THE IMPLEMENTATION OF HESITANT FUZZY SPATIAL CO-LOCATION PATTERN MINING ALGORITHM BASED ON PYTHON

ZhaoWei Liu1, 2, Bo Wei1, 2, *, ChuanLi Kang1, 2, JianWu Jiang1, 2

1 Guangxi Key Laboratory of Spatial Information and Geomatics, Guilin University of Technology, Jiangan Road, Guilin, China - (weect, superweibo)@glut.edu.cn, (275653188, 1126926348@qq.com
2 College of Geomatics and Geoinformation, Guilin University of Technology, Jiangan Road, Guilin, China

Commission VI, WG VI/4

KEY WORDS: Spatial Data Mining, Hesitant Fuzzy Sets, Spatial Co-Location Pattern Mining, Hesitant Fuzzy Spatial Co-Location Pattern Mining, Python, Mining Algorithm, Point of Interest

ABSTRACT:

As one of the important research directions in the spatial data mining, spatial co-location pattern mining aimed at finding the spatial features whose the instances are frequent co-locate in neighbouring domain. With the introduction of fuzzy sets into traditional spatial co-location pattern mining, the research on fuzzy spatial co-location pattern mining has been deepened continuously, which extends traditional spatial co-location pattern mining to deal with fuzzy spatial objects and discover their laws of spatial symbiosis. In this paper, the operation principle of a classical join-based algorithm for mining spatial co-location patterns is briefly described firstly. Then, combining with the definition of classical participation rate and participation degree, a novel hesitant fuzzy spatial co-location pattern mining algorithm is proposed based on the establishment of the hesitant fuzzy participation rate and hesitant fuzzy participation formula according to the characteristics in fusion of hesitant fuzzy set theory, the score function and spatial co-location pattern mining. Finally, the proposed algorithm is written and implemented based on Python language, which uses a NumPy system to the expansion of the open source numerical calculation. The Python program of the proposed algorithm includes the method of computing hesitant fuzzy membership based on score function, the implementation of generating k-order candidate patterns, k-order frequent patterns and k-order table instances. A hesitant fuzzy spatial co-location pattern mining experiment is carried out and the experimental results show that the proposed and implemented algorithm is effective and feasible.

1. INTRODUCTION

Spatial co-location pattern mining is one of the important research directions of spatial data mining, which was first proposed in 2001 (Shekhar, Huang, 2001). Because of the uncertainty of the data, there are more and more scholars begin to devote themselves to the research of co-location mining in uncertain data.

In life, people are often used to describe the spatial relationship between two objects in the form of inaccurate distances such as "far" and "near" and inaccurate directions such as "east", "west", "north" and "south". At the same time, different people have different intuitive judgments in distance, orientation and location. However, in a certain extent, the inaccurate intuitive judgment like this can reflect the position of the object and the relationship between the object and the object. Such description is one of the main research contents of fuzzy mathematics (Xie, 2000).

The concept of fuzzy sets was proposed in 1965 in the first time. Zadeh (1965) used the member functions to express the degree of membership of elements relative to sets. In fact, the fuzzified data is more in line with people's thinking habits and can more intuitively reflect information. After the establishment of the theory of fuzzy sets, it has been rapidly integrated into various disciplines. In order to better reflect the hesitation and uncertainty of the object, scholars in various fields have expanded it in various forms. However, in the actual decision-making process, the decision-makers are often in a state of hesitation, and often wanders in the judgment of several alternative object attributes. And different decision-makers have different numbers of alternative judgments. Therefore, Torro (2010) and Narukawa (2009) defined hesitant fuzzy sets in 2010, which is another extension of fuzzy sets. The hesitant fuzzy set allows the membership degree of each object to have multiple possible fuzzy elements, thus more accurately depicting the embarrassment of the decision-maker's dilemma.

With the introduction of hesitant fuzzy sets, spatial co-location pattern mining methods are greatly enriched. The spatial co-location model based on the hesitant fuzzy set can scientifically solve the problems such as the uncertainty of the location of point space object instances and the possibility of multiple values of point space object instances. This paper takes hesitant fuzzy sets as the basic theory of expressing and processing uncertain data. By combining with spatial co-location pattern mining and introducing the calculation method of the score function (Xia, Xu, 2011), we establish relevant definitions mining strategies and algorithms, and obtain a spatial co-location pattern mining method based on hesitant fuzzy sets.

