Innovative Solutions for Design and Fabrication of Deep Learning Based Soft Sensor

Radhia Khdhir,
rkboujelben@ju.edu.sa
Department of Computer Science
College of Science and Arts in Qurayyat,
Jouf University, Saudi Arabia

Aymen Belghith
a.belghith@seu.edu.sa, aymen.belghith@gmail.com
Computer Science Department
College of Informatics and Computing
Saudi Electronic University, Saudi Arabia

Summary
Soft sensors are used to anticipate complicated model parameters using data from classifiers that are comparatively easy to gather. The goal of this study is to use artificial intelligence techniques to design and build soft sensors. The combination of a Long Short-Term Memory (LSTM) network and Grey Wolf Optimization (GWO) is used to create a unique soft sensor. LSTM is developed to tackle linear model with strong nonlinearity and unpredictability of manufacturing applications in the learning approach. GWO is used to accomplish input optimization technique for LSTM in order to reduce the model's inappropriate complication. The newly designed soft sensor originally brought LSTM's superior dynamic modeling with GWO's exact variable selection. The performance of our proposal is demonstrated using simulations on real-world datasets.

Keywords:

Soft sensor, Long Short-Term Memory (LSTM), Grey Wolf Optimization.

1. Introduction

Soft sensor is a synthetic deductive reasoning projection methodology that provides accurately manageable characteristics to forecast system parameters that are complicated owing to technical and administrative constraints, as well as a context in which it occurs. The soft sensor aims to create a predictive statistical method among accurately manageable quantities and major difficulties set of independent, which is used to address the issue that precludes readings from being used as product testing signal generated. Soft sensor technologies have also been a key technological evolution including both academic and commercial [1], with soft sensor approaches being used more especially in various processes.

Most parameters in genuine manufacturing applications can indeed be evaluated. The assessment regularity is significantly small due to technological limits, sensory specifications, ambient considerations, etc. Soft quantification is an efficient way to objectively convert observable characteristics to tougher ones in statistical models [2–4]. Neural Networks (NNs) are sophisticated approaches for modeling complicated and chaotic systems that have already been widely used during soft sensors [5–7]. In [8], Heidari et al. initiated a separate inter feed-forward neural network which is more appropriate, in existing NN topologies, to estimate boundary layer flow roughly comparable viscosity. In [9], Sheela and Deepa combined personality maps with MLP to create a synthesized approach, which are used to anticipate the wind speed of a green power process. Authors in [10] created a dynamically recursive network for a soft sensor of biological processes. Their implementation to a purification acetaldehyde acid extraction technique showed excellent efficacy. In [11], throughout the case study, soft sensors were created for performance estimation and a different style of automated vector autoregressive analyzation has been developed based on the network info logistic forecasting, that could accurately forecast multimodal aberrations.

Many factors are tracked by online sensors while dealing with industrial processes. Some of these variables, however, are difficult to monitor. Because of slow hardware sensors or laboratory analysis, variables are measured with long delays in some circumstances, making real-time monitoring of the process impossible. Based on online measured variables, inferential models can be developed to estimate these difficult-to-measure parameters including the use of soft sensors. Figure 1 summarizes their functioning principle.

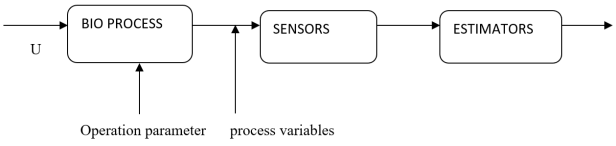

Figure 1: Soft Sensors Functioning Principle

There is now an expanding theme of using statistics machine learning techniques to increase technologies, operations, or commodities throughout several manufacturing industries for at least ten years [12]. Such enhancement potentially manifest in many ways, ranging from lower density or intermediate goods usage [13] to

Manuscript received February 5, 2022
Manuscript revised February 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.2.17

superior equipment utilization [14, 16] or increasing concentrations of automated processes [17] to greater energy reliability. Recently, fighting climate change because of stronger burdensome regulations seems to be a big factor [18, 19].

