Iowa State University

COMS 509

Requirements Engineering for Machine Learning: Perspectives from Data Scientists

Vogelsang, Andreas, and Markus Borg.

IEEE 27th International Requirements Engineering Conference Workshops (REW). IEEE, 2019.

Aishwarya Sarkar

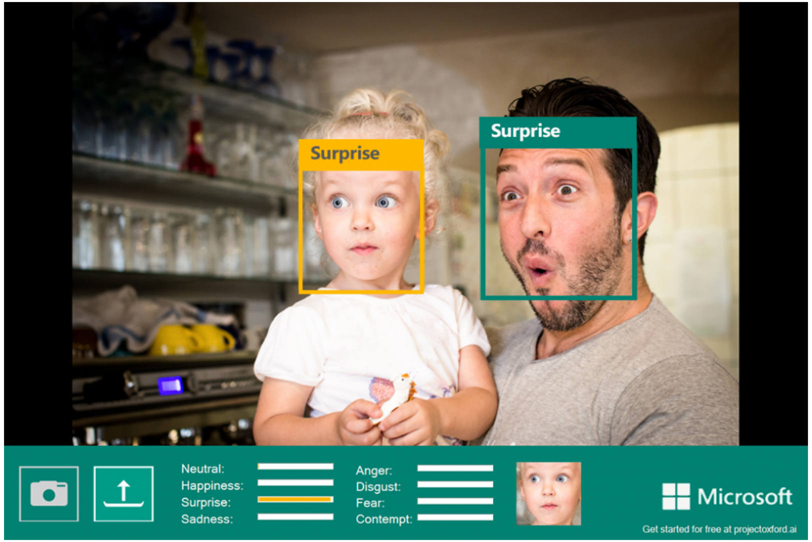
Ph.D. Student

Overview

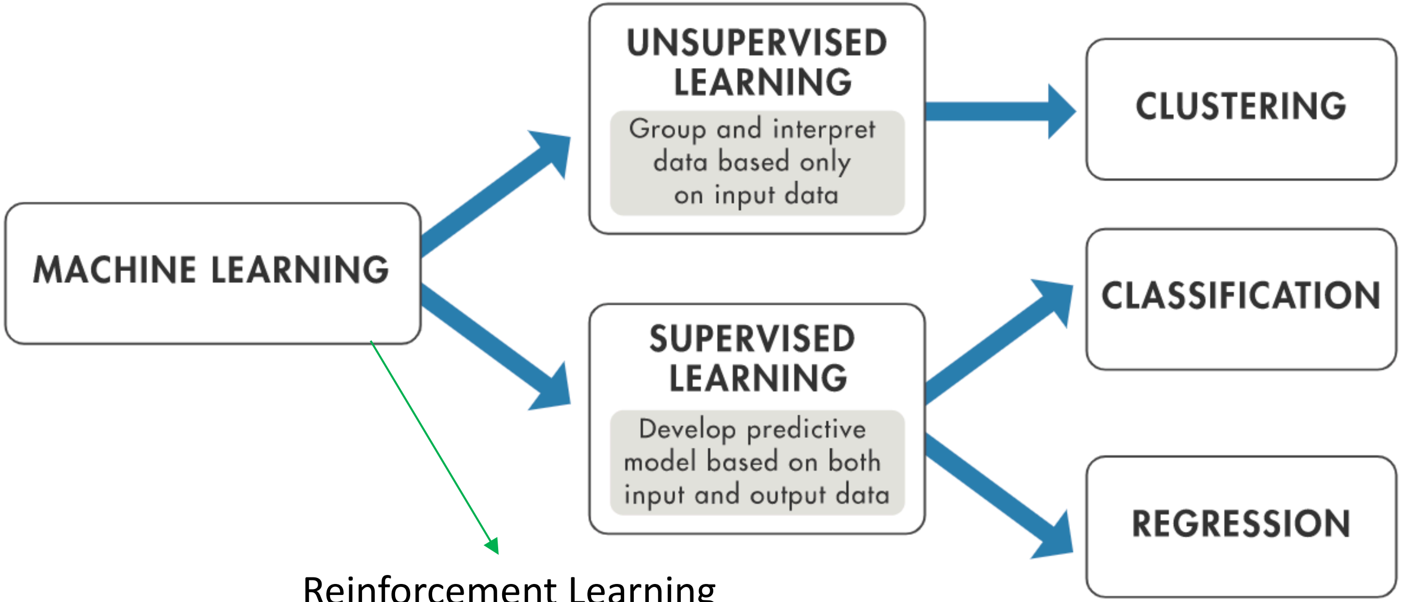
1. Where does the problem start?
2. Types of ML Approaches
3. High-Level Idea of different ML approaches
4. The Problem
5. Contribution of the paper
6. Suggested RE-process for ML Systems
7. An interesting study by Microsoft team
8. Evaluation
9. References

