Exchange Solutions

WHITEPAPER

Machine Learning in Retail/E-commerce

Machine Learning Versus Artificial Intelligence

Although Machine Learning (ML) is a sub-field of Artificial Intelligence (AI), they often used interchangeably these days. General Artificial Intelligence aims to create intelligent machines that are able to perform intellectual tasks, which normally require human intelligence. These tasks include learning from few examples, planning, reasoning, understanding languages and abstract concepts, interacting with the world, and generalizing acquired concepts. Today, general AI remains in the science fiction novels and movies. Depending on whom you ask, the answer to when we should expect a general AI ranges from (hopefully) never to a couple of years. On the other hand, Narrow (or Applied) Artificial Intelligence focuses on performing a single task extremely well, which may even exceed human performance. Narrow AI is everywhere today, ranging from speech and image recognition, to the prediction of health outcomes, stock prices, and consumer behavior. Machine Learning is the science of getting computer systems to learn from data without being explicitly programmed. Machine learning is a central pillar of narrow AI that focuses on the learning concepts and algorithms. Other fields include computer vision, robotics, and natural language processing.

What Differentiates Machine Learning from Traditional Programming?

"If programming is automation, then machine learning is automating the process of automation

-- Bill Gates.

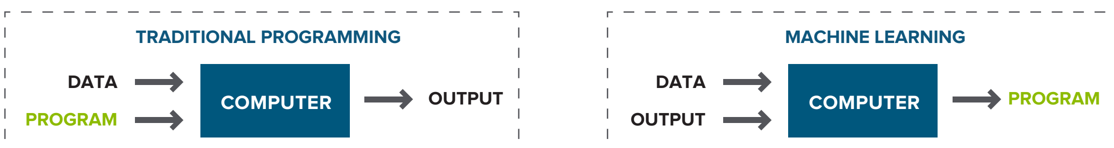

Figure 1. Illustration of Traditional Programming vs. Machine Learning

We essentially write programs to automate tasks, instead of manually repeating the same work. At its most basic level, a program is a set of rules, written in machine-readable instructions, which can be executed by a computer. In traditional programming, we write a program, feed it to the computer together with the data to get the desired output (or target values). By contrast, with machine learning, we feed the data and the desired output to the computer to get our program. Instead of manually designing the rules, we rely on the machine learning algorithm to infer the relevant set rules from the "training" data, and give us a program, or more accurately a "model", that we can then apply to new data to solve similar tasks.

Why Do We Need Machine Learning?

Machine learning is particularly useful for automating tasks where it is difficult to describe the solution, where the desired outputs keep changing over time, or where each example needs a different rule. For instance, it is hard to design a set of rules that allow a program to recognize an object in an image or to drive a car – we don't know how we do it! Detecting fraudulent credit card transactions or predicting stock prices requires combining a large number of dynamically changing rules, which is difficult to maintain manually. While manual creation and manipulation of business rules allows marketers to deliver experiences to specific segments of people, it is not scalable for individual experiences, such as providing individualized content, recommendations and incentives based on a shopper's unique tastes.

Machine Learning Application For E-Commerce

For retailers, machine learning is powering an increasing number of technologies that touch every point of the supply chain. Here are some examples of common machine learning applications for e-commerce.

1. Product recommendation is one of the earliest applications of machine learning in retail. Current recommendation systems can recommend items based on their recency, popularity, profitability, availability, expiration date, etc. They can present similar items in product or category to what a customer has engaged with previously (content-based filtering), or complementary products based on what other customers who share similar tastes have purchased along with those products (collaborative filtering).

2. Search and rank algorithms have become less about listing products that match keywords and more about contextual prediction of what customers might actually want to see at the moment. Recent advances in visual search enabled pixel-by-pixel image search due to deep neural networks. Search by image or by parts of the image will improve the shopping experience by decreasing the time and effort required to find specific items.

3. Anomaly and fraud detection is a popular e-commerce application dealing with fraudulent transactions from stolen credit cards and customers that retract payments via their credit card company after receiving the products.