2. BASIC CONCEPTS

As one of the extensions of co-location pattern mining, the space co-location pattern mining based on hesitation fuzzy set
has similarities and its own characteristics. The data type studied in this paper is point data, that is, spatial object of point type. The spatial objects \(a_1, a_2, b_1\) in Figure 1; \(a_1, a_2, b_1, \text{etc.}\), are defined as point instances, and the category to which a point instance belongs is defined as a feature. That is, A and B are the features of the mined instance objects \(a_1, a_2\) and \(b_1\) respectively, and \(a_1\) and \(a_2\) are the first and second instances of feature A respectively.

An instance of a spatial object is represented by a vector \(<\text{instance ID}, \text{spatial location}, \text{the corresponding spatial object}>\). Then the expression form in hesitant fuzzy sets of location fuzzy instances is \(<\text{instance ID}, \text{space location}, H_i>\), the corresponding spatial object, \(H\) is hesitancy fuzzy set, and \(H = \{<x, h_0(x)> | x \in X\}\), abbreviated as \(<ID, (L, H), O^e>\). When the number of \(h_0(x)\) is 0, the position of the instance is determined. When the number is 1, it is equivalent to the position ambiguity represented by Zadeh fuzzy set. When the number is greater than 1, it indicates that there are multiple possible membership values in the position of the instance. When a spatial object has multiple instances, its position fuzzy instance set is expressed as \(<ID_i, (L, H_i), O^e>(i = 1, 2, ..., n), \ (i = 1, 2, ..., n)\), the \(i\) represents the instance \(i\).

Figure 1 shows A and B, the spatial object feature. When there are three instances of position ambiguity in feature A, \(a_1, a_2\) and \(a_3\) respectively, while feature B has two instances of position ambiguity, \(b_1\) and \(b_2\). The instances in feature A and B represent multiple possible position membership values. For example, the expression \(\{0.7, 0.8\}\) says there are two possible values.

**Figure 1. The spatial object features and the instances**

**Definition 1 (Neighbour Relationship):** Neighbour relationship \(R\) is an input item in data mining, which needs to be defined according to different data domains. For example, it can be defined as adjacent and connected in applying to topological relations, or defined as Euclidean measures in applying to the metric relationships and so on. In this paper, the neighbour relationship is defined as Euclidean distance.

**Definition 2 (Neighbour Collection):** The neighbour collection \(L\) is a set of instances in which any two instances satisfy the neighbour relationship \(R\).

**Definition 3 (Row Instance):** Set a neighbour collection \(L\), when the k-order space co-location mode \(c = \{f_1, f_2, ..., f_k\}\), each feature \(c\) is the feature of the instance in \(L\), and when any subset in \(L\) is no longer a row instance in \(c\), the neighbour collection \(L\) is called a row instance in co-location pattern \(c\). The table instance of \(L\) is the collection of all row instance in \(L\).

**Definition 4 (Score Function):** The hesitant fuzzy set allows the membership degree of each object to have multiple possible fuzzy elements. Therefore, we introduce the score function of hesitant fuzzy elements to calculate the membership degrees of hesitant fuzzy spatial instance, and establish the relevant definition, improves the mining strategy and algorithm.

**Definition 5 (the PR in Score Function based on Hesitant Fuzzy Sets):** Set \(f_i\) as a spatial object and \(HFS_S\_PR(c, f_i)\) represents PR in score function based on hesitation fuzzy sets of \(f_i\) in the k-order space co-location mode \(c\), then it is the ratio between the sum of the score of the non-recurring location fuzzy instance in the table instance of the space co-location mode \(c\) and the total number of instances in \(f_i\). The formula is as follows:

\[
HFS_S\_PR(c, f_i) = \frac{\sum S(a)}{table\_instance\{(f_i)\}}
\]

