SCHOOL OF ADVANCED TECHNOLOGY

SAT301 FINAL YEAR PROJECT

Lifelong Machine Learning for Regression Problems

Interim Progress Report

In Partial Fulfillment

of the Requirements for the Degree of

Bachelor of Engineering

Student Name	:	Yihan Zhang
Student ID	:	1823149
Supervisor	:	Steven Guan
Assessor	:	Rui Yang

Abstract

Humans can learn different knowledges and skills incrementally and obtain them for a long time. Furthermore, humans can utilize the learned knowledges and transfer them to new tasks even if we have no concept about its new domain. However, for classical machine learning models, each model can only deal with one specific task and the accuracy drops if it is applied into another task. Furthermore, accumulated data can solve the lack of data issues in lifelong machine learning models. Though transfer learning and multitask learning can solve this problem, they will not improve the past tasks, which means they are asymmetric. In this work, I will analyze some existing lifelong machine learning framework about classification problems and introduce its key components of lifelong machine learning structure. Then I will propose a simple lifelong machine learning framework which has above properties to solve the regression related problems. The key component LSTM structure will be introduced in detail. Temporal experiment result will be given, starting from data gathering to prediction. Possible improvements of the framework and future work will be introduced.

Key word: Machine learning, Lifelong Machine learning, Regression, Long Short-Term Memory

Contents

1 Introduction

1.1 Motivation, Aims and Objective

1.1.1 Motivation and Advantage of the Work

During these year, online shopping has becoming more and more popular, and sales prediction is widely used in many areas such as estate market and B2B commerce. Therefore, sales volume prediction is becoming a significant feature for sellers so that they can adopt corresponding strategies to keep the balance of number of goods and gain more profits. Overestimated the sales volume could lead to the less actual sales than expected, while underestimated sales volume could lead to larger budget for promotion and then less profit. Sales prediction problem is about to predict the sales in the future based on historical information such as the past sales volume, weather information and promotions activities. If the prediction is accurate enough, seller can make different decisions to sell the goods.

1.1.2 Aims and Objectives

In this report, firstly we will discuss lifelong machine learning in general, including its significant features, then, a novel but simple lifelong machine learning framework for regression problems will be proposed. Important features of LSTM inside the framework will be introduced in detail separately. Furthermore, experiment results about the stock prediction problem will be showed, including the whole process of the experiment, from how to gather the data to the prediction. Meanwhile, the conclusion will be discussed, and the progress of the project will be analyzed, comparing the current work with the project specification. Finally, the work for the future will be given in the format of Gantt chart.

1.2 Literature Review

After machine learning was proposed, many models such as linear regression, decision tree, AdaBoost, Support Vector Machine have been proposed. However, most these models first determine the target task, then use a set of data for algorithm training and model generation, and then apply the trained algorithm to the target task. These models are not intelligent because they have no memory, which is the knowledge base (KB) in LML, as the result, these models are not able to utilize what they have learned into new tasks in the future. Meanwhile, this property leads to the situation that traditional machine learning models require a wide range of data to be used as a training group for effective learning. Figure 1[1] illustrates the brief process of traditional machine learning model, the task is confirmed first, then the trained model will be used for the confirmed task.

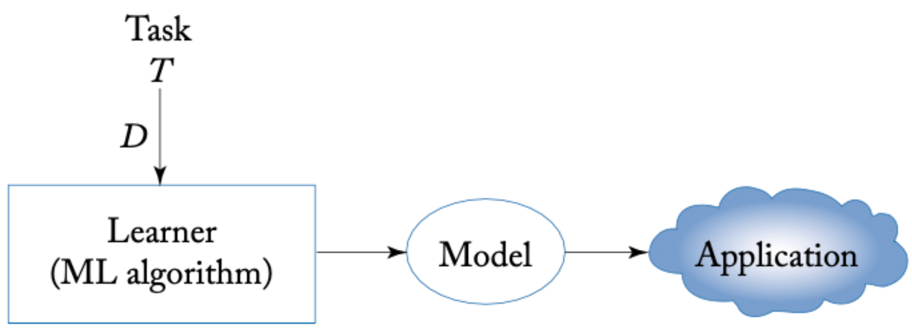


Figure 1: Traditional Machine Learning Model [1]

