Predicting Mining Industry Accidents with a Multi-Task Learning Approach

Rodolfo Palma, Luis Martí, Nayat Sánchez-Pi
Inria Research Center in Chile
Av. Apoquindo 2827, Las Condes, Santiago de Chile, Chile.
{rodolfo.palma,luis.marti,nayat.sanchez-pi}@inria.cl

Abstract

The mining sector is a very relevant part of the Chilean economy, representing more than 14% of the country's GDP and more than 50% of its exports. However, mining is also a high-risk activity where health, safety, and environmental aspects are fundamental concerns to take into account to render it viable in the longer term. The Chilean National Geology and Mining Service (Sernageomin, after its name in Spanish) is in charge of ensuring the safe operation of mines. On-site inspections are their main tool in order to detect issues, propose corrective measures, and track the compliance of those measures. Consequently, it is necessary to create inspection programs relying on a data-based decision-making strategy. This paper reports the work carried out in one of the most relevant dimensions of said strategy: predicting the mining worksites accident risk. That is, how likely it is a mining worksite to have accidents in the future. This risk is then used to create a priority ranking that is used to devise the inspection program. Estimating this risk at the government regulator level is particularly challenging as there is a very limited and biased data. Our main contribution is to apply a multi-task learning approach to train the risk prediction model in such a way that is able to overcome the constraints of the limited availability of data by fusing different sources. As part of this work, we also implemented a human-experience-based model that captures the procedures currently used by the current experts in charge of elaborating the inspection priority ranking. The mining worksites risk rankings built by model achieve a 121.2% NDCG performance improvement over the rankings based on the currently used experts' model and outperforms the non-multi-task learning alternatives.

Introduction

Mining is an essential activity for humankind since prehistoric times. As humanity has grown and developed the need for minerals has expanded. This has lead to an increase in exploration, extraction, and processing. However, modern and, in particular, extensive mining is also a high-impact activity. Mining has a long-established track of negative implications, ranging from the degradation of the natural environment, the production of highly toxic waste and its impact on communities, etc.

Consequently, health, safety, and environment (HSE) issues are priority matters for the mining industry. This industry is frequently in the news. Much of the time it is because of changes in prices of minerals. Other —less frequent but, perhaps, more important— subject of media attention is when disasters strike, as is the case of toxic spills, mining tailing overflows, and underground accidents. These incidents have a high impact on lives, the environment, and public opinion regarding this sector. That is why the correct handling of HSE is a determining factor in this industry's long-term success.

The Chilean mining industry is not an exception. Mining was responsible for 14.2% of Chile's GDP in 2012 and nearly 57% of exports were concentrated in this industry. The country is the largest producer of copper in the world, satisfying the 36% of the world's needs and having 28% of the world's reserves of that mineral (de Solminihac, Gonzales, and Cerda 2018). It is also the world's largest producer of lithium and iodine. However, this intensive and extensive growth has not come without negative ramifications.

Because of this, companies, governments, and communities have been working together to draw policies and methods to make the mining activities as viable and sustainable as possible. The Chilean National Geology and Mining Service (Sernageomin, after its name in Spanish) is the country's national authority regarding mining and geology.

One of its primary objectives is the reduction of accidents and incidents in the mining facilities, as it has a direct impact on the quality of life of workers and the environment. To achieve this, it essential to have near-optimal programming of the inspections of a mining worksite. It can be hypothesized that by optimizing the inspection visits to the facilities some early signs of problems in the riskier areas could be detected and reported and, hopefully, corrected. This programming is particularly complex because of the extreme oblongness of the Chilean geography (as can be perceived in Figure 1), the frequent natural hazards like earthquakes, tsunamis, and flash floods, and the difficulty of accessing sites that are located in the aridest desert of the planet.

