Automation of Feature Engineering for IoT Analytics

Snehasis Banerjee
TCS Research & Innovation
Kolkata, West Bengal
snehasis.banerjee@tcs.com

Arpan Pal
TCS Research & Innovation
Kolkata, West Bengal
arpan.pal@tcs.com

Tanushyam Chattopadhyay
TCS Research & Innovation
Kolkata, West Bengal
t.chattopadhyay@tcs.com

Utpal Garain
Indian Statistical Institute
Kolkata, West Bengal
utpal@isical.ac.in

ABSTRACT

This paper presents an approach for automation of interpretable feature selection for Internet Of Things Analytics (IoTA) using machine learning (ML) techniques. Authors have conducted a survey over different people involved in different IoTA based application development tasks. The survey reveals that feature selection is the most time consuming and niche skill demanding part of the entire workflow. This paper shows how feature selection is successfully automated without sacrificing the decision making accuracy and thereby reducing the project completion time and cost of hiring expensive resources. Several pattern recognition principles and state of art (SoA) ML techniques are followed to design the overall approach for the proposed automation. Three data sets are considered to establish the proof-of-concept. Experimental results show that the proposed automation is able to reduce the time for feature selection to 2 days instead of $4 - 6$ months which would have been required in absence of the automation. This reduction in time is achieved without any sacrifice in the accuracy of the decision making process. Proposed method is also compared against Multi Layer Perceptron (MLP) model as most of the state of the art works on IoTA uses MLP based Deep Learning. Moreover the feature selection method is compared against SoA feature reduction technique namely Principal Component Analysis (PCA) and its variants. The results obtained show that the proposed method is effective.

CCS CONCEPTS

• **Information systems** → **Data analytics**; • **Social and professional topics** → **Automation**;

KEYWORDS

Feature Engineering, Sensor Signal Analytics, Information Processing on Sensor Data, IoT Analytics

1 INTRODUCTION

Internet of Things analytics (IoTA) involves a lot of applications [3] [25] [27] [32] in health, wellness, sustainability, transportation, smart city administration, and urban health while deploying any industrial sensor systems. Some of such applications in health care like blood pressure classification, Coronary Artery Disease (CAD) classification or machine health classification (into good condition or bad condition) involves classification as the machine learning (ML) task [1]. It is required to identify some characters, known as feature, of the input signal in order to classify it using popular ML methods like Support Vector Machine (SVM) or Random Forest (RF). But the method of identifying a set of suitable features for the classifier is a time and cost absorbing task as it needs mostly domain expertise of the signal processing and IoT expert. Hiring people with such niche skill set is costly. Another problem is that some times the features listed by the domain expert are not relevant and if the data set includes less number of instances with a lot of features, the curse of dimensionality comes into play. This results into the need for feature reduction which can be accomplished into two different ways namely feature dimension reduction [18] or reducing the number of features by feature selection [20], [21]. These two steps of listing features and reducing the number of features is known as Feature Engineering (FE) as a whole. Another approach to get rid of this issue of FE is to use some Deep Learning (DL) techniques like Multi Layer Perceptron (MLP) with fully connected layers or Convolutional Neural Network (CNN) which can automatically identify suitable feature set for the classification. But DL methods makes an assumption of having a huge number of annotated data-set to train the system which is mostly not available for IoT tasks. DL remains as a very popular technique for computer vision and natural language processing (NLP) and speech recognition. Another popular method is using Principal Component Analysis (PCA) to derive features that have high correlation to the class. However, in both cases interpretation of the features is an issue, and the domain expert cannot take advantage to carry out causal analysis for a given problem, once suitable features are identified. For example, a cardiologist can justify that the PPG (photoplethysmogram) signals should contain a signal at a particular frequency range to classify a CAD signal. But neither DL methods nor the feature dimension reduction techniques can interpret the features as the feature space do not directly relate to physical world. So in this paper authors have proposed a method of FE for classification using ML techniques where a Feature Listing Database is
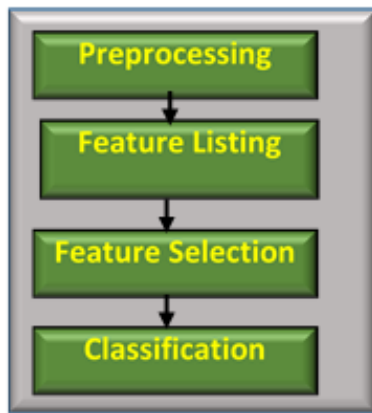
Figure 1: High level workflow for IoT Analytics

