様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　３ 年　 ８ 月 ２０ 日現在

機関番号：８２６２６
研究種目：若手研究
研究期間：2019～2020
課題番号：１９Ｋ２０３５２

研究課題名（和文）Bipartite Graph Embedding: As A Unified Framework

研究課題名（英文）Bipartite Graph Embedding: As A Unified Framework

研究代表者
劉　欣（LIU, XIN）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：２０８０３９３５

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：プロジェクトの成果には、調査方法とオープンソースソフトウェアの両方が含まれます。我々は、単部グラフ、二部グラフ、異種知識グラフ、属性グラフ、マルチグラフなど、さまざまなグラフに埋め込みアプローチを提案しました。また、救急車の需要予測、引用の推奨、脳障害の診断など、実際のアプリケーションへのアプローチの適用にも成功しています。

研究成果の学術的意義や社会的意義
このプロジェクトでは、様々なグラフで、グラフ埋め込みの共通の本質を明らかにします。これまでに、単部グラフ、二部グラフグラフ、異種頂点グラフ、ナレッジグラフ、属性付きグラフ、マルチグラフなど、さまざまなグラフに対する埋め込み手法を提案してきました。さまざまなグラフを研究することは、実世界の複雑なシステムを適切に表現し、幅広い応用につながるため、科学的に意義があります。

研究成果の概要（英文）：The outcomes of the project involve both the research methods as well as open-source softwares. We have proposed embedding approaches for different graphs, including the unipartite graphs, the bipartite graphs, the heterogeneous knowledge graphs, the attributed graphs, and the multigraphs. We also have successfully applied our approaches to real-world applications, including forecasting ambulance demand, citation recommendation, and brain disorder diagnosis.

研究分野：知能情報学関連

キーワード：Graph embedding　Graph neural network　Bipartite graph　Heterogeneous graph　Knowledge graph　Node classification　Node ranking

１．研究開始当初の背景

Graphs (also known as networks) are used in many branches of science as a way to represent the patterns of connections between the components of complex systems, e.g., social networks in social media, citation graphs in research areas, protein-protein interaction graphs in biology area, knowledge graphs in information science area. Analyzing these graphs provides insights into how to make good use of the information hidden in graphs, and thus has received significant attention in the last few decades.

Although graph analytics is practical and essential, most existing methods suffer the high computation and space cost. A lot of research efforts have been devoted to conducting the expensive graph analytics efficiently. In particular, there has been a surge of interest in graph embedding that learns low-dimensional vector representations, or embeddings, for nodes to encode their structural information in the graph. After the embeddings are learned, graph analysis can be easily and efficiently carried out by applying off-the-shelf vector-based machine learning algorithms. This approach removes the need for painstaking feature engineering of the traditional methods and has led to state-of-the-art results in graph analysis tasks, such as node classification, community detection, link prediction, and visualization.

２．研究の目的

Existing graph embedding methods mostly focus on embedding the simple, or unipartite, graphs, where all nodes are of the same type and edges represent a single relation. Yet a large number of real-world data can be naturally represented as bipartite graphs, where the nodes can be divided into two types and edges may only exist between nodes of different types. Moreover, there are other different graph models for representing the complex systems. For example, a social tagging system can be properly represented as a “<user, tag, resource>” tripartite hypergraph, where the nodes can be divided into three types, namely users, resources and tags, and each hyperedge has three end nodes, connecting a user, a resource and a tag that is used by the user to annotate the resource. Another example is the bibliography databases such as DBLP which contain multiple types of objects and relations: the citation relationship between papers, the authorship between authors and papers, the friendship between authors, the affiliation relationship between papers and publication venues. A bibliography database can be properly represented as a heterogeneous multi-relational graph that involves a diversity types of nodes and edges. Other graph models include the multigraphs that allow multiple edges between the same set of nodes for representing, for example, both the friendship and colleague relationship between two persons.

Only a few scattered works have targeted on the different graphs. The proposed approaches are ad hoc for some specific graphs. The key question is thus: under the appearances of various graphs, what is the common essence for graph embedding? The aim of this project is to develop a unified methodology for embedding the different graphs, because they are proper representations of real-world complex systems and lead to wide applications.

