Forecasting of the rate of Industrial Accidents and Labor Disputes by Industry: Application of Ensemble Learner-Based on Fusion Data Model

Introduction

The evaluation of jobs can be done in several ways and for a number of different reasons. Job evaluation, as the term stated, has broadly referred to the organized procedure that industrial engineers and other analysts followed to establish the performance rate of jobs in organizations. The components of a compensation program that shows the emphasis an organization places on actualizing goals differ between organizations such as avoiding labor disputes and discrimination charges among others. Despite the fact that analysts and industrial engineers are followed the process of job evaluation to reduce grievances and labor disputes, there is still an increasing rate of labor disputes by industry. The use of provisional, poorly defined, short-period hire methods allows the dispute to happen and makes it tough to negotiate agreements with employees.

Ergonomic specialists are concerned primarily with mechanical injuries involved in equipment design. A precept that guides them is the designing of safety features that can easily and minimally interfere with the effective operation of the equipment. Another principle that guides them is to label the danger points with a relevant warning. Apart from the guards and safety devices, there is a need for supervision and thorough training of the operators and reminding them to take precautions against anything that can catch in the machine and cause accidents. However, despite the fact that all these efforts resulted in reducing industrial accidents with mechanical injuries, the rate of industrial accidents and the annual economic estimates costs of accidents is still unacceptably high.

The Objective of the Research

The accurate predictions of industrial accident rates and labor disputes rate are challenging tasks and it is an important issue to improve the performance of systems by improving human-machine interaction and labors welfares. In view of these circumstances, this study is intending of looking at the feasibility of forecasting the labor dispute rate and the industrial accident trend simultaneously with the aid of data processing and neural network algorithms by developing powerful ensemble models based on boosting algorithms. Therefore, the novel hybrid forecasting techniques based on the ensemble methods that combine a group of deep-learning models, each known as composite or base learner are developed for both the labor dispute rate and the industrial accident forecasting. The paper intends to reveal that the Ensemble Learner model on fused data has superior forecasting performance than an ensemble model trained on any of the individual data sources and to show that they are capable of forecasting future trends in such an unstable condition. The ultimate objective is to show the trend and the factors/items that contribute most to the increasing rate of accidents and labor dispute trends in other to curb it. The python programming language will be used.

Preliminary Literature Review on Method to be Used

Data fusion is the process of an appropriate combining of data or information obtained from multiple sources to produce complementary and useful information compared to that provided by individual data sources alone. The data obtained as result can improve the forecasting accuracy of the regression or classification. Applications of data fusion and ensemble techniques have been jointly used (Polikar, 2006). Parikh & Polikar (2007) demonstrate that Learn++ -based data fusion has superior performance accuracy than an ensemble classifier trained on any of the individual data sources. Parikh & Chen, (2008) made two observations about intrusion detection with the objective of minimizing the cost of errors and find that the fusion approach has better performance in minimizing the cost. . Fusion of data from various sources of data may produce consistent and accurate data for better understanding of the observed situation thereby decreasing the unreliability that associates with the individual sources (El Faouzi, Leung & Kurian, 2011). The various data fusion methods were drawn from areas such as artificial intelligence. It can also be applied to areas such as bioinformatics, wireless sensor networks, soil mapping, business intelligence, and many other areas including industrial and systems engineering. A data fusion technique integrates the benefits of the individual techniques of the roadside collected acoustic, image, and sensor data (Joshi, 2013). In this proposed research, the different sets of data collected will be analyzed individually using different proposed ensemble algorithms then the best among the proposed models will be used on the fused data. The performance of the approach will check on both arithmetic means and median. Then the percentage errors will be calculated to determine whether the ensemble-based data fusion model outperforms the best ensemble technique trained on the individual data sources

Methodology

The ensemble learner models will be proposed to apply to individual time series data and the best among the proposed models will then be used with fused data. The grand proposed method is the "Fusion Data-Based Ensemble Learner Model".

