JCSE International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.-7, Issue-6, June 2019 E-ISSN: 2347-2693

Transfer Learning: Approaches and Methodologies

Swati Sucharita Barik [1*], Mamata Garanayak[2] , Sasmita Kumari Nayak[3]

[1,2,3] Dept. of Computer Science and Engineering, Centurion University Of Technology and Management, Bhubaneswar, India

[*]Corresponding Author: swatisucharitabarik@gmail.com, Tel.: +91-99382-07636

DOI: https://doi.org/10.26438/ijcse/v7i6.852855 | Available online at: www.ijcseonline.org

Accepted: 10/Jun/2019, Published: 30/Jun/2019

***Abstract*—** Machine learning and Data Mining Techniques are mainly used for many Real world problems. The traditional methods include training the data and test .But it will not be applicable for real world scenario. Some of the reason may be the cost of training data and inability to get those. These drawbacks are giving rise to the concept known as Transfer Learning.It ensures that training data must be independent and distributed identically.Transfer Learning is considered as a solution to the insufficient training data.

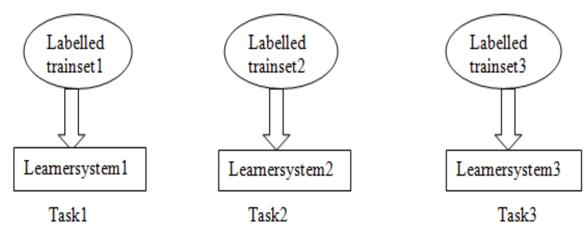
***Keywords*—** *Data Mining, Transfer Learning, Machine Learning.*

I. Introduction

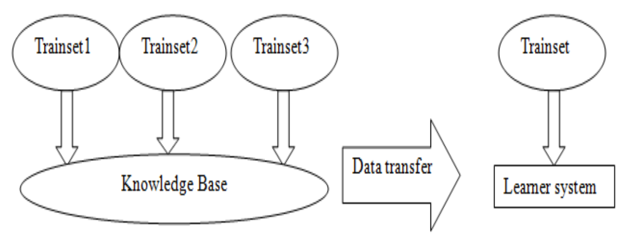
Transfer learning is also known as inductive transfer. It is a major research problem in Machine learning. The concept behind traditional machine learning method is training the data and testing data. Then the data are taken from the same domain, so that the input feature space and data distribution characteristics are same[1][2].It focuses on storing knowledge as well as problem solving. These are applicable to some related problems. Transfer learning is a machine learning method where a model is developed for a task and is reused as the starting point for a model on a second task. It is the knowledge of an already trained Machine Learning model, which is applied to a different but related problem. Traditional machine learning is characterized by training data and testing data with same input feature space and the same data distribution.

Transfer learning and domain adaptation refer to a situation where learning is exploited to improve generalization [3].Transfer learning is an optimization, which allows rapid progress while modelling the second task. Transfer learning is related to problems like multitask learning and concept drift and it is not exclusively an area of study for deep learning. Transfer learning is useful in deep learning with a large amount of resources required to train deep learning models. On this deep learning models are trained. Transfer learning only works in deep learning if the model features learned from the first task are general [4].

Training and testing are two major principles of supervised machine learning method. It can be regarded as training a learner with data and testing them. The Traditional data mining and machine learning algorithms make predictions on the future data using statistical models, trained on previously collected labelled or unlabeled training data. Transfer learning methods appear in data mining, machine learning and applications of machine learning and data mining.



[fig 1.Traditional Learning Method]



[Fig2: Transfer Learning Method]

II. WORKING PRINCIPLE AND APPROACHES TO TRANSFER LEARNING

There are different methods to merge the transfer learning into transfer learning algorithms. Instance transfer approach and parameter transfer approach are among them.

Neural Networks detect edges in their earlier layers, shapes in their middle layer and some task specific features in the later layers [4][5]. With transfer learning, the early and middle layers are used and retrain the later layers. It helps to leverage the labelled data of the task initially trained on.

Following are the approaches to Transfer learning.
1. Training a Model to Reuse: To solve Task A, suppose a user doesn't have enough data to train a Deep Neural Network. so a way is to find a related Task B, where user has an abundance of data. Then it could train a Deep Neural Network on Task B and use this model as starting point to solve initial Task A [5][6].

If user has the same input in both Tasks, then user could just reuse the model and make predictions for new input or could change and retrain different task specific layers and the output layer.

