World Wide Web (2023) 26:1–6
https://doi.org/10.1007/s11280-022-01048-9

Special Issue on Decision Making in Heterogeneous Network Data Scenarios and Applications

Jianxin Li[1] · Chengfei Liu[2] · Ziyu Guan[3] · Yinghui Wu[4]

Published online: 22 March 2022

Decision making is the process of making choices by identifying a decision, gathering information, and assessing alternative resolutions. Using a step-by-step decision-making process can help customers make more deliberate, thoughtful decisions by organizing relevant information and defining alternatives. Traditionally, decision making has been investigated in recommendation in social networks, autonomous operations in multi-agent environments, production planning and scheduling in manufacturing systems, patients' care and treatment in emergency department management in hospitals. However, nowadays, the data used for decision making analysis is often linked and it is in the form of heterogeneity. These heterogeneous relationships may be implicit and cannot be directly processed using the traditional approaches. Therefore, this special issue will establish an emerging forum to attract high-quality research submissions and solve the new challenges of making smart decisions in heterogeneous network scenarios and applications.

This special issue attracted 46 paper submissions, and 22 high-quality research articles were accepted after at least two rounds of vigorous reviews by two or three expert reviewers for each paper. We strongly believe that it is now an ideal time to publish this special issue with the aforesaid selected papers. The contents will provide readers of World Wide Web Journal with cutting-edge and topical information for their related research. These accepted articles make contributions and novel aspects of different research in social media data scenarios, traffic data scenarios, data security scenarios, medical data scenarios, knowledge graph learning, and the other graph based deep learning models.

✉ Jianxin Li
jianxin.li@deakin.edu.au

✉ Chengfei Liu
cliu@swin.edu.au

✉ Ziyu Guan
zyguan@xidian.edu.cn

✉ Yinghui Wu
yxw1650@case.edu

[1] Deakin University, Geelong VIC3220, Australia

[2] Swinburne University of Technology, Hawthorn VIC3122, Australia

[3] Xidian University, Xian, China

[4] Case Western Reserve University, Cleveland, OH, USA

The first line of research in this special issue is to support different decisions in social media application scenarios. In the paper "Sparse Relation Prediction based on Hypergraph Neural Networks in Online Social Networks", Guan et al. targeted to predict potential links in sparsely observed networks. To achieve this goal, they firstly discussed some blurry hyperedges and then proposed a novel hyperedge shrinking method to make the learned hyperedges more hierarchical. This method can learn hypergraph structure automatically from the given sparsely observed links and rely less on manual design. Following this, we further propose a novel hypergraph-based graph neural network to learn potential links in the graph. To address semantic fusion in the heterogeneous networks, they also put forward multi-level reconstruction methods to preserve both specific semantics denoted by meta-paths, and high-level semantics denoted by hypergraphs. In the article "Group Homophily Based Facility Location Selection in Geo-Social Networks", Ma et al. formalized the conditional p-center problem in geo-social networks (GSCpC). To tackle this problem, they developed a homophily-based relaxation algorithm by considering both social constraint and spatial constraint. Specifically, in the first phase, they proposed a partition algorithm based on Voronoi diagram to conquer the social constraint. Next, in the second phase, they proposed a heuristic algorithm to refine the query results by minimizing the distance. To accelerate the performance, they also provided a category-aware user grouping strategy and a spatial distance-based strategy to prune the unpromising results significantly. In the article "Identifying informative tweets during a pandemic via a topic-aware neural language model", Gao et al. proposed a novel Topic-Aware BERT (TABERT) model by leveraging a topic model to extract the latent topics of tweets, and developing a flexible framework to combine topic information with the output of BERT. After that, they also adopted adversarial training to achieve semi-supervised learning, and used a large amount of unlabelled data to improve inner representations of the model. This work can help to address the challenge for establishing an automated system to detect useful information in social media. In the paper "Intra- and Inter-Association Attention Network-enhanced Policy Learning for Social Group Recommendation", Wang et al. aim to address the challenge of how to capture the intra-interaction and inter-association among users, groups, and items in social networks. To do this, they proposed an Intra- and inter-association attention network with Policy learning for Social Group Recommendation (IP-SGR). Specifically, for intra-interaction attention model, they captured the preference of user pair agreement with the representation of their co-interaction items, while a gate filtering component is utilized to aggregate the group agreement with the member representations of the group. To capture the inter-association representation of groups and items, they presented inter-group attention network and inter-item prototype learning model, respectively. Finally, a reinforcement learning-based model was developed to obtain the positive and negative reward for social group recommendation. In the paper "Personalized Tag Recommendation via Denoising Auto-Encoder", Zhao et al. proposed a novel personalized tag recommendation model based on the denoising auto-encoder, namely DAE-PTR, which learns the representations of entities and encodes the complex relationships by exploiting the denoising auto-encoder framework. Specifically, for each user, it firstly generates the corrupted version of the respective tagging information by adding the multiplicative mask-out/drop-out noise into the original input, and then it learns the latent representations from the corrupted input via the auto-encoder frame work by using the cross-entropy loss. Bo doing this, the representations of involved entities (e.g., users, items and tags) and the complex relationships among them can be learnt and captured, which is significant to personalized tag recommender systems.