Data Fusion Processing [summary]

Assuming we have a "q" number of the different datasets, that is
$\{(\boldsymbol{x_t})^{\boldsymbol{q}}, (\boldsymbol{y_t})^{\boldsymbol{q}}\}_{t=1}^{l}$
Where
$\{(\boldsymbol{x_t})^{\boldsymbol{q}}\}_{t=1}^{l}$ are the features, and
$\{(\boldsymbol{y_t})^{\boldsymbol{q}}\}_{t=1}^{l}$ are the labels or responses.
Considering simple linear regression written as follow:
$(y_t)^l = \beta_0 + \beta_t x_t + \varepsilon_t, \ \ \varepsilon_t \sim N(0, \sigma_\varepsilon^2).$

$$Y^1 = \begin{bmatrix} (y_1)^1 \\ (y_2)^1 \\ \cdot \\ \cdot \\ \cdot \\ (y_t)^1 \end{bmatrix} : Y^1 = H_1\beta + \varepsilon, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}. \varepsilon \sim N\big(0, R_\varepsilon = cov(\varepsilon)\big) \qquad , R_\varepsilon = \begin{bmatrix} \sigma_\varepsilon^2 & \cdots & \theta \\ \vdots & \ddots & \vdots \\ \theta & \cdots & \sigma_\varepsilon^2 \end{bmatrix}.$$

$$Y^2 = \begin{bmatrix} (y_1)^2 \\ (y_2)^2 \\ \cdot \\ \cdot \\ \cdot \\ (y_t)^2 \end{bmatrix} : Y^2 = H_2\beta + \eta, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}. \eta \sim N(0, R_\eta = cov(\eta)) \qquad , R_\eta = \begin{bmatrix} \sigma_\eta^2 & \cdots & \theta \\ \vdots & \ddots & \vdots \\ \theta & \cdots & \sigma_\eta^2 \end{bmatrix}$$

$\left(\tilde{\beta}\right)^1 = (H_1^T R_\varepsilon^{-1} H_1)^{-1} H_1^T R_\varepsilon^{-1} Y^1).$ $\hat{w}_1 = cov\left(\tilde{\beta}\right)^1 = (H_1^T R_\varepsilon^{-1} H_1)^{-1}$. Therefore, $\left(\tilde{\beta}\right)^1 = \hat{w}_1 H_1^T R_\varepsilon^{-1} Y^1).$
Also, $\left(\tilde{\beta}\right)^2 = \left(H_2^T R_\eta^{-1} H_2\right)^{-1} H_2^T R_\eta^{-1} Y^2).$ $\hat{w}_2 = cov\left(\tilde{\beta}\right)^2 = \left(H_2^T R_\eta^{-1} H_2\right)^{-1}$. Therefore, $\left(\tilde{\beta}\right)^2 =$ $\hat{w}_2 H_2^T R_\eta^{-1} Y^2)$
Furthermore,

$$\left(\tilde{\beta}\right)^1 = \begin{bmatrix} \left(\tilde{\beta}_0\right)^1 \\ \left(\tilde{\beta}_1\right)^1 \end{bmatrix} \Rightarrow (\hat{y}_t)^1 = \left(\tilde{\beta}_0\right)^1 + \left(\tilde{\beta}_1\right)^1 x_t \ . \ \left(\tilde{\beta}\right)^2 = \begin{bmatrix} \left(\tilde{\beta}_0\right)^2 \\ \left(\tilde{\beta}_1\right)^2 \end{bmatrix} \Rightarrow (\hat{y}_t)^2 = \left(\tilde{\beta}_0\right)^2 + \left(\tilde{\beta}_1\right)^2 x_t$$

$\tilde{\beta}_0$ *is an intecept while* $\tilde{\beta}_0$ *is a slope*
Fusion model

