I'm not robot

reCAPTCHA

Continue

Machine learning taxonomy

Es • es [secure site, there is no need to create an account] Go to the main content download PDF Summary: Despite your great success, Machine learning can have its limits when dealing with data from Insufficient training. A potential solution is the additional integration of prevailing knowledge to the formation process that leads to the information of informed machine learning. In this article, we present a general structured vision of various approaches in this domain. We provide a definition and propose a concept of informed machine learning that illustrates your building blocks and distinguishing conventional machine learning. We have introduced a taxonomy that serves as a classification framework for informed Machine Learning approaches. He considers the source of knowledge, his representation and his integration in the gas pipeline of the machine. Based on this taxonomy, we examine related research and describe how different representations of knowledge, such as alternative equations, rules of logic, or results of the simulation can be used in learning systems. This evaluation of numerous articles on the basis of our taxonomy discovers key all key in the domain of informed Machine Learning. From: Laura von Rueden [see email] [v1] Fri, March 29, 2019 08:37:40 UTC (852 KB) [V2] WED, February 12, 2020 21:10:43 UTC (187 KB) [V3] CRI, May 28, 2021 07:34:41 UTC (676 KB) Introduction in the previous article of this series on Machine Learning, I tried clearly my understanding of learning of Machine. I argued that I understand Machine Learning as a set of algorithms and all to automatically determine the parameters of mathematical models and automatically choose the best template suitable for a particular problem. In this article I will briefly try to present my personal taxonomy (or categorization) of techniques of Machine Learning. Fortunately, this will also give us a general vision of the uses of machine learning. Click on the right control icon table to have a hierachous tadoxls vision at any time. I also suggest to open the internal connections on new tabs so you do not lose your reading position in the article. Taxonomy by data set requirements According to the available data and the purpose of the software package we are developing, we will train our models differently. In the late Sections, we will discuss supervised learning where exemption (expected results) are given, not supervised learning where no example example is given, and learning reinforcement where learning occurs through the intereach With an environment. Supervised Learning A supervised learning algorithm requires the presence, in the formation phase, the correct expected exit for each element of the initial data set. We say that the entries are labeled. Supervised learning illustrations. The algorithm can learn from the labeled images of dogs and cats. The image in the lower right corner is the new entry to be labeled. For example, Let¡ly assume that my dataset is my e-mail box. Leta s assume that I use to manually sort my emails as a spamÂ ¢ or a non-spam ¢ no. Now lettés suppose I want to create a program to automatically filter my emails for me. In this scenario, I have examples to feed my learning algorithm with: manually classified emails. So, I have not only the entries in my dataset (emails), but also the corresponding outputs (if they are in the spam or no) folder. An algorithm that is based on such a data set one is called supervised. It requires entry values and their corresponding outputs. From these examples, it will determine the best forecasting function to estimate the outputs for new A supervised learning analogy is often quoted as the situation where we show an object for a child and we say the name of that object so that the child learns to name the object. Learning not supervised to overseeing supervised learning, a non-supervised ethaniery does not require example exemption (or labels). Learning illustrating not supervised. Colorful areas show Unlabelled Unlabelled grouping of dogs and cats. Algorithm groups together images with similar features. For example, let's assume that I have a database that contains the customer sales historic in an online store along a dance. Let's assume that I want to create a system of recommendation for this store. My goal is, given a product a customer is currently viewing, suggesting another product that has the max chance to be purchased along with what is being viewed. For example, intuitively, I can suggest baby diapers if a customer is visualizing a bottle of baby. I want my learning algorithm to automatically find this type of association in my dataset. In this example, I do not have any correct skirts to train my algorithm with. I just had the inputs, this is, the transactions in the database. Below we discussed the learning of association rules, a non-supervised learning technical configuration designed to extract this type of relationship from a set of transaction data. Semi-supervised learning I see semi-supervised learning as a prolonged non-supervised learning by supervised learning. Semi-supervised learning is usually used when we have a small amount of labeled entries and a much larger amount of non-marked entries. Illustrating semi-supervised learning. A non-supervised phase group illuminated images of dogs and cats (colored areas). A supervised phase, using the exclusive labeled image in the upper left corner, can then label all cats as "cat", and all dogs like "no cat". For example, let's reformulate our email filtering example, so I have only 202020 emails more than 2000020 00020000 that are classified in 222 folders like spam or spam. Most supervised learning algorithms trained only 202020 emails will malfunction with the much larger set of emails not assorted. A solution can be to run a non-supervised algorithm that groups similar emails (222 groups in our case). This grouping algorithm is applied in the complete data set (20000 + 2020000 + 