Thuy Pham
U41243664
EC601 Mini Project 2

I. Introduction to the topic

Unsupervised learning is a branch of machine learning wherein algorithms/models are created to infer patterns from a dataset. Unlike with supervised learning, there is no reference to labeled/known outcomes. Instead, unsupervised learning enables discovery of the underlying structure of data. A few examples of applications of unsupervised learning include: clustering and anomaly detection. Clustering aims to split the dataset into groups, wherein data points are more similar within the group than with points of another group, while anomaly detection aims to detect unusual points in the dataset.

Machines ethics is an emerging field in philosophy concerned with the moral behavior of artificially intelligent beings. As the field of machine learning and artificial intelligence continues to advance at a quick pace, advances in this branch of philosophy grow evermore necessary. Machine ethics is considered briefly in this paper.

II. Summary of references

i. Combining unsupervised and supervised learning in credit card fraud detection

The traditional approach to credit card fraud detection relies on supervised learning algorithms, which assume that fraudulent activity can be detected/predicted based on past transactions. The labels for each transaction (as genuine or fraudulent) tend to be known a posteriori. This research paper proposes that supervised learning is not a robust enough model for fraud detection – a combination/hybrid of supervised and unsupervised learning can improve credit card fraud detection, Researchers applied a "best-of-both-worlds" sequential approach, wherein they used unsupervised learning to augment the dataset, and then fed the augmented dataset into a supervised learning model.

The primary application of unsupervised learning considered in this research paper is outlier detection. Unsupervised learning aims to discover the distribution of the transactions, then outliers of the distribution are considered fraudulent activity. The researchers computed outlier scores at different levels of granularity (i.e. different contexts), then integrated the scores as features into a supervised learning model, and finally assessed the added value of including these scores. Metrics were computed to compare the accuracy of three datasets: 1) the original data, 2) the augmented data (original with the outlier scores), and 3) the outlier scores alone. The Random Forest model was used as a baseline for comparison due to its superiority in credit card fraud detection.

The research paper concludes that augmenting the original dataset with all calculated outlier scores did not benefit fraud detection, and in some cases, even reduced accuracy, across the three different contexts considered (global, local, and cluster). However, they found that the outlier scores may play an important role in risk prediction, as they rank only below two features (shop risk and last shop risk) in analysis of significance of features for fraud detection. In addition, one approach considered – wherein they augmented the dataset with only one outlier score – improved fraud detection accuracy.

ii. Auto insurance fraud detection using unsupervised spectral ranking for anomaly

This research paper proposes using unsupervised spectral ranking for anomaly (SRA) to detect auto-insurance fraud. The researchers consider ranking advantageous in cost and benefit evaluation analysis, especially for this particular application, where manually determining whether a claim is fraudulent or not requires a large amount of resources. Unlike in the previous case, obtaining clearly fraudulent or non-fraudulent labels is costly (and can be impossible), so unsupervised learning approaches are favorable compared to supervised ones.

The approach taken in this research paper is to assess a ranking for auto-insurance fraud claims by detecting anomaly of interdependence relation among features. The interdependence relation is captured by kernel similarity for synthetic data and similarity measures (outlined in more detail in the paper) for categorical data, as in the case with auto-insurance fraud detection. The researchers use a supervised random forest tree model as an upper bound for performance.

The paper concludes that SRA significantly surpasses existing unsupervised outlier-based detection methods for this particular application. Additional validation for the SRA approach is demonstrated by comparing the most important features detected by this approach with those discovered from training the supervised random forest – the top three were identical (base policy, car types, and fault). The researchers also conclude that one can view SRA as a relaxation of the unsupervised support vector machine (SVM) problem.

iii. Machine Ethics: Creating an Ethical Intelligent Agent

Researchers argue for the importance of machine ethics and favor an explicit ethical agent (i.e. machine can calculate best action by knowing ethical principles and applying them) over an implicit one (i.e. machine is programmed to be ethical). With the continuing advancement of AI technology, there is a growing concern that machines may behave unethically and cause detriment to humans. Machine ethics must be fostered in order to address this concern and prevent the alternative response of stifling AI research. The researchers state that the problem is difficult to address because ethics hasn't even been codified in the field of philosophy. In addition, there are other unique challenges to consider in the case of machine ethics, e.g. can ethics be computed?