The current method for doing the programming of the visits is currently defined as a two-step process. First, for each mining facility, it must be determined a perceived risk based on a given set of human experience and intuition factors and, second, given that perceived risk organize a ranking so that

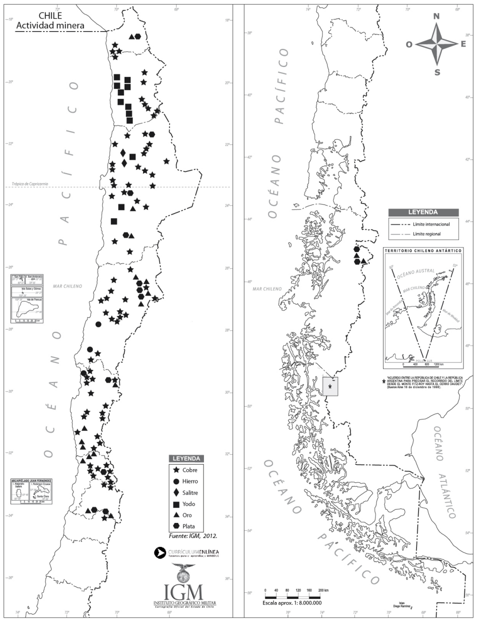

Figure 1: Map of the distribution of mining activities along the Chilean geography. Source: Sernageomin, used with permission.

the worksite with higher risk are visited first.

This method can be improved by enhancing i) how that risk is estimated and, then, ii) how the worksite rank is elaborated. As the risk is currently estimated as a result of a semi-intuitive experience-based process, it seemed like the best way to yield a rapid and substantial improvement without disrupting the procedures and formalisms in a highly-regulated sector.

Today, with the advances of new technologies, accidents, incidents, and occupational health records are stored in heterogeneous repositories. Similarly, the amount of information about HSE that is daily generated is increasing but it is generally stored as unstructured or poorly-structured data and. This poses a challenge that is a top priority for industries that are looking for ways to search, sort, analyze and extract knowledge from masses of data

One important challenge when tackling this task at the level of a regulatory authority, is the lack of detailed data, as the service only receives reports with highly-aggregated data of accidents when severe accidents happen and —as it can be expected— these are infrequent events. However, it can be hypothesized that there are a number of features that can be extracted from previous inspection reports that can act as early indicators and be used to predict if and/or when accidents happen in the year ahead.

In order to overcome the limited-sized and heavily-biased dataset, we pose the problem of determining the risk by posing it as a multitask learning problem (Thrun 1995; Caruana 1997; Crawshaw 2020).

This work reports the progress achieved addressing the problem of transforming the operation of Sernageomin generating a data-based decision-making tool. We succeed at providing a principled method for estimating facility risk that relies on existing data and that outperforms the procedure that is being currently used.

The rest of this paper is organized as follows. In the next section, we present the necessary theoretical foundations and related works that are necessary to understand our proposal. After that, in Section we describe the approach that is currently in place and that has been formalized in an algorithmic way for the first time as part of this work. Subsequently, in Section we introduce the multitask learning model that we propose. Then, in Section we apply and compare different variants of our model as well as the current experience-based approach. Finally, in Section we provide some final remarks and outline our future work.

Related Work

Accident prevention is a priority for any modern industry. It is self-evident that eliminating them or mitigating their impact is a key element of success from the human, environmental, and operational points of view. Using evidence and/or data from previous cases to understand the nature and causes of accidents have been in the foundations of fields like statistics and actuarial science, to just cite two.

One of the characteristics of this class of problems is that each instance has many particular characteristics that call for a very customized approach. Consequently, there have been a myriad of papers proposing solutions that range from ontologies (Sanchez-Pi, Martí, and Bicharra Garcia 2016), feature selection (Lin, Wang, and Sadek 2015), convolutional neural networks (Wenqi, Dongyu, and Menghua 2017), representation learning (Chen et al. 2016), data fusion and time-series prediction (Martí et al. 2014; Sanchez-Pi et al. 2014; Moosavi et al. 2019), just to mention a few.

However, what best characterizes the effort when dealing with these problems is the need to carry out intense work on data preparation and problem understanding. Perhaps this is best understood with two prototypical examples. Rudin et al. (2010) tackle the problem of predicting manhole events in Manhattan. In order to deal with such a problem, it is necessary to process text reports, fuse that information with different sources, and then propose a predictive model. Similarly, Moosavi et al. (2019) deals with the traffic accident prediction problem by consuming multi-source and heterogeneous data like weather, traffic patterns, points of interest, etc. to produce a predictor.

