Lifelong Machine Learning (Day-1)

28th June 2021

On completion of chapter 1 on lifelong machine learning, I understood the various aspects of Lifelong or Continual learning, or just the basic definition of it which goes as, for a given new task say Tn+1 for the given data Dn+1, we use the previous tasks upto Tn also the data till then Dn, the knowledge use from that plays an important role in understanding the new data and task. Now the major problem that occurs here is that of “Catastrophic Forgetting”, here when the model learns new data for its new task, it forgets the previously learnt parameters for the previous task, and since it is a relatively new area of research, some extensive research in this field could be very vital. Speaking about the architecture of LL, first is the primary Knowledge Base(KB) which is used for storing the data and model from the previously learnt tasks now this itself has several subcomponents, which are as follows:

1. Past Information System(PIS) – Storage of past data, intermediate result and model.
2. Meta Knowledge Miner (MKM) – It performs meta mining from the PIS(extraction of higher level knowledge).
3. Meta Knowledge Store(MKS) – The data mined by MKM in PIS is stored here.
4. Knowledge Reasoner (KR) – It makes inference based on the knowledge from MKM and PIS to generate more knowledge.

And other parts of the architecture following the Knowledge base are Knowledge Based Learner(KBL), Task-Based Knowledge Miner(TKM), Model, Application, Task Manager. These are all the major parts of a lifelong learning system, but in reality implementing the algorithm isn’t so sophisticated and some of the overall model are pretty simple, with few components from the mentioned above. The key challenges which are faced by LL is mainly two kinds of problems, one is Applicability of knowledge and the other one is correctness of the knowledge. Seeing the evaluation technique for LL is firstly running the data on previous tasks, running the data on new task , running a baseline algorithm( isolated learning) and finally comparing the results of both to show that LL has better prospects than Isolated learning because of learning from past knowledge and experience.

Implemented basic Tensor operations using PyTorch and also did basic linear regression for prediction of fruit prices with given input and target tensor values a model was built and the loss function used to evaluate the model was Mean Square Error.

Day – 2

29th June 2021

Completed the second chapter from Lifelong Machine Learning, we see various learning paradigms, from transfer learning, multi task learning, Online learning, reinforcement learning and finally meta learning. For transfer learning and multi task learning we also see their implementation with Deep Neural Networks and how these techniques are very different from Lifelong Learning, only two paradigms from the mentioned above come close the definition proposed for lifelong learning, Say we take Multi task learning which is closely performed for Supervised learning when it is joint with Online learning it is a form of Lifelong Learning, similarly even Meta learning, which exactly doesn’t represent the said definition but since it makes use of many tasks to help learn the new task. We go in depth in each method for example in transfer learning we take the example of Naïve Bayes Transfer Classifer(NBTC), it was a proposed method for transfer learning it is employed in two steps:

1. Building an initial Naïve Bayesian Classifer for the labelled Data(Dl) under the Distribution(dl) from the source domain.
2. Run an Expectation-Maximization Algorithm together with the data with the target unlabeled data to find a local optimal model under the target domain distribution(du).

This was also explained in the context of how Deep Neural Networks implementation takes place in Natural Language Processing Applications. Next followed by Multi task learning, how GO-MTL( Group and Overlap Multi Task Learning), which explains how the outliers was ignored in Traditional MTL, but in the proposed model of GO-MTL even the outliers having negative correlation have some information which could be also useful and be part of the cluster which are grouped together based on similar features. Same way for Reinforcement Learning, it is a way where the agent learns on trial and error basis and how it is different from LL.

Apart from this implemented a basic neural network using Pytorch, same as yesterday to predict the yield of fruits of apple and orange, considering three for atmospheric conditions, used loss function and Stochastic Gradient Descent as the optimizer, trained the model for 100 epochs. And able to get a better result than yesterday using neural networks.