

A Data-driven Approach to Improve Control Room Operators' Response

Nicola Tamascelli^{a,b,*}, Giordano Scarponi^a, Nicola Paltrinieri^b, Valerio Cozzani^a

^aDepartment of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, Bologna, Italy

^bDepartment of Mechanical and Industrial Engineering, NTNU, Trondheim, Norway

nicola.tamascelli2@unibo.it

Digitalization has significantly improved productivity and efficiency within the chemical industry. Distributed Control Systems and extensive use of sensor networks enable advanced control strategies and increase optimization opportunities. On the other hand, chemical plants are increasingly complex, equipment is highly interlinked, and it is more difficult to describe the system dynamics through first principles. Finding the root causes of process upsets and predicting dangerous deviations in process conditions is often challenging. Advanced and dynamic tools are needed to grant safe and stable operations in such a complex and multivariate environment. In this context, Machine Learning techniques may be used to exploit and retrieve knowledge from the large amount of data that chemical plants produce and store on a daily basis. Data-driven methods may be adopted to develop predictive models and support a proactive approach to process safety. The study aims to develop Machine Learning techniques to improve the response of control room operators during critical events. Specifically, alarm data originated in an upper-tier Seveso site have been collected, cleaned, and analyzed to identify periods of intense alarm activity. Alarm behavior following operator responses has been evaluated to assess whether the actions were adequate to prevent future alarm occurrences. In doing so, alarm events that reoccur within 30 minutes after an operator acknowledgment have been identified and labeled. Subsequently, a hybrid classification algorithm was trained to predict the probability that a critical alarm reoccurs after being acknowledged by the operator. This predictive tool might be used to support the operator's decision-making process and focus his/her attention on critical alarms that are more likely to occur again in the near future.

1. Introduction

The alarm system is one of the first layers of protection to prevent process deviations from escalating into accidents (Stauffer and Clarke, 2016). Still, there are inherent difficulties in designing, operating, and maintaining an efficient alarm system (Goel et al., 2017), which includes both technical (e.g., sensors, DCS, actuators) and human functions (e.g., operators). Alarms inform control room operators about dangerous deviation from normal operating conditions so that appropriate corrective actions could be taken. On the other side, operators should be provided with enough time to detect the issue, diagnose the situation, and determine/implement corrective actions (ANSI/ISA, 2016). Still, manual intervention by operators is subject to human error; improper procedures, worker fatigue, and lack of operator training may prevent an adequate response (Exida, 2009). In fact, several accident reports have highlighted that improper alarm management and inaccurate operator actions play a significant role in the development of process accidents. For example, poor alarm prioritization and an excessive alarm annunciation rate contributed to the Texaco Milford Haven explosion, where 26 workers were injured (Health and Safety Executive, 1997). Also, the non-detection of a loss of coolant led to the Three Mile Island Accident (United States Nuclear Regulatory Commission, 2018), where alarms rang, warning light flashed, but no operator could diagnose the situation. In an attempt to rationalize and provide a methodology for a more effective design and management of alarm systems, standard manuals have been published, such as ISA 18.2 (ANSI/ISA, 2016) and EEMUA 191 (EEMUA, 2013). Still, much remains to be done (Goel et al., 2017). The advent of the Third Industrial Revolution has already

brought changes and improvements to the industry. Chemical plants are more productive, more automated, more flexible, and safety systems are more advanced. Nevertheless, some issues have arisen as well. Modern DCS allows the configuration of new alarms with few clicks of mouse (Katzel, 2007). As a result, a larger number of alarms are installed, and the workload for operators has increased to the point where they are often overwhelmed by nuisance alarms (Kondaveeti et al., 2013). In addition, chemical plant complexity has increased, and control/safety functions are more intricate. As complexity increases, failures are more likely to occur, and root causes of process upsets are more difficult to detect (Wall, 2009). Therefore, it may be challenging for control room operators to find the appropriate set of corrective actions in a reasonable time. An intelligent tool to assess and predict the effectiveness of operators' actions would be of great support in dealing with critical situations. In this context, advancements in IT, IoT, and computer science have led to the development of intelligent computer-based algorithms to extract knowledge from data and support knowledge-based decision-making. In fact, a massive amount of process and alarm data are produced and stored on a daily basis (Reis and Kenett, 2018). Machine Learning algorithms may be used to "mine" these data and create predictive models for, e.g., fault detection e fault diagnosis (Tian et al., 2015), predictive maintenance (Carvalho et al., 2019), Dynamic Risk Assessment (Paltrinieri et al., 2019), modeling and simulation (Aleixandre et al., 2015).

