Abnormal Event Detection via Heterogeneous Information Network Embedding

Shaohua Fan
Beijing University of Posts and Telecommunications
Beijing, China
fanshaohua92@163.com

Chuan Shi*
Beijing University of Posts and Telecommunications
Beijing, China
shichuan@bupt.edu.cn

Xiao Wang
Beijing University of Posts and Telecommunications
Beijing, China
wangxiao_cv@tju.edu.cn

ABSTRACT

Heteregeneous information networks (HINs) are ubiquitous in the real world, and discovering the abnormal events plays an important role in understanding and analyzing the HIN. The abnormal event usually implies that the number of co-occurrences of entities in a HIN are very rare, so most of the existing works are based on detecting the rare patterns of events. However, we find that the number of co-occurrences of majority entities in events are the same, which brings great challenge to distinguish the normal and abnormal events. Therefore, we argue that considering the heterogeneous information structure only is not sufficient for abnormal event detection and introducing additional valuable information is necessary. In this paper, we propose a novel deep heterogeneous network embedding method which incorporates the entity attributes and second-order structures simultaneously to address this problem. Specifically, we utilize type-aware Multilayer Perceptron (MLP) component to learn the attribute embedding, and adopt the autoencoder framework to learn the second-order aware embedding. Then based on the mixed embeddings, we are able to model the pairwise interactions of different entities, such that the events with small entity compatibilities have large abnormal event score. The experimental results on real world network demonstrate the effectiveness of our proposed method.

KEYWORDS

Heteregeneous Information Network; Abnormal Event Detection; Network Embedding

ACM Reference Format:
Shaohua Fan, Chuan Shi, and Xiao Wang. 2018. Abnormal Event Detection via Heterogeneous Information Network Embedding. In *2018 ACM Conference on Information and Knowledge Management (CIKM'18), October 22–26, 2018, Torino, Italy*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3269206.3269281

*Corresponding author

CIKM '18, October 22–26, 2018, Torino, Italy

ACM ISBN 978-1-4503-6014-2/18/10…$15.00
https://doi.org/10.1145/3269206.3269281

1 INTRODUCTION

Heterogeneous information networks (HINs), consisting of multi-typed entities and links, are ubiquitous in the real world, since various interconnected data in different complex systems can be naturally modeled as HIN [5]. For example, in a collaboration HIN, the types of entities include author and paper, and the links can model the collaboration relations as *author-paper-author*. One important step towards understanding and analyzing the HIN is the abnormal event detection. Abnormal event detection aims to discover the events with different patterns or behaviors from other events in the HIN. Taking the collaboration HIN as an example, it is well known that different authors in a same research area usually work together to publish a paper. However, we can notice that it is also possible for two authors with different research areas to join in a same paper. Discovering such events can make us better understand the relations among different authors and research areas, and further promote a wider research collaboration in the future. Moreover, in a movie recommendation system, a user may always write review on action movies, but he/she suddenly writes a review on an emotion movie. Mining such event enables us to deeply analyze the user's latent interests, and further recommend more diverse movies.

The basic assumption of many previous works is that the abnormal event usually rarely occurs in a HIN, i.e., if the number of co-occurrences of entities in an event are very rare, then this event can be considered as an abnormal event [1, 2]. However, because of the sparsity of HIN, most of the events happen rarely. We analysed on Aminer [7] co-author network, in this network, the links represents the *author-paper-author* relation. We computed the number of co-occurrences of the authors, and we found that at least 60% authors only co-authored once. This fact implies that the links can hardly provide useful information to distinguish the normal and abnormal events. Therefore, we argue that considering the heterogeneous information network structure only is not sufficient for abnormal event detection and introducing additional valuable information is necessary. Besides the network structure, the attribute is another widely used source to describe the entities. It provides the clue to re-define the events, that is, even if two entities have rare relation, if they have similar attributes, their relation still should be normal. High-order structure is also an effective information to alleviate the sparsity issue. Therefore, considering both of the attributes and high-order structure can help us detect events with rare attributes and high-order patterns.

In this paper, we propose a novel deep heterogeneous network embedding method which incorporates the entity attributes and second-order structure simultaneously to detect the abnormal event in HIN. First, due to different entity types have different attribute spaces, we utilize different type-aware Multilayer Perceptron (MLP) components to learn the attribute embeddings. Second, in order to keep computational efficiency, we only select specific type of entities and adopt the autoencoder framework to learn embeddings which encode the second-order information. Finally, based on the mixed embeddings, we are able to model the pairwise interactions of different entities, such that events with small attributes and network structures entity compatibilities may have large abnormal event scores. We conduct extensive experiments, including the quantitative experiment and case study, demonstrate the effectiveness of our proposed method.

