International Journal of Economics and Business Administration
Vol. 7, No. 1, 2021, pp. 11-17
http://www.aiscience.org/journal/ijeba
ISSN: 2381-7356 (Print); ISSN: 2381-7364 (Online)

A Multi-Task Learning Approach for Expenditure Prediction

Carol Anne Hargreaves*, Loh Sheng Xiang

Department of Statistics & Applied Probability, National University of Singapore, Singapore

Abstract

In this paper, we utilize a Multi-Task Learning (MTL) approach to predict tourist expenditure, and compared its performance to other machine learning models. Using the MTL approach, different tasks representing different sub-categories of tourist expenditure was defined for our models, and based on the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), the MTL approach had an upper hand in predicting unseen data compared to the Random Forest and Ridge Regression. We conclude that based on RMSE and the MAE, the Multi-Task Learning approach had a slight advantage over the Random Forest and Ridge Regression models, as the MTL approach was able to utilize the regularization term to facilitate learning and updating of weights from other tasks, thereby gaining an edge on its prediction power compared to the other Single Task Learning (STL) methods. Other than looking at the errors, we wanted to see whether the MTL approach was able to give a good interpretation of the model, such as which features were important in the prediction of expenditure. With regards to the interpretability of the MTL model, the MTL gave similar features of importance as the Ridge Regression model. For example, both models placed emphasis on the characteristics of the tourist's accommodation for the prediction on total expenditure. Despite the Random Forest being a non-linear model, it seems that the MTL's transfer learning ability outweighed the benefit of being a non-linear model. In conclusion, we have demonstrated that the MTL technique improved the prediction performance of the tourist expenditure.

Keywords

Multi-Task Learning (MTL), Transfer Learning, Non-linear Model, Single Task Learning (STL), Prediction, Expenditure, Random Forests, Ridge Regression Model

Received: February 1, 2021 / Accepted: February 18, 2021 / Published online: March 29, 2021

1. Introduction

Many would have a go at the actual prediction of tourist expenditure. From the use of simple regression models such as linear regression [1], to popular methods such as neural networks [2-3]. Upon careful examination, these studies typically carry out prediction on a global scale; they predict tourist expenditure using a model on the entire data. This can pose problems as the weights of the various features in the data are assumed to stay constant, when in fact, the importance of features can vary for different groups of tourists. For example, having a luxurious hotel may be more important to middle aged tourists compared to young tourists who would spend more on shopping. Therefore, breaking the overall tourist expenditure up to sub-expenditures such as expenditure on accommodations or expenditure on food for example, might allow for lower variance in calculating the response variable. We call these sub-expenditures partitions of the overall tourist expenditure space.

To address the issue, we use the method of Multi-Task Learning (MTL) [4-6] to model the tourist expenditure prediction problem. MTL is a machine learning techniques in

* Corresponding author
E-mail address: carol.hargreaves@nus.edu.sg (C. A. Hargreaves)

which N_T learning tasks are solved at the same time, using commonalities and differences across tasks. This approach results in improved learning efficiency and prediction accuracy [22-24], although the possibility of improvement depends on how information is encoded in the data. In this work, MTL will be applied, for the first time, to predict tourist expenditure, where the tourist expenditure is dependent on five quantities, 'Accommodation', 'Dining', 'Transport', 'Sight Seeing' and 'Shopping'.

Further, MTL is a form of transfer learning which can be used when there are multiple related tasks. The MTL architecture is characterized by multiple branches of layers, that get their input from a common set of layers. The general idea of multi-task learning is to find a common representation in the earlier layers of the network, while the individual tasks $\tau \in T$ are solved in their respective single-task branches in the later stages of the network. This is most commonly realized as an encoder-decoder structure, in which each task represents a specialized decoder to the representation provided by the common encoder. While each type of label $Y_T = (y\tau 1, y\tau 2,...) \in Y_T$ favors the learning of certain features in the common part, some of them can be as well. This structure can thus help to boost the performance of the expenditure tasks [26].

The MTL network improves the model prediction performance by jointly learning correlated tasks [24-25]. The model is unique such that it enhances the prediction power by sharing information between tasks to enhance the prediction performance. MTL has seen many applications [7-9], but none in tourist expenditure. Hence, we look at MTL to perform prediction on tourist expenditures, in hopes that it fares better than other machine learning methods.

There are two parts to MTL:

(1) Defining each task and

(2) How to utilize the relatedness amongst the tasks.