The researchers recognize that this is an inherently interdisciplinary field of research: AI researchers will need to interface with theoretical ethicists (and vice versa). They note that this intersection can advance ideas for theoretical ethicists, as it allows them a way to test the ethical principles/approaches proposed. They also apply philosophical ideas surrounding ethics in discussing whether machines should be considered beings that can behave ethically and argue for machines that are explicit ethical agents for intuitive reasons. This research paper concludes with an outline of six steps for building a machine that is an explicit ethical agent.

III. Analysis of results

Pros

- Unsupervised learning does not require labeled datasets,
    - which can be costly or impossible to obtain, i.e. in the case of auto-insurance fraud or
    - can only be known a posteriori, i.e. in the case of credit card fraud.
- Unsupervised learning can adapt to changes in customer behavior as well as novel fraud patterns.
- Unsupervised learning can supplement existing supervised learning models to improve overall accuracy in the case of outlier detection.
- Certain unsupervised learning techniques may perform better than others for a particular application.

Cons

- Accuracy of results can vary depending on what metrics are used to evaluate the inclusion/performance of the unsupervised learning (aspect of the) model.
- Clustering has inherent technical limitations, primarily in the form of choosing hyper-parameters:
    - for example, determining the correct amount of clusters *k* for the k-means algorithm.
    - selecting appropriate contextual attributes
- Depending on implementation choice for the unsupervised learning application, you may have to filter the data, which can result in missed detection of a pattern of fraud, e.g. only considering accounts with a minimum number of transactions.
- There are no learning targets available to guide the learning process for unsupervised learning.
- It can be difficult to identify relevant features for unsupervised learning.
- In general, you will need more data to accurately detect the "true" underlying structure of the dataset.

IV. Conclusion with recommendations

In conclusion, unsupervised learning can be a good approach depending on the type of problem one is trying to solve. It is especially pertinent in the case of fraud detection, as fraudulent activity is modeled in the form of outlier/anomaly detection, a pattern that can be inferred from an unlabeled dataset. As with any choice of model, unsupervised learning approaches have both pros and cons. Those interested in developing such systems should be aware of the extensive experimenting required to tune parameters and the difficulty of identifying relevant features for the model. Those desiring to make use of such systems should be aware of the variability and tradeoff in accuracy with other traditional approaches (i.e. supervised learning models).

In general, for those interested in the pursuit of ML and advancing AI, I would recommend considering the ethical ramifications of the tools being built. In some cases, the potential for negative impact is not obvious and the ever-growing desire for continued advancement and convenience can cloud it further. It is important consider these risks and

complications beforehand in order to actively address (and prevent) them. For example, as a society, we are only now becoming aware of and attempting to retroactively resolve concerns with and issues resulting from Facebook selling our data without our explicit consent. Consider the models IBM uses to automate filtering of their job applicants. It has only recently realized that these models are inherently biased due to the data containing bias that exists systemically in our society. I am of the opinion that we can improve our own morality by anticipating these ethical concerns as they relate to ML.

V. References

1. Anderson, Michael and Anderson, Susan Leigh. "Machine Ethics: Creating an Ethical Intelligent Agent." *AI Magazine,* Volume 28 Number 4, 20017, 15-26.

2. Carcillo, Frabrizio et al. "Combining unsupervised and supervised learning in credit card fraud detection." *Information Sciences*, May 22, 2019, 10-27

3. Nian, Ke et al. "Auto insurance fraud detection using unsupervised spectral ranking for anomaly." *Journal of Finance and Data Science*, 2, 2016, 58-75

VI. Learnings from Peers

i. Bowen Qin

Topics Covered in Report

Tensorflow is an end-to-end open source platform for machine learning. Bowen focuses her report on the use of CNN for image detection and segmentation. With a simple CNN, users can achieve a test accuracy of over 70% on the CIFAR10 dataset. Similarly, users can easily build a CNN to classify images into different classes and to segment images into lower-level information, i.e. pixel-wise masks.