Furthermore, information gathering required for this kind of strategies is fraught with difficulties, including the long-term viability of construction equipment: Authorized deterioration figure comes from (rarely) 6 to more than 30 years, based on the location, item of equipment, and automotive sector [20]. Additionally, anecdotal evidence suggests that, particularly in businesses of all sizes, durable systems can last substantially extended in regular usage. As a result, much of today's modern gear was built but that was before the data-hunger arose. As a result, it frequently lacked at substantially so many of the sensors required to gather it.

Products, on the one hand, are currently inherently complex, with interpreted relationships amongst datum entries. Datasets, on the other hand, are quantitative approach with substantial intermediation distortion and unpredictability, making forecasting with traditional NNs more difficult. Long Short-Term Memory (LSTM), a sophisticated type of NN, was subsequently developed to manage sequences dependence [21–23]. Because its memory cells can keep their state throughout time and standardize the evidence coming in or out of the cell, an LSTM network is so much greater for developing long-term long-range dependencies. As a result, LSTM networks have found success in a variety of domains, including meteorological developments [24], traffic forecasting [25], human activity identification [26], etc. Because of its strengths, LSTM is being used in the construction of soft sensors in manufacturing applications. Yuan et al. designed a directed recurrent neural network for a soft sensor and used two practical industrial scenarios to illustrate the effectiveness of the suggested soft sensor [27]. Sun suggested a new LSTM network for a soft sensor by integrating feature extraction choosing and supervised dynamic methodological approaches, and the network was demonstrated using a real-world Equilibrium adsorption column [28].

This work proposes a new soft sensor technique that combines LSTM and GWO, with the GWO compressing LSTM system parameters. The following is a list of the paper's major aspects:

(1) For LSTM with GWO, a novel feature selection strategy is designed. The proposed technique effectively eliminates the additional functionality induced by duplicated parameter combinations, therefore increasing the modeling efficiency of the LSTM.

(2) The designed soft sensor technique is designed in two important manufacturing applications.

(3) Simulation findings show that the new soft sensor model performs better and is more flexible while performing feedback control.

2. Related works

Various deep neural network techniques using the angular velocity signal to calculate ignition characteristics are presented in this related work. Several papers (e.g., [27–29]) predict the flow rate curve first, and then calculate the ignition parameters (typically p_{Max} and its placement in the rotating angle range) for elements to determine. Other papers, such as [17], proposed to calculate and assess ignition parameters directly using a neural network. Note that we opt the latter strategy. The choice differs greatly in terms of which criteria are deemed especially important for the effectiveness of the techniques.

To calculate cylinder pressure, Bennett et al. utilized a computational model with a semi integrated moving average Arima with level higher (NARX) design [27]. Among 1.7° and 4.3°, the standard deviation varied between 5.3% and 33.6%, due to the operating environment. The same scientists improved this result significantly by utilizing a period computational model in [14]. The mean error fell to just 1.141 percent to 1.323 percent, respectively for 1.651° to 3.082°, under the same conditions utilized in [27]. As a result, the type of network used in these two papers had a significant impact on the techniques' effectiveness. In [29], a multilayer perception is presented that includes data associated with the air ratio, combustion angle, and compressor boost pressure in addition to the rotor position signal. The emphasis was on the artificial network's structure that had a significant impact on the computation effectiveness.

To calculate injection pressure curves, the authors of [28] used a computational model with multilayer perception. They did not use the basic rotor position signal, but instead translated it into the Fourier transform and evaluated just the first 20 overtones, in contrast to earlier research that have been using a Radial Basic Function (RBF) network. They also employed the structure-borne acoustic signal's 21st-50th vibrations. As a result, the preparation of the given data is the most important aspect of this project. The typical errors for p_{Max} and its value in the cranks angle range are 3.4% and 1.5 degrees, respectively.

3. METHODOLOGY

3.1 Design

Enterprises hold information in previous records, which is traditionally performed by an oversight command and information management or Distributed

Control System (DCS). Because the modelling can get some more insight than what is contained in the data, the recorded information should be sufficient to capture the entire dynamic behaviour of the system. The initial phase of the design involves filtering and preparing the data when it has been acquired. This is because unfiltered data from databases has well-known issues including oversampling, outliers, and missing data, as well as consistency issues such discrepancies, weather conditions, and excessive noise. As a result, the designer must treat them with care and prepare data so that it may be used in subsequent design processes. Clustering method, contrast adjustment, feature extraction, and reduction are all common pre-processing procedures.