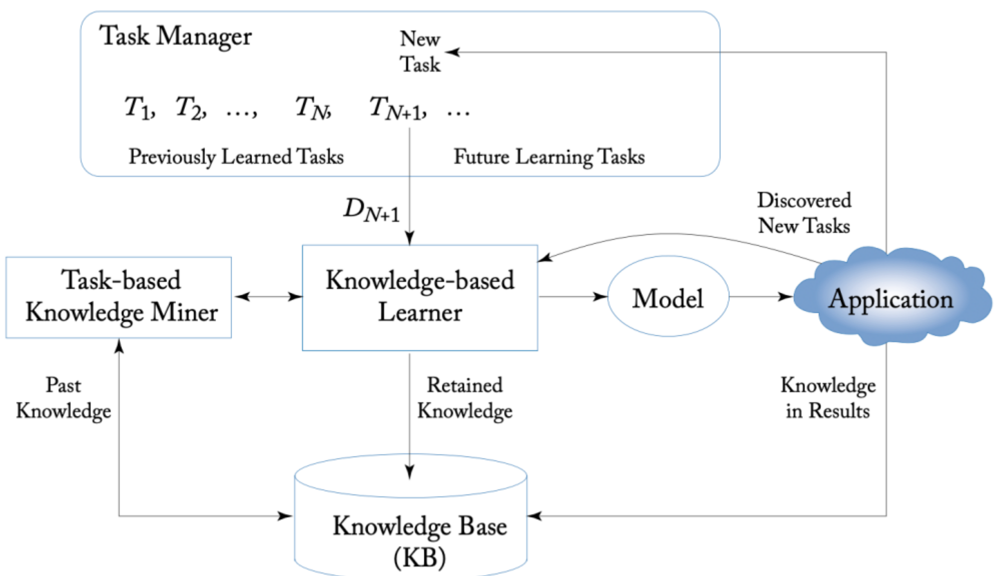


Figure 2: The Lifelong Machine Learning System Architecture [1]

With the idea above, the architecture of lifelong machine learning system has been proposed. Lifelong machine learning was firstly proposed at the end of last century [2] and the goal is to learn continuously. It is clear to observe that the architecture is more complicated compared with the traditional machine learning model. However, this figure is only for illustration purpose, it is not always the case that all the components will be used in the real situations. For traditional machine learning models, each model can only deal with one task, but for LML model, it has a series of tasks rather than a single task, which can be defined as: { $T_1, T_2, \ldots T_N, T_{N+1}$}

Introduction of key components of LML architecture [1]:

- **Knowledge Base (KB):** It is used for storing the learned knowledge from the past
- **Knowledge-based Learner (KBL):** It is used to leverage the knowledge in the knowledge base to learn the new task.
- **Task-based Knowledge Miner:** It is used to mine knowledge from the knowledge base for the new task.
- **Task Manager:** It manages tasks in the system and presents the new learning task to the knowledge-based learner.

According to Ruvolo [3], who proposed an efficient lifelong learning algorithm. The performance can achieve the accuracy identical with multi-task learning models but is three times faster than MTL models.

A survey about lifelong learning for sentiment classification conducted by Chen [4] have showed that the performance of LML is better than multi-task learning model and single task learning model.

Chen [1] summarizes the differences between some paradigms with lifelong machine learning:

- **Transfer Learning:** It only transfers the information from the source domain to the target domain once and does not retain the knowledge for the future, which is different from the accumulative knowledge idea of lifelong machine learning. Furthermore, transfer learning is unidirectional since it is only able to transfer source domain to the target domain restricted by the little knowledge in the target domain.

