|  |  |
| --- | --- |
| **Heilmeier questions** | **Section** |
| 1. What are you trying to do? Articulate your objectives using absolutely no jargon. | 1 |
| 1. How is it done today; what are the limits of current practice? | 2,3 |
| 1. What's new in your approach? Why will it be successful? | 4 |
| 1. Who cares? | 5 |
| 1. If you're successful, what difference and impact will it make, and how do you measure them (e.g., via user studies, experiments, ground truth data, etc.)? | 5, 6 |
| 1. What are the risks and payoffs? | 7 (payoff 5) |
| 1. How much will it cost? | 8 |
| 1. How long will it take? | 8 |
| 1. What are the midterm and final "exams" to check for success? How will progress be measured. | 9 |

Word count ~1350 excluding final table and headings

**Team 17 – Project Proposal (DRAFT V1)**

**[1] Objectives**

The Project is focussed on the application of machine learning to model the condition of rail ballast (the track bed on which sleepers lie) with the aim of better informing maintenance decisions. The project is being undertaken for Queensland Rail, a major Australian public rail network.

The project objectives include:

* 1. Determine whether there is an ability to predict maintenance required on rail ballast using a variety of factors such as track geometry,
  2. Provide a prediction for maintenance required in the near future (12 to 24 months), and
  3. Provide interactive functionality to enable users to select a geographic section of rail and drill-down to understand the localised status of ballast.

**[2] Current methods of measuring rail condition**

Is this where the survey is included?

Much of the research into the application of machine learning for the rail industry has related to rollingstock (such as locomotives) and rail. Limited published work is available on predicting maintenance activity required on ballast.

Nakhaee et al’s survey [1], states that deep learning algorithms have been the most extensively used technique for the detection of structural defects – particularly with regards rail (as opposed to ballast). This survey identifies several impediments to the application of machine learning for rail maintenance. Some of which include:

1. Imbalance in defective and non-defective observations: as the vast majority of rail is non-defective(often more than 99%), a simple classifier will be biased towards selection of the non-defective class.
2. Availability of labelled datasets: it is highly time consuming for skilled staff to label thousands of kilometres of rail as defective or non-defective.
3. Explainability of machine learning models: a significant number of papers published regarding rail maintenance utilise convolutional neural networks which are “black box” solutions difficult to explain to management.

Hajizadeh et al [2] provide proposals to address the first two issues identified. Their paper proposes the use of techniques such as minority over-sampling with noise to balance the defective/non-defective datasets. They also propose the use of semi-supervised techniques to address the lack of labelled data.

Focussing on track geometry, Sharma [3] utilises Markov Decision Processes to propose a condition monitoring maintenance actions of: none; minor; and major. A major difference between Sharma’s work and that proposed by the Project Team is that Sharma considered only major defects that violate regulatory rules. The Project seeks to identify degradation in ballast via track geometry prior to a potential breach of rail standards.

[Other potential literature surveys: research: see list wrt Peng at end; industry e.g. cost of rail maintenance; data collection/cleansing?; dimensionality reduction techniques e.g. PCA?; ML techniques e.g. logistic regression; linear regression wrt maintenance; visualisation e.g. best-practice drill-down of broad geographic into local sections]

**[3] Current Practice of Queensland Rail**

Queensland Rail collect a significant quantity of data relating to the condition of their rail network, focusing on ballast condition. Two primary sources are used to formulate ballast maintenance priorities:

1. Track Recording Car (TRC): the TRC data is captured on a quarterly basis. This data provides information regarding the condition of the rail geometry such as height of left and right rail and the twist in the line. A primary driver for deterioration in track geometry is the degradation of the underlying ballast. Elsewhere, fractal analysis on vertical TRC data has shown a moderate correlation with ballast fouling, indicating that TRC data is an indirect measure for ballast condition [8].
2. Ground Penetrating Radar (GPR): GPR data is collected for the entire network every 3 years. Previously demonstrated to yield up to 100% classification accuracy using an SVM classifier to detect soiling, GPR data is a robust measure of ballast condition [7]. The data informs engineers as to the level of contamination in the ballast, ballast thickness and moisture likelihood.

Queensland Rail’s network extends more than 6,600 kilometres across the state. As such, the quantity of data collected is significant with more than 10 million data fields collected by the TRC split across 16 features and ~350,000 data fields split across 22 features collated during the GPR survey.

Current practice involves engineers conducting a manual visual inspection of the TRC readings, comparing these with prior TRC runs to identify potential degradation and then overlaying the GPR data for the same section of track to provide confirmation of potential ballast contamination (Figure 1).