Where does the Problem start?

- Machine Learning (ML) has accomplished major technical breakthroughs in image recognition, natural language processing, and beating humans in complex games.
- This is made possible by
  - Improved computational power of modern processors (GPUs)
  - Large availability of data
  - Accessibility of these complex technologies by rather simple-to-use frameworks and libraries
- Many companies started using ML to improve their products and processes.

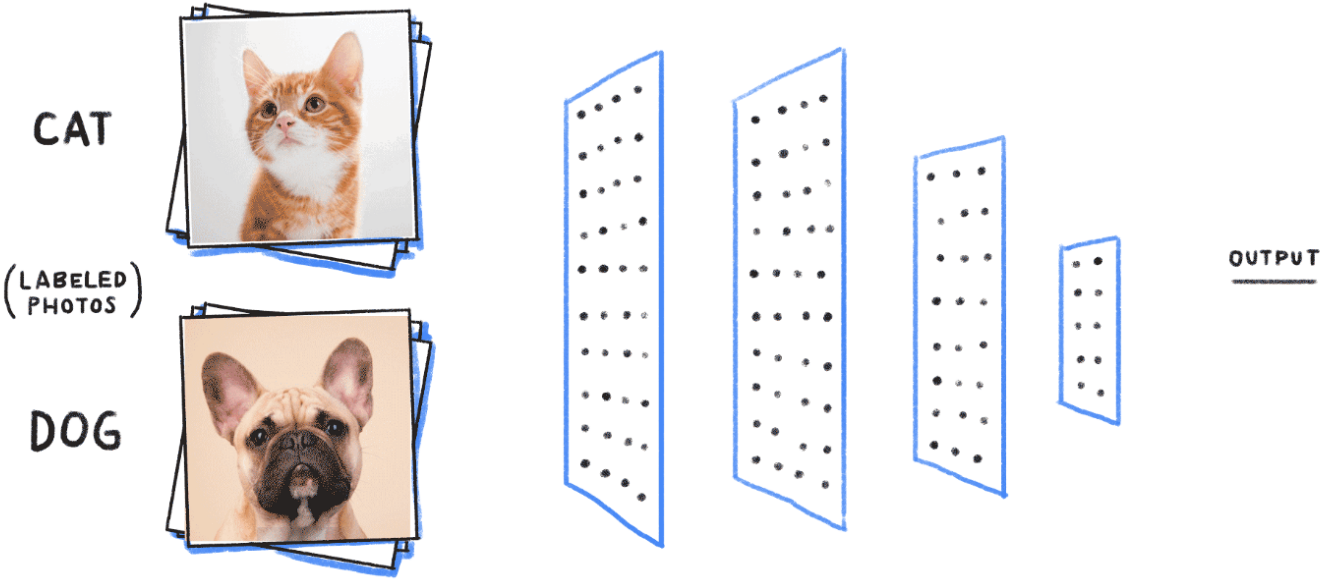


Types of ML approaches

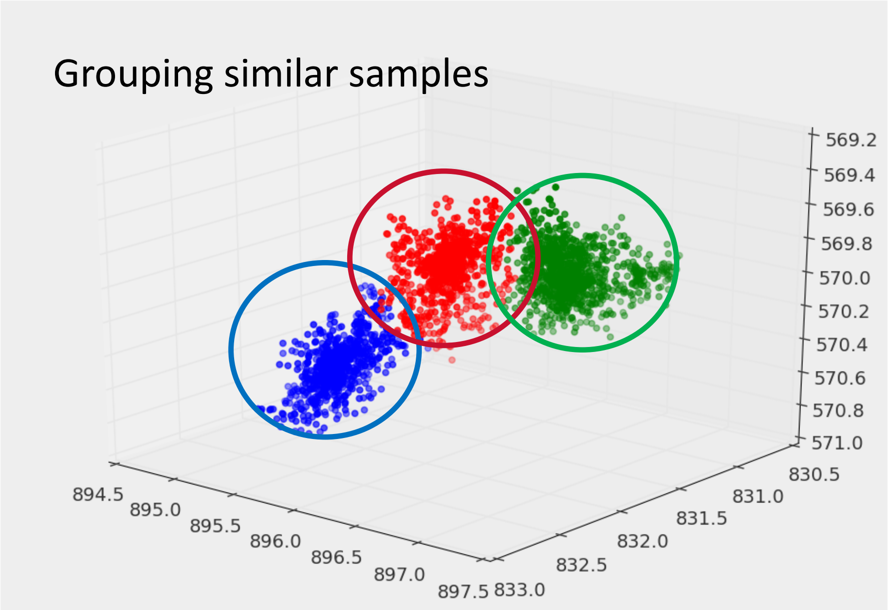


Reinforcement Learning

How does ML work? An Example of Supervised Learning - Classification



Clustering and Regression



Clustering

Effect

Cause

$$y = a + bx$$

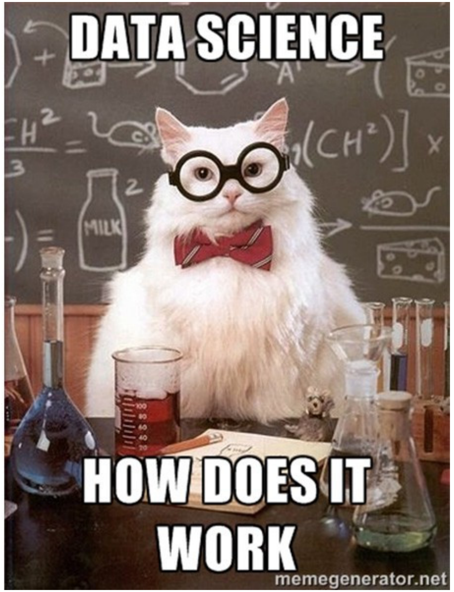
Income

Education

Linear Regression

The Problem – Paradigm shift

- A paradigm shift compared to conventional software engineering.
- In this paper, the authors share results that indicate that RE must evolve to match the specifics of ML systems.
- Right now, Data Scientists decide the **definition of the fitness functions, the selection and preparation of data, and the quality assurance**.
- However, these decisions should be based on the understanding of the business domain and the stakeholder needs.



Contribution of the paper

- Conducted interviews with four data scientists to explore their perceptions on RE.
- The interviews covered specific requirements for ML systems, challenges involved in RE for ML, and how the RE process needs to evolve.
- We were interested in the following research questions:
  - **RQ1:** How do data scientists elicit, document, and analyze requirements for ML systems?
  - **RQ2:** What processes do data scientists follow and which parts relate to requirements?
  - **RQ3:** What are requirements-related challenges that data scientists face?

A. Quantitative Targets a.k.a. Functional Requirements

- Just like for conventional software products, it is **hard to satisfy a client without making expectations explicit**.
- Interviewees mentioned measures to quantify the predictive power including accuracy, precision, and recall. P3: "**The customers don't understand the performance measures**."
- **Selecting and interpreting performance measures appropriately** is crucial for the acceptance of ML systems.
- Interviewees stressed that data scientists need skills to help a client set **reasonable targets**, including domain understanding, statistics, and computer science.
- **In ML systems, the quality of the resulting predictions can be considered a functional requirement**

B. Explainability

- ML systems are **hard to understand for human analysts** since models are adjusted to fit **specific data**.
- The **ability to explain decisions** in ML systems depends on the used technology.
- **A Requirements Engineer should elicit explainability requirements from a user's point of view**

Difference in Performance Measures

Classification

Performance Metric – Accuracy, Precision, Recall

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} = \frac{1 + 90}{1 + 90 + 1 + 8} = 0.91$$