2. Using a PreTrained Model: There are a lot of models. Wehave to find layers to reuse and train.
3. Feature Extraction: Discover the best representation of problem, by finding the most important features. This approach is also known as Representation Learning and can often result in a much better performance obtained with hand designed representation.

In Machine Learning, features are manually handcrafted by researchers and domain experts. Deep Learning can extract features automatically. Neural Networks have the ability to learn which features to put into are important and which aren't. A representation learning algorithm can discover a good combination of features within a very short timeframe, for complex tasks which would need a lot of human effort.

The learned representation can then be used to use the first layers to spot the right representation of features but output of the network cannot be resulted, as it is too task specific. The data can be fed into network and intermediate layer can be used as the output layer. This layer can be interpreted as a representation of the raw data. It is used in Computer Vision as it can reduce the size of dataset, decreases computation time and makes it more suitable for traditional algorithms [7][8].

Examples of Transfer Learning with Deep Learning
Transfer Learning with Image Data**:** It performs transfer learning with predictive modelling problems using image data as input. It may be a prediction task that takes photographs or video data as input. For these types of problems, it is common to use a deep learning model pretrained for a large and challenging image classification task such as the ImageNet 1000-class photograph classification competition [9].

III.PROPOSED METHODS AND ALGORITHMS

Here we will discuss about the different methods and the use of the algorithm.
1.SVM based transfer learning method: In Least Square Support Vector Machine (LS-SVM), the parameters are found by solving the equation with criterion error (ERR), generated from leave one out error [10].
The Proposed algorithms are as follows:
i)Weighted Error Rate(WERR): It is extended from the criterion error (ERR) by introducing the weighting factor related to the number of positive and negative examples.
ii) Adapt W: It is obtained by substituting ERR with WERR in Adapt.
iii) Adapt 2W: other than replacing ERR with WERR, the weighting factor is also introduced in the model adaptation method[12].
Merits:–i) The adaptation can avoid negative transfer.
ii) The adaptation algorithms improve the overall performance in classification.
iii) The proposed algorithm is able to perform one shot learning[13].

Demerits: –The superiority of proposed algorithm exists only in situations where there are few samples of new category.for a learning task with sufficient samples of the new category, the proposed algorithm leads to higher computational cost.

2. Transfer In Inductive Learning: The objective of inductive learning task is to induce a predictive model from a set of training examples[15]. The goal is classification.Examples of classification systems are artificial neural networks and symbolic rule learners. The other type of inductive learning involves modelling probability distributions over interrelated variables with graphical models. Example is Bayesian Networks and Markov Logic Networks[20].The predictive model is inherited by inductive learning algorithm. It makes accurate predictions on training examples[22].In order to produce a model with this generalization capability, a learning algorithm must have an inductive bias about the distribution of the training data [24].The algorithm is based on the hypothesis space of possible models. For example, the hypothesis space of the Naive Bayes model is limited by the assumption that characteristics are conditionally independent. The bias of an algorithm can be determined by its search process through hypothesis space. It determines the order in which hypotheses are considered. Rule learning algorithms construct rules, one predicate at a time and it reflects the assumption that predicates contribute in pairs or more. Transfer in inductive learning works by allowing source task knowledge to affect the target task's inductive bias [25]. It is usually concerned with improving the speed with which a model is learned as well as its generalization capability.

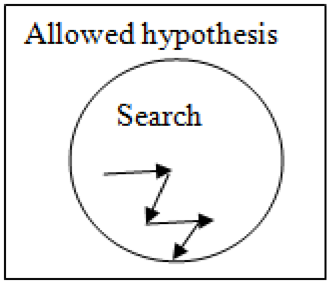


Fig3: Inductive learning

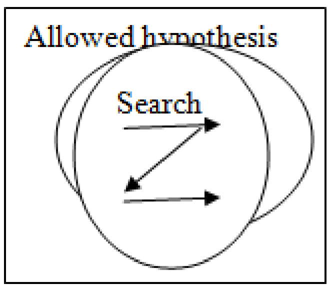
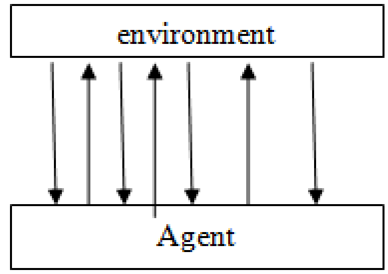


