Survey of Transfer Learning with Relation to Network Traffic

Byron Barkhuizen

1 Context

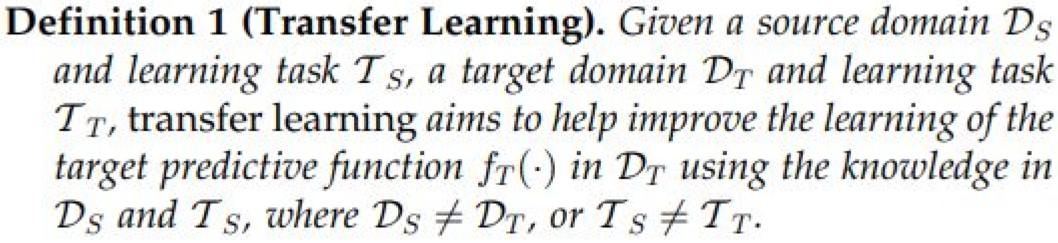
Transfer learning is a research field in machine learning that seeks to ‘learn to learn’ in the same way that humans might approach a new task having some experience in a related one. The advantages of an algorithm that can mimic this ability are much lower computation requirements (time, power), lower complexity of approaching new problems, less feature engineering, and possibly higher accuracy depending on how well the knowledge transfers.

2 Purpose

For the task of anomaly detection within medicine, cybersecurity or less obvious applications in vehicle traffic analysis where there are opportunities for the useful understanding of commonalities between different datasets transfer learning becomes important. High accuracy in traditional machine learning approaches can be achieved with large sets of labelled data and subject expertise, however this is rarely the case for lesser known domains. Furthermore, stale datasets or rapidly evolving anomalies can render these models useless, requiring their re-definition. Traditional deep learning approaches require retraining and a large quantity of new data, this can take a long time and it is sometimes not possible to meet these requirements in a reasonable time.

3 A Formal Definition of Transfer Learning

Transfer learning is the establishing of a relationship between a source domain and a target domain, along with source tasks and target tasks. The existence or assumption that there is some similarity between the source and target is the theoretical basis for the application of transfer learning (1). The features within a source or target domain have some probability distribution. When these features distributions are the same then the new dataset can essentially just be classified in the same way that the original dataset was. If the source and target dataset do not conform to the same probability distribution then this cannot be done accurately. However this does not mean that a model needs to be completely re-trained and re-classified on the new dataset, there theoretically can exist some knowledge that can be transferred between the two datasets. A mathematical definition for transfer learning is shown in the following image.



Definition 1 (Transfer Learning). *Given a source domain $\mathcal{D}_S$ and learning task $\mathcal{T}_S$, a target domain $\mathcal{D}_T$ and learning task $\mathcal{T}_T$,* transfer learning *aims to help improve the learning of the target predictive function $f_T(\cdot)$ in $\mathcal{D}_T$ using the knowledge in $\mathcal{D}_S$ and $\mathcal{T}_S$, where $\mathcal{D}_S \neq \mathcal{D}_T$, or $\mathcal{T}_S \neq \mathcal{T}_T$.*

Some assumptions for the applicability of transfer learning are:

1) Various features (shapes, edges, colours) or knowledge are general and domain agnostic
2) Domains or tasks between source and target are related enough so as not to cause negative transfer, in which case performance/accuracy would be decreased
3) Low quantity of labelled data available, or no labelled data at all in target domain
4) Source knowledge is labeled and rich

Transfer learning seeks to loosen ‘same distribution’ requirements through the application of transfer learning algorithms. These algorithms will seek to expand upon their similarities, finding the shared features according to some similarity conditions, typically in higher-dimension feature space. Once the shared features are established then knowledge transfer has occurred and there exists some information with the same distribution, allowing for meaningful continued classification algorithms to be used. The knowledge of the source data has been imposed upon the target data in a useful way.

4 Applying Transfer Learning

When we decide that we want to apply transfer learning we first have to ask a few questions. We need to know ‘what’ to transfer, ‘how’ to transfer, and ‘when’ to transfer.

What to transfer tells us which parts of the knowledge from the source domain remain relevant and will be a useful piece of knowledge to keep. We do not want to keep knowledge that is specific to source or target tasks, instead we want to find the commonalities.

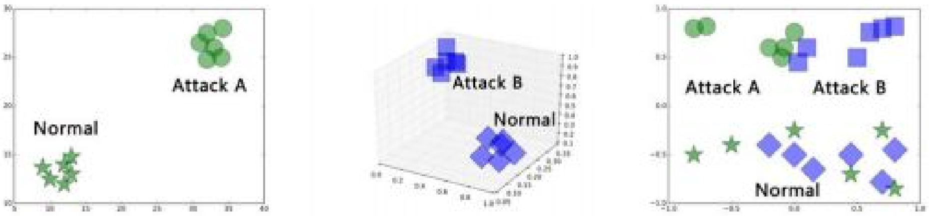
Learning algorithms are then responsible for the ‘how’ to transfer. They operate on the basis that there is some knowledge to be shared and depending on the conditions of the source and target domains and tasks they will find a way to align their knowledge.