**Definition 6 (the PI in Score Function based on Hesitant Fuzzy Sets):** Set \(HFS_S\_PI(c)\) as PR in score function based on hesitation fuzzy sets of \(f_i\) in the space co-location mode \(c\) and also named as participation index in score based on hesitation fuzzy sets. The calculation formula is as follows:

\[
HFS_S\_PI(c) = \min\{HFS_S\_PR(c, f_i)\}
\]

The PI in score function based on hesitant fuzzy sets with weak monotonicity, that is, if \(c_i \subseteq c\), then \(HFS_S\_PI (c_i) \geq HFS_S\_PI(c)\). It means that the frequency of co-location mode has a certain priori, and the join-based algorithm can use this property to conduct feature level pruning to reduce the operation amount.

**Definition 7 (Prevalence Co-Location Patterns):** If the PR in score function based on hesitant fuzzy sets of \(f_i\) in the space co-location mode \(c\) is greater than or equal to the threshold value of user-specified input item min\(HFS_S\_PI\), then the candidate co-location mode \(c\) is called as frequent co-location mode. The key to the problem of co-location mining is frequent co-location mining.

**3. ALGORITHM**

The mining algorithm of spatial co-location pattern is mainly based on the reference feature centre model and event centre model (Agrawal, Srikant, 1994, Shekhar, Huang, 2001, Zhang, Mamoulis, Cheung et al., 2004). Compared with the former, the event centre model requires more to form my feature instance. The co-location pattern mining algorithm studied and improved in this paper is join-based algorithm for event centre type.

**3.1 Join and Pruning in Join-based Algorithm**

The join-based algorithm, proposed by Shekhar, S., Huang, Y. and Xiong, H. (2004), is a fully connected co-location pattern mining algorithm. This algorithm replaces transactions with the concept of neighbourhood, and does not need to specify spatial reference features. In other words, it is a non-transactional algorithm that uses neighbourhood relationship to mine co-location pattern. Finding the frequent co-location pattern is

---

The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-3/W10, 2020
International Conference on Geomatics in the Big Data Era (ICGBD), 15–17 November 2019, Guilin, Guangxi, China
primarily a computational process based on joins between table instances.

Figure 2. The steps of join in join-based algorithm

The join steps (Yoo, Shekhar, 2006) in the join-based algorithm (as shown in Figure 2) are as follows: select the co-location mode \( C_k \) and \( C_{k'} \) from the k-order candidate co-location table \( C_k \), and connect them to generate the k-order co-location mode and insert them into the k-order candidate co-location table \( C_{k+1} \). The realization method is as follows:

1. \( ckp\_pattern.append(c[k+1]['pattern'].iloc[i][f_w+1]) \)
   \( ckq\_pattern.append(c[k+1]['pattern'].iloc[i][f_w+2]) \)
   # The row instances \( p \) and \( q \) are proposed from the k-order co-location pattern table instance, to generated the row instances of k+1
2. if distance['distance'][d_p, d_q] == 1:
   \( tt\_co\_lo.loc[i, 'instance'].append([d_p+1, d_q+1]) \)
   # Judge the condition to generate the row instance of

k+1-order co-location mode and insert to \( C_{k+1} \).

After generating an instance of the k+1 candidate co-location pattern, a pruning was carried out through the participation, and a part of the candidate co-location pattern that did not meet the minimum participation threshold in the input item was cut off. Combined with the weak monotony of participation, the feature level pruning of the k+1 order candidate co-location table was carried out with the k-order co-location table, which could improve the efficiency of the algorithm to some extent.

3.2 Spatial Co-location Pattern Mining Algorithm based on Hesitant Fuzzy

In this paper, Xia and Xu's score function for hesitant fuzzy element is introduced to calculate the participation rate and participation degree of hesitation fuzzy score of spatial instance objects. This function is capable of handling complex factors such as highly uncertain information and multiple correlations of the system. Specific implementation methods are as follows:

