International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-6, April 2020

An Effective Integration of Domain Knowledge into Deep Neural Networks

ArunimaS, KiruthikaS, M.AswinRaj, K.KavinKumar, S.SuryaPrakash

Abstract**: *Machine learning in recent years has become an integral part of our day to day life and the ease of use has improved a lot in the past decade.There are various ways to make the model to work in smaller devices.A modest method to advance any machine learning algorithm to work in smaller devices is to provide the output of large complex models as input to smaller models which can be easily deployed into mobile phones .We provided a framework where the large models can even learn the domain knowledge which is integrated as first-order logic rules and explicitly includes that knowledge into the smaller model by simultaneously training of both the models.This can be achieved by transfer learning where the knowledge learned by one model can be used to teach the other model.Domain knowledge integration is the most critical part here and it can be done by using some of the constraint principles where the scope of the data is reduced based upon the constraints mentioned. One of the best representation of domain knowledge is logic rules where the knowledge is encoded as predicates.This framework provides a way to integrate human knowledge into deep neural networks that can be easily deployed into any devices.***

Keywords* : *Algorithms, Ensemble models, First-Order Logic Rules, Deep Neural Networks.

I. INTRODUCTION

Ensemble of models are the traditional way to extract the information when the datasets are vast and have different functionality.This seems similar to parallel computation.But when the datasets are very huge the computation time also increases and accuracy drops.The most important bottleneck of any machine learning model is data, so we aim to provide a framework that helps us to make the machine learning model work well with less amount of data.We know that the amount of labelled data that is available is very small when compared to the unlabelled data that is present in the real world and also we need a more explainable machine learning models.There are lot of research going in these area where they try to identify causal relationship between data using the domain knowledge that human possess.

Integrating domain knowledge into neural networks or machine learning models makes the machine learning algorithms think exactly like humans which is the ultimate goal of machine learning. So all these research works helps us to achieve at a best possible solution for the hard open problems that are the bottlenecks of a typical machine learning.By this we can actually reduce the computational time and also makes the models to deploy in device without any complexity in execution.The ultimate goal of the framework is to integrate the domain knowledge in a complex network and the smaller model explicitly includes the domain knowledge by the process of distillation.This is like a student learning from the teacher where the small model learns from the large complex models.The main difference between other approaches is the simultaneous training of both the complex and smaller models.The remainder of this paper is organized as follows. We first review related work. Next, we formally introduce the knowledge transfer part and then integration of domain knowledge as first-order logic rules

II. RELATED WORK

Hinge-loss Markov random fields (HL-MRFs)is a probabilistic graphical model that makes the inference convex by generalising models. Randomized algorithms, probabilistic graphical models, and fuzzy logic communities are combined to form the Markov-random fields (1).The posterior regularization framework is to restrict the space of the model posteriors on unlabeled data to make the model to behave in a desired way.This helps to incorporate the domain knowledge into the neural networks (2).Distillation of the knowledge into the smaller networks from the complex knowledge has been proposed by (3) in which he says that it is possible to compress the knowledge in an ensemble into a single model which is much easier to deploy into mobile devices.The knowledge transfer happens when the probabilities of the larger models called the soft targets are used to train the smaller models.These soft targets are used in the loss function to compare the output of the smaller models and the train the smaller model.Another related work that gave a way to combine the first order logic with probabilistic graphic models which was a great achievement for the integration of domain knowledge.Here each clause has assigned with a weight that represents the importance of the first order logic (4).Combining the neural networks with logic rules can be two ways.One is neural symbolic system (11) which is developed based on rules or the first order logic.The other way is to train a network with rules that are assigned as weights to the nodes of a neural net (12).The neural symbolic systems like KBANN are networks that are constructed with rules to obtain the knowledge acquisition and perform reasoning.

Revised Manuscript Received on April 30, 2020.

***** Correspondence Author

ArunimaS*, Pursuing Bachelor’s Degree, Sri Krishna College of Technology, Coimbatore, TamilNadu, India.

KiruthikaS, Assistant Professor, Department of Computer at Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India.

M.AswinRaj, Pursuing Bachelor’s Degree,. Sri Krishna College of Technology, Coimbatore, TamilNadu, India.

K.KavinKumar, Pursuing Bachelor’s Degree,. Sri Krishna College of Technology, Coimbatore, TamilNadu, India.

S.SuryaPrakash, Pursuing Bachelor’s Degree,. Sri Krishna College of Technology, Coimbatore, TamilNadu, India.

