. The utilized methods that are categorized as the classification algorithms have 1 feature and 1303 observations. Six different failure types were identified in the dataset. Table 1 presents the causes of dragline failures and failure numbers associated with each type. The input data are Time to Failure (TTF) and the output data are called types of failures. Table 2 shows the sample datasets used as input for the algorithm.

Table 1. Causes and number of dragline failures.

ID of Failure Type	Failure Type	Failure Number
1	Mechanical Failure	945
2	Electrical Failure	276
3	Energy related Failure	53
4	Major Revision	5
5	Maintenance	14
6	Power Cut	7

Table 2. Sample dataset.

Mean Time to Failure (h)	ID of Failure Type
1.6597	1
3.5625	1
14.4167	2
5.5625	6
0.5938	1

Overburden stripping task are carried out by the draglines which are widely used in opencast coal mine. In this case, not only reliability but also availability of draglines have great effect on overall production rate of the mine. In order to increase the availability of the dragline, it is vital to focus on its major sub-system such as i) hoisting, ii) walking, and iii) swing. A draglines sub-system failure causes production losses which is around one million dollar per day [16]. In order to increase the availability of a dragline, predictive maintenance can be very suitable and beneficial. Therefore, based on the historical data of the dragline's failures, a predictive maintenance model has been developed. As mentioned before, first the dataset is divided into three parts: training dataset, validation dataset, and testing dataset, and then the enhanced KNN is fed by training data. Due to labeled output in the training dataset, the enhanced KNN in this study is used as classifier. In order to evaluate the accuracy of the developed model, the testing dataset was fed to it, so the accuracy of 82 percent has been reached (Fig 1). The accuracy for the classifying the failure type in previous studies [13, 14] was at most 72 percent which had been reached by the conventional method of KNN.

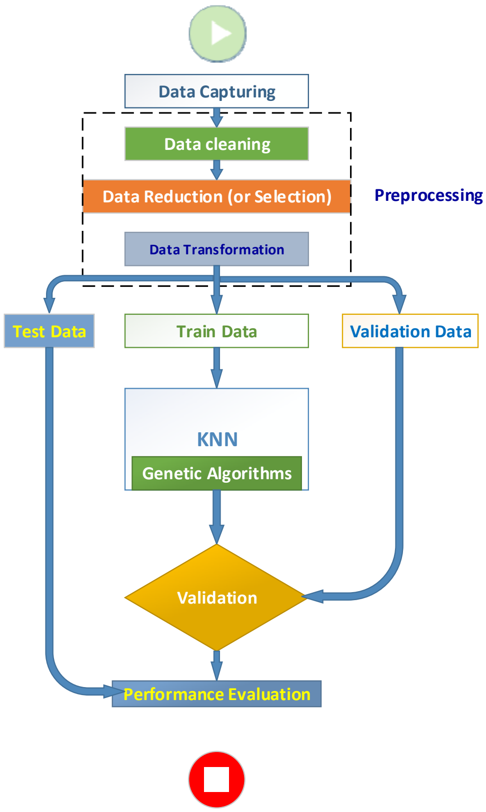

Fig. 1. Flowchart of all the steps.

4 Conclusions

Draglines play major role in the opencast mines particularly in the coal mines. In this case, availability and reliability of dragline is a vital issue that should be considered. Draglines' breakdowns lead to high operation cost (i.e. direct and indirect costs). In order to increase the dragline reliability and its availability as well, predictive maintenance strategies have to be taken into account. In this study, predictive maintenance strategy has been applied, which was tried to predict accurate failure mode in order to carry out preventive maintenance actions. Historical failure data of the dragline which is currently working in a coal mine was obtained and used for making a classifier function (i.e., the mentioned classifier can be applied for unlabeled input which predict the failure mode). Some preprocessing tasks was carried out on the dataset, and the dataset was divided into two parts: training dataset and testing dataset. Finally, hybridized KNN (i.e. KNN has been combined with GAs) applied on the dragline's dataset, which has represented higher accuracy in contrast to other methodologies used by the author on the same dataset. Accuracy of the enhanced KNN model is 82 percent, which is 10 percent more than conventional model of KNN (i.e. the accuracy of the conventional KNN was 72 percent), Table 3. As a result, a better predictor leads to regulation of better proactive maintenance strategy, so planning and scheduling of the preventive maintenance will shaped in an appropriate manner.

Table 3. Applied machine learning methodologies and their accuracies.

Methodology	Regression Analysis (%)
KNN)	72
Enhanced KNN	82

References

1. Ebeling, C.E.: An introduction to reliability and maintainability engineering. Waveland Press Inc. (2010)
2. Gits, C.: Design of maintenance concepts. International Journal of Production Economics 24 (3) (1992). 217-226
3. Herbaty, F.: Handbook of Maintenance Management Cost Effective Practices, 2nd Edition, Noyes Publications, Park Ridge, NJ (1990)
4. Swanson, L.: Linking maintenance strategies to performance. Int. J. Production Economics, (2001). 70. 237–244
5. Taghizadeh Vahed, A. and Demirel, N.: Application of machine learning for dragline failure prediction. The first International Innovative Mining Symposium, (2017)
6. Dindarloo, S. R. and Siami-Irdemoosa, E.: Data mining in mining engineering: results of classification and clustering of shovels failures data. International Journal of Mining, Reclamation and Environment, 31(2), (2017). 105–118.
7. Vagenas, N., Runciman, N. and Clément, R.S.: A methodology for maintenance analysis of mining equipment, Int. J. Surf. Min. Reclam. Environ. 11 (1997). 33–40.
8. Samanta, B., Sarkar, B., and Mukherjee, S.K.: Reliability analysis of shovel machines used in an open cast coal mine, Mineral Res. Eng. 10 (2001). 219–231.
9. Samanta, B., Sarkar B., and Mukherjee, S.K.: Reliability assessment of hydraulic shovel system using fault trees, Institution of Mining and Metallurgy. Trans A Min. Technol. 111 (2002). A129-1135.
10. Roy, S.K., Bhattacharyya, M.M., and Naikan, V.N.A.: Maintainability and reliability analysis of a fleet of shovels. Institution of mining and metallurgy, Trans. A. Min. Technol. 110, (2001). A163-A171.
11. Hall R.A. and Daneshmend, L.K.: Reliability and maintainability models for mobile underground haulage equipment, CIM Bulletin 96 (2013). 159–165.
12. Yuriy, G. and Vayenas, N., Discrete-event simulation of mine equipment systems combined with a reliability assessment model based on genetic algorithms, Int. J. Min. Reclam. Environ. 22 (2008). 70–83.

13. Taghizadeh Vahed, A. and Demirel, N.: Classification of draglines failure type using multilayer perceptron and radial basis function. 25th International Mining Congress and Exhibition of Turkey. (2017)
14. Taghizadeh Vahed, A. and Demirel, N.: Classification of draglines failure types by K-nearest neighbor algorithm.26th International Symposium on Mine Planning and Equipment Selection. (2017)
15. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, Silverman,C.D. R., and Wu,A.Y.: An efficient k-means clustering algorithms: analysis and implementation, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002). 881–892.
16. Townson, P. G., Murthy, D. N., and Gurgenci, H..: Optimization of dragline load. In E. W. Blischke, and D. N. Murthy, Case Studies in Reliability and Maintenance, John Wiley & Sons Inc. (2003).