$\left(\widetilde{\beta}\right)^{fus} = W_1\left(\widetilde{\beta}\right)^{1} + W_2\left(\widetilde{\beta}\right)^{2}$, where, $W_1 = \hat{w}_1(\hat{w}_1 + \hat{w}_2)^{-1}$, and $W_1 = \hat{w}_2(\hat{w}_1 + \hat{w}_2)^{-1}$

$(\hat{y}_t)^{fus} = \left(\widetilde{\beta}_0\right)^{fus} + \left(\widetilde{\beta}_1\right)^{fus} x_t$

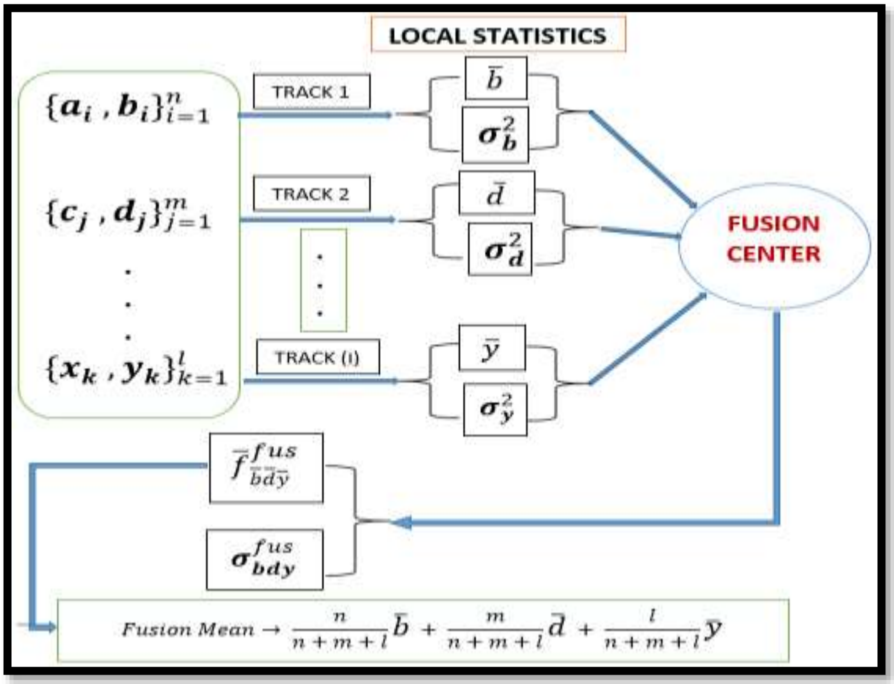

Fig. 1. Graphic analysis of fusion approach to a simple model. The statistical computation will be done on the dataset from various sources to obtain the mean and variances of each dataset, this is termed local statistics. The data will then be fused together and the fusion statistics will be computed on the fused data. **(Source: Original-Author of Proposal)**

The process of fusion to arrive at the fusion output is graphically shown in Fig. 2 below. Using the best among the proposed models with the fused data is what I termed the "Ensemble Learner-Based Fusion Data Model". The performance of the approach will be checked on both mean square error (MSE), Scatter index (SI) which is a normalized RMSE computed by dividing the RMSE by the data mean and accepting the estimations if it is less than one. The other evaluation metric will be both mean arctangent percentage error (MAAPE) and the weighted mean absolute percentage error (WMAPE). These are scale-independent metrics that can be relied on for better performance comparison especially when the scales in the data are different. I will check the output against the result obtained from the individual output of the same model to determine whether the ensemble learner-based fusion data model outperforms the best ensemble technique trained on the individual data sources.