Figure 2: Analysis of pain areas in IoT Analytics

maintained that becomes useful once good features are identified. The contributions of the proposed work is enlisted:

1. A survey has been carried out to understand the workflow of a typical IoT System development and identify the stages that demand maximum effort and niche skills, thereby costing more money and time.

2. The work proposes a method for automation of FE so that the suitable features for classification can be obtained in less time, less cost and the recommended features are physically interpretable to assist the domain expert in making casual analysis.

3. A WIDE architecture for feature listing is proposed that starts to use basic features in the basic layer and extracts more derived features on the higher layer. This layered architecture also helps to reduce the complexity as the features for different layers are computed iteratively. Features of higher layer are not computed once the desired performance is obtained at a lower layer. The performance is measured using different metrics like sensitivity, specificity, F-score.

The proposed method is tested against annotated data sets on man and machine predictive analytics, the use cases of which demand interpretation of features. The used data sets contains (i) NASA's bearing data set that includes good and bad bearing data, (ii) Emotion dataset to classify emotion of users, and (iii) MIMIC-II data set to classify high and low blood pressure from PPG signal. The accuracy obtained by applying some classifier on the recommended feature set are compared against the performance obtained by applying the same classifier on the SoA features reported in literature to solve these point problems. Some of these point problems were also tried in our research lab and hence we were able to compare the development time. We have also compared the proposed method against DL techniques like many layered Multi Layer Perceptron (MLP) method as most of the literatures have used MLP for sensor signal processing [22]. The proposed feature selection technique is compared against a popular feature dimension reduction technique namely Principal Component Analysis (PCA). Finally, through an example we have also given a physical interpretation of the recommended features so that the recommended features can be validated against the physical phenomenon.
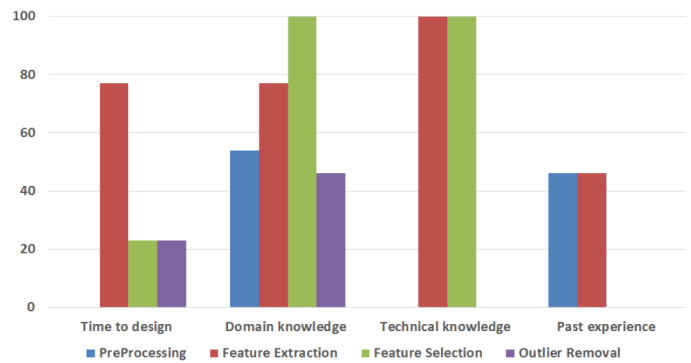
Section 2 discusses about the survey conducted to analyse typical IoTA tasks. Section 3 describes the process and corresponding system architecture. Section 4 discusses the various experiments carried out. Finally section 5 concludes the paper.

2 BACKGROUND

We made a survey over the associates within our organization involved in the IoTA to identify the pain areas of the application developers. In this survey, we have surveyed seven (7) projects which include ninety-five (95) employees (female: 35 and male: 60). The associates under each project form a team. Each team consists of members among whom 30% having more than 10 years of experience and Master's or higher academic qualification and each team is mostly led by a Ph.D. person. Sixty per cent (60%) associates in each team have 5-10 years of experience and expertise in signal processing and the remaining 10% are developers mostly equipped with good coding skill in C/Java/Python but don't have in depth signal processing, IoT or domain knowledge.