- **Multitask Learning:** Though there are many tasks can be completely by MTL, they can be reduced to one big task because it does not accumulate any knowledge over time, which is also different from the continuous learning idea.

According to the opinion of Hong [1], big data plays an important role in training the model because even it is difficult for humans to learn without anything. Therefore, if we need to implement lifelong machine learning model, a diverse range and a large number of domains should be given to the system so that the model can gain and master more knowledges. For image and text classification problems, there are many existing big datasets, as the result, many existing LML frameworks are related to these areas.

1.3 Industrial Relevance

In industrial production area, sometimes the seller of the factory may need to know the quantity of a certain item they require in order to purchase them from suppliers on demand. Otherwise, there could be insufficient supply of goods or lead to the overstocking of goods, which will reduce the revenue of the company. Therefore, this kind of problem can be categorized as a regression problem and can be utilized with lifelong machine learning technique to give the industrial companies confidence to predict how many products they require. Meanwhile, the model should be able to predict both horizontally (different cities) and vertically (different years).

Companies can use these forecasts to take actions, for instance, they can stock more items during the peak selling seasons while promote the sales of the products by carrying out promotional activities during off-seasons. If only focus on the concept of lifelong machine learning and get out the case study, it can be used to predict more things related to industrial such as the stock of the company, help and assist workers to complete their work more efficiently with higher accuracy. It can also predict the weather, which is useful in agriculture production and guarantees human security. Finally, it might produce the machine which is more like humans.

2 Methodology and Proposed Framework

In this section, the methodology of the framework for regression problems will be introduced, including some important components.

2.1 Long Short-Term Memory (LSTM)

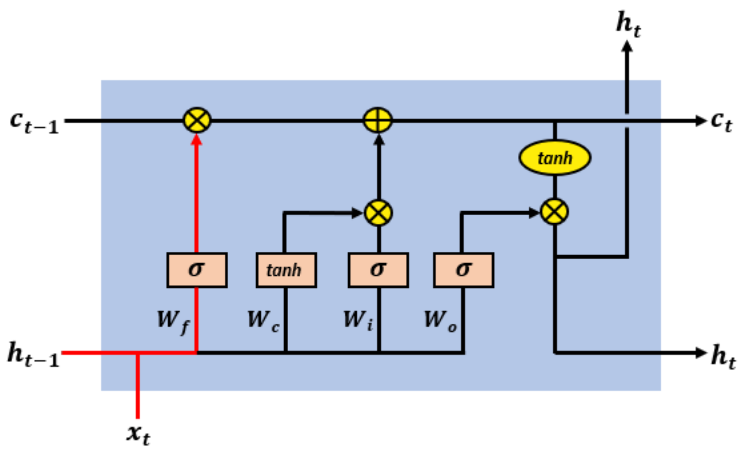


Figure 3. Structure of LSTM [5]

In 1997 [6], the proposal of LSTM solves the gradient vanish or explosion problems of RNN and LSTM has ability to memorize the dependency between input sequence and time. It achieves excellent effects in handwriting recognition and speech recognition.

Input of LSTM: c_{t-1}, h_{t-1} and x_t.

Output of LSTM: c_t and h_t.

Basically, there are three types of gates in LSTM, forget gate, input gate and output gate, which will be introduced below in detail.

