Mining Multivariate Implicit Relationships in Academic Networks

Bo Xu
Dalian University of Technology
Dalian, Liaoning, China
boxu@dlut.edu.cn

Bowen Chen
Dalian University of Technology
Dalian, Liaoning, China
chen.mail.one@gmail.com

Tianyu Zhang
65447
Dalian, Liaoning, China
zhangtianyu_beibei@126.com

Jiaying Liu
Dalian University of Technology
Dalian, Liaoning, China
jiaying_liu@outlook.com

Chunke Liao
Dalian University of Technology
Dalian, Liaoning, China
liaochunke@dlut.edu.cn

Zhehuan Zhao
Dalian University of Technology
Dalian, Liaoning, China
z.zhao@dlut.edu.cn

ABSTRACT

Multivariate cooperative relations exist widely in the academic society. In-depth research on multivariate relationships can effectively promote the integration of disciplines and advance scientific and technological progress. The mining and analysis of advisor-advisee relationships in the cooperative network, as a hot research issue in sociology and other disciplines, is still facing various challenges such as the lack of universal models and the difficulty in identifying multivariate relationships. The traditional advisor-advisee relationship mining methods only focus on the binary relationship, and require secondary processing of node attributes and edge attributes. Therefore, based on the attributes of the node, edge, and network, we transfered the Capsule Network to multivariate relation analysis. The experimental results proved the simplicity and reliability of this model. And we studied the effects of the network feature vectors' dimension, routing iterations, and normalization on the performance of the Capsule Network. Considering that Capsule Network takes a long training time, we adopted Warm Restarts method to speed up the training process. In addition, we also used the model to generate a large-scale multivariate academic genealogy.

CCS CONCEPTS

• **Information systems** → **Data mining**; • **Computing methodologies** → *Artificial intelligence*; Machine learning.

KEYWORDS

social network analysis, relation extraction, scientific collaboration network, advisor-advisee relationship

ACM Reference Format:
Bo Xu, Bowen Chen, Tianyu Zhang, Jiaying Liu, Chunke Liao, and Zhehuan Zhao. 2022. Mining Multivariate Implicit Relationships in Academic Networks. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3487553.3524727

WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France.

ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524727

1 INTRODUCTION

With the rapid development of technology and information globalization, scientific research activities show a trend of diversification, internationalization and complexity. Academic cooperation is the main form of resource sharing, knowledge exchange and innovation of scientific researchers, and it is also an important carrier of human collective wisdom [8]. Compared with personal research, the results produced by multiple cooperative relationships have received widespread attention and recognition in various fields.

Under such a social background, the research on the complex relationships of academic society can help the government plan future scientific development strategies and promote innovation in the scientific and technological revolution. The mining and analysis of the multivariate implicit relationships in academic networks reveal the development law of contemporary science for researchers, provide a solid foundation for scientific research innovation, and also provide new ideas for government to formulate scientific research policies and improve scientific research efficiency.

As a typical representative of network, academic network has also attracted extensive attention, and its research results have also been applied to practical work. For example, citation network is used to study the knowledge flow and knowledge transfer of science and technology; co-authorship network is used to discover scientific research groups and identify cooperation patterns; co-citation network, co-word network and coupling network can be used to judge interdisciplinary research level, identify scientific research experts in various fields, and describe the map of discipline and scientific development. The research of academic network includes evaluation of academic influence and cooperation efficiency, disciplinary and interdisciplinary cooperation, disciplinary knowledge and research topic identification, scientific map drawing, and knowledge path discovery. The future research direction is about hybrid academic network and heterogeneous academic network.

In the past, binary relationship was the main object of academic relationship research. Our work goes one step further, focusing on multivariate relationship and applying the Capsule Network to the new domain of relation extraction.

The rest of the paper is structured as follows. Section 2 introduces the current relevant research. Section 3 formally defines the problem we want to explore, and shows the data attributes and the main experimental process. Section 4 describes the specific experimental details and analyzes the experimental results. Section 5 concludes the paper.

2 RELATED WORK

Multivariate cooperative relations exist widely in the academic society. With the increase of academic data and the fusion of information such as papers, journals, conferences, scholars and institutions, the research on the multivariate cooperative relationships in academic networks has data support [24]. In recent years, the work published in the world's top journals such as Nature, Science and PNAS has shown the academic community's emphasis on the study of multivariate cooperative relations. Since the relationship must exist between two or more entities, the relationship used to describe the relationship between two entities is called a binary relationship. Similarly, used to describe the relationship between three entities is called a ternary relationship. Used to describe the relationship between n entities is called n-ary relationship, and n-ary relationship is also collectively called multivariate relationship [26].

The binary relationship is the most popular studied object in the field of network science in the past few decades. Through the abstraction of the real world, the objects in many systems can be simplified into binary relationship data for modeling and visualization. However, application fields such as social networks, transportation networks, and biomedical networks have brought abundant and ever-increasing amounts of complex data, and the relationships between objects have become more complicated. Traditional binary relationships cannot describe these intricate relationships. In the actual system modeling, it is often necessary to reflect the multivariate relationship characteristics of the system [25]. For example, in a multi-player game, nodes are used to represent players. If the binary relationship is used to represent the association between players, it can only represent the relationship between two people in a game, but can not describe the relationship among all players. Therefore, more and more studies have discovered the importance of multivariate relationships. Multivariate relationships can better and more naturally express the internal connections and patterns hidden in information [27].