Seven projects involved in this survey are (i) Motion capture, object recognition and rendering of articulated objects from Kinect based skeleton data, (ii) Object classification and recognition from Hyper-spectral sensor, multi-spectral sensors (Landsat, digital globe WorldView etc.) (iii) Bio sensors based object recognition, (iv) Outdoor camera based object recognition. (v) Detecting heart rate and blood pressure from PPG, (vi) Thermal imaging based object classification, and (vii) EEG (electroencephalogram) based cognitive load classification. For each project, the project goal, sensors used, and the steps followed to achieve the goal are shown in Table 1. These steps can be combined to construct a superset of steps followed in any sensor signal processing based IoTA as shown in Figure 1. The steps used in this figure are briefly described here:

- Pre-Processing: This module aims to filter out the noises in a signal. It also helps to remove the outlier in the signals. Different pre-processing techniques are available in the literature.
- Feature Listing: Any signal is represented by some features. For example, time domain or frequency domain analysis normally give many features for an input signal.
- Feature Reduction: The possible number of features for a sensor signal is very huge and so it is required to reduce the number of features. Feature reduction techniques should be

Sensor	Signal Type	Goal	Steps
Camera, mobile, Kinect,	Stochastic	Object and Scene Recognition	Sampling→Calibration→Object Classification
Hyper spectral sensors,	NA	Object recognition	Image acquisition→Pre-Processing → Calibration→ Feature listing→ Supervised or Unsupervised learning
Biosensors like ECG, PPG, EEG, EMG, Camera,	NA	Object recognition	Pre-Processing →Feature Listing →Classification
Camera	NA	Object Recognition	Pre-Processing →Feature Listing →Classification
Microphone, Camera, IR Camera	Periodic, Stationary, Non-stationary	Physiological condition estimation	Pre-processing→Signal Quality Checker →Feature Listing→Feature Selection→Computation
Thermal camera, LED/PhotoDiode	Multiple frequencies	Classification, Estimation calculation, imaging beamforming, denoising	Filtering→ Denoising → features detection →classification
EEG, GSR, SpO2, Camera, Eye tracker	Aperiodic, Non-stationary	Classification and clustering, estimation	Pre-processing→Noisy window removal →Feature Listing→Feature Selection →Computation

Table 1: Steps involved in IoT sensor signal processing and analytics for different applications

selected with the following points kept in mind: (i) Interaction among the features, (ii) Interaction among the features with machine learning tool, (iii) Goal to achieve which can be obtained from an annotated data set, (iv) Interpretability of the reduced feature.

- **Classification**: The technique takes the experience which is a representation of the feature set and the label as input and optimizes some parameters like accuracy, sensitivity to accomplish a goal like classification.

The analysis as shown in Figure 2 clearly reveals that most of the associates under survey express that feature listing and feature selection or feature reduction requires the maximum domain knowledge and technical knowledge. The other steps like pre-processing (including noise cleaning and outlier removal) takes the highest time to design. As the pre-processing algorithms are well established and even source code for them are available in the web, we propose to provide the developer with a list of source codes for all those different state of the art preprocessing techniques so that the developer can try any of the methods on their data set and select the best one among them. One such method is described in [4]. But feature selection is a difficult process. The input to our proposed automated system is the labeled data set of the problem to solve. Another aspect of the study was to list down the features, commonly reported in the related literature, that are used in IoTA applications and finally recommend the features to be used for the given use case.

3 PROPOSED METHOD

In this section, we are going to discuss how to automate the feature selection step which is the most time consuming and domain specific expertise dependent step. We plan to use state of the art pattern recognition techniques to achieve this. The key idea behind this is as follows.

Exploration of all possible features: Finding the right set of features for any pattern classification task is still an unsolved problem. Great amount of research is still concentrating on this problem [4]. In our

approach, we, at first, have studied the existing literature [6] - [15] to have a more or less exhaustive list of features which have been used by different researchers for different classification tasks on different sensor data sets.

We have organized these reported features in an hierarchical manner as shown in Figure 3. The features reported in literature can be mainly classified in three types: (i) time domain features (TD) (ii) Fourier transformation based features (FD) (iii) Discrete Wavelet transformation based features (DWT). A challenge of using features from Wavelet Transform is the appropriate selection of a suitable mother wavelet. This is because more than 100 different types of mother wavelets are reported in literature, and a ready made automatic wavelet selection method or tool do not exist. Hence, considering the the different feature types, it is possible to get a large number (say, N) of features (including coefficients of the transform domain) from the sensor signals. This results in $2^N - 1$ possible combinations of features.