In the case of the problem we are dealing with here, there is a particular machine learning scheme denominated multi-task learning that is particularly useful, as it will be demonstrated on subsequent sections. While in machine learning the focus is to optimize the parameters of a given model for a particular metric, whether this is an error score on a certain benchmark or other performance indicators. In order to achieve this, the optimization algorithm is executed in a process generally denominated as training until the performance of the model no longer improves.

While it is possible to achieve an acceptable performance using this approach, it has been shown that doing this it is possible that it is ignored information that might help the model even better on the metric of interest. Specifically, this information comes from the training signals of similar tasks. By sharing representations between related tasks, it enables the model to generalize better on the main task. This approach is called multitask learning.

There are at least four approaches for implementing a multitask learning scenario:

1. task grouping and overlap (Hajiramezanali et al. 2018), where tasks are grouped or exist in a hierarchy, or be related according to some general metric,
2. exploiting *a priori* unrelated tasks (Paredes et al. 2012), where joint learning of unrelated tasks which use the same input data is deemed beneficial as it can lead to sparser and more informative representations for each task grouping, essentially by screening out biases present in the data,
3. knowledge transfer (Yosinski et al. 2014), where a pre-trained model can be used as a feature extractor to perform pre-processing for another learning algorithm, and
4. group online adaptive learning (GOAL) (Zweig and Chechik 2017) where sharing information is deemed particularly useful when models operate in continuously changing environments because a model could benefit from previous experience of another and quickly adapt to a new environment.

Experience-Based Model

One of the main challenges of this wprk was the absence of key performance indicators and baseline algorithms that would allow the evaluation of the risk rankings. Therefore, it was not straightforward to assess the performance of the predictive models that will be implemented.

Additionally, in Sernageomin there was no standardization regarding the mechanism that each regional office must use to determine the worksites to be inspected. In practice, each one of the ten different regional offices has the freedom to develop risk calculation methods for mining worksites, which are then used to schedule inspections. The foundations, formulation, and implementation of these methods are diverse: risk matrices that assign each mining worksite a scalar, matrices with worksites features that allow them to be ordered, among others. What all risk models have in common is that they are designed based on the experience of the inspections coordinators of the regional offices.

Taking the above premises under consideration, one of the first activities carried out with Sernageomin was to formulate a model —not necessarily predictive— that allowed calculating the risk of mining worksites. The construction of this *experience–based model* was done in close collaboration between the authors and the experts in Sernageomin.

This model would help to establish a baseline against which to compare the performance of the future predictive models, to become familiarized with the terms and semantics of the mining industry, which would later prove useful in the construction of a multitask neural network.

An interesting outcome of the construction of the experience-based model was discovering that building a ranking ordered by the risk of mining worksites was more important than the individual risk score itself. This makes sense if we consider the main use case of this tool: to efficiently focus the limited inspections resources of each regional office.

Algorithm 1 Experience-based model for worksite risk ranking.

1: Sort worksites by descending order, using the number of facilities stopped by a Sernageomin issued sanction.
2: Untie worksites sorting by ascending order, using the year of the last inspection.
3: Untie worksites sorting by ascending order, using the number of pending corrective measures to be fulfilled.
4: Untie worksites sorting by descending order, using the sum of number of accidents with lost time plus the number of fatal accidents in the last 24 months.
5: Untie worksites sorting by descending order, using the number of days since the last inspection.

Due to the above, the experience-based model was built using a set of features considered relevant by experts to determine the risk of mining worksites. Then, the features were ranked in terms of how relevant they are to the calculation of the risk. Finally, depending on the feature hierarchy, the worksites are sequentially sorted.

The formulation of the model based on experience is observed in the Algorithm 1. It allows sorting the mining worksites of the different regional offices of the Service to better focus the inspections resources. Although the formulation of the model is not characterized for being one with solid statistical foundations, it formalizes a working modality that generally reflects the operation of the Service before implementing digital transformation strategies. Also, and as mentioned above, it will allow establishing a baseline to evaluate the machine learning model experiments.

Finally, and from a broader project management perspective, this model was successfully implemented in Sernageomin through a web application that is currently being used nationwide shown in Figure 2. This way, the risk of adoption always present in this type of project was mitigated.

Multitask Model

A widely studied family of problems that can be adapted to solve the worksite risk one is those of predictive maintenance (Susto et al. 2015).