2.1.1 Forget Gate

Forget gate decides how much information will be retained or throw through the sigmoid layer, it accepts h_{t-1} and x_t as the input. Output 0 to 1 values for c_{t-1} and label it as f_t. The formular is:

$$f_t = \text{sigmoid}\left(W_f.[h_{t-1}, x_t] + b_f\right)$$

2.1.2 Input Gate

Input gate has two part: one is sigmoid layer, deciding what value should be updated, the probability is i_t and another one is a tanh layer, which will create a candidate vector $\widetilde{\mathrm{C}}_{\mathrm{t}}$, which will be added into the state. The formular of it is:

$$\mathrm{i}_t = \mathrm{sigmoid}(\mathrm{W_i}.[\mathrm{h_{t-1}}, \mathrm{x_t}] + \mathrm{b_i})$$

$$\widetilde{C_t} = tanh(W_C.[h_{t-1}, x_t] + b_C)$$

The formular for updating the state:

$$C_t = f_t * C_{t-1} + it * \widetilde{C_t}$$

2.1.3 Output Gate

decides which part of current state C_t will be output through sigmoid layer. Then use tanh layer to let the value in the range between -1 and 1, then multiply with output of sigmoid layer, final result labels with h_t.
The formular is:

$$o_t = sigmoid(W_o[h_{t-1}, x_t] + b_o)$$

$$h_t = o_t * tanh(C_t)$$

2.2 LML framework for Regression Problem

The framework can be divided into three parts, for the first part, it is a container for storing the data as well as data preprocessing. Then the manipulated data will be put into the LSTM training model, the temporal output will be optimized after each iteration. Finally, the result will be derived through the trained model.

Loss function: mean squared error (MSE) function is utilized as the loss function of the prediction task:

$$\text{loss} = \frac{1}{\text{n}} . \sum \left(\text{y}_\text{n}^2 - \widehat{\text{y}_\text{n}^2} \right)$$

Where n is the number of samples in the training set, y_n is the ground truth value, $\widehat{y_n}$ is the prediction value.

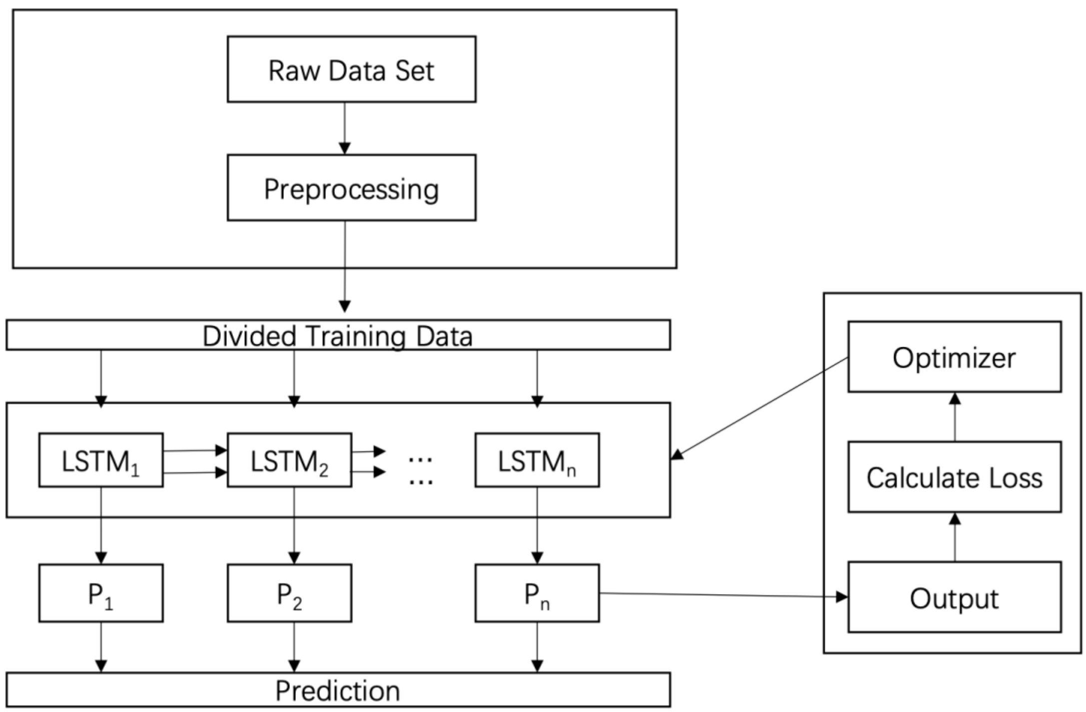


Figure 4: Proposed LML Framework for Prediction Problems

3 Preliminary Results

In this section, the environment of the experiment, the whole process of the implementation starting from data gathering to the final testing will be introduced.