The sample rates of data obtained in vegetation databases are frequently diverse. Convenient characteristics are performed automatically utilizing various online sources sensors, whereas difficult-to-measure parameters are tested infrequently, at a significant expense, and with long delays, as in the case of laboratory analysis. As a result, the previous frequently have sufficiently large sample rates, even greater than the sampling theorem requires, but the later often to be down sampled. High sampling rates might result in large datasets with data co integration. As a result, re-sampling is required to re-synchronize the variables and avoid dealing with large datasets.

In datasets gathered from different businesses, incomplete values and outliers are prevalent issues. Peak, fluid saturation, flattening trends, and interruptions are examples of the earlier, which arise whenever values are absent in a variable's observation; the latter are genuinely incongruous data, with the bulk of the documented ones deviating substantially from the average range of values. Sensor or system failures, as well as parameter uncertainties, can both cause them. They are normally dealt with either eliminating the samples that contain them or by using some sort of imputing approach to fill in the missing observations. Outlier detection, on the other hand, is a difficult operation that can be accomplished using statistical techniques like the 3-rule, or the ultimate confirmation must be carried out properly by a horticultural specialist to minimize anomaly concealing (aberration overloading).

3.2 Fabrication

A. Carbon black composite preparation

For a multitude of conditions, researchers are combining the flexibility of silicone polymers with the electrical capabilities of conductive filler materials [8]. Plasticizers get a high elongation and elastic modulus, as well as good moisture levels resistance. Furthermore, owing to the critical shortage of electrostatic forces, ketones, like other macromolecules, are practically conductive at low voltage. The proportion of the additional composite electrode is among the parameters that defines a composite's conductivity and percolation threshold [9]. Concentrations that surpass the decomposition temperature are premium for ridiculously priced fillers like silver nanoparticles and nanomaterials. Furthermore, producing a homogenous and durable filler distribution is critical for creating the infrastructure for electromagnetic transfer, and this usually necessitates the use of a polar organic solvent to aid dispersion [14]. To optimise this process, several dispersion methods [9] and disseminating solvents [15] have since been devised.

Furthermore, it is required to omit the use of the disseminating fluid in favour of a relatively easier production approach and balanced by raising the filler content. The carbon black to Ecoflex ratio was raised to 1:6 and blended by hand for a few moments. Manufacturing a percentage point superconducting laminate with this methodology costs only 0.12 USD, saving you an additional 18% over the direction of development. This improved conductivity composite production system involves no specialised equipment, involves only a few moments, and is way more efficient than existing methods.

B. Sensor Fabrication

Initially, drop a 1 mm thick couple of layers of sustainability has been defined material into a mould, which is blended 1A:1B by weight. After pouring, pressure services required the mixture to remove any trapped air. The silicone is cured for around 2 hours at 70° C. Just after silicone has hardened, it is needed to overlay it with a 0.8 mm thick laser-cut pattern. The stencil is 2 mm broad, which matches to the sensor's width. The laser cutter was employed for convenience, but considering the stencil's measurements, it could easily have already been made by manually.

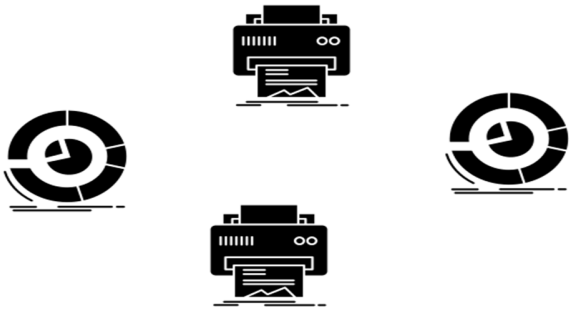

Figure 2: Fabrication of soft sensor

Figure 2 depicts the procedure for fabricating single sensors by embedding the carbon composite V in Ecoflex gel. We start by pouring a 1 mmv thick base layer of Ecoflex rubber into a mould, which is then blended 1A:1B by weight. We vacuum degas the mixture before to dumping to remove any trapped air. We cure the silicone at ambient temperature for about 2 hours or use a 70° C oven to speed up the process. We use a sensor stencil with a diameter of 0.8 mm to cover the rubber after it has hardened. The sensor's width is 2 mm, hence the stencil is 2 mm broad. The laser cutter was employed for convenience, but considering the stencil's proportions, it might have easily been cut via manually.