Some of these models try to predict the remaining useful life (RUL) of machines to prevent halts in industrial processes (Okoh et al. 2014). On the other hand, other models learn the probability distribution of failure of some element or piece within a time window in the future. Applying these notions to our problem, a practitioner could draw the analogy that "machines" or "pieces" are the mining worksites and that "failures" are the accidents. Then, a risk ranking can be constructed by sorting the mining worksites in ascending order according to their RUL or descending order according to their probability of failure in the future.

After experimenting with both tasks and different models,

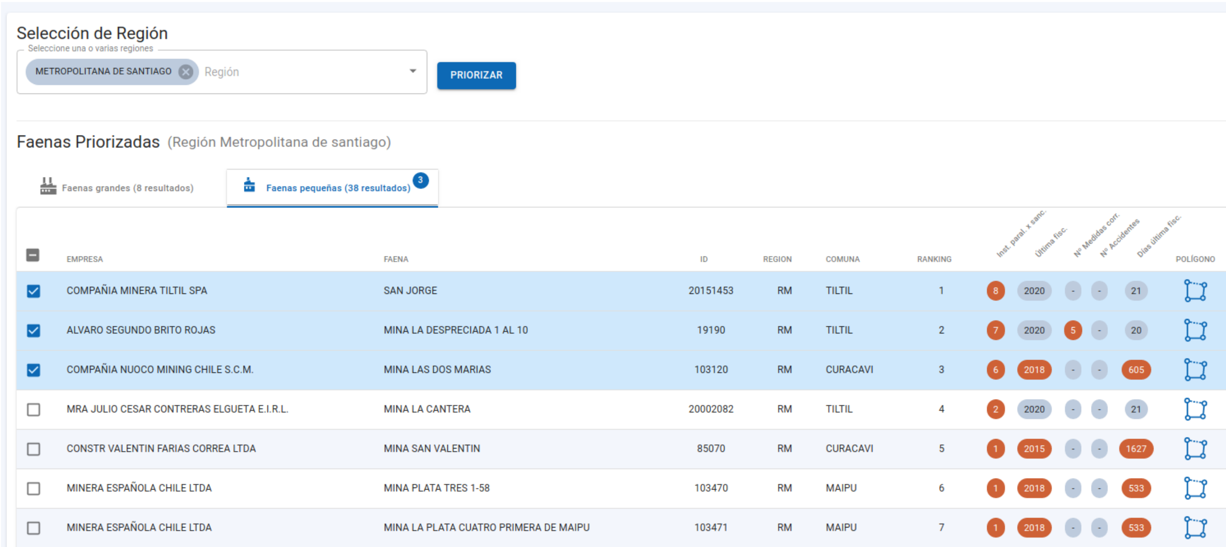

Figure 2: Implementation of experience-based model. The application serves a ranking of worksites with visual cues about the datapoints that explain the worksite position.

such as decision trees and multi-layer perceptrons, we propose a multitasking neural network to solve our risk problem.

The neural network input is a feature vector $\boldsymbol{x} \in \mathbb{R}^n$ from the tabular dataset representing a mining worksite in a particular year–month. This input vector $\boldsymbol{x}$ passes through a shared hidden layer activated by a non-linearity implemented with a hyperbolic tangent function. Then, the computation graph is branched off into the two tasks. The first one, a regressor, learns to predict the approximate amount of days until the next accident at the given mining worksite, similar to an RUL–predictive maintenance model. The second one, a classifier, learns to predict the approximate probability that the given mining worksite suffers accidents in the future, analog to the predictive maintenance models that predict probability distributions.

A set of mining worksite could be a risk–ranked by ordering it in descending order according to the probability that they will have accidents in the future. Even though we don't explicitly use the regressor output to build the risk ranking, we supervise its outputs during training seeing an increase in the overall ranking quality, as we'll see in the experiments section.

Unlike the experience-based model, the multitasking model requires to be properly trained before it can be used. We define the loss function as the loss of a multitasking model parameterized by Θ on a dataset D as

$$\mathcal{L}(D, \Theta) = \lambda \mathcal{L}_{\text{clf}}(D, \Theta) + (1 - \lambda)\mathcal{L}_{\text{reg}}(D, \Theta)\,. \quad (1)$$

The classifier output, which is the probability that the worksite will have accidents in the future, is supervised by a Boolean label. The term $\mathcal{L}_{\text{clf}}(D, \Theta)$ is implemented using a binary cross entropy between the predicted value and the true value column.