2 PRELIMINARIES

Heterogeneous Information Network. A HIN is defined as a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with an entity type mapping function $\phi : \mathcal{V} \to \mathcal{A}$ and a link type mapping function $\varphi : \mathcal{E} \to \mathcal{R}$. $\mathcal{A}$ and $\mathcal{R}$ denote the sets of predefined entity and link types, where $|\mathcal{A}| + |\mathcal{R}| > 2$. In HIN, two entities can be connected via different semantic paths, defined as **meta-paths**. There are multiple specific paths under the meta-path ρ, which is called a **meta-path instance** [6]. In Fig. 1(a), we present an example of HIN. We can see that the HIN contains multiple types of entities connected by links. And meta-path *Author-Paper-Author* (APA) indicates two authors co-author a paper. Then we can get a_i-p_i-a_j and a_j-p_j-a_z instances for the APA meta-path.

Abnormal Event Detection in HIN. Given a HIN, each event (instance) $e = (a_1, \cdots, a_m)$ is collected by specific meta-path, where a_i denotes an entity from the type $\mathcal{A}_i$, and each entity in an event may have the same type with others. Given n events $D = \{e_1, \cdots e_n\}$, abnormal event detection aims to discover the rare pattern events in D.

Because previous work [1, 2] only treats each entity as a categorical value, here, we take Aminer dataset [7] as an example, and we count the number of two authors' collaboration. Specifically, for authors A_1 and A_2, we count the number $N_{1,2}$ of all the meta-path instances A_1-P-A_2, where P denotes all the paper which the two authors have been co-authored. Then for all the authors n, we can get the numbers $\mathcal{N} = \{N_{1,2}, \ldots, N_{n-1,n}\}$. We sort the numbers in $\mathcal{N}$ by ascending order, then we get x-th percentile c, denoting that there are at least x% numbers in $\mathcal{N}$ is equal or smaller than c. As shown in Table 1, we find that at least 60% authors have been co-authored only once. Based on this fact, we think that we need to introduce more extra information to HIN, in order to get more meaningful results.

Table 1: Percentile of APA instances in Aminer

Percentile	10-th	30-th	50-th	60-th	70-th	90-th	99-th
count(c)	1	1	1	1	2	3	11

3 THE PROPOSED MODEL

Our model **AEHE**, which is abbreviated for Abnormal Event detection via Heterogeneous information network Embedding, is shown in Fig.1(b). As can be seen, for the entity attributes of each type, we utilize a specific multi-layer projection to learn the attribute embedding, also, for the second-order structures of each type of entities, we adopt the autoencoder framework to learn the second-order structure preserving embedding. Then we concentrate the embeddings to model the likelihood of events. Next, we will introduce the details.

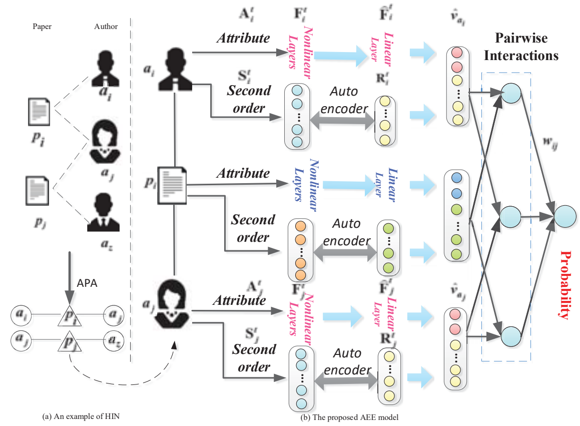

Figure 1: The overall architecture of the proposed model. The same color indicates the shared weights.

3.1 Heterogeneous Attribute Embedding

Each entity in HIN is usually associated with additional attributes, and different types of entities should have different attribute spaces. The main goal of the heterogeneous attribute embedding is to learn the mapping functions to project attributes from different types of entities to a low dimension space, so that the attributes of them become comparable. Specifically, assuming $\mathbf{A}^t$ is an attribute matrix for type-t ($t \in \mathcal{A}$) entities, and i-th row of $\mathbf{A}^t$ is the attribute vector of a_i. We adopt a MLP component to learn its low-dimensional nonlinear d_I^t-dimensional attribute embedding $\mathbf{F}_i^t$ as:

$$\mathbf{F}_i^t = \mathrm{MLP}^t(\mathbf{A}_i^t), \tag{1}$$

where the MLP^t is the nonlinear component for type-t , which consists of two hidden layers with ReLU as activation function. Each typed MLP component can have different structures. As shown in Fig.1(b), same color of components means the shared weights.

