MSIP:Study on multi-source infection pattern mining algorithm in four-dimensional spacetime

Deshun Li[1], Lingyu Li[1], Kaiyi Li[1], Qiuling Yang[1], Jiangyuan Yao[1], Yuyin Tan[2]
[1] School of Computer Science and Technology, Hainan University, Haikou, China
[2] School of Cyberspace Security, Hainan University, Haikou, China
Corresponding author: Yuyin Tan, Email: 990674@hainanu.edu.cn

***Abstract*—The existing contagion patterns lack the mining of dynamic multi-source infection events in real four-dimensional space-time, which reduces the applicability of the current research results. To address the shortcomings and challenges of existing research work, we investigate the multi-source infection pattern mining algorithm for infectious dynamics in four-dimensional space-time. In this paper, the multi-source infection pattern (MSIP) model is first defined in four-dimensional spacetime. The proposed model considers elements that are more accurate to the real environment, including multi-source infection, movement trajectory, and spatio-temporal dimension. The infection pattern mining algorithm based on opportunity (MABO) is proposed under the MSIP model. The result of the experiment demonstrates that MABO performs effectively, which can mine more infected events under multi-source infection conditions.**

***Index Terms*—Four-dimensional spacetime; multi-source infection; infection pattern; mining algorithm**

I. Introduction

Mobile object trajectories can provide accurate data in application fields such as traffic management [1] [2] [3], location prediction [2], advertising [4] [5], and epidemic control [6]. Nevertheless, the exponential growth of trajectory data in recent years has provided a significant barrier for correct data mining in a variety of application domains. During a new crown epidemic, it is impossible to prevent the spread of viruses among numerous vector creatures, where their transmission patterns are highly associated with the movement trajectories of vector organisms. With these data in four-dimensional spacetime, the mining algorithm can effectively mine infected objects. This contributes to the accurate prevention and control of epidemics, which minimizes the cost of social management.

Recently, IMO [6] and MSIE [7] were proposed to describe the infection pattern of mobile objects. The IMO model only examines a one-to-one infection scenario, which cannot describe multi-source infection. This model determines an object is infected by this object and infected source sharing the same continuous period and the spatial distance. Within a given threshold of period and distance, a normal object will be infected by infected source.

In order to overcome the mentioned shortage of IMO [6], MSIE [7] defines a many-to-one infection pattern and proposes the MIPM algorithm to mine infectious events in planar space. The proposed algorithm determines an object is infected by using sliding window, which accurately records the set of infected objects. However, MSIE only examines the multi-source infected pattern for moving objects in a planar space, which cannot reflect the spatio-temporal moving trajectories in four-dimensional space. It is challenging to reproduce infection episodes in real world with the current infection pattern.

In this paper, we first propose the multi-source infection pattern (MSIP) model, which considers the infection pattern and four-dimensional spacetime. The proposed model considers elements in real environment, including multi-source infection, movement trajectory, and spatio-temporal dimension. On consideration of factors above, MSIP is more precise in real word than the existing models in related work. Under the MSIP model, an infection pattern mining algorithm based on opportunity (MABO) is proposed, which determines an object is infected by considering the threshold of period and distance in four-dimensional spacetime. Within a given threshold of period, distance, and length of window of opportunity, the proposed algorithm can accurately mine the infected events in real world. The result of experiment shows that MSIP is comprehensive and can mine more infected events than the existing work.

The contributions of this paper are given as follows:

- MSIP is proposed, which considers multiple infection sources, movement trajectories, and moving objects in four-dimensional spacetime.
- We propose MABO, which mine the infected events based on opportunity window.
- We evaluate the proposed model and algorithm through experiments, which is more precisely than the related work.

The remainder of this paper is organized as follows: Section II defines the model. Section III details the proposed algorithm. Section IV gives the results of the experiment. SectionV concludes the paper.

II. MSIP:Multi-source Infection Pattern

In this section, we discuss the factors of infection in real world, then define the MSIP model.

A. Spatial factors of infection

Infection pattern in real world can be affected by the following three factors: the movement trajectory, infected source, and the spatial and temporal dimensions of the aimed objects. According to recent research, no alternative model has been established that combines such diverse aspects, including multi-source infection, movement trajectories, and the dimensions of spacetime. However, the existing infection models are usually built on one-to-one or many-to-one infection patterns, or mine infected events in planar space. Multi-source infection and the dimension of spacetime have not been considered jointly, which is closer to real environment.