On the other hand, the regressor output is supervised by the true value of days until the next accident. The term of the loss function $\mathcal{L}_{\text{reg}}(D, \Theta)$ is implemented by the mean squared error between the predicted value and the supervised value. We use a hyperparameter λ to linearly combine both terms and balance the importance of each task in this multitasking setting.

Experiments and Results

The data stored in Sernageomin's databases were processed to build a tabular dataset. Essentially, for each year-month existing since the system came into operation, the state of each mining worksite was recorded. The state of each mining site was structured according to the features described in Table 1.

After performing this pre-processing, a table with 191,518 rows was built. The obtained records span a time range from January 2014 until June 2019. All records from January 2014 to November 2017, equivalent to 134,062 rows, were considered part of the training dataset. On the other hand, the records spanned between November 2017 and May 2018, corresponding to 19,151 rows, were assigned to the validation dataset. Finally, the rows from May 2018 to June 2019, equivalent to 38,305 records, were part of the testing dataset. Therefore, the training, validation, and testing sets were approximately 70%, 10%, and 20% of the complete dataset, respectively. Each split feature was individually scaled to a range of $(0, 1)$ using a min–max transformation.

During training, when mining worksites have no accidents in the future, we fill the missing values with $1,460$, equivalent to the number of days in 4 years. Finally, the `MONTH` feature is categorical and represented with a one–hot transformation.

Evaluation Framework

The purpose of applying mining risk models is to build worksite rankings that allow a better usage of the limited

Feature name	Feature description
NUM_FACILITIES	The number of mining facilities operating in the worksite.
DAYS_SINCE_LAST_INSPECTION	Number of days since the worksite's last inspection.
HAS_NEVER_BEEN_INSPECTED	True if the worksite has been inspected in the past, otherwise it is false.
STOPPED_BY_SANCTION	Number of worksite's facilities stopped because of sanction.
PENDING_ACTIONS	Number of worksite's corrective measures whose execution is still pending.
NO_TIME_LOST_COUNT	Worksite's number of accidents with no worker time lost in the last 24 months.
TIME_LOST_COUNT	Worksite's number of accidents with worker time lost in the last 24 months.
FATAL_COUNT	Worksite's number of fatal accidents in the last 24 months.
HOURS_WORKED	Worksite's work hours performed in the last 24 months.
ACCIDENTS_RATE	Average number of accidents with lost time per million hours worked.
MONTH	Worksite recorded month.
TOTAL_ACCIDENTS_COUNT	Sum of NO_TIME_LOST_COUNT, TIME_LOST_COUNT, and FATAL_COUNT.
FATAL_TIME_LOST_COUNT	Sum of TIME_LOST_COUNT and FATAL_COUNT.

Table 1: Dataset features used as input for the multitask learning model along with their descriptions.

organization's resources. Therefore, one risk model is better than another if it builds better risk rankings.

We define a ranking as an ordered set of mining worksites. Also, we define the function rel_i that maps objects from the set of mining worksites to the set of real numbers. In particular, the value of rel_i indicates how relevant a mining worksite is in the context of a risk ranking. For this particular task, the relevance of a mining worksite is defined as the number of accidents it'll have in the following twelve months.

A commonly used metric for measuring ranking quality is the discounted cumulative gain (DCG) (Wang et al. 2013). The DCG of a ranking is calculated as

$$\mathrm{DCG}_p = \sum_{i=1}^{p} \frac{\mathrm{rel}_i}{\log_2(i+1)}\,. \quad (2)$$

In this way, when the most relevant elements are in the first ranking positions, high DCG values will be obtained. On the other hand, when the most relevant elements are far from the top positions, the value of the DCG is worse.

We can define the IDCG_p as the DCG_p of the ideal ranking. The ideal risk ranking is the ordered set of worksites sorted by their relevance in descending order.