The first part would depend on the data set given. Different strategies are used for different data sets. We will discuss more on the definition of each task after looking at the data set. For the second part, we utilize the relatedness by introducing regularization terms. Regularization terms are used to alter the way a model learns by means of reducing errors on a loss function. In this paper, we exploit the use of MTL in our expenditure problem using the following steps:

1. From the data set obtained, we formulate it into a MTL problem by partitioning the data into tasks.
2. We clean the features by categorizing them into different groups to increase the number of features available. Multiple research papers [10-11] use groups of the same feature type for predicting the demand and expenditure, and we do the same. e.g. Instead of using the age of the person, use age groups such as 15 - 20, 21- 30, etc.
3. We apply MTL to our defined problem and compare its performance with other state-of-the-art approaches.
4. We conduct analysis on the tasks and partitions and determine their impact on the MTL model's performance.

2. Literature Review

Over time, many studies were performed to predict tourist expenditure. Brida and Scuderi [12] performed a pure regression model analysis using the classic linear regression and ordinary least squares (OLS) estimates. Comparable models include a scobit based choice model by Wu, Zhang and Fujiwara [13] as well as the application of a neural network on the tourism economy by Law [2] and Yu, et al. [3]. Kim, et al. [14] used a type of multivariate analysis called Tobit analysis to investigate tourist expenditures.

Rudkin, Sharma [15] proposed an unconditional quantile regression (UQR) to obtain quantiles of tourist expenditure that are independent of the covariates. Time series models were also used in the research by Cao, Li and Song [16] which utilizes the autoregression time series to model tourist demands and their responses to the Chinese economy. Pai, Hung and Lin [17] proposed a novel model based on the combination of fuzzy c-means and logarithm least- squares support vector regression (LLS-SVR). As shown there are many models used to predict tourist expenditure, but none utilizing MTL.

Table 1 summarizes the literature mentioned in a timeline.

Table 1. Tourist expenditure prediction timeline.

	Timeline
2000	Neural Network by Law
2011	Tobit Analysis by Kim et al.
2012	Linear Regression by Brida and Scuderi
2013	Hybrid model of fuzzy c-means and LLS-SVR by Pai, Hung and Lin
2013	Scobit based choice model by Wu, Zhang and Fujiwara
2016	UQR by Rudkin and Sharma
2016	Autoregressive model by Cao, Li and Song
2017	Dendritic Neural Network by Yu et al.

Although tourist expenditure has been heavily studied, our work stands out from the existing work for two reasons.

(1) Our tourist data is more comprehensive in terms of its features than some used in the literature. This enables us to investigate deeper the impact of features on tourist expenditure. (2) Most of the studies can be categorized into traditional Single Task Learning (STL) which involves implementing a model on the entire data set to predict a single outcome. In STL, the model learns independently even though there may be different tasks. In MTL, the tasks learn

from each other using their relatedness to better predict their own response variable. In principle, MTL can help better predict tourist expenditure than those in the mentioned literature.

3. Methodology

It is a data set used to provide insight on tourist expenditure in and we will be using it to do predictive analysis of tourist expenditure. The data set comprises of 623 questions given to 48,489 tourists. Questions about demography such as age and nationality, to tour- related specifics such as number of nights in hotel or number of tourist attractions visited, were posed to tourists in the years 2016 to 2017.

Our aim is to predict the expenditures of tourists in 17 different categories. Throughout the paper, we will be using the term 'response' as the variable we want to predict, and the term 'predictors' as the input features to predict the response.

We need to form a sub data set with all non-zero values for the MTL model. To do that, we will need to only include data with non-zero values for all responses. We will only utilize data where all responses are non-zero. This will allow us to achieve a non-zero subset of responses that can be used with MTL. After this selection, we end up with the following data:

Table 2. Predictor Variables.

	New Responses	No. of non-zero data
Data	total, accommodation, dining, transport, sightseeing, shopping	14,956

From Table 2, we have our new data sets for our MTL model. These data set will also be used with our Random Forest and Ridge Regression models for comparison.

3.1. Multi-Task Learning (MTL)

Caruana [5] explains MTL by using a single network to produce 4 outputs simultaneously from 1 input set. We compare the MTL to a STL approach where the same inputs are put into 4 models, each giving a single output. The disadvantage in using STL is that there is no inductive transfer to leverage additional sources of information that can potentially improve the accuracy of our learning on the current task.