3.3 Grey Wolf Optimization

Grey wolf optimization simulates the intelligent behavior of grey wolves, including leadership and hunting behaviors. The apex predator is the Grey Wolf (GW), which implies that it is at the top of the biological food chain. GWs frequently opt to live. The grey wolf lives in a pack of 5 to 12 members with a rigid social structure. The hierarchy of a grey wolf pack is divided into four levels: alpha, beta, delta, and omega. Alpha (α) is the first rank in the hierarchy and is regarded as the pack's first leader. It oversees all decision-making processes, such as pursuing prey, approaching prey, and training the entire pack of wolves. The second level of the hierarchy is beta (ß), which advises the alpha when it passes away. The wolf can be female or male, and if the other wolves become old or die, he or she will most likely be the better one.

Omega (ω) is the lowest rank of the hierarchy, obeying the decisions of the three higher-ranking leaders while also ensuring the wolf pack's safety and integrity. In level delta (δ), the wolves are supposed to subject, and this level dominates. This group includes scouts, elders, and sentinels. Scouts keep an eye on the territory lines and warn the pack if they are in danger. The following is a mathematical representation of the GWO working process:

Retrieval in the GWO methodology is based on the wolf's relocation. Such three wolves are being pursued by the x wolves. During the chase, the GW circles the quarry.

3.3.1 Encircling prey

The encircling behavior of GWO is mentioned in the following equation:

$$E = |FX_P(i) - X(i)| \quad (1)$$

$$X(i+1) = X_P - BF \quad (2)$$

Where:

- i represents the present repetition.
- B and F are coefficient vectors.
- Xp represents the Prey's position vector.
- X represents the GW's position vector.

Vectors B and F are computed as follows:

$$B = 2a * rand(1-a) \quad (3)$$

$$F = 2 * (rand\ 2) \quad (4)$$

Even during searched repetitions, parameter a is progressively lowered between 2 to 0; rand1 and rand2 are procedurally chosen variables in the range of [0, 1].

3.3.2 Hunting

The capacity to recognize prey position aids the GW in encircling the prey. The effective quest is led through a regular basis. Furthermore, it may play a role in hunting behaviors and should be conscious of the whereabouts of prey to investigate GW's hunting behavior. As a result, the very first top three results calculated are saved, and the remaining of the evolutionary algorithms are exposed to updates on their own positions in relation to the number one search agents' whereabouts. We are listing some examples:

$$E_\alpha = |(F_1 X_\alpha) - X| \quad (5)$$

$$E_\beta = |(F_1 X_\beta) - X| \quad (6)$$

$$E_\delta = |(F_1 X_\delta) - X| \quad (7)$$

$$X_1 - X_\alpha(B_2\ E_\alpha) \quad (8)$$

$$X_2 - X_\beta(B_2\ E_\beta) \quad (9)$$

$$X_3 - X_\delta(B_2\ E_\delta) \quad (10)$$

$$X(t+1) = \frac{X1 + X2 + X3}{3} \quad (11)$$

Note that β and δ analyze the direction of prey, while the other wolves adjust their positions around the predators at randomness. The location of prey is then

evaluated, and the rest of the wolves randomly update their locations around the prey.

3.3.3 Attacking Prey

The GW starts the predator hunting procedure when the prey stops moving. The range is reduced to tapering B from the greatest value to quantitatively depict the process of the impending invasion of the prey in the direction of the prey. This improves the GWO's efficiency per $|B| < 1$ by pushing the GW to assault (exploitation) in the direction of constructing a greedy algorithm.

3.3.4 Search for prey

GWs usually look for the best candidate by looking at the locations of α, β, δ and wolves. They disperse (adventure) from one another to look for prey and then reunite (utilization) to attack the prey. To drive the search agent to diverge from prey, the random coefficient vector (B) must be above 1 or below –1. Furthermore, the random coefficient vector F aids GWO in achieving better random behavior during optimization and avoiding local optima. In contrast to parameter B, parameter F is lowered in a non-linear manner. Option F is required to procedurally provide chosen values each time in order to emphasize the exploring phase.

