ISSN 2277-2685
IJESR/March. 2026/ Vol-16/Issue-1s/111-116
Swapnali Sunil Gawali *et. al.,* / International Journal of Engineering & Science Research

A Survey on the State-of-the-art Multi-Task Learning Models: Techniques, Open Problems and Future Directions

Swapnali Sunil Gawali[1], Dr. D. B. Kshirsagar[2]
[1,2]Computer Dept, Sanjivani College of Engineering, Kopargaon Savitribai Phule Pune University, Pune, MS, India
Mail Id: gswap13@gmail.com[1], dbk444@gmail.com[2]

Abstract
Multi-Task Learning (MTL) is a type of machine learning technique where a model is trained to perform multiple tasks simultaneously. In deep learning, MTL refers to training a neural network to perform multiple tasks by sharing some of the network's layers and parameters across tasks. In MTL, the goal is to improve the generalization performance of the model by leveraging the information shared across tasks. By sharing some of the network's parameters, the model can learn a more efficient and compact representation of the data, which can be beneficial when the tasks are related or have some commonalities. MTL can be useful in many applications such as natural language processing, computer vision, and healthcare, where multiple tasks are related or have some commonalities. Despite the recent progress in deep learning, most approaches still go for a silo-like solution, focusing on learning each task in isolation: training a separate neural network for each individual task. Many real-world problems, however, call for a multi-modal approach and, therefore, for multi-tasking models. Multi-task learning (MTL) aims to leverage useful information across tasks to improve the generalization capability of a model. Hence, a detailed survey on the summarization on MTL model will be carried out in this abstract. In this survey, we provide a well-rounded view on state-of-the-art MTL techniques within the context of deep learning. The contributions of the MTL model concern the following. First the MTL from a network architecture point-of-view will be considered. The comprehensive overview as well as examine the pros and cons of the recent familiar MTL techniques will be included. A variety of optimization schemes to undertake the joint learning of many tasks will be examined. The qualitative constituents of the works, their commonalities and the dissimilarities will be explored. Finally, a huge investigational assessment over various databases will be provided to examine the advantages and disadvantages of different models, including both architectural as well as optimization strategies.

Index Terms*—Artificial Intelligence, Robotics, Cloud comput-ing, resource allocation, optimization*

I. Introduction

Multi-Task Learning (MTL) is a learning paradigm in machine learning and it aims to leverage useful information contained in multiple related tasks to help improve the generalization performance of all the tasks [25]. Multitask Learning is an approach to inductive transfer that improves generalization by using the domain information contained in the training signals of related tasks as an inductive bias. It does this by learning tasks in parallel while using a shared representation; what is learned for each task can help other tasks be learned better [10].

Single task learning does not adequately represent the learning process of human beings. Therefore, combining information across domains is an essential component of human intelligence. Generally, multi-task learning should be used when the tasks have some level of correlation [11]. In other words, multi-task learning improves performance when there are underlying principles or information shared between tasks [12]. For example, two tasks involving classifying images of animals are likely to be correlated, as both tasks will involve learning to detect fur patterns and colors. This would be a good use-case for multi-task learning since learning these image features is useful for both tasks [13].

The number of input features, variables, or columns present in a given dataset is known as dimensionality, and the process to reduce these features is called dimensionality reduction [23]. A dataset contains a huge number of input features in various cases, which makes the predictive modeling task more complicated. Because it is very difficult to visualize or make predictions for the training dataset with a high number of features, dimensionality reduction techniques are required in such cases. When the number of tasks is large or the data dimensionality is high, online, parallel, and distributed MTL models, along with dimensionality reduction and feature hashing, provide computational and storage advantages [16].