4 Current Works

(2) demonstrates the application of unsupervised transfer learning. The utilization of a popular pre-trained model called Inception Resnet-v2 by Google as a feature extractor yielded greatly reduced number of significant features with little to no training required. This demonstrates the use of a pre-trained neural network and transferring this to gain knowledge about the features of a target dataset. This is a very popular transfer learning approach that works by freezing the initial layer weights which are associated with basic features such as edges and gradients and replacing the final fully connected layer which deals with classification. This helps to identify common features and define a new feature map or the target dataset. The knowledge of the source domain is passed to the target domain. There are a few modifications to this approach which concern the extent to which existing layers in the pre-trained model are kept frozen, or they can be allowed to slightly modify weights in a fine-tuning process which is also very popular.

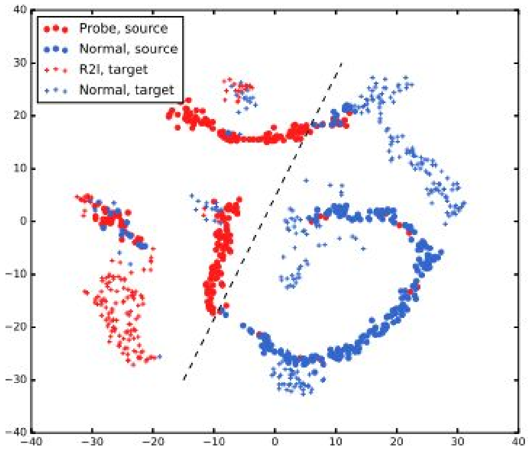
(3) takes a more advanced approach to aligning source and domain datasets by the definition of a feature-based transfer learning algorithm called HeTL. The experiment's aim was to transfer knowledge about known cyber attacks to the introduction of unknown attacks

in order to classify them as malicious or not. A feature extraction process is undertaken through a CNN style approach. The feature spaces of the source and target domain are represented in a common latent space. An optimization function is applied to assess the distances between data of the source and target domains. This optimization function along with a distortion function attempts to keep the structure of the original data as much as possible, while maximizing similarities between the source and target domain by minimizing the difference in the latent space. This is done through a learning process using a gradient approach while observing the changes in the probability distribution of the source and domain data. HeTL performed extremely well in comparison to the situations where no transfer learning was used, and it performed better than other transfer learning algorithms. The two images below are from the paper and demonstrate the transformation of the source and target domain from their individual feature space to a shared latent space where some similar distributions can be observed. HeTL successfully finds a subspace that makes distributions of different attacks similar. The resulting modified feature-based transformations resulted in the second visual representation of attack vs. norma data where there is a clear decision boundary, greatly assisting classification, which would not be otherwise observed without transfer learning.



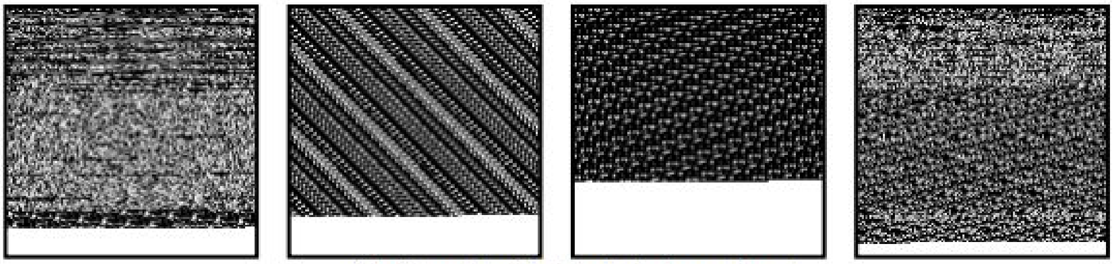
(a) Source domain (b) Target domain (c) Projected space

Fig. 4. Illustration of proposed feature space transformation concept.

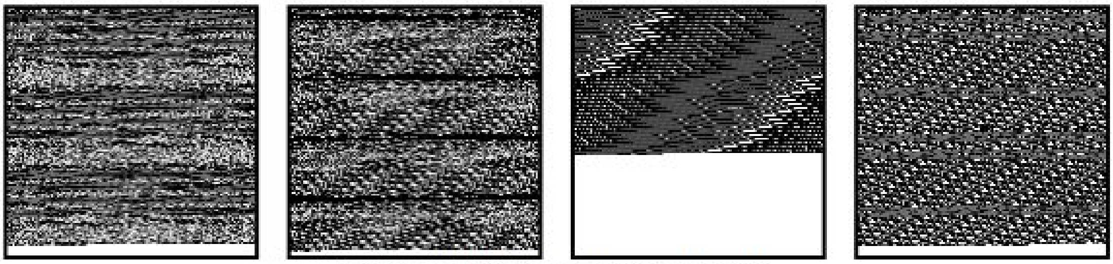


(4) extends the fine-tuning approach to transfer learning. An image representation of network traffic data is used alongside a pre-trained image network and comparisons are made between using no transfer learning and transferring some of the weights. The transfer learning approach performs extremely well compared to the no transfer learning approach, outperforming it by nearly 50%. The images generated for use in the pre-trained network are