Finally, we introduce the normalized DCG (NDCG) (Wang et al. 2013) which is calculated as

$$\mathrm{NDCG}_p = \frac{\mathrm{DCG}_p}{\mathrm{IDCG}_p}\,. \quad (3)$$

The NDCG value is 0 when the built ranking does not contain relevant elements and 1 in case that the built ranking quality is equivalent to the ideal ranking one. In this work, all the mining worksites of the different regional offices are considered in the metric calculation, so the parameter p is equal to the cardinality of the mining set.

The procedure for evaluating the models in a dataset consists of iterating through each of the regional offices and sequentially prioritizing the different year–months for the corresponding subset of mining work sites. This way, the evaluation framework simulates the work of the mining inspections coordinator of a regional office, who selects the critical worksites that must be inspected each month in a given regional office. Finally, for each regional office, we compute the mean NDCG of every year–month group.

Experience-Based Model Experiments

The experience-based model, depicted in algorithm 1, sequentially orders a set of mining worksites according to previously selected features using the organization's experts' knowledge and experience. This model doesn't need to be trained to prioritize mining worksites. Therefore, it is possible to directly evaluate its performance on the testing split of the tabular dataset. The results of the experience-based model are displayed in Table 2.

Baselines

We train and evaluate two classes of baseline models to properly assess the proposed multitask model. On the one hand, we train two separate tree models implemented in the scikit-learn library (Buitinck et al. 2013) on each task: regression and classification. On the other hand, we train two single–task neural networks for each task. Essentially, these neural models are architecturally identical to the multitask model except that we only consider one task-specific head after the shared hidden layer. To build a risk ranking using a regressor, we sort the mining worksites set by descending order using the model's output.

Results of both the decision trees and neural networks baselines are shown in Table 2. We note that with simple models, such as a decision tree regressor and a single–task neural network classifier, we're able to increase the quality of the rankings by an average of 93.8% and 114.4%, respectively.

Multitask Model Experiments

The multitask model shared hidden-layer size was 50 neurons. The classifier head was implemented as a fully connected layer of size 50 activated by a $\tanh$ function followed by a layer of dimension two activated by a $\mathrm{softmax}$ function. Furthermore, the regression head was implemented as another fully connected layer of size 50 activated by a $\tanh$

Regional office	Experience–based model	Decision tree regressor	Decision tree classifier	NN regressor	NN classifier	Multitask model
Arica	0.23	0.96	0.66	0.55	1.00	1.00
Tarapacá	0.32	0.58	0.53	0.36	0.62	0.66
Antofagasta	0.33	0.64	0.48	0.75	0.76	0.77
Atacama	0.27	0.66	0.49	0.50	0.85	0.95
Coquimbo	0.21	0.83	0.41	0.46	0.87	0.97
Centro	0.20	0.68	0.39	0.62	0.69	0.70
O'Higgins	0.19	0.86	0.50	0.98	1.00	1.00
Maule	0.77	0.50	0.63	0.41	0.54	0.57
Sur	0.21	0.41	0.39	0.29	0.60	0.66
Magallanes	0.87	0.85	0.70	0.64	0.79	0.68
Mean NDCG	**0.36**	**0.70**	**0.52**	**0.56**	**0.77**	**0.80**
% improvement		**93.8%**	**43.7%**	**54.6%**	**114.4%**	**121.2%**

Table 2: Mean NDCG score of different models on the dataset's test split. Higher values are better.

function followed by a single-dimensional layer with no activation function.

The network was trained using a stochastic gradient descent algorithm with a batch size equal to 32 over 200 epochs. The best results were obtained with an initial learning rate equal to 0.001 and a hyperparameter $\lambda = 0.99$. At the end of each training epoch, the model was evaluated on the validation dataset. Finally, the model that minimized the loss on the validation set was selected for the final evaluation of the testing split.

The results obtained by the multitasking model are observed in Table 2. Our multitask model outperforms every baseline average performance and achieves an average performance increase of 121.2% with respect to the model based on experience.

We note that there are two regional officeswhere all the machine learning models fail to outperform the experience-based model. An in-depth analysis of the particular characteristics of these regions causing this behavior is necessary to deploy these models in operation.

Conclusions

In this paper, we reported our results when dealing with the prediction of mining accident risks in the Chilean context. The main contribution of this paper is a neural network trained in a multitask learning setup that predicts mining worksite risk. This neural network should eventually replace the current expert's model, to support the agency's decision making processes with the data available on its databases.