Similarly, we adapt this concept of simultaneously completing tasks to our tourist expenditure prediction problem by means of MTL. To formulate our methodology, we look at existing research that have applied MTL in their work, as well as other machine learning techniques that we can use as a comparison to MTL.

We were inspired by the research done by Gao et al. [8] who utilized MTL with regularization terms to predict housing prices. Gao regarded the housing prediction problem as a multi-task prediction problem. The data was consisted of characteristics of houses which had different categories such as number of schools, facilities, transportation options in the vicinity; categories that had relations to price of housing. MTL requires the definition of tasks. For their definition of tasks, they split each category in the data set into smaller partitions, each leading to one task. For example, distance to nearest train stations can be split into distances ranging from 0 to 1000 meters, 1001 to 2000 meters, etc. The MTL model will then predict the housing prices of houses that fit into each sub-category. This was done to guarantee sufficient data in each group.

To ensure there is sufficient relations between each group, more tasks were defined such that multiple categories were in a single task. For example, instead of only train station distances, it could be further specified into train station and schools for distances of 0 to 1000 meters. However, these overlapping categories were defined such that there is enough data for prediction in each subgroup.

The MTL model comes in to infer a linear function for each task, but instead of estimating the weights of a linear function, we estimate a weight matrix, where the weight variables come from the p number of linear functions from p number of tasks.

For MTL, the objective function to minimize is:

$$\min_W \sum_{p=1}^{P} \left\| x_p w_p - y_p \right\|_F^2 + \Omega(\mathrm{W})$$

Table 3. Variable notations.

Notation	Explanations
P	Number of tasks
p	A particular task
W	Weight matrix
w_p	Weight vector for a task p
x_p	Input predictors for task p
y_p	Output responses for task p

Table 3 defines the notations that we will use. Among the regularization terms used, Gao concluded that the Graph regularization term is too strict for some tasks, and among the l1- norm and l2,1-norm, l2,1-norm gives better results in terms of prediction power. Hence, we will use the l2,1-norm as our $\Omega\,(W)$.

Therefore, the objective function of our research is:

$$\min_W \sum_{p=1}^{P} \left\| x_p w_p - y_p \right\|_F^2 + \alpha \left(\sum_i \sqrt{\sum_j w_{ij}^2} \right)$$

The performance metrics used by Gao to evaluate the prediction performance of the model was the Root Mean Square Error (RMSE) and the Mean Absolute Error (MAE), we will use RMSE and MAE as our evaluation tool as well.

As we require a comparison of our methodology, we defined two baseline approaches, namely, a tree-based method called Random Forests, as well as a linear model called Ridge Regression.

3.2. Random Forest

Random Forest [19] is basically an ensemble of individual decision trees which are merged together to get a more accurate and stable prediction. This method of merging is known as bagging - a combination of trained models to improve the overall accuracy. This is true as averaging a set of observation reduces variance, and thus reducing the model's uncertainty in unseen data sets [20]. Random Forest is an improvement of bagging which adds an additional layer of randomness. By constructing each tree using different number of bootstrap samples, it allows better accuracy when there is an over-powering predictor in the data, leading to less overfitting. Liaw and Wiener [19] used Random Forest in an experiment on prediction of median prices of houses in Boston. Their algorithm on Random Forest is as follows:

Their algorithm on Random Forest is as follows:

1. Draw n bootstrap samples from original data.
2. For each of these samples, randomly sample m of the predictors and perform a regression tree.
3. Predict the new data by aggregating the predictions of the n trees.

3.3. Ridge Regression

Ridge Regression [20] is linear model which uses a l_2 regularization term as a penalty. This penalty term controls the extent of the loss term. For a penalty of the usual Residual Sum of Squares (RSS) on a multiple linear regression model, the objective function to be minimized would be:

$$\text{RSS} + \alpha \sum_{j=1}^{p} \beta_j{}^2$$

where $\alpha \geq 0$ is a tuning parameter and β_j's are the coefficients. The penalty term will cause some of the estimates to shrink close to zero, reducing the corresponding parameter's impact on the fitted values. Mahajan, K. Jain and Bergier [21] used this regression in their research to identify shopping patterns in residences in a metropolitan city. The study explains the advantages of using Ridge Regression on data with high multi-collinearity. They proposed that the instability of a normal least squares estimator can be corrected using a Ridge Regression. By minimizing the improved OLS objective function as above, the experiment noticed that it is better to use all of the variables rather than completely taking out some of the variables. This is supported by the theory of using Ridge Regression as it shrinks the estimates close to 0 and not exactly 0, hence all the predictors remain intact.

