GRAPH POWERED MACHINE LEARNING : A SURVEY

Deepansha Singh - West Windsor Plainsboro High School South

Abstract

In the 21st century, the very existence of our universe is formed on the basis of interconnectedness. From modeling social media networks to returning the most optimal GPS route, graph algorithms have been an integral part of our daily lives. Likewise, in recent years, by identifying patterns and making predictions, machine learning is evolving to replace any efforts humans need to make to complete tasks. In the intersection of both of these disciplines lies a very cutting edge technology - Graph Based Machine Learning. This domain empowers Machine Learning by introducing the concept of relationships in context to make more accurate predictions. Current research is focused on obtaining more efficient methods to various tasks such as node embeddings, community detection, clustering, and label prediction in the supervised, semi-supervised, and unsupervised machine learning settings.

Keywords : Graph Theory, Machine Learning, Graph Powered/Based Machine Learning, Supervised, Semi-supervised, Unsupervised, Node embeddings, Clustering, Community Detection, Label Prediction

1. Introduction

This review paper gives an overview of the recent research being conducted in the supervised, semi-supervised, and unsupervised areas (three sections of paper) of graph based machine learning.

Particularly, much research has been conducted to obtain accurate graph embeddings, link predictions, clustering, and community detection.

While we have developed several frameworks which perform these tasks, further enabling ML to make much more accurate predictions, I believe that we still have much scope for future research in this field. I believe that we are progressing towards drawing accurate predictions with little or even no data (zero or one shot learning) on graphs. Another potential future area of research is changing the training phase to be more efficient when large amounts of data in the graph is present. Furthermore, another future research pathway would be to handle the fact that graphs are dynamic and ever-changing - nodes can be deleted and added at any time (temporal) when creating new algorithms. Apart from applying these revolutionary findings in real world applications such as the corporate ecosystem, we are beginning to apply our findings in Natural Language Processing to get more precise results.

This review paper depicts our progress in this realm by providing summaries of the benefits of current findings, synthesizing a multitude of research papers, and commenting on various potential areas of improvement in the near future.

2. Supervised Learning

A major research area in graph based supervised learning is link prediction. [14] delves into predicting hidden links between entities in a graph. This research has great potential in identifying malicious intent in a social network, such as terrorist activity or any type of fraudulent activity. [14] achieves this by identifying and evaluating features essential for prediction as well as comparing

the computational analyses of various link prediction algorithms. One challenge of link prediction that [14] doesn't address the fact that links can also temporally change, where links are constantly destroyed and created over time. [15] takes temporal link prediction into account and proposes a solution.

[15] builds off of supervised link prediction and creates an inductive framework which generates links between entities based on the graph linkage over a multitude of graphs. Furthermore, [15] experimentally supports their findings are profound by demonstrating that the supervised setting generated great results.

DeepGL, a deep learning supervised framework, solved many different issues in this research domain such as handling large graphs as well as higher order graph structures. [16] introduces both a space efficient solution in terms of handling sparse graphs and scalable framework. Instead of looking at only node embeddings like previous research, [16] explores both node and edge embeddings. To further exemplify its impressive scalability, "In particular, node2vec takes 1.8 days (45.3 hours) for 10 million nodes, whereas DeepGL finishes in only 15 minutes"(Rossi, Zhou, Ahmed, 2017).

3. Semi-supervised Learning

One of the earlier research studies conducted in graph based semi-supervised learning uses a transductive learning approach to produce more accurate translations. Furthermore, [4] took into account the lexical similarity of words when constructing the graph, resulting in a log-linear model. While one drawback of a transductive approach is its unscalability (as this approach involves repetitively rebuilding the graph), [4] solved this issue by constructing "...separate graphs for each test sentence without losing global connectivity" (Alexandrescu and Kirchhoff, 2009) and introducing a transitive closure, which consists of "...all hypotheses for that test sentence " (Alexandrescu and Kirchhoff, 2009). Because transductive approach is not able to create a general conclusion for data that has not been tested during this training phase, one potential improvement for the research [4] developed is to use an inductive method.

A major topic in semi-supervised machine learning and graph theory intersection is graph embeddings. [2] develops both a transductive and inductive approach to learn embeddings. Furthermore, during the training phase, [2] uses the Stochastic Gradient Descent (SGD) optimization algorithm for the loss function at each iteration. Because the SGD method is used, one critical component is that the data for each iteration should be inputted randomly to eliminate any source of bias.

Another advancement in graph based semi-supervised machine learning domain has been utilized to classify whether emergency situation related tweets are relevant or irrelevant. [1] develops an inductive approach, avoiding the need to iteratively rebuild the graph. Furthermore, unlike [2], [1] takes into account how similar a particular pair of tweets are when constructing the graph. While [1] did incorporate various efficient choices such as using Euclidean distance to calculate how similar two tweets are based on word

embeddings and use the k-nearest neighbor approach ([3] used a similar methodology), one way [1] could improve its findings is by drawing accurate conclusions with a much smaller training dataset (Note : Their dataset had "...about 364 million tweets and about 3 billion words" (Firoj, Shafiq, and Muhammad, 2018). While significant machine learning research throughout the past has made conclusions based on very large datasets, one significant challenge currently is making accurate conclusions with little or no data (one shot or zero shot learning).