The advisor-advisee relationship and the citation relationship are common binary relationships in academic networks [9]; the ternary relationship is more complex, such as conference closure ("scholar-meeting-scholar" refers to two scholars participating in the conference); multivariate relationships include teamwork relationships, colleague relationships, and so on. Generally speaking, multivariate relationships are very common in academic society, and have more practical significance and research value, but they are often heterogeneous and more complex [7, 11, 12]. Currently, the mining and analysis of multivariate cooperative relationships in academic society are mainly based on the following three types of methods: statistical analysis, relational algebra, and network science.

• Statistical Analysis

Statistical analysis, as the earliest research method of academic multi-relation mining and analysis, was based on questionnaires in the early stage. As the scale of data increases and technology advances, the statistical characteristics of large-scale academic data are gradually revealed. Various technical methods in statistical analysis, such as data collection, data cleaning, data storage, data analysis, are gradually used in the mining and analysis of multivariate relationships in academic society. Commonly used statistical analysis methods include hypothesis testing, correlation analysis, regression analysis, factor analysis, parameter estimation, and so on. Statistical analysis methods can be used to mine and analyze a variety of relationships, including cooperative relationships, co-citation relationships, and advisor-advisee relationships. Wang et al. [21] analyzed the cooperative relationship among researchers with different academic ages and found that researchers with similar academic ages cooperate more closely.

• Relational Algebra

Relation is considered to be the connection between entities, which is used to describe the behavior between two or more entities in a specific time. The algebraic constructions of social network structure make relational algebra the main natural framework in algebraic network analysis [13]. Boorman et al. [1] gave an algebraic expression for the individual structure embedded in social networks based on algebraic semigroups. Since then, a series of works have been developed, and relational algebra has also been widely used in relationship mining.

• Network Science

As an important application field and research branch of network science, the objects of relationship mining are diverse. Zhang et al. [5] studied their social cooperative relations based on the scholars' cooperative network. Combined with actual data, the study found that the co-authoring network is a self-organizing scale-free network. Based on network statistics, it is proved that the scientific research cooperation network has the characteristics of a small-world network. Based on the evolution process of academic cooperation network, Zhu [29] proposed a model about entropy theory to analyze the dynamic network evolution. Liu et al. [28] analyzed the cooperative network of highly cited scholars and found that there are many forms of cooperative relations and groups. Cross-institutional or transnational cooperation has become the mainstream trend of current scientific research.

In recent years, researchers have paid more and more attention to the mining of multiple relational data. Mukhopadhyay et al. [11] proposed a method to mine multivariate relationships on the Medline online database and gave the results of lung cancer mining. The result contains 168 ternary relationships, including the relationships among biological entities such as diseases, organs, drugs, chemicals, and pathways, and each relationship has a weight. In addition, some researchers apply multivariate relationships to the fields related to knowledge systems, such as modeling multivariate relationships in article-article networks, article-user networks, and article-topic word networks.

3 PROBLEM DEFINITION AND EXPERIMENTAL DESIGN

This section first introduces the main symbols and their definitions involved in the process of multivariate relationships mining, and formalizes the problem. Then, we describe the experimental procedure.

3.1 Problem Definition

Since we are mainly concerned with the multiple advisor-advisee relationships in the cooperation network, we first need to build the collaboration network $G = (V, E)$. The node $v_i \in V$ represents the author i, each edge in E represents a cooperative relationship between nodes, and C_i represents the set of collaborators of the author i.

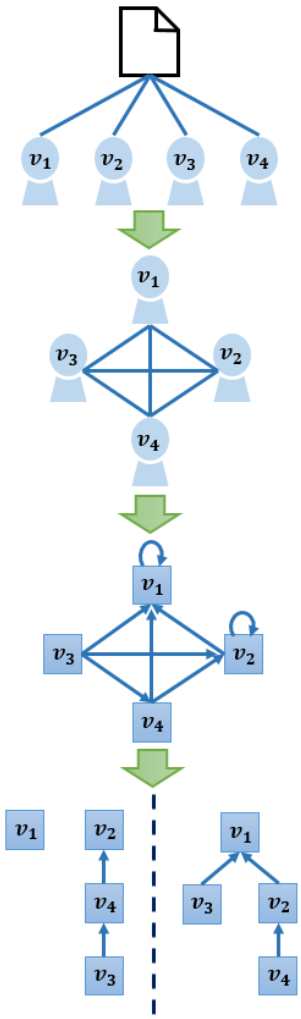


Figure 1: An illustration of the problem definition

Figure 1 shows a simple example of problem analysis. Through the publication information of the paper, the author's cooperation network is obtained. As mentioned above, each node represents an author, and each directed edge points to his potential advisor. If the arrow points to the node itself, it means that it might be the root node. The cooperative network constructed in this paper is a directed acyclic graph (DAG). If it is not a DAG, other efforts are needed [2]. We use the author's earliest paper publication time to determine the direction of the edge. After obtaining the directed candidate graph, the model is trained through labeled data to predict other unlabeled edges.

The possible multiple advisor-advisee relationships shown at the bottom of Figure 1.

It has been observed that both network topology and author attributes are important in academic relationship mining [17, 20]. Next, we describe the author and network attributes in detail.

