Research Paper - Presentation

G Vojeswitha - AI20BTECH11024

July 4, 2021

Title

OBS network blocking probability prediction using ensemble technique

Authors

1. Srija Chakraborty, Dept. of C.S.E.,National Institute of Technology, Rourkela, India, 518CS6016@nitrkl.ac.in

2. Ashok Kumar Turuk, Dept. of C.S.E.,National Institute of Technology, Rourkela, India, akturuk@nitrkl.ac.in

3. Bibhudatta Sahoo, Dept. of C.S.E., National Institute of Technology, Rourkela, India, bdsahu@nitrkl.ac.in

OBS Networking and Blocking Probability

1. **Optical Burst Switching Networks(OBS NETWORK)**:
   It is a optical networking technique that allows sub-wavelength switching of data. Among other existing switching techniques, OBS performs better for network with high traffic.

2. **Data Burst Packet** - The packet containing the data to be transmitted.

3. **Burst Control Packet(BCP)** - A control signal containing all the routing instructions of data burst. BCP is transmitted prior to data burst.

4. Burst contention occurs whenever at the time of transmission burst gets blocked. If the blocking probability can be predicted for a network, then the performance of the network can be improved

5. The **Blocking probability** of any network can be determined by calculating the proportion of the count of the bursts (lost) divided by the aggregate count of burst (sent).

Abstract

1. . In the proposed technique, different predictor models and feature reduction algorithms are used to predict the blocking probability of upcoming traffic in a OBS network which will help us improve the performance of the network drastically

2. The proposed model is able to predict the blocking probability with higher accuracy, along with lowering the number of burst loss by wavelength conversion strategy. Thus, it will help future network designers to have an advanced idea about the performance of the network under specific framework.

3. Any Eventdriven simulator can be used to calculate the blocking probability simulation. Erlang fixed point approximation(EFPA) is used for analytical approach. However, these traditional approaches are much more time-consuming and tedious. Therefore, different AI techniques are used for prediction.

Proposed architecture of this approach
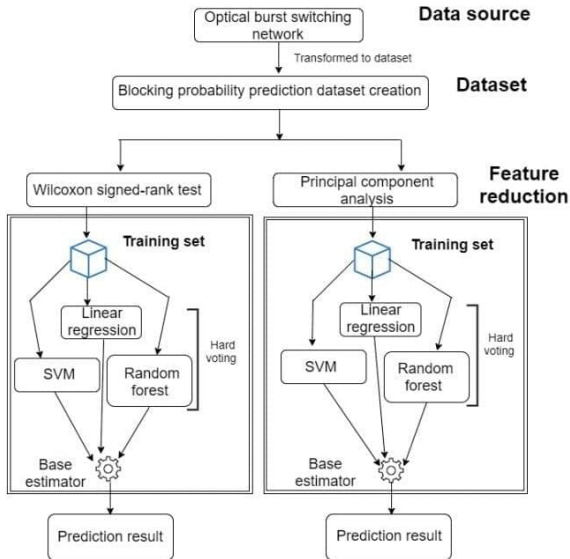
Dataset creation

Seven independent input parameters are taken as input. With help of these input parameters, proposed ensemble model is trained to study the mapping from the condition of the network and predict the blocking probability.

E	Mean traffic load in S-D pair
D	Diff b/w max, min traffic load
C	Number of channels per wavelength
W	Number of wavelengths
T_{off}	Offset time
APL	Average path length
CR	Concentration of rout

Table

Feature reduction techniques

1. For faster prediction, dimensionality of large data sets are reduced, by transforming a large set of variables into a smaller set that still contains most of the information in the large set.

2. Reducing the number of variables of a data set naturally comes at the expense of accuracy, but smaller data sets are easier to explore and visualize and make analyzing data much easier and faster for machine learning algorithms.

3. In this proposed model, Wilcoxon signed-rank test and Principal component analysis (PCB) are the feature reduction techniques used to remove least important features.

