Interpreting the features for domain experts Using WIDE learning

asd

Abstract

In this paper, authors have proposed an interpretable feature recommendation method for solving sensor signal analytics problem machine maintenance domain. The basic Wide learning based architecture for feature recommendation is out of the scope of discussion in this paper and authors have emphasized on the interpretation of the recommended features and how this human in loop interpretation system can can be used as a prescriptive system. The proposed system was deployed in solving a regression problem for one internal data set of machine maintenance record, as well as a prescriptive system on the popular bearing data-set from NASA prognostic repository. The proposed system is also used to analyze the casuality of a machine maintenance problem.

1. Introduction

Development of a sensor data based descriptive and prescriptive system involves machine learning tasks (Goodfellow et al., 2016) like classification and regression. Any such system development requires the involvement of different stake holders like:
Domain expert: who understand the problem domain, like doctor in case of health care
Signal processing (SP) expert: who can suggest the suitable signal processing algorithm and corresponding parameters
Machine Learning (ML) expert: who can design the classifier or the regression model
Coder or developer: who can construct a deplorable solution
Now the problem of developing such a system is that each of the stake holders speaks their own languages and that is often difficult to understand for others. So, while we were trying to make a classifier or a regression problem in health-care or machine maintenance domain we found that making such a system so that any domain expert can explain and understand requires the following steps:

- Domain expert explains the goal of the problem to the SP and ML person
- SP expert provides a list of algorithms that can be used as features for this problem
- ML expert recommends the optimal feature set based on the available data
- SP expert tunes the parameters of those algorithms (like window size for an FFT algorithm), and the ML expert tunes the hyper parameters for the ML task.
- Recommended feature set is presented to DE for validation and verification
- Final system is deployed

The penultimate step is difficult in a Deep Learning based approach though some works (Kim et al., 2015) can be found in this area. In this paper we are going to present how to interpret the recommended features by a wide learning approach as presented in (Banerjee et al., 2016) can be used and verified by a domain expert to make a robust system. This proposed human in loop interpretable feature recommendation system can be used in a prescriptive manner also.

Wide learning is a new term in the ML community with very less number of related work can be found on it. Initially this term was used in (Pandey & Dukkipati, 2014) in 2014. Almost similar two methods has been shown in (Cheng et al., 2016) and (Banerjee et al., 2016). In the latter, authors has shown the proposed wide architecture and its applicability in health care. In this paper we are going to present the method of interpretation of the recommended features in domain expert understandable format and its advantages.

2. Proposed Method

The proposed machine learning based critical feature set recommendation framework (Banerjee et al., 2016) as shown in Figure 1 accepts a set of input sensor data. The input feature set is a combination of derived features obtained by transforming the raw time series data in diverse domains such as time, frequency and time-frequency domains. Any feature set recommendation framework would in general recommend only the corresponding indices of the relevant features. Such feature identification mechanism is sufficient to trace back the recommended features from the generated feature pool. However such practice does not leave any room for any further refinement of the recommendation through incorporation of domain expert