3.2 Heterogeneous Second-order Structure Embedding

The second-order Proximity of HIN measures the proximity of two entities with respect to their neighborhood structures, and we use this information to represent entities, then further find abnormal events with rare second-order patterns. Given a heterogeneous network, we can use symmetry meta-path (e.g. APA) to degenerate meta-path instances to links, which connect two authors. Then we get a homogeneous network $G = (V, E)$, which V is the set entities with the same type, E is the set of links connect the nodes in V, and S denotes its adjacency matrix, S_i^t is adjacency vector of entity

a_i^t. We use $\mathbf{S}$ as our input feature and an autoencoder [4] as the model to preserve the neighborhood structures. The autoencoder is composed by an encoder and decoder. The encoder is a non-linear mapping from feature space $\mathbf{S}$ to latent representation space $\mathbf{R}$ and decoder is a non-linear mapping from latent representation $\mathbf{R}$ space back to origin feature space $\hat{\mathbf{S}}$, autoencode step of i-th entity with type-t is shown as follows:

$$\mathbf{R}_i^t = \sigma(\mathbf{W}_1^t * \mathbf{S}_i^t + \mathbf{b}_1^t),\ \hat{\mathbf{S}}_i^t = \sigma(\hat{\mathbf{W}}_1^t * \mathbf{R}_i^t + \hat{\mathbf{b}}_1^t), \tag{2}$$

where σ is the ReLU function, $\mathbf{R}_i^t$ is latent representation of entity a_i^t, and $\hat{\mathbf{S}}_i^t$ is reconstruct representation of $\mathbf{S}_i^t$. $\{\mathbf{W}_1^t, \mathbf{b}_1^t\}$ and $\{\hat{\mathbf{W}}_1^t, \hat{\mathbf{b}}_1^t\}$ are the parameters of encode and decode layer.

However, due to the sparsity of the input adjacency matrix $\mathbf{S}$, the number of zero elements is much larger than that of non-zero elements. To encode more valuable information, we set different weights to different elements, and the loss function for all entities in an event is shown as follows:

$$\mathcal{L}_{ae}(e) = \sum_{1\le i\le m} \|\beta * sign(\mathbf{S}_i^t) \odot (\mathbf{S}_i^t - \hat{\mathbf{S}}_i^t)\|, \tag{3}$$

where m is the number of entity in a single event, *sign* is sign function, "$\odot$" means the Hadamard product, and $\beta > 1$ is the weight for reconstructing the non-zero elements. Furthermore, in HIN, the entities often have various types, considering the special characteristics of different types, in our model, each heterogeneous entity type has their own autoencoder model as shown in Fig. 1(b).

3.3 The Probabilistic Model for Event

In order to utilize different type of entities' embeddings to compute the normality of events, we need to project different entities' embeddings to the same dimension. Specifically, we set L as the dimension of mixed embeddings of attribute and second-order embeddings, and set L_R^t as the dimension of $\mathbf{R}_i^t$, and a linear layer is used to transform dimension of attribute embedding as $L_a^t = L - L_R^t$, the linear transformation matrices, denoted by $\mathbf{P_t} \in \mathbb{R}^{d_I^t \times L_a^t}$. The transformed attributes are denoted by $\hat{\mathbf{F}}_i^t$, where we have:

$$\hat{\mathbf{F}}_i^t = \mathbf{P}_t^T \cdot \mathbf{F}_i^t, \tag{4}$$

where $\mathbf{P}_t^T$ is the transpose of $\mathbf{P}_t$. Then, we combine the attribute embedding and second-order structure embedding into a unified representation of the current entity a_i as follows:

$$\hat{\mathbf{v}}_{a_i} = \hat{\mathbf{F}}_i^t \oplus \mathbf{R}_i^t, \tag{5}$$

where "$\oplus$" denotes the vector concatenation operation.