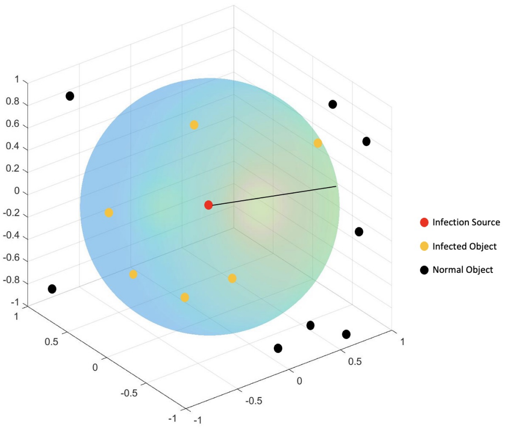

Fig. 1. Factors of infection in spatial dimension

Fig.1 shows the factors of infection in spatial dimension, with a given distance threshold $d = 1.5$. Let s_0 denote infection source. In the two-dimensional plane with the height $z = 0$, o_1 is an infected object infected by s_0. Normal object o_2 is not infected by s_0 with the plane of $z = 0$. By considering the stereo space, o_2 will be infected by s_0, which can reflect infection events in real environment. Actually, within the given consistent time period, all the normal objects in the sphere space will be infected by s_0. The movement trajectory data in four-dimensional spacetime is compatible with infection events mentioned above.

B. Multi-source infection

Moving objects will be exposed to the multiple infection scope of infected sources.This requires us to consider movement trajectories and multi-source infection. Movement trajectories are sequences of locations gathered at a certain period, or an ordered set of spatial point recordings collected on a set of timestamps $t = 1, 2, 3, ...$ The movement trajectory is defined in space and time. It is essential to get the movement trajectories in construction of infection model.

Fig.2 shows infection factor of multiple source and movement trajectories. Let s_1 and s_2 are two different infected

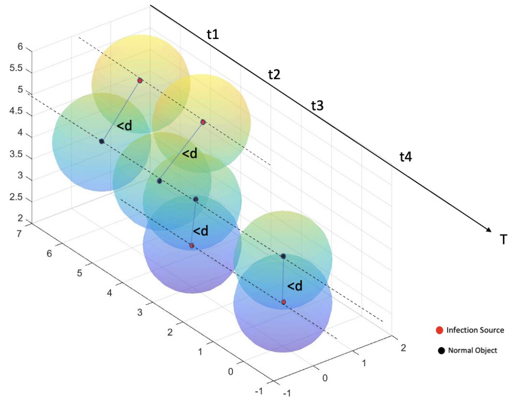

Fig. 2. Infection factor of multiple source and movement trajectories

sources with their movement trajectories from t_1 to t_5. Let o_1 and o_2 denote objects which share the same spacetime with s_1 and s_2 in observation. In the consecutive time period t_3 to t_5, the distance between o_1 and s_1 is less than the infection distance d. According to the model IMO, o_1 will be infected by s_1, which is referred to as one-to-one infection. Based on the rule of infection, o_2 will not be infected by s_1 or s_2 respectively, even though the distances are less than d. However, from t_1 to t_5, o_2 can be infected because it locates in the infection scope of s_1 and s_2 jointly. Under the circumstance, o_2 is infected in the many-to-one pattern.

Ignoring the multi-infection source infection event is inconsistent with the actual situation, which will have a significant impact on the model's accuracy. Obviously, the many-to-one mining pattern performs better than the one-to-one mining pattern in real environment.

From the perspective of spacetime, the existing work ignore the altitude coordinate in movement trajectories of objects, which can cause misjudgment in mining infection event. Let xy denote the planar coordinate of an object. Let z denote the altitude coordinate of an object. For example, within the infection distance d in xy coordinate, an object will not be infected if it is out of the infection distance in xyz coordinate. Obviously, obtaining and calculating the height can make the mining infection events more complex. With the development of positioning, sensing and computation, it is simple to access and process movement trajectories with z coordinate in four-dimensional spacetime. This can significantly enhance the precision and efficiency of mining infection events. In this paper, we consider the factor of altitude coordinate in the analysis of movement trajectories.