One application of semi-supervised learning in graphs is in the corporate ecosystem. [6] proposes a methodology which identifies relationships between entities using both structured and structured data, where each node (resembling different entities such as organizations, corporations, etc.) are given a "fingerprint, an abstract representation of the company which can be compared to any other fingerprint to calculate a similarity score" (Malik, MacGillivray, Olof-Ors, Sun, and Saroha, 2011). [6] also solved many real life issues; for instance, it constructed a directed multigraph (exactly two edges between nodes to depict the real life circumstances where entities may have different significance scores for each other). A future extension, which could solve problems encountered during the entity extraction phase, could involve incorporating Natural Language Processing techniques such as chunking, dependency parsing, and POS tagging.

Another application of graph based deep learning delves into learning and generating distributed routing protocols by utilizing Graph-Query neural nets. [5] tests the protocol generated in both Djikstra's shortest path and max-min routing algorithms, where the neural network gets exposed to graph topology changes and packet losses. By taking into account these real world problems protocols encounter, this training phase mechanism further strengthens the significance of the research. [5]'s research exemplifies the power of using a graph network, where the nodes resemble the routers, to develop a more resilient, accurate, and efficient automated mechanism to generate distributed protocols. This is further supported by the fact that, "In average, accuracies of 98%, respectively 95% could be reached for shortest path routing, respectively min-max routing" (Geyer and Carle, 2018). Apart from applying various graph neural networks to several different fields, another method to enhance this research is by taking into account how to handle much larger graphs (where there could be many more routers present in the network), and changing the training phase methodology to accommodate for this.

For instance, Stanford's GraphSAGE, a framework using inductive representation learning on graphs, utilizes a different approach to overcome the obstacle of handling large graphs - neighborhood sampling. While [17] trained its aggregator functions, which learned "...to aggregate feature information from a node's local neighborhood..instead of training a distinct embedding vector for each node" (Hamilton, Ying, and Leskovec, 2017), in both an unsupervised and supervised setting, this methodology has potential to be applied to semi-supervised settings in the near future.

4. Unsupervised Learning

GRAPH POWERED MACHINE LEARNING : A SURVEY

Deepansha Singh - West Windsor Plainsboro High School South

For a long time, the spectral clustering approach has been deemed as one of the greatest research findings in the graph theory and machine learning area. From explaining how to construct Laplacian matrices to selecting eigenvectors (key steps in the spectral clustering algorithm), [8] succinctly summarizes our progress in spectral clustering.

A major subtopic in graph based unsupervised learning is graph clustering. [7] proposes a model called GraphEncoder, which efficiently clusters graphs by employing an autoencoder (unsupervised neural network) to embed the graph and then the k-means algorithm on this result. In the past, many research articles like [8] explored the potential of spectral clustering. However, through conceptual and experimental evidence, [7] has demonstrated that their methodology of incorporating an autoencoder with a sparsity regularization term in the objective function to efficiently cluster graphs is more computationally sound than the traditional graph clustering method. One drawback of using an autoencoder is its inability to give accurate results for varying datasets different from what it was initially trained on. A possible avenue for research in this realm would be to add a computationally efficient feature which makes autoencoders even more powerful in terms of giving accurate output for all types of input data. Another possible branch of this research could involve incorporating a feature which discards irrelevant input, for autoencoders generate their own labels for all of the input data, whether or not some of the raw data is relevant/crucial.

One of the pioneering research findings in scalable unsupervised machine learning ML algorithms which learns social representations in graph uses random walk approach. [10] Particularly, one prominent finding is that “In some cases, DeepWalk’s representations can outperform its competitors even when given 60% less training data” (Perozzi, Al-Rfou, and Skiena, 2014).

Similar to [10], [9] also incorporates random walk algorithm to deduce the word embedding, where each individual node resembles a word, and the walk itself is a sentence. Because [9]’s propagation algorithm learns vector embeddings by considering a particular node’s neighbors and transmitting both forward and backward messages between neighboring vertices, for graphs which are massive and more densely interconnected, it is necessary to constrict the number of neighbors. . Overall, this research paper experimentally depicts how it computationally outperforms other techniques such as [10]. However, one drawback of this approach is that the “...loss is significantly higher for smaller values of k (parameter that measures the subset size from the set of all the node’s neighbors)”(García-Durán and Niepert, 2017). Thus, future research areas could aim to minimize this loss.

[11] introduces a new approach, different from [10], to learn node embeddings. While [10]’s framework produces great output when limiting the number of parameters, [11] altogether replaces these hyperparameters through backpropagation. The very basis of their research involves a new type of attention model which is employed on the input, honing on certain

segments of the data (of the random walk), instead of the inference phase. [11]'s findings clearly demonstrate better results in learning node embeddings by incorporating a new type of attention model. Another benefit of [11]'s findings is that it replaced hyperparameters, opening up gateways to applying the research to a multitude of real life applications.