3.1.1 *Node Attributes.*

Node attributes represent the inherent attributes of each node itself. We mainly select node attributes for author nodes, including institution, academic age, number of published papers, and so on. These attributes can be used to measure the academic influence of the author and the similarity with other nodes. The node features we selected are shown in Table 1.

Table 1: Description of node features

Notation	Description
aa_i	the academic age of the author i
aa_j	the academic age of the author j
org_i	the organization of the author i
org_j	the organization of the author j
np_i	the number of publications of i before collaborating with j
np_j	the number of publications of j before collaborating with i

3.1.2 *Edge Attributes.*

The edge describes the collaboration intensity between two authors, including cooperation times, cooperation duration, and so on. The selected edge features are shown in Table 2.

Table 2: Description of edge features

Notation	Description
ad_{ij}	difference of academic age between i and j
ct_{ij}	the collaboration times of i and j
cd_{ij}	the collaboration duration of i and j
ft	the number of i and j being the first two authors
lf	the number of i and j being the first and the last authors
$kulc_{ij}^t$	the collaboration similarity between i and j

$kulc_{ij}^t$ is the collaboration similarity between author i and collaborator j after t years since their first collaboration [22].

3.1.3 *Network Attributes.*

The random walk method is used to sample the collaboration network structure to obtain the structural attributes of the network [19, 23]. The dimension of the output vector is a hyperparameter.

Based on the above parameters, we can describe the advisor-advisee relationship mining problem as follows:
Input: An author i with the node, edge, and network attributes, the collaboration network G, and C_i.
Output: i's advisees.

3.2 Experimental Design

For the mining of implicit multiple relationships in academic networks, the design of the experimental process is as follows:

• (1) Crawl the correct advisor-advisee relationship.
Crawl the tree-like advisor-advisee relationship listed on The Academic Family Tree website. The Academic Family Tree is an online database compiled by registered users, which saves the academic genealogy in different fields. However, there is no one-to-one correspondence between the names of AFT and those in the MAG data set. It is necessary to judge whether the person is in the MAG data set according to other conditions, and deal with duplicate names.

The first step of the experiment is shown in Figure 2.

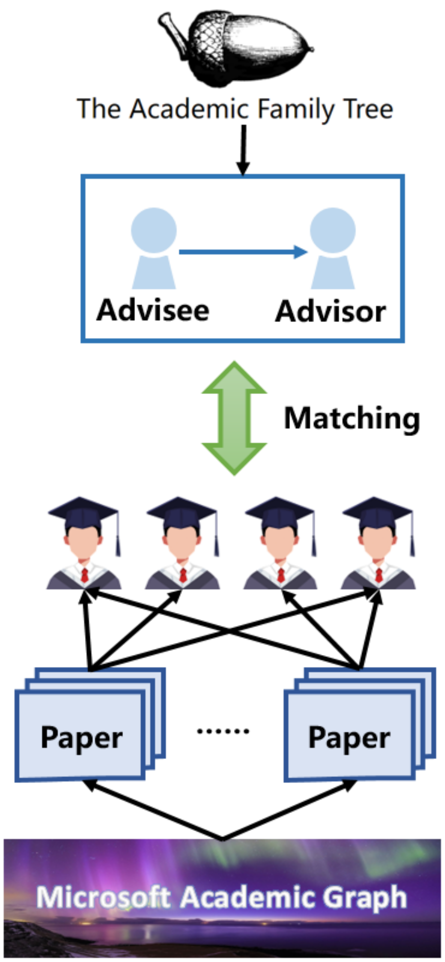


Figure 2: Experimental procedure of step(1)

• (2) Construct collaboration network through the Microsoft Academic Graph Dataset (MAG).
This step is to find the published papers through the person's name, and then put all the authors into a collection to get the collaborator set of this person. The graph obtained at this time is undirected, but the advisor-advisee relationship is directed.

• (3) Construct the candidate directed graph and the feature vector.
The third step is to find the earliest publication time of each author. Determine the direction of the relationship through the earliest publication time. Advisor's earliest paper is usually published earlier than advisee. Then we crawl the required node attributes and edge attributes in the data set. Network feature vectors are generated by node2vec [4], according to the network structure.

The idea of node2vec is the same as that of DeepWalk [14]: the random walk is used to sample the combination (node, context), and then the method of processing word vector is used to model such combination to obtain the representation of network nodes [15]. However, node2vec improved the random walk generation method in DeepWalk, so that the generated random walk can reflect the characteristics of depth-first and breadth-first sampling, and improve the effect of network embedding.

• (4) Use the Capsule Network [16] to classify and output multivariate relations.
Capsule Network is proposed because CNN ignores the relationship between features when extracting features [30]. For example, in face recognition, if there are features such as nose, eyes, and mouth, the CNN model usually activates neurons to determine that it is a human face. However, it does not consider low-level features and the orientation, position and size of the higher-level features. Each neuron of the Capsule Network is no longer a scalar, but a vector. Each vector implicitly contains some information, such as probability, direction, size. Just like the relationship between the eyes and the face, the size of the eyes is 1/10 of the face, and the direction is consistent with the face. Only when these features of eyes satisfy this corresponding relationship with the face to be predicted, they can make a contribution to the prediction.