3.1.1 Environment of the experiment

Hardware: M1 chips Macbook (Simple model training), Google Colab (For CUDA acceleration purpose)

Programming Language: Python 3

Software: Jupyter Notebook, PyCharm

Packages: Numpy, Pandas, Keras, sklearn, Matplotlib, Pytorch (On the macbook), TensorFlow(On the Google Colab)

3.1.2 Data Gathering and Introduction

The data was collected from Kaggle [7], which is the historical stock price of IBM from 2006 to 2018, containing 3020 rows of data with 6 columns of attributes.

In [3]:
```
dataset.head()
```

Out[3]:

Date	Open	High	Low	Close	Volume	Name
2006-01-03	82.45	82.55	80.81	82.06	11715200	IBM
2006-01-04	82.20	82.50	81.33	81.95	9840600	IBM
2006-01-05	81.40	82.90	81.00	82.50	7213500	IBM
2006-01-06	83.95	85.03	83.41	84.95	8197400	IBM
2006-01-09	84.10	84.25	83.38	83.73	6858200	IBM

In [4]:
```
dataset.shape
```

Out[4]:
```
(3020, 6)
```

Figure 5: Data Attributes Observation

3.1.3 Data Preprocessing

Fill null values and data normalization are manipulated in this step.

3.1.4 Split Training, Testing Set and Model Construction

Figure 6: Training Set (blue) and Testing Set (orange)

3.1.5 Training, Make Prediction and Visualization

Figure 7 below illustrates the loss curve with epoch of the LSTM model. The model is trained within 20 epoch and the batch size is 32. It is clear that in the early stage, the loss drops quickly and after 6 iterations, the MSE loss becomes flat and drops slowly, around 0.004. In the last epoch, the loss is 0.0018.