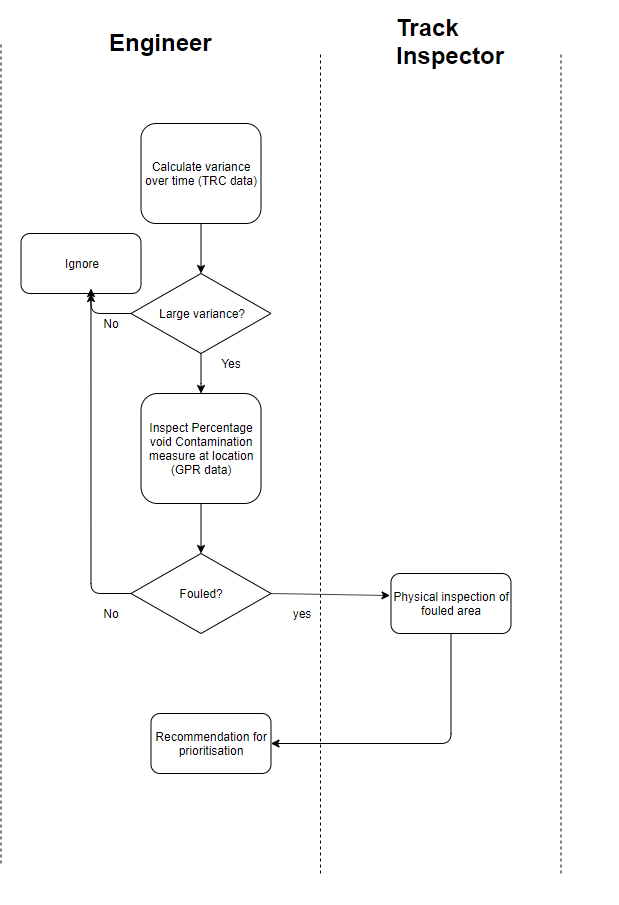


Figure 1Flowchart of QR's maintenance decision process

The process is highly labour intensive. QR’s engineers trialled a “heatmap” within Excel whereby the process was semi-automated but lacked visualisation of the geographical context of the track.Additionally, a 10km section of track involved approximately 2,000 rows of data and over 20,000 standard deviation calculations. This rendered the spreadsheet unstable. As a result, the engineers did not make plans to roll-out the “heatmap” process.

**[4] Novelty in Approach**

The approach proposed includes several extensions intended to enhance QR’s current methods:

1. Augmenting current processes by automating the production of “heatmaps”: the Project intends to read in TRC and GPR data and collate these for comparable sections of track, automate the standard deviation calculations to identify degradation in geometry and then colour code the track state to ease visual inspection. The data would be written to a file without calculations ensuring the file was stable and fast to load.
2. Providing further insights via combining additional factors: the Project Team has requested QR provide additional attributes that may impact ballast degradation such as the location of culverts and embankments (which impact drainage); train load (e.g. in tonnage); and maintenance history.
3. Further informing decision-making through machine learning: Using ML algorithms to predict track degradation in time (e.g. in 12 to 24 months’ time) and/or deterioration in space (e.g. further along the line adjacent to a current defect). This could potentially reduce the first two decision points into a single step (Figure 1), streamlining the decision-making process.

Lasisi and Attoh-Okine [9] have shown that a principal components analysis (PCA) of TRC data used in combination with SVM outperforms industry-standard weighting of TRC output in predicting track condition. Incorporating PCA into our analysis of the TRC data may therefore yield a more accurate assessment of track condition in the QR dataset than the current practice.

1. Re-shaping user experience with an interactive drill-down capability: ideally a network map for the entire state would be developed and users would have the ability to drill-down to assess the current status of a localised section of rail.

**[5] Potential Benefits**

Queensland Rail’s annual maintenance and capital works expenditure on ballast and resurfacing work alone is ~$50 million p.a. From interviews the team has conducted, executives anticipate ~10% of this expenditure could be saved annually if maintenance activities could be predicted with more reliability. Additionally, ~$30 million in capital expenditure could be avoided.

On a global scale, Peng, Kang et al [4], identified that in a rail context, preventive maintenance costs are excessive when undertaken too early or too late. From a safety perspective, Peng [5] identified that of 1,890 train accidents in 2009, 658 were caused by track defects.

As such, the global benefits offered by successful innovation in track maintenance include:

1. Potential to harvest financial returns in excess of $1 billion p.a, and
2. Materially improved safety via reduced accidents.

**[6] Assessing Success**

The following are planned to objectively measure the success of the Project:

1. Train/test split with K-fold cross-validation: the machine learning algorithms will be trained on data (e.g. relating to periods 2-3 years prior) and will be tested on unseen test-data (e.g. relating to recent periods). This process will inform the Project Team as to the predictive power (if any) of the model.
2. End-user satisfaction: the usability of the interactive visualisation will be assessed via a brief survey of at least 2 end-users.