Wilcoxon Signed Rank Test

1. The null hypothesis of this test is 'probability distribution of set of pair wise differences is centered at 0'.For this procedure, differences comes from a continuous distribution

2. Test statistics for this procedure are calculated by either adding the ranks assigned to the positive differences T_+ or by adding the ranks assigned to the negative differences T_-.

$$T_+ + T_- = \frac{n(n+1)}{2} \tag{1}$$

where n = no.of samples of the data.

Wilcoxon Signed Rank Test (contd)

1. However for large samples, Z is used as standard normal distribution to test the hypothesis.

$$Z^+ = \frac{\frac{T^+ - n(n+1)}{4}}{\left\{ \frac{n(n+1)(2n+1)}{24} \right\}^{\frac{1}{2}}} \tag{2}$$

2. Null hypothesis is analysed by forming a rejection region for test statistic T_+ or T_-.
   A two-tailed rejection region for the null hypothesis based on T_+ is used.

Principal Component Analysis

It is a technique used to convert a set of possibly co-related variables to that of linearly uncorrelated variables using orthogonal transformation. Using eigen values and eigen vectors, data is categorised based on their similarities and differences.

The following equation is used to achieve the decreased dimensional data

$$Y = Z^T X \tag{3}$$

where Y = Final Data Set, X = Original Data Set, Z is the orthogonal matrix containing eight eigen vectors.

Wilcoxon Signed Rank Test vs PCA

For a dataset of OBS network, PCA reduced more unwanted features than wilcoxon signed-rank test.

Predictor techniques

1. In this work, ensemble of predictors like Linear Support Vector Machine (SVM), Linear Regression (LR) and Random Forest (RF) are used to train and predict burst contention or burst blocking probability

2. To acomplish ensemble, we use bagging and as we predict the exact result instead of probable result, we use hard voting. In our model, hybridization is done by bagging.

3. In bagging, we use same predictor many times and the training instances are sampled.

4. After using each model, we will get accuracy for each of them. Using each of these results, we can make a better prediction by aggregating the predicted results and by voting for the final class, which has more number of votes

Linear Support Vector Machine

1. $w^T x$ is the hyperplane of linear SVMs which segregate data points of two classes optimally
   Here w is a hyperplane that learns from data using SGD (stochastic gradient descent).

2. In this algorithm, the objective function E(w) used in SGD as a feature vector,

$$E(w) = \frac{1}{p} \sum_{i=1}^{p} L(y_i, w^T x_i) + a||w||^2 \tag{4}$$

where $x_i \in$ X and respectively y_i lies between 0 to 1
a is regularisation constant
and L is the loss function defined as

$$L(t, y) = max(0, 1 - t(y)) \tag{5}$$

Linear regression

1. Equation of this model : $Y = aX + b$
   where X is the independent variable, a is the slope and b is the y-intercept. Here a and b can be calculated using equation 5 and 6.

$$a = \frac{(\sum y)(\sum x^2) - (\sum x)(\sum xy)}{n.(\sum x^2) - (\sum x)^2} \quad (6)$$

$$b = \frac{n(\sum xy) - (\sum x)(\sum y)}{n.(\sum x^2) - (\sum x)^2} \quad (7)$$

Random Forest

1. In this algorithm, best features are not searched while splitting a node rather than it is searched among a random subset of features.

2. The classification formula for this model is:

$$J(k, t_k) = \frac{m_{left}}{m} G_{left} + \frac{m_{right}}{m} G_{right} \tag{8}$$

Where G evaluates the impurity of the subsets left and right and m represents the total number of instances available in the subsets.

Training model

1. Random forest is prefered to use since it works efficiently on large dataset while SVM and linear regression are weak learners.

2. After training is performed utilizing the linear regression, next we train the model using a linear support vector machine.At the last stage of algorithm, random forest algorithm is used.

3. With linear regression, the samples can be predicted with higher or lower accuracy.

4. Linear SVMs work efficiently most of the times, many datasets are not even being linearly separable. For getting that situation to regularise the SVM, we have set C=1 and loss hyperparameter to "hinge"

Training model(contd)

1. The parameters which are used in random forest are the maximum depth of tree, random state and also the minimum number of samples needed for splitting an internal node and the minimum number of samples needed to be at a leaf node.