2020000 + 20 emails). Then we assign any email that your next neighbor's paste is no longer older (according to some medical) among the 202020 classified emails that resides in the same group. More clearly, denote an e-mail by a number ranging from 111 to 200202002020020, where 111 to 202020 denote classified emails. If for example, email 222 (which is between the legislatized e-mails 202020), it has the slightest distance of any e-mail denotado III and the e-mail 222 is a spam too, so we will consider an email III To be a spam too. In this example, we actually use two non-supervised algorithms: the first to create the groups is a grouping algorithm and another to find the nearest neighbors is a classification algorithm, 2 taxions that we discussed some sero Later later. I will mention that some authors have the reverse point of view - seeing semi-supervised learning as a prolonged supervised learning by a non-supervised learning - where non-supervised learning can be seen as a preprocessing. Finally, we will observe that semi-supervised learning is often used in configurations where data collection is cheap, but labeling is expensive (usually time consuming). Reinforcement learning does not require an initial set of data, but an environment (or context). The data are collected by the learning algorithm, exploring the environment. The algorithm tries to optimize a decision process toward a goal. Illustrating reinforcement learning. The algorithm explores the environment and learns to reach its goal (reached in Figure 6). Problems for learning They include optimizing the behavior of autonomous agents, decision making and investments in the market of actions, route optimization, game strategy, business strategies and more. Taxonomy by mathematic representation Some techniques of Machine Learning simply compare new instances for those that have been trained. Other techniques try to find a general representation of relationships in the data set. The first is called an instance based and the last model based. based. Learning in order to make estimates, a predictor trained by an instance-based learning algorithm compares new data for training examples. Instance-based learning illustrator. The red dot is the "new" entry to sort. Your nearby neighbors are in red circle. It will be classified as "star" because this rigger is the most common among its nearest neighbors. Generally a medical distance or sometimes a measure of similarity is needed. Distance Methods include Euclidean distance, hamming distance, lightweight distance (used for text comparison), etc. Strategies to estimate property of interest for new data include voting and mother. Voting means that the algorithman estimation will be equal to the most common value between the kkk training examples that are the closest of the new data (according to the media chosen distance). METHER means giving as our estimate of the marital value of interest of these neighbors nearest kkk. In case of the mother's mother, if k = 1k = 1k = 1 then we just give the value of the neighbor closer as our estimate. In our example of an email filtering Semisupervised, the non-supervised phase where they are classified at 200002000020000 emails not marked by one only comparing them for the 202020 marked emails is an example of An instance-based ethics. Model-based learning illustrator. The algorithm chooses a hypothesis, a mathematical representation. It then determines the parameters of this hypothesis based on the available data. Our discussion in getting a Machine Learning Intuition gave us (I hope) a good notice than learning a model is. Taxonomy of training data provisioning scheme intuitively, we feel that learning techniques should not have a memory of the entire set of data that were trained onwards. However, the iterative adjustments that the learning technique makes on the mathematic representation is to determine if they are based on the data it is equipped with. For these antagonistic reasons, we can intuitively understand that many learning techniques will not be able to adjust in new data a trained representation, maintaining it consistent with their previous formation (Because there is no previous data memorial). Learning Techniques That require the entire dataset for your formation are called all-learning batch. Some learning algorithm can actually adjusts a trained representation for new data. They are called all online learning. Learning lot in Lotting Lot, all examples should be provided during the traning phase. The resulting formation is then used in production and no further learning occurs. In this scenario, we obtain new examples, we need to train a new template from scratch in the complete set of enriched data. For a large set of data that can not fit into the memory a computer, there are some techniques to deal with batch learning. They are called out-of-core. Some of them address the issue of the data access management and, for example, allowing the load in the memory only the part of the data required for a TTT time calculation. They usually turn the set of data into a new representation that facilitates these optimizations. Other MÂ © All Lot Variants Learn to accept the data of small pieces. mini-batch techniques still require the entire dataset in time of But the consume data for small subsets. The handsets consume an element at a time and choose this element according to a probability distribution. Online learning learning batch to contrary, an online learning technique can be provided with new examples of training progressively and changes your accordinly representations, even though you are standing in production. An online learning algorithm takes the best current predictor as an additional entry. For many underlying representations, true online learning is not possible. However, depending on the formulation, we can often find a pseudo-online algorithm based on recursive adjustments. In this case, the new predictor depends on the best predictor current and all previous examples (already