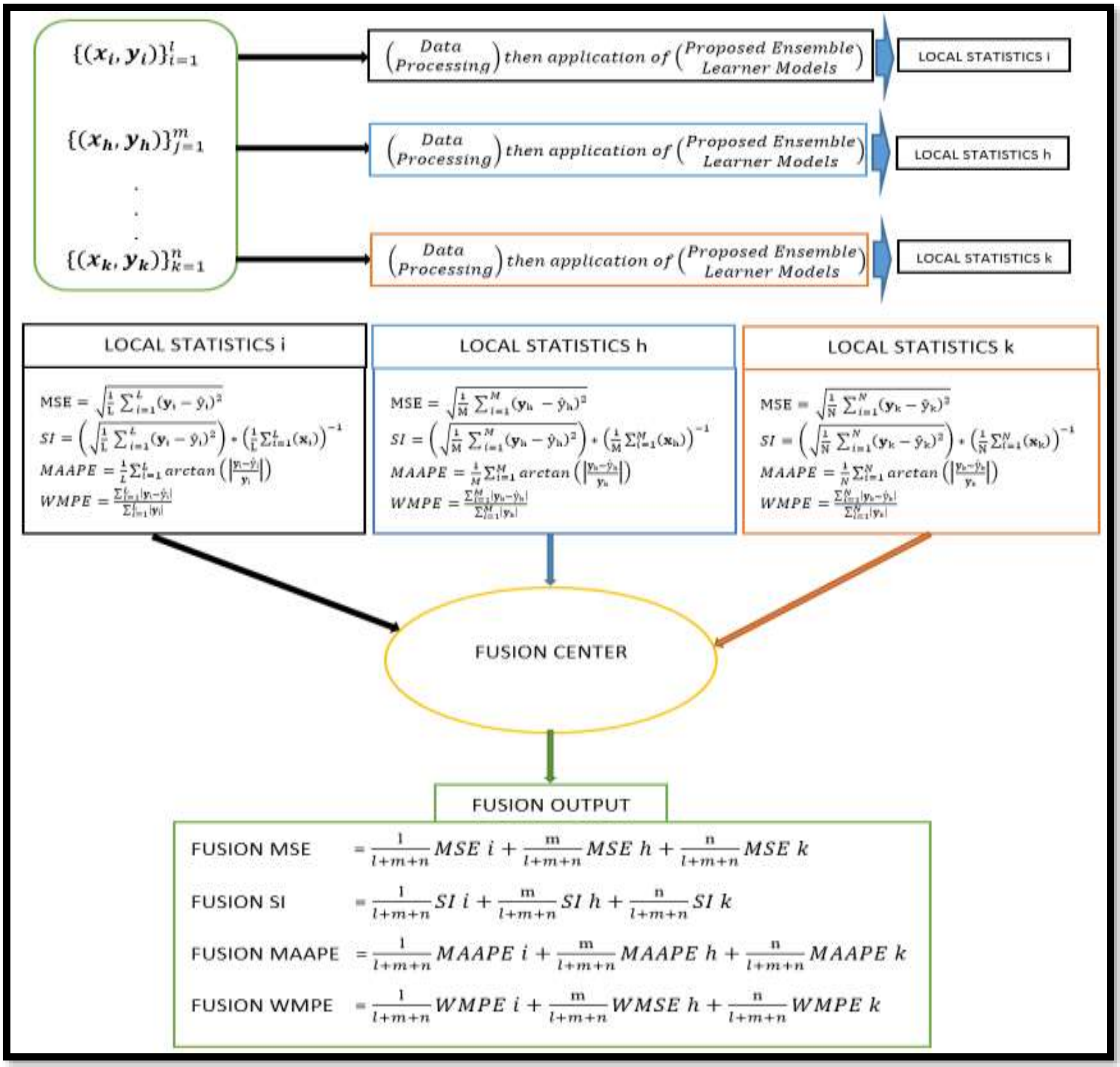

Fig. 2. Graph shows the stages involve in determining the fusion output. The stages include the collection of datasets from various sources. The processing of the data followed, including data normalization, fixing of window size and split of data into a training set and a test set, and concatenation. The best among the proposed ensemble-learner model is used to train and predict the fused data in other to obtain the fusion output **(Source: Original-Author of Proposal)**