After getting a single event $e = \{a_1, \dots\ a_m\}$ in event space, similar as [2], we model the scoring function by pairwise interactions among embedded entities to quantify its normality:

$$S_\theta(e) = \sum_{i,j:1\le i<j\le m} w_{ij}(\hat{\mathbf{v}}_{a_i} \cdot \hat{\mathbf{v}}_{a_j}), \tag{6}$$

the dot product between a pair of entity embedding $\hat{\mathbf{v}}_{a_i}$ and $\hat{\mathbf{v}}_{a_j}$ encodes the compatibility of two entities co-occurrence in a single event. w_{ij} is the nonnegative weight for pairwise interaction between entity type $\mathcal{A}_i$ and $\mathcal{A}_j$. Then we model the event probability using the following parametric form:

$$P_\theta(e) = \frac{exp(S_\theta(e))}{\Sigma_{e'\in\Omega} exp(S_\theta(e'))}, \tag{7}$$

where θ is the set of parameters, and Ω is the event space. Similar as [2], we generate negative samples e' by randomly replacing one entity in positive sample e, and our objective function is:

$$\mathcal{L}_p = log\sigma(S_\theta(e)) + \sum_{e'} log\sigma(-S_\theta(e')). \tag{8}$$

As can be seen, by maximizing Eq. 8, the first term makes the embeddings $\hat{\mathbf{v}}_{a_i}$ and $\hat{\mathbf{v}}_{a_j}$ from positive sample e closer, so the normal score of the event will become large. The second term forces the embeddings $\hat{\mathbf{v}}'_{a_i}$ and $\hat{\mathbf{v}}'_{a_j}$ from negative samples e' far away from each other, so the normal score will become small.

3.4 Overall Architecture

To preserve the ability of modeling event probability, as well as the reconstruction ability of second-order information , we combine the objectives in Eqs. 3 and 8. For each e and its negative sample e', our model jointly optimizes the objective function is:

$$\mathcal{L} = -\mathcal{L}_p + \alpha[\mathcal{L}_{ae}(e) + \sum_{e'} \mathcal{L}_{ae}(e')] + \eta\mathcal{L}_{reg}, \tag{9}$$

here, two hyper-parameters α and η are used to trade-off different parts. Besides, $\mathcal{L}_{reg}$ is an $L2$-norm regularizer of MLP components in section **3.1** to prevent overfitting. At last, we adopt Adam algorithm to minimize the objective in Eq. 9.

Discussion. Usually, introducing attribute and second-order information both will obtain much better results. However, in order to keep computational efficiency, it is easy to check that our model is very flexible for one to consider only one source (attribute information or high-order structure or both) based on their domain knowledge. Furthermore, when the number of entity types grow, we can easily add corresponding typed attribute and second-order components, then the mixed embeddings are easily to be feed in a pairwise model. All the operation discussed above without changing the objective function Eq.9.

4 EXPERIMENTS

4.1 Datasets

Aminer [7] is an online academic website. In this network, authors collaborate with different people on different topics, and we detect the abnormal co-authored event (APA) which is different from others. We construct co-authored attributed heterogeneous information network which consists of 12755 authors, 13795 papers and 73957 APA instances. For each author, we match its research topic distribution, which is a 200 dimension vector, and each dimension denotes the weight of the topic for the author. For each paper, we extract its abstract, and we use Term Frequency - Inverse Document Frequency (TF-IDF) to represent abstracts. Then we reduce the dimension of topic distribution to 50 and TF-IDF vector to 100 via SVD. In order to evaluate our method, similar to [2], for each event in the testing data, we randomly replace one of entities with other entities of the same type, and make sure the newly generated events do not occur in both training and testing datasets, so that they can be considered more abnormal than the observed events.

4.2 Experimental Settings

Baselines. We choose the following state-of-the-art methods for abnormal event detection and embedding. (1) **Count Number**: this is a straightforward method count the number of co-occurrences of entity pairs in the HIN. The abnormal score of an event can be computed by negative of the sum of the number of its entity pairs’ co-occurrences in our counting results. (2) **metapath2vec [3]**: we first use it to learn the entity embeddings, and then use Eq.6 to get the abnormal scores. (3) **APE**: it is an entity embedding-based abnormal detection method for heterogeneous categorical events. (4) **AEHE**: this is the proposed method. Note that we use the negative of Eq. 6 output as the abnormal score. (5) **AEHE (attribute only)**: this method is our proposed AEHE without using the second-order information.

For baselines, we set as default parameters. For our method, we set the dimension of attribute embeddings to 30, the dimension of second-order embeddings to 5, the learning rate to 0.001, and selected best-performed hyper-parameters α and β on validation sets. We vary the size of the training set from 20% to 80% and the remaining instances are as testing sets. We adopt MAP and AUC [2] for evaluation.

4.3 Experimental Results

Table 2 shows the AUC and MAP of different methods on detecting abnormal event. Note that our method AEHE outperforms all the baselines in terms of both metrics. AEHE achieves 0.020-0.080 improvements in terms of MAP and 0.010-0.054 gains in terms of AUC over APE. For count number and metapath2vec, they do not have training precess, so all the results of different percent training set are same.