Feature Selection: The value of N could be as high as $6,000,000$ as explained later in this section. Therefore, in order to find the optimal feature set, exploitation of all such $2^N - 1$ combinations is practically infeasible as this would require testing of machine performance for each combination. This can be achieved mostly by the domain experts from their experience and understanding about the physical phenomena. Typically feature selection method takes 4-6 months for such problems. In pattern recognition literature, there are several feature reduction techniques available which can give a reduced set of features giving optimum performance [24], [26], [28]. Principle Component Analysis (PCA) is one such commonly used technique for feature reduction. Actually, what PCA does is feature extraction, i.e. the resultant features are not interpretable. However, we need interpretable features for our tasks. For instance, in health analytics, if we know that certain features are very strong in predicting CAD (coronary artery disease) patients, the doctors need to know what these features are. Therefore, in order to reduce the number of feature to make the feature recommendation task
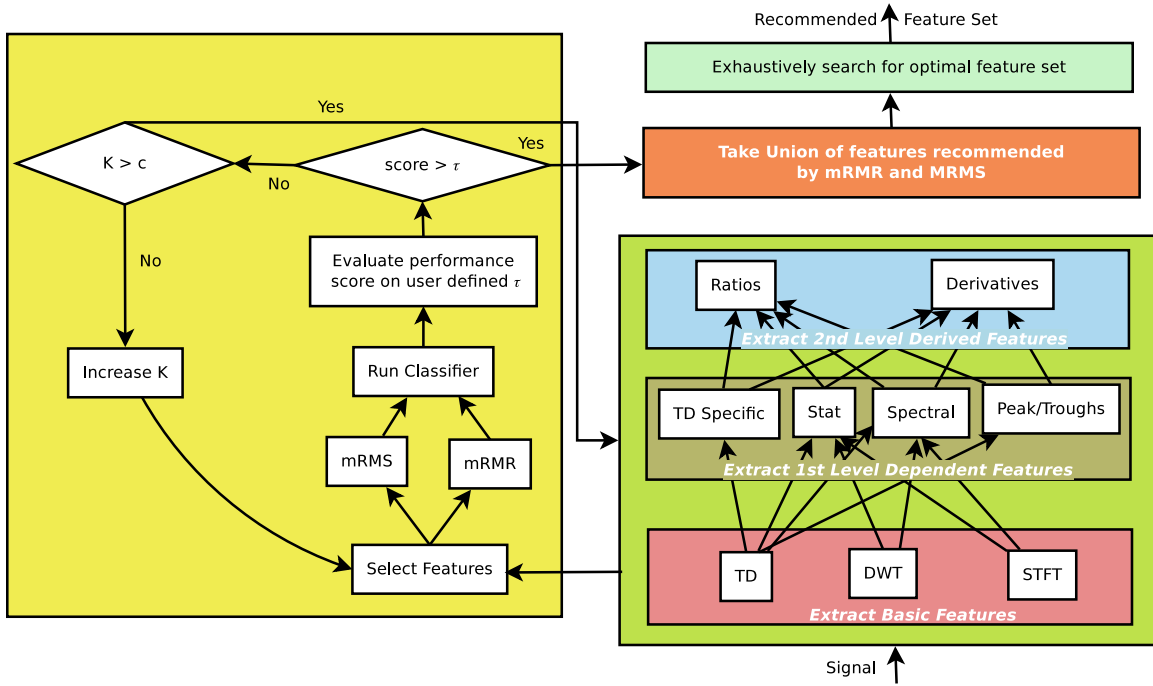
Figure 3: Proposed Method of Feature Recommendation

tractable we follow feature selection instead of feature extraction as feature selection methods do not alter the original features.

In our method, we followed an iterative feature selection where k-features are selected at each iteration and system performance (e.g. classification accuracy) is checked for this feature set. If the selected feature set results in *expected* performance, that feature set is recommended and process halts. Else, another set of k-features is selected in the succeeding iteration and similar steps are carried out. For checking the classification accuracy, we choose SVM-(support vector machines) based classification with different kernels. Different values of k are fitted in progressive increments to get a good result. For a given value of k, features are selected using two techniques namely, MRMS [19] and mRMR [21]. The reason for choosing these two techniques is their impressive efficiency in feature selection as demonstrated in [20]. Details on mRMR and MRMS used for this framework can be found in [31].