In fact, the application of the Capsule Network in Natural Language Processing is more natural than in Computer Vision. It uses vectors to represent a feature, which makes the feature expression more abundant. This corresponds to the method of using word vectors instead of one hot to represent a word in NLP.

The dynamic routing algorithm flow of the Capsule Network proposed by Zhao et al. [30] is as follows:

Algorithm 1: Dynamic Routing Algorithm

procedure *ROUTING*($\hat{u}_{j|i}, \hat{a}_{j|i}, r, l$)
Initialize the logits of coupling coefficients $b_{j|i} = 0$
for r *iterations* **do**
  for all capsule i in layer l and capsule j in layer $l+1$:
    $c_{j|i} = \hat{a}_{j|i} \cdot leaky - softmax(b_{j|i})$
  for all capsule j in layer $l+1$:
    $v_j = g(\sum_i c_{j|i}\hat{u}_{j|i}), a_j = |v_j|$
  for all capsule i in layer l and capsule j in layer $l+1$:
    $b_{j|i} = b_{j|i} + \hat{u}_{j|i} \cdot v_j$
end
return v_j, a_j

Other experimental procedures are shown in the figure below.

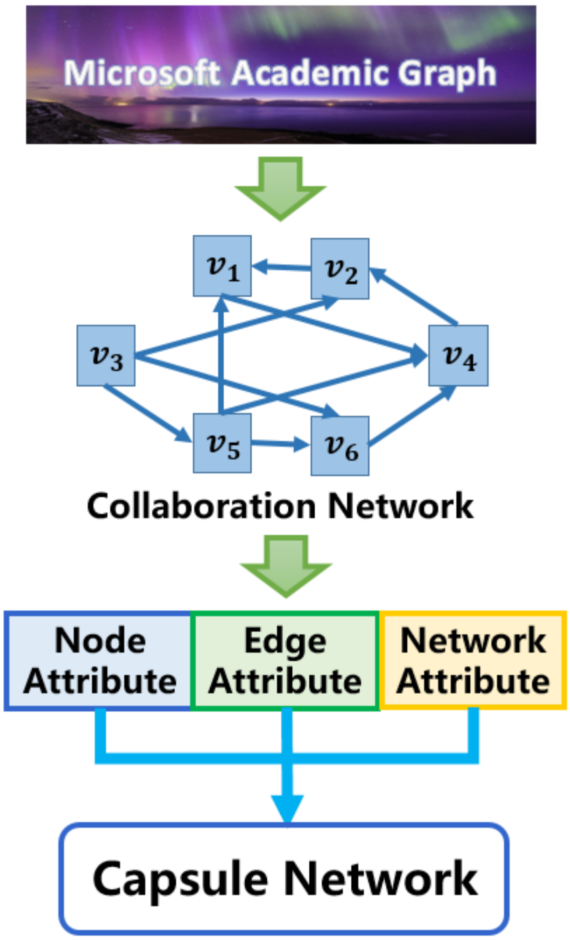


Figure 3: Other experimental procedures

4 EXPERIMENTAL RESULTS AND ANALYSIS

This section describes the experimental details, including experimental dataset, data preprocessing method, and evaluation metrics. In order to ensure the effectiveness of the model, we compare the results with other algorithms and verify the accuracy of the mined multivariate relationship through evaluation metrics.

4.1 Experimental Setup

First of all, it is necessary to solve the problem of duplicate names. The names in the MAG data set cannot correspond to the names in the Academic Family Tree (AFT). We believe that authors with the identical name are the same person, if one of the following conditions is satisfied:
(1) They have at least one same institution.
(2) They have at least one co-author.

Then build a dictionary corresponding to the author names in the AFT and the MAG. The form of each dictionary element is {"the author's name in AFT": "the author's name in the MAG"}.

As mentioned earlier, we extract the advisor-advisee relationship from the AFT as the ground truth. We start from a node, crawl the advisees of the node, add these advisees to the list to be crawled, and cycle this operation, adjust the amount of crawled data by controlling the depth. We crawled 5894 pieces of data from computer science and 4682 pieces of data from neuroscience.

In the experiment, we used 150-dimensional network feature vectors and a fixed learning rate of $1e-3$ to train the model. We use 3 iterations of routing for all data sets because it can optimize the loss faster and eventually converge to a lower loss.

4.2 Baseline Methods

In experiments, we evaluated and compared our model with several widely used baseline methods.
Methods based on machine learning include:
• Logistic Regression (LR) [31]: using maximum likelihood function to estimate and learn parameters, which is commonly used in classification problems.
• Support Vector Machine (SVM) [18]: this is a commonly used binary classifier, which learns the decision boundary by solving the maximum margin hyperplane for the input samples.
Methods based on neural network include:
• Convolutional Neural Network (CNN) [6]: neural networks have simple judgment capabilities like humans, and can give better results in text classification.
• Long Short-Term Memory (LSTM) [3]: LSTM is a special type of RNN, and three gates control the information in the entire time series. Different from the way that CNN extracts features in space, LSTM extracts features in time.