Table 2: Performance of abnormal co-author event detection. The number in bold is the best performance.

Metric	Method	20%	40%	60%	80%
MAP	Count Number	0.020	0.020	0.020	0.020
	Metapath2vec	0.911	0.911	0.911	0.911
	APE	0.825	0.919	0.940	0.952
	AEHE (attribute only)	**0.924**	0.940	0.941	0.950
	AEHE	0.915	**0.943**	**0.960**	**0.973**
AUC	Count Number	0.780	0.780	0.780	0.780
	Metapath2vec	0.903	0.903	0.903	0.903
	APE	0.875	0.930	0.951	0.962
	AEHE (attribute only)	0.916	0.934	0.941	0.944
	AEHE	**0.929**	**0.950**	**0.963**	**0.972**

4.4 Case Study

We conduct a case study to demonstrate the effectiveness of introducing attribute information. We first compute the abnormal scores of all the instances, then take the top N instances, and then compute similarity score by the following formula:

$$\mathrm{AS} = \sum_{1 \le i \le N} (\mathbf{A}_{i1}^t \cdot \mathbf{A}_{i2}^t), \tag{10}$$

where $N = 1000$ is the top number, and $\mathbf{A}_{i1}^t$ and $\mathbf{A}_{i2}^t$ denotes attributes of entities z_{i1} and z_{i2} with same type t in event i. The similarity of entities in an abnormal event is usually small. We can find that the AS of APE is 25.66, while ours is 18.37, indicating that our model can detect the event which have more abnormal attribute patterns.

Table 3 shows some abnormal events detected by our model. In the first event, Daniel A. Keim and D. Kokkinakis co-author a paper named “Fingerprint Matrices”, which studies about text processing. Daniel A. Keim’s research areas are Data Mining (DM) and Information Visualization (IV). However, D. Kokkinakis is interested in Computer Networks (CN) and Wireless Communication (WC). And the second event also indicates an interdisciplinary coauthor. From the two cases, we can see that our method can find abnormal events which have rare attribute patterns. These patterns may have further collaboration potential, and promote a wider research collaboration in the future.

Table 3: Detected abnormal events examples

entity	1	2
A_1	Daniel A. Keim	Jan Zizka
	(DM, IV)	(CN, WC)
P	Fingerprint Matrices	SpeckleSense
	(Text Processing)	(Communication Hardware)
A_2	D. Kokkinakis	Ramesh Raskar
	(CN, WC)	(Physical, Digital)

5 CONCLUSIONS

In this paper, we track a challenging problem of abnormal event detection in HIN. Different from previous work only considering the heterogeneous information structure, we propose a novel deep heterogeneous network embedding method which incorporates the entity attributes and second-order structures simultaneously to address this problem. The experimental results on real world network demonstrate the effectiveness of our proposed method.

ACKNOWLEDGMENTS

This work is supported in part by the National Natural Science Foundation of China (No. 61772082, 61702296, 61375058), the National Key Research and Development Program of China (2017YFB0803304), and the Beijing Municipal Natural Science Foundation (4182043).

REFERENCES

[1] Leman Akoglu, Hanghang Tong, Jilles Vreeken, and Christos Faloutsos. 2012. Fast and reliable anomaly detection in categorical data. In *CIKM*. 415–424.
[2] Ting Chen, Lu An Tang, Yizhou Sun, Zhengzhang Chen, and Kai Zhang. 2016. Entity embedding-based anomaly detection for heterogeneous categorical events. In *IJCAI*. 1396–1403.
[3] Yuxiao Dong, Nitesh V Chawla, and Ananthram Swami. 2017. metapath2vec: Scalable Representation Learning for Heterogeneous Networks. In *SIGKDD*. 135–144.
[4] Y Lecun, Y Bengio, and G Hinton. 2015. Deep learning. *Nature* 521, 7553 (2015), 436.
[5] Chuan Shi, Yitong Li, Jiawei Zhang, Yizhou Sun, and Philip S. Yu. 2016. A Survey of Heterogeneous Information Network Analysis. *TKDE* 29, 1 (2016), 17–37.
[6] Yizhou Sun, Jiawei Han, Xifeng Yan, Philip S. Yu, and Tianyi Wu. 2011. PathSim: Meta Path-Based Top-K Similarity Search in Heterogeneous Information Networks. In *VLDB*. 992–1003.
[7] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. ArnetMiner:extraction and mining of academic social networks. In *SIGKDD*. 990–998.