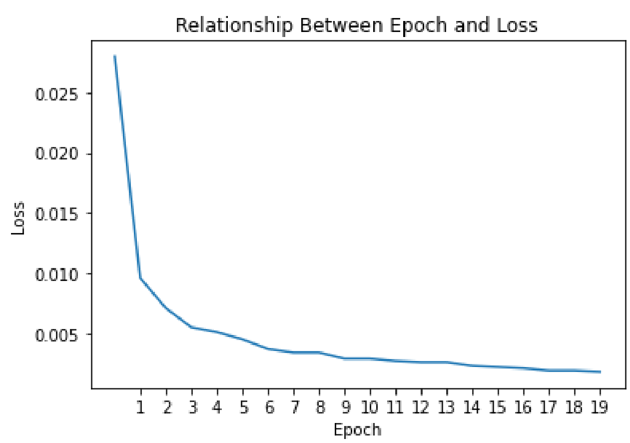


Figure 7: Relationship between Epoch and Loss Value

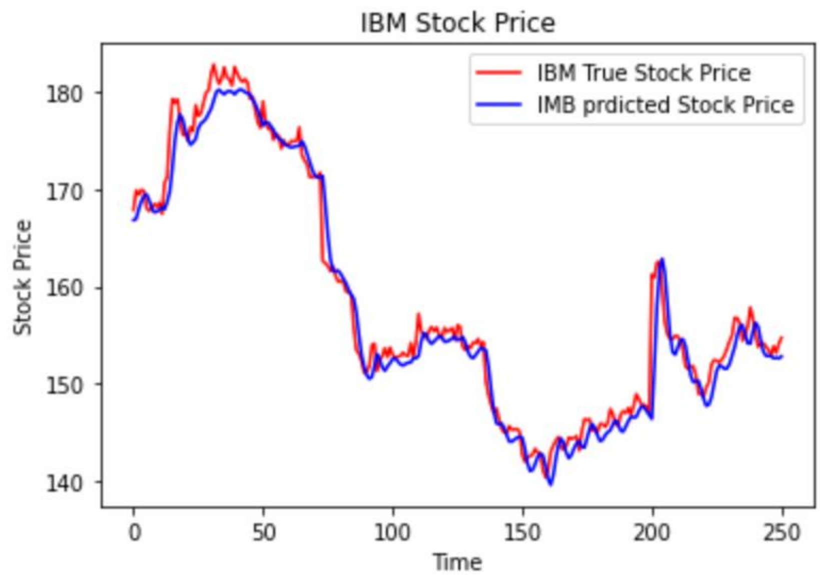


Figure 8: Prediction and the Ground truth Value

The prediction value is similar with the ground truth value in the testing data set.

4 Conclusion and Future Work

4.1 Conclusion and Limitation

Lifelong machine learning can make models behave and mimic humans, which is the type of intelligence. In the last few years, there were more researchers paying attention to this area but mainly on classification or NLP problems. Therefore, this report proposes a LML framework on regression problems. The experiment result illustrates that it is possible for model to predict the stock price, which is more related to the time series. But it is difficult to predict those with more unpredictive features such as sales promotion if we want to predict the sale volume of goods. Moreover, it is tough to detect, for instance, the sales volumes of shirt and cotton-padded jacket, if the model firstly learn the weather can decrease the sales volume of shirt, when it turns to predict cotton-padded jacket tasks, this knowledge could be a negative factor for it. There is still a long way to go for the fully intelligent.

4.2 Progress Analysis

Overall, the result has meet the original goal of the project specification. However, some milestone dates are not exact the same with those in the Gantt chart of the specification. To be more specific, In the specification report, it is estimated that the coding part can be done quickly within 20 days, but when I did the research of lifelong machine learning framework, most of papers were about NLP or image classification problems and there is no research about lifelong machine learning for regression problem. Therefore, I spent too much time on reading papers in order to find transfer learning or multitask learning code to refer. Therefore, the paper reading and literature review part accounts for most of time of the project. Figure 9 shows the real time table of the project in semester and the plan for the semester.