The second line of research in this special issue is to support different decisions in the traffic scenarios and data security scenarios. In the article "Glider: Rethinking Congestion Control with Deep Reinforcement Learning", Xia et al. proposed GLIDER, a new congestion control protocol that uses deep reinforcement learning to be more versatile and adaptive to dynamic environments. In particular, GLIDER uses a framework based on Deep Q-Network, that a sender keeps adapting its congestion control strategies by continuously interacting with the network environment. In addition, the sender constantly sends data, making it challenging to apply reinforcement learning algorithms that require step-by-step state computation to congestion control. Therefore, they designed a Dynamic Bisection Division Algorithm (DBDA) to discretize the packet transmission process into steps to ensure GLIDER's feasibility on congestion control. In the paper "Optimization of Maintenance Personnel Dispatching Strategy in Smart Grid", Chen et al. presented a new optimal strategy, namely adaptive NSGAII (NSGAII/A), for dispatching maintenance personnel in the event of security risk that can improve manpower efficiency and maintain the security of power grid at a minimum cost. To solve this optimization problem, they modelled the optimization objectives and constraints in the process of maintenance personnel scheduling, and improved the legacy nondominated sorting genetic algorithm NSGAII to combat its shortcomings in the calculation of congestion and the strategy of individual selection. In the paper "A novel feature-based framework enabling multi-type DDoS attacks detection", Zhou et al. made a comprehensive analysis about the characteristics of various types of DDoS (Distributed Denial of Service) attacks and innovatively proposed five new features from heterogeneous packets, including entropy rate of IP source flow, entropy rate of flow, entropy of packet size, entropy rate of packet size, and number of ICMP destination unreachable packet, to detect not only various types of DDoS attacks, but also the mixture of them.

The third line of research in this special issue is to support different decisions in commercial data scenarios and medical data scenarios. In the paper "A Multi-attribute Decision Making Approach based on Information Extraction for Real Estate Buyer Profiling", Zhu et al. investigated the problem of real estate buyer profiling and proposed a novel multi-attribute decision making (MADM) approach, trying to solve the needs of enterprises to locate target customers accurately. They first deal with the dataset by integrating structured with unstructured data, where an Enriched Bi-directional long short-term memory (Bi-LSTM) Conditional Random Field (EB-CRF) model is proposed to extract important information in the unstructured data. Based on four general dimensions (i.e., basic information, family situation, purchase intention, financial situation), they then designed an entropy-based weight allocation algorithm to obtain attribute weights, which helps explore implicit heterogeneous relationships. In the paper "Clustering-enhanced Stock Price Prediction with Deep Learning Approach", Li et al. designed a clustering-enhanced deep learning framework to predict stock prices with three matured deep learning forecasting models. The proposed framework considers the clustering as the forecasting preprocessing, which can improve the quality of the training models. To achieve the effective clustering, they extended the Weighted Dynamic Time Warping (WDTW) method by modifying the cost weight function with logistic probability density distribution function. They also proposed a new similarity measure, Logistic WDTW (LWDTW), to capture the relative importance of price observations, and implemented the clustering-based forecasting framework with the deep learning models such as Long Short-Term Memory (LSTM), Recurrent Neural Network (RNN) and Gated Recurrent Unit (GRU). In the paper "Memory-Augmented Meta-Learning Framework for Session-Based Target Behavior Recommendation", Yu et al. proposed a memory-augmented meta-learning framework