4. Churn prediction allows companies to predict when customers will stop using a service to buy some time to address the issue.

5. Dynamic Pricing is extensively used in hotel and airline industry, where the prices are continuously updated based on demand, seasonality, supply, competitors' prices, known or unknown visitors, time of day, etc.

Other applications include inventory forecast by predicting market demands, anticipatory shipping (before an order is placed), customized website layouts (to match each user's preferences), and adaptive chat bots and personal shopping assistants.

Types of Learning

There are three main types of machine learning depending on the desired output:

- Supervised learning: Desired outputs are provided with the training data
- Unsupervised learning: Outputs are not provided (unknown or difficult to get)
- Reinforcement learning: No explicit outputs, but positive or negative rewards are provided after a sequence of actions.

In this paper we focus on supervised learning, which is the most mature and most commonly used type of learning in current machine learning applications. When the output is provided for every input in the training data, the learning approach is called supervised learning because the human who defined the desired output can be considered as a supervisor or teacher directing the learning process. The learning algorithm attempts to learn a mapping (or approximate a function) from input to output, and then use this mapping to predict the output given new (unseen) inputs.

Supervised learning distinguishes between regression and classification problems depending on the output type. If the output is a numerical, continuous variable, then the task is called regression, for instance, predicting the order value for an online retailer based on clickstream data (see Figure 2, top). Regression algorithms include linear and nonlinear regression, decision trees, random forest, gradient boosted trees, and deep neural networks.

On the other hand, classification algorithms attempt to predict a categorical variable or class label, such as the type of visitors to a retailer's website (e.g., buyers or browsers) or the probability of each class (e.g., the probability of convergence of buyers versus that of browsers). The classification algorithms can also be categorized according to the number of predicted classes into one-, two-, or multi-class classifiers. One-class classifiers are trained on data belonging to one class only, for example, in fraud and anomaly detection problems this type of classifier is trained on data representing the "expected" behavior, and during operations it considers events with low probability as fraud or anomalies. Two-class classifiers, also called binary classifiers, are designed to discriminate between two classes, after being trained on data from both classes, such as buyers and browsers in retails, spam and ham (legitimate emails), etc. Finally, multi-class classifiers[1] are designed and trained on data with more than two-classes and for each input they output probabilities or scores for each class (highest score wins). For instance, classification of products into different categories based on a text description or an image.

Regression Task: Predicting Order Values

Examples	Features							Target/Output
	H	S	C	P	...	I	ATC	Order Value
S_1	1	1	0	1		1	1	\$250
S_2	1	0	2	1		3	1	\$700
S_3	0	0	0	1		2	2	\$150
S_4					...			...

Classification Task: Distinguishing Buyers from Browsers

Examples	Features							Target/Output
	H	S	C	P	...	I	ATC	Converted
S_1	1	1	0	1		1	1	Yes
S_2	1	0	2	1		3	1	No
S_3	0	0	0	1		2	2	Yes
S_4					...			...

Figure 2. Supervised Learning Approaches Using Simple Feature Vectors, Extracted from Clickstream Data based on the Frequency of Page Views. (H: Home, S: Search, C: Category, P: Product, I: Information, ATC: Add-To-Cart). Each Row is One Session.

Algorithms in this category include logistic regression, support vector machines, naïve Bayes classifier, nearest neighbors, decision trees, random forest, gradient boosted trees, and deep neural networks. Many algorithms may, with a slight variation, be used for both regression and classification problems.

Feature Engineering

The first stage in designing machine learning systems is the collection and selection of representative data to the problem at hand. Raw data must be then transformed into features (also called predictors, attributes, characteristics or variables), which are convenient inputs for the learning algorithms. This step is called feature engineering and it is crucial for building effective machine learning models.