Fig4: Inductive transfer

3. Transfer in Reinforcement Learning: A reinforcement learning (RL) agent operates in a sequential control environment. It is called Markov decision process(MDP).It performs actions, changing the state and trigger rewards. The objective is to learn a policy for acting in order to maximize its cumulative reward. It involves solving a temporal credit assignment problem, since an entire sequence of actions may be responsible for a single immediate reward. At time step t, it observes the current state and consults its current policy to choose an action, π(st) = at. After taking the action, it receives a reward rt and observes the new state st+1 and it uses that information to update its policy before repeating the cycle [26].RL consists of a sequence of episodes, which end when the agent reaches one of a set of ending states. During learning, the agent must balance between exploiting the current policy, exploring new areas to find potentially higher rewards[27]. A common solution is the greedy method. Here the agent takes random exploratory actions a small fraction of the time. Usually it takes the action recommended by the current policy[28][29]. There are several categories of RL algorithms. Some types of methods are only applicable when the agent knows its environment model. In this case dynamic programming can solve directly for the optimal policy without requiring any interaction with the environment. In most RL problems, the model is unknown. The learning approaches use interaction with the environment to build an approximation of the true model and the Model free approaches learn to act without ever explicitly modelling the environment.



[Fig5: Reinforcement Learning]

IV. Conclusion and Future Scope

In this paper, we have reviewed about the transfer learning approaches, methodologies. The approaches are based on the fact about transfer in learning. It becomes difficult to achieve transfer as the source and target tasks become more complex. Practical applications of reinforcement learning are highly complex. Transfer learning has become a sizeable subfield in area of machine learning. It has numerous benefits.It is regarded as an important aspect of human learning as well as practical benefits.It can make machine learning more efficient.As computing power increases and researchers apply machine learning to more complex problems, knowledge transfer can also become more desirable.

References

[1]Yosinski J,Clune J,Bengio Y and Lipson H,How transferable are features in deep neural networks?in Advances in Neural Information Processing Systems 27 (NIPS '14), NIPS Foundation, 2014

[2] Dai, Wenyuan et al. "Translated Learning: Transfer Learning across Different Feature Spaces." NIPS (2008).

[3]Mohsen Kaboli,"A Review of Transfer Learning Algorithms",Technische Universität München, 2017

[4] Hinton G, Osindero S, Teh Y W. A Fast Learning Algorithm for Deep Belief Nets, J. Neural Computation, 2006, **18**(7): 1527–1554

[5] Taghi M. Khoshgoftaar and DingDing Wang," A survey of transfer learning"

[6] K. Driessens, J. Ramon, and T. Croonenborghs,"Transfer learning for reinforcement learning through goal and policy parametrization".In ICML Workshop on Structural Knowledge Transfer for Machine Learning, 2006

[7] Rakesh Kumar Saini, "Data Mining tools and challenges for current market trends-A Review", International Journal of Scientific Research in Network Security and Communication, Vol.7, Issue.2, pp.11-14, 2019

[8]Madhusmita Dey, Swati Sucharita Barik,"Security Enhancement in ATMs through Helmet Detection using Inductive Transfer Learning",IJSRET,Volume 7, Issue 4, April 2018

[9]Sinno Jialin Pan and Qiang Yang."A survey on transfer learning. IEEE Transactions on knowledge and data engineering", 22(10):1345–1359, 2010

[10]Hands-On Machine Learning with Scikit-Learn and Tensor Flow: Concepts, Tools, and Techniques for Building Intelligent Systems