This work focuses on the application of Machine Learning techniques for predicting the probability that a critical alarm reoccurs after being acknowledged by an operator. In this way, the operator's attention would be driven to alarms that are more likely to occur again in a short time. Alarm data from an ammonia production site have been used to evaluate the proposed methodology.

2. Alarm database: structure and features

Alarm records originated in an ammonia production plant have been used to test the proposed methodology. Alarms that occurred between July and November 2017 have been extracted and stored in a CSV file (i.e., the alarm database), which contains 26,473 alarm records described by means of 39 different attributes. That is, the database may be considered a 26,473x39 matrix, where each row represents an alarm event, and each column represents an attribute (i.e., a feature) of an alarm event. A reduced version of the database was described and used by Tamascelli et al. (2020a, 2020b). In the present study, the whole dataset has been used. Most of the alarms in the database occurred between 09/09/2017 and 09/10/2017, when a total power outage forced an emergency plant shutdown. During this critical event, the alarm annunciation rate often exceeded 1000 alarms/day, and the workload for control room operators increased drastically.

Each alarm event is described by a list of attributes. A comprehensive description of the attributes may be found in Tamascelli et al. (2020b). However, only three attributes are needed for uniquely identifying an alarm event:

1. Timestamp;
2. Source;
3. Identifier.

The Timestamp is the date and time of the alarm occurrence. The Source represents the instrument or logic function that generated the alarm. An example of a Source is LI315 (i.e., the level indicator in the control loop 315). The Identifier indicates the alarm status. Nine different identifiers are found in the database, as shown in Table 1.

Table 1: Alarm Identifiers

Identifier	Meaning
HHH	The measured variable has exceeded the high alarm setpoint
HTRP	The measured variable has exceeded the very-high alarm setpoint
LLL	The measured variable has exceeded the low alarm setpoint
LTRP	The measured variable has exceeded the very-low alarm setpoint
ALM	Generic alarm
IOP	Instrument failure or out-of-range measure
ACK	The operator has acknowledged the alarm
NR	A generic alarm is terminated (it refers to an earlier ALM alarm)
Recover	Alarm terminated (it refers to an earlier HHH, HTRP, LLL, LTRP, or IOP alarm)

In addition to alarms per se, the database keeps track of two different events: the acknowledgment of an alarm by an operator and the recovery of an alarm. The former is described by the Indicator “ACK”, the latter by “Recover” or “NR” depending on the Identifier of the original alarm.

3. Methodology

The method follows three main steps: Data pre-processing, Target identification, and Machine Learning simulations.

3.1 Data pre-processing

Alarm attributes (i.e., columns of the database) that have not been deemed relevant for the analyses have been discarded. For example, empty columns have been removed, as well as columns that show the same value for each event in the database. Also, redundant attributes have been removed.

Next, missing values have been substituted by the value “0”. This has been done because most Machine Learning algorithms do not tolerate missing values (Brink et al., 2016). The choice of the value “0” is arbitrary; a different numerical or categorical value (i.e., a text string) would be equally effective (Han et al., 2012). In doing so, one must ensure that the chosen value is outside of the domain of the attribute affected by missing values (i.e., the attribute should never take values equal to the one selected for the substitution).