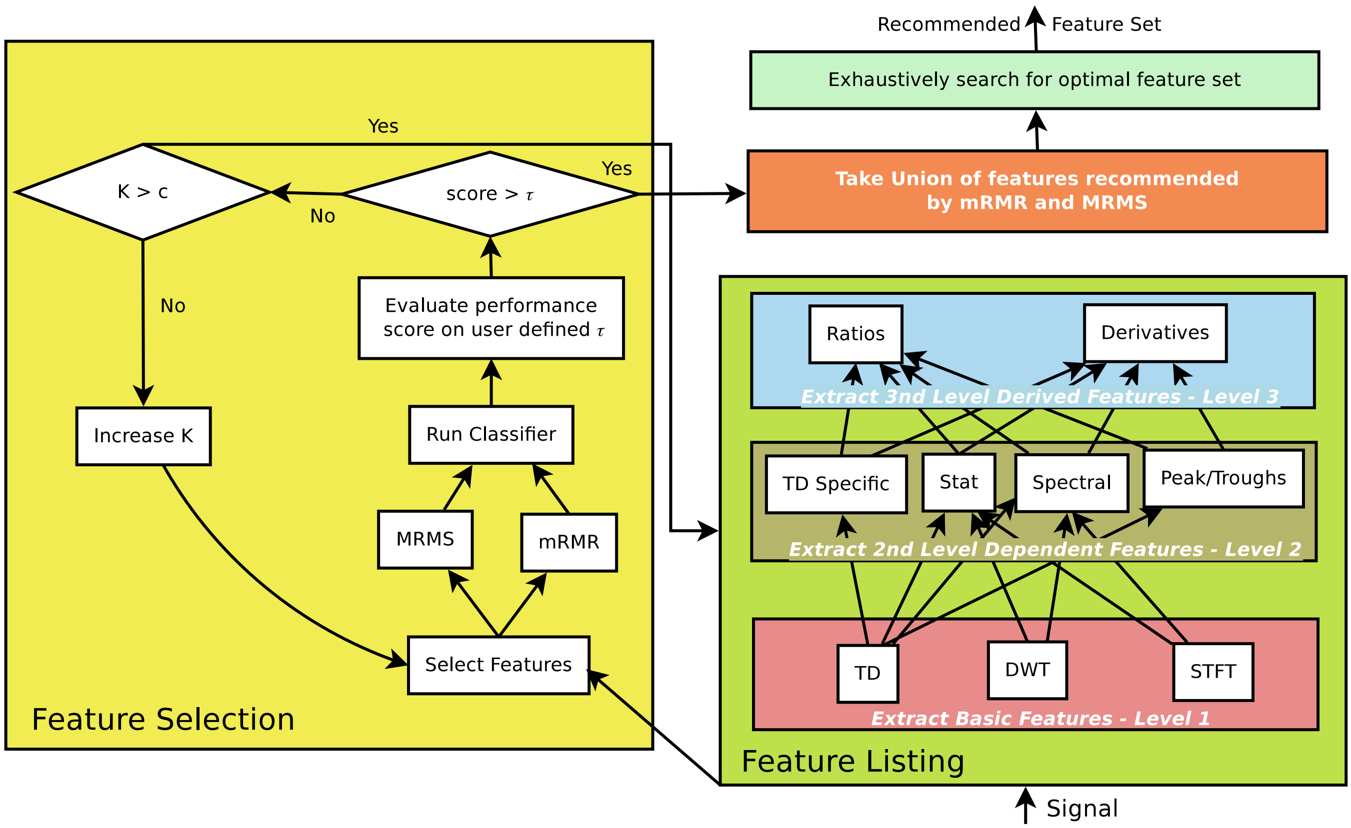

Figure 1. Proposed Method of Feature Recommendation

opinion, which may be of paramount importance, especially in safety-critical domains such as machine pyrognostics. In order to address the issue, the proposed feature recommendation framework consists of a feature interpretation module. The feature interpretation module accepts the recommended feature indices as input and returns any granular information that can be obtained by analysis its genesis methodology. While feature values were derived to form input derived feature pool, a mapping table is iteratively maintained that stores the details of the steps though which each indexed feature value is being generated. The steps of each indexed value generation would typically include information regarding domain of transformation, transformation technique, location of the feature value in the transformed vector, etc. Utilizing such a mapping table, the steps followed to arrive at the recommended feature indices can be traced back reliably. Once the complete information regarding the genesis of each recommended feature is fetched, such information is again passed on another module that is designed to return further granular data that may further aid in proper representation of the recommended features to the domain experts.

Similarly on the multi-sensor data-set D4 (3.1), the axis load is required to be predicted by using rest of the sensor data like relative position, feed data, speed. Here, we have applied MIC (Reshef et al., 2011) to find the mutual information content and the ρ^2 parameter between the observed load data and rest of sensor data. These highly related sensors are then verified by the domain experts and the modified list of sensors is taken as the input for predicting the load. Now, similar method is applied on each of the sensor data to identify the suitable features out of them.

3. Experiments

3.1. Datasets

The experiment is performed on two well known openly available 1-D sensor signal data sets, the specification of each being described in table 2 and one internal titanium machinery data-set:

(i) D1 and D2: NASA Bearing[1] data set contains 4 bearing data instances each having 984 records, while the first bearing fails after 700th record among the total 984 recorded readings. The last two readings are not considered due to

[1]NASA Bearing Set 3 at https://ti.arc.nasa.gov/tech/dash/pcoe/ prognostic-data-repository/publications/#bearing

Table 1. Description of data sets used for experiments

Datasets (D)	Total No. of Instances	Class-0 No. of Instances	Class-1 No. of Instances	No. of Samples	Sampling Rate (Hz)	Time Window Size (seconds)
D1: NASA All	3932	282	3650	20480	20,000	0.5
D2: NASA Subset	647	282	365	20480	20,000	0.5
D3: Mobifall	258	132	126	230	50	1

* the number of samples per data instance varied in the range of 10612 to 71332, hence truncated for uniformity

presence of missing values. So, we get 282 ‘bad bearing’ (class 0) records as ground truth for a class, while the rest 700 of the first bearing and 982 values each from rest 3 bearings that do not fail form the ‘good bearing’ class 1. To handle data unbalancing and see its effects, we have created two data-sets: D1: that contains the full dataset instances, D2: that contains a randomly selected small subset of the ‘good bearing’ instances along with all the ‘bad bearing’ instances.