4.3 Evaluation Metrics

In order to quantitatively evaluate the effect of the proposed model, the following metrics are used to evaluate the prediction effect: accuracy, precision, recall, and F1-Score. If there are m positive samples and n negative samples in the test set, after using the model to predict, TP represents the number of correct matches (a positive sample is correctly judged as a positive example), and FP represents a false positive that a negative sample is incorrectly judged as a positive example. TN represents the number of correct non-matches in which a negative sample is correctly judged as a negative example, and FN represents a false negative in which a positive sample is incorrectly judged as a negative example.
• **Accuracy**: the proportion of samples that are correctly classified among all samples.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

• **Precision**: the proportion of true positive samples in all samples judged as positive examples.

$$Precision = \frac{TP}{TP + FP}$$

• **Recall**: the proportion of correctly classified positive examples in all positive samples.

$$Recall = \frac{TP}{TP + FN}$$

• **F1-Score**: calculating the harmonic average of precision and recall.

$$F1-Score = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

4.4 Results and Analysis

Table 3: Advisor-advisee relationship identification performance of different methods

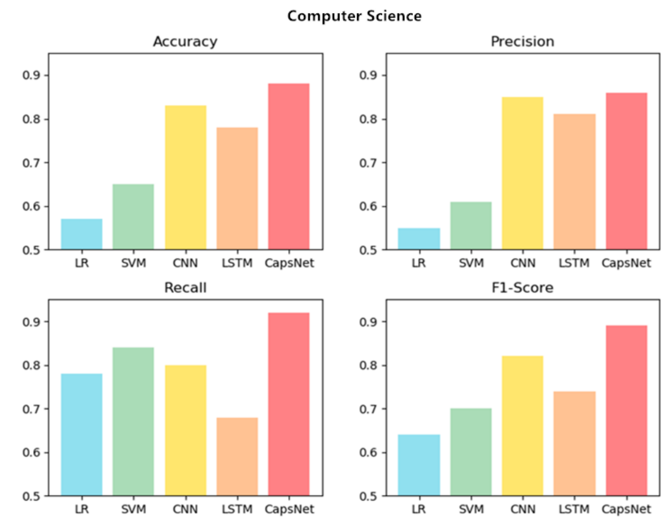
Metrics \ Method	LR	SVM	CNN	LSTM	CapsNet
	Computer Science				
Accuracy	0.57	0.65	0.83	0.78	0.88
Precision	0.55	0.61	0.85	0.81	0.86
Recall	0.78	0.84	0.8	0.68	0.92
F1-Score	0.64	0.7	0.82	0.74	0.89
	Neuroscience				
Metrics \ Method	LR	SVM	CNN	LSTM	CapsNet
Accuracy	0.53	0.66	0.77	0.86	0.91
Precision	0.53	0.63	0.82	0.74	0.89
Recall	0.6	0.79	0.85	0.79	0.95
F1-Score	0.54	0.7	0.83	0.76	0.92

The data in Table 3 proves that the Capsule Network performs well on data in different fields. By comparing with different models, the effectiveness of the Capsule Network is demonstrated. At the same time, for data of different disciplines, the performance shown by the Capsule Network is also different.

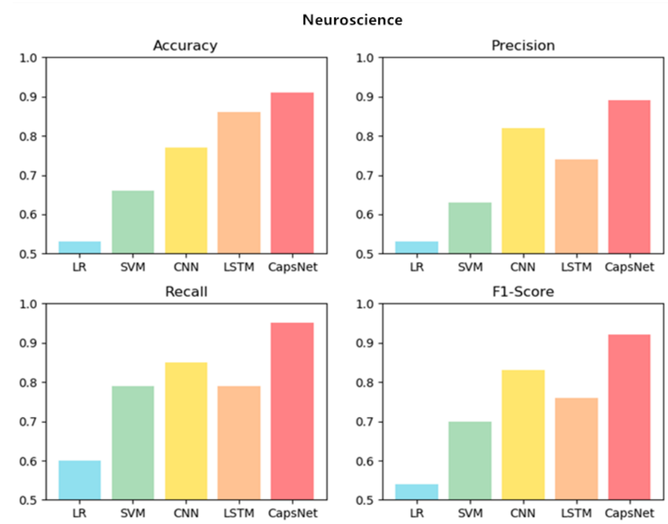
The Capsule Network's ability to learn features is stronger than other methods since it constantly filter features, and key features can be rapidly strengthened in the dynamic routing process. And Capsule Network can correctly classify low-level features to produce high-level features. CNN also has feature enhancements, but it is blind exploration and its efficiency is relatively low. LSTM is not as good as CNN, probably because these features are not related to time series.

Among all the comparison methods, the methods based on neural network is better than the methods based on machine learning, which shows that neural network can effectively identify data features in experiments and better fit data.

In order to show the data more directly, we drew a histogram, as shown in Figure 4.



(a) Experimental results in Computer Science



(b) Experimental results in Neuroscience

Figure 4: Experimental results of different methods

4.5 Parameter Sensitivity

We analyze the influence of parameter setting on the experimental results from the following aspects: (1) The dimension of network feature vector; (2) The number of routing; (3) Vector normalization.

• **The dimension of network feature vector**

In order to verify the influence of the dimension of network feature vector on the experimental results, we set the dimension to 10, 43, 75, 108, 140.

We can observe the influence of the dimension of the network feature vector on the experimental results through the data in Figure 5. Whether using low-dimensional vectors or high-dimensional vectors, Capsule Network can achieve good results. However, for the data of different disciplines, the performance of CapsNet is different.