**[7] Project Risks**

The primary risks relate to the novelty of the Project. In particular:

1. Data availability: will the client provide the data (including additional parameters) in a sufficiently manner to adhere to the Project schedule?
2. Subject matter expertise: will the client provide the necessary resources to inform the Project team on areas requiring subject matter expertise?
3. Predictive power of a model: as the application of machine learning to the client’s environment is novel, it is uncertain whether a model can be developed that has any predictive capability.

While the Project Team believes each and all of the above risks have the ability to jeopardise the project, the risks are weighed against the following mitigators:

1. The risks have been discussed with QR and support has been offered to the Project Team,
2. The Project is novel and as such of interest to the Project Team (versus a “safer” application using an established dataset), and
3. The benefits (identified in Section 5) of successful innovation in the industry are substantial.

**[8] Project Cost and Timeframe**

No financial outlay has been provisioned:

* Queensland Rail has offered to provide data and subject matter expertise at no cost, and
* No external consulting or contract services (such as video production) have been planned.

The Project is anticipated to be completed within 3 months. Team members commenced work on the Project on 1 September 2019 and the Project will be completed in accordance with the schedule required by CSE6242 i.e. 29th November 2019.

**[9] Workplan, Milestones and Team-member Contribution**

The Project Workplan is outlined below:

|  |  |  |
| --- | --- | --- |
| **Week Ending** | **Core Activity** | **Target Milestone** |
| 6th September | * Meet with QR Executive general Manager to discuss potential scope | * Identify potential project opportunities |
| 13th September | * Meet with QR Senior Manager Track and Civil Infrastructure to refine scope | * Refine scope * Arrange meetings with line management |
| 20th September | * Meet with Senior Civil Engineer to review current process | * Obtain initial datasets |
| 27th September | * Attempt to automate a sample of “heatmap” * Present status to Executive General Manager * Identify opportunities for machine learning * Identify additional parameters required | * Initial heatmaps * Communication of “next steps” to QR team and request for additional information |
| 4th October | * Obtain QR feedback on proposed approach * Pursue additional information requested | * Finalise workplan and initial approach |
| 11th October | * Finalise project proposal | * Project Proposal submitted * Proposal Presentation Slides and Video |
| 18th October | * Finalise parameters to be used within the model | * Additional datasets obtained |
| 25th October | * Develop model | * Initial model |
| 1st November | * Develop model | * Objective tests of predictive power |
| 8th November | * Develop visualisation | * Submit Progress Report |
| 15th November | * Develop visualisation | * Complete visualisation |
| 22nd November | * User-testing * Refine user-interactivity model | * Obtain results of usability survey |
| 29th November | * Finalise submission * Identify limitations and potential future improvements | * Submit Poster Presentation Video * Submit Final Report |

*NB: items in green are mandatory deliverables*

All team-members have contributed equally to the Project.

**Appendix: Literature Survey**

1. Nakhaee, Hiemstra , Stoelinga, van Noort: *The Recent Applications of Machine Learning in Rail Track Maintenance: A Survey*
2. Hajizadeh, Nunez, Tax: *Semi-supervised Rail Defect Detection from Imbalanced Image Data*
3. Sharma: *Data-Driven Optimization of Railway Track Inspection and Maintenance Using Markov Decision Process* MS to review in detail
4. Peng, Kang, Ouyang: *A Heuristic Approach to the Railroad Track Maintenance Scheduling Problem.* Not sourced
5. Peng: *Scheduling of Track Inspection and Maintenance Activities in Railroad Networks* Not sourced
6. Eriksen, Gascoyne, Mangan, Fraser: *Practical Applications of GPR Surveys for Trackbed Characterisation in the UK, Ireland, USA and Australia* MS to review in detail
7. Shao, W., Bouzerdoum, A., Member, S. and Phung, S. L. (2011) ‘Automatic Classification of Ground-Penetrating-Radar Signals for Railway-Ballast Assessment’, *IEEE Transactions on Geoscience and Remote Sensing*. IEEE, 49(2), pp. 3961–3972. doi: 10.1109/TGRS.2011.2128328.
8. Landgraf, M. and Hansmann, F. (2018) ‘Fractal analysis as an innovative approach for evaluating the condition of railway tracks’, 233(6), pp. 596–605. doi: 10.1177/0954409718795763.
9. Lasisi, A. and Attoh-okine, N. (2018) ‘Principal components analysis and track quality index : A machine learning approach’, *Transportation Research Part C*. Elsevier, 91(March 2018), pp. 230–248. doi: 10.1016/j.trc.2018.04.001.
10. Queensland Rail: *Queensland Rail Annual Report 2018-2019*