３．研究の方法

Our idea is that the various graph models can be transformed to node-edge bipartite graphs. Therefore, we start from bipartite graph embedding and then extend the method to various graphs. Specifically, we design methods based on the following techniques: (a) The random walk based technique; (b) The matrix factorization based technique; (c) The graph neural networks (GNN) based technique. Moreover, we concentrate on developing methods for usage in practical settings. Finally, we apply the proposed methods to real-world applications and validate the effectiveness.

４．研究成果

The outcomes of the project involve both the research methods as well as open-source softwares, which consist the following parts.

- We proposed the first GNN method for embedding bipartite graphs by fully considering the structure characteristics of bipartite graphs. Our method achieves 77.3% - 87.7% accuracy in binary label prediction task, which significantly outperforms traditional machine learning algorithms, statistical models, and the latest graph-based methods. This work has been published in IEEE Access journal. We have released the source code in https://github.com/Tracy-King/BiGCN
- We proposed a new approach that adapts GNN to graphs containing missing features. Our approach integrates the process of imputing missing features and graph learning within the same neural network architecture. As a result, our approach is robust against low level of missing

features. Experiments demonstrate that our approach achieves improvement of up to 11.82% in node classification and 14.54% in link prediction. This work has been published in FGCS journal. We have released the algorithm source code in https://github.com/marblet/GCNmf/

- We proposed the first GNN based model to approximate betweenness and closeness centrality, which are two commonly used node ranking measures to find out influential nodes in the graphs in terms of information spread and connectivity. A benefit of our approach is that the model is inductive, which means it can be trained on one set of graphs and evaluated on another set of graphs with varying structures. Our model consistently performs well on real-world as well as synthetic datasets with up to 44% improvement for real-world graphs and up to 70% for synthetic graphs with respect to the best KT Score on comparison methods. This work has been published in CIKM 2019 and ACM TKDD journal. We have released the software in https://github.com/sunilkmaurya/GNN_Ranking/
- We introduced an approach for temporal knowledge graph embedding that gracefully adjusts to time validity of virtually any granularity. Our model is robust to non-contiguous validity periods. It is generic enough to adapt to many existing non-temporal models and its size (number of parameters) does not depend on the size of the graph (number of entities and relations). Our approach achieves the state-of-the-art results, with 0.819 and 0.841 accuracy in time prediction in Wikipedia and YAGO datasets. This work has been published in CIKM2020. We have released the algorithm source code in https://gitlab.com/jleblay/tokei/
- We proposed an unsupervised approach for embedding attributed graphs. Our approach effectively preserves the graph structure and attribute information and overcomes the shortcoming of existing methods which deal with graph structure and attributes separately. Our approach significantly outperforms existing methods in label classification and node clustering tasks. This work has been published in Computing journal.
- We have studied how much graph structure is preserved by graph embedding techniques. We proposed four evaluation methods to evaluate the amount of information preserved by the embeddings. We found that the current graph embedding approaches can only preserve part of the topological structure. It is insufficient to rely on the embeddings to reconstruct the original graph, to discover communities, and to predict links at a high precision. Thus, we conclude that there is still some room for improving the current graph embedding techniques. This work has been published in Computer Science and Information Systems journal.
- We applied graph embeddings to real-world applications and achieved considerable success. One is multi-graph embedding for forecasting ambulance demand. This is meaningful to urban public health emergency management and is helpful for local governments to better allocate emergency medical service (EMS) resource and decrease the emergency risk. The second is knowledge graph embedding based method for automatic recommendation of citations that enhances search engines for scholarly digital libraries. Another is top-n sparse sequential recommendation that fulfils users' needs and increases the retention rate. The last is graph-based multi-model ensemble method for brain disorder diagnosis using functional magnetic resonance imaging (fMRI) data. These works have been published in ICDE2021, JCDL2020, IEEE Access journal, and Sensors journal.