Finally, it may happen in industrial databases that different measurements have different units. For example, one may find that some of the pressure measurements are expressed in “bar”, while others in “atm”. Whenever this happens, it is critical to ensure that attribute values referring to homogeneous physical quantities are expressed into common measurement units. Also, numerical values should be normalized in order to suppress scale effects (e.g., using min-max scaling) (Brink et al., 2016).

3.2 Target identification

The database must be analyzed to find and highlight events where an operator has acknowledged an alarm, but still, another alarm from the same Source occurs within 30 minutes. Events that meet this criterion are called Target events. The time window has been selected in accordance with the approach mentioned in the PETRO-HRA Guideline, which evaluates the 30 minutes criterion as the time required for action from the operator (Stauffer and Clarke, 2016). A binary categorical variable is assigned to each event in the database to highlight Targets. The binary variable is called the Label of an alarm event, and it assumes the values “YES” or “NO” depending on whether the event is a Target or not. Therefore, if the database contains n events, n Labels are generated, stored in a vector, and appended to the alarm database. Table 2 clarifies the role of Labels.

Table 2: Fictitious alarm sequence from LI315.

Timestamp	Source	Identifier	Attribute 4	...	Attribute n	Label
01/01/2021 00:00:00	LI315	LLL	---	...	---	NO
01/01/2021 00:03:00	LI315	LLL ACK	---	...	---	NO
01/01/2021 00:19:00	LI315	LLL Recover	---	...	---	NO
01/01/2021 01:30:00	LI315	LLL	---	...	---	NO
01/01/2021 01:15:00	LI315	LLL ACK	---	...	---	YES
01/01/2021 01:40:00	LI315	LTRP	---	...	---	NO

The table shows a fictitious alarm sequence from LI315. Data are organized as described in section 2, except for the last column, which contains the Labels. The second-last event of the series has “YES” as a label since another alarm from the same Source (LTRP) has occurred less than 30 minutes after acknowledging the previous low-level alarm (LLL). On the contrary, the first ACK event in Table 2 has “NO” as a label because the alarm has been recovered after 16 minutes (LLL Recover).

3.3 Machine Learning simulations

A Wide&Deep classification model has been trained and evaluated on the alarm database. The purpose of the algorithm is to classify alarm events into two categories: Target (i.e., Label = “YES”) and Not Target (i.e., Label = “NO”). That is, the model would predict whether an acknowledgment will be followed by another alarm from the same Source (Label = “YES”) or not (Label = “NO”). The workflow to set up and perform the Machine Learning simulations is illustrated in Figure 1. Two steps must be followed to complete a classification task: training and evaluation. During training, $\frac{2}{3}$ of the alarm database is fed to the Wide&Deep model (arrow T1 in Figure 1), which “learns” the relationship between the features of an event (i.e., the attributes) and its Label.

The process involves the joint optimization of two distinct models (Cheng et al., 2016): a Linear model and a Deep Neural Network. The structure of a Wide&Deep model is illustrated in Figure 1. Mathematics and technical details behind the model are out of the scope of this work and may be found in Cheng et al. (2016).

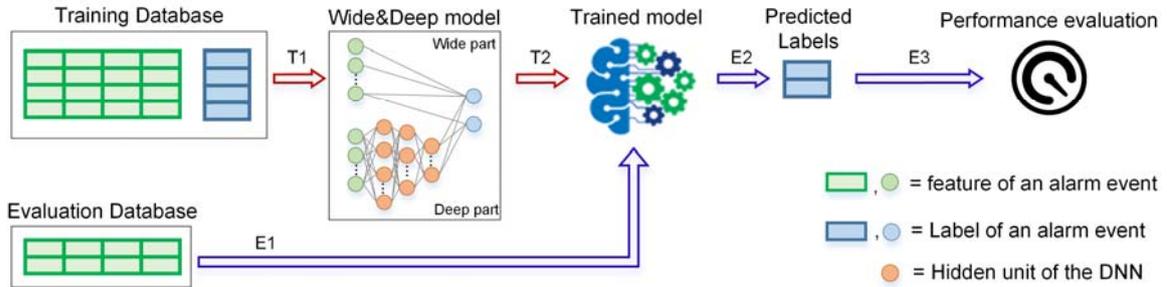


Figure 1: Training and evaluation of the Wide&Deep model.