2. All the selected samples should be considered at the time of splitting the internal node. In a tree, if it leaves at least minimum training samples in each of the left and right branches, then it is considered as a split point at any depth.

Algorithmn: Blocking probability prediction model

Algorithm 1: Ensemble blocking probability prediction model for OBS network

---

Input: A data set X

Output: A predictor model which predicts blocking probability of OBS network.

1 Reduce unwanted features using wilcoxon signed-rank test X_W;

2 Reduce unimportant features using PCA X_P;

3 **Function** `Training`(X_W, X_P):

4   **for** *training instances in dataset X_W* **do**

5   Apply bagging on SVM, LR, RF one by one on dataset X_W;

6   Apply hard voting to predict final blocking probability W_B;

7         $W_B \leftarrow$ blocking probability decided using hard voting ;

8   **for** *training instances in dataset X_P* **do**

9   Apply bagging on SVM, LR and RF one by one on dataset X_P;

10   Apply hard voting to predict final blocking probability P_B;

11         $P_B \leftarrow$ blocking probability decided using hard voting ;

12   **return**;

13 **end function**
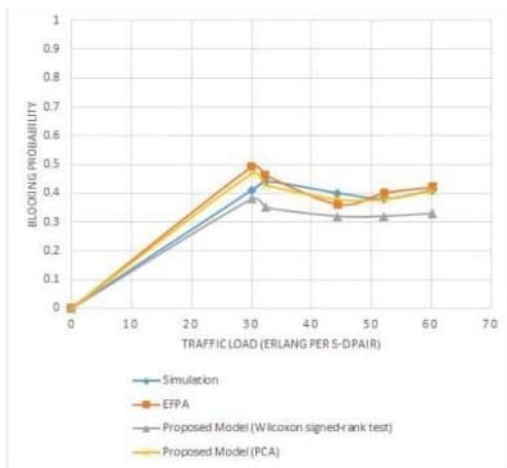

Experiment results

Figure: Blocking probability estimation by Simulation, EFPA, proposed ensemble model with wilcoxon signed-rank test and proposed ensemble model with PCA with setting D = 6, C= 40, W= 3, Toff = 0.0000086, APL= 3.75, CR= 3
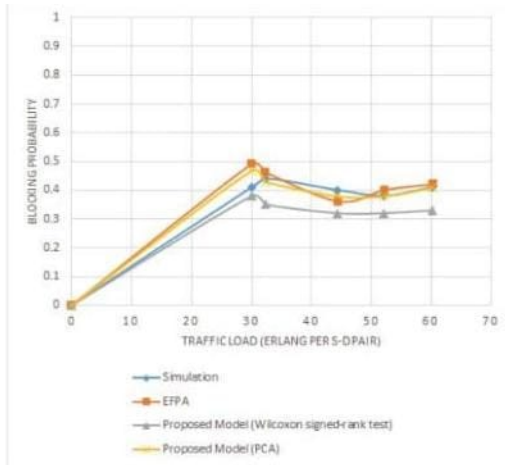
Experiment results

Figure: Blocking probability estimation by ensemble proposed model, ADMM-I-ELM, ADMM-Log-I-ELM, EFPA, Simulation with setting D=10, C=18, W=5, Toff = 0.000009, APL=2.628, CR=2.25

Experiment results

1. While applying ensemble with wilcoxon signed-rank test technique as feature reduction, the root-mean square error(RMSE) is calculated as 7.4×10^{-4} and with PCA technique as feature reduction, the root-meansquare error(RMSE) is calculated as 2.3×10^{-4}

2. The first feature reduction technique used is wilcoxon signed-rank test which does not perform well for this model and gives only 80 % accuracy

3. However, for the second technique which is principal component analysis, it performs much better and gives 97 % accuracy for the dataset obtained from OBS network