3.1 The proposed Architecture for Feature Recommendation

The proposed method is applicable in extracting and recommending features from 1D sensor signals. The sensor signal is subjected to time domain, frequency domain and time frequency domain analysis to extract features which for example counts to almost 12 million features for input data of size 20,000. Let us assume a 1D signal of length n sampled at fs frequency. So after subtracting the mean from the signal we get the same n number of TD features. Now the entire signal is splitted into multiple overlapping windows to extract the FD features. So if we consider a 1sec window with 50% overlap then we get (n*2)/fs number of windows. Considering the STFT window

size to be 256, number of FD features can be obtained is (256*n*2)/fs. Similarly by applying DWT we obtain another n features. So extracting first level features we get n+n+(512*n)/fs ≈ 3n number of features. The level two features are derived from each window of the level one features. Thus in level 2 around 20*(3n/fs)*2≈120n features are extracted. Therefore around 3n+120n=123n features are extracted at the end of layer2. Similarly in 3rd layer features are derived from each window of 2nd level features which is around 2*(120n/fs)*2 ≈ 480n. So after 3rd layer the number of features extracted is of the order 123n+480n ≈ 600n. Now in dataset 1, value of n is nearly 20,000. So number of features extracted after 3rd layer is nearly 600X20,000≈ 12,000,000. Level two features are applied on top of each of the level one features. Thus after level 2 this number of features extracted is becoming nearly of the order of 4n*4=16*n. Similarly after 3rd layer this number becomes of the order or 30*n. Now in data set 1 this n is nearly 20,000. So after 3rd layer of feature extraction total number of feature extracted from the signal is nearly $30 \times 20,000 \approx 6,00,000$. So, it is not possible to apply any exhaustive search on top of this 12,000,000 features so that one can optimize the performance metrics say sensitivity or specificity. The feature selection architecture used in this paper is shown in Figure 3. The proposed method is as described below:

Input: Time domain signal
Output: Recommended feature list

- Compute the mean subtracted time domain signal by subtracting the mean from the original signal. This entire signal is considered as the time domain (TD) feature.
- Compute the short term Fourier transform(STFT) for each window where window size is provided as an user input.

- The given signal is subjected to four level Discrete Wavelet Transform (DWT) using the most optimal mother wavelet. The optimal mother wavelet is selected using the property of maximum energy to entropy ratio [23].
- Take the union of the all three TD, FD, and DWT features. Let this set be defined as FL_1
- Apply minimum redundancy maximum relevance (mRMR) and Maximum Relevance Maximum Significance (MRMS) for FL_1. Each of these methods recommends a different feature set say x and y.
- Iteratively select the number of features to be recommended by each of these methods so that the recommended features applied on a classifier can exceed the performance metric score defined by the user, say τ.
- Take union of the features recommended by both mRMR and MRMS methods. Let the recommended feature set be z where $z = x \cup y$.
- Features from level 2 (FL_2) and level 3 (FL_3) are extracted using the algorithms shown in Figure 3.
- In level 2 we have extracted spectral features, statistical features, and peak-trough features. Spectral features used in the proposed method are Spectral Centroid, Spectral Crest Factor, Spectral Decrease, Spectral Flatness, Spectral Flux, Spectral Kurtosis, Spectral Rolloff, Spectral Skewness, Spectral Slope and Spectral Spread which are computed for each window of mean subtracted signal. Statistical Features used here are mean, variance, standard deviation, root mean square, skewness, kurtosis. Average peak amplitude, average trough amplitude, average peak to peak distance and average trough to trough distance are computed as the peak-trough features. FL_3 includes different ratios and derivatives of the FL_2 features.
- These FL_2 and FL_3 features are also reduced using mRMR and MRMS similarly.
- Once the feature space is reduced, exhaustively generate all possible combinations of features and apply classifier with different parameters.
- Apply Support Vector Machines (SVM) classifiers with different kernels namely (i) linear, (ii) radial basis function (RBF), (iii) sigmoid, and (iv) polynomial and each of the kernels is tested with different parameter values.
- Thus construct the model for classification by selecting the feature list and the SVM kernel for which the recognition accuracy maximizes. SVM is selected as a classifier as the problems at hand are binary classification where SVM is known to excel and converge quickly.