C. Definition of MSIP

Let V denote a set of normal objects and MS denote a set of infected sources. For each element i in V and MS, it has movement trajectory tra_i on a time series T. Let t denote the timestamp on T, Δt denote a set of timestamps.

At the above preconditions, we can mine infection events with a given threshold $(\Delta t, d)$.

For convinence, we first define one-to-one infection pattern. Let $i \in V$ denote a normal object, let $j \in MS$ denote an infected source. In one-to-one infection pattern, object i is infected by infected source j when the distance between tra_i and tra_j is less than d at the period Δt. Let tra_i and tra_j denote the movement trajectories of object i and object j respectively. Let $x_{i,k}$, $y_{i,k}$, $z_{i,k}$ denote the x, y, z coordinate of object i at the timestamp t_k, respectively. Let $xyzt_k^i$ denote a record of the movement trajectory of object i at the timestamp t_k. During the observation, $tra_i = \{xyzt_1^i, xyzt_2^i, ..., xyzt_n^i\}$, $tra_j = \{xyzt_1^j, xyzt_2^j, ..., xyzt_n^j\}$. Let $dis(i,j)$ denote the distance of object i and object j. If $dis(i,j) = \sqrt{|x_{i,k} - x_{j,k}|^2 + |y_{i,k} - y_{j,k}|^2 + |z_{i,k} - z_{j,k}|^2} < d$ at the period Δt consistently, object i is infected by object j in one-to-one infection pattern.

Based on one-to-one infection pattern, we define many-to-one infection pattern in four-dimensional spacetime. In many-to-one pattern, object i is infected by a set of infection sources $S_j \subseteq MS$ jointly in time period Δt. Infected object i is consistently in the infection distance of each $j \in S_j$, where the continuous time exceeds the Δt.

Through combing the one-to-one pattern and many-to-one pattern, we obtain the MSIP model. The MSIP is multi-source infection pattern, which considers the two patterns above in four-dimensional spacetime. This model takes multi-source infection, movement trajectory, and spatio-temporal dimension into account, which is more precise to the real environment.

III. INFECTION PATTERN MINING ALGORITHM BASED ON OPPORTUNITY

In this section, we first describe the principle of mining infection pattern, then present an algorithm on opportunity. Finally, we analyze the performance of algorithm.

A. Principle of mining infection pattern

According to the analysis above, the infection event mining in MSIP is compatible with the window of opportunity. The principle of opportunity algorithm is taking advantage of every possible opportunity to realize a goal. In multi-source infection event mining, every infection source makes use of possibilities to infect normal objects. By taking into account movement trajectory, various infection sources, and the spatio-temporal dimension of moving objects, the algorithm based on opportunity achieves its objective of precise mining of infection events.

To facilitate the mining of multi-infection sources, we give the definition of the set of candidate infection sources $C_(i,j,t) \subseteq MS$, which denote all possible infection sources of object i in time period t. $C_{i,j,t}$ is used to store candidate infection source objects, with $C_(i,j,t) = \{j | j \in MS \bigwedge dis(i,j) \leq d\}$ in time period t. The elements of the candidate infection source set must satisfy that they belong to the set MS, where the distance between i and j is not greater than d at time period t.

Based on the planar space, we introduce the elevation dimension to mine multi-source infection events in MSIP model. The random moving of infected sources and normal objects facilitates multi-source infection event mining based on opportunity algorithm. We will give the details of the opportunity algorithm in MSIP.

B. Mining Algorithm Based on Opportunity (MABO)

The proposed MABO is an infection pattern mining algorithm based on opportunity, which considers the threshold of period and distance in four-dimensional spacetime. Within a given threshold of period, distance, and length of window of opportunity, MABO algorithm is shown below.

For convenience, we first describe some variables used in this paper as shown in Table I.