Let us consider the example of an online retailer, where the task is to predict the probability of conversion and the order value based on the clickstream data collected during user interaction with the site. Typically, data scientists consult with business domain experts to design a set of features with a strong predictive power.[2] These features may include general, high level indicators, such as the recency of the visit, number of visits, and average time on site. More specialized features would include information about specific page views, such as the number and frequency of category and product views. Advanced features involve granular details such as frequency, variety, repeated views, time spent per page about product, category, brand, style, information pages, the level of engagement based on mouse and scrolling activities, etc., which could reveal interests and intents.

Standard machine learning algorithms require fixed-size feature vectors of numeric values as inputs for training, as shown in Figure 2. Each element of the feature vector should, ideally, describe the desired output independently from the others.[3] Each element in the feature vector can be thought of as an observation that is independent from the order and time of its occurrence. In fact, training data and feature vectors are also called training set and feature set to hint that the order in which the elements are observed is not important. These algorithms are therefore not well-suited for learning from sequences of events or observations. In sequential data, the order of the elements in the sequences is important and may convey temporal and contextual information. Sequential data include, text, video streams, DNA/proteins sequences, and web usage or clickstream events.

[1] One-class classifiers could be used for multi-class problems by training one classifier for each class. Similarly, two-class classifiers, such as Support Vector Machines (SVM), can be used for multi-class problem by using one-versus-one or one-versus-all strategies, which divides the multi-class problem into several two-class classification tasks.

[2] In some applications, such as speech and image recognition, deep learning algorithms can learn these features from the training data.

[3] Many statistical-based learning algorithms, such linear and logistic regression, rely on the independence assumption among the features, which are also called independent variables while the desired output is called the dependent variable (to emphasize relationship to the independent variables).

Sequential machine learning algorithms are designed to capture the information embedded in the input data streams. For instance, Markov models, a probabilistic model for categorical sequences, can capture the dependency of the current event on the previous events in terms of a compact matrix of transition probabilities. For instance, Figure 3 shows a Markov model trained on sequences of clickstream events collected from converted sessions. In contrast with the two-class classifier presented in Figure 2 (bottom), Markov model can be considered as a one-class sequential classifier that models the class of interest (the buyers in our case). Once trained, it can be used to assign probabilities to new clickstream sequences. The lower the probability assigned to a sequence, the less likely it was generated by a user belonging to the buyers class. Similar models could be constructed for the browsers or any other classes.

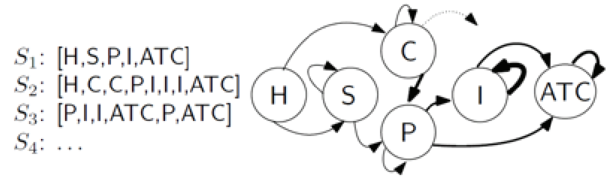

Figure 3. Markov Model Trained On Sequences of Clickstreams Collected from Converted Sessions.

Markov model assumes the future event is independent from the past and only depends on the current state (Markov assumption). Nth-order Markov models extend the dependency of the current event to the previous N events, which gives the model more predictive power but exponentially increases its storage and computational complexity.

Sequential models also allow to predict the next event in a sequence given the previous ones. For instance, the item with highest probability to be added to a shoping card. Other more powerful sequential models include hidden Markov models, conditional random field, recurrent neural network, and long short-term memory (the last two are a family of deep neural network for sequential data).

ES Engage

At Exchange Solutions, the ES Engage platform provides a scalable, efficient, easy-to-integrate, and adaptive solution that combines multiple (diverse and accurate) machine learning models.

ES Engage encourages known buyers to buy more, incents known browsers to buy before leaving the site, and dynamically categorizes new users using advanced in-session features and combination of sequential and non-sequential models.

Our solution is margin-aware: we accurately predict the probability of conversion and the order value and combine it with the cart content and margin to present a highly optimized offer for each customer.

Email us at marketing@exchangesolutions.com and we can share with you how Machine Learning through ES Engage can work for you.

Contact Us

Phone
416.646.7000

LinkedIn
linkedin.com/company/exchange-solutions

Email
marketing@exchangesolutions.com

www.exchangesolutions.com