shown below. The transfer learning approach is measured at different levels of fine-tuning, and it performs at its best with only a few first layers freezed. When too many layers become frozen the features being explored are source domain specific and it does not perform well, going down to 33.51% accuracy. The testing phase for one packet data is extremely quick and it shows promise of this type of promise being utilized in real time for intrusion detection.



(a) Normal traffic data in the image format.



(b) Botnet traffic data in the image format

(1) is another demonstration of a feature-based transfer learning algorithm. Similar to (3) the approach seeks to find similarities between the source and target domain feature distributions, or it seeks to ‘force’ them. The source and target features are mapped to a high dimensional space in order to assess their similarity by a method such as Euclidean distance. This work extends the feature mapping component of transfer learning by using a convolutional neural network to classify the newly identified features as source or target. This optimizes the parameters of the mapper so that after it is finished the traffic data in the target domain is more similar to the source domain. The output of the network is binary, and features will be classified as being part of the source domain or not. A loss function is defined by the similarity between the two, and they seek to minimize this. When it is minimized it means that the maximum possible similarity between the two domains has been reached. The method showed impactful results, effectively being able to transfer knowledge between the source and target domains in a short time (1 epoch).

5 IoT Identification Recommendation

The transforming of network traffic data to image, as supported by (4) and (5), as an application agnostic approach to network flow analysis and therefore an approach for uniquely and accurately identifying IoT devices has high accuracy and it is something that has an abundance of resources to build upon. Python with Keras is highly performant for computer vision problems and there is a large amount of documentation as well as research papers available on the manipulation of image data for deep learning. The abstraction away from textual information taken from our PCAP files allows a more adaptable approach, better suited for transfer learning.

The issue becomes whether the image representation of the pcap file, similar to (5), can be enriched somehow in order to show more unique information about the network flow from an IoT device. This would allow more accurate IoT device identification. (6) demonstrates the relevance of using other flow level information apart from the packet data. Semantic relationships are created between statistical attributes such as activity cycles, port numbers and cipher suites to find features that best represent unique IoT devices and therefore act as a sort of signature. The issue with this is that many of these features can theoretically be altered by vendors and be unreliable. Incorporating these features also requires a domain expert for feature engineering, and the architecture involved is more complex than some other approaches. Nevertheless it would be useful to find a meaningful way to include some of these important features alongside the image of network traffic flow.

Transfer learning should be implemented between the source and target domains as the representation of flow data as an image allows a better understanding of the features and therefore will improve the feature extraction phase. Once the extracted features are understood then a dataset from a target domain, such as a new environment with new iot devices, the probability distributions of features does not have to be identical. A feature-based approach between the two domain datasets can be used to normalize some key features and align their feature distributions in such a way that it can be easily classified. The knowledge of the known IoT devices can be transferred to a new network environment with some minor adjustments being made through a transfer learning algorithm.

References

(1) Xiong P., Cui B., Cheng Z. (2021) Anomaly Network Traffic Detection Based on Deep Transfer Learning. In: Barolli L., Poniszewska-Maranda A., Park H. (eds) Innovative Mobile and Internet Services in Ubiquitous Computing. IMIS 2020. Advances in Intelligent Systems and Computing, vol 1195. Springer, Cham

(2) P. Krishnakumari, A. Perotti, V. Pinto, O. Cats and H. van Lint, "Understanding Network Traffic States using Transfer Learning," *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, 2018, pp. 1396-1401, doi: 10.1109/ITSC.2018.8569450.

(3) J. Zhao, S. Shetty and J. W. Pan, "Feature-based transfer learning for network security," *MILCOM 2017 - 2017 IEEE Military Communications Conference (MILCOM)*, Baltimore, MD, 2017, pp. 17-22, doi: 10.1109/MILCOM.2017.8170749.

(4) Taheri, S., Salem, M., & Yuan, J. (2018). Leveraging Image Representation of Network Traffic Data and Transfer Learning in Botnet Detection. *Big Data And Cognitive Computing*, *2*(4), 37. doi: 10.3390/bdcc2040037

(5) Kotak, Jaidip & Elovici, Yuval. (2020). IoT Device Identification Using Deep Learning.

(6) Sivanathan, A., Gharakheili, H., Loi, F., Radford, A., Wijenayake, C., Vishwanath, A., & Sivaraman, V. (2019). Classifying IoT Devices in Smart Environments Using Network Traffic Characteristics. *IEEE Transactions On Mobile Computing*, *18*(8), 1745-1759. doi: 10.1109/tmc.2018.2866249