Regression

Performance Metrics - RMSE/MAE/MAPE

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (Predicted_i - Actual_i)^2}{N}}$$

TP – True Positive, TN – True Negative, FP – False Positive, FN – False Negative
RMSE – Root Mean Square Error, MAE – Mean Absolute Error, MAPE – Mean Absolute Percentage Error

C. Freedom from Discrimination

- ML systems are designed to discriminate - identify recurring patterns in data (i.e., stereotypes) and apply these patterns to judge about unseen data.
- Some forms of discrimination are considered as unacceptable by our society or by law (e.g., race or gender)
- Introduced by unrepresentative training data
- Availability of tools and algorithms that can balance discrimination
- Need to know which attributes you must balance
- **Hence, a requirements engineer must elicit and identify the “protected” characteristics that must not be used by the ML algorithm to discriminate samples.**

D. Legal and Regulatory Requirements

- An example is how the General Data Protection Regulation (GDPR) constrains that personal data can only be used in ways specified by an explicit consent.
- Revoked Consent leads to retraining of the whole model
- **Requirements engineers working on ML systems must stay on top of legal requirements, to ensure no illegal features have influenced the final dataset used for training the ML models**

E. Data Requirements

- Training data is an integral part of any ML system.
- Training data need specified and validated requirements like code
- **Data Quantity and Data Quality**
- Requirements on data quantity should not be specified on the number of examples but rather on the **diversity of the examples**
- The **higher the quality** of the data, the better the application will work.

RE-PROCESS FOR ML SYSTEMS

A. Elicitation
B. Analysis
C. Specification
D. Verification and Validation (V & V)

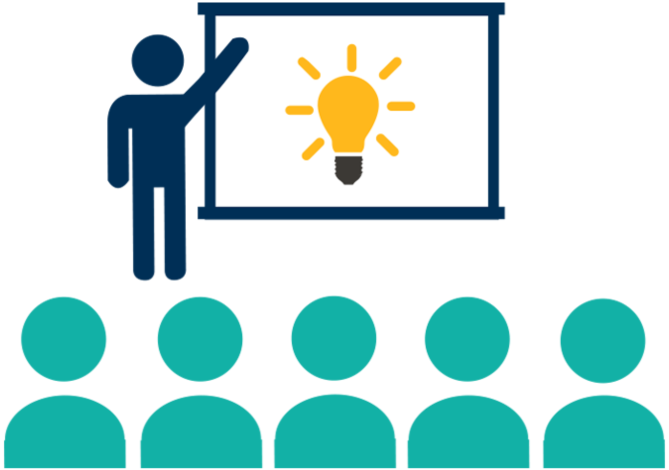
TABLE I
RE FOR ML: IMPACT ON RE ACTIVITIES

Elicitation	Analysis	Specification	V&V
• Elicit additional data sources • Important stakeholders: Data scientists and legal experts • "Protected" characteristics?	• Discuss performance measures • Discuss conditions for data preparation, definitions of outliers, and derived data	• Quantitative targets • Data requirements • Explainability • Freedom from discrimination • Legal and regulatory constraints	• Analyze operational data • Look for bias in data • Retrain ML models • Detect data anomalies

A few more observation from a study by Microsoft team[2]

- More education and training
- **ML modularity is not the same as software modularity** –
  - Need for maintaining strict module boundaries
  - Each model's results will affect one another's training and tuning processes
- **Customization and Reuse** –
  - straightforward only in similar domain
  - model may require retraining, or worse, may need to be replaced with another model
- **Data Discovery and management –**
  - Engineers must find, collect, curate, clean, and process data for use in model training and tuning.
  - While there are very well-designed technologies to version code, the same is not true for data

Evaluation



- If these changes are made, it would benefit significantly
- One of the main challenges will be to provide training to the Requirement Engineers.
- Knowledge transfer
- Hyperparameter Tuning
- REs would need substantial experience to understand where to look to find out requirements.
- And of course, more communication should be encouraged between the REs and the Data Scientists.

References

[1] Vogelsang, Andreas, and Markus Borg. "Requirements Engineering for Machine Learning: Perspectives from Data Scientists." *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*. IEEE, 2019.

[2] Amershi, Saleema, et al. "Software engineering for machine learning: A case study." *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 2019.

Questions