The mining worksites risk rankings built by our multitask-trained neural network's scores achieve a 121.2% NDCG performance improvement over the rankings based on the expert's model. This is an important result that should eventually lead to a reduction in accidents.

However, these results are by no means final. On the one hand, we are interacting with mining companies, government agencies, and communities to try to enhance the amount of data available and, therefore, to be able to address more ambitious goals involving casual inference of the causes of accidents, or more accurate predictions. We also plan to incorporate exogenous variables related to weather, market price, production rates, etc. that could indicate possible modifications of the regular operation regimes of the mining worksites.

On the other hand, in the future there should be efforts to add explainable attributes to our proposed model. The presence of these attributes would positively impact the organization's change management tactics and were no t the main focus of this work.

Another important aspect that remains unaddressed is how to generate the inspection schedule. This problem can be posed as a Traveling Salesperson Problem with priorities, where these priorities are the estimated risk or probability of an accident. This implies that, potentially, there would not a unique optimal schedule but a set of trade-off solutions, and therefore, becoming a multi-objective optimization problem.

This neural network can be progressively ensembled with or eventually replace the current experience-based model, to support the agency's decision making processes leveraging data they already have. Assimilating a novel technology takes an arduous path in a conservative industry like mining, even more, if it has to do with accident prevention. The novel models being proposed are currently under assessment and it is expected that they start to be used after being accepted by the different stakeholders. So far, the outcome of this work has been welcomed by Sernageomin, with public high praises from the service's direction (Minería Chilena 2020).

The datasets and source code of our experiments is available online at https://github.com/Inria-Chile/mining-risk-multitasking-model.

Acknowledgments

This work was funded by project CORFO 10CEII-9157 Inria Chile and funded in part by Sernageomin under the direction of A. Domeyko and lead by Data Intelligence Office headed by H. Cano and involving G. Jeldres and B. Alvarado.

References

Buitinck, L.; Louppe, G.; Blondel, M.; Pedregosa, F.; Mueller, A.; Grisel, O.; Niculae, V.; Prettenhofer, P.; Gramfort, A.; Grobler, J.; Layton, R.; VanderPlas, J.; Joly, A.; Holt, B.; and Varoquaux, G. 2013. API design for machine learning software: Experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, 108–122.

Caruana, R. 1997. Multitask Learning. *Machine Learning* 28(1): 41–75. doi:10.1023/A:1007379606734. URL https://doi.org/10.1023/A:1007379606734.

Chen, Q.; Song, X.; Yamada, H.; and Shibasaki, R. 2016. Learning Deep Representation from Big and Heterogeneous Data for Traffic Accident Inference. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI'16, 338–344. AAAI Press.

Crawshaw, M. 2020. Multi-Task Learning with Deep Neural Networks: A Survey URL http://arxiv.org/abs/2009.09796.

de Solminihac, H.; Gonzales, L. E.; and Cerda, R. 2018. Copper mining productivity: Lessons from Chile. *Journal of Policy Modeling* 40(1): 182–193.

Hajiramezanali, E.; Dadaneh, S. Z.; Karbalayghareh, A.; Zhou, M.; and Qian, X. 2018. Bayesian multi-domain learning for cancer subtype discovery from next-generation sequencing count data. In *Advances in Neural Information Processing Systems*, 9115–9124.

Lin, L.; Wang, Q.; and Sadek, A. W. 2015. A novel variable selection method based on frequent pattern tree for real-time traffic accident risk prediction. *Transportation Research Part C: Emerging Technologies* 55: 444–459. ISSN 0968-090X. doi:https://doi.org/10.1016/j.trc.2015.03.015. Engineering and Applied Sciences Optimization (OPT-i) - Professor Matthew G. Karlaftis Memorial Issue.

Martí, L.; Sanchez-Pi, N.; Molina, J. M.; and Bicharra Garcia, A. C. 2014. High-Level Information Fusion for Risk and Accidents Prevention in Pervasive Oil Industry Environments. In Corchado, J. M.; Bajo, J.; Kozlak, J.; Pawlewski, P.; Molina, J. M.; Gaudou, B.; Julian, V.; Unland, R.; Lopes, F.; Hallenborg, K.; and García Teodoro, P., eds., *Highlights of Practical Applications of Heterogeneous Multi-Agent Systems — The PAAMS Collection*, volume 430 of *Communications in Computer and Information Science*, 202–213. Berlin/Heidelberg: Springer International Publishing. ISBN 978-3-319-07766-6. doi:10.1007/978-3-319-07767-3_19.