(ii) D3: Mobifall[2] data set is a popular fall detection data-set created by volunteers aged 22-47 years. Although the data-set contains various levels of activities, however we have portioned the data-set into ‘fall’ (class 0) and ‘not fall’ (class 1), in order to restrict to binary classification.

(iii) D4: Titan PED data-set is used for predicting the axis load or spindle load at any instance of time depending on the other sensor data. This data set includes 27 different sensor data captured at 1 Hz sampling rate but of total duration of 26 days. So the total amount of data is nearly 21,00,000. This data set is not a public one.

3.2. Results and Analysis

Tables [2-4] show some of the sample feature sets obtained for the classification tasks in the respective data-sets. This listing of features along with ranges of values obtained helps the domain experts who maps the obtained values to the physical world and the problem domain at hand, so that causal analysis of the problem can be made and deeper insights can be gained.

Table 2. Recommended features for D1 (win=0.5s)

Sl.	Feature description
1	Difference of root mean square values of windowed DWT coefficients
2	Difference of Standard deviation values of windowed DWT coefficients
3	Flux of spectral coefficients
4	Mean of STFT coefficients
5	Root mean square of STFT coefficient
6	Variance of STFT coefficients

[2]http://www.bmi.teicrete.gr/index.php/research/mobiact

Table 3. Recommended features for NASA Bearing data-set, window size = 0.5 sec

Sl No.	Feature description	
1	STFT	Frequency: 1851.1851 Hz Frequency: 1853.1853 Hz Frequency: 1153.1153 Hz Frequency: 1837.1837 Hz Frequency: 1845.1845 Hz
2	Difference of standard deviation values of windowed discrete wavelet transform (DWT) coefficients	
3	Standard deviation of STFT coefficients	

Table 4. Recommended features for NASA Bearing data-set, window size = 1 sec

Sl No.	Feature description	
1	STFT	Frequency: 1613.5807 Hz Frequency: 1829.5915 Hz Frequency: 1830.5915 Hz Frequency: 1837.5919 Hz
2	Kurtosis of DWT coefficients	
3	Standard deviation of DWT coefficients	
4	Standard deviation of STFT coefficients	
5	Zero crossing of DWT coefficients	

3.3. Physical interpretation

Traditionally feature selection method is a manual effort where a domain expert identifies some features from his domain expertise and experience and then plot them for both the classes to conclude the feature is relevant or not. Due to lack of space, we have selected the NASA Bearing Data-set for interpretation analysis in this paper. Similar interpretation were also found in the other data-sets. Our proposed automated feature recommendation method also predicts the features at 14 Hz (DWT features) as well as in the even (6^{th}) harmonic space of the fundamental frequencies of the bearings rotating elements as reported below. Thereby the recommended features can be mapped to the