Modern critical infrastructures are characterized by a high degree of complexity in terms of vulnerabilities, threats, and interdependencies. The possible causes of a digital assault or cyber attack are not simple to identify, as they may result from a chain of seemingly insignificant incidents whose combination produces cascading effects at multiple levels [16]. Similarly, the rapid growth of digital technologies associated with critical infrastructure and the technical characteristics of their sub-systems continuously generate enormous volumes of heterogeneous data. This requires the adoption of intelligent techniques for critical analysis and optimal decision-making. In many applications, such as network traffic monitoring and share market analysis, data is received at a very high frequency over time [17]. Consequently, it is not feasible to store all historical samples, making real-time processing essential while preventing the review of previous samples because of the one-pass constraint [18].

In particular, different tasks may have conflicting requirements. Handling high-dimensional data is extremely challenging in practice and is commonly referred to as the curse of dimensionality [20]. As the dimensionality of the input dataset increases, machine learning algorithms and models become increasingly complex. With the growth in the number of features, the required number of training samples also increases proportionally, thereby increasing the possibility of overfitting [21]. Consequently, models trained on high-dimensional datasets often become overfitted and exhibit poor generalization performance.

Existing MTL methods are commonly categorized into two groups: hard parameter sharing and soft parameter sharing. Hard parameter sharing refers to sharing model weights among multiple tasks so that each weight is optimized jointly to minimize multiple loss functions [?]. Current techniques for constructing multi-task neural networks include shared feature extractors with task-specific decoders, different parameter-sharing strategies in existing network architectures, sharing and recombination of neural network modules, learning what to share, and fine-grained parameter sharing [22]. The most prominent optimization approaches include per-task loss weighting based on uncertainty or learning speed, gradient modulation and replacement to avoid conflicting gradients among tasks, and multi-objective optimization [23]. Although soft parameter sharing does not explicitly share parameters, it encourages the task-specific models to maintain similar parameters [24].

By reducing the dimensionality of features, the storage space required for the dataset is significantly decreased, and the computational cost as well as training time of machine learning models are substantially reduced [25].

II. Related Work

In 2019, Yousefi et al. [1] have proposed MTL model based on Gaussian processes for joint learning of variables that have been aggregated at different input scales. Their model represented each and every task as the linear combination of the realizations of latent processes that are integrated at a different scale per task. They were then able to compute the cross-covariance between the different tasks either analytically or numerically. They also allow each task to have a potentially different likelihood model and provide a variational lower bound that could be optimized in a stochastic fashion making their model suitable for larger datasets. They showed examples of the model in a synthetic example, a fertility dataset and an air pollution prediction application.

In 2018, Anezakis et al. [2] have initiated MTL model for Real-Time Large-Scale Data Analytics to differentiate the memory and towards the cyber protection of critical infrastructure. More specifically, it suggested the Multi Overlap LEarning STReaming Analytics (MOLESTRA) which was a standardization of the "Kappa" architecture. Their aim was the analysis of large data sets where the tasks were executed in an overlapping manner. This was done to ensure the utilization of the cognitive or learning relationships among the data flows. The proposed architecture used the k-NN Classifier with Self Adjusting Memory (k-NN SAM). MOLESTRA provided a clear and effective way to separate the short-term from the long-term memory. In this way the temporal intervals between the transfer of knowledge from one memory to the other and vice versa are differentiated.

In 2020, Hajj et al. [?] have developed a MTL model with time series data to enrich annotated Time-Series (TS) sensing data by way of transfer learning with new MTL models. Compared to previous MTL approaches for TS, this work introduced three contributions. First, they proposed a systematic method to examine the efficiency of MTL approaches by exploring options for three key characteristics of MTL with TS: the choice of features that efficiently capture temporally dynamic information, the similarity measure that effectively modeled the commonality and uniqueness across tasks being learned, and the choice of regularization for achieving the best tradeoff between a model's generalizability and accuracy. Second, they presented an MTL deep learning model that was shown to achieve state-of-the-art performance for personalized human activity recognition from time-series. Experimental results on three benchmark activity recognition datasets and one activity recognition in-the-wild dataset showed that the proposed framework provided performance gained over prior work while presenting a unified approach for designing MTL solutions for personalized time-series classification problems.

