Requirements Engineering for Machine Learning

Prof. Dr. Andreas Vogelsang
Technische Universität Berlin

andreas.vogelsang@tu-berlin.de
@andivogelsang
The Message of this Talk

Applications that use **Machine Learning**?
Should I care as a **Requirements Engineering**?
**YES!**
Machine Learning Applications Everywhere

- Virtual Personal Assistants
- Traffic Predictions and Routing
- Social Media Services
- Product Recommendations
- Online Fraud Detection
- Email Spam and Malware Filtering
- Online Customer Support
- Search Engine Result Refining
Machine Learning

• In traditional programming, we write algorithms to solve problems
  • Sorting, searching, calculating function derivatives, solving the towers of Hanoi, navigation route computation, ...

• Task: Identify numbers in hand-written notes
  • Not so easy!

• The Machine Learning Approach:
  • „Train“ a mathematical model to solve this task
  • Training data:
    • Many hand-written notes with correct numbers
Machine Learning

• Training phase:
  • Adjust variables to minimize an error function

• Prediction phase:
  • Use trained model to calculate output based on input
Types of Machine Learning

Supervised Learning
- Data with labels
- Error
- Mapping/Prediction

Unsupervised Learning
- Data without labels
- Classes

Reinforcement Learning
- States and actions of environment
- Reward
- Next Action
Development Changes...

**Traditional Programming**
- Input \( x \) Program \( \rightarrow \) Output
- Knowledge is in the program
- Program quality is important
- Focus on correctness

**Machine Learning**
- Input \( x \) Output \( \rightarrow \) Program
- Knowledge is in the data
- Data quality is important
- Focus on uncertainty

**What about RE?**
Machine Learning Applications

Hybrid systems with ML and traditional parts
Machine Learning Applications from the View of a Requirements Engineer
ML from the View of a Requirements Engineer

RE just as usual?
ML from the View of a Requirements Engineer
ML from the View of a Requirements Engineer

Part of the system?

Training Data

Training

Performance

Input

Trained Model / Program

Prediction

Output
New Types of Requirements for Machine Learning Applications
Types of Requirements

- Process Requirement
- System Requirement
- Project Requirement

- Functional Requirement
  - Data Quantity
  - Data Quality
  - Performance Measures

- Quality Requirement
  - Discrimination
  - Explainability
  - Accessibility and Confidentiality

- Constraint

M. Glinz: “On Non-Functional Requirements”, RE’07
Functional Requirements for ML Applications

How to describe behavior, data, input, or reaction to input stimuli of ML applications?

Data Quantity, Data Quality, Performance Measures
Data Quantity and Quality Requirements

Training

How much data is available for training?

How accurate is the training data?

How representative is the training data?

Input

Output

Performance

Training Data

Trained Model / Program

Prediction
Performance Requirements

What is the demanded performance on the training data?

What is the expected performance in the application?

How is performance measured?
Performance Measures for ML

The truth is

<table>
<thead>
<tr>
<th>Case A</th>
<th>not case A</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positives (TP)</td>
<td>False Positives (FP)</td>
</tr>
<tr>
<td>False Negatives (FN)</td>
<td>True Negatives (TN)</td>
</tr>
</tbody>
</table>

Accuracy = \( \frac{TP+TN}{TP+TN+FP+FN} \)

Precision = \( \frac{TP}{TP+FP} \)

Recall = \( \frac{TP}{TP+FN} \)
Performance Measures for ML

• Example: Identify cancer in X-ray images
• Requirement: “The app shall have an accuracy of > 90%”

• Warning: Imbalanced training data
• What if the training data consists of
  • 95% images without cancer
  • 5% images with cancer
• A (trivial) algorithm that always predicts “no cancer” has an accuracy of 95%

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}
\]

Change the requirement:
“The app shall have an accuracy of > 90% on a balanced training set”

Change the requirement:
“The app shall have a recall for detecting cancer of 100%”
Performance Measures for ML

• Example: Identify cancer in X-ray images

• Requirement:
  “The app shall have a recall for detecting cancer of 100%”

• Warning: Precision vs. Recall Trade-off

• A (trivial) algorithm that always predicts “cancer” has a recall of 100%

• Precision is only 5%. Does that algorithm help?

