IMPROVING AUTOMATED TESTING ACCURACY WITH MACHINE LEARNING

Presented by Justin Lawler & Mirjana Radovanov
Introduction to Autodesk
Localization in Autodesk

- 90+ Products & Services on different technologies
- 16 languages
- Content: Software, Documentation, Videos, Websites, Forums, AKN (Autodesk Knowledge Network), Marketing campaign, Training material
Localization in Autodesk

Videos de présentation des nouvelles fonctions et de l'interface utilisateur.

Nouvelles fonctionnalités
Avez-vous essayé...
Zoom et panoramique plus rapides
Utilisation d'AutoCAD où que vous soyez
Barre d'état
Suite...
Apprentissage
Nouveautés
Fondamentaux sur les fonctions de base

Présentation de l'interface utilisateur
Aide-mémoires
Commandes | Nouveau | Mis à jour | Obsoleté
Variables système | Nouveau | Mis à jour | Obsoleté
Express Tools
Raccourci | Touches de remplacement temporaire
Ressources

「新功能」和「使用者介面概述」視訊

新功能
學習
連接

資料
自訂 | CAD 管理
開發者文件
發行說明 | 系統需求
常見問題
下截

Autodesk Community

User Conference on
Advanced Automated Testing
Automated Localization Testing

- LAGT (Localization Autodesk UI Testing)
  - Test Cases: 963 (AutoCAD)
  - Products: 30
  - Languages: 16
  - Builds per year: 3

- Manual Analysis and Verification
  - TestDesk
  - 20 Years Old - Need to Retire
  - Manually Very Intensive
Automated Localization Testing - 2018

~4 Million Errors Flagged by LAGT Tool (All Products)
TestDesk Automated Localization Testing
Localization Errors – false positive examples

Truncation Error (French)

Overlapping Error (Chinese)
The Problem

TestDesk Results Management System

LAGT (UI Testing Engine)

Autodesk Products
The Solution – Augmented Intelligence
The Solution - Jarvis

JARVIS Machine Learning Engine

LAGT (UI Testing Engine)

Autodesk Products
JARVIS - Architecture
Building Machine Learning Model

- 8 years of Labeled Data
- >20 Error Types
- Data Stored as
  - Oracle Database
  - XML Files
Machine Learning Pipeline

Data Extraction → Data Cleaning → Feature Extraction → Model Training → Model Validation
ML Pipeline - Data Cleaning

- Remove Duplicate Rows
- Remove Null Values
- Remove Rows with Bad Data
- Remove Constant Columns
- Remove Irrelevant Columns
ML Pipeline - Feature Selection

• Manually Selected
  • Relevant Inputs Only
• Calculated
  • Remove Highly Correlated Features
ML Pipeline - Algorithm Selection

- Binary Classification Problem
- > 60 Input Parameters
- Highly Imbalanced Data Set
- Cannot Miss-Diagnose a True Error
ML Pipeline: Algorithm Selection

- Algorithms Tried
  - Random Forest
  - XGBoost
  - CatBoost
ML Pipeline - Algorithm Selection - CatBoost

- Open-sourced machine learning algorithm
- Why CatBoost
  - Lots of categorical features in the data set
  - Faster Development
    - Better Defaults
  - Handles Imbalanced Data Sets
ML Pipeline – Model Training

- Truncation Model
- 2M Rows Raw Data
- Data Split
  - Training on 75%
  - Validation on 25%
- Trained on AWS Sagemaker
## ML Results CATBoost – Confusion Matrix*

<table>
<thead>
<tr>
<th></th>
<th>Predicted True</th>
<th>Predicted False</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual False</td>
<td>3.6% (35,415)</td>
<td>96.4% (956,073)</td>
</tr>
<tr>
<td>Actual True</td>
<td>87.4% (6,195)</td>
<td>12.6% (896)</td>
</tr>
</tbody>
</table>

*Results for Truncation Errors Only*
# Jarvis Result Server – User Interface

<table>
<thead>
<tr>
<th>No.</th>
<th>Error Type</th>
<th>Element Name</th>
<th>Element Type</th>
<th>Description</th>
<th>Error Status</th>
<th>Conf. %</th>
<th>Comment</th>
<th>Defect</th>
<th>Support File</th>
<th>Analyzed by</th>
</tr>
</thead>
<tbody>
<tr>
<td>142</td>
<td>DLG, UNKN...</td>
<td>DUMMY</td>
<td>DialogBox</td>
<td>Unable to ...</td>
<td>TO BE ANALYZED</td>
<td>90 %</td>
<td>Create</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>143</td>
<td>HARNESS_E...</td>
<td>SilkTest</td>
<td>SilkTest</td>
<td>This testca...</td>
<td>TO BE ANALYZED</td>
<td>70 %</td>
<td>Create</td>
<td>Harness_Er...</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Challenges Faced

- Data Set Size
- Imbalanced Data Sets
- Extracting Legacy Data
Takeaways – Was It Worth It?

- Reduce Turn-Around Time
- Reduces Human Touch-Points
- Replaces Legacy (20 year old) System
  - New User Interface
  - Cloud Based
  - New Database
  - Fix Security Problems
- Cost Savings are a By-Product
Q&A