4. Experiments

In this section, we evaluate the performance of MTL against two other Single Task Learning (STL) methods, namely Random Forest and Ridge Regression. To prevent multi-collinearity, we look to remove predictors that have high correlation with each other. We will take a correlation coefficient threshold of > 0.80 as an indicator of high correlation. We will be removing one predictor from each pair of predictors with high correlation above the threshold.

Our workflow is as follows:

1. Perform MTL method over data set, predicting 6 responses at the same time.
2. Perform STL approaches over data set, predicting all 6 responses separately. i.e. One STL model for oneresponse.
3. Compare the performances of all models using evaluation methods such as RMSE and MAE.

For each model, we do a 70% - 30% train-test split. We train the model using the training set, and perform a prediction on the test set to obtain a test RMSE and test MAE as our evaluation tool

4.1. Multi-Task Learning (MTL)

We use apply a gradient descent algorithm to find the matrix W such that our objective function

$$\sum_{p=1}^{P} \left\| x_p w_p - y_p \right\|_F^2 + \alpha \left(\sum_i \sqrt{\sum_j w_{ij}^2} \right)$$

is minimized, where α is a penalty term.

To get the optimal α for training, a 5-fold Cross-validation (CV) is used to obtain the MSE from each model using a different α, and we choose the α that gives us the lowest MSE in our actual model. The optimal α was 0.80. Using this optimal α, we train the MTL model using the data training set and obtain its training RMSE and MAE. After which, the model will predict the test data to get its test RMSE and MAE.

4.2. Random Forest Results

For our random forest STL model, we perform a random forest model on 6 separate data sets for our 6 responses. For each of the models, we do an optimization on the hyper-parameters as well. We also use a 5-fold CV using a grid search of values that we want to try out.

The parameters that we want to optimize are the depth of the trees, minimum samples per leaf, maximum features to be included at each split, and minimum samples for each splits. We do the grid search for each of the mentioned parameters with the following values:

1) max depth: [5, 10, 15, 20]

2) minimum samples per leaf: [50, 70, 100]

3) maximum no. of features: [349, 300, 200, 100, 10]

4) minimum samples per split: [50, 100, 150]

After getting the best parameters based on MSE, we use these parameters in each model, and gather their training and test errors. We use a fixed number of trees of 500 in this Random Forest model.

4.3. Ridge Regression Results

In Ridge Regression, we look to find a set of βs that minimizes

$$\text{RSS} + \alpha \sum_{j=1}^{p} \beta_j^{\ 2}$$

where $\alpha \geq 0$ is a tuning parameter.

For Ridge Regression, just like MTL, using a 5-fold cross validation, we need to find the optimal α term be MSE for each single task, i.e. we find 6 optimized αs, one for each data set containing each response. After which, we train the model using the training samples with the optimized α, and get the training and test errors.

5. Results

Upon execution of our models on the data set, we obtain our test predictions as well as our evaluation scores (test RMSE and test MAE). We compare our results from the MTL model with STL models.

Table 4. RMSE of training and test sets of our models.

Non-Package	Training RMSE			Test RMSE		
	MTL	RF	Ridge	MTL	RF	Ridge
Total	562.24	582.75	560.42	577.83	619.66	579.72
Accommodation	258.12	276.88	256.34	258.25	276.99	258.77
Dining	155.00	151.33	155.45	154.50	156.49	154.75
Transport	49.14	47.78	49.22	50.00	50.43	50.03
Sightseeing	111.47	110.02	112.49	116.54	116.46	117.54
Shopping	315.38	304.66	316.80	328.27	326.74	328.83

Table 5. MAE of training and test sets of our models.

Non-Package	Training RMSE			Test RMSE		
	MTL	RF	Ridge	MTL	RF	Ridge
Total	421.86	440.84	420.98	435.41	473.46	437.62
Accommodation	187.74	202.86	186.67	186.61	202.67	187.09
Dining	116.87	113.38	117.46	117.92	118.96	118.30
Transport	37.41	36.37	37.45	38.15	38.70	38.17
Sightseeing	82.91	81.85	84.21	87.64	87.39	88.73
Shopping	227.98	216.91	228.66	238.69	233.55	238.77

Tables 4 and 5 show our training and test scores for our models, where the bold and green-coloured scores represent the model with the lowest error. At first glance, we see that MTL does not perform as well in the training prediction but managed to surpass the others in most of the test predictions.

