NAC Executive Insights

Technology
June 10, 2020

Verification and Validation of Project Management Artificial Intelligence

Key Points

- Proper reliance on artificial intelligence (AI) in project management requires strong AI predictive tools.
- Data
  - Project management AI systems require measures to assess the quality of the domain populations (training and testing).
  - AI data validation and model selection require rigorous attention.
  - Individual users who contribute data to a multi-enterprise training data set must retain sufficient rights over their data while the broader (multi-enterprise) insights gained are derivative.
- Project management AI is founded on predictive analytics, at its best diagnostic, but with an expert viewpoint.
- Verification and validation
  - The process of verification and validation (V&V) is about establishing trust.
  - AI does not lend itself to checking using conventional V&V methodologies.
  - AI V&V will increasingly rely on model checking and measurement.
- Bias
  - We can identify several types of bias with potential impacts on project management AI, among them sample, exclusion, observer, and prejudice.
  - Verification and validation of PM-employed AI need to include an important element of bias testing.
- Project management AI will benefit from an anomaly detector.
- Verification must ensure that the developed intelligent system conforms to its specifications and that its knowledge base is consistent and complete within itself.
- Validation is the process of ensuring that the output of the intelligent system is equivalent to those of human experts given the same inputs.
- We must remain cognizant that a learnt model in one domain may not apply to another.

Introduction

Project management AI is founded on predictive analytics. It reflects uncertainty, which is best described as a confidence level. Confidence will change over time as the project progresses and new data are considered by the AI model.

Project management AI is at its best diagnostic, but with an expert viewpoint. It is a decision aid and as such is part of the broader project management system. AI can help the project team make better decisions.

It is unlikely that users will trust an AI system no matter how impressive some of its demonstrations of competence may be without some causal explanation of its behavior.

Verification and validation are part of that causal explanation process.

Verification and Validation (V&V)

There is a need to continuously define verification and validation since a range of thoughts exist across the literature. This definitional challenge can be seen in the differences regarding V&V between the Institute of Electrical and Electronic Engineers (IEEE) and the U.S. Department of Defense (DOD).

Verification

Verification deals with satisfying specifications. It involves the structural correctness of the knowledge base and is internally consistent and complete. Among the challenges are a taxonomy that addresses key milestones and activities in the project execution process. Such a taxonomy must be capable of transcending project types. In addition, a tendency to redefine the taxonomy for each project or data subset must be avoided.

Verification of AI-enabled project management systems must move beyond static rule testing given the non-deterministic nature of AI programs. We must check for inconsistencies and incompleteness, discrepancies, ambiguities, and redundancy. Verification must ensure that all portions of the rules base are exercised in testing. Testing with known results may leave us susceptible to errors because of weak or incomplete coverage of the test set. Verification of AI-enabled project management systems requires us to think about underrepresented data subsets and late stage or other temporal failure regimes. For example, how does our AI model behave when presented with a test data set of all project successes and asked to look for failure?

Accuracy depends on the training data set. Issues of correctness, completeness, and appropriateness of source data quality can be failure points. Thus, automated data quality checks become a necessity.

Project management AI systems require coverage measures to assess the quality of the domain populations (training and testing) and meta-knowledge to provide guidance on:

- Fitness for a specific use case.
- Likelihood they are in the training population.
- How representative the test set is of the intended population.

Standardized measures and ontologies need to be developed to allow the proposed project management AI systems to be evaluated.

Verification must happen first, and only then can validation proceed.

Validation

Validation involves exercising the system and testing the project management AI. This is a dynamic process in which we test for functional correctness, comparing the behavior and predictions of the project management AI to the real world, or at least our interpretation of it. We define failure. We define an acceptable level of predictive confidence. We define what constitutes a valuable lead time over timely human prediction. All these considerations and others go into our validation of the developed AI. Data validation and model selection to date, however, have not received enough rigorous attention and much more remains to be done.

One other consideration deals with validation for intended use. This is discussed later.

Role of V&V

The process of verification and validation is about establishing trust. Users must trust adaptive, non-deterministic, or complex AI systems.

Trust = fitness for use, reliability, and robustness

Because AI does not lend itself to checking using conventional V&V methodologies, however, building trust is harder. AI is complex and therefore more difficult to understand and test. Methods that have been developed for assessing normal software still apply, but are best described as good but not sufficient.

Traditional testing is typically scenario based, seeking to consider the maximum number of potential situations with the minimum number of tests. AI testing is compounded by its complexity and could conceivably take more resources than its initial development. With AI, the number of potential situations is too large.

AI V&V will increasingly rely on model checking. The distinct advantage of model-based AI systems is that the high-level description is the system. Declarative knowledge describes facts and relationships within a domain, making it easier to understand/use/communicate, and—importantly—build trust. Simple queries for this declarative model reduces V&V efforts, but only if we have necessary confidence in both the model and the data used to train it.

Several early efforts in project predictive analytics have shown an ability to detect failure but have fallen short on predicting success. Is the inability to predict success a model weakness? Does an opportunity exist to test the sensitivity to failure by changing one variable at a time over a range? What does model performance and our V&V activities tell us about our own project execution methodologies?