In 2020, Zhang et al. [25] have initiated MTL using laser processing based deep learning in extracting representative features from laser processing images with a multi-task loss that consisted of cross-entropy loss and logarithmic smooth L1 loss. In the experiment, AlexNet with multi-task learning proved to be better than deeper models. This framework of deep feature extraction also has tremendous potential to solve more laser machining problems in the future.

In 2018, Nian and Wan [5] have presented a methodology for structural health monitoring data recovery of Bayesian multi-task learning with multi-dimensional Gaussian process prior. The proposed methodology stood to model a series of tasks simultaneously rather than modeling each task independently while explicitly encoding the correlations among tasks that could be learnt efficiently from data. The primary advantage of Bayesian multi-task learning for data reconstruction was that it makes more efficient use of the data available and it advanced the reconstruction capability by making use of the underlying task relatedness. Since the modeling performance of the Gaussian process-based Bayesian approach heavily relied to the selected covariance function, particular focus has been laid on the influences of various kinds of covariance functions including the unblended and composite (hybrid) ones on reconstruction performance. The instrumented Canton Tower of 600 m high is used as a test bed to illustrate the effectiveness of the proposed method in reconstruction of structural health monitoring data. The traditional Bayesian single-task learning approach is also implemented for comparison purpose. The reconstruction results of the structural health monitoring data show that the proposed Bayesian multi-task learning methodology affords an excellent performance, while the Bayesian single-task learning method was unreliable in certain cases; yet, the selection of covariance function has a significant impact on the reconstruction performance of the proposed methodology. The work presented in this study also gained insight into how to chose an appropriate covariance function for reconstruction of missing structural health monitoring data.

In 2017, Strezoski and Worring [6] have initiated a OmniArt method for multi-task learning for artistic data analysis with a shared representation applied in the artistic domain. They continue to show how different multi-task configurations of our method behave on artistic data and outperformed handcrafted feature approaches as well as convolutional neural networks. In addition to the method and analysis, they propose a challenge like nature to the new aggregated data set with almost half a million samples and structured meta-data to encourage further research and societal engagement.

In 2017, Jonathan et al. [7] have introduced a Pose-Aware Multi-Task Re-Identification (PAMTRI) framework for labeling images with detailed pose for vehicles. This approach included two innovations compared with previous methods. First, it overcome viewpoint dependency by explicitly reasoning about vehicle pose and shape via keypoints, heatmaps and segments from pose estimation. Second, it jointly classified semantic vehicle attributes (colors and types) while performing ReID, through multi-task learning with the embedded pose representations. Since manually labeling images with detailed pose and attribute information was prohibitive, they create a largescale highly randomized synthetic dataset with automatically annotated vehicle attributes for training. Extensive experiments validate the effectiveness of each proposed component, showing that PAMTRI achieved significant improvement over state-of-the-art on two mainstream vehicles ReID.

In 2022, Gong et al. [8] have initiated the new multi-task learning-based time series anomaly detection (MTAD) method for anomaly detection, which captures the spatial-temporal correlation of the data to learn the generalized normal patterns and hence facilitated anomaly detection. First, four proxy tasks are implemented for feature extraction through joint learning: (1) Long short-term memory-based data prediction; (2) autoencoder-based latent representation learning and data reconstruction; (3) variational autoencoder-based latent representation learning and data reconstruction; and (4) joint latent representation-based data prediction. Proxy Tasks 1 and 4 capture the temporal correlation of the data by fusing the latent space, whereas Tasks 2 and 3 fully capture the spatial correlation of the data. The isolation forest algorithm then detects anomalies from the extracted features. Application to a real spacecraft dataset revealed the superiority of our method over existing techniques, and further ablation testing for each task proves the effectiveness of fusing multiple tasks. The proposed MTAD method demonstrates promising potential for effective in-orbit anomaly detection for spacecraft.

MTFL-TS [?].

III. Problem Definition

The multi-task learning (MTL) in big data dimensions have severe problems like increased computational complexity and time complexity while processing large amount of data. MTL is not good at learning representations and thus reduces the overall performance. Thus, it affects the whole system. To overcome these challenges a big data dimensionality reduction in multi-task learning is proposed.