Unlike other research such as [11] and [10], which rely on a random walk methodology , [12] considers local mutual information by using graph convolutional encoders in transductive and inductive settings to obtain node embeddings. One future possibility to extend this research could be to adapt a new framework which would take temporal link prediction into account.

5. Conclusion & Future Work

There has been much progress in the graph theory based machine learning area. From developing new methods for graph embeddings to link prediction, supervised, semi-supervised and unsupervised forms of machine learning have made algorithms much faster and predictions more accurate.

We still have much scope for future work in this emerging domain. Producing accurate shots with less data (zero shot and one shot learning) is becoming increasingly important because of the lack of sample data available in some real world cases. Additionally, future research will consider the temporal property of graphs when constructing more efficient algorithms. In some cases, it is also essential for the ML training phase to accommodate for larger datasets by tweaking this training methodology. Currently, graph powered machine learning is being applied in a multitude of applications and will continue to be a powerful tool for more real world applications, in terms of algorithm efficiency and output accuracy.

6. Citations

[1] Alam, Firoj, and Joty, Shafiq, and Imran, Muhammad. "Graph Based Semi-supervised Learning with Convolution Neural Networks to Classify Crisis Related Tweets". ICWSM. 2018

[2 Yang, Z., Cohen, W. and Salakhutdinov, R. (2016)."Revisiting Semi-Supervised Learning with Graph Embeddings". [online] arXiv.org. Available at: https://arxiv.org/abs/1603.08861.

[3] W. Dong, C. Moses, and K. Li, "Efficient k-nearest neighbor graph construction for generic similarity measures," *World Wide Web*, pp. 577–586, Mar. 2011.

[4]A. Alexandrescu and K. Kirchhoff, "Graph-based Learning for Statistical Machine Translation," *North American Chapter of the Association for Computational Linguistics*, pp. 119–127, 2009.

[5] F. Geyer and G. Carle, "Learning and Generating Distributed Routing Protocols Using Graph-Based Deep Learning," *Big Data Analytics for Network Traffic Monitoring and Analysis*, pp. 40–45, Aug. 2018.

[6] H. Malik, Hassan & MacGillivray, Ian & Olof-Ors, Måns & Sun, Siming & Saroha, Shailesh. (2011). Exploring the Corporate Ecosystem with a Semi-Supervised Entity Graph. International Conference on Information and Knowledge Management, Proceedings. 1857-1866. 10.1145/2063576.2063844.

[7] F. Tian, B. Gao, Q. Cui, E. Chen, and T.-Y. Liu, "Learning deep representations for graph clustering," *AAAI*, pp. 1293–1299, 2014.

GRAPH POWERED MACHINE LEARNING : A SURVEY

Deepansha Singh - West Windsor Plainsboro High School South

[8]Jia, H., Ding, S., Xu, X. et al. Neural Comput & Applic (2014) 24: 1477. https://doi.org/10.1007/s00521-013-1439-2

[9] García-Durán, A., & Niepert, M. (2017). Learning Graph Representations with Embedding Propagation. *NIPS*.

[10] Perozzi, Bryan & Al-Rfou, Rami & Skiena, Steven. (2014). DeepWalk: Online Learning of Social Representations. Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 10.1145/2623330.2623732.

[11] S. Abu-El-Haija, B. Perozzi, R. Al-Rfou, and A. Alemi, "Watch your step: learning node embeddings via graph attention," *NIPS'18 Proceedings of the 32nd International Conference on Neural Information Processing Systems*, pp. 9198–9208, 2018.

[12] Veličković, Petar & Fedus, William & L. Hamilton, William & Lio, Pietro & Bengio, Y & Devon Hjelm, R. (2018). Deep Graph Infomax.

[13] Hasan, Mohammad & Chaoji, Vineet & Salem, Saeed & Zaki, Mohammed. (2006). Link Prediction Using Supervised Learning.

[14] M. Dunlavy, Daniel & G. Kolda, Tamara & Acar, Evrim. (2010). Temporal Link Prediction Using Matrix and Tensor Factorizations. ACM Transactions on Knowledge Discovery from Data. 5. 10.1145/1921632.1921636.

[15] Trivedi, Rakshit & Sisman, Bunyamin & Ma, Jun & Faloutsos, Christos & Zha, Hongyuan & Luna Dong, Xin. (2018). LinkNBed: Multi-Graph Representation Learning with Entity Linkage. "

[16] Rossi, R.A., Zhou, R., & Ahmed, N.K. (2017). Deep Feature Learning for Graphs. *ArXiv, abs/1704.08829*.

[17] William L Hamilton, Rex Ying, and Jure Leskovec. Inductive representation learning on large graphs. Neural Information Processing Systems (NIPS), 2017.