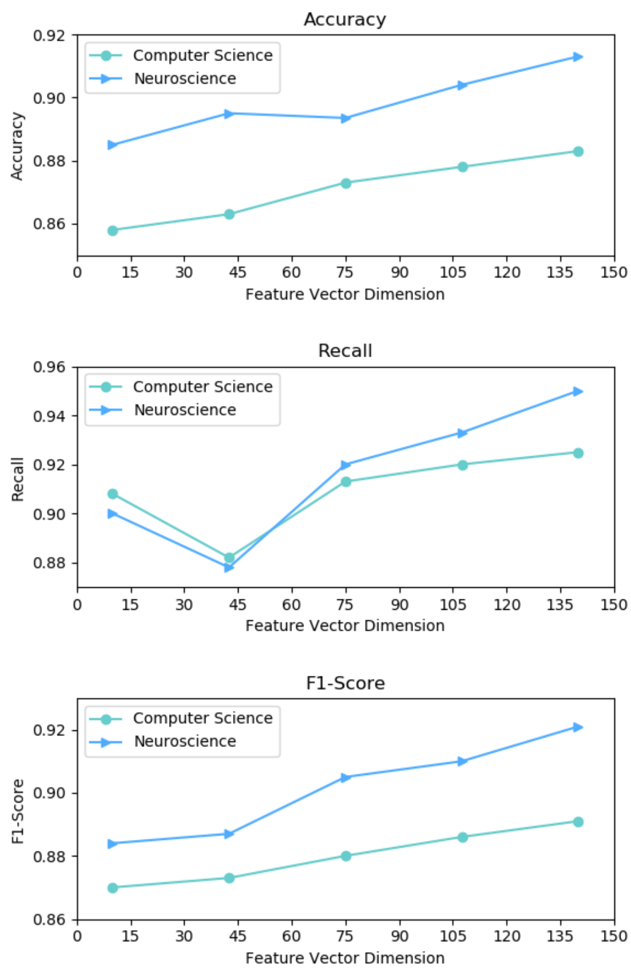


Figure 5: The influence of network feature vector dimension

• The number of routing and vector normalization

We set up different control groups to verify the influence of routing and normalization on the results of the experiment.

Table 4 reports the experimental results of different CapsNet settings and shows the importance of routing and normalization.

The data of network attributes is between 0 and 1. Compared with that, the values of node attributes and edge attributes are much larger. The results show that it is necessary to normalize the input vector.

When the number of routing iterations is small, the accuracy and F1-Score show a linear upward trend with the increase of the number of iterations. But after three iterations, the increase rate of accuracy gradually slowed down. And we observe that more routing iterations increase network capacity and the training time, and tend to overfit the training data set.

• Dynamic learning rate

The learning rate is an important hyperparameter for the rapid convergence of neural networks in the training process. If the learning rate is large, the optimization process may stop at a local minimum or diverge. But the low learning rate will lead to very slow convergence. Although CapsNet has a higher accuracy rate than CNN, it also requires longer training time because of more parameters that need to be calculated. In order to solve this problem, we use the warm restarts technique to dynamically adjust the learning rate and successfully reduce the training time.

Table 4: The impact of routing times and normalization

Computer Science				
Method	Routing	Normalization	Accuracy	F1-Score
CapsNet	1	no	0.81	0.81
CapsNet	1	yes	0.84	0.83
CapsNet	3	no	0.84	0.85
CapsNet	3	yes	0.88	0.88
Neuroscience				
Method	Routing	Normalization	Accuracy	F1-Score
CapsNet	1	no	0.83	0.84
CapsNet	1	yes	0.85	0.87
CapsNet	3	no	0.87	0.88
CapsNet	3	yes	0.91	0.92

In the Stochastic Gradient Descent with Warm Restarts [10] (aka warm restarts), the learning rate is initialized to the maximum value, and then reduced by cosine annealing until the lower limit of the selected interval is reached. When the learning rate reaches the minimum value, set it to the maximum value again to realize the step function.

The learning rate of warm restarts is shown in the figure below.

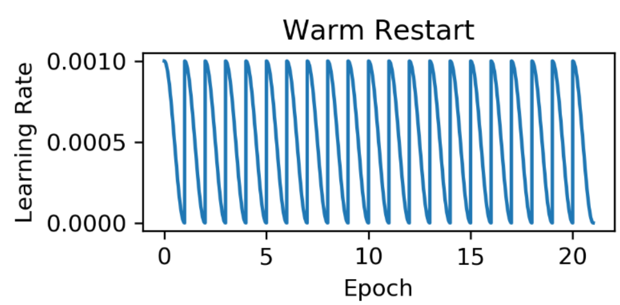


Figure 6: Learning rate of Warm Restarts

The cosine annealing function is given in the following equation, where lr_{min} and lr_{max} is the boundary of the learning rate range, ts is the training step, and T_i is the number of training steps in each cycle. When $ts = T_i$, ts is set to 0, and the cycle starts again. The process repeats cyclically during the training time, and the cycle period needs to be set appropriately to optimize the training time and accuracy.

$$lr = lr_{min} + \frac{1}{2}(lr_{max} - lr_{min})(1 + \cos \pi \cdot \frac{ts}{T_i})$$

Changing a small learning rate to a large learning rate in a short period may cause the training loss suddenly increase. Therefore, warm restart usually causes the model to diverge. But the results show that by adding some controllable divergence, the model can bypass the local minimum and find a better global minimum. This is similar to finding a valley, then climbing a nearby hill and finding a deeper valley. In this process, the same global minimum can be found faster because its path generally has a higher gradient.