IV. Problem Statement

Gaussian process [1] was able to compute the cross-covariance between the different tasks both analytically or numerically. Also, it provides a variational lower bound that can be optimized in a stochastic fashion that is suitable for larger datasets. However, in high dimensional spaces the efficiency becomes low when the number of features exceeds the dozens and sample and feature requirement is usually large for the prediction.

Kappa Architecture [2] helps to execute the task in overlapping manner to ensure the usage of the cognitive or learning relationships among the data flows and the long term memory and short term memory is separated to differentiate the intervals between one memory and the other memory. Moreover, a particular type of back layer optimization is absent in the Kappa architecture and the architecture always needs the requirement of single data scheme.

MTFL-TS [3] achieve best tradeoff between a model's generalizability and accuracy and the learning time series data capture the dynamic information efficiently and models the commonality and uniqueness. Yet, the tasks are not similar or not closely related to each other and MTL is not good at learning representations and thus the overall performance is reduced.

Laser Processing [25] has great capacity to solve more laser machining problems and it helps in extracting representative features from laser processing images with a multi-task loss. However, there will be dependencies on the labeled information since the method is supervised and there is a complexity in modelling a physical process and computing with analytical information.

Bayesian multi-task learning [5] makes more efficient use of the available data and gives rise to improved reconstruction capability by making use of the underlying tasks and it efficiently encodes the correlations among tasks that can be learnt efficiently from data. Yet, there is no universally acknowledged method for constructing networks from data and specifying a prior is extremely difficult and it can produce posterior distributions that are heavily influenced by the priors.

OmniArt [?] improves the quality of categorical analysis in artistic domain for multi-task learning and different multi-task configurations were performed and this method outperformed in handcraft feature. Moreover, varying the parameters scale was not available and it only depends on artistic data and not adapted and transferred to different domains.

PAMTRI [7] method overcomes the viewpoint dependencies by reasoning about vehicle pose and heatmaps. In addition, it jointly classifies the vehicle colour and type via MTL with embedded representations. But, the domain gap between the real and synthetic data is not bridged efficiently and for large scale datasets, there is no sufficient annotated vehicle information.

MTAD [8] method captures the temporal correlation of the data to learn the generalized normal patterns and facilitates anomaly detection and for each task there is ablation testing and it proves the effectiveness of fusing multiple tasks. However, the large numbers of datasets are not provided with different characteristics and the changes in the weights are not determined properly.

Hence, these challenges promote us to design a new deep learning-based big data dimensionality model in multi-task learning.

A. Research Gap

- Existing work in this area is focused on point estimation of the input data.
- Existing systems have limited to focus on feature selection methodologies.
- Less work available which addressed feature transformation methodology for multi-task learning.
- Existing systems are limited to address data dimensionality reduction classification in multi-task learning for large dataset.

V. Research Methodology

Over the past decades, the multi-task learning problem has attracted much attention in the artificial intelligence and machine learning communities. However, most published work in this area focuses on point estimation; that is, estimating model parameters and/or making predictions. Deep Learning algorithms extract high-level, complex abstractions as data representations through a hierarchical learning process. Complex abstractions are learnt at a given level based on relatively simpler abstractions formulated in the preceding level in the hierarchy in multi-task learning process.

Hence, a deep learning-based multi-task learning model will be introduced based on the big data for to process towards data dimensionality. At first, the big data will be obtained from the traditional authentic online databases and then it will be subjected to data pre-processing to improve the quality of data. The resultant pre-processed data will be applied to the multi-task learning algorithm. The block diagram of the developed big data dimensionality-based multi-task learning model is given in Figure 1.