We note that the differences in errors in not large. Nonetheless, it provides proof that MTL can work well as compared to others in this prediction problem.

5.1. Two-sample T-tests

We want to see if our predicted values are different from the actual expenditure values. To do that, we perform a 2-sample T-test, assuming unequal population variances.

Given a set of actual expenditure values X_1 and a set of predicted expenditure values X_2, with size n_1 and n_2 respectively. In our case, $n_1 = n_2$. The null hypothesis of the test is as follows:

H_0: $\mu_1 = \mu_2$

H_1: $\mu_1 \neq \mu_2$

where μ_1 and μ_2 are the means of X_1 and X_2 respectively.

Degrees of freedom for the models turns out to be > 1000. For an α significance level of 0.05, the 2-tailed T Test

Statistic critical value is 1.96. Therefore, we reject the null hypothesis if the T-statistic is greater than 1.96.

Table 6. T-test statistics.

Non-Package	T-test Statistics		
	MTL	RF	Ridge
Total	0.142	0.607	0.178
Accommodation	0.474	0.445	0.521
Dining	0.186	0.524	0.158
Transport	0.125	0.263	0.059
Sightseeing	0.002	0.068	0.094
Shopping	0.010	0.374	0.065

Table 6 shows the T-test statistics for all the models. We see that all the values are lesser than 1.96. Hence, we do not reject the null hypothesis. We conclude that there is no difference in the actual expenditure values and the predicted expenditure values at a 5% significance level for all of the models.

5.2. Predictions

Following up on the T-test, we look at how our predicted expense value differ from our actual expenditure values.

Table 7. Actual Expenditure vs Predicted Expenditure (non-exhaustive).

	Total Actual	MTL Prediction	RF Prediction	Ridge Prediction
Tourist 1	2550	2531.82	2702.74	2527.29
Tourist 2	2180	1961.36	2346.81	2034.51
Tourist 3	2700	2016.53	2272.81	2032.28
Tourist 4	2020	2285.77	2419.74	2306.34
Tourist 5	495	1062.23	964.02	1080.21
Tourist 6	985	891.34	1372.45	863.32
Tourist 7	905	1252.65	1235.31	1290.84
Tourist 8	3260	2103.38	2378.57	2124.28
Tourist 9	930	1279.54	1509.9	1272.77
Tourist 10	4200	3068.43	2886.5	3088.42

Table 7 shows the predictions of our models on the test response, total. We see that the 3 models are giving close answers to the response and the results are meaningful. i.e. All predictions seem to be close to the actual values. Therefore, these observations are in line with the result from the T-test.

5.3. Overall Evaluation

Based on RMSE and MAE, MTL has an upper hand in predicting unseen data compared to Random Forest and Ridge Regression. Quality of the predictions are also checked by T-tests and by looking at the predictions vs actual responses. The combination of these results is a good indicator of MTL being able to utilize the regularization term to facilitate learning and updating of weights from other tasks.

The MTL and Ridge Regression have similar important features. Although with differences in the coefficients, both the MTL and Ridge Regression place emphasis on characteristics of the tourist's accommodation for the prediction on total expenditure. Unlike the Random Forest which has 'number of days on holiday' as the most important feature. This could be due to the MTL and Ridge Regression models being linear models and are unable to partition the predictor space to capture the non-linear trends.

6. Conclusion

In this paper, 3 models were used to predict expenditures of tourists from 2016 to 2017. The different tasks representing different sub-categories of tourists expenditure were defined for our models. We first demonstrated that using MTL in theory on the tourist expenditure problem can improve prediction performance. We then looked at its interpretation to see if it gives a new perspective on the tourist expenditure problem.

After careful evaluation, we conclude that for this prediction problem, MTL has a slight advantage over Random Forest and Ridge Regression based on RMSE and MAE. This is due to MTL's ability to improve its weights by learning from the prediction of other responses, thereby giving an edge on the prediction power compared to other STL methods. Despite the Random Forest being a non-linear model, it seems that MTL's transfer learning ability outweighs the benefit of being a non-linear model. With regards to the interpretability of MTL, it gives similar features of importance as another linear model, such as the Ridge Regression, compared to the non-linear Random Forest model.