learned). In cases of luck, the dependence of the previous examples can be expressed using an aggregation function. In this case, we do not store these data elements, but rather the aggregation function and information required for your computation (including your previous result). When none such trick is as possible, no warranty can be given on the storage complexity or memorine of recursive pseudo-online algorithms. Taxonomy By use or purpose Common tasks of machine learning include: finding the value of a property of a phenomenon that depends on other properties of this phenomenon (or other instances). This task is called regression. Finding limits to separate data elements into groups. When groups are taxed from the beginning (they are part of the algorithm entries), the task is called the classification when the groups should be discovered by the algorithm we call task pool. Optimizing decision processes. Rules of mining association (learning of association rules). Disassembling a measurement (usually a temporal season) in its basic independent components. This is the separation of the blind source. Reducing the number of variables - input of a learning problem, a.k.a dimensionality reduction. Regressing regression illustrations. The input data only has 1 feature. Left, a degree 3 polynomial regression and right a linear regression. The regression tries to find the value of a property of a phenomenon, depending on the values of other properties or instances of the same learning. Instance-based learning illustrator. The red dot is the "new" entry to sort. Your nearby neighbors are in red circle. It will be classified as "star" because this ice cream vendor wants to predict your rents based on Tempurature forcasts. We will be learning the parameters (model e) of a regression if we are tried to create a software package for this requirement. The classification classification tries to find limits in the data set in order to separate the elements into several known (or defined) classes before training. Illustration of the classification. Both models illustrated above draw linear limits (data is linearly separable). Créda: [1] Typically classification fall under supervised learning. However, there are non-supervised classification, such as anomaly detection or outlier detection. For example, if we had a set of data labeled data consisting of photos of handwritten dips ranging from 000 to 999, training an algorithm to recognize dips in new photos would be a classification task where classes are 1,2 ã, Â €¦, 91, 2, \ Points, 91.2 ã¡, 9. In addition, our e-mail filtering example is a task of classification With two classes: spam or no spam. I see the classification as a different beast from the regression, but some authors present them as two similar tasks. For these authors, the only conceptual difference is that the departure of the regression is contained as the exit of the classification is discreet. I do not share this vision, because as stated above, I see the regression as trying to estimate a value while the classification attempts to estimate a limit. Clustering clustering tries to group observations so that the elements belonging to the same group (or cluster) are more similar - according to some similarity measure - and theses belonging to different groups are more different. Clustering illustrations. Plot represents the color clusters. Crédo: [2] Grouping It is a non-supervised learning task. We present grouping while we talk about semisupervised learning through our modified e-mail filtering example. In real life, grouping helps bioplogs create their taxonomy, marketing marketing professionals Finding groups of customers sharing similar behavior, search engines in pageable web categorization | Decision making decision making attempts to find what decision to take in order to optimize a gain, use of a resource or an evolution toward a goal. Illustrating reinforcement learning. To explore the medium, the algorithm to know which decisions to take in order to achieve their goal. Otiethime decision making can be learned by using any supervised learning, semisupervised or reinforcement. He is, however, associated with reinforcing learning. Machine learning for decision making is a good fit for dealing in the actions markets, the construction of autonomous agents (such as Rovers Planeta Mars), decision-making In politics, insurance and many more domans where strategy is important. Rule Association Learn association attempts of learning rule to determine as associated items actors. Illustrations of association rules. The rules on this example represent the product associations often present in customer carts in a store. Customers who buy milk and diapers often buy feet :) The fundamental conceptual difference with clustering, in my opinion, is that grouping grouping is based on the properties of the elements, while association rules It is based on like other agents or elements of the group actors together in some transactions. Association rules is usually based on techniques of learning without supervision. We gave an example of association rules when we discussed learning not supervised above. Another use of algorithms resulting from association rules include sorting books in a library, product placement in a store, composing service packages (for insurance, for example) Â ¢ | An analogy for association rules is like it to decipher an ancient tongue. Knowing the common thong structures, their relationships with each other (ascendancy, similarity Â ¢ |), other statumical properties and some historical facts, we learned the association rules used by the old users of the Language for the scripts that left to make sense for us. Blind separation source attempts to separate blind source to decompose a measurement in its basic components. Usually the measurement is a time time, as a sound disc. Blind source separation illustration. The signals are properly recovered (image below), but the colors are different. This illustrates the "blind" aspect of