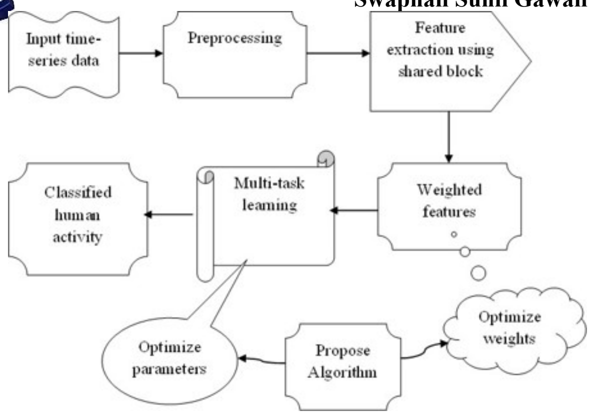

Fig. 1. Block diagram of the developed big data dimensionality-based multi-task learning model

References

[1] Fariba Yousefi, Michael T. Smith and Mauricio A´lvarez, "Multi-task Learning for Aggregated Data using Gaussian Processes," Department of Computer Science, 2019.

[2] V. D. Anezakis, K. Demertzis and L. Iliadis, "MOLESTRA: A Multi-Task Learning Approach for Real-Time Big Data Analytics," *Innovations in Intelligent Systems and Applications (INISTA)*, pp. 1–8, 2018.

[3] Hajj, H., Mahmoud, R. A. and Karameh, F. N., "A Systematic Approach to Multi-Task Learning from Time-Series Data," *Applied Soft Computing*, pp. 106586, 2020.

[4] Quexuan Zhang, Zexuan Wang, Bin Wang, Yukio Ohsawa and Teruaki Hayashi, "Feature Extraction of Laser Machining Data by Using Deep Multi-Task Learning," Department of Systems Innovation, School of Engineering, 2020.

[5] Wan, H. P. and Nian, Y., "Bayesian Multi-Task Learning Methodology for Reconstruction of Structural Health Monitoring Data," *Structural Health Monitoring*, vol. 18, no. 4, 2018.

[6] Gjorgji Strezoski and Marcel Worring, "OmniArt: Multi-task Deep Learning for Artistic Data Analysis," *Computer Science Multimedia*, 2017.

[7] Birchfield Jonathan, Zheng Tang, Milind Naphade, Stan Tremblay, William Hodge, Ratnesh Kumar, Shuo Wang and Xiaodong Yan, "PAMTRI: Pose-Aware Multi-Task Learning for Vehicle Re-Identification Using Highly Randomized Synthetic Data," *Computer Vision Foundation*, 2017.

[8] Jianglei Gong, Kaifei Yang, Yakun Wang, Xiaodong Han, Yuehua Cheng and Lifang Guo, "Unsupervised Anomaly Detection for Time Series Data of Spacecraft Using Multi-Task Learning," *Applied Sciences*, 2022.

[9] Y. Zhang and Q. Yang, "A Survey on Multi-Task Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 34, no. 12, pp. 5586–5609, Dec. 2022.

[10] Geng Tu, Jintao Wen, Hao Liu, Sentao Chen, Lin Zheng and Dazhi Jiang, "Exploration Meets Exploitation: Multitask Learning for Emotion Recognition Based on Discrete and Dimensional Models," *Knowledge-Based Systems*, vol. 235, pp. 107598, 2022.

[11] Sami Alabed and Eiko Yoneki, "High-Dimensional Bayesian Optimization with Multi-Task Learning for RocksDB," *Machine Learning and Systems*, pp. 111–119, Apr. 2021.

[12] H. Suk, S. W. Lee and Shen, "Deep Sparse Multi-Task Learning for Feature Selection in Alzheimer's Disease Diagnosis," *Brain Structure and Function*, pp. 2569–2587, 2015.

[13] C. Chen, Q. Liu, X. Wang, C. Liao and D. Zhang, "semi-Traj2Graph Identifying Fine-Grained Driving Style With GPS Trajectory Data via Multi-Task Learning," *IEEE Transactions on Big Data*, vol. 8, no. 6, pp. 1550–1565, Dec. 2022.

[14] X. Wu, M. Xu, J. Fang and X. Wu, "A Multi-Attention Tensor Completion Network for Spatiotemporal Traffic Data Imputation," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20203–20213, 2022.