7. Limitations and Future Works

One limitations of the MTL model is that, it is unable to capture non-linear trends in the data, hence it's results were similar to the Ridge Regression model. To capture such non-linear trends, one could incorporate the MTL model with neural networks with a non- linear activation function.

Nonetheless, the MTL model is a good step towards creating a better prediction model in solving the tourist expenditure problem.

References

[1] Emanuela Marrocu, Raffaele Paci, and Andrea Zara. "Micro-economic determinants of tourist expendi- ture: A quantile regression approach". In: (2015).

[2] Rob Law. "Demand for Hotel Spending by Visitors to Hong Kong: A Study of Various Forecasting Tech- niques". In: (2000).

[3] Ying Yu et al. "Statistical Modeling and Prediction for Tourism Economy Using Dendritic Neural Network". In: (2017).

[4] Boyu Wang and Joelle Pineau. "Online Boosting Al- gorithms for Anytime Transfer and Multitask Learn- ing". In: (2015).

[5] Rich Caruana. "Multitask Learning". In: (1997).

[6] Yu Zhang and Qiang Yang. "An overview of multi- task learning". In: (2018).

[7] Xavier Gilbert, Vishal M. Patel, and Rama Chellappa. "Deep Multitask Learning for Raiway Track Inspec- tion". In: (2017).

[8] Guangliang Gao et al. "Location-Centered House Price Prediction: A Multi-Task Learning Approach". In: (2019).

[9] Han Yuan et al. “Multitask learning improves predic- tion of cancer drug sensitivity”. In: (2016).

[10] Juan L. Nicolau and Francisco J. Mas. “Heckit mod- elling of tourist expenditure: evidence from Spain”. In: (2005).

[11] Frederick Dayour, Charles Atanga Adongo, and Fran- cis Taale. “Determinants of backpackers’ expendi- ture”. In: (2015).

[12] Juan Gabriel Brida and Raffaele Scuderi. “Determi- nants of tourist expenditure: a review of microecono- metric models”. In: (2012).

[13] Lingling Wu, Junyi Zhang, and Akimasa Fujiwara. “Tourism participation and expenditure behaviour: Analysis using a scotbit based discrete-continuous choice model”. In: (2013).

[14] Woo Gon Kim et al. “Factors affecting the travel ex- penditure of visitors to Macau, China”. In: (2011).

[15] Simon Rudkin and Abhijit Sharma. “Enhancing un- derstanding of tourist spending using unconditional quantile regression”. In: (2016).

[16] Zheng Cao, Gang Li, and Haiyan Song. “Modelling the interdependence of tourism demand: The global vector autoregressive approach”. In: (2016).

[17] Ping-Feng Pai, Kuo-Chen Hung, and Kuo-Ping Lin. “Tourism demand forecasting using novel hybrid sys- tem”. In: (2013).

[18] Matthieu Kowalski. “Sparse Regression Using Mixed Norms”. In: (2009).

[19] Andy Liaw and Matthew Wiener. “Classification and Regression by Random Forest”. In: (2002).

[20] Gareth James et al. *An Introduction to Statistical Learning*. 7th ed. Springer, 2017.

[21] Vijay Mahajun, Arun K. Jain, and Michel Bergier. “Parameter Estimation in Marketing Models in the Presence of Multicollinearity: An Application of Ridge Regression”. In: (1977).

[22] Argyriou, A.; Evgeniou, T.; Pontil, M. Multi-task feature learning. In Proceedings of the 19th International Conference on Neural Information Processing Systems (NIPS’06), Vancouver, BC, Canada, 4–7 December 2006; MIT Press: Cambridge, MA, USA, 2006; pp. 41–48.

[23] Michelucci, U, and Venturini, F. “Multi-Task Learning for Multi-Dimensional Regression: Application to Luminescence Sensing”, MDPI, Applied Sciences, (2019), 9, pp. 4748.

[24] Zhang, Y.; Yang, Q. A Survey on Multi-Task Learning. arXiv 2018, arXiv: 1707.08114.

[25] Thung, K. H.; Wee, C.-Y. A brief review on multi-task learning. Multimed. Tools Appl. 2018, 77, 29705–29725.

[26] Liebel, L and Körner, M. “Auxiliary Tasks in Multi-task Learning”, Computer Science, Computer Vision and Pattern Recognition (2018). arXiv: 1805.06334 [cs.CV].