Machine learning vs fuzzy inference methods for predicting the oil spill consequences with small data sets

Diana Kalibatienė[1], Anastasiya Burmakova[2]

Department of Information Systems, Faculty of Fundamental Sciences, Vilnius Gediminas Technical University, [1]diana.kalibatiene@vilniustech.lt, [2]anastasiya.burmakova@vilniustech.lt

ABSTRACT

- Big data is usually needed for machine learning and fuzzy inference methods to predict consequences in different application areas. However, in practice, learning, inferencing, and prediction from small data remain a key challenge in machine learning and fuzzy inference. No exception is the prediction of the oil spill consequences on the ground environment, for which only a limited data set is available.
- In this study, we have used several machine learning methods (support vector regression (SVR), Decision trees, Ensembles, and Gaussian Progress Regression) and the proposed adaptive neural fuzzy inference system (ANFIS) to predict the oil spill consequences on the ground environment from small data sets of real oil spill objects.
- An additional pre-training of methods was performed with the synthetic data obtained from the mathematical model.
- Results obtained during the experiments have shown that the proposed method can efficiently provide an accurate and sufficient prediction of the oil spill consequences on the ground environment.

MAIN ISSUES

1. The analysis of the existing approaches on predicting an accidental situations with small data shows that it is meaningful to apply deep learning and fuzzy-based prediction:
   - In **vague situations with small data** (i.e., not all initial data is known or interval values are known only), it is impossible to make an accurate predictions using linear equations and mathematical model.
   - Only **limited data sets** and **not all variable in a set** are available for accidental situations (i.e., oil spill on the ground), since we cannot intentionally create a situation in the real world.
   - The more fuzzy dataset input we have, the **more rules for prediction** are developed and the more linear calculations are needed for prediction.
2. There is lack of research on prediction of accidental situations with a small data → More research is needed to investigate existing approaches with small datasets.

THE LINEAR MATHEMATICAL MODEL

How much oil will get into groundwater?

- **A linear model** is created to simulate accidental oil spill situations and to generate synthetic data.
- Each layer is characterized by a set of variables, used for prediction of oil spill penetration consequences:
  - the shape of the pollution spot and the weight of the evaporated oil product;
  - the depth of penetration of oil products (OP) into Soil and Ground;
  - the weight of oil product adsorbed by Ground and its concentration;
  - the residual weight of the oil product that can reach Groundwater;
  - the time to reach the maximum concentration at the Groundwater level;
  - describe the horizontal redistribution of oil product with Groundwater.

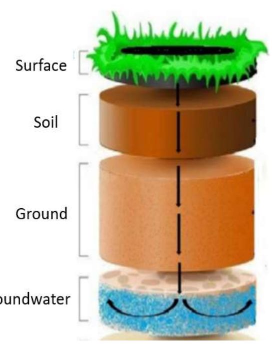

Figure 1. The Linear mathematical prediction model for accidental oil spill prediction

APPROACH

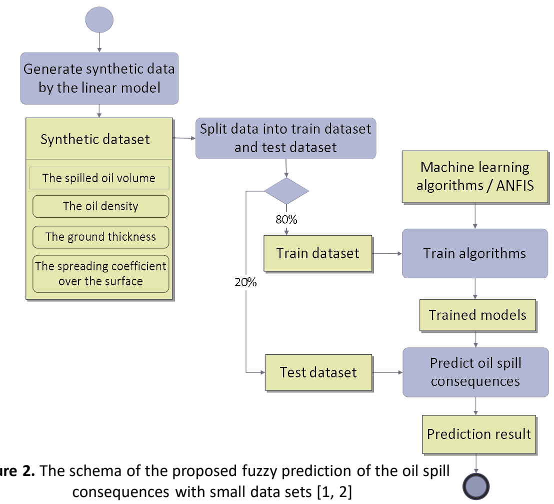

Figure 2. The schema of the proposed fuzzy prediction of the oil spill consequences with small data sets [1, 2]

ANFIS

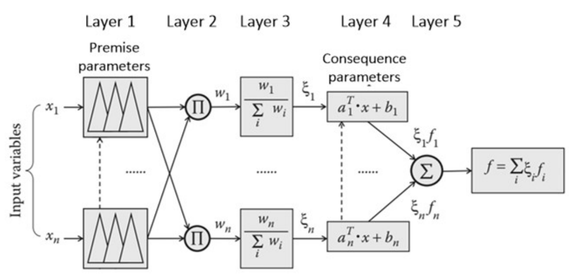

Figure 3. The basic the adaptive neuro-fuzzy inference system (ANFIS) structure [3]