TABLE I
DEFINITIONS OF PARAMETERS

Parameter	Definition
V	A set of normal objects
MS	A set of infected sources
d	Distance threshold
Δt	A set of timestamps
tra	Movement trajectories
$dis(i,j)$	Distance of object i and object j
$C(i,j,t)$	Candidate infection event
S_{dt}	Temporary set of infected objects
E	Infection event
Res	Result of mining

Algorithm 1 Mining Algorithm Based on Opportunity

Input:V,MS,tra,d,Δt
Output: Res
1. Let $Res = \emptyset$, $C(i,j,t) = \emptyset$;
2. In each window of time Δt, calculate the $dis(i,j)$ between $i \in V$ and $j \in MS$;
3. Add j to $C(i,j,t)$ where i located in the range of infection distance;
4. Create an infection event $E = (i, S_{dt}, \Delta t)$;
5. Add E to Res;
6. Repeat step 2-5 until iterating each $i \in V$;
7. Return Res;

In this algorithm, we first initialize the set Res and $C(i,j,t)$ (line 1). Then, for each window of time, we calculate the distance between each uninfected object i and infected source j (line 2). If i fulfill the requirement of infection distance, add j to $C(i,j,t)$ (line 3). Create an infection event $E = (i, S_{dt}, \Delta t)$, which indicates object i is infected by infected source set S_{dt} at time period Δt under four-dimensional spacetime (line 4). Then add E to Res (line 5). Traverse each i in V, repeat step 2-5 (line 6). Finally, return Res set, which contains a set of E (line 7).

C. Analysis of the algorithm

Based on the execution process of the mentioned algorithm, we analyse the time complexity of the algorithm. The algorithm consists of two main parts, moving the window of opportunity and determining the infection event. We give the time complexity analysis process of the two parts respectively.

In the part of moving the window of opportunity, let $n = |V|$, $m = |MS|$, which denote the number of objects in normal object sets and infection objects sets respectively. Let t denote the length of time we observed. Thus, the time complexity is $O(mnt)$ when calculating the dis in each window of opportunity in line 2. Let k denote the average number of infection objects which located in the range of infection distance. The time complexity is $O(nk)$ when adding j to $C(i, j, t)$. In the part of mining infection events, the time complexity is definitely less than $O(mnt)$ and $O(nk)$.

To sum up, the MABO algorithm spends the majority of its time calculating the distance within the moving time windows of opportunity, and the total time complexity of the algorithm is $O(mnt)$. The complexity of the algorithm is polynomial time, which can efficiently mine infection events.

IV. SIMULATION EXPERIMENTS

A. Experiment Setup

This experiment uses the GeoLife GPS Trajectories dataset [8] [9] [10] obtained from the GeoLife project of the Microsoft Asian Research Institute. This dataset contains 182 user trajectories in three years including timestamp, longitude, latitude, and altitude. We retrieved experimental objects from a part of trajectories. In this paper, the mining pattern algorithm MABO is implemented in Matlab. Two aspects are considered during experimental analysis, algorithm mining results and mining efficiency.

The hardware environment of this study is comprised of an AMD R7 processor and 16GB of RAM, while the software environment is a 64-bit version of the Windows 11 operating system.

The default experimental parameters are as follows: the maximum number of moving objects is 150, the minimum distance for determining infection events is $d = 40$, the time period threshold is 16, and the total length of timestamps is 400 in each movement trajectory.

B. Results of Experiments

The Fig.3 depicts the mining results of the MABO algorithm.

As shown in Fig.3, the number of mining infection events increases with the increment of the number of moving objects in three algorithms IMO, MSIE, and MABO. The mining algorithm mines more infection events as the length of timestamps increases. In the result of the three algorithms, the MABO algorithm mines fewer infection events than the MSIE algorithm and more than the IMO algorithm regardless of the number of moving objects.

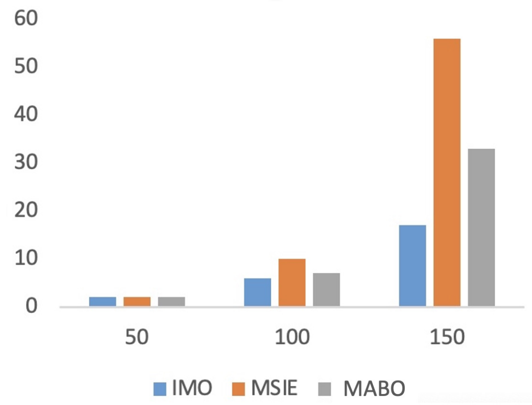

Fig. 3. Result of mining infection events by different algorithms

C. Analysis of the Experiment

According to the result of experiments, the MABO algorithm can successfully mine possible infection events and accurately imitate the real environment. Given the number of moving objects, the number of result sets mined by the MABO algorithm is more than that of the IMO algorithm because the multi-source infection factor is taken into consideration in the model. The results of MABO algorithm is fewer than that of the MSIE algorithm, because the MABO algorithm extends spacetime to a four-dimensional level. The MSIE algorithm ignores the z-coordinate, which possibly results in mining uninfected objects. IMO ignores the multi-source and four-dimensional spacetime, which results in missing some infected objects. Therefore, MABO is more accurate than IMO and MSIE, which facilitates the mining of infection events.