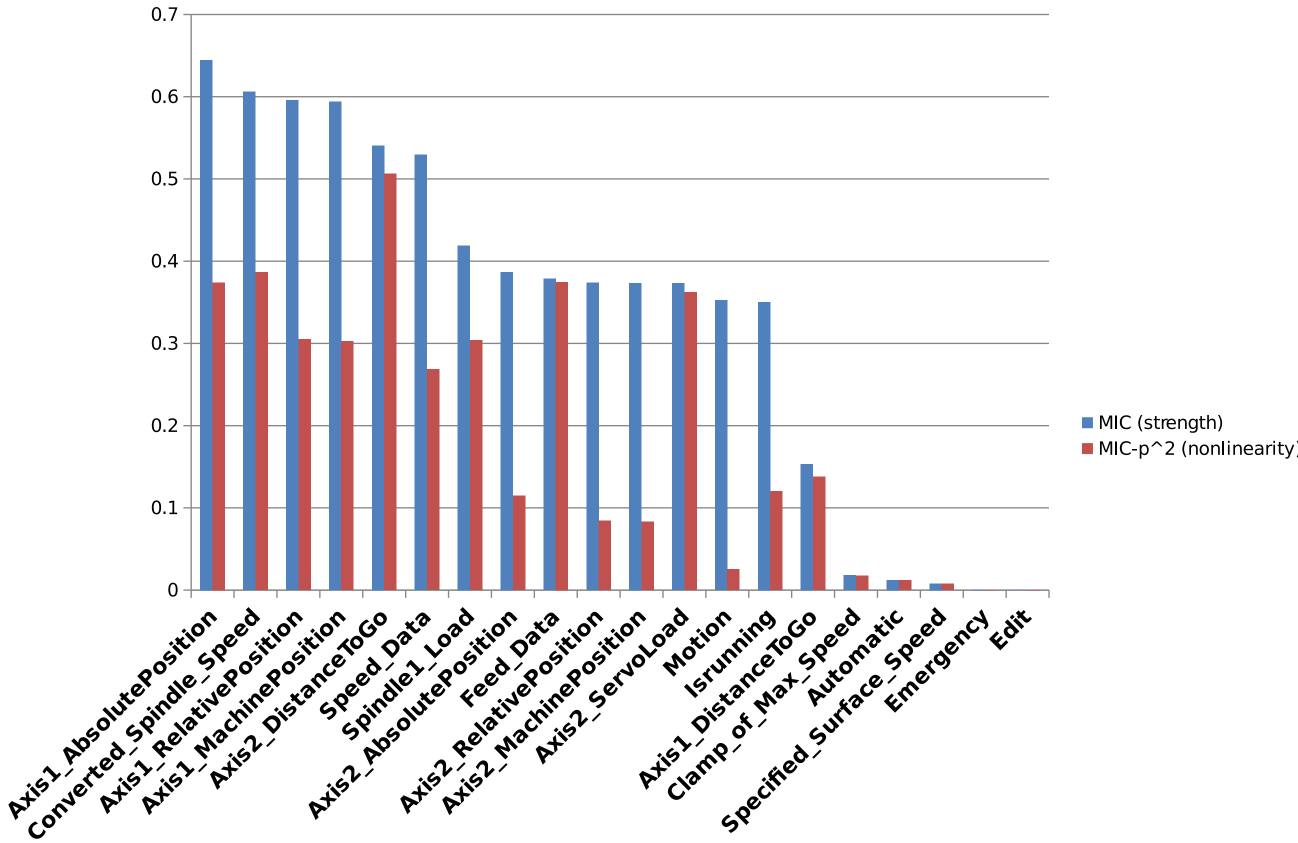

Figure 2. Sensors Affecting Axis 1 Servo Load for D4

physical world elements for further introspection and analysis of the allied domain expert. Now the bearing Physics suggests fundamental frequencies as:

Outer Race Frequency =236.4 Hz
Inner Race Frequency =296.9 Hz
Rolling Element Frequency =279.8 Hz
Shaft Frequency = 33.33 Hz
Bearing Cage Frequency = 14.7 Hz

Further we can suggest the manufacturer the reason of exact failure obtained from the physical interpretation of the recommended features. In this case we can predict that the error may arise because of all possible reasons other than the problem in Shaft frequency.

Figure 2 shows (for the D4 data-set) that the sensors that are correlated with axis 1 servo load are feed data, spindle load, speed data, axis 1 positions, axis 2 positions, but, as per theory, orthogonal axes can not have impact on each other. Thus the axis 2 position or load should not have any impact on axis 1 load, subsequently it was validated by the domain experts that these situation occurs only when the cutting is performed at a 45^o angle to the axis. This is exactly used for the operation of rough cut on the outer diameter. Thus, the interpretable features can also assist the domain expert to get an insight about the mechanical process performed.

4. Conclusions

In this paper, we have presented a method to recommend features using Wide Learning technique that can be interpreted to the domain experts. In case of NASA bearing data-set, this interpretation helps them to analyze the cause of failure. On the other hand, the recommended sensors for a regression problem assist the production manager (domain expert) of a mechanical plant by providing an insight about the mechanical process. So this proposes system may not be optimal but practical for making any deplorable machine learning based sensor signal analytics prognostic system.

References

Banerjee, Snehasis, Chattopadhyay, Tanushyam, Biswas, Swagata, Banerjee, Rohan, Choudhury, Anirban Dutta, and Pal, Arpan. Towards wide learning: Experiments in healthcare. *Machine Learning for Health, NIPS 2016, arXiv preprint arXiv:1612.05730*, 2016.

Cheng, Heng-Tze, Koc, Levent, Harmsen, Jeremiah, Shaked, Tal, Chandra, Tushar, Aradhye, Hrishi, Anderson, Glen, Corrado, Greg, Chai, Wei, Ispir, Mustafa, et al. Wide & deep learning for recommender systems. In *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*, pp. 7–10. ACM, 2016.

Goodfellow, Ian, Bengio, Yoshua, and Courville, Aaron. *Deep Learning*. MIT Press, 2016. http://www.deeplearningbook.org.

Kim, Been, Doshi-Velez, Finale, and Shah, Julie A. Mind the Gap: A generative approach to interpretable feature selection and extraction. In *Advances in Neural Information Processing Systems*, 2015.

Pandey, Gaurav and Dukkipati, Ambedkar. To go deep or wide in learning? In *AISTATS*, pp. 724–732, 2014.

Reshef, David N, Reshef, Yakir A, Finucane, Hilary K, Grossman, Sharon R, McVean, Gilean, Turnbaugh, Peter J, Lander, Eric S, Mitzenmacher, Michael, and Sabeti, Pardis C. Detecting novel associations in large data sets. *science*, 334(6062):1518–1524, 2011.