Minería Chilena. 2020. Inria y Sernageomin colaboran para lograr cero accidentes en minería. *Minería Chilena* URL https://www.mch.cl/2020/02/10/inria-y-sernageomin-colaboran-para-lograr-cero-accidentes-en-mineria/. In Spanish.

Moosavi, S.; Samavatian, M. H.; Parthasarathy, S.; Teodorescu, R.; and Ramnath, R. 2019. Accident risk prediction based on heterogeneous sparse data: New dataset and insights. *GIS: Proceedings of the ACM International Symposium on Advances in Geographic Information Systems* 33–42. doi:10.1145/3347146.3359078.

Okoh, C.; Roy, R.; Mehnen, J.; and Redding, L. 2014. Overview of Remaining Useful Life Prediction Techniques in Through-life Engineering Services. *Procedia CIRP* 16: 158 – 163. ISSN 2212-8271. doi:https://doi.org/10.1016/j.procir.2014.02.006. URL http://www.sciencedirect.com/science/article/pii/S2212827114001140. Product Services Systems and Value Creation. Proceedings of the 6th CIRP Conference on Industrial Product-Service Systems.

Paredes, B. R.; Argyriou, A.; Berthouze, N.; and Pontil, M. 2012. Exploiting Unrelated Tasks in Multi-Task Learning. In Lawrence, N. D.; and Girolami, M., eds., *Proceedings of the Fifteenth International Conference on Artificial Intelligence and Statistics*, volume 22 of *Proceedings of Machine Learning Research*, 951–959. La Palma, Canary Islands: PMLR. URL http://proceedings.mlr.press/v22/romera12.html.

Rudin, C.; Passonneau, R. J.; Radeva, A.; Dutta, H.; Ierome, S.; and Isaac, D. 2010. A process for predicting manhole events in Manhattan. *Machine Learning* 80(1): 1–31. ISSN 08856125. doi:10.1007/s10994-009-5166-y.

Sanchez-Pi, N.; Martí, L.; and Bicharra Garcia, A. C. 2016. Improving ontology-based text classification: An occupational health and security application. *Journal of Applied Logic* 17: 48–58. ISSN 1570-8683. doi:10.1016/j.jal.2015.09.008. URL http://www.sciencedirect.com/science/article/pii/S1570868315000774.

Sanchez-Pi, N.; Martí, L.; Molina, J. M.; and Bicharra Garcia, A. C. 2014. An information fusion framework for context-based accidents prevention. In *17th International Conference on Information Fusion (FUSION)*, 1–8.

Susto, G. A.; Schirru, A.; Pampuri, S.; McLoone, S.; and Beghi, A. 2015. Machine learning for predictive maintenance: A multiple classifier approach. *IEEE Transactions on Industrial Informatics* 11(3): 812–820. ISSN 15513203. doi:10.1109/TII.2014.2349359.

Thrun, S. 1995. Is Learning the N-Th Thing Any Easier than Learning the First? In *Proceedings of the 8th International Conference on Neural Information Processing Systems*, NIPS'95, 640—-646. Cambridge, MA, USA: MIT Press.

Wang, Y.; Wang, L.; Li, Y.; He, D.; Liu, T.; and Chen, W. 2013. A Theoretical Analysis of NDCG Type Ranking Measures. *CoRR* abs/1304.6480. URL http://arxiv.org/abs/1304.6480.

Wenqi, L.; Dongyu, L.; and Menghua, Y. 2017. A model of traffic accident prediction based on convolutional neural network. In *2017 2nd IEEE International Conference on Intelligent Transportation Engineering (ICITE)*, 198–202.

Yosinski, J.; Clune, J.; Bengio, Y.; and Lipson, H. 2014. How transferable are features in deep neural networks? In *Advances in neural information processing systems*, 3320–3328.

Zweig, A.; and Chechik, G. 2017. Group online adaptive learning. *Machine Learning* 106(9-10): 1747–1770. ISSN 15730565. doi:10.1007/s10994-017-5661-5.