4 EXPERIMENTS

4.1 Dataset

The experiment is performed on 3 data sets out of which two data set are publicly available i.e. the Bearing dataset from the Prognostics Data Repository and second one is the MIMIC data-set for classifying blood pressure. The third data-set is internal to our organization and this aims to classify the emotion of a person into three classes namely happy, sad, and neutral. As there is the rest two data sets are two class classification problem, we have used our own

data set to prove that our proposed system works on multi-class use-cases also. The first data set used as a machine automation case study. The rest two data sets are used in health care and emotion detection. 2nd and 3rd data sets are based on photoplethysmogram (PPG) sensor signal. The bearing data set (named IMS-Rexnord) consists of three datasets describing a test to failure experiment. Each dataset comprises of several files each of which has a record of 1 sec vibration signal snapshot which are recorded at specific intervals. The sampling rate is 20KHz, recording interval is 10 min and there are 20480 data points in each of the file. In our experiment we have used the second dataset. It has 984 files. Each individual file holds the record of 4 channels representing 4 bearings where the first bearing eventually turns faulty due to outer race failure.

The experiment is performed in two ways. In the first case only the bearing 1 data is considered. State of the art [30] shows that the bearing 1 starts degrading after the 700th point where each file of 1 minute readings is denoted as a point. So here we have formulated a two class classification problem where the first 700 files are considered to be healthy and the rest 282 files are considered to be faulty (the last 2 files are discarded due to presence of many noisy signal values).

In the second case we have considered all the bearings to formulate the two class classification problem. The bearing dataset has 984 files. Each file has record of 4 bearings. Therefore the total number of good bearing samples is 3652 (1st 700 samples of 1st bearing and 984 samples of each of the three other bearings) and the number of faulty bearing samples is 282 (last 282 files of 1st bearing). To avoid any biasness we have segregated the data into 5 folds. Each of the dataset consisting of 282 faulty bearing samples and 730 good bearing samples.

The second dataset which is used to classify the blood pressure into high and low, records the PPG signal of 118 subjects from Bangalore and Gujarat. Among the 118 subjects, 15 subjects have high systolic blood pressure and 103 have low systolic blood pressure. To avoid any biasness due to imbalance in the dataset, the dataset is segregated into three datasets. Each of the dataset consists of the PPG signal record of 49 subjects, 15 subjects having high systolic blood pressure and 34 subjects having low systolic blood pressure. The sampling rate is 60 Hz.

The third dataset (used to classify the emotion into happy and sad) records the fingertip pulse oximeter data of 33 healthy subjects (13F and 20M) with average age 27. No two emotion elicitation video was shown to a subject in one single day. The Pulse Oximeter is used to detect and record the PPG signal. We used standard video stimuli which itself served as ground-truth and the rigorous experimentation procedure ensured that the time synchronization error between the stimuli and recorded physiological data is always less than 1sec.

Table 2 shows the performance of proposed method against state of the art results, proving the efficacy of the method. It is to be noted that time and cost is a huge factor when comparing the performance. In many cases, a trade-off fits the solution well when performance gain is not much when automotion is compared to the manual effort.

Table 2: Comparison of proposed method against state of the art for different data sets

Data set	SOA no of features	Recommended no of features	SOA accuracy	Recommended accuracy	SOA effort	Recommended effort
NASA data set	15	10	99.38	100	6 months	2 days
Emotion data set	16	11	82.3	90.91	4 months	2 days
BP data set	23	15	79.5	87.8	6 months	2 days

Table 3: Comparison of proposed method against 5-layered MLP

Data set	Activation Function	No. of Epoch	MLP Accuracy	SoA Accuracy	Proposed Method Accuracy
NASA data set	softmax	5	71.00	99.38	100
Emotion data set	relu	15	50.0	82.3	90.91
BP data set	softmax	10	50.0	79.5	87.8

Table 4: Comparison of proposed method against PCA

Data set	Dimensionality Reduction Algorithm	No. of Principal Components	PCA Accuracy	SoA Accuracy	Proposed Method Accuracy
NASA data set	svd	5	94.00	99.38	100
Emotion data set	svd	10	50.0	82.3	90.91
BP data set	eig	5	62.50	79.5	87.8

4.2 Comparison against MLP

Learning representations from raw data is an emerging field. Theano has been used as the software platform to carry out experiments.