3.4 LSTM model

The internal hidden state model is maintained by the Recurrent Neural Network (RNN) which is a class of neural network. The cyclic connection between its units is directed by the temporal behavior of the arrangement with random length. The long short-term memory is also known as the hidden Markov model extension. The modeling long-term temporal dependencies are achieved by using a non-linear transition function. By adding three gates in the RNN neuron, the LSTM is extended to the three layers: the forget gate neuron, input gate neuron, and output gate neuron. The forget gate neuron controls the current state whether to forget or not. The output gate neuron states the output while the input gate neuron states whether the input should be read or not. The sequence in the long-term dependencies is learned when these gates are enabled. The recurrent hidden layer is effectively proliferated with the help of these three gates. This does not affect the output. The drawbacks in the RNN have been overcome by the LSTM, which is an effective method when compared to RNN.

$$y_t = \gamma\left(A_s.s(t-1) + A_r R_g(t)\right) \quad (12)$$

$$ip_t = \gamma\left(A_{ips}\, s(t-1) + A_{ipr} R_g(t)\right) \quad (13)$$

$$fg_t = \gamma\left(A_{fgs} s(t-1) + A_{fgr} R_g(t)\right) \quad (14)$$

$$op_t = \gamma\left(A_{ops} s(t-1) + A_{opr} R_g(t)\right) \quad (15)$$

$$s(t) = fg_t \odot r(t-1) + ip_t \odot y_t \quad (16)$$

$$op(t) = s(t) \odot op(t) \quad (17)$$

In the above equation, “fg”, “ip”, and “op” denote the forget gate, input gate, and output gate, respectively. Parameter γ denotes the activation function, symbol $\odot$ denotes the product of the gate value, and R denotes the parameter of the matrices.

3.5 Proposed LSTM-GWO

The assessment system is crucial in the identification of MI characteristics, which has a serious influence on the computation final performance. A better direct way is to determine the constant with the highest MI of result variable Y and source variable using the following equation:

$$R = I(X_i; Y). \quad (18)$$

The penalized phrases in the MIFS [30] technique is predicated on a measure of significance that takes into account correlations and duplication between variables. The efficient algorithm is described below:

$$R = I(X_i; Y) - \beta \sum_{x_{ss}} I(X_i; X_s) \quad (19)$$

Where S is the extracted features subset, Xs denotes the selected feature, and variable Y denotes the compensation for duplicated entries.

[30] presented the MIFS-U approach to lessen the reliance on parameters. The important components are expressed as follows:

$$R = I(X_i; Y) - \beta \sum_{x_{ns}} \frac{T(\lambda_s, Y)}{H(X_s)} I(X_i; X_s) \quad (20)$$

$$R = I(X_i; Y) - \frac{1}{|S|}\beta \sum_{x_{ms}} I(X_i; X_s) \quad (21)$$

$$R = I(X_i; Y) - \frac{1}{|S|}\beta \sum_{x_{ms}} NI(X_i; X_s) \quad (22)$$

$$NI(X_i; X_s) = \frac{I(X_i; X_s)}{\min\{H(X_i), H(X_s)\}} \quad (23)$$

The proposed methodology seeks to minimize redundant variables and enhance model accuracy by combining GWO and LSTM. Then, the Root Mean Square Error (RMSE) of the LSTM network is utilized as the assessment standard. The GWO-LSTM algorithm's operational principle is broken into two sections. Moreover, a model for prediction was developed using LSTM with GWO for variable selection in the procedure, and the LSTM NN is trained to find network hyper parameters and structure. The starting set of *n* variables is set to *F*, and the empty set is assigned to *S*. The modeling approach is GWO with the LSTM of RMSE. In this approach, the first variable is picked. Then, in each phase, the model parameters are identified until the simulation deteriorates or fulfils the stop requirement. Finally, the chosen subset is modeled and a value prediction is created.

The pseudo code is presented as follows:

Step 1 (Initialization): Using the training dataset, create a new upgraded Elman NN and determine the network's starting weight values.

Step 2: Choose a dataset of input variables.

Step 2.1: Combine ISOMAP and LSTM to create a variable data preprocessing approach.

Step 2.2: Determine the estimate geographical separation for the prevailing parameter dataset to find the highest Eigen value of the grid and construct the Mercer kernel matrix using equation (8).