Clustering analysis will allow us to find related parameters, yielding new insights for the project manager. AI-enabled planning and scheduling tools can be treated similar to model checking in that they all explore a state space[1] described by a model.

V&V of AI-enabled project management systems should ensure that we see sanity properties such as consistency, absence of ambiguity, or expected properties such as functional dependency between variables. The resultant tools must provide trustful diagnosis such as being able to infer accurate and sufficient information on the state of a project from its observed behavior.

Bias

We must address the subject of bias and the notion of artifacts or artificial patterns that are caused by deficiencies in the data collection process. For example, in our quest to understand project failure we preferentially populate the training data set with failed projects, not providing the tool with a representative sampling of successful projects. Alternately, the data quality of many of our worst performing projects excludes them from our training data, limiting our ability to potentially see the worst of the worst.

Other examples of potential data deficiencies could include under- or over-representation of one client or client type, geographic region, or project delivery method (E, P, C; EPCM; EPC; PPP/PFI[2].

Appendix 1 defines several types of bias potentially impacting project management AI.

Verification and validation of PM-employed AI needs to include an important element of bias testing. Model predictions must be consistent for all possible inputs. Did the system learn something we are unaware of, such as failure linkage to size, which becomes problematic in a different currency or after escalation? We need to test for true positives, true negatives, false positives, and false negatives.

Special Challenges

Non-determinism is a V&V issue for AI. AI systems are prone to error due to their complexity. AI, specifically AI based on machine language, is non-deterministic. That is, uncertainty may be found on the system's future behavior. It can behave differently for different runs.

Non-deterministic choices generate from incoming external events, scheduling of concurrent tasks, and intentional random choices. This result is exponentially many possible executions, creating a state space explosion.

[1] The state space of a dynamic system is the set of all possible states of the system. It models a physical or project system as a set of input, output, and state variables.
[2] Public Private Partnership (PPP); Private Finance Initiative (PFI)

Concurrency is a natural source of non-determinism since the order and timings of independent processes can vary. We see this in project execution and recognize that a new set of risks emerges as we increase project concurrency through techniques such as modularization and parallel execution. It is important that we understand and minimize concurrency risks.

Stochastic non-determinism can arise in AI-enabled project management systems through adaptation of system behavior as a result of new project data. We can learn from the changes these data additions have on our model, and thus understand how the system responds to these new internal choices. This allows us to increasingly turn them into external choices and opportunities. Within a non-deterministic program, a small number of assumptions control which non-deterministic option the program will take. We see this in master-variable scheduling and machine learning in other domains. This provides a project management opportunity to focus on control of a few variables[3].

Adaptive systems change in response to expanded data sets[4]. In a sense they can improve themselves, becoming more “confident” and executing faster. This evolving nature creates a challenge for verification and validation. Adaptation may render obsolete any pre-adaptation certification, perhaps necessitating dynamic comparison between the current model being used and the certified model. Understanding how the model migrates over time can provide new insights.

Test Methods

We have previously discussed the need for quality unbiased data. Now, however, we need enough data to adequately test the model and its adaptation with new project examples. In a sense, the first step in the validation process is ensuring the adequacy of the training and test set. The model developed from the training data will be initially tested with data not used to build it. We must remain cognizant that a learnt model in one domain may not apply to another.

How does learnt theory change over time as more data are processed? When does learning stabilize (that is, when do more data not change the model)? Have we collected too much data in a given domain? Sequence studies can aid in answering these questions. One may also consider artificial distributions such as planned behavior.

Project management AI will benefit from an anomaly detector that tests if new data differ from data previously managed. A pre-filter could reject new input if too anomalous. A post-filter could recognize unusual output and stop that from effecting the rest of the system. This addresses the need to ensure stability of the output theory.

Such an anomaly detector could also be relevant for an ongoing fit-for-purpose test. This could reveal if the current model deployment is out of bounds of fit-for-purpose.

---

[3] This is referred to as the funnel assumption. It is argued that in searches through a space containing uncertainties, most of the reachable conclusions will be reached via a small number of "master variables" in a "narrow funnel."
[4] When the data mined indicate systemic changes, the self-modifying prognostic system refines the previously developed algorithm. (7)

Testing methodologies should investigate the value of sensitivity analysis, robustness to small perturbations in inputs, and identify-useful metrics for AI-based software, and then validate their value as predictors of important cost, performance, and reliability measures. Alternative models, such as regression models, may be derived from sufficiently large data sets and can be used in conjunction with the AI-developed model and provide a reasonableness test.

Summary

Proper reliance on artificial intelligence in project management requires strong AI predictive tools, with known confidence levels at various time frames (less confident prediction of failure early on but with a strengthening predictive confidence as more time lapses), including:

- Transparent and robust AI algorithms, trained on known, relevant data sets and validated for intended use.

- Knowledgeable deployment of validated AI to use cases verified to be consistent with the validated AI.

