PARASCRIPT®

PARASCRIPT SMART LEARNING PRIMER

Parascript FormXtra.AI Smart Learning

OVERVIEW

Parascript has introduced the first combination of true machine learning and intelligent capture with FormXtra.AI. FormXtra.AI is a document automation-focused machine learning platform that provides automated document classification, separation and extraction. It is configured and tuned using a variety of applied machine learning algorithms. This document provides the background of Smart Learning and how it can be applied.

Artificial Intelligence: Traditional AI versus Machine Learning AI

Much has been made over artificial intelligence in the general media and within the technology solution marketplace. Artificial intelligence is being applied to everything, but there are few details regarding what types of AI are used or how they actually work to benefit the user.

Most AI software in-place today is still of the “expert systems” form where developers encode rules based upon subject matter expertise in order to automate specific activities. As one would expect, encoding expertise into a software application can be time-consuming and result in brittle models that only work when expected inputs are encountered with little flexibility.

This problem of excessive effort and brittleness has led to an increased interest in using machine learning algorithms to create rules automatically that are more adaptive and forgiving. Whether the underlying algorithms are Bayesian or neural network-oriented is not the main focus. Automation of work—using a system that can *learn* and *adapt* — is the promise of true machine learning systems.

The primary way to discern between the two major types of AI—traditional expert system versus machine learning—is the manner in which the system is configured.

Artificial Intelligence: Traditional AI vs. Machine Learning AI

Expert systems often use a learn-by-example model where users provide specific instructions in their day-to-day activities. Or, they employ a GUI-based rule-building capability. Or, these two are combined. The result is a gradual building of a knowledgebase of rules that are used to automate work.

Machine learning systems, on the other hand, focus on input data called ***ground truth data.*** Ground truth data is simply tagged data that provides the ***answer key***. For example, it could be a labeled set of pictures containing a cat, along with labeled pictures not containing a cat. The machine learning system uses this base level of input data to identify key features of the *pictures with cats* that are not present in the *pictures without cats* in order to develop a model of how to identify cats.

Machine learning systems need a lot of ground truth data. Unlike expert systems, which use the knowledge of humans to encode rules, machine learning algorithms are hungry for data to develop this knowledge on their own. Deep learning neural networks are the most data hungry. They often require hundreds of thousands to several million samples to create models that deliver reliable results.

While there is effort in curating a sizable sample set, the benefit is a model that trains itself faster and more-reliably than an alternative expert system rules-based model. Machine learning systems can identify and associate key patterns that humans cannot accomplish. Even more importantly, once a trained machine learning system is put into production, it can continue to learn and adapt. A rules-based system cannot without active human participation. Some rules-based systems have implemented learn-by-example in a production environment to try to overcome this deficiency, but this introduces potential human errors and more cumbersome workflows.

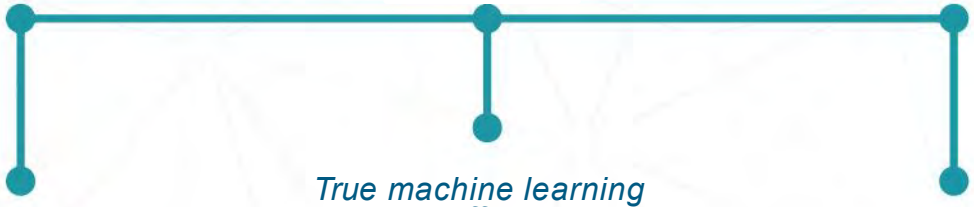
What is Smart Learning?

Smart Learning takes a machine learning approach to configuring and maintaining intelligent capture. This includes common tasks such as image clean-up and preprocessing or ***image perfection***, document classification, document separation and data extraction.

The distinction between Smart Learning and general machine learning platforms is that Smart Learning is purpose-built for document-based information. It includes a significant amount of human-based knowledge about documents and specific document-oriented tasks encoded into the system in addition to applied machine learning algorithms. Rather than rely on a single machine learning technique such as Deep Learning, Smart Learning uses different algorithms, each selected based on its ability to learn and perform a specific task, such as locate a randomly-placed Total Amount field on a purchase order. In all, several dozen pre-trained machine learning algorithms are used for different tasks, each chosen for their precision.

This approach minimizes the ground truth data required to build reliable models and ensures that results are comprehensive and optimized for accuracy. Whereas a general purpose machine learning platform might require 100,000 or more document samples and a significant amount of configuration, Smart Learning only requires a few hundred to a few thousand samples to produce results in hours that are similar to what can be accomplished with a rules-based system and rules-based systems, which take hundreds of hours to configure.