5．主な発表論文等

〔雑誌論文〕　計15件（うち査読付論文　15件／うち国際共著　4件／うちオープンアクセス　7件）

1．著者名 Choong Jun Jin、Liu Xin、Murata Tsuyoshi	4．巻 22
2．論文標題 Optimizing Variational Graph Autoencoder for Community Detection with Dual Optimization	5．発行年 2020年
3．雑誌名 Entropy	6．最初と最後の頁 197～197
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/e22020197	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Chairatanakul Nuttapong、Murata Tsuyoshi、Liu Xin	4．巻 7
2．論文標題 Recurrent Translation-Based Network for Top-N Sparse Sequential Recommendation	5．発行年 2019年
3．雑誌名 IEEE Access	6．最初と最後の頁 131567～131576
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2019.2941083	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Liu Xin、Zhuang Chenyi、Murata Tsuyoshi、Kim Kyoung-Sook、Kertkeidkachorn Natthawut	4．巻 16
2．論文標題 How much topological structure is preserved by graph embeddings?	5．発行年 2019年
3．雑誌名 Computer Science and Information Systems	6．最初と最後の頁 597～614
掲載論文のDOI（デジタルオブジェクト識別子） 10.2298/CSIS181001011L	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Pornprasit Chanathip、Liu Xin、Kertkeidkachorn Natthawut、Kim Kyoung-Sook、Noraset Thanapon、Tuarob Suppawong	4．巻 -
2．論文標題 ConvCN: A CNN-Based Citation Network Embedding Algorithm towards Citation Recommendation	5．発行年 2020年
3．雑誌名 Proceedings of the 2020 ACM/IEEE Joint Conference on Digital Libraries (JCDL'20)	6．最初と最後の頁 433～436
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3383583.3398609	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Taguchi Hibiki、Murata Tsuyoshi、Liu Xin	4．巻 -
2．論文標題 BiMLPA: Community Detection in Bipartite Networks by Multi-Label Propagation	5．発行年 2020年
3．雑誌名 Proceedings of the 6th International Winter School and Conference on Network Science (NetSci-X’20)	6．最初と最後の頁 17～31
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-38965-9_2	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Choong Jun Jin、Liu Xin、Murata Tsuyoshi	4．巻 -
2．論文標題 Optimizing Variational Graph Autoencoder for Community Detection	5．発行年 2019年
3．雑誌名 Proceedings of the IEEE International Conference on Big Data Workshops (IEEE BigData Workshop'19)	6．最初と最後の頁 5353～5358
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/BigData47090.2019.9006123	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名 Maurya Sunil Kumar、Liu Xin、Murata Tsuyoshi	4．巻 -
2．論文標題 Fast Approximations of Betweenness Centrality with Graph Neural Networks	5．発行年 2019年
3．雑誌名 Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM’19)	6．最初と最後の頁 2149～2152
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3357384.3358080	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名 Kertkeidkachorn Natthawut、Liu Xin、Ichise Ryutaro	4．巻 -
2．論文標題 GTransE: Generalizing Translation-Based Model on Uncertain Knowledge Graph Embedding	5．発行年 2020年
3．雑誌名 Advances in Artificial Intelligence	6．最初と最後の頁 170～178
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-39878-1_16	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 －

1．著者名	4．巻
Maurya Sunil Kumar、Liu Xin、Murata Tsuyoshi	15
2．論文標題 Graph Neural Networks for Fast Node Ranking Approximation	5．発行年 2021年
3．雑誌名 ACM Transactions on Knowledge Discovery from Data	6．最初と最後の頁 1～32
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3446217	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名	4．巻
Taguchi Hibiki、Liu Xin、Murata Tsuyoshi	117
2．論文標題 Graph convolutional networks for graphs containing missing features	5．発行年 2021年
3．雑誌名 Future Generation Computer Systems	6．最初と最後の頁 155～168
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.future.2020.11.016	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名	4．巻
Jin Ruidong、Xia Tianqi、Liu Xin、Murata Tsuyoshi、Kim Kyoung-Sook	9
2．論文標題 Predicting Emergency Medical Service Demand With Bipartite Graph Convolutional Networks	5．発行年 2021年
3．雑誌名 IEEE Access	6．最初と最後の頁 9903～9915
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2021.3050607	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