The truth is

<table>
<thead>
<tr>
<th>Case A</th>
<th>not case A</th>
</tr>
</thead>
<tbody>
<tr>
<td>True Positives</td>
<td>False Positives</td>
</tr>
<tr>
<td>(TP)</td>
<td>(FP)</td>
</tr>
<tr>
<td>False Negatives</td>
<td>True Negatives</td>
</tr>
<tr>
<td>(FN)</td>
<td>(TN)</td>
</tr>
</tbody>
</table>

Recall = \( \frac{TP}{TP+FN} \)

Precision = \( \frac{TP}{TP+FP} \)
Performance Measures for ML

Specifying performance requirements for ML applications demands a rigorous analysis of the problem to be solved (and that is RE work!)

Solution 1: 60% Precision 90% Recall
Solution 2: 90% Precision 60% Recall

Task 1: Detect credit card fraud
Task 2: Recommend interesting articles to customers
Task 3: Identify cancer in X-ray images
Quality Requirements for ML Applications

How to describe specific qualities of ML applications?

Freedom of Discrimination, Explainability, Accessibility and Confidentiality
Quality Requirements for ML Applications

New types of qualities for ML applications
- Freedom from Discrimination
- Explainability
- Accessibility and Confidentiality

ISO 25010: System and software quality models
<table>
<thead>
<tr>
<th>ENGLISH</th>
<th>TÜRKİŞCH</th>
</tr>
</thead>
<tbody>
<tr>
<td>She is a doctor.</td>
<td>O bir doktor.</td>
</tr>
<tr>
<td>He is a nurse</td>
<td>O bir hemşire.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TÜRKİŞCH</th>
<th>ENGLISH</th>
</tr>
</thead>
<tbody>
<tr>
<td>O bir doktor.</td>
<td>He is a doctor.</td>
</tr>
<tr>
<td>O bir hemşire.</td>
<td>She is a nurse.</td>
</tr>
</tbody>
</table>
Freedom from Discrimination

Wrong

Right for the Right Reasons

Right for the Wrong Reasons

Right for the Right Reasons

Baseline: A **man** sitting at a desk with a laptop computer.

Our Model: A **woman** sitting in front of a laptop computer.

Baseline: A **man** holding a tennis racquet on a tennis court.

Our Model: A **man** holding a tennis racquet on a tennis court.

Burns et al.: “Women also Snowboard: Overcoming Bias in Captioning Models”
Freedom from Discrimination

• ML applications are designed to discriminate
• However, some forms of discrimination are considered unacceptable (disability, race, sexuality, gender, pregnancy)

**Freedom of Discrimination:**
Using only logics of discrimination that are societally acceptable

• ML applications amplify biases in data (especially for underrepresented input)

• **RE task:** What are “protected” characteristics?
The component conditionally drives an external fan. This fan is required for active ventilation of the headlight. The duration until the switch is recognized as hanging must be a configurable parameter.

Winkler, Vogelsang: “Automatic Classification of Requirements Based on Convolutional Neural Networks”, AiRE’16
Explainability Requirements

• **Explainability**: The ability to provide hints or indication on the reasons why an application made a decision.

• Explaining decisions in ML applications is hard but not impossible
• Explainability must be implemented into an application from the start

• **RE task:**
  • Which decisions need to be explained?
  • Who needs explanation?
Accessibility and Confidentiality

What is the influence of laws and regulations towards data?

Legal and regulatory data requirements/constraints
RE for ML Applications

Elicitation
- Important stakeholders
  - Data Scientists and Data Engineers
  - Experts in Data Protection Laws
- Scoping
  - Training inside or outside the system?

Analysis
- Get legal approval
- Discuss and define performance measures

Specification
- Functional Requirements
  - Necessary/Available training data (amount, accuracy, representativeness)
  - Demanded performance on training data
  - Expected performance on real data
  - Use well-understood performance measures
- Quality Requirements
  - Is discrimination critical?
    - What are “protected” characteristics?
  - Are there decisions that need to be explained?
  - Influence of laws and regulations on data availability

V&V
- Define measures for data analysis
  - Look for bias
  - Assess quality
- Define measures to control quality in production
  - Outlier detection
  - Field data analysis
Summary

The Message of this Talk

RE + ⚪ = ?

Applications that use Machine Learning?
Should I care as a Requirements Engineering?
YES!

Prof. Dr. Andreas Vogelsang
Technische Universität Berlin

andreas.vogelsang@tu-berlin.de
@andivogelsang