for session-based target behavior recommendation. It adopts meta-learning to learn well-generalized global sharing initialization parameters for all sessions, and derives personalized local parameters for each session through fine tuning. Specifically, they utilized separated self-attention network to capture sequential patterns of different behaviors. Then, in order to capture user's dynamic preferences, they treated all multi-behavior interactions in the session as a sequence, and adopted temporal convolutional network to extract current intentions. On top of meta-learning framework, a soft-clustering method is implemented based on the representation of session's intention, so that sessions with similar intentions could share initialized parameters from similar clusters. In the article "Auxiliary Signal-Guided Knowledge Encoder-Decoder for Medical Report Generation", Li et al. explored auxiliary signals' power to facilitate generating medical reports. To do so, they proposed an Auxiliary Signal-Guided Knowledge Encoder-Decoder (ASGK) to mimic radiologists' working patterns. Specifically, the auxiliary patches are explored to expand the widely used visual patch features before fed to the Transformer encoder, while the external linguistic signals help the decoder better master prior knowledge during the pre-training process. In the article "Dynamic Path Learning in Decision Trees using Contextual Bandits", Ju et al. presented a novel online decision-making solution, where the optimal path of a given decision tree is dynamically found based on the contextual bandits analysis. At each round, the learner finds a path in the decision tree by making a sequence of decisions following the tree structure and receives an outcome when a terminal node is reached. At each decision node, the environment information is observed to hint on which child node to visit, resulting in a better outcome. To address the issue, they proposed Dynamic Path Identifier (DPI), a learning algorithm where the contextual bandit is applied to every decision node, and the observed outcome is used as the reward of the previous decisions of the same round. A set of theoretical proofs has been provided in this manuscript.

The fourth line of research in this special issue is to contribute knowledge graph-based decision making. In this article "TransO: A Knowledge-Driven Representation Learning Method with Ontology Information Constraints", Li et al. proposed a novel ontology information constrained knowledge representation learning model, TransO, which can efficiently model relations explicitly and seamlessly incorporate rich ontology information to improve model performance and maintain low model complexity. Moreover, specific constraint strategies are proposed for entity types, relations, and hierarchical information to effectively implement reasoning and completion of KGs and construct knowledge-driven decisions that are more consistent with the logic of human knowledge in complex network applications. In the paper "PreKar: A Learned Performance Predictor for Knowledge Graph Stores", et al aimed to design an automatic performance predictor, PreKar, to make smart decisions in the selection of appropriate knowledge graph stores. Specifically, they developed a novel candidate stores generator, which not only discovers all possible candidate stores for model training, but also multiplies the number of training instances. Based on the generated stores, they can derive an effective and lightweight encoder to not only embed the main features of workloads and stores into the model, but also guarantee the high efficiency of the performance predictor. In the paper "Example Query on Ontology-Labels Knowledge Graph Based on Filter-Refine Strategy", Ding et al. studied a novel example query on the ontology-labels knowledge graph to better capture the query interest of users and improve the semantic relevance of query results. They also developed a filter-refine strategy-based solution to solve the example queries, which includes an ontology-labels tree index for reducing the search space, a bidirectional index for improving query efficiency, and an effective candidate results combination method for identifying top-k results.