Layer 1 – Fuzzification of inputs (x_1, ..., x_n)) using triangular membership function (MF). **Layer 2** – Evaluation of the rule strength (w_i). **Layer 3** – Normalizing the strengths of all rules ($\overline{w}_i$). **Layer 4** – Applying the rule (R_i) to obtain the output f_i. **Layer 5** – Computing the global model response (f).

$$R_i : if\ \left(x_1\ is\ A_1^{(i)}\right) \ldots and \ldots \left(x_n\ is\ A_n^{(i)}\right)\ then\ f_i = a_i^T \cdot x + b_i$$

Initial parameters for ANFIS-model [2]:

1. The Gaussian membership function.
2. Five terms were set (how many interval values there will be for the input data).
3. Five learning epochs (how many times the ANFIS model will be retrained).

RESULTS

1. For predicting we have used machine learning methods (linear regression, support vector regression (SVR), Decision trees, Ensembles, and Gaussian Progress Regression). The best results among those methods is shown by Gaussian Progress Regression with RMSE ~ 4,3%.
2. The proposed approach of fuzzy prediction of the oil spill consequences with small data sets is implemented by incorporating ANFIS and an experiment is conducted (RMSE ~ 1%).
3. Experiment results are as follows:
   - **Result A:** shows the dependency between ***input1 (the ground thickness)***, ***input2 (the spilled oil volume*** (experimentation interval [10m^3; 10000m^3]***)*** and ***output*** (***the penetrated OP***). The output significantly depends on input1, i.e., the ambiguity of the OP penetration into the groundwater is caused by a change in the ground type, which is expressed through the ground thickness.
   - **Result B:** shows the dependency between ***input2 (the spilled oil volume)***, ***input4 (the oil density)*** and ***output*** (***the penetrated OP***). The output depends on input4, i.e., the same volume of oil of different density is spread on different surface, and the larger the surface is, the less oil will penetrate into the ground. Ultimately, when a very large oil volume spilled (~5000m^3), it will pass into the groundwater.

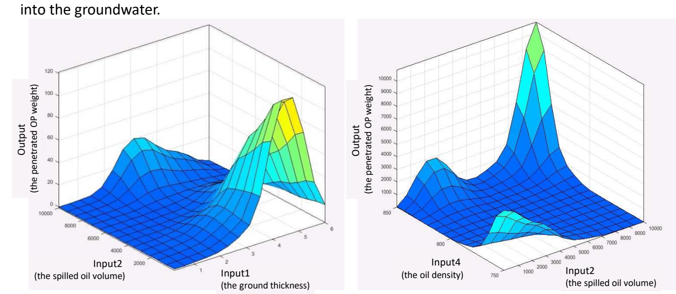

Figure 4. Results of oil penetration prediction with ANFIS (Result A and Result B)

Table 1. Quality of prediction results of machine learning algorithms and the proposed model with ANFIS

Method	R [4]	R² [4]	RMSE [5]
Linear Regression	0.67	0.45	14.2%
Support vector regression (SVR)	0.74	0.54	19.1%
Decision trees	0.57	0.32	15.8%
Ensembles	0.37	0.14	17.7%
Gaussian Progress Regression	0.97	0.95	4.3%
The proposed model with ANFIS	0.99	0.99	1.0%

REFERENCES

1. Kalibatiene, D., & Burmakova, A. (2021). Fuzzy Model for Predicting Contamination of the Geological Environment During an Accidental Oil Spill. *International Journal of Fuzzy Systems*, 1-15.
2. Burmakova, A., & Kalibatiene, D. (2021, April). An ANFIS-based Model to Predict the Oil Spill Consequences on the Ground. In *2021 IEEE Open Conference of Electrical, Electronic and Information Sciences (eStream)* (pp. 1-6). IEEE.
3. Jang, J. S. (1993). ANFIS: adaptive-network-based fuzzy inference system. *IEEE transactions on systems, man, and cybernetics*, 23(3), 665-685.
4. Ranković, V., Radulović, J., Radojević, I., Ostojić, A., & Čomić, L. (2010). Neural network modeling of dissolved oxygen in the Gruža reservoir, Serbia. *Ecological Modelling*, 221(8), 1239-1244.
5. Willmott, C. J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate research*, 30(1), 79-82.