- Recognition of AI limitations due to excluded data (external ecosystem data) and an assessment of the relevance of its consideration in the particular use case (project).

Verification must ensure the developed intelligent system conforms to its specifications and that its knowledge base is consistent and complete within itself. Special attention on data is required, including understanding and confirming any of a range of potential biases. In addition, transparency and verification of user data rights emerges as a core issue.

Individual users who contribute data to a multi-enterprise training data set must retain sufficient rights over their data while the broader (multi-enterprise) insights gained are derivative. User data must be protected and users must maintain control over access and usage of their data.

Other data, of uncertain or unknown provenance, should not be used in an AI algorithm or service without verifying rights, applicability, and appropriateness for intended use and embedded bias.

Users should ensure that both their data and any AI algorithms they make available are robust against manipulation.

Validation is the process of ensuring that the output of the intelligent system is equivalent to those of human experts given the same inputs. The robust validation regime that is suggested is necessary to achieve a high degree of confidence in the validity of the AI algorithm and therefore confidence that it has been represented and used with a high degree of explanability.

The explanability should have as a minimum the following attributes:

- Decision making process should be explainable.
- Recommendations should include sufficient explanations, data used and limitations, and reasoning.
- AI decision processes should be verifiable.
- AI intent should be transparent.

− AI algorithms may be powerful and scalable, but also transparent to inspection.
− Understand how added data may change expected outcomes.

Appendix 1

Bias Potentially Impacting Project Management AI

Sample bias – Sample bias arises when training and test data do not represent the project environment in which we are deploying the project management AI tool. For example, we train the tool on oil and gas projects but deploy it on infrastructure projects. This suggest a potential need to parse our universe of data into robust subsets trained for the intended use. While a broadly trained tool may provide different and even potentially more valuable insights than one trained on a subset of the data, it would not be unreasonable to expect the more specialized data subset to provide higher confidence insights. An evaluation of these two approaches is warranted as the industry moves forward. We must extend validation theory to define data validation—does the data provide a valid representation of the problem space?

Exclusion bias – During initial tool development and testing, it would not be unreasonable to limit the number of project features to be considered. As tool development proceeds, it will be more valuable to include more fields of data. For example, when considering overall success/failure of large complex projects one may note that delayed completion of the process engineering stage correlates very strongly with overall project failure. As we expand the range of data considered, it would be advantageous to incorporate more data related to the precursors of process engineering, external influencers, and actual process engineering performance. Such inclusion would allow for diagnosis of process engineering challenges earlier in the execution process, focusing corrective action at an earlier point, even before overall project diagnosis might otherwise have flagged a problem. In the data collection and preparation stage we must take care in cleaning any data and be aware that early reporting period data may actually reflect plan versus actual as the project management systems are being stood up, or worse, the project is experiencing poor project startup.

Observer bias – The training and testing teams see what they expect to see. Their thinking is anchored, and weaker but important insights may be overlooked. In the training and testing stage it is important to screen for potential biases by involved humans. Later in tool deployment, we see a different form of observer bias, denial, which acts to reject what does not fit our current mental models.

Prejudice bias – This arises in the selection of training and testing data where stereotypes are unconsciously reflected in the data. There is, therefore, a need for a more even-handed distribution of examples by avoiding input of perhaps not enough successful projects or not enough in a given size range within the scope of intended use. Proper representation at the extremes of the intended use range is particularly important.

Measurement bias – This reflects a systemic bias in how/what we measure. For example, we measure funds expended versus commitments made; actual productivity versus rate of productivity improvement; and request for information (RFIs) or number of holds versus drawings issued for construction.

References

1. Iyengar, V., Subramanian G, S. (2018), The Right Testing Strategy for AI Systems, An Infosys Viewpoint
2. Prieto, R. (2019). Proper Reliance on Artificial Intelligence in Project Management; PM World Journal, Vol. VIII, Issue VIII, September
3. Gonzalez, A., Barr, V. (2000). Validation and verification of intelligent systems – what are they and how are they different? Journal of Experimental and Theoretical Artificial Intelligence, Volume 12
4. Ghoneim, S., 5 Types of bias & how to eliminate them in your machine learning project; Data Science in the Real World
5. Kohli, P., Dvijotham, K., Uesato, J., Gowal, S. (2019) Identifying and eliminating bugs in learned predictive models, Deep Mind blog
6. Pullum, L., Darrah, M., Skias, S., Tso, K. S., Tai, A. T. (2004) Developing a Data Driven Prognostic System with Limited System Information, IEEE
7. Menzies, T., Pecheur, C. (2004) Verification and Validation and Artificial Intelligence
8. Rushby, J. (1988) Quality Measures and Assurance for AI Software; Technical Report CSL-88-7R performed for the National Aeronautics and Space Administration under contract NAS1 17067
9. Prieto (2019), Verification & Validation of Project Management AI; PM World Journal, Vol. VIII, Issue X, November

About the Author

Bob Prieto was elected to the National Academy of Construction in 2011. He is a senior executive who is effective in shaping and executing business strategy and a recognized leader within the infrastructure, engineering, and construction industries.