The algorithm procedure used in the experiment is shown in Algorithm 2.

Algorithm 2: Warm Restarts for CapsNet

▷ WR stands for Warm Restarts.
procedure $WR(lr_{min}, lr_{max}, T_{curr}, T_i)$
  ▷ Learning rate update
  $lr \leftarrow lr_{min} + \frac{1}{2}(lr_{max} - lr_{min})(1 + \cos \pi \frac{T_{curr}}{T_i})$
  if $T_{curr} = T_i$ **then**
    ▷ Warm Restart after T_i training steps
    $T_{curr} \leftarrow 0$
  else
    ▷ Current step update
    $T_{curr} \leftarrow T_{curr} + 1$
  return T_{curr}

The experimental results are shown in Figure 7. Warm Restarts reached convergence after 10-11 training epoches, and BaselineCapsNet needed about 15 epoches.

Compared with BaselineCapsNet, using Warm Restarts can reduce about 4 training epochs. The amount of data used in this experiment is relatively small, and perhaps the acceleration effect of Warm Restarts is more obvious for larger data sets. However, Warm Restarts has little effect on the improvement of accuracy.

4.6 Application

We applied the trained model to the entire MAG database to construct an academic genealogy of neuroscience, but authors who published less than ten papers were automatically excluded. And we plan to further analyze the social relationships among them based on the current academic genealogy.

5 CONCLUSION

In this work, we use the Capsule Network to extract advisor-advisee relationships based on node, edge, and network attributes. Unlike the traditional methods that only focus on binary relations, we obtain multivariate relations by aggregating binary relations. Through comparison with other methods, the effectiveness of the Capsule Network is proved. By adjusting the parameters several times, the influence of the dimension of the network feature vector, the number of routing, and normalization on the experimental results are found. We also proved the effectiveness of Warm Restarts technique for accelerating the CapsNet training process. Finally, the trained model is applied to the entire MAG dataset to generate academic genealogy.

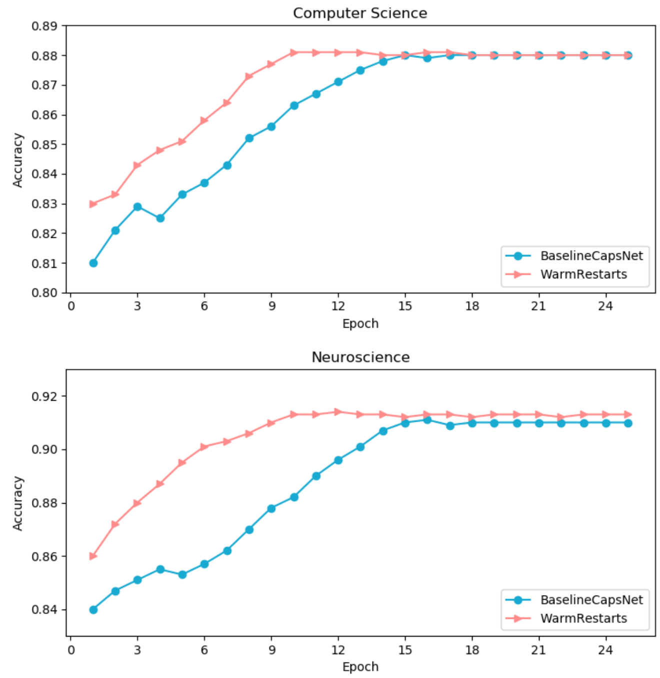


Figure 7: Accuracy results obtained with BaselineCapsNet and WarmRestarts

The following topics may be good choices for future work in this research field.
(1) Based on the existing data to identify other relationships in the collaboration network, such as friends and colleagues.
(2) Further optimize the current relationship mining model.
(3) The data source is relatively single, and author information in other databases can be added to improve accuracy.

ACKNOWLEDGMENTS

This work is partially supported by the Natural Science Foundation of China under Grant 62072073, Grant 61906028, and partially by the Fundamental Research Funds for the Central Universities under Grant DUT20RC(4)010.

REFERENCES

[1] Scott A Boorman and Harrison C White. 1976. Social structure from multiple networks. II. Role structures. *American journal of sociology* 81, 6 (1976), 1384–1446.
[2] W. Chi, J. Han, L. Qi, L. Xiang, W. Lin, and H. Ji. 2012. Learning Hierarchical Relationships among Partially Ordered Objects with Heterogeneous Attributes and Links. In *Proceedings of the 2012 SIAM International Conference on Data Mining*.
[3] Klaus Greff, Rupesh K Srivastava, Jan Koutník, Bas R Steunebrink, and Jürgen Schmidhuber. 2016. LSTM: A search space odyssey. *IEEE transactions on neural networks and learning systems* 28, 10 (2016), 2222–2232.
[4] Aditya Grover and Jure Leskovec. 2016. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 855–864.
[5] Zhang Jing, Wang Fusheng, and Yang Hongyong. 2013. Research on Group Emerging Behavior in Social Cooperation Network. *Complex Systems and Complexity Science* 10, 3 (2013), 45–49.