V. CONCLUSION

In order to address the shortcomings and challenges of the current infection pattern, it is necessary to extend mining algorithm in four dimensional spacetime. In this paper, we define multi-source infection pattern (MSIP) model, which comprehensively considers multi-source infection, movement trajectory, and spatio-temporal dimension. We give the definition of one-to-one and many-to-one mining infection pattern, which fully take the dimension of spacetime and multiple sources into consideration. Then, the infection pattern mining algorithm based on opportunity (MABO) is proposed to mine infection events. The proposed algorithm determines an infection event based on the threshold of time and distance in four-dimensional spacetime. MABO is a new and effective algorithm, which contributes to the epidemic control, traffic management, location prediction etc. In further work, we will optimize the mining algorithm to decrease the complexity and increase the accuracy. Besides, in order to facilitate social management, we will apply the model and algorithm to practice in the future.

VI. ACKNOWLEDGMENT

This work was supported by the Hainan Provincial Natural Science Foundation of China (2019RC096, 620RC560, 620RC562), the Scientific Research Setup Fund of Hainan University (KYQD(ZR)1877), the Program of Hainan Association for Science and Technology Plans to Youth RD Innovation (QCXM201910), and the National Natural Science Foundation of China (62162021, 61862020).

REFERENCES

[1] X. Wang, L. T. Yang, X. Chen, M. J. Deen, and J. Jin, "Improved multi-order distributed hosvd with its incremental computing for smart city services," *IEEE Transactions on Sustainable Computing*, vol. 6, no. 3, pp. 456–468, 2021.

[2] K. Collins and G.-M. Muntean, "A vehicle route management solution enabled by wireless vehicular networks," in *IEEE INFOCOM Workshops 2008*, 2008, pp. 1–6.

[3] U. Lee and M. Gerla, "A survey of urban vehicular sensing platforms," *Comput. Netw.*, vol. 54, no. 4, p. 527–544, mar 2010. [Online]. Available: https://doi.org/10.1016/j.comnet.2009.07.011

[4] X. Wang, L. T. Yang, L. Kuang, X. Liu, Q. Zhang, and M. J. Deen, "A tensor-based big-data-driven routing recommendation approach for heterogeneous networks," *IEEE Network*, vol. 33, no. 1, pp. 64–69, 2019.

[5] L. Ren, Z. Meng, X. Wang, L. Zhang, and L. T. Yang, "A data-driven approach of product quality prediction for complex production systems," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 9, pp. 6457–6465, 2021.

[6] J. Xu, H. Lu, and Z. Bao, "Imo: A toolbox for simulating and querying "infected" moving objects," *Proc. VLDB Endow.*, vol. 13, no. 12, p. 2825–2828, sep 2020. [Online]. Available: https://doi.org/10.14778/3415478.3415485

[7] Y. Chen, H. Dai, G. Yang, and Y. Chen, "Multi-source infection pattern mining algorithms over moving objects," in *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2022, pp. 1102–1107.

[8] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining interesting locations and travel sequences from gps trajectories," in *Proceedings of the 18th International Conference on World Wide Web*, ser. WWW '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 791–800. [Online]. Available: https://doi.org/10.1145/1526709.1526816

[9] Y. Zheng, X. Xie, and W.-Y. Ma, "Geolife: A collaborative social networking service among user, location and trajectory," *IEEE Data(base) Engineering Bulletin*, June 2010. [Online]. Available: https://www.microsoft.com/en-us/research/publication/

[10] Y. Zheng, Q. Li, Y. Chen, X. Xie, and W.-Y. Ma, "Understanding mobility based on gps data," in *Proceedings of the 10th International Conference on Ubiquitous Computing*, ser. UbiComp '08. New York, NY, USA: Association for Computing Machinery, 2008, p. 312–321. [Online]. Available: https://doi.org/10.1145/1409635.1409677