the algorithm, which in fact knowns nothing of the initial signs (middle image). Crédo: [3] Source of blind separation is usually a task without supervision, but can be supervised too. A frequently quoted example of blind source separation is highlighting the discourses of different people from a recording where everyone speaks at the same time. Leta s imagine that we are at a party and get to the kitchen, where there are 5 people talking at the same time. We have a record of this cacophony and usually want to separate each personan speech to a separate track. This would be a task for blind source separation. Here means Blinda â ¢ that the algorithm has no information on the sources of the signals or about the mixing process. An analogy of the blind source separation is like us, human beings, in a multitude with several people talking, are capable of isolating someone else's voice in particular that we want to hear. Dimensionality reduction Attempts to reduce dimensionality to reduce the number of variables â ¢ â € ® input of a learning problem while maintaining the most relevant information. Dimensionality reduction illustrations. Each image represents a handwritten dip. Both selection of resources and extraction gave these pixellate gray shades images. The classification algorithm adds to work very well in the dataset Criterion: [4] Dimensionality reduction exit can be used as input of other learning algorithms or for viewing (if the number of entries can be reduced to 333 or less). When used as input of other learning algorithms, it can improve yours (Reduce your error generalizaÃ§Ã the £) Â Because less variÃ¡veis learning algorithms usually means less complex, less noise and less ¢ INSTANCE redounds in the data. The variÃ¡veis â â entry of learning algorithms sÂ £ o call features. Typically, dimensionality reduction consists of 3 resource reduction types: Resource selection tries to find a subset of resources that are most relevant to a particular problem. The extraction of resources is often used in conjunction with the selection of resources and tries to infer new features of the existing ones. These new features will be more informative and less redundant than the initials and will also describe with the initial data set. The projection of resources reduces the dimension of the entries, designing them in a smaller dimensional Â ¢ â â € œe GeometryÂ ¢ â € ™1. The projection of resources is useful when we assume that the inputs are described by a smaller dimensional "geometry", but doubled in a higher dimensional space. For example, in the figure below, the shaped entrances on the left make a 2D surface â € "evaluated â € in a 3D space. The projection illustration of Resource. Various projections on a plane of a 2D surface folded in a 3D space. said: [5] We must note that a major challenge in the learning of the machine is, sometimes finds or choose the representation of our resources. For example, in the example of email filtering, how can we represent a numerically such that it can be a relevant entry for a learning algorithm? A method can be the : 1) Create an Â ¢ â € "Addituation Is the indictment of the e-mail dictionary. Sword. Another option maybe every component of the vector is the number of occurrences in the email of each word of dictionary. Another representation is Possible. A Basic Illustrator Dimensionality reduction can be seen in our example of section dip recognition on the above classification. Maybe the need £ color My Information to recognize digits, so poderÃamos give 222 color components and reduce the dimensionality of our data 333 (channels R, G, B) to 111 (grayscale value) . On the same idea, we can reduce the size of the image and compress it to further reduce your dimensionality. Let's consider another example: a program that says if a photo shows a face or not. Us human, we only need some lines of drawings in order to recognize a face, so we can intuitively think that our algorithm also needs only a small amount of information from a photo to say if it is a face or not. In fact, there are neural networks capable of extracting these features from an image and allow other algorithms to rate the image as face or do not face. CONCLUSION There were many information in this article, then we will recapitulate. There are different axes of taxonomy of makers learning algorithms. We can differentiate techniques of machine learning based on the type of data they need to learn. In this taxonomy, to distinguish supervised learning, semi-supervised, do £ from the supervised and reforÃ§o. Another axis of taxonomy is the mathematic representation that they trust. From this axis, we distinguish techniques of machine learning based on models and instances based. An alternative of differentiation of machine learning algorithms is pursuing, consume data for training. In this axis, we have all of learning in batch and online. Finally, we can base our taxonomy about the type of task that the algorithm resulting from the training tries to reach. These tasks sÂ £ notably the return £ o, £ the binding, cluster, DECISION making the £, rule learning Association £ o, £ Separation of the blind origin and £ reduÃ§Ã the dimensionality . These axes of They are not exclusive. For example, a supervised learning algorithm can be an instance-based learning technique. In addition, the taxonomy presented in this article is my personal attempt to link various qualifications of Machine Learning And it is not exhausting. Taxonomy is useful when we try to frame a problem and try to determine what apprenticeship to use for this specific problem. Knowing these taxonomies can guide us by finding a reduced subset of candidate learning algorithms between which a validation technique (for example) can choose. At the next article, discussing linear regression, we will make a big step to actually apply the learning of the machine in practice. Written in Sun January 13, 2019, 12:00 GMT + 01: 00. Last update in Sun Mar 10, 2019, 12:18 GMT + 01: 00. GMT + 01: 00.