1．著者名	4．巻
Liao Zihan、Liang Wenxin、Cui Beilei、Liu Xin	-
2．論文標題 Structure-guided attributed network embedding with “centroid” enhancement	5．発行年 2021年
3．雑誌名 Computing	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s00607-021-00916-y	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Zhaonan Wang、Tianqi Xia、Renhe Jiang、Xin Liu、Kyoung-Sook Kim、Xuan Song、Ryosuke Shibasaki	4．巻 -
2．論文標題 Forecasting Ambulance Demand with Profiled Human Mobility via Heterogeneous Multi-Graph Convolution Network	5．発行年 2021年
3．雑誌名 Proceedings of the 37th IEEE International Conference on Data Engineering (ICDE'21)	6．最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Leblay Julien、Chekol Melisachew Wudage、Liu Xin	4．巻 -
2．論文標題 Towards Temporal Knowledge Graph Embeddings with Arbitrary Time Precision	5．発行年 2020年
3．雑誌名 Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM'20)	6．最初と最後の頁 685～694
掲載論文のDOI（デジタルオブジェクト識別子） 10.1145/3340531.3412028	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1．著者名 Rakhimberdina Zarina、Liu Xin、Murata Tsuyoshi	4．巻 20
2．論文標題 Population Graph-Based Multi-Model Ensemble Method for Diagnosing Autism Spectrum Disorder	5．発行年 2020年
3．雑誌名 Sensors	6．最初と最後の頁 6001～6001
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/s20216001	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計7件（うち招待講演　0件／うち国際学会　6件）

1．発表者名 Chanathip Pornprasit、Xin Liu、Natthawut Kertkeidkachorn、Kyoung-Sook Kim、Thanapon Noraset、Suppawong Tuarob
2．発表標題 ConvCN: A CNN Based Citation Network Embedding Algorithm towards Citation Recommendation
3．学会等名 ACM/IEEE Joint Conference on Digital Libraries (JCDL 2020)（国際学会）
4．発表年 2020年

1．発表者名 Taguchi Hibiki、Murata Tsuyoshi、Liu Xin
2．発表標題 BiMLPA: Community Detection in Bipartite Networks by Multi-Label Propagation
3．学会等名 The 6th International Winter School and Conference on Network Science (NetSci-X’20)（国際学会）
4．発表年 2020年

1．発表者名 Choong Jun Jin、Liu Xin、Murata Tsuyoshi
2．発表標題 Optimizing Variational Graph Autoencoder for Community Detection
3．学会等名 The IEEE International Conference on Big Data Workshops (IEEE BigData Workshop'19)（国際学会）
4．発表年 2019年

1．発表者名 Maurya Sunil Kumar、Liu Xin、Murata Tsuyoshi
2．発表標題 Fast Approximations of Betweenness Centrality with Graph Neural Networks
3．学会等名 The 28th ACM International Conference on Information and Knowledge Management (CIKM’19)（国際学会）
4．発表年 2019年

1．発表者名 Kertkeidkachorn Natthawut、Liu Xin、Ichise Ryutaro
2．発表標題 GTransE: Generalizing Translation-Based Model on Uncertain Knowledge Graph Embedding
3．学会等名 Annual Conference of the Japanese Society for Artificial Intelligence
4．発表年 2020年

1．発表者名 Zhaonan Wang、Tianqi Xia、Renhe Jiang、Xin Liu、Kyoung-Sook Kim、Xuan Song、Ryosuke Shibasaki
2．発表標題 Forecasting Ambulance Demand with Profiled Human Mobility via Heterogeneous Multi-Graph Convolution Network
3．学会等名 The 37th IEEE International Conference on Data Engineering (ICDE'21)（国際学会）
4．発表年 2021年

1．発表者名 Leblay Julien、Chekol Melisachew Wudage、Liu Xin
2．発表標題 Towards Temporal Knowledge Graph Embeddings with Arbitrary Time Precision
3．学会等名 The 29th ACM International Conference on Information & Knowledge Management (CIKM'20)（国際学会）
4．発表年 2020年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

6．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

7．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

8．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関			
中国	Dalian University of Technology			
中国	Southern Uni. of Science and Technology			
オランダ	Utrecht University			