Step 2.3: Using the cost function, for control quasi cumulative distribution matrix and use algorithm to retrieve the kernel matrix using equation (10).

Step 2.4: Equation (11) is used to determine cost function based on Step 2.2 and Step 2.3.

Step 2.5: Solve the equation (12).

Step 2.6: Remove duplication of effort and distracting noises from the input categorical variables.

Step 3: Apply equation (19) to the optimization and use equations (5) and (6) to obtain the additional features for equation (20).

Step 4: Using equation (17), modify the weight values and create a new computational model for equation (18).

4. Results and Discussion

Data samples were obtained and categorized into two sections for enhanced LSTM ongoing coaching and feedback. 3000 operation datasets of secondary variables and related temperatures in the devolatilization claiming process were chosen for training, while 1500 processes datasets were used as the testing dataset. In addition to increasing forecasting accuracy, a data preprocessing strategy based on the integration of GWO and LSTM is employed to remove the background and inconsistent data from the training dataset.

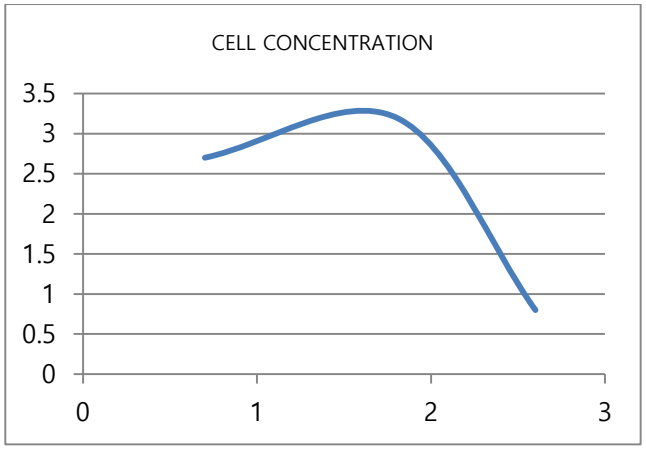

Figure 3: Cell concentration

Figure 3 displays the SO2 concentration prediction curve generated by the GWO-LSTM method. Clearly, GWO-LSTM can effectively follow the dynamic change of the attribute value, demonstrating the effectiveness of our technique. After repeated sessions in the same simulation testing process, all techniques in this study use a shared dataset with the same parameter optimization method. In almost the same engineering environment, several recognized concepts were emulated. The system for algorithm simulation was programmed using MATLAB 2019. The following standards are used to record the theoretical values:

(1) The set of possible analyzed parameters in the final procedure is referred to as overall size (MS)

(2) Predictive Mean Squared Error (PMSE) is a metric that depicts the gap between the estimated and expected value. It is determined using the following equation:

$$\mathrm{PMSE} = \frac{1}{n_t}\sum_{i=1}^{n_t}\left(y_i - \hat{y}_i\right)^2 \qquad (24)$$

where *yi* and *yi*^ are the actual and predicted values of the output variable, respectively. Parameter n_t is the number of datasets in the testing samples.

(3) The square of the collection correlation coefficients seen between actual worth and the forecast is known as the coefficient of determination (R^2).

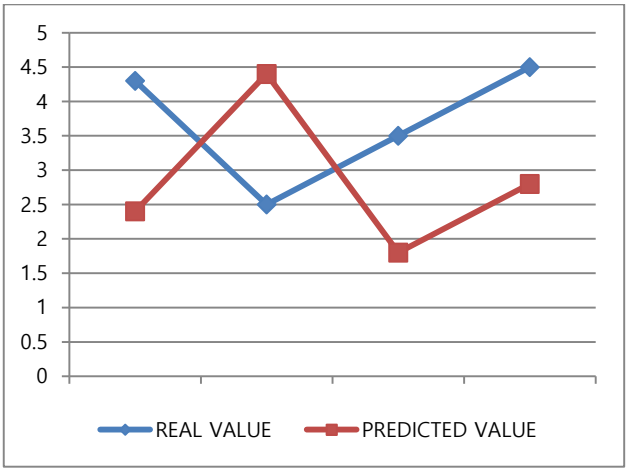

Figure 4: Real value versus Predicted values

Figure 4 illustrates the real and predicted values of the target attribute using the GWO-LSTM model. The curves clearly show that our method can successfully track variations in butane concentration, significantly demonstrating its effectiveness.

