Automatic Selection of Models for Cost-effective and Sustainable ML Workflows

What It Is

In order to automate complicated processes including document processing, infrastructure monitoring, traffic management, and software development, this project aims to streamline and enhance machine learning (ML) workflows. It presents the Popper system, which balances cost and runtime efficiency while improving the error correction and reliability of these procedures.

The Problem It Solves

ML models, such as OCR engines or classifiers, can discreetly produce errors while processing unexpected or unfamiliar inputs, making it difficult to develop ML workflows. By eliminating troublesome inputs, traditional techniques frequently compromise recall while prioritizing precision. Programmers also have to deal with the tedious and ineffective task of manually choosing the optimum models and settings to optimize operations.

The Technology It Uses

- **ML Workflows and Frameworks (e.g., LangChain)**

  Combining calls to ML models with conventional software logic allows for automated procedures such as identity card processing, which analyzes photos, extracts text and classifies information.

- **OCR Engines (e.g., EasyOCR, Tesseract)**

  Used to recognize text in scanned pictures. To deal with problems from out-of-distribution inputs; Popper enables programmers to define alternative OCR engines.

- **ResNet Classifier**

  Used to categorize pictures or documents. When errors occur, alternative classifiers can be developed to increase accuracy.

- **Popper System**

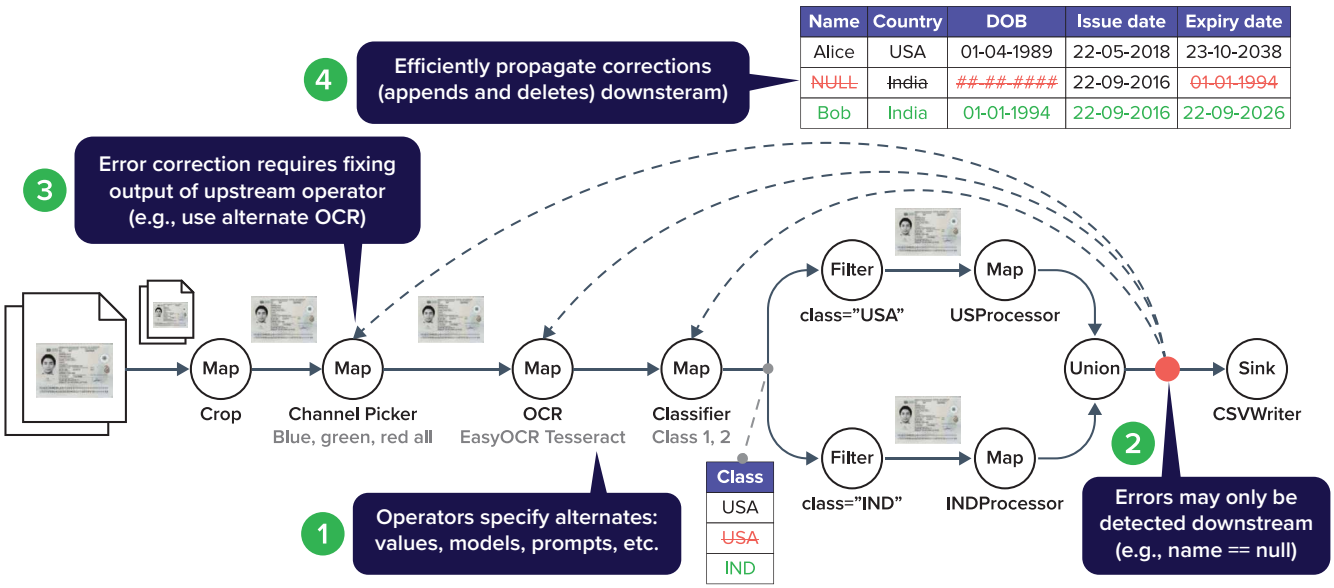
  Incorporates the idea of assertions (such as "name must not be null") to validate the results of workflows. Popper improves precision and recall by retrying the workflow with different models or configurations when assertions fail.

- **Profile-Guided Optimizer**

  Optimizes cost and efficiency for subsequent process steps by dynamically rearranging models based on their performance during runtime.

Key Goals

- Increase accuracy and reliability by permitting alternative models and assertions to solve ML model faults.
- Preserve data that can normally be discarded by retrying unsuccessful operations with different setups, data that would normally be discarded can be preserved.
- Automate the process of identifying the optimal model to save execution time and expense without compromising accuracy.
- Reduce manual load for programmers by giving Popper the ability to dynamically adjust and improve routines using acquired performance insights.



THANK YOU!

R Systems is a leading digital product engineering company that designs and builds next-gen products, platforms, and digital experiences empowering clients across various industries to overcome digital barriers, put their customers first, and achieve higher revenues as well as operational efficiency. We constantly innovate and bring fresh perspectives to harness the power of the latest technologies like cloud, automation, AI, ML, analytics, Mixed Reality etc.

Contact Us

For more information about our solutions or to discuss how we can help your business, please contact us at:

marketing@rsystems.com
www.rsystems.com