Emission prediction of a thermal power plant

Juraj Jurčo
KD Lab FI MU
173001@mail.muni.cz

Lubomír Popelínský
KD Lab FI MU
popel@mail.muni.cz

Karel Křehlík
OT Energy Services a.s., Třebíč
kkrehlik@otenergy.eu

Abstract. The task of prediction of emissions is very challenging and also important. We argued that simple learning techniques that learn only one predictive model are not powerful enough in more complex situations. Better predictive results can be achieved by splitting data into smaller parts and for each part to learn a sub-model. We proposed and tested a novel method that combines meta-learning and ensemble learning. We showed that there is significant increase in prediction accuracy.

Keywords: meta-learning, model prediction, boiler, NO_x

1 Introduction

The production of emission is nowadays huge problem around the world. Latest results published by Intergovernmental Panel on Climate Change (IPCC) group have discovered that human influence is the dominant cause of the observed warming with confidence of 95% [1]. The main contributor of pollution is burning fossil fuel, including coal. Reduction of emissions produced by humans together with preserving current power consumption standards is hard to reach without an extra financial aid.

Description of combustion. Emission produced by boiler is not only carbon dioxide but there are various types of exhalations. During incomplete combustion carbon monoxide (CO) is produced which is poisonous for living beings. Other dangerous gas produced is nitrogen oxide (NO_x) that destroys ozone layer. There are various ways how to decrease these pollution causing exhalations. One of the ways are secondary methods for reduce NOx emission. First of them is selective non catalytic method (SNCR), which through the reagent spraying system (NH_3 or CH_4N_2O) into the optimal temperature window, reduces NO_x emission. SNCR method has negative aspects, which are residual NH_3 emissions in flue gas in the form of NH_3 gas and solid particles after sulfur and NH_3 reaction. Second way is selective catalytic method (SCR), which use catalyst for the reduction of the NO_x emission. This is more suitable for environmental but also more expensive investment and high operational costs.

Optimization of combustion. The elementary (primary) way for the NO_x emission reduction is improved combustion. There are several ways, from completely rebuilding combustion system (high investment costs), to improve existing control system algorithms, in order to improve parameters of combustion system.

This main idea of better combustion, which has been elaborated in this paper, is to build the mathematical model from historical data. This model is then used for emission prediction and that knowledge is further exploited for finding combustion optimization criteria, i.e. for the finding better control curves. Implementation of an improved version

of the control system is the final step. There are several ways how to validate such a system. First of them is to test it at the real technology what requires a very good knowledge of the behavior of technology. The other way of testing is indirect - on CFD models. CFD models can provide very good insight into technology behavior (combustion in boiler). We can also easily make several tests to check the optimal of the control system.

Machine learning approach. If we were able from historical data to predict emissions in future it would be easier to optimize them. The focus of our work is to find model for emission prediction with support of machine learning techniques. For model validation we use historical data that were not used for learning model. We implemented this method on one of six boilers of coal power plant Melnik I. There are six pulverized granulation boilers with hammers mill. By optimization there are no extra fees required and burning is more economy so overall budget may be reduced. This solution is also nature-friendly because we can optimize tradeoff between exhalations and energy production.

The structure of the paper is the following. In Section 2 we describe the raw data and explain basic data transformation. In Section 3 we introduce a novel method for emission prediction based on learning multiple models. The results are discussed in Section 4. We conclude with concluding remarks in Section 5.

Related work. Prediction in time series is a common task in time series analysis [2, 3]. A use of Support Vector Machines for predicting NO_x, the flue gasses emitted by power generation plant is solved in [4]. Staviri et al. created mathematical model for prediction of emissions of a coal boiler where they reached constant difference between their mathematical model and real values 1 - 1.8% [5].