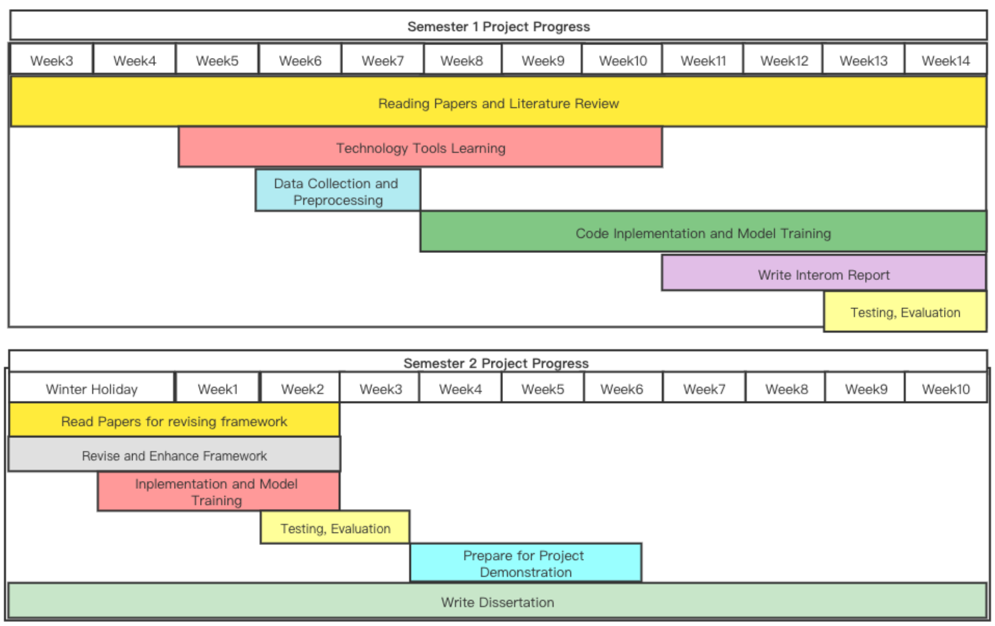


Figure 9: Gantt Chart of semester 1 and semester 2

The only budget is the account of Google Colab pro, since the speed of my computer's GPU is quite slow, and the free account of Google Colab can only get limited resources. It costs 9.9 dollar each month.

4.3 Future Work

At present, there is no universal lifelong machine learning model that can do all of tasks in all fields limited by the current technology and algorithm. Only a few tasks can be learned together. There is still a long way to realize the real intelligent for machines. It can be predicted that a large amount of data related to various field should be given in order to train and realize the universal lifelong machine learning model.

In the semester 2, more dataset about the shopping goods will be trained into the model and the original framework could be revised, another variant of LSTM, GRU structure could be added into the original framework and the corresponding implementation code, models will also be revised.

References

[1] Z. Chen, and B. Liu, "Lifelong machine learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 10, no. 3, pp. 1-145, 2016.

[2] Thrun. S, Mitchell.T.M., "Lifelong robot learning, "Robotics and autonomous systems, 1995, 25-46.

[3] Ruvolo, P.; Eaton, E. ELLA: An efficient lifelong learning algorithm. In Proceedings of the International Conference on Machine Learning, Atlanta, GE, USA, 16–21 June 2013, pp. 507–515.

[4] Z. Chen, N. Ma, and B. Liu, "Lifelong learning for sentiment classification," arXiv preprint arXiv:1801.02808, 2018.

[5] Nir. A, (2018, Dec. 10). How LSTM networks solve the problem of vanishing gradients [Online]. Available: https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577

[6] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[7] Szrlee, (2017). Historical stock data for DIJA 30 companies [Online]. Available: https://www.kaggle.com/szrlee/stock-time-series-20050101-to-20171231

Appendix A. Acronyms

LML	Lifelong Machine Learning
KB	Knowledge base
PIS	Past Information Store
KBL	Knowledge-based Learn
KM	Knowledge Miner
KR	Knowledge Reasoner
SVM	Support vector machine
MSE	Mean squared error
RNN	Recurrent neural network
LSTM	Long short-term memory
GRU	Gate recurrent Unit
f_t	Forget gate
I_t	Input gate
O_t	Output gate
C_t	Memory cell
h_t	Hidden layer state
w	Weight matrices
b	Bias vector

Appendix B. Poster

Lifelong Machine Learning For Regression Problems

Xi'an Jiaotong-Liverpool University 西交利物浦大学

Yihan Zhang

School of Advanced Technology Supervisor: Steven Guan

Abstract— Humans can learn different knowledges and skills incrementally, obtain them for a long time and we can utilize the learned knowledges, transfer them into new tasks even if we have no concept about its new domain. However, for classical machine learning models, each model can only deal with one specific task and the accuracy drops if it is applied into another task. Furthermore, accumulated data can solve the lack of data issues in lifelong machine learning models. In this work, I will analyze some existing lifelong machine learning framework and introduce its key components of lifelong machine learning. Then I will propose a simple lifelong machine learning framework which has above properties to solve the regression related problems.

Introduction of Lifelong Machine Learning

Lifelong machine learning had been proposed [1] in 1995 and developed over the past 65 years. The goal of this idea is to let the machine to complete various tasks with the help of the accumulated knowledges.

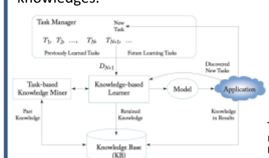


Figure 1: Architecture of LML [2]

Knowledge Base(KB):
Used for storing the knowledge

Knowledge-based Learner(KBL):
Leverage the knowledge in the KB to learn the new task.