The gradient descent method is used to train the upgraded LSTM NN in the investigations with information extraction. The GWO and LSTM methods are used to choose the databases. Furthermore, the values are updated and determined using the Bayesian criterion to compare false detection curves [30]. The machine's learning model is used to retrain the algorithm once the weight characteristics have been quite well. To achieve a good effectiveness, the training rate is adjusted according with loss's alteration.

5. Conclusion

The purpose of this research is to create and build soft sensors using artificial intelligence approaches. In this research, a unique soft sensor is created using a mix of a Long Short-Term Memory (LSTM) network and Grey Wolf Optimization (GWO). In the learning technique, LSTM was created to deal with linear models with substantial nonlinearity and unpredictability of manufacturing applications. To reduce the model's improper complication, GWO is employed as an input optimization strategy for LSTM. Originally, the newly created soft sensor combined LSTM's outstanding dynamic modeling with GWO's precise variable selection. The performance of the proposed approach is demonstrated using simulations on real-world datasets.

References

[1] W. Yan, D. Tang, and Y. Lin, "A data-driven soft sensor modeling method based on deep learning and its application," IEEE Transactions on Industrial Electronics, vol. 64, no. 5, pp. 4237–4245, 2017.

[2] Z. Ujevic, I. Mohler, and N. Bolf, "Soft sensors for splitter product property estimation in CDU," *Chemical Engineering Communications*, vol. 198, pp. 1566–1578, 2011.

[3] K. Sun, J. Liu, J.-L. Kang, S.-S. Jang, D. S.-H. Wong, and D.-S. Chen, "Development of a variable selection method for soft sensor using artificial neural network and nonnegative garrote," *Journal of Process Control*, vol. 24, pp. 1068–1075, 2014.

[4] S. Gao, J. Yang, and J. Wang, "D-FNN based modeling and BP neural network decoupling control of PVC stripping process," *Mathematical Problems in Engineering*, vol. 2014, Article ID 681259, 13 pages, 2014.

[5] A. Nawaz, A. S. Arora, C. M. Yun, H. Cho, S. You, and M. Lee, "Data authorization and forecasting by a proactive soft sensing tool-anammox based process," *Industrial & Engineering Chemistry Research*, vol. 58, no. 22, pp. 9552–9563, 2019.

[6] S. Zhang, X. Chen, and Y. Yin, "An ELM based online soft sensing approach for alumina concentration detection," *Mathematical Problems in Engineering*, vol. 2015, Article ID 268132, 8 pages, 2015.

[7] K. Sun, X. Wu, J. Xue, and F. Ma, "Development of a new multi-layer perceptron based soft sensor for SO_2 emissions in power plant," *Journal of Process Control*, vol. 84, pp. 182–191, 2019.

[8] E. Heidari, M. A. Sobati, and S. Movahedirad, "Accurate prediction of nanofluid viscosity using a multilayer perceptron artificial neural network (MLP-ANN)," *Chemometrics and Intelligent Laboratory Systems*, vol. 155, pp. 73–85, 2016.

[9] K. G. Sheela and S. N. Deepa, "Neural network based hybrid computing model for wind speed prediction," *Neurocomputing*, vol. 122, pp. 425–429, 2013.

[10] Y. He, Y. Xu, Z. Geng, and Q. Zhu, "Soft sensor of chemical processes with large numbers of input parameters using auto-associative hierarchical neural network," *Chinese Journal of Chemical Engineering*, vol. 23, no. 1, pp. 138–145, 2015.

[11] M. Zabadaj, K. Chreptowicz, J. Mierzejewska, and P. Ciosek, "Two-dimensional fluorescence as soft sensor in the monitoring of biotransformation performed by yeast," *Biotechnology Progress*, vol. 33, no. 2, pp. 299–307, 2017.

[12] J. Rehrla, A. P. Karttunenb, N. Nicolaïc, T. Hörmanne, and M. Hornd, "Control of three different continuous pharmaceutical manufacturing processes: use of soft sensors," *International Journal of Pharmaceutics*, vol. 543, no. 1-2, pp. 60–72, 2018.

[13] Z. Ge, "Supervised latent factor analysis for process data regression modeling and soft sensor application," *IEEE Transactions on Control Systems Technology*, vol. 24, no. 3, pp. 1004–1011, 2015.