The above works are different ways to use the first order logic to integrate domain knowledge and the process of transferring knowledge from larger neural network to smaller one

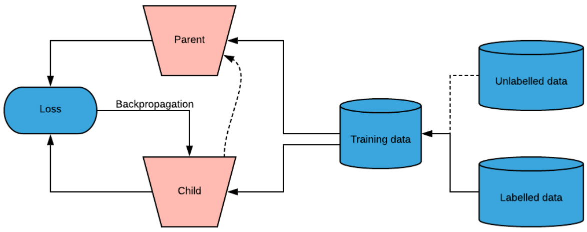
III. KNOWLEDGE TRANSFER

Neural network with more than two layers are the deep neural networks.Complexity and computation increases as the layer increases in a deep neural network.So we go for a more optimal way to reduce the computational complexity.The transfer of knowledge from the larger neural nets where the logic rules are integrated to more smaller models is the ultimate goal of the framework.The larger model is called the parent model and the smaller one is called the child.Parent model is first trained and the loss function of the parent consists of soft logits of parent and the child .Here the loss function Eq. (1) would be the KL divergence which makes the child to stay close to the parent model.Child is trained with the loss function of itself and the parent so that it explicitly incorporate the logic rules into it Eq. (2)

$$\theta(P) = \min \mathrm{KL}(p(Y|X)\|c(Y|X)) \qquad (1)$$

$$\theta(C) = \mathrm{Loss}(Y, C(x)) + \mathrm{Loss}(S(p),C(x)) \quad (2)$$

Here in Eq. (2) Y is label of the child and C(x) is the actual output of the child.S(p) is the soft targets of the parent model.The child back-propagates as shown in “Fig 1”.The logic rules that are added to the parent which explicitly get integrated into child.



1. Parent-Child framework

IV. FIRST ORDER LOGIC

First order logic is where the sentences are represented as predicates.In the case of text data we can consider the sentences to be predicates.Consider an example sentence. “He is rich but **he is unhappy**”.Sentence after “but” is consider to be B and before to be A.B has higher priority when compared to A.Therefore weight of B determines the output y. Soft logic allows continuous truth values from the interval [0, 1] instead of {0, 1}.The given example for sentiment analysis.We can encode the soft logic for every application we use.

Encoding First-Order Logic

$(1 + W(B)+)/2$ when Output=Positive
$(2 - W(B)+)/2$ when Output=Negative

V. INTEGRATING INTO NEURAL NETWORKS

The logic rules are integrated into parent through posterior constraint principle where the sentence that have more importance is given with more weight and then trained.This actually reduces the posteriors depending upon the constraint we propose in the given example.This helps in reducing the amount of posteriors that a network has to look into to make decisions.This eventually reduces the time of the network which along with knowledge transfer makes it more efficient framework.The rules are integrated directly into the parent model and the child gets the explicit integration of the logic rules.This also helps in the reduction of the complexity where the large model gets the logic rules which makes the posteriors of the parent model to get reduced based on the constraints.

VII.RESULT AND DISCUSSION

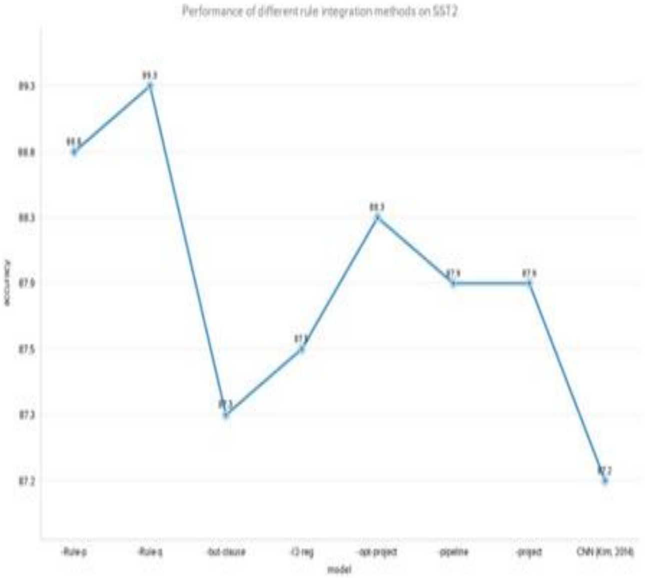
We performed experiments with SST2 with different amount of labelled data and unlabelled data and it showed performance improvement which is showed in table 1 and table 2.Performance measures with various methods used in sentiment analysis and NER tasks are shown in “Fig 3” and “Fig 4”.