Smart Learning also includes automated performance management. If a machine learning system is allowed to adapt based on dynamic, ongoing data, there is always a risk that the data used will adversely impact performance. However, with Smart Learning, data is collected and analyzed to ensure that it is the proper quality. It then uses this data to adjust, optimize and adapt, comparing results with previous configurations. This function automatically configures output in your system so that data is only ***accepted*** if it is statistically accurate, based upon your needs.



*Turn configuration and tuning into a "**compute time**" problem.*

True machine learning creates efficient and portable models of data.

Process more document types – even ones that have not been seen before.

PARASCRIPT

What are the benefits of Smart Learning?

Intelligent Capture is like no other business solution in that the primary value is in achieving a high rate of *unattended* automation. This type of automation requires a high degree of reliable accuracy along with the ability to apply automation to a variety of document types.

While most other business solutions can provide value through automation of tasks or enabling more efficient processes, few business solutions are so dependent upon automated accuracy. Once in production, results from intelligent capture will often "***drift***" due to changes in production documents including document types, types of data, locations of data and if documents are scanned, image quality.

Unfortunately, the reality of today's intelligent capture solutions is that it takes a lot of time and expertise to create and maintain a reliable and accurate system. This imposes a significant amount of upfront costs both in terms of time required and skilled staff.

Smart Learning enables the delivery and management of highly-tuned systems without the requisite costs associated with initial configuration and ongoing management. There is no need to train a developer to create rules.

With Smart Learning, you only need to provide tagged data and the system does the rest. Smart learning software automatically analyzes each document and page to create models that can reliably optimize image quality, classify and separate documents, and locate and extract needed data. The result is a model that is optimized to your documents and outputs data that meets your stringent accuracy requirements to enable true straight-through processing of data.

Once in production, there is no need to spend time analyzing and correcting rules. Smart Learning software adjusts and tunes itself in the background by automatically collecting and curating production data. What's more, the learned models are completely portable and can be used on other projects with similar document types and requirements.

What are Appropriate Applications for Smart Learning?

While Smart learning takes your tagged samples and creates fully-tuned models, there are some use cases that make sense over others. There are areas where Smart Learning excels. These include document types where there are a significant variety of layouts or formats such as invoices, remittances, purchase orders, explanation of benefits and bills of lading. These types of documents are typically referred to as "***semi-structured.***" The data is often similar, but the labels and placement of this data can vary widely in semi-structured documents. Rather than have developers analyze samples and create rules, Smart Learning intelligently and quickly creates reliable models.

Structured forms can also benefit from smart learning. However, if you have 30 or fewer forms, you can simply use FormXtra.AI's automatic configuration to discover and create the necessary fields. Using your blank forms, it automatically locates the fields, data types and field names, and creates the structured rules for them. For structured forms, strict zone-based configurations are typically more reliable than a machine learning approach—especially for a smaller number of forms—because you can supply explicit rules. If you have many more forms
(some organizations have several hundred to hundreds), then Smart Learning is the obvious choice to automate processing for all of them since you can forego the arduous process of creating rules hundreds of times.

What is Required for Smart Learning?

Data! As discussed, this is specifically ***ground truth data***. For document-oriented machine learning, ground truth data constitutes a representative sample set of documents along with the corresponding field data values for each document or page. Since Smart Learning is an applied machine learning approach for document-based data, you don't need as many samples as would be required from a generic machine learning platform. Parascript scientists have already imbued the platform with knowledge about document-oriented topics and features.

The required format is a comma-delimited file where the columns describe the generic field names and the rows represent the field values found on each page. Smart Learning does not need to know the actual field labels found on each document/page, only the values.

Summary of Capabilities and Use Cases

Smart Learning Capabilities

Smart Image Perfection	Smart Document Classification	Smart Variance Detection	Smart Data Extraction	Smart Measurement
Smart image perfection automatically applies image correction algorithms to ensure good quality.	Smart document classifiers support every document type from text-heavy to documents with pictures.	Smart variance detection enables grouping like documents in order to streamline data location and extraction.	Smart document analysis enables creation of data element location hypotheses to cover a broad range of document types.	Smart measurement analyzes results and refines the extraction model presenting improvement options to admin roles.

Use Cases

Invoices	Remittances	Purchase Orders	Forms	Receipts	Bills of Lading
US and foreign invoices where header-footer data is needed. Line items are supported for single-page invoice variants.	US and foreign remittance advice for both retail and wholesale.	US and foreign purchase orders for header-footer data. Line items are supported for single-page PO variants	Ideal for when the number of structured form forms exceeds 30 which makes initial configuration and maintenance a problem.	US and foreign retail receipts for data including date, vendor, and total amount. Data without labels such as items purchased are not supported.	US and foreign bills where there is a high degree of variance for data such as addressee, account number, address, dates, and phone.

PARASCRIPT

PARASCRIPT®

THANK YOU

Smart Learning Primer | Parascript FormXtra.AI

www.parascript.com

888.225.0169

info@parascript.com