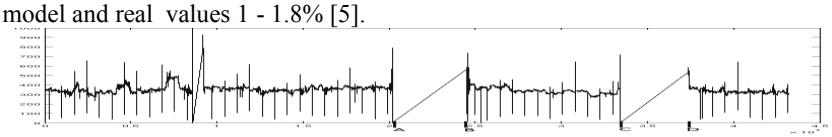

Fig. 1. Course of emissions of NO_x

2 Preliminary analysis and data transformation

2.1 Data description and baseline

We used real data from a coal power plant. Boilers in this power plant are pulverized, granulation boilers with hammer mills. Data contained 26 input attributes about temperature of combustion, air pressures, air flows, settings of flaps and output values of flue gases O_2, CO, NO_x and SO_2. If the process of combustion is varying over time, i.e. the process may be influenced by different factors, the task becomes more difficult. There are at least two different processes in data. You can see them in the Fig. 1– one between time A and B, and C and D as linearly increasing amount of emissions, and second one as all other data. The first process determines state when sensors are out-of-order or boiler broken.

2.2 Mean values of input attributes

Some of 26 attributes can be aggregated because they measure similar input attributes, e.g. Setting of the secondary air inlet 1 – 4; Primary air flow, corresponds to the fuel supply from the mill 1 -- 4 (setpoint); Current motor mill; Setting of the secondary air inlet

Order	*Attribute name*
1	Temperature of flue gas L+R
2	Steam power of boiler (steam pressure in the steam piping)
3	Air pressure before central air fun
4	Primary air flow intake (mill 1 – 4)
5	Current motor mill 1 – 4 (setpoint)
6	Setting of the secondary air inlet flap number 1 – 4
7	Setting of over fire air inlet valve number 1 – 2
8	Setting of flaps of core air intake, fuel supply of mill 1 – 4
9	Setting of flaps of tertiary air supply number 1 – 2
10	Air pressure behind LUVO2 L+P
11	NO_x in the flue gas

Fig. 2. List of new input attributes

valve number 1 -- 4 and so on). Because of the similarity of attributes we transformed data and computed new features -- mean values of similar attributes. Thus we received a dataset with ten input attributes and one predictive value – NO_x. This transformation resulted with more robust description of the combustion process. The list of new input attributes is in the Fig. 2.

2.3 Data transformation

The combustion is not impulsive. It is rather a process that requires some time before the effect of input changes can be observed. Dataset in its original form was composed of single-point values. Based on experience, it is worth to consider 10 minutes interval for prediction. As it takes time while a setting of boiler takes effect - about 5 minutes in our case - we predict, from 10 minutes interval, the emissions with 5 minutes delay. To consider that 10 minutes history for analysis we computed six statistics from this time window - mean value, standard deviation, slope of regression line, Root Relative Squared Error (RRSE) from regression line and two quartiles – lower and upper, see Fig. 3. RRSE, sometimes also referred as Root Mean Squared Deviation, is a measure of the differences between predicted values and the observed values. Quartiles are values that split sorted data into four intervals. Mean of the lower quartile is that 25% of values are lower than this value; Upper quartile means that 75% of data are lower than this value. Those statistics were computed only for input attributes; there were no statistic window for target variable.

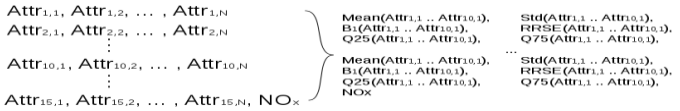

Fig. 3. Computation of 10 minutes statistics and 5 minutes class shift

Then we learned various classifiers on this dataset. As a machine learning tool we used Weka [5]. We split data into training and test set in 66:34 ratio so that two thirds of data were used for training and the rest for validation of the learned model. As a learning algorithm we used REPTree, the fast regression learner that builds pruned regression tree using information gain. For pruning it uses reduced-error pruning. To increase efficiency it sorts values of numerical attributes only once. When applied REPTree to the data described above the correlation of a new model with mean dataset was 0.53. When comparing with correlation 0.42 on original data – i.e. without aggregation – there was increase of correlation about 11%. As shown before, the transformations of the dataset improved results of prediction. When we considered time window and class attribute shift as it was proposed by domain expert, error prediction was lowered. As we can observe there are some improvements from previous results. Unfortunately these improvements are not sufficient enough and they are not useful for real prediction of emissions from boiler's settings.