In general, the algorithm aims at optimizing the internal parameters of a function f to best represent the relationship between features of an event (X), and its Label (Y):

$$f(X) \approx Y \quad (1)$$

The function f in Eq(1) comprises a linear part, where features are linearly combined and mapped into Labels, and a Deep part, where features are linearly combined and transformed into derived features (i.e., hidden units or “neurons” of the DNN) through nonlinear transformations. The parameters used to set up the model are the same as those used by Tamascelli et al. (2020b). After training, a trained model is obtained (T2 in Figure 1). Next, the model is evaluated. Labels are removed from the rest of the database, which is fed to the trained model (E1 in Figure 1). The algorithm takes as an input the features of each event and returns the probability of the Label being “YES” or “NO” (E2 in Figure 1). By default, a probability decision threshold equal to 0.5 is used to convert Label probabilities into Labels (i.e., if the probability of Label “YES” is greater than 0.5, the event will be labeled as “YES”). Finally, predicted Labels are compared with true Labels to assess the model performance.

4. Results

The target identification procedure (step 3.2) highlighted that a total of 119 events meet the requirements to be classified as Target. The training database comprises 17,649 alarm events, of which 78 belong to the Target category. The evaluation database contains 8824 events, of which 41 belong to the Target category.

After the evaluation phase, three metrics have been calculated in order to assess the performance of the model:

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \mathbf{0.995} \quad (2)$$

$$\text{Precision} = \frac{TP}{TP+FP} = \mathbf{0.5} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} = \mathbf{0.049} \quad (4)$$

Where TP identifies a True Positive (i.e., the model predicted the label “YES”, and the true Label of the event is “YES”), a TN identifies a True Negative (i.e., predicted Label = “NO”, real Label = “NO”), FP identifies a False Positive (predicted Label = “YES”, real Label = “NO”), and FN identifies a False Negative (i.e., predicted Label = “NO”, real Label = “YES”). Therefore, the sum of TP and TN indicates the number of correct predictions, while FN and FP show the number of wrong predictions.

5. Discussion

Metrics presented in Eq(2), Eq(3), and Eq(4) indicates that 99.5 % of the prediction were correct. Nevertheless, this result does not imply satisfactory performance. In fact, only 41 out of 8824 events in the evaluation database have “YES” as a true Label. Therefore, the model would have achieved an Accuracy greater than 99 % by always predicting the Label “NO”. This happens because the dataset used for the simulation is imbalanced, meaning that there are only a few examples of Target events within the database. In

this situation, Precision and Recall offer more information than Accuracy. Furthermore, it should be considered that mislabelling a Target event as a non-Target one is more critical than labeling a “NO” event as “YES”. Thus, Recall is the most meaningful metric to consider in this specific task. A large precision is desirable but not as important as a high Recall. Precision in Eq(2) indicates that 50 % of the “YES” predictions were correct, and the Recall in Eq(4) shows that only 4.9 % of the Target event were correctly identified (i.e., 2 out of 41). Evidently, the rarity of Target events must have affected the learning process and contributed to the uncertainty of predictions. However, these metrics have been calculated using a standard probability decision threshold equal to 0.5. There is no guarantee that this value will lead to the best performance. Thus, the decision threshold has been varied from 0 to 1; every time the threshold has been changed, predicted labels are calculated again, and so are Precision and Recall. Figure 2 illustrates how Precision and Recall change with the decision threshold. Lowering the threshold to 0.012 would increase the Recall from 0.049 to 0.9 and decrease the Precision from 0.5 to 0.34. As previously mentioned, Recall is the most important metric to consider in this problem. Thus, the performance would significantly improve since the increase in Recall is five times larger than the decrease in Precision.