1. for i in range(0, N):
   membership.loc[i, 'mem'] = EM.iloc[i, 2]  # Recording uncertainty fuzzy membership table
   fuzz = membership.loc[abs(ac)-1, 'mem']
   fuzz1 = recfindall(r"d+", fuzz)
   # Extracting hesitation fuzzy membership degree
   fuzzAllCount = fuzzAllCount + float(fuzz1[0])
   fuzz_member = fuzzAllCount/fuzzCount
   count_all = count_co_all + fuzz_member
   count = f_i_pr.loc[find_f, 'pattern_instance_count']
   pr_data = count_all / count
   # Calculating hesitation fuzzy PR in score function

2. for fu_i in range(0, fuzzCount):
   fuzzAllCount = fuzzAllCount + float(fuzz1[0])
   fuzz_member = fuzzAllCount/fuzzCount
   count_all = count_co_all + fuzz_member
   count = f_i_pr.loc[find_f, 'pattern_instance_count']
   pr_data = count_all / count
   # Calculating hesitation fuzzy PR in score function

The complete algorithm of spatial co-location pattern mining based on hesitant fuzzy set is shown in Figure 3.
4. EXPERIMENT AND RESULT ANALYSIS

The spatial co-location pattern mining algorithm based on hesitation fuzzy set was written on Python and developed on PyCharm. Development and experimental environment: Intel® Core™ i7-8650u CPU @1.90GHz 2.11ghz; 16 GB of memory; Windows 10.

4.1 Experimental Data

The actual data set used in the experiment is the sina weibo user Point of Interest (POI) check-in data, which are in Macao special administrative region. The data set contains 40 spatial object features and 3193 hesitant fuzzy instances, in which each instance contains 0 or more fuzzy elements (0 means that the membership degree of the spatial instance is 1). The data is stored in the excel files, as shown in Table 1 and Table 2.

| ID | Information                          | Longitude  | Latitude  | Name   | Hesitation fuzzy membership |
|----|--------------------------------------|------------|-----------|--------|-----------------------------|
| 1  | Macau international airport(arrive)  | 113.57977  | 22.15473  | Airport | [0.7, 0.75, 0.85, 0.9, 0.9] |
| 2  | Macau international airport(departure) | 113.57977  | 22.15473  | Airport | [0.8, 0.9]                 |
| 3  | Macau heliport                       | 113.56471  | 22.19386  | Airport | [0.9, 0.95, 0.95]          |
| 4  | The bank of China                    | 113.56147  | 22.1486   | Bank   | {}                         |
| 5  | China construction bank (Macau)      | 113.55314  | 22.19913  | Bank   | [0.95]                     |
| ... | ...                                  | ...        | ...       | ...    | ...                         |
| 3192 | Ho li wan performing arts theatre | 113.55254  | 22.20964  | Theatre | [0.6, 0.7, 0.8] |
| 3193 | Dom Pedro V Theatre                 | 113.54325  | 22.18897  | Theatre | [0.5, 0.6, 0.6, 0.7, 0.9] |

Table 1. Attributes of hesitant fuzzy instances

| Instance name (3193) | Feature name (40) |
|----------------------|-------------------|
| Airport              | Airport           |
| Airport              | Bank              |
| Airport              | Bar               |
| Bank                 | Beach             |
| Bank                 | Book shop         |
| Bank                 | Café              |
| Bank                 | Cinema            |
| Bank                 | Convenience Store |
| Bank                 | Cultural Palace   |
| ...                  | ...               |

Table 2. Spatial object feature names

4.2 Operation Results and Data Analysis

The experiment parameter is: the characteristic number $K = 40$, fuzzy instance number $N = 3193$, relationship between neighbouring threshold $d = 500$, minimum participation threshold $\text{min prev} = 0.18$. The results show that the frequent co-location pattern set mined by the spatial co-location pattern mining algorithm based on hesitation fuzzy set on the POI data set of Macao special administrative region is shown in Table 3 and Table 4.

| 3-order frequent patterns |
|---------------------------|
| {Cinema, Mall, Café}      |
| {Cinema, Mall, Parking lot}|
| {Cinema, Mall, Restaurant}|
| {Cinema, Mall, Square}    |
| {Living quarter, School, Convenience Store} |
| {Living quarter, School, Restaurant} |
| {Living quarter, School, Supermarket} |
| {Mall, Hotel, Station}    |
| {Park, Square, Book shop} |
| {Park, Square, Café}      |
| {Park, Square, Parking lot}|
| {Scenic spot, Hotel, Parking lot} |
| ...                       |