[11]Rich Caruana."Multitask learning.Machine Learning", 28(1):41–70, 1997
[12] Dai W, Yang Q, Xue GR, Yu Y (2007) Boosting for transfer learning. In: Proceedings of the 24th international conference on machine learning. p. 193–200.
[13]A. Quattoni, M. Collins, and T. Darrell."Transfer learning for image classification with sparse prototype representations",CVPR, 2008
[14]Trevor Hastie, Robert Tibshirani, and Jerome Friedman. "The Elements of Statistical Learning: Data Mining, Inference, and Prediction". New York:Springer Verlag, 2001
[15]S. J. Pan, V. W. Zheng, Q. Yang, and D. H. Hu, "Transfer learning for wifi-based indoor localization," in Proceedings of the Workshop on Transfer Learning for Complex Task of the 23rd AAAI Conference on Artificial Intelligence, Chicago, Illinois, USA, July 2008
[16]K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell, "Text classification from labelled and unlabeled documents using EM " Machine Learning, vol. 39, no. 2-3, pp. 103–134, 2000
[17]L. I. Kuncheva and J. J. Rodrłguez, "Classifier ensembles with a random linear oracle," IEEE Transactions on Knowledge and Data Engineering, vol. 19, no. 4, pp. 500–508, 2007
[18]X. Zhu, "Semi-supervised learning literature survey," University of Wisconsin–Madison, Tech. Rep. 1530, 2006
[19]H. DaumeIII and D. Marcu, "Domain adaptation for statistical classifiers," Journal of Artificial Intelligence Research, vol. 26, pp. 101–126, 2006
[20]T. Evgeniou and M. Pontil, "Regularized multi-task learning," in Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Seattle, Washington, USA: ACM, August 2004, pp. 109–117
[21]J. Gao, W. Fan, J. Jiang, and J. Han, "Knowledge transfer via multiple model local structure mapping," in Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, Nevada: ACM, August 2008, pp. 283–291
[22]L. Mihalkova and R. J. Mooney, "Transfer learning by mapping with minimal target data," in Proceedings of the AAAI-2008 Workshop on Transfer Learning for Complex Tasks, Chicago, Illinois, USA, July 2008
[23]S. Bickel, M. Bruckner, and T. Scheffer, "Discriminative learning for ¨ differing training and test distributions," in Proceedings of the 24th international conference on Machine learning. New York, NY, USA: ACM, 2007, pp. 81–88
[24]W. Dai, Y. Chen, G.-R. Xue, Q. Yang, and Y. Yu, "Translated learning," in Proceedings of 21st Annual Conference on Neural Information Processing Systems, 2008
[25]S. Ramachandran and R. J. Mooney, "Theory refinement of Bayesian networks with hidden variables," in Proceedings of the 14th International Conference on Machine Learning, Madison, Wisconson, USA, July 1998, pp. 454–462
[26]X. Wu, V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. F. M. Ng, B. Liu, P. S. Yu, Z.-H. Zhou, M. Steinbach, D. J. Hand, and D. Steinberg, "Top 10 algorithms in data mining," Knowl. Inf. Syst., vol. 14, no. 1, pp. 1–37, 2008
[27] Zehai Gao, Cunbao Ma, Zhiyu She, Xu Dong, "An Enhanced Deep Extreme Learning Machine for Integrated Modular Avionics Health State Estimation", Access IEEE, vol. 6, pp. 65813-65823, 2018
[28] Huan Liu, Zheng Liu, Haobin Dong, Jian Ge, Zhiwen Yuan, Jun Zhu, Haiyang Zhang, Xuming Zeng, "Recurrent Neural Network-Based Approach for Sparse Geomagnetic Data Interpolation and Reconstruction", Access IEEE, vol. 7, pp. 33173-33179, 2019.
[29]Dhar Gupta K., Pampana R., Vilalta R., Ishida E. E. O. and de Souza R. S. 2016 IEEE Symposium on Computational Intelligence and Data Mining (CIDM-16) (Athens, Greece) Automated Supernova Ia Classification Using Adaptive Learning Techniques

Authors Profile

Swati Sucharita Barik is currently working as Asst. Professor in Centurion University of Technology and Management,Bhubaneswar,Odisha.
She holds M.Tech Degree in Computer Science & Engineering She is having 13+ years of teaching experience.Her area of research interests include computational intelligence,artificial intelligence,Machine Learning, Data mining and Cryptography & Network Security.

Mrs.Mamata Garanayak is currently working as an Asst. Professor in Centurion University of Technology and Management, Bhubaneswar, Odisha.
She received her B.Tech from BPUT, Odisha,India in 2005. She completed her M.Tech from KIIT, Bhubaneswar.Odisha,India in 2009.
She is pursuing her Ph.D Degree in Computer Science & Engineering. She is having 15 years of teaching experience. Her area of research interests include Big Data,DataMining.

Mrs. Sasmita Kumari Nayak, received her B.Tech in IT from BPUT, Odisha, India in 2005 and M.Tech in IT from CET, Bhubaneswar, Odisha, India in 2010. Presently she is pursuing her Ph.D research under the guidance of Dr. Chandra Sekhar Panda and Dr. Sasmita Kumari padhy.She is working as an Assistant Professor at Computer Science & Engineering Department of CUTM (Centurion University of Technology and management), Bhubaneswar, Odisha, India.Her research interest includes Grid Computing, Multiprocessor scheduling, Soft and Evolutionary Computing.