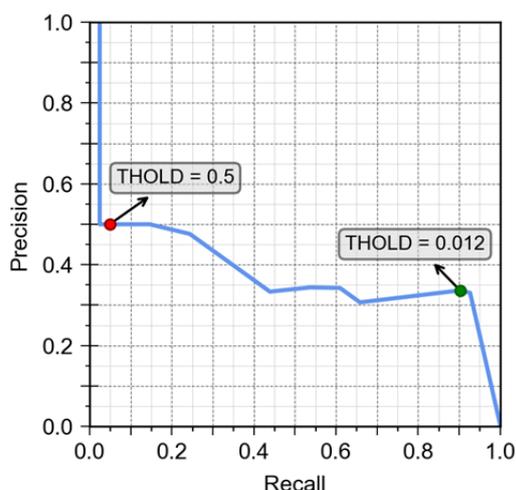


Figure 2: Precision–Recall curve produced by the Machine Learning simulation. Coordinates of points on the curve represent Precision and Recall obtained using different decision thresholds (THOLDS). The Red mark represents the point at threshold = 0.5. The green mark represents the point at threshold = 0.012.

It is worth noting that 0.012 is a relatively small threshold. However, most of the predicted probabilities are lower than 0.012. More than 80 % of the prediction probabilities are smaller than 0.1, and only 110 events over 8824 have a probability larger than 0.012. This suggests that the model is relatively unconfident, which may be due to the rarity of the event considered. Still, 90 % of the Target event lies within those 110 events, which is an encouraging result considering the size of the dataset. In this situation, lowering the threshold to such a low value seems an acceptable compromise considering how probabilities are distributed. Future works should investigate whether training the algorithm with more alarm data would partially overcome the issues related to the rarity of Target events and eventually improve the model confidence. Additional tests should also be performed to assess whether different sets of features or different Machine Learning models would be better suited for the problem under assessment. Moreover, it is worth stressing that the analyses rely entirely on historical alarm data. Further tests are needed to assess the algorithm performance in a real environment. For example, the model may be integrated into the plant DCS in order to analyze live streams of alarm data; this would allow evaluating the model effectiveness and highlighting its possible limitations.

6. Conclusions

This work proposes a data-driven method to extract knowledge from historical alarm data and perform predictions on the effectiveness of control room operators’ actions. A real industrial database has been used to support the analyses. A Wide&Deep classification model has been trained and evaluated on the database. The model aims at predicting whether or not the operator’s acknowledgment of an alarm will be followed by another alarm from the same Source within 30 minutes. In this way, the model would indirectly predict the effectiveness of the operator’s action and eventually drive his/her attention to alarms that are more likely to occur again in a short time. The issues related to the identification of rare unwanted events (such as those

considered in this work) have been discussed. Results show that even if performance may seem inadequate, a high Recall value may be obtained by lowering the decision threshold. After this simple adjustment, the model performance has improved considerably, and more than 90 % of the Target events have been correctly identified. Further investigations should be performed to evaluate the viability of the study in real-time applications. However, the approach suggests that Machine Learning may be used to extract relevant information from historical alarm data and use the acquired knowledge to support the control room operators proactively.

Acknowledgments

The authors would like to acknowledge Yara International for supplying the data and for the valuable support. Also, we wish to extend our thanks to Davide Santini, whose work has contributed significantly to the analyses described in this paper.