Table 1. Accuracy (%) on SST2 dataset

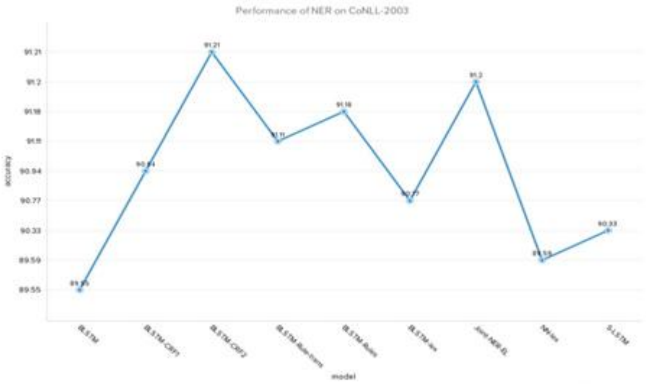
Models	Accuracy
CNN	87.2
-Rule p	88.2
Rule q	**89.3**
Semi Rule p	84.6
Semi Rule q	**85.7**

Table 2. Integration methods vs Accuracy

Integration Methods	Accuracy
But clause	87.3
-l2 reg	87.5
Rule p	**88.8**
Rule q	**89.3**



2. Performance Measures of Various Methods - Sentiment Analysis



3. Performance Measures of Various Methods - NER

VIII. CONCLUSION

A good machine learning algorithm is one that is more accurate and easily accessible for each and every person.Accuracy has been the target for the past decade.So now

it is the time to concentrate in reducing the time complexity and using less data which helps to deploy in smaller devices. Knowledge transfer is a good method for reducing the complexity and is an optimised way to deploy in mobile devices.The integration of the domain knowledge helps to reduce the use of more labeled data since the network learns half the functionalities from the domain knowledge.This even helps to use more unlabelled data for training the network.So by integrating the domain knowledge we tend to decrease the use of more labeled data which has been the most tedious talk for the machine learning models.

REFERENCES

1. Bach, S. H., Broecheler, M., Huang, B., and Getoor, L. (2015). Hinge-loss Markov random fields and probabilistic soft logic.
2. Ganchev, K., Grac ̧a, J., Gillenwater, J., and Taskar, B. (2010). Posterior regularization for structured latent variable models. JMLR, 11:2001–2049.
3. Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531.
4. Richardson, M. and Domingos, P. (2006). Markov logic networks. Machine learning, 62(1- 2):107–136.
5. T. G. Dietterich. Ensemble methods in machine learning. In Multiple classifier systems, pages1–15. Springer, 2000.
6. G. E. Hinton, L. Deng, D. Yu, G. E. Dahl, A. Mohamed, N. Jaitly, A. Senior, V. Vanhoucke,P. Nguyen, T. N Sainath, and B. Kingsbury. Deep neural networks for acoustic modelling inspeech recognition: The shared views of four research groups. Signal Processing Magazine,IEEE, 29(6):82–97, 2012.
7. R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton. Adaptive mixtures of local experts.Neural computation, 3(1):79–87, 1991.
8. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. JMLR, 12:2493–2537.
9. Franc ̧a, M. V., Zaverucha, G., and Garcez, A. S. d. (2014). Fast relational learning using bottom clause propositionalization with artificial neural networks. Machine learning, 94(1):81–104
10. Garcez, A. S. d., Broda, K., and Gabbay, D. M. (2012). Neural-symbolic learning systems: foundations and applications. Springer Science & Business Media.
11. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., and Kuksa, P. (2011). Natural language processing (almost) from scratch. JMLR, 12:2493–2537.

AUTHORS PROFILE

Arunima S is pursuing her Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, TamilNadu, India. Her research interests include Data mining and Machine Learning

Kiruthika S is currently working as Assistant Professor in the Department of Computer at Sri Krishna College of Technology, Coimbatore, Tamil Nadu, India. She received her Master’s degree in the year 2016 from Anna University, Chennai from Anna University, Chennai, Tamilnadu, India

M. Aswin Raj is pursuing her Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, TamilNadu, India. Her research interests include Data mining and Machine Learning

K. Kavin Kumar is pursuing her Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, TamilNadu, India. Her research interests include Machine Learning

S. Surya Prakash is pursuing her Bachelor’s Degree in Engineering at Sri Krishna College of Technology, Coimbatore, TamilNadu, India. Her research interests include Data mining and Machine Learning