3 Learning multiple models

This led us to idea to learn not only one model but more of them. Models would be learned on shorter time period, e.g. 1 or 3 hours. Such models could predict emissions better because they would be used only for prediction on data that would be similar to those on what they were trained. The method lies in splitting training set into smaller parts of the same length called bins. The length of bins was estimated to preserve real world situation. For our tests we used 1, 3, 6, 12 and 24 hours. A new model for each time interval was learned and tested separately. The best results were obtained for interval of length 1 hour so in the rest of the paper we will consider just this bin size. Separate models were trained for all bins in the training set. The size of the whole training set was 3 weeks and number of bins and models trained here was 504 (3 weeks × 7 days × 24 hours). All trained models were stored in memory. The phase of training of all modes was followed by testing phase. The test set was of size 9 days and it was also split into bins of the same size like training set. Total number of tested bins was 216. On each of those 216 bins was tested all 504 models built in learning phase. Then from all models that have been tested the best one was selected. For the real application the selection of the best model could be based on meta-learning technique [6]. There are two main streams in meta-learning. One uses meta-characteristic of data (e.g. the number of examples or simple statistics about the data) and aims at generalization of behavior of learners in order to predict the best one for a new data set that has been described with the same meta-characteristics. The other is based on landmarking [7] and its variant, sampling-based landmarking. The idea of sampling-based landmarking [8] is simple, from performance (usually from accuracy) of a collection of learning algorithms on smaller portions of data to predict which algorithm will perform the best on the whole data set. It has sense only in the case of prediction of the meta-learner being reliable and the price or landmarking (usually the training time) being significantly smaller than testing every algorithm from the collection on the full data. In the time when we would like to predict emissions we already know the recent past history. When we use bin size of one hour, this history is also size of one hour. Then we compute meta-attributes of this recent history. We look into the past, where the boiler already was in the most similar state. We train model for prediction on the following bin of this past bin and we use it for the prediction of following values.

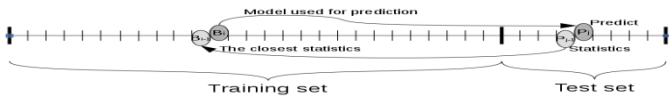

Fig. 4. Selection of model based on meta-attributes

The proposed method is illustrated in the Fig. 4. 0For predicting emissions of test bin P_j we need to find the closest model for the previous interval P_{j-1} – let it be B_{i-1}. Now we use model that had the best performance on the bin B_i. The goal now is to find a method how to choose the best models. A choice of the best N models is based on similarity of training and test set (using statistical characteristics of two bins) and ensemble classification.

4 Results

The best result was obtained when splitting training set into bins of one hour size as mentioned. The selection of the proper model for prediction helped to increase correlation and also to lower error of the prediction (RRSE). The best average correlation was 0.66 with standard deviation 0.16. More detailed information – all correlations for all bins – is in

the Fig. 5. Axis X represents time (bins) and axis Y correlation. Horizontal line represents value of 0.75 that is some kind of border line for real application. When we compare it to classical method with correlation of 0.53 (see Section 2) the increase reached 13%.

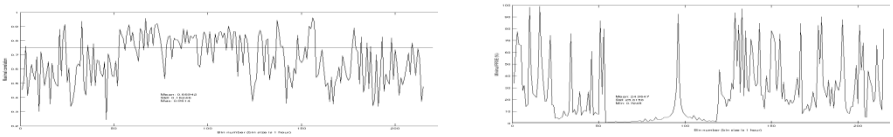

Fig. 5. Correlation (left) and RRSE (right) of predicted values for 10 minutes statistics and 5 minutes class shift

Mean value is around 25.8% with standard deviation of 25.9. From the graph it is clear that for some bins there was no one good model trained and the error is quite high (value of RRSE over 50%). Our dataset was limited to 30 days but in the case of longer history there could be trained more models that would fit better. Results of this method were better than classical method of values prediction but still not satisfying. The character of the dataset is that there is difference between sampling frequency and preprocessing of input attributes and class attribute. Input attributes are measured with frequency of one second while class attribute in frequency of one minute. This difference was equalized by computing moving average of input attributes to one minute. The class attribute remain in raw values. It means that this attribute may contain noise and it could confuse classifiers and make prediction harder. To lower an effect of noise in class attribute – NO_x – we computed 10 minutes interval of those values. The learned function now predicts mean value of NO_x on 10-minutes interval. A result for this data was even better; the mean correlation increased to value 0.77 and the standard deviation has lowered to 0.11. Results for all tested bins are shown in the Fig. 6.