[15] J. Sulam, B. Ophir, M. Zibulevsky and M. Elad, "Trainlets: Dictionary Learning in High Dimensions," *IEEE Transactions on Signal Processing*, vol. 64, no. 12, pp. 3180–3193, June 15, 2016.

[16] Salam Hamdan, Sufyan Almajali, Moussa Ayyash, Haythem Bany Salameh and Yaser Jararweh, "An Intelligent Edge-Enabled Distributed Multi-Task Learning Architecture for Large-Scale IoT-Based Cyber–Physical Systems," *Simulation Modelling Practice and Theory*, vol. 122, Jan. 2023.

[17] Zeyu Gao, Bangyang Hong, Yang Li, Xianli Zhang, Jialun Wu, Chunba Wang, Xiangrong Zhang, Tieliang Gong, Yefeng Zheng, Deyu Meng and Chen Li, "A Semi-Supervised Multi-Task Learning Framework for Cancer Classification with Weak Annotation in Whole-Slide Images," *Medical Image Analysis*, vol. 83, Jan. 2023.

[18] Zeng, Y., Cao, H., Ouyang, Q. and Qian, "Multi-Task Learning and Data Augmentation for Negative Thermal Expansion Materials Property Prediction," *Materials Today Communications*, 2021.

[19] Mehdi Katranji, Sami Kraiem, Laurent Moalic, Guilhem Sanmarty, Ghazaleh Khodabandelou, Alexandre Caminada and Fouad Hadj Selem, "Deep Multi-Task Learning for Individuals Origin–Destination Matrices Estimation from Census Data," *Data Mining and Knowledge Discovery*, 2019.

[20] Mignone, P., Pio, G., Džeroski, S. and Ceci, "Multi-Task Learning for the Simultaneous

Reconstruction of the Human and Mouse Gene Regulatory Networks," *Scientific Reports*, 2020.

[21] Qing Liao, Ye Ding, Zoe L. Jiang, Xuan Wang, Chunkai Zhang and Qian Zhang, "Multi-Task Deep Convolutional Neural Network for Cancer Diagnosis," *Neurocomputing*, vol. 348, pp. 66–73, July 5, 2019.

[22] P. Yang, P. Zhao and X. Gao, "Robust Online Multi-Task Learning with Correlative and Personalized Structures," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 11, pp. 2510–2521, Nov. 1, 2017.

[23] Wu, Shengbin and Wang, Yibai, "A Federated Multi-Task Learning Model Based on Adaptive Distributed Data Latent Correlation Analysis," *Journal of Information Processing Systems*, vol. 17, pp. 441–452, 2021.

[24] A. Liu, N. Xu, W. Nie, Y. Su and Y. Zhang, "Multi-Domain and Multi-Task Learning for Human Action Recognition," *IEEE Transactions on Image Processing*, vol. 28, pp. 853–867, Feb. 2019.

[25] B. Zhang, E. M. Provost and G. Essl, "Cross-Corpus Acoustic Emotion Recognition with Multi-Task Learning: Seeking Common Ground While Preserving Differences," *IEEE Transactions on Affective Computing*, vol. 10, pp. 85–99, Mar. 1, 2019.

[26] Ammar Kamal Abasi, Sharif Naser Makhadmeh, Mohammed Azmi Al-Betar, Osama Ahmad Alomari, Mohammed A. Awadallah, Zaid Abdi Alkareem Alyasseri, Iyad Abu Doush, Ashraf Elnagar, Eman H. Alkhammash and Myriam Hadjouni, "Lemurs Optimizer: A New Metaheuristic Algorithm for Global Optimization," *Applied Sciences*, vol. 12, pp. 10057, 2022.

[27] Amali D. G. B. and Dinakaran M., "Wildebeest Herd Optimization: A New Global Optimization Algorithm Inspired by Wildebeest Herding Behaviour," *Journal of Intelligent & Fuzzy Systems*, pp. 1–14, 2019.