References

- Aleixandre, J., Alvarez, I., García, M., Lizama, V., 2015. Application of Multivariate Regression Methods to Predict Sensory Quality of Red Wines. *Czech Journal of Food Sciences* 33, 217–227. <https://doi.org/10.17221/370/2014-CJFS>
- ANSI/ISA, 2016. ANSI/ISA–18.2–2016 Management of Alarm Systems for the Process Industries. ANSI/ISA.
- Brink, H., Richards, J., Fetherolf, M., 2016. *Real-World Machine Learning*, first. ed. Manning Publications, Shelter Island.
- Carvalho, T.P., Soares, F.A.A.M.N., Vita, R., Francisco, R. da P., Basto, J.P., Alcalá, S.G.S., 2019. A systematic literature review of machine learning methods applied to predictive maintenance. *Computers and Industrial Engineering* 137, 106024. <https://doi.org/10.1016/j.cie.2019.106024>
- Cheng, H.-T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., Anderson, G., Corrado, G., Chai, W., Ispir, M., 2016. Wide & deep learning for recommender systems, in: *Proceedings of the 1st Workshop on Deep Learning for Recommender Systems*. pp. 7–10.
- EEMUA, 2013. EEMUA Publication 191 Alarm systems - a guide to design, management and procurement.
- Exida, 2009. *Saved by the Bell: Using Alarm Management to make Your Plant Safer*. Sellersville, PA.
- Goel, P., Datta, A., Mannan, M.S., 2017. Industrial alarm systems: Challenges and opportunities. *Journal of Loss Prevention in the Process Industries* 50, 23–36. <https://doi.org/10.1016/j.jlp.2017.09.001>
- Han, J., Kamber, M., Pei, J., 2012. *Data Mining: Concepts and Techniques*, *Data Mining: Concepts and Techniques*. Elsevier Inc. <https://doi.org/10.1016/C2009-0-61819-5>
- Health and Safety Executive, 1997. *The Explosion and Fires at the Texaco Refinery, Milford Haven, 24 July 1994: A Report of the Investigation by the Health and Safety Executive Into the Explosion and Fires on the Pembroke Cracking Company Plant at the Texaco Refinery, Milford Haven on 24 J, Incident Report Series*. HSE Books.
- Katzel, J., 2007. *Control Engineering | Managing Alarms [WWW Document]*. URL www.controleng.com/articles/managing-alarms (accessed 1.23.20).
- Kondaveeti, S.R., Izadi, I., Shah, S.L., Shook, D.S., Kadali, R., Chen, T., 2013. Quantification of alarm chatter based on run length distributions. *Chemical Engineering Research and Design* 91, 2550–2558. <https://doi.org/10.1016/j.cherd.2013.02.028>
- Paltrinieri, N., Comfort, L., Reniers, G., 2019. Learning about risk: Machine learning for risk assessment. *Safety Science* 118, 475–486. <https://doi.org/10.1016/j.ssci.2019.06.001>
- Reis, M.S., Kenett, R., 2018. Assessing the value of information of data-centric activities in the chemical processing industry 4.0. *AIChE Journal* 64, 3868–3881. <https://doi.org/10.1002/aic.16203>
- Stauffer, T., Clarke, P., 2016. Using alarms as a layer of protection. *Process Safety Progress* 35, 76–83. <https://doi.org/10.1002/prs.11739>
- Tamascelli, N., Arslan, T., Shah, S.L., Paltrinieri, N., Cozzani, V., 2020a. A Machine Learning Approach to Predict Chattering Alarms. *Chemical Engineering Transactions* 82. <https://doi.org/10.3303/CET2082032>
- Tamascelli, N., Paltrinieri, N., Cozzani, V., 2020b. Predicting Chattering Alarms: a Machine Learning Approach. *Computers & Chemical Engineering* 107122. <https://doi.org/https://doi.org/10.1016/j.compchemeng.2020.107122>
- Tian, Y., Fu, M., Wu, F., 2015. Steel plates fault diagnosis on the basis of support vector machines. *Neurocomputing* 151, 296–303. <https://doi.org/10.1016/j.neucom.2014.09.036>
- United States Nuclear Regulatory Commission, 2018. *Backgrounder on the Three Mile Island Accident*, United States Nuclear Regulatory Commission Library.
- Wall, K., 2009. Complexity of chemical products, plants, processes and control systems. *Chemical Engineering Research and Design* 87, 1430–1437. <https://doi.org/10.1016/j.cherd.2009.03.007>