Requirement Engineering of Machine Learning – Enhancing Evaluation of Machine Learning Models

Paper by Sky Qiao

Presented by Sky Qiao

UNIVERSITY OF WATERLOO | FACULTY OF MATHEMATICS

Paper Overview

- Why we need Requirement Engineering for Machine Learning?
  - To have more efficient communication with the other team
  - Ensure the data meet desired property for ML systems
  - Specify clear and measurable objectives for LM.
  - Document and manage requirements throughout the system's lifecycle.

Paper Overview

- Requirements is always there, why RE?
  - Need to find a way to make sure requirements are met.
- This paper we aim to solve this problem by find a way that is
    - Quantifiable
    - Meaningful
    - Cost-efficient

Paper Overview

- Problem Statement
  - Given a set of requirement, we want to create a quantitative framework to evaluate the fulfillment of requirements.

Assumption

- ML System
  - Addressing models in the context of multi linear regression
- Requirements
  - Only look at the Predetermined Requirements defined later

Requirements

- Common Requirement in ML project
  - Data Requirements
  - Performance Requirements
  - Maintainability Requirements.

Data Requirements

- Data Availability
    - Sufficient data availability is a prerequisite for ML projects.
- Data Quality
    - Ensuring high-quality data is vital, as it directly influences the model's performance.
- Data Balance
    - Make sure the data does not contain class imbalance.
- Problem is that defining the term “Sufficient”, “High-quality” and “Balance” can be subjective.

Data Requirements

- Solution:
  - Collaboration with Domain Experts: Ask domain experts to define those terms for us.
    - Not Cost-efficient, does not scale
  - Standardized Metrics: Establish standardized metrics and guidelines for data availability, data quality, and data privacy.
    - Objectivity, Scalability, and Consistency

Metrics for Data Quality

- Data Availability
  - To estimate the number of sample we need, we can calculate confidence interval limits
- Data Quality
  - To ensuring the data's quality, we can ensure the data meet some properties.
  - For Multi-Linear Regression, the desire properties are: Normality of residuals, No multicollinearity and no functional misspecification.
- Data Balance
  - To ensure data balance, we can ensure data have the same number of samples for each class

Metrics for Data Quality

- Data Availability
  - Formula of CI:

Unlimited population:

$$CI = \hat{p} \pm z \times \sqrt{\frac{p(1-p)}{n}}$$

Finite population:

$$CI' = \hat{p} \pm z \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n'} \times \frac{N-n'}{N-1}}$$

  - To calculate the sample size,

Unlimited population: $$n = \frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2}$$

Finite population: $$n' = \frac{n}{1+\frac{z^2 \times \hat{p}(1-\hat{p})}{\varepsilon^2 N}}$$

where
- **z** is z score
- **p̂** is the population proportion
- **n** and **n'** are sample size
- **N** is the population size

Metrics for Data Quality

- Data Quality - To Check:
  - Normality of residuals - Shapiro-Wilk Test
    - If p-value is greater than the significance level (0.05), the normality assumption is satisfied.
  - No multicollinearity - Variance Inflation Factor (VIF)
    - If VIF values is less than 5, no multicollinearity assumption is satisfied.
  - No functional misspecification - RESET (Regression Specification Error Test)
    - If p-value is greater than the significance level (0.05), no strong evidence of functional misspecification

Metrics for Data Quality

- Data Balance:
  - Use automation tool compute the sample counts for each class
    - Reject the data if the counts are not equal for all classes.
    - Otherwise, approve the data.

Performance Requirements

- Accuracy
  - Achieve a good accuracy on the test set
- Training time
  - To be able to finish training in a short time
- Testing time
  - Model is able to give out a result in a short time
- Problem: What is “good accuracy”, “short time”?

Metrics for Performance Requirements

- Accuracy
  - We define that in MLR, a **good accuracy** means the MSE value is smaller than 0.05.
- Training time
  - We define that in MLR, a **short training time** means the time it takes to train a model is less than 1 hour
- Testing time
  - We define that in MLR, a **short testing time** means the time it takes to test one data point is less than 10mins

Metrics for Performance Requirements

- Accuracy - MSE value is smaller than 0.05
  - The train-test split we will use is 0.8 to 0.2.
  - MSE (Mean Squared Error) < 0.05
    - means the **squared difference** between each predicted value and its actual value is less than 0.05.
  - After testing, if the MSE value is strictly less than 0.05, we say our model has a good accuracy.

Metrics for Performance Requirements

- Training time & Testing time
  - We will use the **time** function to measure the time it takes.
  - The machine we will use is the standard setup of the author's machine.
  - After training and testing, if the training time it takes is strictly less than 1h, we say our model has a short training time; if the testing time it takes is strictly less than 10mins, we say our model has a short testing time .

Maintainability Requirements

- Documentation
  - We want to have explanation on every function in the code
- Code Commenting
  - A good maintainability measure is that we have lots of comment to explain what each line is doing
- Code Quality
  - The code is both easy to understand and straight forward

Metrics for Maintainability Requirements

- Documentation
  - We define **a good documentation** means
  - $\forall$ functions f, $\exists$ a comment C such that C explains the behavior of f
- Code Commenting
  - We define a project as '**well-commented**' if and only if the number of lines containing comments is greater than one third of the total number of lines in the project.
- Code Quality
  - We define a project with high code quality means that each commit is reviewed by at least two people.

Maintainability Requirements

- Documentation
  - To check if the requirement is meet, we will use automation tool to detect if there is a comment block before every function in the code.
  - If we did not detect comment block before a function
    - Return Bad Documentation
    - Otherwise, return Good Documentation.

Maintainability Requirements

- Code Commenting
  - To check if the requirement is meet, we will use automation tool to count the number of line that has comments **c**, and total number of lines **n**, in the project.
  - If $n/3 > c$:
    - Return Not well-commented
    - Otherwise, return well-commented

Maintainability Requirements

- Code Quality
  - To check if the requirement is meet, we will set up the git tool, to ask for two person's approval before merge request into the main.
  - With this setup, we can say that each commit is reviewed by at least two people

Conclusion

- RE is an important part of ML
- We need this framework to evaluate the fulfillment of requirements objectively.
- With Stats method and modern tool, we can find a solution to measure/ensure the fulfillment of the requirement.

Future Work

- Expand the framework to apply it to more requirements.
- Eliminate the hard threshold we set up with some methods that can be justify.
- Do experiments to testing this system:
  - Cost it takes to use this framework.
  - Friendliness of using this framework
  - Effectiveness to maintain the requirement of the project as it grows larger

Thank You!

Reference

- (2023). A Survey of Data Quality Requirements That Matter in ML Development Pipelines. https://dl-acm-org.proxy.lib.uwaterloo.ca/doi/pdf/10.1145/3592616

Q&A Session