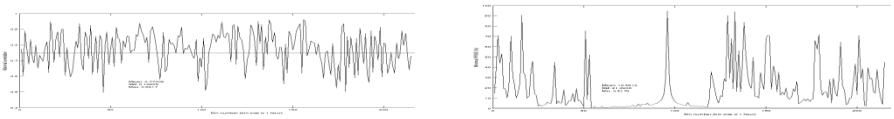

Fig. 6. Correlation (left) and RRSE (right) of predicted values for 10 minutes statistics and 10 minutes average value with 1 minute class shift

Computation of the mean value of class attribute had also impact on RRSE. From previous value 25.8 it decreases to 18.6 with also lower standard deviation of 21 with comparison to previous 25.9. As we can see there are still some bins for that prediction was not as successive. But results show that noise present in class attribute can be partially removed by computation of mean value. We provided also some additional tests, with computation of shorter mean value of class attribute, but results were worse. Comparison of real values with the values predicted by the best classifier is shown in the Fig. 7.As we can see predicted values approximates data quite good.

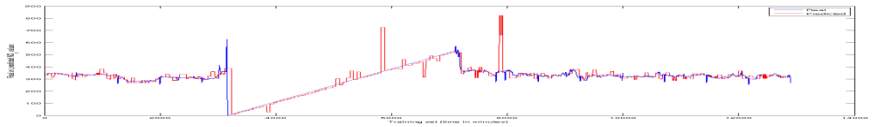

Fig. 7. The best real predicted values for 10 minutes statistics and 10 minutes average value with 1 minute class shift

5 Conclusion and future work

The task of prediction of emissions is very challenging and also important. We argued that comparisons to the classical simple learning methods that learn only one predictive model are not powerful enough in more complex situations. The best correlation achieved by classical methods was 0.42 or 0.53 respectively for mean values of attributes. Better predictive results can be achieved by splitting data into smaller parts and where for each part to learn a sub-model. We proposed a novel method that combines meta-learning and ensemble learning. We showed that there is significant increase in prediction accuracy. By this method we achieved even better results and the increase was quite significant. The mean value of correlation for 10 minutes statistical window and 5 minutes class shift was 0.66 and when we predicted also 10 minutes mean value of output emissions we reached correlation 0.77 with original values. The proposed novel method for prediction of emissions of the boiler lighted the path how prediction of values can be done. During our analysis were used data contained both states of the boiler as described in Chapter 2. It is hard to learn one classifier that is able to classify well both processes. It is a reason why it would be better first to recognize a current process and then learn a classifier for that particular part of the time series.

6 Acknowledgements

The work has been supported by Faculty of Informatics, Masaryk University in Brno, Czech Republic. We also thank to all members of Knowledge Discovery Lab FI MU for fruitful discussions and to Rastislav Mirek for data cleaning and to Zdeňka Sitová for preliminary experiments.

7 References

1. IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
2. Han, J., and M. Kamber. 2011. Data Mining: Concepts And Techniques: Concepts And Techniques.
3. Mitsa, Theophano. 2010. Temporal Data Mining. 1st ed. Boca Raton, FL: Chapman & Hall/CRC.
4. Saiful Idzwan B., Chen Chai Phing, Tiong Sieh Kiong. 2014. Prediction Of Nox Using Support Vector Machine For Gas Turbine Emission At Putrajaya Power Station, Journal of Advanced Science and Engineering Research Vol 4, No 1 March (2014) 37-46
5. Staviri, V.-F.; Eduard, G. Comparative study on the calculation of nitrogen oxide emissions, Electrical and Power Engineering (EPE), 2012 International Conference and Exposition on, Issue Date: 25-27 Oct. 2012
6. Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, Ian H. Witten (2009); The WEKA Data Mining Software: An Update; SIGKDD Explorations, Volume 11, Issue 1.
7. Brazdil, Pavel B. 2009. Metalearning. 1st ed. Berlin: Springer.
8. Pfahringer Bernhard and Bensusan, Hilan: Meta-learning by Landmarking Various Learning Algorithms, ICML 2000.
9. Leite, R. and Brazdil, P.: Active Testing Strategy for Best Classication Algorithm via Sampling and Metalearning. In Coelho H et al. (Eds.) Proceedings of ECAI 2010