[14] B. Maschler, D. White, and M. Weyrich, "Anwendungsfälle und Methoden der künstlichen Intelligenz in der anwendungsorientierten Forschung im Kontext von Industrie 4.0," University of Stuttgart, 2020, DOI: 10.18419/opus-10740.

[15] S. Yin, W. Ji, and L. Wang, "A machine learning based energy efficient trajectory planning approach for industrial robots," Procedia CIRP, vol. 81, pp. 429–434, 2019, DOI: 10.1016/j.procir.2019.03.074.

[16] T. Müller, N. Jazdi, J.-P. Schmidt, and M. Weyrich, "Cyber-Physical Production Systems: enhancement with a self-organized reconfiguration management," Procedia CIRP, 2020 (accepted).

[17] A. Mayr et al., "Machine Learning in Production – Potentials, Challenges and Exemplary Applications," Procedia CIRP, vol. 86, pp. 49–54, 2019, DOI: 10.1016/j.procir.2020.01.035.

[18] G. Hussain, M. Jabbar, J.-D. Cho, and S. Bae, "Indoor positioning system: a new approach based on LSTM and two stage activity classification," *Electronics*, vol. 8, no. 4, p. 375, 2019.

[19] L. Yang, Y. Li, J. Wang, and Z. Tang, "Post text processing of Chinese speech recognition based on bidirectional LSTM networks and CRF," *Electronics*, vol. 8, no. 11, p. 1248, 2019.

[21] J. Chen, Q. Jin, and J. Chao, "Design of deep belief networks for short-term prediction of drought index using data in the Huaihe river basin," *Mathematical Problems in Engineering*, vol. 2012, Article ID 235929, 16 pages, 2012.

[22] S. Xingjian, Z. Chen, H. Wang et al., "Convolutional LSTM network: a machine learning approach for precipitation nowcasting," in *Proceedings of Advances in Neural Information Processing Systems*, pp. 802–810, Montreal, Canada.

[23] Z. Zhao, W. Chen, X. Wu, P. C. Y. Chen, and J. Liu, "LSTM network: a deep learning approach for short-term traffic forecast," *IET Intelligent Transport Systems*, vol. 11, no. 2, pp. 68–75, 2017.

[24] J. Liu, A. Shahroudy, D. Xu et al., "Spatio-temporal lstm with trust gates for 3d human action recognition," in *Proceedings of European Conference on Computer Vision*, pp. 816–833, Amsterdam, Netherlands.

[25] X. Yuan, L. Li, and Y. Wang, "Nonlinear dynamic soft sensor modeling with supervised long short-term memory network," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 5, pp. 3168–3176, 2019.

[26] Q. Sun and Z. Ge, "Probabilistic sequential network for deep learning of complex process data and soft sensor application," *IEEE Transactions on Industrial Informatics*, vol. 15, pp. 2700–2709, 2018.

[27] C. Bennett, J. F. Dunne, S. Trimby, and D. Richardson, "Engine cylinder pressure reconstruction using crank kinematics and recurrently-trained neural networks," *Mechanical Systems and Signal Processing*, vol. 85, pp. 126–145, 2017, DOI: 10.1016/j.ymssp.2016.07.015.

[28] S. Trimby, J. F. Dunne, C. Bennett, and D. Richardson, "Unified approach to engine cylinder pressure reconstruction using time-delay neural networks with crank kinematics or block vibration measurements," *International Journal of Engine Research*, vol. 18, no. 3, pp. 256–272, 2017, DOI: 10.1177/1468087416655013.

[29] R. Johnsson, "Cylinder pressure reconstruction based on complex radial basis function networks from vibration and speed signals," *Mechanical Systems and Signal Processing*, vol. 20, no. 8, pp. 1923–1940, 2006, DOI: 10.1016/j.ymssp.2005.09.003.

[30] H. Mrabet, E. Giacoumidis, I. Dayoub, and Aymen Belghith, "A Survey of Applied Machine Learning Techniques for Optical Orthogonal Frequency Division Multiplexing Based Networks", Emerging Telecommunications Technologies, ETT 2022, ISSN: 2161-3915, Wiley, November 2021, DOI: https://doi.org/10.1002/ett.4400.