Article Review: Requirements Engineering for Machine Learning: Perspectives from Data Scientist

Machine learning is used in many applications these days. ML is used by many companies to improve its products. There are many challenges that takes place in requirement engineering when used with ML applications. Recent study suggests that requirement engineering for ML based system is very difficult. Currently all the decisions of a ML system is taken by Data Scientist. However, it should be based on domain needs and business requirement and due to which it is responsibility of requirement engineer.

The paper performed interview of 4 data scientist to understand better about requirement engineering and the findings were that the requirement engineer must know the new requirements which are there due to ML paradigms. Machine learning including Supervised, Unsupervised and reinforcement learning and self-adapting system. The main concern in requirement engineering for ML-based system is that in traditional system activities such as requirement analysis, specification at the start and acceptance and inspection at end do not work with ML based system as prior estimation or accuracy is not possible in ML.

Self-adaptive systems collect data continuously and monitor their behavior based on the data the system decides whether any adaptation is required or not. As RE for self-adapting system is incomplete many decisions must wait until run-time and that is a challenge. The study research tries to understand how data scientist understand and document requirements for ML system what is the process that they follow to get requirements and what challenges do they face. As ML is new field there are specialized people who work in it and hence for research, they considered data scientists who are experts on this field. Each author coded one interview and reviewed and validated the codes of the interview of other author and these results were discussed in a meeting and then combined all the data to see if entire data is considered and covered. Certain threats is that if the interviewer do not collect the entire data and to solve this issue the interviews were taped.

For any successful system the requirement must be clear in ML resulting prediction quality is functional requirement. Data scientists need skills to make client set reasonable targets and this skill helps in RE. Mostly in ML it is difficult to understand the code but in most ML system we do not need to change the code, but we need to change the training data. Also, there is requirement the data scientist must know the ML model but there is a chance that entire data is not known by the data scientist. In ML we need to train data and data requirement is also a requirement in ML based systems. Increasing quality and quantity of data increases the value generated by ML based systems. Also, another important part of ML based RE is training data is close to actual data present.

To conclude I think ML based system are different from traditional software system and hence RE for ML based system is different and difficult. Many factors discussed come into picture for requirement engineering for ML.