Task-based Knowledge Miner(KM)
Mines knowledge from the KB specifically for the new task.

Task Manager:
Manages tasks in the system and presents the new learning task to the KBL in a lifelong manner.

LML Framework for Regression Problems

Structure of LSTM [3], which is the key component in the framework

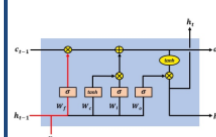


Figure 2: LSTM Structure[4]

Forget Gate:

$$f_t = \mathrm{sigmoid}\left(W_f . [h_{t-1}, x_t] + b_f\right)$$

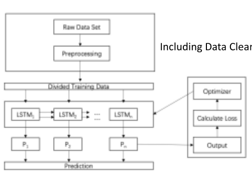
Input Gate:

$$i_t = \mathrm{sigmoid}(W_i . [h_{t-1}, x_t] + b_i)$$
$$\widetilde{C}_t = tanh(W_C . [h_{t-1}, x_t] + b_C)$$
$$C_t = f_t * C_{t-1} + it * \widetilde{C}_t$$

Output Gate:

$$o_t = sigmoid(W_o [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t * tanh(C_t)$$

Figure below is the proposed framework for regression problems.



Including Data Cleaning and Data Normalization

MSE (Mean Squared Error) Loss Function:

$$\mathrm{loss} = \frac{1}{n} . \sum \left(y_n^2 - \widehat{y_n^2}\right)$$

Figure 3: Proposed Framework for Regression Problems

Experiment Result

Data Gathering and Splitting

The data comes from the real-world data, which is the stock price of IBM from 2006 to 2018, the figure below shows how the data is split into training data and testing data based on time.

Figure 4:Data Visualization and Split

Training and Testing

The model is trained within 20 epoch and the batch size is 32. It is clear to observe that, during the early stage, the loss drops quickly and after 6 iterations, the MSE loss becomes flat and drops slowly, around 0.004. In the last epoch, the loss is 0.0018.

According to the testing result, the prediction curve is similar with the ground truth curve in the testing data set.

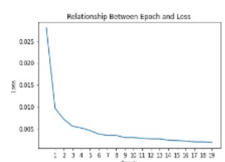


Figure 5:Relationship between Epoch and Loss Value

Figure 6:Prediction Curve and Ground Truth Curve

Conclusion and Future Work

This work introduces the existing LML framework, including its significant components and proposes a simple LML framework with LSTM structure for regression problems. The experiment is based on stock price prediction based on a real-world stock price data set.

In the semester 2 or in the future, more datasets will be train into the model and the original model could be revised in order to enhance the performance.

Realizing the lifelong machine learning is the ultimate goal for machine learning, but it is difficult to realize the fully intelligent restricted by current technology and algorithm. There is still a long way to achieve full intelligence like human beings.

Selected References

[1]: Thrun. S, Mitchell.T.M., "Lifelong robot learning, "Robotics and autonomous systems, 1995, 25-46.
[2]: Z. Chen, and B. Liu, "Lifelong machine learning," Synthesis Lectures on Artificial Intelligence and Machine Learning, vol. 10, no. 3, pp. 1-145, 2016.
[3]: S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
[4]: Nir. A, (2018, Dec. 10). How LSTM networks solve the problem of vanishing gradients [Online]. Available: https://medium.datadriveninvestor.com/how-do-lstm-networks-solve-the-problem-of-vanishing-gradients-a6784971a577

Acknowledgement

This project is carried out by Yihan Zhang as the undergraduate final year project, which is supervised by Dr. Steven Guan, Department of Advanced Technology. Thanks for Dr. Steven Guan, who has provided a lot of inspirations and guidelines about this project. And thanks for his Ph.D. student Xianbin.Hong, who always give me supports and feasible instructions when I meet problems.