New Things Learned

- Tensorflow only became open source on November 9th, 2015
- Tensorflow is developed and maintained by Google Brain
- CIFAR10 dataset of labeled images with mutually exclusive classes
- Oxford-IIIT Pet Dataset consists of images of pets, labels, and pixel-wise masks
- U-Net model consists of an encoder and decoder

ii. Mengting Song

Topics Covered in Report

Mengting presents a logical "bottom-up" approach to how the problem of object segmentation is being addressed. First, you must consider image classification, wherein an image of a singular object is attached to a label identifying said object. Then, the next tier up addresses object detection, i.e. where in the image is a specific object? The results tend to be provided in bounding box annotations for a particular object (or multiple objects). Then, finally, one can consider object segmentation, wherein an exact outline of the object (or multiple objects) is the end-goal.

Similarly, the state-of-the-art methods to address object detection build on top of each other. The approach taken by traditional CNNs require that images be broken down into smaller (and smaller) regions, until the regions are able to be labeled with a single class (image classification). R-CNNs aimed to improve upon the CNN approach by first determining regions of interest in an image using some proposal method. Fast R-CNNs reduce the amount of time taken (compared to R-CNNs) to detect objects by applying a pooling layer to the regions of interest. Fast R-CNNs then evolved into Faster R-CNNs.

Finally, the report examines YOLO (You Only Look Once), a specific model, in some detail. YOLO performs a single evaluation on the entire image to predict bounding boxes and class probabilities for each box. YOLO understands generalized object representation and is extremely fast. Essentially, YOLO takes the input image and segments it into some n x n grid. For each "square" in the grid, image classification and localization are applied. The result is that YOLO will output a label vector that includes whether an object is detected (yes/no), bounding box annotations, and probabilities for specific classes, for example.

New Things Learned

- Traditional CNN approach to object detection was abandoned in favor of Region-based-CNNs (R-CNNs) due to excessively large computational costs
- R-CNNs are comprised of three components: 1) CNN for feature extraction, 2) linear SVM classifier for identifying objects, and 3) regression model for tightening bounding boxes
- Selective search is used to identify regions of interest in an image; it effectively uses a sliding window to search the entire image
- Step-by-step understanding of how YOLO detects objects in an image

iii. Ningrong Chen

Topics Covered in Report

Object detection is a branch of computer vision for detecting objects of a certain class in images or videos. Ningrong lists a series of approaches aimed at solving the problem of object detection, including: R-CNNs, Fast R-CNNs, Faster R-CNNs, R-FCN, YOLO, SSD, FPN, Mask R-CNN, Cascade R-CNN, RetinaNet, Sniper, PANet, TridentNet, and CBNet. Brief comparisons are made between each, touching upon improving accuracy and speed.

The report also touches on Model Zoos. Model Zoos are a collection of pre-trained models on different datasets using various algorithms. Three Model Zoos are mentioned: Tensorflow Detection Model Zoo, SimpleDet Model Zoo, and Detectron Model Zoo. Learning cost is high for the model zoos with larger models; the ones mentioned are aimed at object detection.

New Things Learned

- Object Detection has multiple sub-branches, including 3D Object Detection, Real-time Object Detection, and Salient Object Detection
- Faster R-CNN is a combination of Fast R-CNN and RPN (Region Proposal Network)
- RPN simultaneously predicts object bounds and objectness scores at each position
- YOLO predicts bounding boxes and class probabilities directly from full images in one evaluation instead of repurposing classifiers to perform detection

iv. Shubham Aroras

Topics Covered in Report

The report lists some standard tasks in computer vision that include various types of recognition (image classification, image captioning, segmentation, object detection, etc.) and motion analysis (tracking, optical flow). It briefly introduces a few key models in the history of image classification: LeNet-5, AlexNet, and ResNet. LeNet-5 (a CNN) was one of the first neural networks that utilized back propagation. CNN architecture are optimized for images (it significantly reduces the number of parameters that need to be learned). AlexNet was able to build upon LeNet-5 and improve image classification capabilities to a large scale due the availability of exponentially more data and parallel computing using GPUs. ResNet set a new standard by outperforming traditional, plain neural networks.

New Things Learned

- LeNet-5 was one of the first NN that utilized back propagation
- AlexNet skips the step of feature extraction
- ResNet utilized the idea of skip connections across non-sequential layers