The last line of research in this special issue is to contribute deep learning-based decision making. In the paper "Structured Anchor-inferred Graph Learning for Universal Incomplete Multi-view Clustering", He et al. devised a novel Structured Anchor-inferred Graph Learning (SAGL) method to tackle the challenging universal incomplete multi-view spectral clustering problem, which can handle arbitrary view missing cases. Specifically, instead of using the fixed distance-based weighting matrix in the existing incomplete MVSC, they constructed a structural anchor-based similarity learning model to formulate a learnable asymmetric intra-view similarity matrix. Meanwhile, the inter-view similarities can be successfully bridged by the paired anchor samples, which can skilfully overcome the limitation of insufficient information operations on incomplete multi-view data. Moreover, they also derived a complete anchor-inferred graph learning scheme to enhance the efficiency of the spectral clustering process. In the paper "Gated Graph Convolutional Network with Enhanced Representation and Joint Attention for Distant Supervised Heterogeneous Relation Extraction", Ying et al. studied the challenging issue of existing distant supervised methods, which inevitably accompany with the wrong labelling problem because they often used attention mechanism to select valid instances while ignore the core of relation extraction, i.e., entity pairs and relations. To address the issue, this work is to incorporate the enhanced representations into a gated graph convolutional network for enriching the background information and further improve the attention mechanism to focus on the most relevant relation. In the paper "Bipartite Graph Capsule Network", Zhang et al. proposed a novel method, named Bipartite Capsule Graph Neural Network (BCGNN), to achieve classification performance better on bipartite graphs. To preserve the structure, nature and labeling information of the bipartite graph, BCGNN creates the connections between vertices of the same type to build it's one-mode projection. Then, it captures the features and performs better by using the hierarchical capsule network. In the paper "Attention-based Hierarchical Denoised Deep Clustering Network", Dong et al. pointed out that the traditional GCN has not fully learnt the structural information of the neighbours. Therefore, in this paper, they proposed an attention-based hierarchical denoised deep clustering (AHDDC) algorithm to solve the problem, which enables GCN to learn multiple layers of hidden information and uses the attention mechanism to strengthen the information. Besides, they also used a denoising autoencoder to reduce the influence of the data noise on the clustering. In the paper "Effective Rule Mining of Sparse Data Based on Transfer Learning", Sun et al. designed a transfer learning-based algorithm to mine rules on sparse data effectively, named TL-ERMSD. The algorithm can detect the knowledge of a common structure as well as the rules and logic between the source and target domain. Then, rule transfer is carried out by establishing the mapping mechanism between the two domains. It advances the rule mining research of existing studies that mainly focused on dense data. In the paper "Multi-Center Federated Learning for Personalized Decision Making with Client Clustering", Long et al. developed a novel multi-center aggregation mechanism to cluster clients using their models' parameters. It learns multiple global models from data, and simultaneously derives the optimal matching between users and centers. They then formulated it as a bi-level optimization problem that can be efficiently solved by a stochastic expectation maximization (EM) algorithm. In the paper "LIFOSS: A Learned Index Scheme for Streaming Scenarios", Yu et al. presented a learned index scheme for streaming scenarios (LIFOSS), where the workloads insert, delete, and query data arbitrarily. Precisely, LIFOSS consists of three parts: an adaptive packed-memory array which stores data and handles updates with lower bound of performance guaranteed; a middle-layer model group, used to fit

the cumulative distribution function of data; a feedback mechanism designed to update parameters of the model group above in real-time locally.

We would like to thank all the authors and anonymous reviewers for their timely and important contribution to this special issue. Also, we would like to express our appreciation to the editor-in-chiefs Professor Yanchun Zhang and Professor Marek Rusinkiewicz and editorial office of the World Wide Web Journal (Springer) journal for their support.