Table 3. 3-order frequent patterns

| 4-order frequent patterns |
|---------------------------|
| {Cinema, Mall, Café, Parking lot} |
| {Cinema, Mall, Café, Square} |
| {Living quarter, School, Convenience Store, Supermarket} |
| {Scenic spot, Hotel, Parking lot, Restaurant} |
| ...                       |

Table 4. 4-order frequent patterns

Analysis of mining results of spatial co-location mode mining algorithm based on hesitation fuzzy set on POI data set of Macao special administrative region: for example, the results of 3-order frequent patterns {Cinema, Mall, Parking lot} means that there is an 18% chance of having a cinema, shopping mall and parking lot in the same location in the Macao special administrative region. And there is an 82% chance that a location with both cinema and shopping mall will have parking lot in 500 meters around.

The co-location mode can be used to explore the spatial correlation between different types of urban infrastructure and provide important decision support for urban planning, regional management, commercial layout and other applications.

5. CONCLUSIONS

In this paper, the expression form of space object location based on hesitation fuzzy is obtained by determining the relation between hesitant fuzzy set and spatial co-location mode mining, and the definition and formula of hesitant fuzzy participation rate and participation degree based on score is proposed. Based on the join-based algorithm, a spatial co-location pattern mining algorithm based on hesitation fuzzy sets
is implemented by using python. The feasibility of the algorithm is verified by the experiment.

In order to mine spatial co-location patterns more effectively based on hesitant fuzzy sets, the research will be done on the improvement of the score function, the minimum distance threshold $d$ and the minimum participation threshold $\text{min}_\text{prev}$ in the future.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation of China under grant number 41461085; the Natural Science Foundation of Guangxi Province under grant number 2016GXNSFAA380035; the Foundation of Guangxi Key Laboratory of Spatial Information and Geomatics under grant number 16-380-25-04; the “BaGui Scholars” Special Funds of Guangxi Province under grant number 2019-79; and the Doctoral Foundation of Guilin University of Technology under grant number 1996015.

REFERENCES

Agrawal, R., Srikant, R., 1994: Fast algorithms for mining association rules. The 20th International Conference on Very Large Data Bases, Santiago, Chile.

Huang, Y., Zhang, P., Zhang, C., 2008. On the relationships between clustering and spatial co-location pattern mining. International Journal on Artificial Intelligence Tools, 17(1), 55-70.

Shekhar, S., Huang, Y., 2001. Co-location rules mining: A summary of results. In Proc. Intl. Symposium on Spatio and Temporal Database.

Shekhar, S., Huang, Y., Xiong, H., 2004. Discovering colocation patterns from spatial data sets: a general approach. IEEE Transactions on Knowledge & Data Engineering, 16(12), 1472-1485.

Torra, V., 2010. Hesitant fuzzy sets. International Journal of Intelligent Systems, 25(6), 529-539.

Torra, V., Narukawa, Y., 2009. On hesitant fuzzy sets and decision. IEEE International Conference, 1378-1382.

Xie, J., 2000: Fuzzy mathematics method and its application. Central China University of Science and Engineering.

Xia, M., Xu, Z., 2011. Hesitant fuzzy information aggregation in decision making. International Journal of Approximate Reasoning, 52(3), 395-407.

Yu, C., 2014. Implementation of Join-based algorithm in Matlab. Computer development and application, 27(7), 24-30.

Yoo, J. S., Shekhar S., 2004: A partial join approach for mining co-location patterns. ACM Int’l Symposium on Advances in Geographic Information Systems(ACM-GIS).

Yoo, J. S., Shekhar, S., 2006. A Join-less Approach for mining spatial co-location patterns. TKDE, 18(10), 1323-1337.

Zadeh, L. A., 1965. Fuzzy sets. Information and Control, 8(3), 338-353.

Zhang, X., Mamoulis, N., Cheung, D., Shou, Y., 2004: Fast mining of spatial collocations. Proceedings of the ACM SIGKDD International Conference in Knowledge Discovery and Data Mining, 3842393.