Different number of layers (based on standard thumb rules of input size) for Multi-Layer Perceptron (MLP) with Dropout feature has been tried out so that automatic feature learning can take place. Different activation functions like tan hyperbolic, softmax, sigmoid, rectified linear unit (relu) etc. has been investigated at different layer levels to get a suitable architecture for classification task for the given problems.

Table 3 lists the configurations obtained for a 5 layered MLP for which the best performance was obtained. It is seen that MLP techniques fail in comparison to proposed method as well as state of the art. Also, MLP has some drawbacks when put into the IoT spectrum - a) need for a lot of data for training and b) non-interpretable feature extraction.

4.3 Comparison against PCA

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to derive principal components representative of the features under consideration. This has two outcomes: a) dimension of feature space can be reduced by selecting most prominent principal components b) derived features is supposed to represent the feature space better. Gaussian kernel is used for SVM based classification on the principal component features extracted. Various dimension reduction techniques has been used like Alternating Least Squares (als), Eigen Value Decomposition (eig) and the traditional Singular Value Decomposition (svd). Table 4 lists the configurations leading to the best results. It is apparent from the table that PCA based methods do not perform well in comparison to proposed method. Another drawback of PCA is the derived features are not interpretable.

5 CONCLUSION

This paper has carried out a survey to understand the workflow time and cost needs of a typical IoTA task. In the survey, Feature Engineering came out as the most taxing task among all endeavors that comprise IotA. To tackle the challenge, a system was built to automate this sub-task of IoTA. The system has been tested on three datasets and has been found to give good results when compared to state of the art. The proposed method is compared with PCA and MLP, which are two divergent paths of feature engineering. Feature interpretation is another notable aspect of the system. Future work will look into automation of parameter tuning and selection of machine learning models. Automatic window selection [2] for a given dataset is also planned. Integration of knowledge bases and use of reasoning [34] [33] [29] for ease of interpreting features and causality analysis is also planned.

REFERENCES

[1] Ian Goodfellow, Yoshua Bengio and Aaron Courville *Deep Learning*, MIT Press, 2016.
[2] Banerjee, Snehasis, et al. *Windowing mechanisms for web scale stream reasoning*, Proceedings of the 4th international workshop on Web-scale knowledge representation retrieval and reasoning, CIKM, 2013.
[3] Banerjee, Snehasis, et al. *Semantic exploration of sensor data*, Proceedings of the 5th International Workshop on Web-scale Knowledge Representation Retrieval and Reasoning, CIKM, 2014.
[4] T. Chattopadhyay et.al. *A Smart Framework for IoT Analytic Workflow Development*, In Proc. of demo of Sensys, South Korea, Nov 2015
[5] Wang et. Al. *A Feature Subset Selection Algorithm Automatic Recommendation Method*, Journal of Artificial Intelligence Research 47 (2013) 1-34, May, 2013.
[6] Hong Lu et.al. *SoundSense: Scalable sound sensing for people-centric applications on mobile phones.* In Proceedings of MobiSys'09.

[7] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D. Lane, Tanzeem Choudhury, and Andrew T. Campbell. *The jigsaw continuous sensing engine for mobile phone applications.* In Proceedings of SenSys'10.

[8] Emiliano Miluzzo et.al. *Sensing meets mobile social networks: The design, implementation and evaluation of the CenceMe application.* In Proceedings of SenSys'08.

[9] Nicholas D. Lane et.al. *BeWell: A smartphone application to monitor, model and promote wellbeing.* In Proceedings of Pervasive Health'11.

[10] Andreas Krause et.al. *Context-awaremobile computing: Learning context-dependent personal preferences from a wearable sensor array.* IEEE Transactions on Mobile Computing 5, 2, 113-127.

[11] Roger G. Eston et.al. *Validity of heart rate, pedometry, and accelerometry for predicting the energy cost of children's activities.* Journal of Applied Physiology 84, 1, 362-371.