[6] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. 1998. Gradient-based learning applied to document recognition. *Proc. IEEE* 86, 11 (1998), 2278–2324.

[7] Ronghua Liang and Xiaorui Jiang. 2016. Scientific ranking over heterogeneous academic hypernetwork. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 30.

[8] Jiaying Liu, Jing Ren, Wenqing Zheng, Lianhua Chi, Ivan Lee, and Feng Xia. 2020. Web of scholars: A scholar knowledge graph. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2153–2156.

[9] Jiaying Liu, Feng Xia, Lei Wang, Bo Xu, Xiangjie Kong, Hanghang Tong, and Irwin King. 2019. Shifu2: A network representation learning based model for advisor-advisee relationship mining. *IEEE Transactions on Knowledge and Data Engineering* (2019).

[10] Ilya Loshchilov and Frank Hutter. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983* (2016).

[11] Snehasis Mukhopadhyay, Mathew Palakal, and Kalyan Maddu. 2010. Multi-way association extraction and visualization from biological text documents using hyper-graphs: applications to genetic association studies for diseases. *Artificial intelligence in medicine* 49, 3 (2010), 145–154.

[12] Martin Oti and Han G Brunner. 2007. The modular nature of genetic diseases. *Clinical genetics* 71, 1 (2007), 1–11.

[13] Philippa Pattison. 2011. Relation algebras and social networks. *The sage handbook of social network analysis* (2011), 447–458.

[14] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. 701–710.

[15] Yao Rui. 2018. *A Study of Network Featuure Resentation Using Node2Vec Model*. Ph.D. Dissertation. Nanjing University.

[16] Sara Sabour, Nicholas Frosst, and Geoffrey E Hinton. 2017. Dynamic routing between capsules. *arXiv preprint arXiv:1710.09829* (2017).

[17] Ke Sun, Lei Wang, Bo Xu, Wenhong Zhao, Shyh Wei Teng, and Feng Xia. 2020. Network representation learning: From traditional feature learning to deep learning. *IEEE Access* 8 (2020), 205600–205617.

[18] Jak Suykens. 2001. Support Vector Machines: A Nonlinear Modelling and Control Perspective. *European Journal of Control* 7, 2–3 (2001), 311–327.

[19] Lei Wang, Jing Ren, Bo Xu, Jianxin Li, Wei Luo, and Feng Xia. 2020. Model: Motif-based deep feature learning for link prediction. *IEEE Transactions on Computational Social Systems* 7, 2 (2020), 503–516.

[20] Wei Wang, Jiaying Liu, Tao Tang, Suppawong Tuarob, Feng Xia, Zhiguo Gong, and Irwin King. 2020. Attributed collaboration network embedding for academic relationship mining. *ACM Transactions on the Web (TWEB)* 15, 1 (2020), 1–20.

[21] Wei Wang, Shuo Yu, Teshome Megersa Bekele, Xiangjie Kong, and Feng Xia. 2017. Scientific collaboration patterns vary with scholars' academic ages. *Scientometrics* 112, 1 (2017), 329–343.

[22] Tianyi Wu, Yuguo Chen, and Jiawei Han. 2010. Re-examination of interestingness measures in pattern mining: a unified framework. *Data Mining and Knowledge Discovery* 21, 3 (2010), 371–397.

[23] Feng Xia, Ke Sun, Shuo Yu, Abdul Aziz, Liangtian Wan, Shirui Pan, and Huan Liu. 2021. Graph learning: A survey. *IEEE Transactions on Artificial Intelligence* 2, 2 (2021), 109–127.

[24] Feng Xia, Wei Wang, Teshome Megersa Bekele, and Huan Liu. 2017. Big scholarly data: A survey. *IEEE Transactions on Big Data* 3, 1 (2017), 18–35.

[25] Jin Xu, Shuo Yu, Ke Sun, Jing Ren, Ivan Lee, Shirui Pan, and Feng Xia. 2020. Multivariate relations aggregation learning in social networks. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020*. 77–86.

[26] Jun Yu, Dacheng Tao, and Meng Wang. 2012. Adaptive hypergraph learning and its application in image classification. *IEEE Transactions on Image Processing* 21, 7 (2012), 3262–3272.

[27] Shuo Yu, Feng Xia, Chen Zhang, Haoran Wei, Kathleen Keogh, and Honglong Chen. 2021. Familiarity-Based Collaborative Team Recognition in Academic Social Networks. *IEEE Transactions on Computational Social Systems* (2021).

[28] Liu Yun, Bai Xu, Yang Fangjuan, Yan Zhe, Gui Bingxiu, and Yu Meng. 2016. Research on Collaboration of Highly Cited Scientist Papers. *Scientific Research Management* 37, 4 (2016), 328–336.

[29] Zhu Yunxia. 2017. Entropy-based Research on the Evolution Measurement of Scientific Research Cooperation Network. *JOURNAL OF INTELLIGENCE* 36, 4 (2017), 183–188.

[30] Wei Zhao, Jianbo Ye, Min Yang, Zeyang Lei, Suofei Zhang, and Zhou Zhao. 2018. Investigating capsule networks with dynamic routing for text classification. *arXiv preprint arXiv:1804.00538* (2018).

[31] Zhou Zhihua. 2016. Machine Learning. (2016).