[12] Mashfiqui Rabbi et.al.*Passive and in-situ assessment of mental and physical well-being using mobile sensors.* In Proceedings of UbiComp'11.

[13] Saeed Abdullah et.al. *Towards population scale activity recognition: A scalable framework for handling data diversity.* In Proceedings of AAAI'12.

[14] Nadav Aharony et.al. *The social fMRI: Measuring, understanding, and designing social mechanisms in the real world.* In Proceedings of UbiComp'11.

[15] Celine Coutrix and Nadine Mandran et.al.*Identifying emotions expressed by mobile users through 2D surface and 3D motion gestures.* In Proceedings of UbiComp'12.

[16] Shaojiang Dong et.al. *Bearing Degradation Process Prediction Based on the Support Vector Machine and Markov Model*, Hindawi Publishing Corporation, Shock and Vibration, Volume 2014, Article ID 717465, 15 pages

[17] Hai Qiu et.al. *Robust Performance Degradation Assessment Methods for enhanced rolling element bearings prognostics*

[18] Anand Rajaraman, Jeffrey David Ullman, *Mining of Massive Datasets*, Cambridge University Press New York, 2010.

[19] Pradipta Maji et. al *Fuzzy-Rough MRMS Method for Relevant and Significant Attribute Selection*, 2012, Advances on Computational Intelligence: 14th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems.

[20] Pradipta Maji and Sushmita Paul *Dependence Issues in Knowledge-Based Systems Rough set based maximum relevance-maximum significance criterion and Gene selection from microarray data*, 2011, International Journal of Approximate Reasoning.

[21] Peng, Hanchuan, Fuhui Long, and Chris Ding. *Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy.* Pattern Analysis and Machine Intelligence, IEEE Transactions on27.8 (2005): 1226-1238.

[22] Voyant, Cyril, et al. *Meteorological time series forecasting based on MLP modelling using heterogeneous transfer functions.* Journal of Physics: Conference Series. Vol. 574. No. 1. IOP Publishing, 2015.

[23] W. K. Ngui et al., *Wavelet Analysis: Mother Wavelet Selection Methods*, Applied Mechanics and Materials, Vol. 393, pp. 953-958, 2013.

[24] T. Batabyal, T. Chattopadhyay, D. P. Mukherjee, *Action Recogntion using joint co-ordinates of 3-D Skeleton Data,* ICIP 2015, Quebec, Canadada, Sept, 2015.

[25] Banerjee, Snehasis, et al. *What affects me?: a smart public alert system based on stream reasoning,* Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication. ACM, 2013.

[26] Chattopadhyay, T. et al. *Automatic selection of binarization method for robust ocr,* ICDAR, pp. 1170-1174, IEEE, 2013.

[27] Mukherjee, Debnath, et al. "Ad-hoc ride sharing application using continuous sparql queries." Proceedings of the 21st International Conference on World Wide Web. ACM, 2012.

[28] Chattopadhyay, T., et al. *Recognition of channel logos from streamed videos for value added services in connected TV,* ICCE, IEEE, 2011.

[29] Mukherjee, Debnath and Misra, Prateep and Banerjee, Snehasis. *System And a Method For Reasoning And Running Continuous Queries Over Data Streams*, US Patent 20,150,039,648. 2015.

[30] S. Dong et. al., *Bearing degradation process prediction based on the support vector machine and Markov model,* Journal of Shock and Vibration, Hindawi Publishing Corporation, 2014.

[31] Banerjee, Snehasis, et al. *Towards Wide Learning: Experiments in Healthcare.* Machine Learning for Health Workshop, NIPS 2016. arXiv preprint arXiv:1612.05730 (2016).

[32] Banerjee, Snehasis, et al. *Towards a universal notification system,* In Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2013 IEEE/WIC/ACM International Joint Conferences on (Vol. 3, pp. 286-287). IEEE, 2013.

[33] Mukherjee, Debnath, et al. *Towards efficient stream reasoning,* ODBASE, On the Move to Meaningful Internet Systems. Springer Berlin Heidelberg, 2013.

[34] Mukherjee, Debnath, et al. *System and a method for reasoning and running continuous queries over data streams,* U.S. Patent Application No. 14/387,540.